UNIVERSITY
of
GLASGOW

Molecular evolution of *Mannheimia* (*Pasteurella*) *haemolytica*, *Mannheimia glucosida*, and *Pasteurella trehalosi*, and characterization of temperate bacteriophages

A thesis presented for the degree of
Doctor of Philosophy
of the University of Glasgow

by

InKyoung Lee BSc

Institute of Biomedical and Life Sciences
Division of Infection and Immunity
University of Glasgow
Glasgow, UK

March, 2005.

DECLARATION

The thesis is the original work of the author:

I. Lee

DEDICATION

This three year study is dedicated to the wishes of my father, DongYang Lee and my mother, JiHong Kim.

ACKNOWLEDEMENTS

I express my gratitude to all those who support me to complete this thesis. Without them, I could not have come this far. I am deeply indebted to my supervisor Dr Robert Davies for his support, guidance, and excellent supervision and for reading several drafts of this thesis and offering suggestions for improvement. Thanks are also due to my colleagues in the Division of Infection and Immunity.

I would also like to thank Professor HoGun Rhie and Professor JeongChil Yu, Kyung Hee University, for their advice and letters of recommendation. I also want to thank the Overseas Research Students Award Scheme (ORSAS) and the University of Glasgow for providing the funding for this study.

I sincerely thank my family and friends for being so understanding and supportive during this study. Especially, I would like to give my special thanks to my parents KyoHee Lee, JongYea Ji, DongYang Lee and JiHong Kim and my husband JaeBum Lee whose patient love enabled me to complete this study.

PRESENTATIONS/PUBLICATIONS

1. Davies, R. L. & Lee, I. (2004). Sequence diversity and molecular evolution of the heat-modifiable outer membrane protein gene (*ompA*) of *Mannheimia(Pasteurella) haemolytica, Mannheimia glucosida,* and *Pasteurella trehalosi*. *Journal of Bacteriology* **186,** 5741-5752.

2. Davies, R. L. & Lee, I. Evidence for host-directed evolution of the heat-modifiable outer membrane protein (*ompA*) of *Mannheimia haemolytica.* Poster presentation at the 6th International *Pasteurellaceae* Conference, 5th - 10th May 2002, Banff, Canada.

3. Lee, I. & Davies, R. L. Comparative molecular evolution of housekeeping and virulence-associated genes of *Mannheimia haemolytica, Mannheimia glucosida*, and *Pasteurella trehalosi.* Poster presentation at the 6th International *Pasteurellaceae* Conference, 5th - 10th May 2002, Banff, Canada.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	Adenine
attL	Attachment site Left
attR	Attachment site Right
BHI	Brain Heart Infusion
BHIA	Brain Heart Infusion Agar
bp	Base pair(s)
BRD	Bovine respiratory disease
BSA	Bovine serum albumin
C	Cytosine
°C	Degrees Celsius
CO_2	Carbon dioxide
COG	Clusters of orthologous groups
cos	Cohesive end
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytosine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dNTP	2'-deoxynucleoside 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	Ethylenediaminetetra acetic acid
ET	Electrophoretic type
et al.	*et alios* (and others)
F	Forward
G	Guanine
g	Gram(s)
× g	A unit of force equal to the force exerted by gravity
h	Hour(s)
HCl	Hydrochloric acid
ICTV	International Committee for the Taxonomy of Viruses
Ig	Immunoglubulin
IHA	Indirect haemagglutination assay
IS	Insertion sequence
KAc	Potassium acetate
kb	Kilobase(s)
KCl	Potasium chloride
kDa	Kilodalton

kg	Kilogram(s)
LPS	Lipopolysaccharide
M	Molar
mA	Milliampere
Mg	*M. glucosida*
mg	Milligram
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
Mh	*M. haemolytica*
µg	Microgram
µl	Microlitre
min	Minute(s)
ml	Millilitre(s)
MLEE	Multilocus enzyme electrophoresis
mM	Millimolar
mm	Millimetre(s)
mRNA	Messenger RNA
NaCl	Sodium chloride
NCTC	National Collection of Type Cultures
ng	Nanogram(s)
nm	Nanometer(s)
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
PFU	Plaque forming units
pH	A measure of the acidity of a solution
pmol	Picomole
Pt	*P. trehalosi*
R	Reverse
RE	Restriction endonuclease
RecA	Recombinase A
RNA	Ribonucleic acid
RNAse	Ribonuclease
rpm	Revolutions per minute
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate
s	Second(s)
T	Thymidine
Tris	2-amino-2-(hydroxymethyl)1,3-propandiol
tRNA	Transfer RNA
U	Units

UV	Ultraviolet
V	Volts
VIC	Veterinary Investigation Centre
w/v	Weight per volume

SUMMARY

Comparative nucleotide sequence analysis of nineteen genes of different functional classes was carried out in representative isolates of *Mannheimia haemolytica* (10 to 32 isolates), *Mannheimia glucosida* (1 to 6 isolates), and *Pasteurella trehalosi* (1 to 4 isolates). The genes include one DNA repair and recombination gene (*recA*), nine metabolic enzyme genes (*aroA*, *asd*, *galE*, *gap*, *gnd*, *g6pd*, *mdh*, *mtlD*, and *pmm*), one secreted protein gene (*gcp*), four periplasm-associated lipoprotein genes (*plpA*, *plpB*, *plpC*, and *plpD*), and four outer membrane protein genes (*ompA*, *tbpB*, *tbpA*, and *wza*). This study was inspired by previous work which had revealed that recombination involving *M. glucosida* and *P. trehalosi*, together with host switching of isolates from cattle to sheep, have played important roles in the evolution of the *M. haemolytica* leukotoxin operon (Davies *et al.*, 2001; 2002).

Comparative nucleotide sequence analysis of the nineteen genes indicated that only single nucleotide substitutions had occurred in the *recA*, *asd*, *gnd*, *g6pd*, *mtlD*, and *gcp* genes of *M. haemolytica*. A phylogenetic tree based on their concatenated sequences supported the evolutionary relationships of isolates of *M. haemolytica* based on previous MLEE studies. In contrast, the *aroA*, *gap*, *mdh*, *pmm*, *plpA*, *plpB*, *plpC*, *plpD*, *ompA*, *tbpB*, *tbpA*, and *wza* genes showed evidence of varying degrees of mosaic structure which suggests that horizontal DNA transfer and intragenic recombination have played important roles in the evolution of these genes. Recombinant segments of *M. haemolytica* genes have been derived from *M. glucosida* and other unidentified sources indicating that DNA from other bacterial species has contributed to the evolution of *M. haemolytica*. Extensive recombinant exchanges have occurred in the outer membrane protein genes *ompA*, *tbpB*, *tbpA*, and *wza*. The different tree topologies and different patterns of nucleotide and amino acid diversity

of these four genes indicate that they have different evolutionary histories. However, the evolutionary histories of these four genes support the previously proposed view that *M. haemolytica* diversity has been influenced by the transmission of isolates from cattle to sheep and *vice versa*, which is probably linked to the domestication of these species.

The temperate bacteriophages of representative isolates of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* were also investigated in this study since phages are known to mediate horizontal DNA transfer. Induction of bacteriophages with mitomycin C and examination by electron microscopy revealed that temperate phages were more prevalent in *M. haemolytica* (24 of 32 isolates) than *M. glucosida* (one of six isolates) and *P. trehalosi* (one of four isolates). Genetic variation of the phages was assessed by restriction endonuclease analysis and host range was determined by plague assay. Phage DNA was successfully isolated from 15 *M. haemolytica*, one *M. glucosida*, and one *P. trehalosi* isolates and restriction endonuclease analysis identified nine different RE types (A to I) in *M. haemolytica*, one RE type (J) in *M. glucosida*, and one RE type (K) in *P. trehalosi*. Plaque assay revealed that 13 *M. haemolytica*, one *M. glucosida*, and one *P. trehalosi* lysates could produce lysis zones on one to six *M. haemolytica*, two *M. glucosida*, and one *P. trehalosi* indicator isolates, respectively.

Using restriction endonuclease analysis and genomic analysis of the *M. haemolytica* genome sequence, phages of RE type A were shown to be very similar to P2 phages of *E. coli*. The annotated genome of this phage type was derived from the bacterial unannotated genomic sequence and compared with the genome of P2 phages.

CHAPTER 1: INTRODUCTION

1.1 The organisms

1.1.1 History and nomenclature of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Mannheimia haemolytica, *Mannheimia glucosida*, and *Pasteurella trehalosi* were formerly recognized as one species, *Pasteurella haemolytica*, which was proposed by Newson and Cross when they isolated a bacterium that caused pneumonia in calves in 1932 (Newsom & Cross, 1932). However, *P. haemolytica* showed highly divergent characteristics and two general typing methods, biotyping and serotyping, were adopted to distinguish isolates within the *P. haemolytica* complex. Biotyping identified two biotypes of *P. haemolytica* based on a number of differences in phenotypes such as cultural and biochemical traits, pathogenicity, growth dynamics, and antibiotic sensitivity. The two biotypes were designated A and T which reflect the ability to ferment L-arabinose and trehalose, respectively (Smith, 1961). Serotyping has recognized 17 serotypes of *P. haemolytica* based on indirect haemagglutination assay (IHA) that detects differences in capsular polysaccharide (see section 1.4.1.1 for further details) (Fodor *et al.*, 1988; Fraser *et al.*, 1982; Pegram *et al.*, 1979; Younan & Fodar, 1995). There is a consistent association between serotypes and biotypes (Table 1.1). Serotypes 1, 2, 5 to 9, 11 to 14, 16, and 17 are associated with biotype A, whereas serotypes 3, 4, 10, and 15 are associated with biotype T (Biberstein & Gills, 1962).

The T biotype of *P. haemolytica* was separated from *P. haemolytica* and reclassified as *Pasteurella trehalosi* in 1990 based on numerical taxonomic analysis and DNA-DNA hybridization (Sneath & Stevens, 1990). Subsequently, the A biotype of *P. haemolytica* was assigned to a new genus, *Mannheimia*, in 1999 based on ribotyping, 16S rRNA sequencing, DNA-DNA hybridization, and phenotypic data

Table 1.1 Reclassification of the *P. haemolytica* complex

New species name	Previous biotypes and serotypes of *P. haemolytica*
M. haemolytica	A1, A2, A5, A6, A7, A8, A9, A12, A13, A14, A16, A17
M. glucosida	A11
P. trehalosi	T3, T4, T10, T15

(Angen *et al.*, 1999b). Serotype A11 isolates were reclassified as *Mannheimia glucosida*, and the remaining 12 serotypes were reclassified as *Mannheimia haemolytica*. The reclassification of the *P. haemolytica* complex is summarized in Table1.1.

The growing information on genetic relationships has led to the reclassification of other members of the family *Pasteurellaceae*, and the three genera *Pasteurella*, *Actinobacillus*, and *Haemophilus* that originally formed the family now include the new genera *Mannheimia* (Angen *et al.*, 1999b), *Phocoenobacter* (Foster *et al.*, 2000), *Lonepinella* (Osawa *et al.*, 1995), *Histophilus* (Angen *et al.*, 2003), *Gallibacterium* (Christensen *et al.*, 2003), and *Volucribacter* (Christensen *et al.*, 2004a). Sequence analysis of the 16S rRNA gene and several housekeeping genes within the *Pasteurellaceae* has provided additional evidence to support this new classification and also suggests that some species such as *[P]. trehalosi*, *[Actinobacillus] capsulatus*, *[Haemophilus] ducreyi*, *[Haemophilus] parainfluenzae*, *[Actinobacillus] actinomycetemcomitans*, and *[Haemophilus] parasuis* need further reclassification since they are on branches that are distinct from the other major branches of *Pasteurella*, *Haemophilus*, and *Actinobacillus*, respectively (Figure 1.1) (Christensen *et al.*, 2004b; Dewhirst *et al.*, 1992; Korczak *et al.*, 2004).

According to the topology based on 16S rRNA sequence data of members of the *Pasteurellaceae* (Figure 1.1), *Mannheimia* is more closely related to the genus *Actinobacillus* than to the genus *Pasteurella* (Angen *et al.*, 1999b; Davies *et al.*, 1996; Dewhirst *et al.*, 1992).

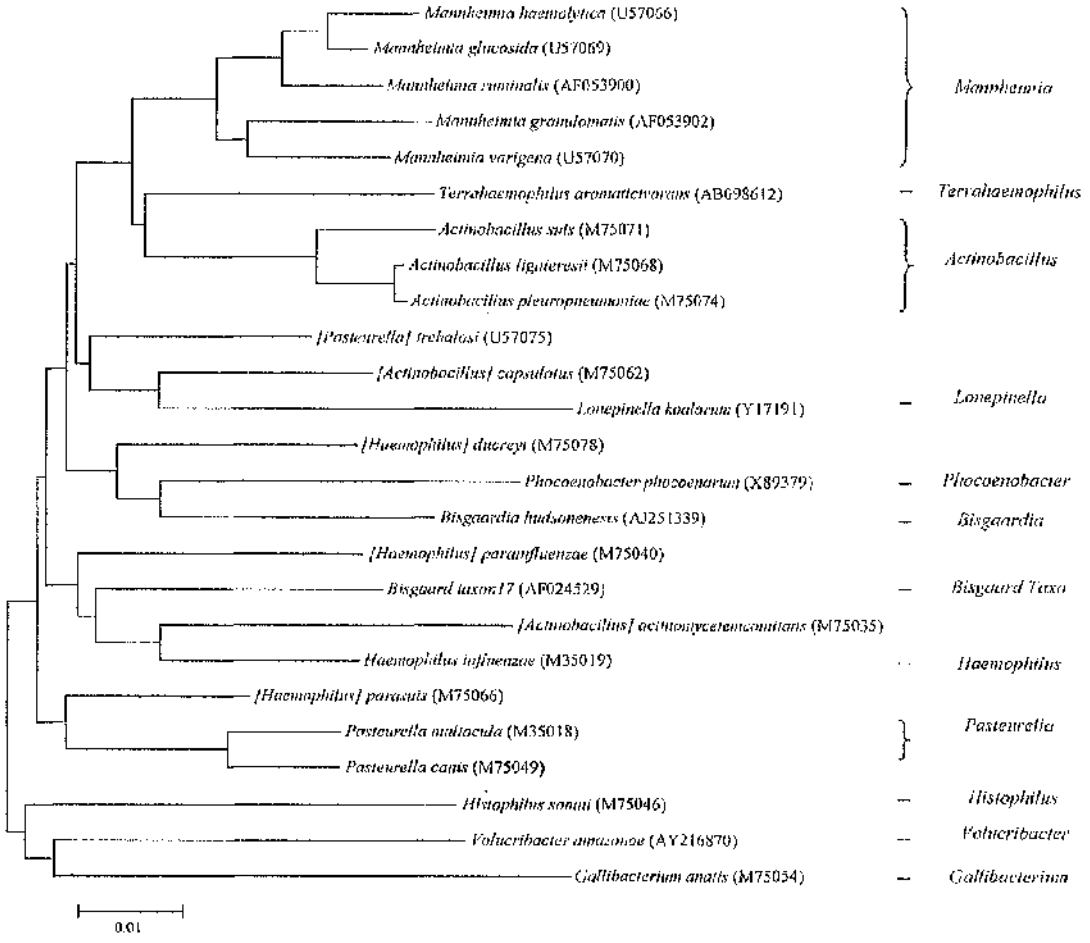


Figure 1.1 Neighbour-joining tree for the 16S rRNA gene of members of the *Pasteurellaceae*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions. The species name and GenBank accession numbers for the sequences used are shown. Genera in square brackets need reclassification because they are on branches that are distant from the major branches.

1.1.2 Morphological and biochemical characteristics of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Cell morphology. *M. haemolytica*, *M. glucosida*, and *P. trehalosi* have a morphology common with other members of the *Pasteurellaceae* (Angen *et al.*, 1999b; Sneath & Stevens, 1990). Cells are Gram-negative, non-motile, rods or coccobacilli. Two types of fimbriae have been demonstrated on the surface of A1 isolates of *M. haemolytica* (Potter *et al.*, 1988), but their presence on other serotypes of *M. haemolytica* as well as on *M. glucosida* and *P. trehalosi*, is unknown.

Cultural morphology. The cultural traits of *P. trehalosi* are slightly different from those of *M. haemolytica* and *M. glucosida* (Smith, 1961). After 24 hours the colonies of *P. trehalosi* are yellowish in colour and measure up to 2 mm in diameter. When smears are made in distilled water, lace-like patterns appear. On the other hand, the colonies of *M. haemolytica* and *M. glucosida* are smaller than those of *P. trehalosi* and show are lightish-grey in colour. Smears in distilled waters show even dispersal.

Biochemical characteristics. The biochemical properties of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* that have been reported by previous investigators (Angen *et al.*, 1999b; Mutters *et al.*, 1989; Sneath & Stevens, 1985; Sneath & Stevens, 1990) are compared in Table 1.2. Three species contain common characteristics for haemolysis, D-sorbitol, maltose, dextrin, indole and D-melibiose. However, *P. trehalosi* can be distinguished from *M. haemolytica* and *M. glucosida* by fermenting trehalose and not fermenting D-xylose. *M. haemolytica* and *M. glucosida* can be differentiated by the NPG test. Although *M. glucosida* and *M. haemolytica* are referred to as the 'A biotype' (Smith, 1961), no isolate of these species ferments L-

Table 1.2 Biochemical characteristics of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Property	*M. haemolytica* [a]	*M. glucosida* [a]	*P. trehalosi* [b]
Haemolysis	+	+	+
Ornithine decarboxylase	-	D	-
Trehalose	-	-	+
L-Arabinose	-	D	-
D-Sorbitol	+	+	+
D-Xylose	+	+	-
Maltose	+	+	+
Dextrin	+	+	+
Glucosides	-	D	D
Gentiobiose	-	D	NT
NPG (β-glucosidase)	-	+	D
Meso-Inositol	D	+	D
ONPF (α-fucosidase)	+	D	-
ONPX (β-xylosidase)	D	D	-
ONPG (β-galactosidase)	D	+	-
Indole	-	-	-
D-Melibiose	-	-	-

[a] See references (Angen *et al.*, 1999b; Mutters *et al.*, 1989)

[b] See references (Sneath & Stevens, 1985; Sneath & Stevens, 1990)

Symbols: + = positive; - = negative; D = + or -; NT = not tested

arabinose in recent experiments (Angen *et al.*, 1999b; Mutters *et al.*, 1989).

1.2 The diseases

M. haemolytica, M. glucosida, and P. trehalosi are commensal parasites, but they have pathogenic potential and can multiply quickly under conditions of immunodeficiency and stress (Biberstein, 1990). The three species have differences in virulence and host specificity (Frank, 1989; Gilmour & Gilmour, 1989). *M. haemolytica* is recovered from various ruminants such as cattle, sheep, goats, and deer, whereas *M. glucosida* occurs mostly in sheep, and *P. trehalosi* occurs only in sheep (Biberstein & Thompson, 1966; Quirie *et al.*, 1986). *M. haemolytica* is responsible for pneumonic pasteurellosis (Frank, 1989), while *P. trehalosi* is responsible for systemic pasteurellosis and *M. glucosida* is not normally associated with disease conditions (Gilmour & Gilmour, 1989)

1.2.1 Pneumonic pasteurellosis caused by *M. haemolytica*

M. haemolytica is the etiological agent of bovine and ovine pneumonic pasteurellosis that cause considerable economic losses to the beef and sheep industries worldwide (Frank, 1989; Gilmour & Gilmour, 1989; Martin, 1996). Bovine pneumonic pasteurellosis has various names such as bovine respiratory disease (BRD), shipping fever, transit fever, dust pneumonia, bronchial pneumonia, fibrinous pleuropneumonia, and bovine epizootic pneumonia. It is believed that losses due to pneumonic pasteurellosis are greater than the losses due to all other diseases of cattle in the United States (De Alwis, 1993).

Serotypes A1, A2, and A6 of *M. haemolytica* are carried in the nasopharynges and

tonsils of healthy and unstressed cattle, whereas all 12 serotypes of *M. haemolytica* are associated with sheep. However, only small numbers occur because they are kept in check by host defence mechanisms (Frank, 1989; Gilmour & Gilmour, 1989). However, under stressful conditions such as physical and physiological stress or viral and bacterial infections, the compromised defence mechanisms allow rapid multiplication of specific serotypes of *M. haemolytica* in cattle and sheep (Biberstein & Thompson, 1966; Frank, 1989; Martin, 1996). The pathogenicity of different serotypes varies in cattle and sheep. For example, serotype A1 isolates are predominant in bovine pneumonic pasteurellosis (Frank, 1989; Gonzalez & Maheswaran, 1993), whereas serotype A2 isolates are the major course of ovine pneumonic pasteurellosis (Gilmour & Gilmour, 1989; Odugbo *et al.*, 2003). This increased proliferation has been shown to increase the number of aerosolised *M. haemolytica* (Jericho *et al.*, 1986; Purdy *et al.*, 1989), thus allowing bacteria to be aspirated deeper into the lung (DeRosa *et al.*, 2000). Once pulmonary alveolar colonization has begun, bacterial toxins together with the animal's immunopathologic responses result in pulmonary damage (Ackermann & Brogden, 2000; McBride *et al.*, 1999). Death can occur within two or three days or the infection can proliferate and lead to chronic lung damage (Brogden *et al.*, 1998). Thus, early recognition and treatment are important. The early clinical signs of pneumonic pasteurellosis are characterized by fever, dyspnea, cough, depression, anorexia, and nasal and eye discharge (Frank, 1989; Gilmour & Gilmour, 1989).

1.2.2 Systemic pasteurellosis caused by *P. trehalosi*

Systemic pasteurellosis of sheep caused by *P. trehalosi* is an important disease in Britain, but it appears to be less common worldwide than pneumonic pasteurellosis (Biberstein & Thompson, 1966). Most outbreaks occur sporadically in weaned

lambs from the period September to December. Stressful environmental conditions such as cold and wet weather and management factors such as marketing and transporting have been implicated as predisposing causes (Gilmour & Gilmour, 1989).

Since lesions were consistently found in the tonsils and in the pharyngeal and oesophageal mucosa, it is postulated that the bacteria multiply at these sites and spread via the venous and/or lymphatic drainage to the anterior vena cava and then to the lung capillary bed forming the characteristic emboli (Gilmour & Gilmour, 1989). The bacteria also spread haematogenously to all of internal organs including the liver, spleen and kidneys (Martin, 1996). This disease is essentially an acute enterotoxaemia and at the start of an outbreak some sheep are usually found dead, while others may be seen to have collapsed and to be dyspnoeic and frothing at the mouth (Gilmour & Gilmour, 1989; Martin, 1996)

1.3 Bacterial virulence factors

A variety of virulence factors are associated with *M. haemolytica* (Confer *et al.*, 1990; Highlander, 2001; Lo, 2001). The major virulence factors of *M. haemolytica* include secreted proteins such as leukotoxin (LKT), glycoprotease (Gcp), and neuraminidase, and cell surface factors such as capsular polysaccharide (CPS), lipopolysaccharide (LPS), and outer membrane proteins (OMPs).

1.3.1 Secreted proteins

1.3.1.1 Leukotoxin (LKT)

M. haemolytica LKT is a pore-forming cytolysin that is a member of the RTX (repeats

in toxin) family of Gram-negative bacterial cytotoxins (Strathdee & Lo, 1987). However, *M. haemolytica* LKT is distinguishable from other RTX toxins by its target specificity for ruminant leukocytes and platelet. Cytolysis requires higher LKT concentrations to form transmembrane pores, which results in osmotic lysis. On the other hand, at lower concentrations, *M. haemolytica* LKT modifies leukocytes and platelets to release toxic oxygen products and protease which cause tissue necrosis (Nyarko *et al.*, 1998).

Nucleotide sequence analysis of the leukotoxin (*lktA*) gene of *M. haemolytica* has led to an understanding of its molecular evolution (Davies *et al.*, 2001). The mosaic structure of *lktA* and the presence of identical *lktA* gene sequences in isolates of different evolutionary lineages suggest that horizontal transfer and intragenic and assortative (entire gene) recombination have occurred. In addition, host switching of isolates between cattle and sheep has also been an important factor in the evolution of *lktA* since ovine alleles contain recombinant segments derived from bovine serotype A2 strains. Therefore, horizontal DNA transfer and recombination, together with host switching of isolates from cattle to sheep, have led to the rapid evolution of LktA and have possibly contributed to host adaptation (Davies *et al.*, 2001)

1.3.1.2 Glycoprotease (Gcp)

The glycoprotease (Gcp) of *M. haemolytica* has been characterized and is a neutral protease which is not toxic or lytic to host cells. However, it has activity against O-linked sialoglycoproteins on the surface of host cells such as macrophages or on the mucosal epithelium of the host respiratory tract (Abdullah *et al.*, 1991; Abdullah *et al.*, 1992). *M. haemolytica* Gcp is also capable of inducing platelet activation, adhesion, and aggregation, which are observed in the lungs of cattle with pneumonic

pasteurellosis (Nyarko *et al.*, 1998).

The glycoprotease of a *M. haemolytica* A1 isolate has been sequenced (Abdullah *et al.*, 1991), and the presence of the *gcp* gene and enzyme activity have been examined in isolates of the various serotypes (Lee *et al.*, 1994; Watt *et al.*, 1997). According to PCR, southern blot hybridization, and western blot analyses, all *M. haemolytica* serotypes possess the *gcp* gene and have glycoprotease activity. However, *M. glucosida* contains the *gcp* gene, but exhibits no glycoprotease activity whereas *P. trehalosi* possesses neither the gene nor glycoprotease activity (Abdullah *et al.*, 1990; Lee *et al.*, 1994; Watt *et al.*, 1997). Vaccination with recombinant glycoprotease of *M. haemolytica* A1 was shown to protect cattle against experimental challenge with live *M. haemolytica* A1 (Shewen *et al.*, 2003).

1.3.1.3 Neuraminidase

There is evidence that *M. haemolytica* A1 isolates produce neuraminidase during a naturally acquired infection in cattle (Straus *et al.*, 1998) and neuraminidases produced by various serotypes of *M. haemolytica* are quite similar (Straus *et al.*, 1993). However, the exact mechanism of neuraminidase in pathogenicity has not yet been elucidated. It has been suggested that neuraminidase may exert a pathogenic role by removing sialic acid from mucus or cell surface glycoproteins (Gottschalk, 1960; Kelly & D., 1970). This can potentially cause a reduction of the protective capability of these proteins against potential pathogens and promote bacterial survival during an active infection.

1.3.2 Cell wall proteins and carbohydrates

1.3.2.1 Capsular polysaccharide (CPS)

The capsular polysaccharides are composed of repeating monosaccharides linked covalently to the cell surface by phospholipid or lipid A molecules (Roberts, 1996). Capsular polysaccharides are an incredibly diverse range of molecules that may differ not only in monosaccharide units but also in how these units are joined together (Roberts, 1996). The roles of bacterial capsules of a number of Gram-negative pathogens in avoiding host defensive systems and in adherence to mucus or epithelial cells have been reported (Boyce & Adler, 2000; Unkmeir *et al.*, 2002).

In *M. haemolytica*, 12 different polysaccharide capsule serotypes have been identified (Table 1.1) and further details of the capsule serotypes are provided in section 1.4.1.1. Studies of the A1 capsular type polysaccharide indicate that it is involved in adherence of *M. haemolytica* to alveolar surfaces (Brogden *et al.*, 1989), resistance to serum bacteriolysis (Chae *et al.*, 1990; McKerral & Lo, 2002), as well as inhibition of the phagocytic and bactericidal activities of neutrophils (Czuprynski *et al.*, 1989).

The capsule gene cluster of *M. haemolytica* serotype A1 has been characterized (Lo *et al.*, 2001) and has a structure typical of group II capsules of Gram-negative bacteria (Roberts, 1996). It consists of three regions which are involved in capsule transport (region 1), capsule biosynthesis (region 2), and phospholipid substitution (region 3) (Figure 1.2). The nucleotide and inferred amino acid sequences of regions 1 and 3 are homologous to those of a number of bacteria (Table 1.3) that produce group II capsules (Chung *et al.*, 1998; Frosch *et al.*, 1992; Kroll & Booy, 1996; Ward & Inzana, 1997). In contrast, the sequences of four uncharacterised

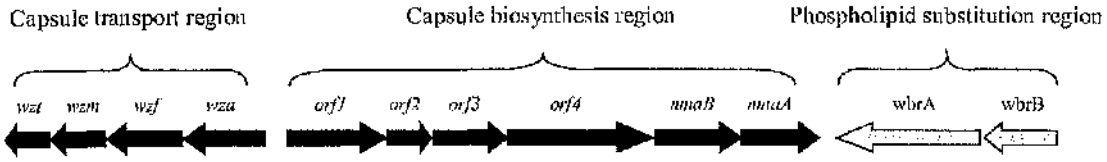


Figure 1.2 Genetic organization of the *M. haemolytica* serotype A1 capsule gene cluster. Green, red, and yellow arrows indicate the capsule transport, capsule biosynthesis, and phospholipid substitution regions, respectively.

Table 1.3 Capsular genes of *M. haemolytica* and homologous genes in other species

Gene name					Gene function
M. haemolytica	*H. influenzae*	*N. meningitidis*	*A. pleuropneumoniae*	*P. multocida*	
wzt	*bexA*	*ctrD*	*cpxA*	*hexA*	ATPase
wzm	*bexB*	*ctrC*	*cpxB*	*hexB*	Inner membrane protein
wzf	*bexC*	*ctrB*	*cpxC*	*hexC*	Periplasm-spanning protein
wza	*bexD*	*ctrA*	*cpxD*	*hexD*	Outer membrane protein
nmaB	-	-	-	-	
nmaA	-	-	-	-	
wbrA	-	*lipA*	-	*phyA*	Phospholipid substitution
wbrB	-	*lipB*	-	*phyB*	Phospholipid substitution

ORFs in region 2 are unique and are thought to have been acquired from an unknown source (Lo *et al.*, 2001). This is consistent with the general view that regions 1 and 3 are conserved, but region 2 is highly diverged due to recombinational exchanges in many bacterial species (Frosch *et al.*, 1992; Roberts, 1996).

Wza is located in the outer membrane and is involved in the translocation of capsular polysaccharide molecules across the outer membrane (Drummelsmith & Whitfield, 2000). Comparative sequence analysis of this protein indicates that it is highly conserved among different Gram-negative bacteria (Paulsen *et al.*, 1997; Rahn *et al.*, 1999; Whitfeld & Roberts, 1999). Wza homologs have been predicted to have a secondary β-barrel structure, (Frosch *et al.*, 1992; Paulsen *et al.*, 1997; Whitfeld & Roberts, 1999) and electron microscopy of *E. coli* Wza reveals that the protein consists of ringlike multimers in the outer membrane (Drummelsmith & Whitfield, 2000; Nesper *et al.*, 2003). Although there is a lack of sequence similarity, the images of the Wza multimers resemble those obtained with members of the secretin protein family (Genin & Boucher, 1994; Russel, 1998), which are involved in the transport of phage particles, DNA, and type II and III proteins across bacterial membranes. The secretins are thought to consist of two major domains, the C-terminal domain embedded in the outer membrane and the N-terminal domain which extends in to the periplasm and is involved in substrate binding and signal transduction (Brok *et al.*, 1999).

1.3.2.2 Lipopolysaccharide

The lipopolysaccharide (LPS) is the major glycolipid molecule present in the cell wall of Gram-negative bacteria and generally consists of three distinct regions (Preston *et al.*, 1996). The lipid A region is composed of sugars and fatty acids and anchors the

LPS in the outer membrane. The core oligosaccharide region consists of approximately 10 monosaccharides and connects the lipid A and O-antigen regions. The O-antigen consists of repeating units of from one to seven monosaccharides (Hitchcock *et al.*, 1986). However, a large number of Gram-negative bacteria contain LPS that lacks O-antigen. These two different LPS types, with and without O-antigens, have been termed smooth and rough-type LPS, respectively (Hitchcock *et al.*, 1986). In general, smooth-type LPS is associated with enteric bacteria and rough-type LPS with mucosal bacteria, although both LPS types can be present in enteric and mucosal bacteria (Preston *et al.*, 1996).

All three regions of LPS are involved in the pathogenesis of Gram-negative bacterial infections. Lipid A has roles in the typical endotoxic activities and host immunopathologic responses (Khan *et al.*, 1998; Luderitz *et al.*, 1973). The core oligosaccharide and O-antigen regions provide resistance to complement-mediated serum killing and phagocytic engulfment by blocking the access of complement to the outer membrane (Allen *et al.*, 1998; Pluschke *et al.*, 1983). The O-antigen also has a role in adherence (Bilge *et al.*, 1996; Jacques & Paradis, 1998).
M. haemolytica LPS is unusual in that rough and smooth forms occur independently in different isolates (Ali *et al.*, 1992; Davies & Donachie, 1996; Lacroix *et al.*, 1993). However, the presence of only a single O-antigen type, in isolates of relatively different diverged core-oligosaccharide regions, suggests that the O-antigen genes may have been obtained recently by horizontal DNA transfer (Davies & Donachie, 1996). Eight different LPS types have been identified in *M. haemolytica* (see section 1.4.1.2) and there is evidence that the LPS of serotype A1 isolates of *M. haemolytica* is also involved in endotoxic activities, adherence, and host immunopathologic responses (Adlam, 1989; Breider *et al.*, 1990).

1.3.2.3 Periplasm-associated proteins

1.3.2.3.1 Lipoproteins (PlpA, PlpB, and PlpC)

Three tandemly arranged lipoprotein genes encoding three 28-30 kDa lipoproteins, have been described separately in *M. haemolytica* by two different groups (Cooney & Lo, 1993; Murphy & Whitworth, 1993). Consequently, two different names, *plpA*, *plpB*, and *plpC* (Cooney & Lo, 1993; Murphy & Whitworth, 1993), and *lpp1*, *lpp2*, and *lpp3* (Cooney & Lo, 1993; Murphy & Whitworth, 1993) have been assigned to these genes. The genes appear to be transcribed from a single promoter located in upstream of *plpA* (or *lpp1*). They encode similar lipoproteins, which are homologous to the 28 kDa lipoprotein genes of *E. coli* (Yu *et al.*, 1986) and *H. influenzae* (Chanyangam *et al.*, 1991). These three lipoproteins are highly immunogenic (Dabo *et al.*, 1994), but examination of the predicted amino acid sequences of *plpA*, *plpB* and *plpC* of *M. haemolytica* suggests that these proteins may be localized to the inner membrane (Cooney & Lo, 1993). Data from southern blot analysis showed that all *M. haemolytica* and *M. glucosida* isolates contain the *plpA*, *plpB*, and *plpC* genes whereas *P. trehalosi* isolates do not appear to contain these genes (Cooney & Lo, 1993).

Although the exact role of the *M. haemolytica* PlpA, PlpB, and PlpC (or Lpp1, Lpp2, and Lpp3) proteins has not been demonstrated, a mutant isolate lacking these three lipoproteins resulted in enhanced susceptibility to bovine complement-mediated killing and reduced capacity for survival *in vivo* (Dabo *et al.*, 1994).

1.3.2.3.2 Lipoprotein (PlpD)

The *plpD* gene encodes lipoprotein PlpD, but it is located in a different part of the chromosome from *plpA*, *plpB*, and *plpC* (Nardini *et al.*, 1998). Sequence analysis indicates that the N-terminal region of PlpD contains typical lipoprotein sequences, but the C-terminal region is similar to the C-terminal region of the outer membrane protein OmpA (Nardini *et al.*, 1998). Therefore, Nardini *et al.* (1998) suggested that PlpD may be an outer membrane protein. However, PlpD may be a periplasmic-spanning protein rather than an outer membrane protein because the C-terminal region of OmpA corresponds to the periplasmic domain (Pautsch & Schulz, 2000).

1.3.2.4 Outer membrane proteins (OMPs)

About 50% of outer membrane mass consist of proteins, either in the form of integral membrane proteins or as lipoproteins that are anchored to the membrane (Koebnik *et al.*, 2000). Outer membrane proteins (OMPs) are essential not only because they maintain the integrity (Sonntag *et al.*, 1978) and selective permeability of the membrane (Lugtenberg & Van Alphen, 1983), but also because they are involved in bacterial pathogenesis. OMPs play a major role in immune complex disease, because many of the proteins are surface exposed (McBride *et al.*, 1999).

1.3.2.4.1 Heat modifiable outer membrane protein (OmpA)

The heat-modifiable outer membrane protein (OmpA) is highly conserved in Gram-negative bacteria (Beher *et al.*, 1980). The OmpA protein of *M. haemolytica* shows homology to that of *E. coli* and other Gram-negative bacteria (Mahasreshti *et al.*, 1997). Studies have suggested the heat-modifiable outer membrane protein (OmpA)

play roles in serum resistance (Weiser & C., 1991) in adherence (Dabo *et al.*, 2003; Reddy *et al.*, 1996; Torres & Kaper, 2003), in conjugation (Schweizer & Henning, 1977), and as a bacteriophage receptor (Morona *et al.*, 1984; Morona *et al.*, 1985) in many Gram-negative bacteria. The structure of OmpA has been very well studied in *E. coli*. The three-dimensional structure of OmpA has been determined by X-ray crystallography and nuclear magnetic resonance spectroscopy and indicates that the protein consists of eight membrane-traversing anitiparallel β-strands and four relatively long, mobile, hydrophilic surface-exposed loops (Arora *et al.*, 2001; Pautsch & Schulz, 1998; Pautsch & Schulz, 2000).

The OmpA proteins of bovine and ovine *M. haemolytica* isolates have been shown to exhibit interisolate molecular mass heterogeneity that correlates with the host of origin (Davies & Donachie, 1996). The *ompA* gene of a bovine serotype A1 *M. haemolytica* isolate has been cloned and sequenced (Zeng *et al.*, 1999). Although the function of OmpA has not been demonstrated in *M. haemolytica*, the significant homology of the *M. haemolytica* OmpA protein with that of other Gram-negative bacteria (Mahasreshti *et al.*, 1997; Zeng *et al.*, 1999) suggests that it may have similar functions. A strong IgG response in cattle to the surface-exposed domains of the *M. haemolytica* OmpA protein demonstrates that the protein has vaccine potential (Zeng *et al.*, 1999).

1.3.2.4.2 Transferrin-binding proteins (TbpA and TbpB)

Because the host provides an iron limited environment to potential pathogens, successful bacterial pathogens require mechanisms for obtaining iron from the host (Wooldridge & Williams, 1993). Therefore, a number of different iron acquisition systems have evolved in bacterial pathogens. These include the production of

siderophores and the presence of specific receptors for iron-containing glycoproteins such as lactoferrin and transferrin (Gray-Owen & Schryvers, 1996). For *M. haemolytica*, which does not produce siderophores, the transferrin binding proteins TbpB and TbpA are thought to be key in acquiring iron from the host (Ogunnariwo & Schryvers, 1996).

The transferrin binding protein genes of *M. haemolytica* have been cloned and sequenced (Ogunnariwo *et al.*, 1997). The *tbpB* and *tbpA* genes are in an operonic arrangement, with *tbpB* preceding *tbpA* and putative regulatory and promoter sequences upstream of the *tbpB* gene. The transferrin binding proteins (TbpA and TbpB) have molecular weights of 100 000 and 71 000, respectively (Deneer & Potter, 1989) and together form a receptor which is specific for ruminant transferrin (Ogunnariwo & Schryvers, 1990; Schryvers & Gonzalez, 1990; Yu *et al.*, 1992). The *tbpB* and *tbpA* genes from isolates of *A. pleuropneumoniae* (Gerlach *et al.*, 1992), *H. influenzae* (Gray-Owen *et al.*, 1995; Loosmore *et al.*, 1996), and *N. meningitidis* (Anderson *et al.*, 1994; Legrain *et al.*, 1993) have been cloned and sequenced. The TbpA proteins are generally highly conserved within a species, while the TbpB proteins tend to be more variable (Cornelissen & Sparling, 1994).

The structure and function of the transferrin binding proteins are similar to the lactoferrin binding proteins, suggesting that these iron-uptake mechanisms have arisen from a common evolutionary precursor (Gray-Owen & Schryvers, 1996; Pettersson *et al.*, 1994). The amino acid sequences of TbpA are similar to the membrane-spanning siderophore receptor FbpA (Buchanan *et al.*, 1999) and other members of the TonB-dependent family of receptor proteins (Cornelissen & Sparling, 1994). In contrast, the amino acid sequences of TbpB lack any obvious membrane spanning stretches (Gerlach *et al.*, 1992) indicating that it is lipid modified and anchored to the

outer membrane (Gray-Owen & Schryvers, 1996). Comparative sequence analysis of transferrin binding proteins and lactoferrin binding proteins from divergent species detect regions of identity that are likely to represent functionally conserved domains. Primers designed from these conserved regions were able to amplify segments of transferrin binding protein (or lactoferrin binding protein) genes from divergent species possessing there receptors (Ogunnariwo & Schryvers, 1996).

Although antigenic heterogeneity of transferrin binding proteins, particularly TbpB, limits its utility in some bacterial species (Linz *et al.*, 2000; Loosmore *et al.*, 1996; Myers *et al.*, 1998), common antigenic domains of TbpB and TbpA have been identified in several species (Holland *et al.*, 1996; Stevenson *et al.*, 1992), and their efficiency to produce protective antigens in various animal challenge models (Loosmore *et al.*, 1996; Rossi-Campos *et al.*, 1992) suggest that they are potential vaccine candidates . In *M. haemolytica*, protective antibodies were produced against TbpB and the antibodies appear to cross-react with TbpB from several different serotypes including serotypes of ovine pathogens (Potter *et al.*, 1999).

1.4 Variation in *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

1.4.1 Classification based on phenotypic relationships

1.4.1.1 Serotyping

Serotyping is used in many countries for differentiating isolates of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* and is based on the capsular polysaccharide (Biberstein & Thompson, 1966; Frank, 1989; Fraser *et al.*, 1983; Gilmour & Gilmour, 1989; Odugbo *et al.*, 2003; Sisay & Zerihun, 2003). Serotyping is performed by the

indirect haemagglutination assay (IHA) using antisera raised against the reference capsular polysaccharide types (Fraser *et al.*, 1983). To date, twelve capsular serotypes have been recognized in *M. haemolytica*, one in *M. glucosida*, and four in *P. trehalosi* (Table 1.1). However, untypeable isolates of *M. haemolytica* and *M. glucosida* isolates are also frequently isolated (Donachie *et al.*, 1984; Gilmour & Gilmour, 1989; Odugbo *et al.*, 2003). The untypeable isolates have been described as mutants which have lost their ability to produce capsular polysaccharide (Gentry *et al.*, 1988). In addition, capsular polysaccharides of types 6, 9, and 16 occur among other *Mannheimia* species (Angen *et al.*, 1999a). This suggests that serotyping is of limited use for differentiating isolates of *M. haemolytica* and *M. glucosida*.

Adlam characterized the capsular structures of *M. haemolytica* serotypes A1 (Adlam *et al.*, 1984), A2 (Adlam *et al.*, 1987), and A7, (Adlam *et al.*, 1986) and *P. trehalosi* serotypes T4 (Adlam *et al.*, 1985b) (Adlam *et al.*, 1985a) and T15 (Table 1.4). The authors reported that the capsular antigens of *M. haemolytica* and *P. trehalosi* could be found in other bacterial species. The capsular polysaccharide of *M. haemolytica* serotype A1 is similar to the widely distributed 'enterobacterial common antigen'. The *M. haemolytica* serotype A2 polymer is identical to the capsular polysaccharide of *Neisseria meningitidis* serogroup B and *Escherichia coli* K1. The serotype *M. haemolytica* A7 capsule has a similar structure to the *N. meningitidis* serogroup L and *Haemophilus influenzae* type f capsules. The teichoic acid like capsular antigen of *P. trehalosi* serotype T15 also occurs in the capsule of *E. coli* K2 and K62, *N. meningitidis* serogroup H, and *Actinobacillus pleuropeumoniae* serotype 9 (Beynon *et al.*, 1992). The *P. trehalosi* serotype T4 polymer is also similar to these capsular structures, but it differs from them in the linkage of galactose to the phosphate group, which is via C6 rather than C4. This difference in configuration is sufficient to make the polymers non cross-reacting in immunological tests with antisera

Table 1.4 The capsule composition of serotypes A1, A2, and A7 of *M. haemolytica* and serotypes T4 and T15 of *P. trehalosi*.

Serotype	Capsule Composition	Similar capsule types in other pathogens
A1	→3)-*O*-(2-acetamido-2-deoxy-4-*O*-acetyl-*β*-D-mannopyranosyluronic acid)-(1→4)-*O*-(2-acetamido-2-deoxy-*β*- D-mannopyranose)-(1→	Enterobacterial common antigen
A2	→2)-*α*-D-*N*-acetylneuraminic acid-(8→(and a dextran polymer)	*Neisseria meningitidis* serogroup B *Escherichia coli* K1
A7	→3)-*β*-2-acetamido-2-deoxygalactopyranose-(1→3)-*α*-2-acetamido-2-deoxy-6-*O*-acetly-glucopyranose-(1-phosphate→	*Neisseria meningitidis* serogroup L *Haemophilus influenzae* type f
T4	→(2-glycerol-1)→(phosphate)→(6-*α*-D-galactopyranose-1)→(partially *O*-acetylated on C2 and C3 of galactopyranose)	
T15	→(2-glycerol-1)→(phosphate)→(4-*α*-D-galactopyranose-1)→(partially *O*-acetylated on C2 and C3 of galactopyranose)	*Escherichia coli* K2 and K62 *Neisseria meningitidis* serogroup H *Actinobacillus pleuropneumoniae* serotype 9

(Adlam *et al.*, 1985b). Adlam (Adlam, 1989) interpreted that the production of similar or identical capsular structures among different pathogenic bacterial species is an example of convergent evolution to elude the defence mechanisms of the host under attack.

1.4.1.2 LPS profiles

The LPS profiles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* were analysed by SDS-PAGE (Ali *et al.*, 1992; Davies & Donachie, 1996; Davies & Quirie, 1996; Lacroix *et al.*, 1993). Eight LPS types were identified in *M. haemolytica*, one in *M. glucosida*, and six in *P. trehalosi* (Davies & Donachie, 1996; Davies & Quirie, 1996; Davies *et al.*, 1996). The variations of *M. haemolytica* LPS profiles were mainly due to four different types of low molecular mass bands representing core-oligosaccharide (1 to 4) and the presence or absence (A or B) of high molecular mass bands representing a single O-antigen type (Davies & Donachie, 1996). There was a strong correlation between LPS profiles and capsular serotype (Table 2.1). For example, serotypes A1, A5, A6, A9, and A12 were associated with LPS type 1A, serotypes A2, A8, A14, and A16 with LPS types 1B and 3B, and serotypes A7 and A13 with LPS type 4A. The LPS profiles of *M. glucosida* were represented by a single LPS type, 4C, which had a similar core-oligosaccharide to the type 4 LPS of *M. haemolytica* but had a distinctive O-antigen type, C (Davies & Donachie, 1996). The LPS profiles of *P. trehalosi* represented by six different LPS types, 1 to 6, were very different to those of both *M. haemolytica* and *M. glucosida* in their core-oligosaccharide and O-antigen regions (Davies & Quirie, 1996).

1.4.1.3 OMP profiles

The OMP profiles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* were analysed by SDS-PAGE (Davies & Donachie, 1996; Davies & Quirie, 1996). Twenty OMP profiles were identified in *M. haemolytica*, two in *M. glucosida*, and four in *P. trehalosi*, respectively. The variations OMP profiles of *M. haemolytica* were strongly correlated with capsular serotypes (Table 2.1). For example, serotypes A1, A5, A6, A8, A9, and A12 were associated with OMP type 1, serotypes A2, A14 and A16 were associated with OMP type 2, and serotypes A7 and A13 were associated with OMP type 3. The OMP profiles of *M. glucosida* showed similarity to those of serotype A7 and A13 isolates of *M. haemolytica* but the OMP profiles of *P. trehalosi* were very different to those of *M. haemolytica* and *M. glucosida*. In addition, OMP analysis demonstrated an association between specific OMP profiles and host specificity (i.e. cattle or sheep). For example, bovine isolates of serotype A1 and A6 were associated with OMP type 1.1, whereas ovine isolates of serotypes A1 and A6 were associated with OMP type 1.2. Similarly, bovine isolates serotype A2 were associated with OMP type 2.1, whereas ovine serotype A2 isolates were associated with OMP type 2.2 (Table 2.1).

1.4.2 Classification based on genetic relationships

In contrast to phenotypic classification, genotypic classification reflect evolutionary relationships and is a more reliable means of differentiating species and isolates within species.

1.4.2.1 Sequence analysis of the 16S rRNA gene

Sequence analysis of the 16S rRNA gene has been successfully used to discriminate species within *Pasteurellaceae* (see section 1.1.1) and clearly reveals that *M. haemolytica*, *M. glucosida*, and *P. trehalosi* are distantly related (Angen *et al.*, 1999b; Davies *et al.*, 1996). However, sequence analysis of the 16S rRNA gene is of limited use for differentiating isolates within a species because the gene is highly conserved. For example, various serotypes of *M. haemolytica* contain only two different types of 16S rRNA gene (Angen *et al.*, 1999b; Davies *et al.*, 1996). Serotype A1, A5 to A9, A12 to A14, and A16 isolates have identical 16S rRNA sequences but differ from serotype A2 at two nucleotide positions. However, the 16S rRNA sequence data for *M. glucosida* and *P. trehalosi* isolates indicated that all isolates were clearly differentiated as distinct species (Angen *et al.*, 1999b; Davies *et al.*, 1996).

1.4.2.2 Multilocus enzyme electrophoresis (MLEE)

Multilocus enzyme electrophoresis (MLEE), which measures allelic variation in housekeeping genes by detecting changes that cause amino acid substitutions affecting the net electrostatic charge or configuration of polypeptides, has been a standard method in both bacterial and eucaryotic population genetics and systematics (Selander *et al.*, 1986). The special advantage of MLEE is that variation in mobility can be directly related to allelic variation at specific genes encoding specific proteins. In particular, the variation is not likely to be due to convergent evolution because housekeeping proteins are selectively neutral and provides a reliable phylogeny in the absence of significant recombination (Musser, 1996; Selander & Smith, 1990; Selander *et al.*, 1986)

MLEE has been successfully used for analyzing the genetic diversity and population structure of *M. haemolytica* and provides an evolutionary framework for the variation of other molecules in *M. haemolytica* (Davies *et al.*, 1997a). In comparison with MLEE data, serotyping is of limited value in defining genetic relationships (Davies *et al.*, 1997a). MLEE showed that a small number of clones are associated with a large proportion of disease cases. For example, serotype A1 and A6 isolates of ET 1 were responsible for 75% of bovine disease, while serotype A2 isolates of ETs 21 and 22 and serotype A7 isolates of ETs 12 to 14 were responsible for the majority of ovine disease (Davies *et al.*, 1997a). This suggests that different evolutionary lineages have become adapted to cattle and sheep (Davies *et al.*, 1997a).

MLEE studies have also been carried out for *M. glucosida* (Davies *et al.*, 1997a) and *P. trehalosi* (Davies *et al.*, 1997b). The mean genetic diversity per locus (H) indicates that *M. glucosida* is a relatively divergent species ($H = 0.485$) whereas *M. haemolytica* ($H = 0.297$) and *P. trehalosi* ($H = 0.289$) are less divergent species (Davies *et al.*, 1997a; 1997b).

1.5 The molecular evolution of bacterial populations

1.5.1 Genetic variation occurs by three mechanisms

Genetic variation, which is continually arising in bacterial populations, is key to the evolution of bacteria, as it enables bacterial isolates to exploit new environments and defines their particular niche in the species (Lan & Reeves, 2000; Reeves, 1992). Three qualitatively different strategies are involved in the generation of genetic variation: local sequence changes, genetic rearrangements, and horizontal DNA transfer.

Local sequence changes. The strategy of local sequence changes includes point mutations (Musser, 1995) and small deletions/insertions (Zinser *et al.*, 2003), which are generated by mutagens or replication errors. The variation occurs spontaneously, and affects one specific trait that may confer an advantage in a changing environment. Since the mutations occur randomly throughout entire genes, the local sequence changes in essential genes such as ribosomal RNA genes or housekeeping genes are used to estimate evolutionary distances (Christensen *et al.*, 2004b; Korczak *et al.*, 2004).

Genetic rearrangements. Insertion sequences (IS) which encode only features necessary for their own mobilization form a part of bacterial genomes as repetitive DNA sequences and participate in bacterial DNA rearrangements in bacterial genomes (Mahillon & Chandler, 1998). Their DNA rearrangement activities result in gene conversion, inactivation, deletion, inversion, and duplication (Arber, 2003). Gene duplication provides the substrates for further mutational evolution such as local sequence changes or horizontal gene transfer without destroying the function of the product encoded by the original gene (Teichmann & Babu, 2004). In addition, IS also occur on plasmids, phages and in composite transposons where the IS often forms the ends of the elements and are involved in new antigenic variants (Ziebuhr *et al.*, 1999).

Horizontal DNA transfer. Bacteria can change rapidly by acquisition of foreign genes from other organisms through three mechanisms (Morschhauser, 2000; Saunders *et al.*, 1999; Spratt & Maiden, 1999). Transformation involves the uptake of free DNA from the environment; conjugation is a cell contact-dependent process that involves DNA transfer via plasmids or transposons; transduction is a phage mediated DNA transfer process. The acquisition of foreign genes allows the rapid

generation of new variants, unlike the first and second mechanisms which result in more slowly evolving genes (Lawrence, 1999; Ochman *et al.*, 2000; Saunders *et al.*, 1999). Therefore, the generation of new isolates and species by the first and second strategies is limited and is a very slow process, but horizontal gene transfer plays a major role in bacterial evolution (Garcia-Vallve *et al.*, 2000; Lan & Reeves, 1996). In particular, the importance of horizontal gene transfer has been highlighted since its role in the emergence of epidemic pathogens has been reported (Mooi & Bik, 1997; Reid *et al.*, 2000; Whittam *et al.*, 1993).

1.5.2 Limitation of bacterial diversity

Since genetic variation of bacteria lead only rarely to an increase in fitness of the organism, the frequency of such changes must be kept to tolerable levels (Begley *et al.*, 1999). This is brought about by DNA repair systems that fix most variation to ensure a certain degree of genetic stability (Arber, 2000). Similarly, restriction modification systems reduce foreign DNA acquisition to low levels by cutting the foreign DNA. However, these small DNA fragments are also recombinogenic. This suggests that the restriction modification systems also stimulate DNA variation in small steps (Arber, 2000; Edwards *et al.*, 1999; Milkman *et al.*, 1999). In addition, the host range of vectors such as bacteriophages and plasmids will limit the extent of horizontal DNA transfer and recombination.

1.5.3 Purifying and diversifying selection

The genetic variation which escapes the repair processes or gene loss is now subject to natural selection and this comprises purifying (negative) and diversifying (positive) selection. Purifying selection removes deleterious variants which reduce the fitness

of organisms whereas diversifying selection favours essential variants which enhance survival, spread, and transmission of an organism within a specific ecological niche. In this way, only advantageous changes can be incorporated into the population and lead bacterial species to improve their fitness in the encountered environment. The level of purifying and diversifying selection can be accessed by comparing the rates of synonymous and nonsynonymous nucleotide substitutions (Fay & Wu, 2001; Kimura, 1977; Nei & Kumar, 2000b). These are defined as the number of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per nonsynonymous sites (d_N), respectively. Since synonymous substitutions do not affect amino acid changes, they are free from natural selection and the d_S value is similar for many genes. On the other hand, nonsynonymous substitutions, which cause amino acid changes, are subject to selective pressure and the d_N value varies extensively from gene to gene, and at different sites within a gene, depending on the protein or domain function. Housekeeping genes are under severe functional constraints are subject to high levels of purifying selection and have low d_N/d_S (or high d_S/d_N) values. In contrast, outer membrane proteins are under strong selective pressure from the host immune system, experience high levels of positive selection, and have high d_N/d_S (or low d_S/d_N) values (Jordan *et al.*, 2002).

1.5.4 Evolutionary rates of different species

Bacterial population structures range from "highly clonal", where high levels of linkage disequilibrium occur to "non-clonal" or "panmictic", where alleles are at linkage equilibrium and recombinational exchanges are sufficiently frequent to obscure the evolutionary relationships (Feil & Spratt, 2001; Smith *et al.*, 1993; Spratt & Maiden, 1999). For example, *N. meningitidis* is a naturally transformable bacterial species (Kroll *et al.*, 1998; Morschhauser, 2000; Saunders *et al.*, 1999) and

has a non-clonal population structure, whereas *E. coli* and *Salmonella enterica* are not naturally transformable and have clonal structures (Pupo *et al.*, 1997; Selander & Smith, 1990). Therefore, the rate of horizontal DNA transfer in a non-clonal bacterial population is much higher than in a clonal bacterial population (Smith *et al.*, 1993). Nevertheless, bacteria such as *E. coli* contain a significant percentage of genes due to horizontal DNA transfer by conjugation or transduction (Lawrence & Ochman, 1998).

1.5.5 Evolutionary rates of genes of different functional classes

The relative level of purifying selection is thought to be the most important factor governing the rate of evolution for protein-coding genes (Jordan *et al.*, 2002). The essential proteins under strong functional constraint are subject to high levels of purifying selection, while non-essential proteins under less functional constraint experience lower levels of purifying selection and evolve more rapidly through horizontal DNA transfer and recombination (Garcia-Vallve *et al.*, 2000; Gogarten & Olendzenski, 1999; Jain *et al.*, 1999; Nelson & Selander, 1992; Radman *et al.*, 2000; Rivera *et al.*, 1998). Comparative sequence analysis (Figure 1.3) of different functional classes of genes among closely related genomes shows that different functional classes of genes evolve at significantly different rates (Jordan *et al.*, 2002). Among the 18 specific functional categories of genes based on the clusters of orthologous groups of proteins (COG) database, genes involved in energy production and conversion as well as in translation are highly conserved, whereas genes of unknown function and genes involved in DNA replication, recombination and repair are divergent. In addition, different functional groups show different levels of d_s value suggesting that synonymous substitutions are also subject to purifying selection based on functional constraint, although some effect of mutational bias cannot be

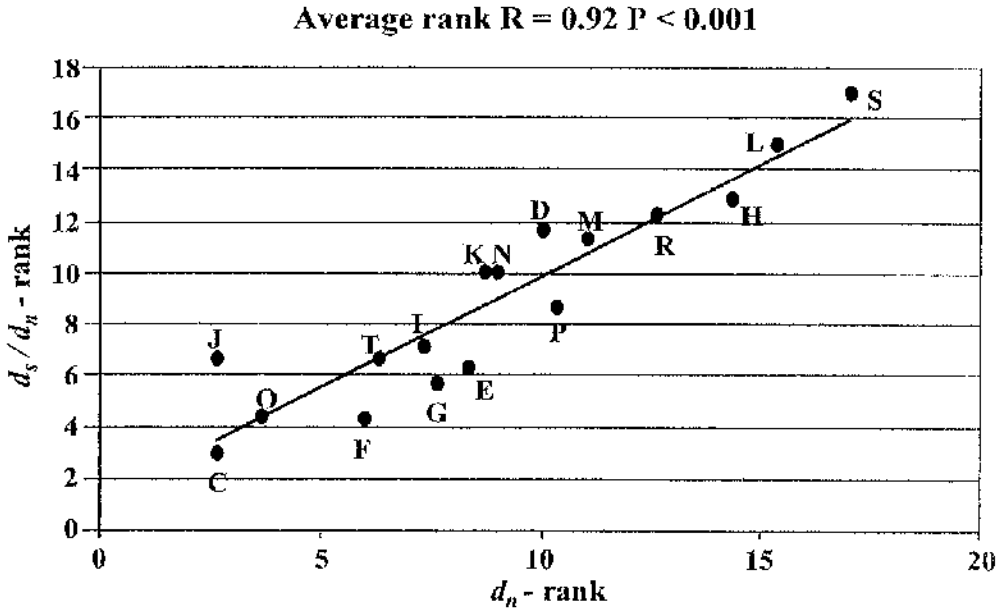


Figure 1.3 Rank correlation of average d_N and d_S/d_N for orthologs from different COG functional categories (Jordan *et al.*, 2002). **Metabolism**: C- energy production and conversion, G - carbohydrate transport and metabolism, E - amino acid transport and metabolism, F - nucleotide transport and metabolism, H - coenzyme metabolism, I - lipid metabolism. **Information storage and processing**: J - translation, ribosomal structure and biogenesis, K - transcription, L - DNA replication, recombination and repair. **Cellular processes**: D - cell division and chromosome partitioning, O - post-translational modification, protein turnover, chaperones, M - cell envelope biogenesis, outer membrane, N - cell motility and secretion, P - inorganic ion transport and metabolism, T - signal transduction mechanisms. **Poorly characterized**: R - general function prediction only, S - function unknown.

ruled out (Jordan *et al.*, 2002).

1.5.6 Housekeeping and virulence-associated genes

Housekeeping genes, which are associated with metabolism and information storage and processing, are selectively neutral (Feil & Spratt, 2001; Musser, 1996). Variation in these genes is generated mostly by local sequence changes and very low rates of horizontal DNA transfer and recombination occur. Thus, these genes have been used for elucidating the genetic structure of bacterial populations (Boyd *et al.*, 1994; Christensen *et al.*, 2004b; Stine *et al.*, 2000). In contrast, genes that encode virulence factors are under strong selective pressure from, for example, the host immune system, pH, temperature, and osmolarity and have been subject to frequent horizontal gene transfer events that obscure the clonal population structure (Guiney, 1997; Gupta & Maiden, 2001; Robertson & Meyer, 1992). These genes include those encoding outer membrane proteins (Hobbs *et al.*, 1994), capsular polysaccharides (Mooi & Bik, 1997), penicillin binding proteins (Spratt *et al.*, 1992), and exotoxins (Davies *et al.*, 2001).

1.5.7 Emergence of new pathogens by horizontal gene transfer

In the last two decades there have been large improvements in bacterial disease prevention and treatment due to advances in immunization and antibiotics. However, at the same time, not only have a series of newly recognized bacterial diseases emerged, but a number of diseases have re-emerged in drug resistant forms (Feil & Spratt, 2001). Most of these diseases are associated with new variants that have acquired virulence or drug resistant genes by horizontal gene transfer. The emergence of *E. coli* O157: H7 and *V. cholerae* O139 are well known examples of

these events.

1.5.7.1 Emergence of *E. coli* O157: H7

Enterohaemorrhagic *E. coli* (EHEC) O157: H7 is an important epidemic pathogen, which probably evolved from an enterophathogenic *E. coli* (EPEC) (Riley *et al.*, 1983). EHEC O157:H7 is distinguished from other *E. coli* by unique characteristics. For example, EHEC O157:H7 does not possess the classical toxins of enterotoxigenic *E. coli* (ETEC), lacks the invasive abilities of enteroinvasive *E. coli* (EIEC), and is serotypically distinct from EPEC, an important cause of infantile diarrhoea in the developing world (Riley *et al.*, 1983). However, EHEC O157: H7 isolates have EHEC-specific Shiga toxin genes encoded by prophages, carry EHEC-specific plasmids that encode adhesins, and have a EPEC-specific chromosomal *eae* gene that medicates attachment to epithelial cells (Donnenberg & Whittam, 2001). Phylogenetic analysis of *E. coli* isolates has shown that EHEC O157:H7 isolates are more closely related to EPEC O55:H7 isolates than to other EHEC isolates (Whittam *et al.*, 1993) (Reid *et al.*, 2000). Therefore, Reid (2000) and Donnenberg (2001) suggested that EHEC O157:H7 isolates evolved from an EPEC O55:H7-like ancestral clone by the addition of the EHEC specific Shiga-like toxins and plasmids.

1.5.7.2 Emergence of *Vibrio cholerae* O139

Among various serotypes of *Vibrio cholerae*, only *V. cholerae* serotype O1 has been responsible for epidemic cholera until serotype O139 was recovered in the cholera epidemic in India in 1992 (Albert *et al.*, 1993; Ramamurthy *et al.*, 1993). *V. cholerae* O139 produces LPS that has a distinct O-antigen with a shorter side-chain and a different sugar composition compared to that of *V. cholerae* serotype O1

(Comstock *et al.*, 1995; Johnson *et al.*, 1994). However, MLEE of *V. cholerae* (Johnson *et al.*, 1994) and *recA* sequence analysis (Stine *et al.*, 2000) have shown that *V. cholerae* O139 isolates are closely related to *V. cholerae* O1 strains. This suggests that *V. cholerae* O139 evolved from *V. cholerae* O1 after acquisition of the O139 type O-antigen genes by horizontal DNA transfer (Mooi & Bik, 1997).

1.6 Candidate genes for comparative sequence analysis

1.6.1 Genes encoding DNA repair and recombination enzymes

1.6.1.1 Recombinase A (*recA*)

The *recA* gene encodes RecA which is involved in nearly every type of homologous recombination event (Kowalczykowski *et al.*, 1994; Radman *et al.*, 2000). Mutations in *recA* affect not only recombination but also DNA repair, SOS mutagenesis, cell division, and chromosomal segregation. RecA is highly conserved among Gram-negative bacteria (Karlin & Brocchieri, 1996) and has been used to study the phylogenic relationships of *V. cholerae* (Byun *et al.*, 1999; Feil *et al.*, 1996; Stine *et al.*, 2000; Zhou & Spratt, 1992) and *Neisseria* species (Byun *et al.*, 1999; Feil *et al.*, 1996; Stine *et al.*, 2000; Zhou & Spratt, 1992).

1.6.2 Genes encoding metabolic enzymes

1.6.2.1 5-enolpyruvylshikimate-3-phosphate synthase (*aroA*)

The *aroA* gene encodes the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (AroA) which catalyzes the condensation of shikimate 3-phosphate (S3P) and

phosphoenolpyruvate (PEP) to form EPSP, a precursor for the aromatic amino acids, p-aminobenzoic acids, and dihydrobenzoate. AroA is an important enzyme for bacterial growth. Disruption of the *aroA* gene leads to auxotrophy of the bacteria and *aroA* mutants have been studied as vaccine candidates in several bacteria (Hernanz Moral *et al.*, 1998; Tabatabaci *et al.*, 2002; Tatum *et al.*, 1994).

1.6.2.2 Aspartate-semialdehyde dehydrogenase (*asd*)

The *asd* gene encodes the enzyme aspartate semialdehyde dehydrogenase (Asd) which catalyses the conversion of aspartyl phosphate to aspartate semialdehyde and is required for the biosynthesis of diaminopimelic acid (DAP), lysine, methionine, threonine and isoleucine (Harb & Abu Kwaik, 1998; Hatten *et al.*, 1993; Hoang *et al.*, 1997; Le *et al.*, 1996; Rees & Hay, 1995). DAP is involved in peptide cross-linking of peptidoglycan which is required for cell wall structure and mutations in *asd* of *Legionella pneumophila* have been shown to affect survival in the intracellular environment of host cells (Harb & Abu Kwaik, 1998).

Sequence comparison of *asd* in *Psudomonas aeruginosa*, *E. coli*, and *H. influenzae* has revealed greater than 63% identity among these bacteria (Hoang *et al.*, 1997), demonstrating the conserved nature of this gene in Gram-negative bacteria. However, comparative sequence analysis of the *asd* gene from pathogenic clones of *V. cholerae* demonstrated a high rate of intragenic recombination without a suitable explanation (Byun *et al.*, 1999).

1.6.2.3 UDP-galactose 4-epimerase (*galE*)

The *galE* gene encodes the enzyme UDP-galactose 4-epimerase (GalE) which

catalyzes the interconversion of UDP galactose and UDP glucose in the Leloir pathway of galactose metabolism (Potter & Lo, 1996). However, it is also important in LPS biosynthesis because galactose is a constituent of the O-antigen and core oligosaccharide regions of LPS. Thus, GalE plays an important role in an organism's virulence. *H. influenzae galE* mutants have been shown to be less virulent than the wild type (Maskell *et al.*, 1992). The inferred amino acid sequence of the *M. haemolytica* A1 GalE shows a high level of homology, 81.6%, with that of the *H. influenzae* type b GalE (Potter & Lo, 1996)

1.6.2.4 Glyceraldehyde-3-phosphate dehydrogenase (*gap*)

The *gap* gene encodes the enzyme glyceraldehyde-3-phosphate dehydrogenase (Gap) which catalyzes the reversible interconversion of glyceraldehyde-3-phosphate and 1,3-bisphosphoglycerate in the glycolytic pathway (Fothergill-Gilmore & Michels, 1993). Comparative nucleotide sequence analysis of the *gap* gene in *E. coli* and *S. typhimurium* have shown that the tree topologies were generally similar to those based on MLEE suggesting that allelic variation has been generated largely by point mutations rather than intragenic recombination (Nelson *et al.*, 1991).

1.6.2.5 6-phosphogluconate dehydrogenase (*gnd*)

The *gnd* gene encodes the enzyme 6-phosphogluconate dehydrogenase (Gnd) which catalyzes the conversion of 6-phosphogluconate to ribulose-5-phopate in the pentose phosphate pathway (Sprenger, 1995). Ribulose-5-phosphate is the final product of the pentose phosphate pathway and is involved in the biosynthesis of RNA, DNA, nucleotide coenzymes, and cell wall constituents. The *gnd* gene of *E. coli* and *S. enteria*, in spite of its housekeeping nature, showed high levels of variation due to

intragenic recombination (Bisercic *et al.*, 1991). It was suggested that this was due to its location next to the highly polymorphic *rfb* genes, which encode O-antigen structures (Bisercic *et al.*, 1991; Nelson & Selander, 1994; Thampapillai *et al.*, 1994). However, the *gnd* gene is not located in the *rfb* operon of *Actinobacillus actinomycetemcomitans* (Bisercic *et al.*, 1991) and in the MLEE study of *M. haemolytica* Gnd showed no amino acid polymorphism (Davies *et al.*, 1997a).

1.6.2.6 Glucose-6-phosphate-1-dehydrogenase (*g6pd*)

The *g6pd* gene encodes the enzyme glucose-6-phosphate 1-dehydrogenase (G6pd) which catalyzes the conversion of glucose 6 phosphate to 6-phosphoglucono-δ-lactone in the first step of the pentose phosphate pathway (Sprenger, 1995). There is less information about sequence analyses of this gene, significant variation of G6PD was demonstrated in the MLEE study of *M. haemolytica* (Davies *et al.*, 1997a)

1.6.2.7 Malate dehydrogenase (*mdh*)

The *mdh* gene encodes the enzyme malate dehydrogenase (Mdh) which catalyses the conversion of malate to oxaloacetate in several metabolic processes, including the tricarboxylic acid and glyoxylate cycles. Previous studies of *mdh* have shown little or no evidence of recombination (Boyd *et al.*, 1994), and the *mdh* gene has been used to determine the genetic relationships of *E. coli* (Pupo *et al.*, 1997). Significant variation of Mdh was demonstrated in the MLEE study of *M. haemolytica* (Davies *et al.*, 1997a).

1.6.2.8 Mannitol-1-phosphate dehydrogenase (*mtlD*)

The *mtlD* gene encodes the enzyme mannitol-1-phosphate dehydrogenase (MtlD) which catalyses the conversion of mannitol-1-phosphate to fructose 6-phosphate (Teschner *et al.*, 1990). There is less information about sequence analyses of this gene, but significant variation of MtlD was demonstrated in the MLEE study of *M. haemolytica* (Davies *et al.*, 1997a).

1.6.2.9 Phosphomannomutase (*pmm*)

The *pmm* gene encodes the enzyme phosphomannomutase (Pmm) which converts mannose-6-phosphate to mannose-1-phospate. Mannose-1-phospate is required for the synthesis of a full-length core polysaccharide in addition to O-antigen (Allen *et al.*, 1998; Comstock *et al.*, 1996; Goldberg *et al.*, 1993). *Brucella abortus* mutants in *pmm* had decreased virulence in mice, confirming the role of this enzyme in the biosynthesis of smooth-type LPS and its role in the intracellular survival of this organism (Allen *et al.*, 1998). In *Sphingomonas paucimobilis* and *P. aeruginosa* Pmm is a bifunctional protein with phosphomannomutase (Pmm) and phosphoglucomutase (Pgm) activity. This protein converts glucose-6-phosphate to glucose-1-phosphate as well as mannose-6-phosphate to mannose-1-phospate. Pmm/Pgm is also involved in synthesis of O-antigen and the core origosaccharide region of LPS, (Regni *et al.*, 2000; Videira *et al.*, 2000; Ye *et al.*, 1994). However, Pmm and Pgm are separate proteins in *Brucella abortus* and *Brucella suis* (Allen *et al.*, 1998; Foulongne *et al.*, 2000; Ugalde *et al.*, 2000). Pmm is also required for alginate production (Goldberg *et al.*, 1993; Ye *et al.*, 1994), which is believed to protect the infecting bacterial cells from phagocytosis, as well as from antibiotic therapy (Govan & Deretic, 1996). The Pgm protein of *M. haemolytica* was shown to

have significant variation by MLEE analysis (Davies *et al.*, 1997a).

1.6.3 Genes encoding secreted proteins

1.6.3.1 Glycoprotease (*gcp*)

The *gcp* gene encodes glycoprotease (Gcp) and was selected based on its putative virulence characteristics and because it is produced in secreted form. Further details of the glycoprotease of *M. haemolytica* are provided in section 1.3.1.2.

1.6.4 Genes encoding periplasm-associated proteins

1.6.4.1 Lipoproteins (*plpA*, *plpB*, and *plpC*)

The *plpA*, *plpB*, and *plpC* genes encode the lipoproteins PlpA, PlpB, and PlpC, respectively, and were selected because the proteins have a highly antigenic nature. Further details of the lipoproteins of *M. haemolytica* are provided in section 1.3.2.3.1.

1.6.4.2 Lipoprotein (*plpD*)

The *plpD* gene encodes the lipoprotein PlpD and was selected because the protein has a highly antigenic nature. Further details of the lipoprotein of *M. haemolytica* are provided in section 1.3.2.3.2.

1.6.5 Genes encoding outer membrane proteins

1.6.5.1 Heat modifiable outer membrane protein (*ompA*)

The *ompA* gene encodes the heat modifiable outer membrane protein OmpA and was selected because of previously recognized molecular mass variation in bovine and ovine isolates of *M. haemolytica* and its potential role in virulence. Further details of the OmpA protein of *M. haemolytica* are provided in section 1.3.2.4.1.

1.6.5.2 Transferrin binding proteins (*tbpB* and *tbpA*)

The *tbpB* and *tbpA* genes encode the transferrin binding proteins TbpB and TbpA and were selected because the transferrin receptor is an important virulence factor, and more importantly is known to exhibit specificity for host transferrin. Further details of the TbpB and TbpA proteins of *M. haemolytica* are described in section 1.3.2.4.2.

1.6.5.3 Capsule transport protein (*wza*)

The *wza* gene encodes the capsule transport protein Wza and was selected because it is located in the outer membrane and exposed to the host immune system. Further details of the Wza protein of *M. haemolytica* is described in section 1.3.2.1.

1.7 Analysis of nucleotide sequence data

1.7.1 Methods of phylogenetic tree reconstruction

There are numerous methods for constructing phylogenetic trees from molecular data

(Nei & Kumar, 2000a). Of these, the unweighted pair-group method using arithmetic averages (UPGMA), the Neighbour-joining method, and the Minimum Evolution method are commonly used for nucleotide sequence data.

UPGMA. UPGMA, which was originally used for phenotypic characters, can be used for constructing phylogenetic trees from nucleotide sequence data when the rate of nucleotide or amino acid substitution is the same for all evolutionary lineages. However, the rate of nucleotide substitution varies in different evolutionary lineages and, therefore, UPGMA often gives an incorrect topology (Nei & Kumar, 2000a).

Neighbour-joining method. The Neighbour-joining method generates a tree topology by finding neighbours, which are two taxa connected by a single node in an unrooted tree. This method is a simplified version of the Minimum Evolution method and can be used instead of the Minimum Evolution method when a small number of sequences are used (Nei & Kumar, 2000a).

Minimum Evolution method. The Minimum Evolution method has useful statistical properties because the method examines all possible topologies to find a topology showing the smallest value of the sum of all branches (S). However, the number of possible topologies rapidly increases with the number of sequences and it becomes difficult to examine all topologies. Therefore, it is suggested to construct a Neighbour-joining tree first, and then the neighbourhood of the Neighbour-joining tree is examined to find a tree with a smaller S value (temporary Minimum Evolution tree). Next, the neighbourhood of the temporary Minimum Evolution tree is examined to find an even smaller S value. This close-neighbour-interchange (CNI) search will be continued until the Minimum Evolution tree is obtained, which is the tree with the smallest S value. However, if many trees represent the data almost

equally, the Minimum Evolution tree may not be supported.

1.7.2 Distance measures to be used for phylogenetic reconstruction

Although there are various distance measures for estimating the number of nucleotide or amino acid substitutions (d), there is no general statistical method for choosing an appropriate distance measure for constructing tree topologies. However, computer simulations and empirical studies have supported the following guidelines for the purpose of topology construction (Nei & Kumar, 2000c)

1. When the Juke-Cantor estimate of the number of nucleotide substitution per site (d) is 0.05 or less, the P or Jukes-Cantor distance is used whether there is a transition/transversion (R) bias or not or whether the substitution rate (r) varies with nucleotide site or not.

2. When $0.05 < d < 1.0$, the Jukes-Cantor distance is used unless the transition/transversion (R) ratio is high ($R > 5$). When the R ratio is high and the sequences examined are long, the Kimura distance or gamma distance should be used. However, if the number of sequences is large and the sequences examined are relatively short, the P distance often gives better results.

3. When $d > 1.0$, the phylogenetic tree constructed is generally unreliable.

1.7.3 Handling of sequence gaps and missing data

Gaps and missing data are generally ignored in the distance estimation, but there are two different ways to treat these sites (Nei & Kumar, 2000a). One approach, called

the complete-deletion, is to delete all of these sites from the analysis. This is generally desirable because different regions of DNA or amino acid sequences often evolve under different evolutionary pressures. However, if the number of nucleotides involved is small and gaps are distributed more or less at random, the other option, called the pairwise-deletion, may be used. This computes a distance for each pair of sequences ignoring only those gaps that are involved in the pairwise comparison.

1.7.4 Reliability of phylogenetic trees obtained

When a phylogenetic tree is constructed, it is important to examine the reliability of the tree obtained (Nei & Kumar, 2000d). The interior branch test and the bootstrap test are two major methods of testing the reliability of the tree for topological errors. Both tests examine the reliability of each interior branch of the tree and compute the confidence probability (P_C) or bootstrap value (P_B). If this P_C or P_B is over 95%, the interior branch is considered to be significantly positive. Even if the interior branches are not well supported by the high P_C or P_B, the phylogenetic tree could be a correct one because it is the best tree obtainable under the principles of reconstruction used.

1.8 Bacteriophages and their life cycle

1.8.1 Bacteriophages of the *Proteobacteria*

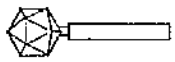
Bacteriophages are viruses that infect bacteria. According to the International Committee for the Taxonomy of Viruses (ICTV), one order, nine families, and seventeen genera of bacteriophages are currently identified in the *Proteobacteria* (Van

Regenmortel *et al.*, 2000) (Table1.5). The current classification of phages is based mainly on phage morphology (tailed, cubic, or filamentous) and the type of phage genome (ds DNA, ss DNA, ds RNA, or ss RNA); it does not reflect hierarchical or evolutionary relationships. A genome-based evolutionary taxonomy for phages has also been proposed (Rohwer & Edwards, 2002) and showed roughly similar classification to that proposed by the ICTV. However, it has been argued that a definition of viral species is not meaningful in the presence of excessive gene transfer (Lawrence *et al.*, 2002; Weinbauer & Rassoulzadegan, 2004).

The taxonomic names of orders, families, and genera are latinized, ending in *–virales* (e.g., *Caudovirales*), *–viridae* (e.g., *Myoviridae*), and *-virus* (e.g., *Inovirus*), respectively. However, the tailed phage genera have vernacular names based on the type species, ending in "-like virus" (e.g., P2-like virus) (Van Regenmortel *et al.*, 2000).

The most abundant types of bacteriophage are ds DNA tailed phages which account for 96% of reported bacteriophages; the other phages comprise less than 4% of bacteriophages (Ackermann, 2003). Tailed phages consist of three families, the *Myoviridae* (61%), *Siphoviridae* (25%), and *Podoviridae* (14%), which have long noncontractile, long contractile tails, and short noncontractile tails, respectively (Ackermann, 2003). They are highly diverse in their DNA size, DNA composition, dimensions and fine structure, and in their physiology, but their common morphological and replication features indicate that they are fundamentally related and they have recently been classified into the order *Caudovirales* (Maniloff & Ackermann, 1998). Tailed phages are arguably very ancient and some estimates suggest that they are older than the separation of life into Bacteria, Archaea and Eukarya (Hendrix *et al.*, 1999).

Table 1.5 The bacteriophages of the *Proteobacteria*

Family	Nature of the genome	Envelope	Genome size (kb)	Morphology
Myoviridae	dsDNA	No	39-169	
Siphoviridae	dsDNA	No	22-121	
Podoviridae	dsDNA	No	19-44	
Tectiviridae	dsDNA	No	15	
Corticoviridae	dsDNA	No	9	
Inoviridae	ssDNA	No	7-9	
Microviridae	ssDNA	No	4-6	
Cystoviridae	dsRNA	Yes	13	
Leviviridae	ssRNA	No	3-4	

1.8.2 Virulent and temperate bacteriophages

Bacteriophages can be grouped into two classes, virulent and temperate, based on the ability of lysogenization (Barksdale & Arden, 1974). Virulent phages are associated with all family types of bacteriophage (Van Regenmortel *et al.*, 2000) and the T series of dsDNA phages are well-studied examples (Birge, 1994a; Birge, 1994b). Virulent phages multiply inside the host cells and escape by lysing the cell or, in the case of filamentous phages, by a budding process. In contrast to lytic phages, temperate phages are associated exclusively with dsDNA tailed phages (*Caudovirales*) with the exception of *Inoviridae* phages that contain ssDNA (Van Regenmortel *et al.*, 2000). Temperate phages differ from virulent phages in that the phage genome is able to integrate into the bacterial chromosome without causing lysis of the host cell.

1.8.3 Major families of temperate phages

1.8.3.1 Myoviridae (dsDNA)

Myoviridae phages have long contractile tails, which are thick and more or less rigid (Figure 1.4A). However, the tail tubes surrounding the sheath become shorter and thicker during penetration (Caspar, 1980). Head and tails are assembled in separate pathways. *Myoviridae* phages seem to be more sensitive to freezing and thawing and to osmotic shock than other tailed phage families. Properties of temperate P1-like, P2-like, and Mu-like viruses are listed in Table 1.6.

1.8.3.2 Siphoviridae (dsDNA)

Siphoviridae phages have long noncontractile tails, which are thin and flexible (Figure

Table 1.6 Properties of temperate tailed bacteriophages

Family	Genus	Morphology	Head size (nm)	Tail size (nm)	Genome size (kb)	GC (%)	DNA topology	Pac site	Integration sites	UV Inducibility
Myoviridae	P1-like	Contractile tail Base plate 6 long fibers	85	228×18	100	46	Circular permuted Terminally redundant	Yes	Plasmid	Yes
	P2-like	Contractile tail Base plate, collar 6 short fibers	60	135×18	34	52	Circular permuted Cohesive end	Yes	Chromosome (Site specific recombination)	No
	Mu-like	Contractile tail Base plate 6 short fibers	60	120×18	39	50	Colinear	Yes	Chromosome (Transposition)	No
Siphoviridae	λ-like	Noncontractile tail 4 short fibers	60	150×8	49	52	Circular permuted Cohesive end	No	Chromosome (Site specific recombination)	Yes
Podoviridae	P22-like	Noncontractile tail Base plate 6 spikes	65	18×8	44	47	Circular permuted Terminally redundant	Yes	Chromosome (Site specific recombination)	Yes

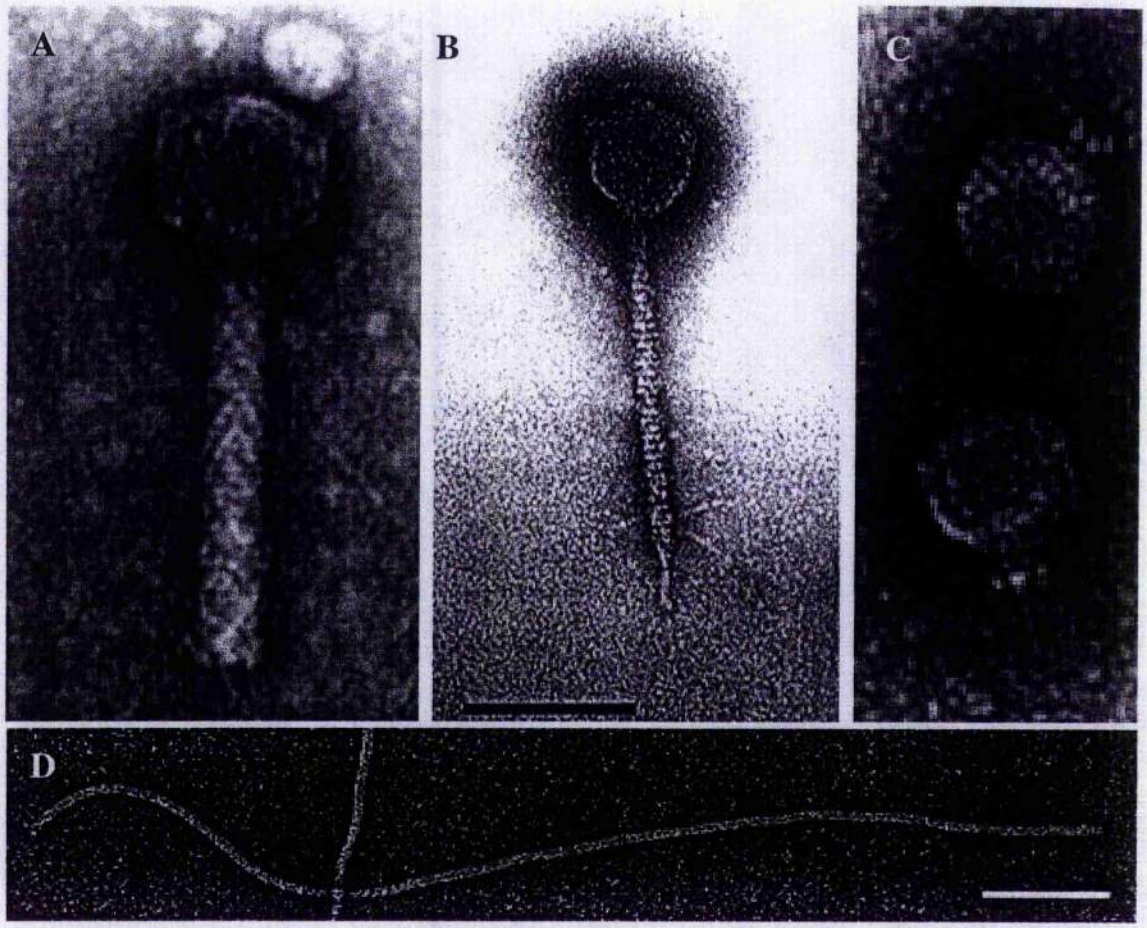


Figure 1.4 Negatively-strained electron micrographs of temperate phages of the major morphological types (families). (A) Enterobacteria phage P2 of the family *Myoviridae*, (B) Enterobacteria phage λ of the family *Siphoviridae*, (C) Enterobacteria phage P22 of the family *Podoviridae*, and (D) Enterobacteria phage M 13 of the family *Inoviridae*.

The sources: (A) http://www.biochem.wisc.edu/inman/empics/0021a.jpg

(B) http://www.virustaxonomyonline.com

(C) http://www.asm.org/division/m/foto/P22Mic.html

(D) http://www.virustaxonomyonline.com

1.4B). Head and tails are assembled in separate pathways. Properties of temperate λ-like virus are listed in Table 1.6.

1.8.3.3 Podoviridae (dsDNA)

Podoviridae phages have short, noncontractile tails (Figure 1.4C). The head is assembled first and the tail subunits are added to the completed head. Properties of temperate P22-like virus are listed in Table 1.6.

1.8.3.4 Inoviridae (ssDNA)

Inoviridae phages are filamentous (Figure 1.4D). They are approximately 7 nm in diameter and 700 to 2,000 nm in length. Although they are not stable, the genomes of *Inoviridae* phages are able to integrate into bacterial genomes.

1.8.4 Life cycles of temperate phages

1.8.4.1 Phage receptor recognition

Before phage DNA can enter through the bacterial cell wall, phages must attach to specific attachment sites, or receptors, on the cell surface (Lindberg, 1973). Common receptor sites include OMPs (Etz *et al.*, 2001; Lang, 2000), LPS (Nesper *et al.*, 2000; Quirk *et al.*, 1976), flagella (Samuel *et al.*, 1999), and pili (Lubkowski *et al.*, 1999; Malmborg *et al.*, 1997). In contrast to these phage receptors, capsule may normally block the access of bacteriophages to the receptors present in an underlying cell wall structure (Weiner *et al.*, 1995). Therefore, phages infecting encapsulated bacteria often produce capsular depolymerases which show substrate

specificity (Hughes *et al.*, 1998; Pelkonen *et al.*, 1992).

1.8.4.2 Penetration

Successful attachment is followed by penetration of the phage DNA into the cell. The sheath contraction of *Myoviridae* T4 phages is the best-studied mechanism of penetration for contracted tail types of bacterial phages (Caspar, 1980). This process is initiated by conformational changes in tail fibers bound to the cell wall. These trigger a rearrangement of the hexagon-shaped base plate into an extended star-shaped conformation. As a result, the sheath of the helical tail slips and forms a shorter helix and the tail tube slides through the cell wall and contacts the cytoplasmic membrane. A pore forms in the membrane and the DNA enters into the cell.

1.8.4.3 DNA circularisation

After attachment and penetration the phage DNA, which is mostly linear, becomes circularized by pairing of complementary cohesive ends (P2 and λ) or by generalized/site-specific recombination of terminally redundant ends (P1 and P22) (Keppel & Georgopoulos, 1988). Keppel and Georgopoulos (1988) suggested that a circular genome has several advantages: (1) it allows complete replication of the genome through a single, unidirectional DNA initiation event; (2) it is protected from degradation by host-coded exonucleases; (3) it is easily integrated into the bacterial genome through a single reciprocal recombinational event (see section 1.8.4.4.1); and (4) it is easily unwound or overwound through the action of topoisomerases

1.8.4.4 Lysogenization or lysis

An overall view of the life cycle of a temperate bacteriophage is shown in Figure 1.5. Temperate phages can enter into a lysogenic cycle or a lytic cycle. In the lysogenic cycle, phage DNA integrates into the bacterial chromosome and the phage genome is multiplied along with the bacterial genome at the time of cell division. However, in the lytic cycle, phage DNA is multiplied inside the host cells and mature phage escape by lysing the cell, in the same way as virulent phages (see section 1.8.2).

The genetic events leading to the lysogenic or lytic cycles have been well studied in phage lambda and are shown in Figure 1.6 (Birge, 1994c; Voyles, 2002b). First, immediate-early mRNA is transcribed to produce mainly N and Cro proteins (Figure 1.6A). Then, the antiterminator N protein allows transcription and translation of CII and CIII (Figure 1.6B). CII is a positive transcriptional regulator and CIII protects CII from rapid degradation by a host protease called HflA (for high frequency of lysogenation). CII/CIII proteins activate P_{RE}, causing transcription of CI (Figure 1.6C).

The Cro and CI proteins, which are the immediate-early and delayed-early gene products, respectively, are involved in a commitment to the lysogenic or lytic cycles. There are three binding sites between P_R and P_{RM}, which are accessible to both the CI and Cro proteins. CI proteins bind to the three sites from right to left, while the CI repressor proteins bind from left to right. If one CI protein fills the right site, it will block the transcription of Cro proteins (Figure 1.6D), but allow synthesis of CI until all three sites are occupied with CI proteins and lysogeny will be established (Figure 1.6E). In contrast, if one Cro protein fills the left site (Figure 1.6D´), it will block CI protein synthesis, but allow transcription of P, O, and Q proteins (Figure

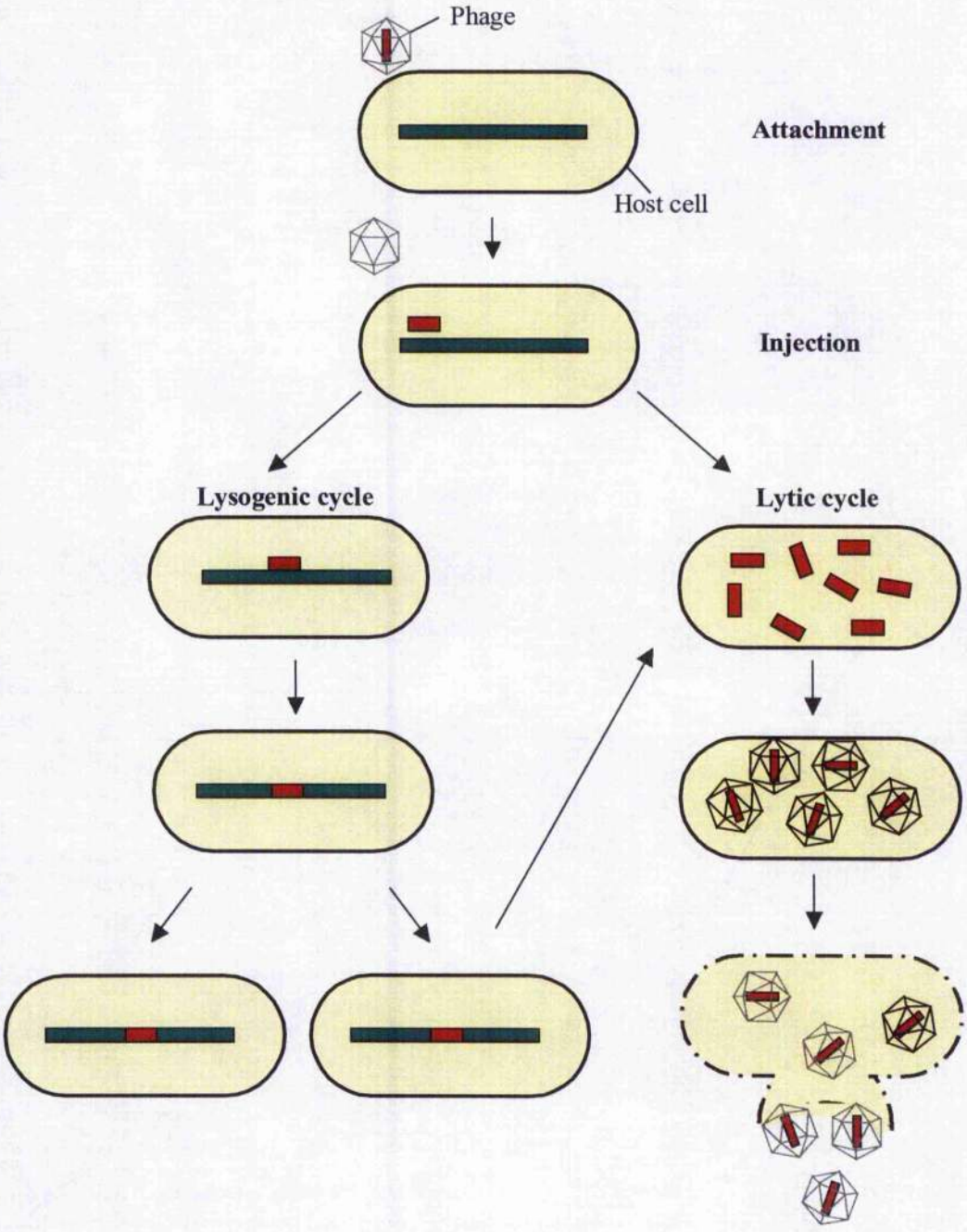


Figure 1.5 Consequence of infection by a temperate bacteriophage. The alternatives upon infection are integration of the virus DNA into the host DNA (lysogenization) or replication and release of mature virus (lysis). The lysogenic cell can also be induced to produce mature virus and lyse.

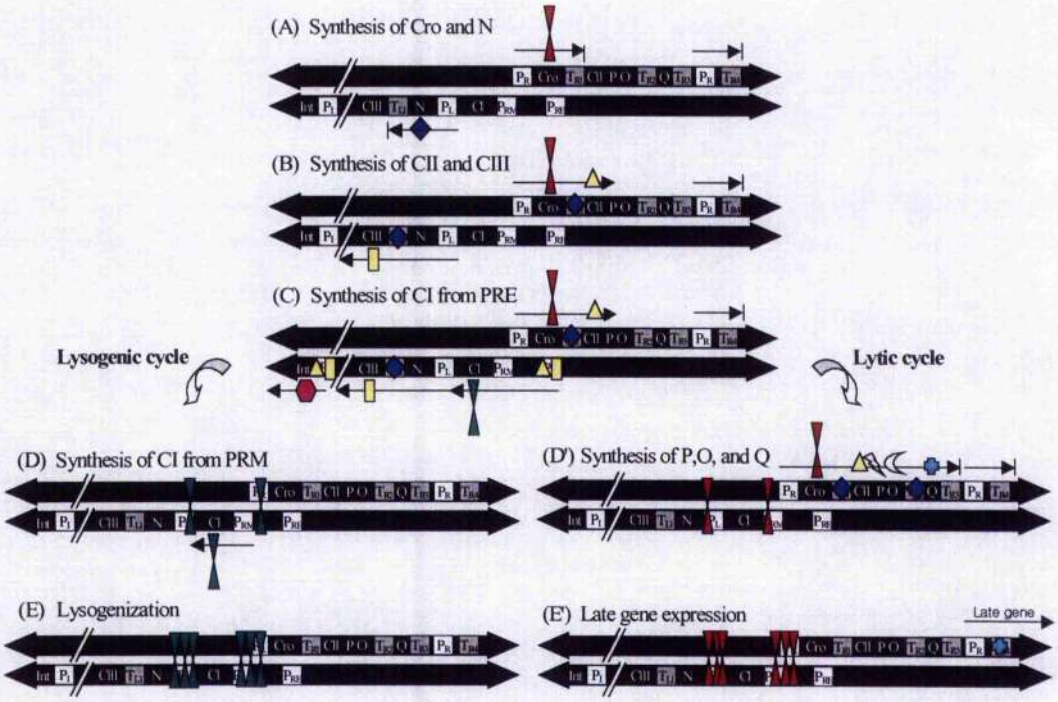


Figure 1.6 The molecular events of the lambda phage lysogenic or lytic cycle. (A), immediate-early gene expression of the Cro and N proteins. (B), antitermination by N permits synthesis of the CII and CIII gene activator proteins. (C), synthesis of the CI repressor and Int proteins begins at PRE and PI, respectively. **Lysogenic cycle:** (D), CI repressor protein blocks further transcription from PR and PL, but acts as gene activator permitting transcription of the CI gene form PRM. (E), concentrated CI proteins finally block PRM. **Lytic cycle**: (D'), the N protein permits transcription to continue and O, P, and Q proteins are synthesized. (E'), the activator protein Q allows transcription of the late genes and concentrated Cro proteins block PR, PL, and PRM.

1.6E´). P and O proteins are required for phage DNA replication, and the specific antiterminator Q proteins are required for transcription of late genes needed for the lytic cycle.

Although the Cro protein is made before the CI protein, the lysogenic cycle has frequently arisen due to two factors (Voyles, 2002b): 1) The efficiency and stability of the CI proteins on binding sites is better than those of the Cro proteins. 2) Low levels of energy sources activate adenylate cyclase and the concentration of cAMP becomes high, resulting in inhibition of transcription of the *hflA* gene. The lower concentration of HflA in turn reduces the degradation of CII and increases CI proteins. In starving cells, lysogenization is more advantageous than the lytic cycle to the phages because the overall biosynthetic rates of starving cells are low and the progeny virions will not be made quickly.

1.8.4.4.1 The lysogenic cycle

When the lysogenic cycle is established (Figure 1.6E), the phage genome integrates into the bacterial genome, via the Campbell model, by site-specific recombination between a bacterial *attB* site and the *attP* site of the circularized phage DNA (Figure 1.7) (Sadowski, 1986). There are various phage integration sites in the bacterial chromosome (Campbell, 2003). Some phages insert into intergenic DNA whilst others insert within genes; these may be either protein structural genes or *tRNA* genes. Intragenic insertion might be expected to inactivate the disrupted gene, but such inactivation is avoided by the presence within the phage of that part of the gene 3' to the crossover point. Thus, the lysogenic chromosome includes a complete copy of the gene and an incomplete pseudogene on the other end of the prophage. In general, insertion sites have an interrupted dyad symmetry. For example, the anticodon loop

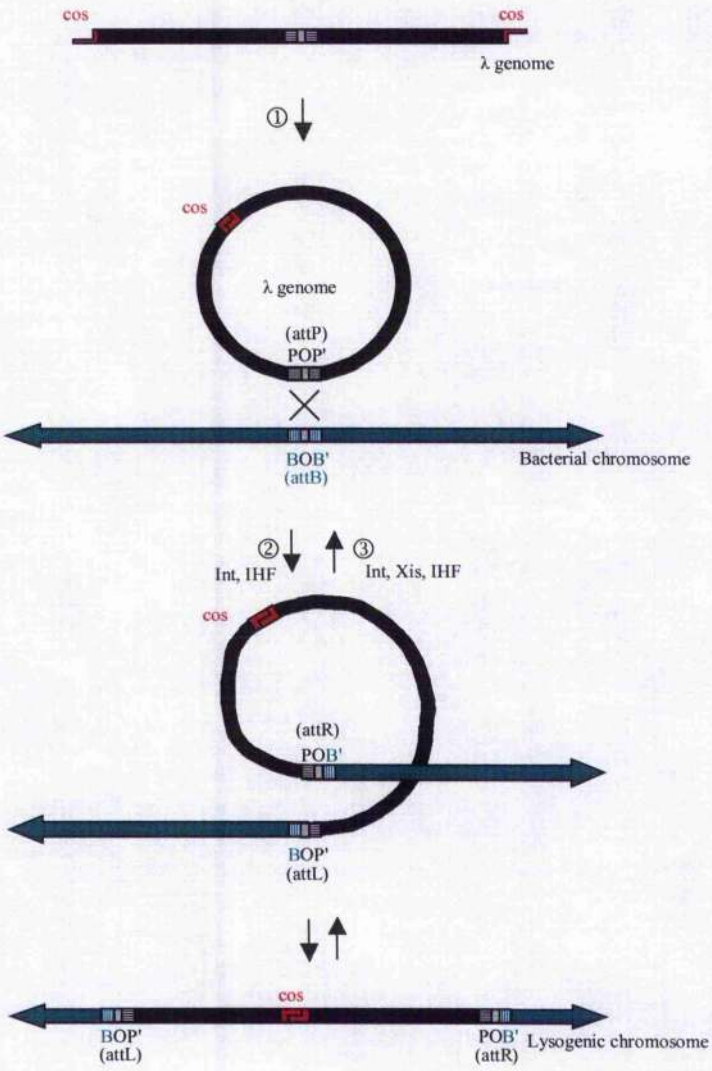


Figure 1.7 Mechanism of integration and excision of phage lambda (Sadowski, 1986). After injection, the linear lambda genome circularizes (reaction 1), and the phage attachment (*attP*) and bacterial attachment (*attB*) sites undergo reciprocal site-specific recombination promoted by Int and IHF (reaction 2). This creates two unique hybrid attachment sites (*attL* and *attR*) each of which differs from *attP* and *attB*. The integrated lambda phage DNA can be cut out by reversal of this reaction in the presence of the Int, IHF, and Xis proteins (reaction 3). The actual crossover takes place in the grey regions of each attachment sites.

region of *tRNA* genes is one of the common phage insertion sites.

1.8.4.4.2 The lytic cycle

When sufficient quantities of O and P proteins are produced (Figure 1.6D′), DNA replication is initiated and concatemers, multimeric DNA molecules, are made by rolling circle replication (Keppel & Georgopoulos, 1988). Then, late mRNA is transcribed to produce late proteins such as phage head and tail protein and lytic proteins, and the progeny viral DNA is cleaved from concatemeric DNA at *pac* sites (P1 and P22) or at unique sites to produce identical DNA molecules with *cos* sites (P2 and λ). The heads and tails are assembled independently (Voyles, 2002a). When DNA is packaged into the assembled head, the assembled tail and tail fibers are added. Ultimately, when assembly is finished, the virions are released from the host cell by lysing them.

1.8.5 Phage induction

Although lysogenic isolates rarely produce phages in nature, temperate phages can easily be induced by artificial treatments. The most popular agents for these treatments are mitomycin C (Humphrey *et al.*, 1995; Inoue & Iida, 1968; Jiang *et al.*, 1995; Pullinger *et al.*, 2004) and ultraviolet radiation (UV) (Inoue & Iida, 1968; Nordeen & Currier, 1983; Richards *et al.*, 1985) but also other agents such as danofloxacin (Froshauer *et al.*, 1996), ciprofloxacin (Strauch *et al.*, 2004), hydrogen peroxide (H_2O_2) (Figueroa-Bossi & Bossi, 1999), and X-ray (Boling & Randolph, 1977) can also be used. However, not all prophages are inducible by all agents and some temperate phages can only be induced by natural events.

Phage induction is associated with the SOS regulatory system that is activated as a result of DNA damage (Walker, 1984). Inducing agents such as mitomycin C and UV light damage DNA and activate the SOS system. In the SOS system, DNA damage induces a protease called RecA. RecA inactivates LexA, the repressor of a number of DNA repair genes, in order to induce expression of DNA repair functions. Since the phage specific repressors of the lytic cycle such as CI (see section 1.8.4.4) are structurally similar to LexA, these proteins also become inactivated and the lytic cycle will begin (Walker, 1984).

1.8.6 Phage immunity

Lysogenic bacteria can be infected by other phages, but they cannot be infected by additional members of the same phage type (Voyles, 2002b). This immunity is conferred by intracellular repression mechanisms under the control of virus genes. When cells are lysogenized with phages, there will be plenty of CI proteins in the cytoplasm (Figure 1.6E). Thus, if new lambda phages are introduced to these cells, the CI proteins can immediately bind to P_R and P_L in the newly introduced phages to block any expression of that virus's genome. Thus, there will be no lysis. However, the lysogenic cells can be cured by heavy irradiation (Barksdale & Arden, 1974). Presumably this treatment causes the prophage to excise from the host chromosome and be lost during subsequent cell growth. These cured isolates are no longer immune to the virus and they can be infected again.

1.9 The role of bacteriophages in bacterial evolution

1.9.1 Evolutionary arms race between temperate phages and bacteria

Many temperate phages carry genes that can increase the fitness of the lysogen (see section 1.9.3) (Brussow *et al.*, 2004). However, they are dangerous to their lysogenic hosts because the lysogenized phages will inevitably enter their lytic cycles in descendants of the initial lysogen. High genomic deletion rates are maintained to remove dangerous genetic parasites in the bacterial chromosome, and control the size of the bacterial genome, despite a constant bombardment with parasitic DNA over evolutionary time (Lawrence *et al.*, 2001). The deletion rates are determined by natural selection where the benefits realized by somewhat promiscuous deletion are offset by the detriments incurred by occasional removal of useful DNA.

1.9.2 Prophages in bacterial genomes

Since bacterial genome data have become available, one of the most surprising findings is the enormous extent of prophage DNA within bacterial genomes (Canchaya *et al.*, 2003a). A number of bacterial genomes contain multiple phage genomes which result in the introduction of numerous unique genes to the lysogen. The most extreme case is currently represented by the food pathogen *E. coli* O157:H7 isolate Sakai (Ohnishi *et al.*, 2001; Perna *et al.*, 2001). This *E. coli* isolate contains 18 prophage genome elements, which account for half of the isolate specific sequences. The sequenced *Streptococcus pyogenes* serotype M3 isolate also contains large number of prophage sequences which constitute about two-thirds of the strain-specific genes (Beres *et al.*, 2002). These findings suggest that bacteriophages have played a predominant role in the emergence of these pathogens. However, a number

of sequenced bacterial pathogens do not possess phage related genes suggesting that prophages may not be required for the evolution of a pathogenic life-style in every bacterial species (Brussow *et al.*, 2004).

1.9.3 Bacteriophages mediate gene transfer

Prophages are not only quantitatively important genetic elements of the bacterial chromosome, but also the most efficient virulence gene provider to bacterial chromosomes (Brussow *et al.*, 2004; Canchaya *et al.*, 2003b; Miao & Miller, 1999). Phage-mediated gene transfer can be achieved by two mechanisms, specialized transduction and generalized transduction (Birge, 1994d).

1.9.3.1 Specialized transduction

Sometimes, temperate phages are imprecisely excised from the host chromosome and small segments of flanking bacterial DNA can be co-packaged with the phage DNA (Birge, 1994d). Upon infection of the next host, this bacterial DNA, together with the phage DNA, can be incorporated into the bacterial chromosome by site-specific recombination. These extra genes frequently encode important virulence factors (Table 1.7) and allow the lysogen to enlarge its host range and increase its fitness in an environmental niche by phage conversion which promotes evasion of host immune defences or provide mechanisms to breach host barriers (Boyd & Brussow, 2002; Brussow *et al.*, 2004). Toxins are common phage-encoded virulence factors. A variety of major bacterial toxins such as cholera toxin, cytotoxin, exfoliative toxin A, and *Pasteurella multocida* toxin are located within the genome of temperate bacteriophages. In addition to toxins, bacteriophages carry O-antigen modification genes (*gtr*), type III secretion effector protein genes (*sopE*), superoxide dismutase

Table 1.7 Phage-encoded virulence factors

Virulence factor	Gene	Phage name	Phage classification	Bacterial host	Reference
Exotoxins					
Cholera toxin	*ctxAB*	CTXΦ	*Inoviridae*	*Vibrio cholerae*	(Waldor & Mekalanos, 1996)
Cytotoxin	*ctx*	ΦCTX	*Myoviridae*	*Pseudomonas*	(Nakayama *et al.*, 1999)
Exfoliative toxin A	*eta*	ΦETA	*Siphoviridae*	*Staphylococcus aureus*	(Endo *et al.*, 2003)
Toxin	*toxA*	-	*Siphoviridae*	*Pasteurella multocida*	(Pullinger *et al.*, 2004)
Proteins that alter antigenicity					
Glucosylation	*gtr*	P22	*Podoviridae*	*Salmonella enterica*	(Vander Byl & Kropinski, 2000)
Effector proteins involved in invasion					
The type III effector	*sopE*	SopEΦ	*Myoviridae*	*Salmonella enterica*	(Zhang *et al.*, 2002)
Enzymes required for intracellular survival					
Superoxide dismutase	*sodC*	Gifsy-2	*Siphoviridae*	*Salmonella enterica*	(Figueroa-Bossi & Bossi, 1999)
Serum resistance					
OMP	*lom*	Lambda	*Siphoviridae*	*Escherichia coli*	(Barondess & Beckwith, 1990)
OMP	*bor*	Lambda	*Siphoviridae*	*Escherichia coli*	(Barondess & Beckwith, 1995)
Adhesions for bacterial host attachment					
TCP pilus	*tcp*	VPIΦ	*Inoviridae*	*Vibrio* cholerae	(Karaolis *et al.*, 1999)

genes (*sodC*), outer membrane protein genes (*lom* and *bor*), and adhesion genes (*tcp*). These phage-encoded virulence factors are always linked to the viral DNA.

1.9.3.2 Generalized transduction

Sometimes, bacteriophages commit the error of packaging lysed bacterial DNA along with or instead of phage DNA and upon infection of the next host, this bacterial DNA can be incorporated into the bacterial chromosome (Birge, 1994d). Almost any size of bacterial genome can be found inside the head of a transducing phage particle. Upon infection of the next host, the foreign bacterial DNA is injected and may be incorporated into the bacterial chromosome by homologous recombination. Using genetically marked genes, the ability of generalized transduction has been examined and a number of bacteriophages capable of generalized transduction have been identified. They include temperate phages from *Pseudomonas aeruginosa* (Budzik *et al.*, 2004), *Streptomyces coelicolor* (Burke *et al.*, 2001), *V. cholerae* (Fidelma Boyd & Waldor, 1999; Hava & Camilli, 2001; Ogg *et al.*, 1981), *Pseudomonas syringae* (Nordeen & Currier, 1983), *Salmonella* and *E. coli* (Schicklmaier *et al.*, 1998), and *A. actinomycetemcomitans* (Willi *et al.*, 1997). Frequencies of transduction in many phages range from 10^{-5} to 10^{-8} per plaque-forming unit (PFU) (Burke *et al.*, 2001; Ogg *et al.*, 1981).

1.10 Objectives of research

The main objective of this study was to investigate the molecular evolution of genes of different functional classes in *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The study was inspired by previous work on the *M. haemolytica* leukotoxin operon (Davies *et al.*, 2001; 2002). These revealed that recombination involving

M. glucosida and *P. trehalosi*, together with host switching of isolates from cattle to sheep, have played important roles in generating the complex mosaic structure of the leukotoxin operon (Davies *et al.*, 2001; 2002). There were a series of questions to address in this study. Is there evidence of horizontal DNA transfer and recombination in other genes? Is there a correlation between the extent of recombination and gene function? Does the intensity of purifying and diversifying selection vary among different functional classes genes? How do the various evolutionary processes relate to differences in virulence and host specialization? Is there further evidence that host-switching of isolates has contributed to the evolution of other genes? With respect to these questions, comparative sequence analysis of one DNA repair and recombinase gene (*recA*), nine metabolic enzyme genes (*aroA*, *asd*, *galE*, *gap*, *gnd*, *g6pd*, *mdh*, *mtlD*, and *pmm*), one secreted protein gene (*gcp*), four periplasm-associated protein genes (*plpA*, *plpB*, *plpC*, and *plpD*), and four outer membrane protein genes (*ompA*, *tbpB*, *tbpA*, and *wza*) was performed in representative *M. haemolytica*, *M. glucosida*, and *P. trehalosi* strains.

A further goal of the study was to identify and characterize temperate bacteriophages among representative *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates since bacteriophages are known to mediate horizontal DNA transfer. Bacteriophages were characterized by electron microscopy, plaque assay, and restriction endonuclease analysis. In particular, a P2-like phage genome was identified and characterized in the bacterial genome of *M. haemolytica* isolate PH213 and the presence of this phage examined in selected isolates.

CHAPTER 2: MATERIALS AND METHODS

2.1 Bacterial isolates

Thirty two isolates of *M. haemolytica*, six isolates of *M. glucosida*, and four isolates of *P. trehalosi* were investigated in this study. These isolates were selected because they represent the various evolutionary lineages, serotypes, and host species of origin (bovine or ovine) and have been described in previous studies (Davies & Donachie, 1996; Davies & Quirie, 1996; Davies *et al.*, 1996; 1997a; 1997b; 2001; 2002). The properties of the isolates are shown in Table 2.1 and the genetic relationships of the *M. haemolytica* isolates are shown in Figure 2.1.

2.2 General bacteriological procedures

2.2.1 Growth media

Brain heart infusion agar (BHIA; Oxoid, UK) containing 5 % (v/v) sheep's blood (E & O laboratories Ltd) was used as routine solid medium for growth of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Brain heart infusion broth (BHIB; Oxoid, UK) was used as the liquid medium.

2.2.2 Sterilisation

All culture media, glassware, and plasticware were sterilised by autoclaving at 15 lbs p.s.i (121˚C) for 15 min.

Table 2.1 Properties of 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Isolate	ET [a]	Capsular serotype	Host species	LPS [b] type	OMP [c] type	*lktA* allele [d]	Origin [e]
M. haemolytica							
PH2	**1**	**A1**	**Bovine**	**1A**	**1.1.1**	***lktA1.1***	**University of Glasgow**
PH30	1	A1	Bovine	2A	1.1.1	*lktA1.1*	University of Glasgow
PH376	1	A6	Bovine	1A	1.1.4	*lktA1.1*	Dumfries VIC
PH346	1	A12	Ovine	1A	1.2.3	*lktA1.2*	Auchincruvie VIC
PH540	2	A1	Bovine	2A	1.1.3	*lktA1.1*	Germany
PH338	3	A9	Ovine	1A	1.2.3	*lktA1.2*	Dumfries VIC
PH388	4	A7	Ovine	3A	1.2.1	*lktA1.1*	Edinburgh VIC
PH50	5	A5	Ovine	1A	1.2.3	*lktA1.3*	Edinburgh VIC
PH56	5	A8	Ovine	1B	1.2.3	*lktA1.4*	Edinburgh VIC
PH238	5	A9	Ovine	1A	1.2.3	*lktA1.4*	Edinburgh VIC
PH8	6	A1	Ovine	1B	1.2.1	*lktA1.5*	Moredun Research Institute
PH398	7	A1	Ovine	1A	1.2.1	*lktA1.5*	Aberdeen VIC
PH284	8	A6	Ovine	1A	1.2.3	*lktA1.2*	Edinburgh VIC
PH232	9	A6	Ovine	2A	1.2.3	*lktA1.5*	Edinburgh VIC
PH66 (NCTC 11303)	**10**	**A14**	**Ovine**	**1B**	**2.3.1**	***lktA9***	**Ethiopia**
PH706	**11**	**A16**	**Ovine**	**1B**	**2.3.1**	***lktA7***	**Edinburgh VIC**

Isolate	ET [a]	Capsular serotype	Host species	LPS [b] type	OMP [c] type	*lktA* allele [d]	Origin [e]
PH296	12	A7	Ovine	4A	3.1.1	*lktA8.1*	Penrith VIC
PH396	13	A7	Ovine	4B	3.1.1	*lktA8.1*	Aberdeen VIC
PH484	14	A7	Ovine	4A	3.1.1	*lktA8.1*	St. Boswells VIC
PH588	**15**	**A13**	**Ovine**	**4A**	**3.3.2**	***lktA6***	**Sutton Bon VIC**
PH494	**16**	**A2**	**Ovine**	**3B**	**2.2.4**	***lktA2.1***	**St. Boswells VIC**
PH550	**17**	**A2**	**Bovine**	**1B**	**2.1.2**	***lktA2.1***	**St. Boswells VIC**
PH196	**18**	**A2**	**Bovine**	**1B**	**2.1.1**	***lktA3***	**University of Glasgow**
PH786	18	A2	Bovine	1B	2.1.1	*lktA3*	Aberdeen VIC
PH526	19	A2	Ovine	1B	2.2.2	*lktA8.1*	Newcastle VIC
PH598	20	A2	Ovine	1B	2.2.1	*lktA8.1*	Shrewsbury VIC
PH202	**21**	**A2**	**Bovine**	**3B**	**2.1.2**	***lktA2.2***	University of Glasgow
PH470	21	A2	Bovine	3B	2.1.2	*lktA2*	Aberdeen VIC
PH278	**21**	**A2**	**Ovine**	**3B**	**2.2.2**	***lktA10.1***	**Penrith VIC**
PH372	21	A2	Ovine	3B	2.2.2	*lktA10.1*	Edinburgh VIC
PH292	**22**	**A2**	**Ovine**	**1B**	**2.2.1**	***lktA8.1***	**Edinburgh VIC**
PH392	22	A2	Ovine	1B	2.2.1	*lktA8.2*	Aberdeen VIC
M. glucosida							
PH344	**1**	**A11**	**Ovine**	**4C**	**3.2.2**	***lktA4.1***	**Auchincruvie VIC**
PH498	3	A11	Ovine	4C	3.2.2	*lktA4.2*	St. Boswells VIC

Isolate	ET [a]	Capsular serotype	Host species	LPS [b] type	OMP [c] type	*lktA* allele [d]	Origin [e]
PH240	5	A11	Ovine	4C	3.2.2	*lktA4.3*	Oslo
PH496	7	UG3	Ovine	4A/C	3.3.2	*lktA4.4*	St. Boswells VIC
PH574	10	UG3	Ovine	4A/C	3.3.2	*lktA4.5*	Penrith VIC
PH290	16	UG3	Ovine	4A/C	3.3.2	*lktA4.6*	Penrith VIC
P. trehalosi							
PH246 (NCTC 10626)	**2**	**T4**	**Ovine**	**3**	**1**	***lktA5.1***	**Edinburgh**
PH252 (NCTC 10641)	4	T10	Ovine	2	2	*lktA5.2*	Edinburgh
PH254 (NCTC 10370T)	15	T15	Ovine	3	3	*lktA5.3*	Edinburgh
PH68 (NCTC 11550)	19	T3	Ovine	1	1	*lktA5.4*	Edinburgh

[a] ET: electrophoretic type

[b] See references Davies *et al.* (1997a; 1997b)

[c] See reference Davies & Donachie (1996) and Davies & Quirie (1996)

[d] See reference Davies *et al.* (2001)

[e] VIC: Veterinary Investigation Centre

Isolates in bold type represent the 12 preliminary isolates

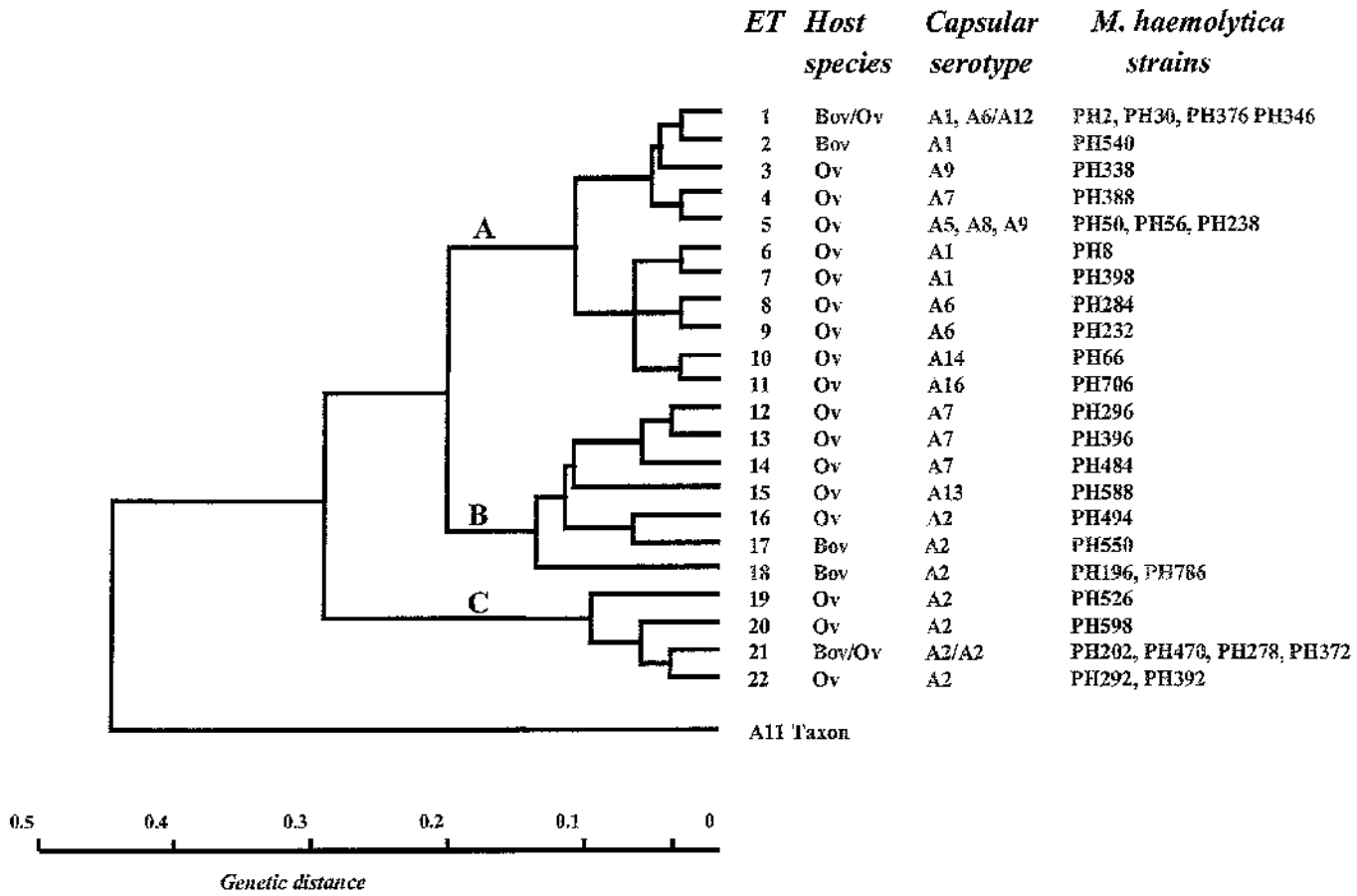


Figure 2.1 Genetic relationships of the 32 *M. haemolytica* isolates investigated in the study (Davies *et al.*, 1997a). Red: bovine isolates, blue: ovine isolates

2.2.3 Storage and growth of bacteria

Bacteria were stored at -85 °C in 50 % (v/v) glycerol in BHIB. Bacteria were subcultured from -85 °C stock cultures onto blood agar and incubated at 37 °C overnight. The following day, 4 to 5 colonies were inoculated into 15 ml BHIB in Universal tubes and incubated overnight at 37 °C with shaking (120 rpm).

2.2.4 Culture purity checks

All agar plates and broth cultures were subjected to thorough purity checks. Agar plate cultures were checked for any unusual colonies after 2 to 3 days. Broth cultures were subcultured onto blood agar and grown at 37 °C overnight.

2.3 Comparative sequencing analysis

2.3.1 General sequencing procedures

2.3.1.1 Selection of genes

Nineteen genes were selected based on the differing functions and usage of the encoded proteins (Table 2.2). These include one DNA repair and recombinase gene (*recA*), nine metabolic enzyme genes (*aroA*, *asd*, *galE*, *gap*, *gnd*, *g6pd*, *mdh*, *mtlD*, and *pmm*), one secreted protein gene (*gcp*), four inner membrane or periplasmic spanning lipoprotein genes (*plpA*, *plpB*, *plpC*, and *plpD*), and four outer membrane protein genes (*ompA*, *tbpB*, *tbpA*, and *wza*).

Table 2.2 Details of the genes selected for study

Gene name	Gene product	Function (usage)
Genes encoding DNA repair and recombination enzymes		
recA	RecA recombinase	DNA repair and recombination
Genes encoding metabolic enzymes		
aroA	5-enolpyruvylshikimate 3-phosphate synthase	Amino acid biosynthesis
asd	Aspartate-semialdehyde dehydrogenase	Amino acid biosynthesis Peptidoglycan biosynthesis
galE	UDP-glucose 4-epimerase	Carbohydrate metabolism LPS biosynthesis
gap	Glyceraldehyde-3-phosphate dehydrogenase	Carbohydrate metabolism
gnd	6-phosphogluconate dehydrogenase	Carbohydrate metabolism (MLEE data)
g6pd	Glucose-6-phosphate 1-dehydrogenase	Carbohydrate metabolism (MLEE data)
mdh	Malate dehydrogenase	Energy production and conversion (MLEE data)
mtlD	Mannitol-1-phosphate dehydrogenase	Carbohydrate metabolism (MLEE data)
pmm	Phosphomannomutase	Carbohydrate metabolism LPS and capsule biosynthesis
Genes encoding secreted proteins		
gcp	Glycoprotease	Neutral protease
Genes encoding periplasm-associated proteins		
plpA, *plpB*, and *plpC*	Lipoproteins PlpA, PlpB, and PlpC	Serum resistance
plpD	Lipoprotein PlpD	Not known
Genes encoding outer membrane proteins		
ompA	OmpA	Outer membrane structural integrity; adherence; phage receptor
tbpB and *tbpA*	TbpB and TbpA	Iron acquisition
wza	Wza	Capsule polysaccharide transport

2.3.1.2 Primer design for PCR amplification and sequencing

The GenBank database was searched to identify the 19 genes in *M. haemolytica* and the closely related species *Actinobacillus actinomycetemcomitans*, *Haemophilus ducreyi*, *Haemophilus influenzae*, and *Pasteurella multocida*. The available gene sequences from each of the above species were aligned with the Lasergene Megalign (DNA star, Inc.) software application, and conserved regions were identified and used to design two pairs of universal primers (two forward and two reverse) using the computer program Primer Designer (version 2.0). Primers were synthesized by Sigma-Genosys (Cambridge, UK) and Hybaid Limited (Middlesex, UK).

The primers were tested in four combinations for PCR amplification with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair amplifying the largest fragment without non-specific bands was used for further PCRs. The successful PCR primer pair was also used for the first stage of sequencing and additional primers (forward and reverse) were designed as sequence data became available.

2.3.1.3 PCR amplification

2.3.1.3.1 Preparation of template DNA

Template DNA was prepared with the InstaGene matrix kit (Bio-Rad) according to the manufacturer's instructions. One millilitre of overnight broth culture was centrifuged for 1 min at 13,000 × g, the supernatant removed, and the bacterial pellet resuspended in 1 ml of sterile dH_2O. The sample was centrifuged again for 1 min at 13,000 × g and the supernatant was removed. The pellet was resuspended in 200 µl of

InstaGene matrix and incubated for 30 min at 56 °C. The mixture was then vortexed at high speed for 10 s and heated in a boiling waterbath for 8 min. After allowing the tubes to cool for 2 to 3 min, the samples were vortexed at high speed for 10 s and centrifuged for 3 min at 13,000 × g. The DNA samples were stored at -85 °C.

2.3.1.3.2 Components of PCR

PCRs were carried out in 0.5 ml Gene Amp reaction tubes (Applied Biosystems). Each PCR reaction mixture contained 32 µl of sterile dH_2O, 5 µl of 10x PCR buffer (100 mM Tris-HCl , 500 mM KCl, 15 mM $MgCl_2$, pH 8.3), 4 µl of dNTPs (2.5 mM dATP, dTTP, dCTP, and dGTP) (Pharmacia Biotech), 4 µl of each primer (12.5 pmol), 1 µl of template DNA, and 0.2 µl (1 unit) of Taq DNA polymerase (Roche, UK) in a final volume of 50 µl. One drop of liquid paraffin was used to overlay each reaction.

2.3.1.3.3 PCR conditions

PCRs were carried out in a Perkin-Elmer 480 DNA thermal cycler. For each new primer set, the following standard PCR conditions were used first: 30 cycles of denaturation at 94 °C for 45 s, annealing at 55 °C for 45 s, and extension at 72 °C for 2 min with a final extension step of 72 °C for 10 min. However, if non-specific amplimers were present on the agarose gel, the optimum annealing temperature was determined experimentally for each primer set.

2.3.1.3.4 Agarose gel electrophoresis

Five microlitres of the PCR mixture were mixed with 1.2 µl of 10x loading dye (Invitrogen) and loaded onto a 1 % (w/v) agarose (Invitrogen) gel in 1x TAE buffer

(20 mM Trizma, 1mM EDTA, pH8.0) containing ethidium bromide (0.5 µg/ml). 5 µl of a 1 kb ladder (Promega) [50 ng/µl] were loaded onto one or two wells. The samples were electrophoresed in a horizontal submarine electrophoresis apparatus (Owl, USA) at 80 V for 1.2 h and the amplimers were visualised and photographed under UV light.

2.3.1.4 DNA sequencing

2.3.1.4.1 Purification of PCR products

PCR products were purified with the QIAquick PCR purification kit (Qiagen, Chatsworth, Calif.) following the manufacturer's protocol. Two hundred and fifty microlitres of buffer PB were mixed with 45 µl of PCR mixture and transferred to a QIAquick spin column. The columns were centrifuged for 1 min at 13,000 × g and the supernatants discarded. The bound DNA was washed by adding 750 µl of buffer PE and centrifuging for 1 min at 13,000 × g. The columns were centrifuged for an additional 1 min 13,000 × g. The DNA was eluted from the column by adding 30 µl dH_2O and, after 1 min of standing, centrifuging for 1 min at 13,000 × g. The effectiveness of purification and approximate DNA concentrations were determined by running 1 µl of purified sample on an agarose gel. For sequencing, the purified DNA was diluted by adding 10 µl sample to 50 µl dH_2O. The DNA samples were stored at -20°C.

2.3.1.4.2 Components of sequencing

The sequence reactions were carried out in 0.2 ml Micro Amp reaction tubes (Applied Biosystems). The reaction mixture contained 4.5 µl of diluted (1:6) purified PCR

product, 1.5 µl of primer (2 pmol/µl), and 4 µl of ABI Prism Big Dye Terminator Ready Reaction Mix (PE Applied Biosystems).

2.3.1.4.3 Conditions for sequencing

The sequence reactions were carried out in a GeneAmp PCR system 9600 thermal cycler (Applied Biosystems), using 30 cycles of denaturation at 96 °C for 10 s, annealing at 50 °C for 5 s, and extension at 60 °C for 4 min.

2.3.1.4.4 Ethanol precipitation

After cycle sequencing, the samples were ethanol precipitated as follow. Twenty five microlitres of 100 % ethanol and 1.5 µl of 2 M sodium acetate pH 4.5 (Applied Biosystems) were added to each reaction tube. The tubes were vortexed gently for 1 s and left for 15 min at room temperature. The samples were centrifuged for 15 min at 13,000 × g and the supernatants were removed. The pellets were rinsed with 100 µl of 70% ethanol and centrifuged for 5 min at 13,000 × g. The supernatants were removed and the samples centrifuged again briefly. The residual ethanol was carefully removed and the samples were air dried for 10 min. Finally, 4 µl of sequence loading buffer were added to each sample and gently vortexed for 15 min.

2.3.1.5 Sequence analysis

Sequence analysis was performed in an Applied Biosystems 373A DNA sequencer by the University of Glasgow sequencing service.

2.3.1.5.1 Analysis of nucleotide and amino acid sequence data

Nucleotide sequence data were analysed and edited with the SEQED™ computer software (Applied Biosystems; version 1.0.3). Statistical and phylogenetic analyses were carried out with MEGA2.1 (Kumar *et al.*, 2001) in conjunction with an amino acid conversion program (AASEQ), amino acid alignment program (CLUSTALX), and a realignment program (REALIGW). Secondary structure predictions were performed with the Psipred secondary structure prediction method (Jones, 1999) (http://bioinf.cs.ucl.ac.uk/psipred/).

For the *ompA* gene (section 3.1.1.5.1), the *M. haemolytica* and *M. glucosida* OmpA sequences were aligned and compared with the *E. coli* OmpA three-dimensional structural models (MMDB 16249 and PDB 1G90 (Arora *et al.*, 2001); MMDB 9208 and PDB 1BXW (Pautsch & Schulz, 1998)) using the computer program Cn3D (http://www.ncbi.nlm.nih.gov/structure/CN3D/cn3d.shtml).

For the *tbpB* and *tbpA* genes (section 3.1.1.5.2), the numbers of nucleotide polymorphic sites were computed using the program PSFIND (version 3). These statistics were then used to draw a similarity plot (Figure 3.41) using the HAPPLOT program (version 4).

The programs AASEQ, REALIGW, CLUSTALX, PSFIND, and HAPPLOT were kindly provided by T. S. Whittam (Michigan State University).

2.3.2 Specific sequencing procedures

Preliminary PCR amplification and sequencing were carried out with 10 diverse *M. haemolytica* isolates (PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, and PH706), one *M. glucosida* isolate (PH344), and one *P. trehalosi* isolate (PH246) for the 19 genes. If the sequences showed significant variation, a larger number of isolates, up to a total of 42 (Table 2.1), were investigated. The following sections contain details of different primer sets and their optimum annealing temperatures specific for the 19 genes. GenBank sequences were used to describe the locations of primers for each gene (see below) but, if appropriate sequences were unavailable, sequences from the unannotated *M. haemolytica* genome sequence (http://www.hgsc.bcm.tmc.edu/microbial/Mhaemolytica/) were used.

2.3.2.1 Genes for encoding DNA repair and recombination enzymes

2.3.2.1.1 RecA recombinase (*recA*)

Initially, two forward (#200 and #201) and two reverse (#202 and #203) primers were designed within the conserved regions of the aligned *recA* sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.2 and 2.3, and Table 2.3). The primers were tested in four combinations (#200/#202, #200/#203, #201/#202, and #201/#203) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #200/#202 was used for PCR and sequencing of the *recA* gene in 10 representative isolates of *M. haemolytica* and one isolate of *M. glucosida* (Figure 2.2). Attempts to amplify *recA* in *P. trehalosi* with the four primer combinations failed. Therefore, additional forward (#387) and reverse (#388) primers were designed and used for the *recA* gene

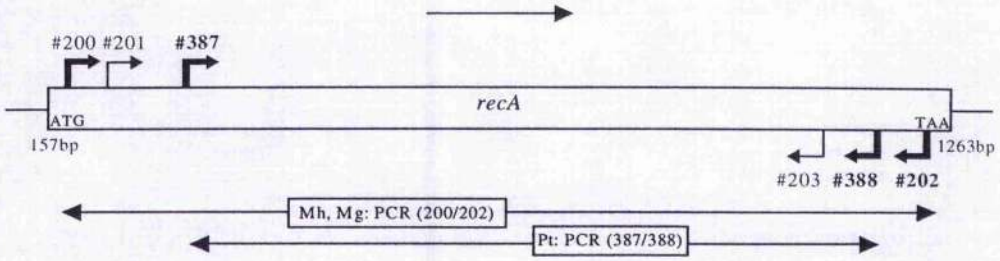


Figure 2.2 Diagrammatic representation of the *recA* gene of *M. haemolytica* taxon 746 (AF176376) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the sequence. The large arrow above *recA* indicates the direction of transcription.

ATGGCAGAGA AAAAAAGTCA AAAGAACACT CCGGTTAAAC AGATTGATCC GGAACAAAAA GAAAAAGCGT TGGCAGCCGC ATTAGCTCAA ATTGAAAAGC
#200 #201
AATTTGGCAA AGGCTCAATT ATGAAGTTAG GCGATACTCA AGCCCTTGAT ATTGAAGCAG TTTCAACCGG TTCTCTTGGT TTAGACTCAG CATTAGGCAT
#387
TGGTGGTTTA CCAATGGGTC GTATTGTGGA AATTTATGGA CCGGAGTCTT CAGGTAAAAC TACATTAACT TTATCTGTGG TGGCTCAAGC CCAAAAAAAT
GGTAAAACTT GTGCCTTTAT TGATGCCGAA CACGCACTTG ATCCTATTTA CGCCCGTAAG TTAGGGGTAG ATACAGATGG ATTATTAATT TCACAACCGG
ATAATGGCGA ACAAGCATTA GAAATTTGTG ATGCGTTAGT TCGTTCCGGT GCGGTTGATG TAATTATTGT GGACTCTGTT GCGGCACTTA CACCAAAAGC
AGAGATTGAA GGCGATATGG GTGATTCCCA TATGGGCTTA CAAGCTCGTT TAATGTCACA AGCATTGCGT AAATTAACCG CAAATATCAA AGCGACCAAC
TGCTTGGTGA TTTTCATTAA CCAAATTCGT ATGAAAATTG GAGTAATGTT CGGTAACCCT GAAACCACAA CTGGTGGTAA TGCCCTTAAA TTCTATGCTT
CAGTACGTTT AGATATTCGT CGTTCAGGCG TAGTAAAAGA CGGTGATGAG GTTATCGGTA GCGAAACAAA AGTGAAAATT GTGAAAAACA AGGTTGCTCC
ACCGTTCCGT GAAGTCCAAT TTGATATTAT GTACGGTGAG GGCATTGCCC GTATGAATGA ATTACTCATT TTAGCGGAAT CTCACGGCTT CATTAATAAA
GCAGGGGCTT GGTTCTCTTA TGAAGGGGAG AAAATCGGTC AGGGTAAAAA TAATGCCATC AAATGGTTAA AAGAACACCC TGAAGTGGCG AGTAAAATTG
#203 #388 #202
AGCAAGATAT TCGAAATCTA CTCATTTCAA ACCCAACTTT TACGGCTACT CCAGATTCTG AAAATGCAGA CAATGCAGAT GATGAATTTA GTGAAGAAGA
ACTCTAA

Figure 2.3 Nucleotide sequence of the *recA* gene of *M. haemolytica* taxon 746 (AF176376) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.3 Details of primers used for PCR amplification and sequencing of the *recA* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
P, S	#200	RecA/F/1	*Mh, Mg*	ACAGATTGATCCSGAACA	39 - 56
	#201	RecA/F/1		AGAAAAAGCRTTRGCAGC	60 - 77
P, S	#387	RecA/F/1	*Pt*	CAATTTGGCAAAGGCTC	100 - 116
P, S	#202	RecA/R/1	*Mh, Mg*	CCACTTCAGGRTKTTCTT	988 - 971
	#203	RecA/R/1		AWGAGAACCAWGCYCCTG	919 - 902
P, S	#388	RecA/R/1	*Pt*	CCYTGACCRATTTTHTC	944 - 928

[a] P: PCR amplification, S: Sequencing

[b] RecA/F/1: RecA/Forward/1, RecA/R/1: RecA/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706

Mg: *M. glucosida* isolate PH344

Pt: *P. trehalosi* isolate PH246

[d] S: G+C, R: G+A, Y: C+T, K: G+T, W: A+T, H: C+T+A

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *recA* of the *M. haemolytica* GenBank sequence (AF176376) (see Figure 2.3)

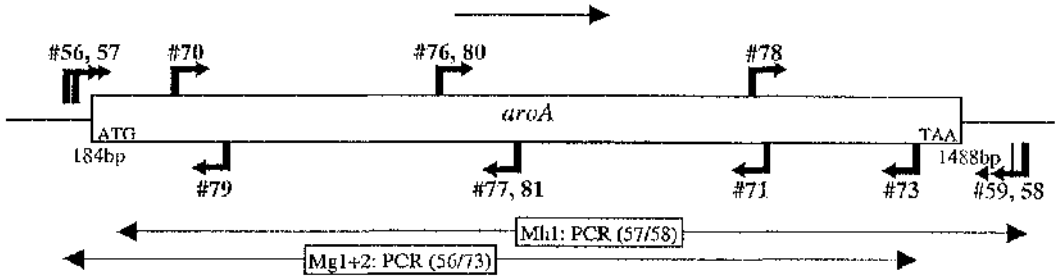
of *P. trehalosi* (Figure 2.2). The optimum annealing temperature for all the PCRs was 55 °C. The positions of all primers on the nucleotide sequence of the *recA* gene of *M. haemolytica* are shown in Figure 2.3 and the details of all primers are listed in Table 2.3.

2.3.2.2 Genes encoding metabolic enzymes

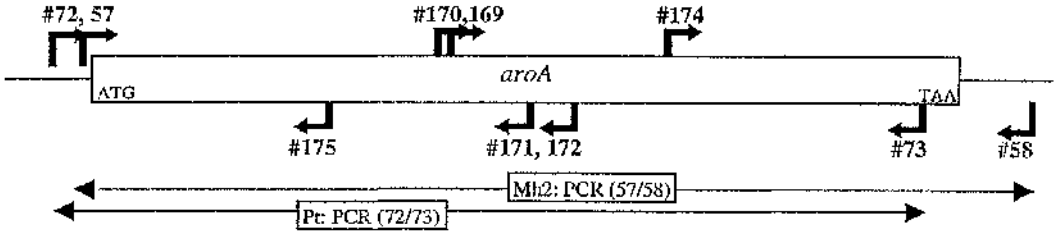
2.3.2.2.1 5-enolpyruvylshikimate-3-phosphate synthase (*aroA*)

Initially, two forward (#56 and #57) and two reverse (#58 and #59) primers were designed within the conserved regions of the aligned *aroA* flanking sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.4 and 2.5, and Table 2.4. The primers were tested in four combinations (#56/#58, #56/#59, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #57/#58 was used for PCR and for the first stage of sequencing of the *aroA* gene in 32 isolates of *M. haemolytica* (Figure 2.4A, B, and C). Attempts to amplify the *aroA* gene of *M. glucosida* and *P. trehalosi* with the four combinations of primers failed. Therefore, additional forward (#72) and reverse (#73) primers were designed. The primer pair #56/#73 was used for PCR and for the first stage of sequencing of the *aroA* gene in six isolates of *M. glucosida* (Figure 2.4A and C), and the primer pair (#72/#73) was used for PCR and for the first stage of sequencing of the *aroA* gene in four isolates of *P. trehalosi* (Figure 2.4B). As sequence data became available, different sequencing primers were designed and used for each group of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Figure 2.4). Four primer pairs were used to sequence *aroA* in *M. haemolytica* group *Mh1* and *M. glucosida* groups *Mg1* and *Mg2* (Figure 2.4A), whereas three primer pairs were used to sequence *aroA* in *M. haemolytica* groups *Mh2* and *Mh3*, *M. glucosida* groups *Mg3*

(A) *M. haemolytica* group *Mh1* and *M. glucosida* groups *Mg1+2*



(B) *M. haemolytica* group *Mh2* and *P. trehalosi* group *Pt*



(C) *M. haemolytica* group *Mh3* and *M. glucosida* groups *Mg3+4*

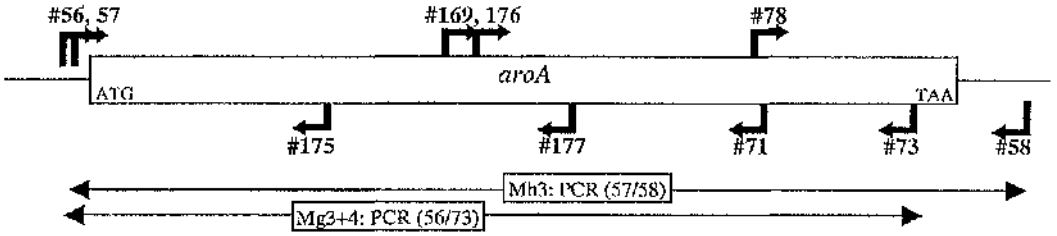


Figure 2.4 Diagrammatic representation of the *aroA* gene of *M. haemolytica* taxon 746 (U03068) showing locations of the PCR and sequencing primers. (A) *M. haemolytica* group *Mh1* and *M. glucosida* groups *Mg1* and 2, (B) *M. haemolytica* group *Mh2* and *P. trehalosi* group *Pt*, and (C) *M. haemolytica* group *Mh*3 and *M. glucosida* groups *Mg3* and *4*. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrow above *aroA* indicates the direction of transcription.

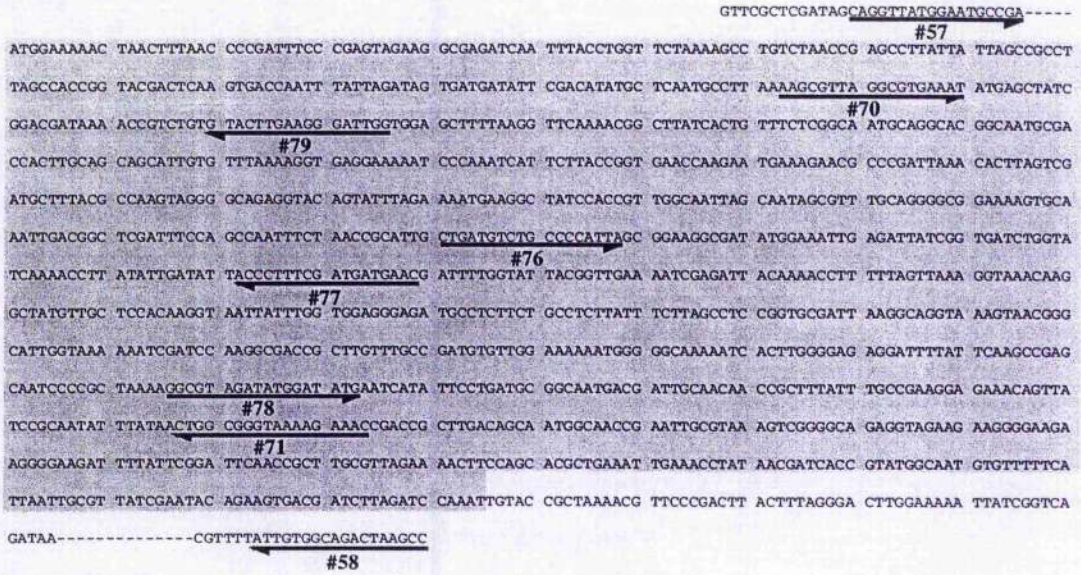


Figure 2.5 Nucleotide sequence of the *aroA* gene of *M. haemolytica* taxon 746 (U03068) showing locations of the PCR and sequencing primers for *M. haemolytica* group *Mh1*. The primers used for PCR and sequencing are in bold type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.4 Details of primers used for PCR amplification and sequencing of the *aroA* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer [a] use	Primer No.	Primer [b] name	Isolates [c]	Sequence (5′ to 3′) [d]	Position [e]
Four stages of sequencing (Figure 2.4A)					
P, S1	#56	AroA/F/1	*Mg1+2*	GCTCGATAGCAGGTTATG	Not shown
P, S1	#57	AroA/F/1	*Mh1*	CAGGGTATGGAATGCCGA	Franking region
S2	#70	AroA/F/2	*Mh1*, *Mg1+2*	AAGCGTTAGGCGTGAAAT	173-190
S3	#76	AroA/F/3	*Mh1*, *Mg1*	CTGATGTCTGCCCCATTA	541-558
S3	#80	AroA/F/3	*Mg2*	CTGATGTCTGCTCCACTA	Not shown
S4	#78	AroA/F/4	*Mh1*, *Mg1+2*	GGCGTAGATATGGATATG	916-933
P, S1	#58	AroA/R/1	*Mh1*, *2+3*	GGCTTAGTCTGCCACAAT	Franking region
	#59	AroA/R/1		GTCTGCCACAATAAAACG	Not shown
P, S1	#73	AroA/R/1	*Mg1+2*	GGAAACGTTTTTGCGGTACA	Not shown
S2	#71	AroA/R/2	*Mh1*, *Mg1+2*	GTTTCTTTTACCCGCCAG	1034-1017
S3	#77	AroA/R/3	*Mh1*, *Mg1*	GTTCATCATCGAAAGGGT	639-622
S3	#81	AroA/R/3	*Mg2*	CCAAAATCTTTCATCATCG	Not shown
S4	#79	AroA/R/4	*Mh1*, *Mg1+2*	CCAATCCCTTCAACTAC	236-220
Three stages of sequencing (Figure 2.4B and C)					
P, S1	#56	AroA/F/1	*Mg3+4*	GCTCGATAGCAGGTTATG	Not shown
P, S1	#57	AroA/F/1	*Mh2+3*	CAGGGTATGGAATGCCGA	Not shown
P, S1	#72	AroA/F/1	*Pt*	GTGCGTCCATTGCAGGTT	Not shown
S2	#169	AroA/F/2	*Mh2+3*, *Mg3+4*	TTAGTCGATGCTTTACGC	Not shown
S2	#170	AroA/F/2	*Pt*	CATTTGGTGGATGCTTTA	Not shown
S3	#78	AroA/F/3	*Mh3*, *Mg3*	GGCGTAGATATGGATATG	Not shown
S3	#174	AroA/F/3	*Mh2*, *Pt*	GGCAAGGTAAAAGTAACG	Not shown
P, S1	#58	AroA/R/1	*Mh2+3*	GGCTTAGTCTGCCACAAT	Not shown
P, S1	#73	AroA/R/1	*Mg3+4*, *Pt*	GGAAACGTTTTTGCGGTACA	Not shown
S2	#71	AroA/R/2	*Mh3*, *Mg3+4*	GTTTCTTTTACCCGCCAG	Not shown
S2	#171	AroA/R/2	*Mh2*	ATCTCCCTCCACCAAATA	Not shown
S2	#172	AroA/R/2	*Pt*	CCTTCCACCAAATAACGA	Not shown
S3	#175	AroA/R/3	*Mh2+3*, *Mg3*, *Pt*,	GGGCGTTCTTTCATTCT	Not shown
S3	#176	AroA/F/3	*Mg4*	ATGAAGGTTATCCGCCGT	Not shown
S3	#177	AroA/R/3	*Mg4*	AAGAGGCAGAAGAGGCAT	Not shown

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing, S3: Third stage of sequencing, S4: Fourth stage of sequencing

[b] AroA/F/1: AroA/Forward/1, AroA/R/1: AroA/Reverse/1

[c] *Mh1*: *M. haemolytica* isolates PH2, PH8, PH66, PH196, PH202, PH232, PH238, PH278, PH284,

PH292, PH296, PH338, PH388, PH396, PH398, PH484, PH494, PH526, PH540, PH550, PH588, PH598, PH706, PH786

Mh2: *M. haemolytica* isolate PH470

Mh3: *M. haemolytica* isolates PH30, PH376, PH346, PH50, PH56, PH372, PH392

Mg1: *M. glucosida* isolate PH290

Mg2: *M. glucosida* isolate PH344

Mg3: *M. glucosida* isolates PH240, PH496, PH498

Mg4: *M. glucosida* isolate PH574

Pt: *P. trehalosi* isolates PH246, PH252, PH254, PH68

[e] Positions of the *aroA* primers for *M. haemolytica* group *Mh1* only are shown. Nucleotide position corresponding to the first 5′ bp of the primer within *aroA* of the *M. haemolytica* GenBank sequence (U03068) (see Figure 2.5)

and *Mg4*, and *P. trehalosi* (Figure2.4B and C). The optimum annealing temperature for all the PCRs was 55 °C. The positions of representative *M. haemolytica* group *Mh1* primers on the nucleotide sequence of the *aroA* gene of *M. haemolytica* are shown in Figure 2.5, and details of all primers are listed in Table 2.4.

2.3.2.2.2 Aspartate-semialdehyde dehydrogenase (*asd*)

Since a GenBank sequence was not available and the genome sequence was not published at the time for the *M. haemolytica asd* gene, two forward (#182 and #183) and two reverse (#184 and #185) primers were designed within the conserved regions of the aligned *asd* sequences from closely related species of *M. haemolytica* (see section 2.3.1.2) (Figures 2.6 and 2.7, and Table 2.5). Attempts to amplify *asd* in *M. haemolytica* isolate PH2 with four primer combinations (#182/#184, #182/#185, *etc*) failed, and additional forward (#232) and reverse (#233) primers were designed. The new primers were tested in five combinations (#232/#184, #232/#185, #232/#233, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #183/#233 was used for PCR and sequencing of the *asd* gene in 10 representative isolates of *M. haemolytica* and one isolate of *M. glucosida*, and the primer pair #182/#233 was used for PCR and sequencing of the *asd* gene in one isolate of *P. trehalosi* (Figure 2.6). The optimum annealing temperature for all the PCRs was 55 °C. The positions of all primers on the nucleotide sequence of the *asd* gene of *M. haemolytica* are shown in Figure 2.7, and the details of all primers are listed in Table 2.5.

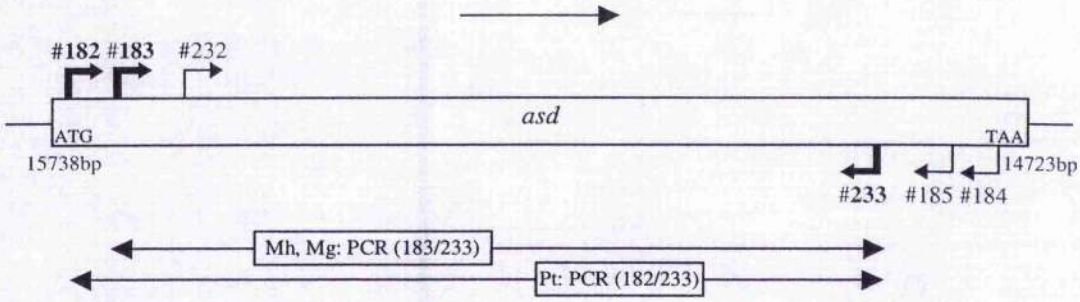


Figure 2.6 Diagrammatic representation of the *asd* gene of *M. haemolytica* isolate PHL213 (complementary contig 138 of the genome sequence as of 2/1/04) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the contig. The large arrow above *asd* indicates the direction of transcription.

Figure 2.7 Nucleotide sequence of the *asd* gene of *M. haemolytica* isolate PHL213 (complementary contig 138 of the genome sequence as of 2/1/04) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.5 Details of primers used for PCR amplification and sequencing of the *asd* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
P, S	#182	Asd/F/1	*Pt*	GGYTTTATYGGYTGGCG	13 - 29
P, S	#183	Asd/F/1	*Mh*, *Mg*	CMGTHTTRATGGATCGTA	44 - 61
	#232	Asd/F/1		CATTATCGTSACBTGCCA	204 - 221
	#184	Asd/R/1		ARTAATTGRTCGCCMAC	1065 - 1048
	#185	Asd/R/1		GMYAAATATTCMGGYCC	1037 - 1021
P, S	#233	Asd/R/1	*Mh*, *Mg*, *Pt*	GGRATCACTTTYACCCA	926 - 910

[a] P: PCR amplification, S: Sequencing

[b] Asd/F/1: Asd/Forward/1, Asd/R/1: Asd/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706

Mg: *M. glucosida* isolate PH344

Pt: *P. trehalosi* isolate PH246

[d] M: C+A, H: C+T+A, R: G+A, Y: C+T

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *asd* of contig 138 of the *M. haemolytica* genome sequence as of 01/02/04

2.3.2.2.3 UDP-galactose 4-epimerase (*galE*)

Initially, two forward (#192 and #193) and two reverse (#194 and #195) primers were designed within the conserved regions of the aligned *galE* sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.8 and 2.9, and Table 2.6). The primers were tested in four combinations (#192/#194, #192/#195, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #193/#194 was used for PCR and for the first stage of sequencing of *galE* gene in 10 representative isolates of *M. haemolytica*, one isolate of *M. glucosida*, and one isolate of *P. trehalosi* (Figure 2.8). The optimum annealing temperature for all the PCRs was 55 °C. For the second stage of sequencing, primers specific to individual isolates were designed as sequence data became available. The positions of all primers on the nucleotide sequence of the *galE* gene of *M. haemolytica* are shown in Figure 2.9 and the details of all primers are listed in Table 2.6.

2.3.2.2.4 Glyceraldehyde-3-phosphate dehydrogenase (*gap*)

Since a GenBank sequence was not available and the genome sequence was not published at the time for the *M. haemolytica gap* gene, two forward (#64 and #65) and two reverse (#66 and #67) primers were designed within the conserved regions of the aligned *gap* sequences from closely related species of *M. haemolytica* (see section 2.3.1.2) (Figures 2.10 and 2.11, and Table 2.7). The primers were tested in four combinations (#64/#66, #64/#67, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #65/#66 was used for PCR and for the first stage of sequencing of the *gap* gene in 32 isolates of *M. haemolytica*, six isolates of *M. glucosida*, and four isolates of

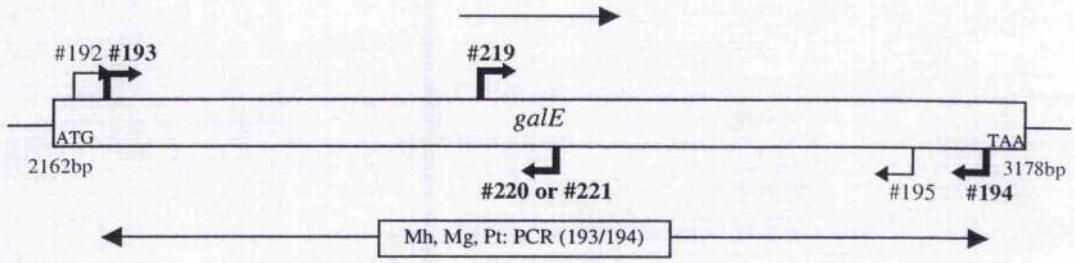


Figure 2.8 Diagrammatic representation of the *galE* gene of *M. haemolytica* isolate taxon 746 (U39043) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrow above *galE* indicates the direction of transcription.

ATGGCAATTT TAGTTACAGG TGGTGCAGGC TACATCGGTT CACATACATT AGTTGAATTA TTAAACGAAA ATCGTGAAAT CGTGGTGTTA GATAATCTTT
192 # 193
CTAATTCCTC TGAAGTATCA CTTGAGCGTG TGAAGCAGAT CACCGGCAAG AGCGTAAAAT TCTATCAAGG CGATATTTTA GATCGTGATA TTTTACGCAA
AATTTTTGCG GAAAATCAGA TTGAATCGGT AATTCACTTT GCCGGTTTAA AAGCTGTAGG CGAAACGTCA GAGAACCGCT ACGTTACTAT CAAAAACAAT
GTAACCGGCT CGATTGTATT GGTAGAAGAG ATGCTAAAAG CCAATGTGAA TACCATTGTG TTTAGTTCAT CGGCGACCGT TTATGGTGAT CCGCAGATTA
TCCCGATTGT GGAATCTTGC CCTGTCGGTG GCACAACCAA CCCTTACGGT ACGTCCAAAT ATATGGTTGA ACGCATTCTT GAAGATACCG TTAAAGCCTT
219
CCCACAATTA AGTGCGGTAG TATTGCGTTA CTTCAACCGG GTAGGGGCTC ACGAAAGCGG TTTAATCGGT GAAGATCCAA ACGGTATCCC AAATAACTTA
#220 or 221
ATGCCGTTTA TCAGCCAAGT GGCTGTGGGC AAATTACCTC AATTATCGGT GTTTGGTGGC GATTATAATA CTCACGATGG CACAGGTGTG CGTGATTATA
TCCACGTGGT AGATTTAGCA CTTGGTCACT TAAAAGCCTT AGACAAGCAC CAAAATGATG CCGGTTTCCA CGTTTACAAT TTAGGCACAG GAACAGGTTA
TTCGGTGCTA GATATGGTAA AAGCCTTTGA AGCCGCAAAT GGCATTACTA TTCCATACAA AGTGGTAGAT CGCCGCCCGG GCGATATTGC CGTTTGCTAC
195
TCCGCTCCAC AAAAAGCGTT AGAGCAACTG GGTTGGGAAA CCGAGCGTGG GCTAGAACAA ATGATGAAAG ACACTTGGAA TTGGCAAAAA AATAACCCGA
#194
ATGGTTATAA AGGCTAA

Figure 2.9 Nucleotide sequence of the *galE* gene of *M. haemolytica* isolate taxon 746 (U39043) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.6 Details of primers used for PCR amplification and sequencing of the *galE* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
	#192	GalE/F/1		TTAGTWACAGGYGGKGC	10 - 26
P, S1	#193	GalE/F/1	*Mh*, *Mg*, *Pt*	GGYGGKGCAGGCTAYAT	19 - 35
S2	#219	GalE/F/2	*Mh*, *Mg*, *Pt*	GGCACAACCAACCCTTAC	430 - 447
P, S1	#194	GalE/R/1	*Mh*, *Mg*, *Pt*	TGCCARTKCCAMGTGTC	986 - 970
	#195	GalE/R/1		GTARCAARYGGCAATATC	900 - 893
S2	#220	GalE/R/2	*Mh*, *Mg*	CTTCACCGATTAAACCGC	574 - 557
S2	#221	GalE/R/2	*Pt*	CTTCACCAATTAAACCGC	574 - 557

[a] P: PCR amplification, S1: First stage of sequencing, S2:Second stage of sequencing

[b] GalE/F/1: GalE/Forward/1, GalE/R/1: GalE/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706

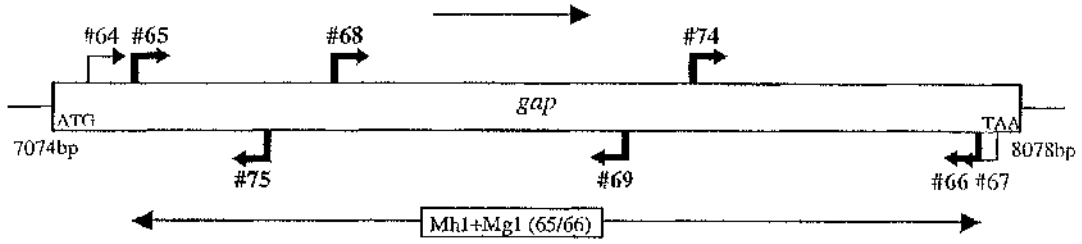
Mg: *M. glucosida* isolate PH344

Pt: *P. trehalosi* isolate PH246

[d] W: A+T, Y: C+T, K: T+G, R: G+A, M: C+A,

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *galE* of the *M. haemolytica* GenBank sequence (U39043)

(A) *M. haemolytica* group *Mh1* and *M. glucosida* group *Mg1*



(B) *M. haemolytica* group *Mh2*, *M. glucosida* group *Mg* 2, and *P. trehalosi* group *Pt*

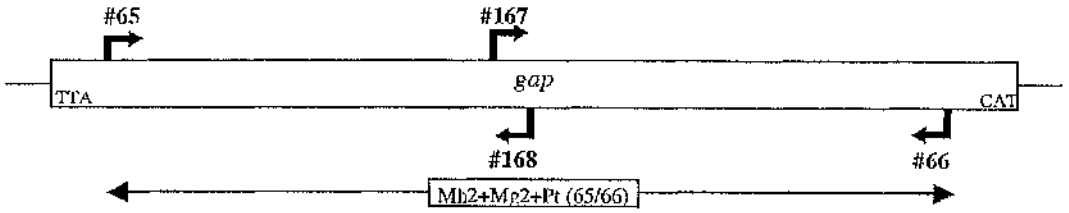


Figure 2.10 Diagrammatic representation of the *gap* gene of ***M. haemolytica*** **isolate PHL213 (contig 156 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers.** (A) *M. haemolytica* group *Mh1* and *M. glucosida* group *Mg1* and (B) *M. haemolytica* group *Mh2*, *M. glucosida* group *Mg2*, and *P. trehalosi* group *Pt*. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the contig. The large arrow above *gap* indicates the direction of transcription.

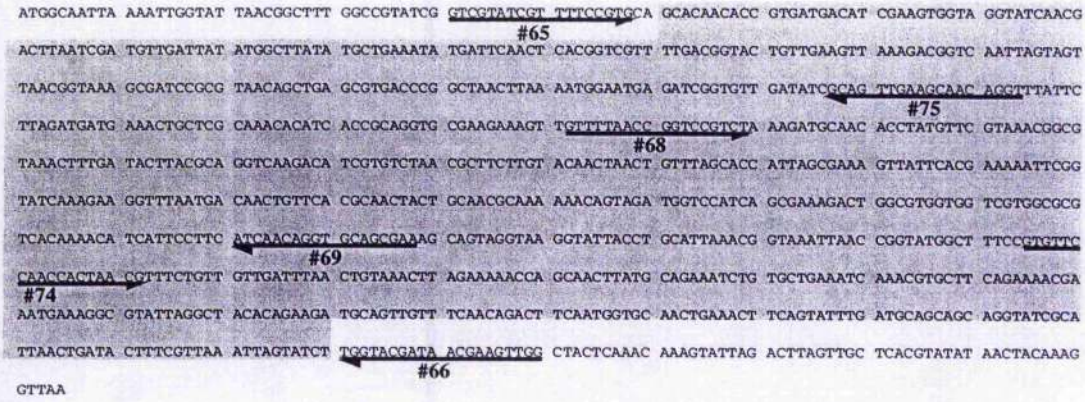


Figure 2.11 Nucleotide sequence of the *gap* gene of *M. haemolytica* isolate PHL213 (contig 156 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers for *M. haemolytica* group *Mh1*. The primers used for PCR and sequencing are in bold type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.7 Details of primers used for PCR amplification and sequencing of the *gap* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Strains[c]	Sequence (5′ to 3′)[d]	Position[e]
Three stages of sequencing (Figure 2.10A)					
	#64	Gap/F/1		ATGGTTTTGGTCGTATCG	Not shown
P, S1	#65	Gap/F/1	*Mh1*, *Mg*1	GCCGTATCGTATTCCGTG	41-58
S2	#68	Gap/F/2	*Mh1*, *Mg*1	GTTTTAACCGGTCCGTCT	352-369
S3	#74	Gap/F/3	*Mh1*, *Mg*1	GTGTTCCAACCACTAACG	695-712
P, S1	#66	Gap/R/1	*Mh1*, *Mg*1	ACCCGTTTGTTATCGTACCA	950-931
	#67	Gap/R/1		GTTTGAGTAACCCGTTTC	Not shown
S2	#69	Gap/R/2	*Mh1*, *Mg*1	TTCGCTGCACCTGTTGAT	638-621
S3	#75	Gap/R/3	*Mh1*, *Mg*1	ACCTGTTGCTTCAACTGC	294-277
Two stage of sequencing (Figure 2.10B)					
P, S1	#65	Gap/F/1	*Mh*2, *Mg*2, *Pt*	GCCGTATCGTATTCCGTG	Not shown
S2	#167	Gap/F/2	*Mh*2, *Mg*2, *Pt*	TGACAACTGTTCACGCAA	Not shown
P, S1	#66	Gap/R/1	*Mh*2, *Mg*2, *Pt*	ACCCGTTTGTTATCGTACCA	Not shown
S2	#168	Gap/R/2	*Mh*2, *Mg*2, *Pt*	GTTGCGTGAACAGTTGTC	Not shown

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing, S3: Third stage of sequencing

[b] GAP/F/1: GAP/Forward/1, GAP/R/1: GAP/Reverse/1

[c] *Mh1*: *M. haemolytica* isolates PH2, PH8, PH66, PH196, PH202, PH232, PH238, PH278, PH284, PH292, PH296, PH338, PH388, PH396, PH398, PH484, PH494, PH526, PH540, PH550, PH588, PH598, PH706, PH786

*Mh*2: *M. haemolytica* isolates PH470, PH30, PH376, PH346, PH50, PH56, PH372, PH392

*Mg*1: *M. glucosida* isolates PH290, PH344

*Mg*2: *M. glucosida* isolates PH240, PH496, PH498, PH574

Pt: *P. trehalosi* isolates PH246, PH252, PH254, PH68

[e] Positions of the *gap* primers for *M. haemolytica* group *Mh1* only are shown. Nucleotide position corresponding to the first 5′ bp of the primer within *gap* of contig 156 of the *M. haemolytica* genome sequence as of 01/02/04 (see Figure 2.11)

P. trehalosi (Figure 2.10). As sequence data became available, different sequencing primers were used for each group of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolate (Figure 2.10). Three primer pairs were used to sequence *gap* in *M. haemolytica* group *Mh1* and *M. glucosida* group *Mg1* (Figure 2.10A), whereas two primer pairs were used to sequence *gap* in *M. haemolytica* groups *Mh2*, *M. glucosida* group *Mg2*, and *P. trehalosi* group *Pt* (Figure 2.10B). The optimum annealing temperature for all the PCRs was 50 °C. The positions of representative *M. haemolytica* group *Mh1* primers on the nucleotide sequence of the *gap* gene of *M. haemolytica* are shown in Figure 2.11, and details of all primers are listed in Table 2.7.

2.3.2.2.5 6-phosphogluconate dehydrogenase (*gnd*)

Since a GenBank sequence was not available and the genome sequence was not published at the time for the of *M. haemolytica gnd* gene, three forward (#238 to #240) and three reverse (#241 to #243) primers were designed within the conserved regions of the aligned *gnd* sequences from closely related species of *M. haemolytica* (see section 2.3.1.2) (Figures 2.12 and 2.13, and Table 2.8). The primers were tested in nine combinations (#238/#241, #238/#242, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #240/#241 was used for PCR and for the first stage of sequencing of the *gnd* gene in 10 representative isolates of *M. haemolytica*, one isolate of *M. glucosida*, and one isolate of *P. trehalosi* (Figure 2.12). The optimum annealing temperature for all the PCRs was 57 °C. For the second stage of sequencing, primers specific to individual isolates were designed as sequence data became available. The positions of all primers on the nucleotide sequence of the *gnd* gene of *M. haemolytica* are shown in Figure 2.13, and the details of all primers are listed in Table 2.8.

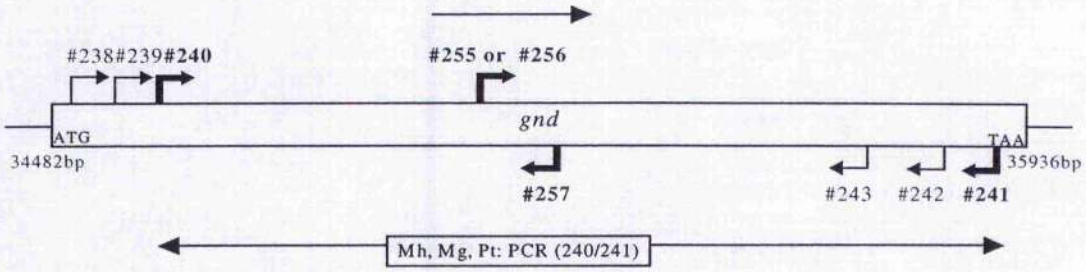


Figure 2.12 Diagrammatic representation of the *gnd* gene of *M. haemolytica* isolate PHL213 (contig 160 of the genome sequence as of 2/1/04) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the contig. The large arrow above *gnd* indicates the direction of transcription.

```
ATGTCACAAA AAGGCGATAT CGGCGTGATT GGACTCGCCG TAATGGGTCA AAATTTAATT TTGAATATGA ATGACAACGG CTTTAAAGTG GTGGCGTACA
                #238       #239       #240
ACCGCACTAC TTCAAAAGTG GATGAGTTCT TAGAAGGTGC GGCGAAAGGC ACGAACATTA TCGGTGCTTA CTCGTTAGAA GATTTAGCGG CGAAGTTGGA

AAAACCGCGT AAAGTGATGC TAATGGTGCG TGCCGGTGAG GTGGTGGATC AGTTTATCGA GGCTCTTCTG CCACATTTAG AAGAAGGCGA CATCATCATT

GATGGCGGTA ACTCAAACTA TCCGGACACC AACCGCCGTG TGAAAGCGTT AGCGGAAAAA GGCATTCGCT TTATCGGCTC AGGTGTATCG GGCGGCGAAG

AGGGTGCTCG TCACGGGCCG TCAATTATGC CGGGCGGTGA TGAATCTGCG TGGCAATATG TGAAGCCGAT TTTCCAAGCA ATTTCAGCCA AAACCGACAA

AGGTGAGCCT TGTTGTGACT GGGTTGGTCG TGACGGAGCA GGGCATTTTG TGAAAATGGT TCACAACGGT ATCGAATATG GCGATATGCA GTTAATTTGT

GAAGCCTACC AATTCTTAAA AGATGGTTTA GGTTTAAGCT ATGATGAAAT GCACGAAATC TTTAAAGAGT GGAAAAACAC CGAATTAGAC AGCTATTTGG
                                                #255 or #256
TGGATATTAC CACTGATATT CTTGCCTACA AAGACGAAGA CGGCGAGCCA TTGGTAGAGA AAATTCTCGA CACCGCAGGT CAAAAAGGAA CGGGTAAATG
                                #257
GACTGGTATT AACGCCCTTG ATTTCGGCAT TCCATTAACA CTGATTACCG AGTCGGTATT CGCCCGTTGC GTATCTGCGT TTAAAGATCA ACGTGTTGCG

GCTTCAAGAT TGTTTAACAA AGAAATTGGT AAAGTTGAAG GCGACAAAAA AGTATGGGTA GAAGCGGTAC GCCGTGCGTT ACTTGCTTCC AAAATCATCT

CTTATGCACA AGGCTTTATG TTGATTCGTG AGGCTTCTGA GCAATTCGGT TGGGACATCA ACTACGGCAA TACCGCCCTA TTATGGCGTG AAGGTTGTAT
                                                                                                #243
TATCCGCAGC CGTTTCTTAG GCAACATTCG TGATGCGTAT GAAGCCAACC CATATTTAGT GTTCTTAGGC TCAGACCCAT ACTTCAAAGA TATTTTGGAA

AACTGCTTGG CAGATTGGCG TAAAGTAGTG GCAAAATCGG TGGAAATCGG TTTACCTGTA CCTTGTATGG CGTCAGCCAT TACCTTCTTA GATGGCTACA
                #242
CCTCGGCACG TTTACCGGCA AACTTACTCC AAGCACAACG TGACTACTTT GGTGCTCATA CTTATGAGCG TACAGACAAA GCTCGCGGTG AGTTCTTCCA
                                #241
CACCAACTGG ACAGGACGTG GCGGTAATAC CGCATCGACC ACTTATGATG TGTAA
```

Figure 2.13 Nucleotide sequence of the *gnd* gene of *M. haemolytica* isolate PHL213 (contig 160 of the genome sequence as of 2/1/04) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.8 Details of primers used for PCR amplification and sequencing of the *gnd* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
	#238	Gnd/F/1		AAGGYGAYATCGGYGTTA	11 - 28
	#239	Gnd/F/1		TCGGYGTTATYGGCTTAG	20 - 37
P, S1	#240	Gnd/F/1	*Mh*, *Mg*, *Pt*	GGCTTAGCSGTRATGGG	31 - 47
S2	#255	Gnd/F/2	*Mh*, *Mg*	GCTATGATGAAATGCACG	638 - 655
S2	#256	Gnd/F/2	*Pt*	GCTACGATGAAATGCAAG	638 - 655
P, S1	#241	Gnd/R/1	*Mh*, *Mg*, *Pt*	RTAGTCRCGTTGTGCTTG	1347 -1330
	#242	Gnd/R/1		GCCACYACTTTRCGCCA	1232 - 1216
	#243	Gnd/R/1		ATACAACCTTCACGCCA	1100 - 1084
S2	#257	Gnd/R/2	*Mh*, *Mg*, *Pt*	CCGTCTTCGTCTTTGTA	733 - 727

[a] P: PCR amplification, S1: Fist stage of Sequencing, S2: Second stage of sequencing

[b] Gnd/F/1: Gnd/Forward/1, Gnd/R/1: Gnd/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706

Mg: *M. glucosida* isolate PH344

Pt: *P. trehalosi* isolate PH246

[d] R: G+A, S: G+C, Y: C+T

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *gnd* of contig 160 of the *M. haemolytica* genome sequence as of 01/02/04 (see Figure 2.13)

2.3.2.2.6 Glucose-6-phosphate-1-dehydrogenase (*g6pd*)

Since a GenBank sequence was not available and the genome sequence was not published at the time for the *M. haemolytica g6pd* gene, three forward (#206 to #208) and two reverse (#290 and #210) primers were designed within the conserved regions of the aligned *g6pd* sequences from closely related species of *M. haemolytica* (see section 2.3.1.2) (Figures 2.14 and 2.15, and Table 2.9). The primers were tested in six combinations (#206/#290, #206/#210, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #207/#210 was used for PCR and for the first stage of sequencing of the *g6pd* gene in 10 representative isolates of *M. haemolytica* (Figure 2.14). Attempts to amplify *g6pd* in *P. trehalosi* and *M. glucosida* with the six primer combinations failed. Therefore, additional forward (#385 and #395) and reverse (#386 and #396) primers were designed, but no combination of any primer pair amplified the *g6pd* gene of *M. glucosida* and *P. trehalosi* (Figure 2.14). The optimum annealing temperature for all the PCRs was 54 °C. For the second stage of sequencing, primers specific to individual isolates were designed as sequence data became available. The positions of all primers on the nucleotide sequence of the *g6pd* gene of *M. haemolytica* are shown in Figure 2.15, and the details of all primers are listed in Table 2.9.

2.3.2.2.7 Malate dehydrogenase (*mdh*)

Since a GenBank sequence was not available and the genome sequence was not published at the time for the *M. haemolytica mdh* gene, three forward (#186, #187, and #191) and three reverse (#188, #189, and #190) primers were designed within the conserved regions of the aligned *mdh* sequences from closely related species of *M. haemolytica* (see section 2.3.1.2) (Figures 2.16 and 2.17, and Table 2.10). The

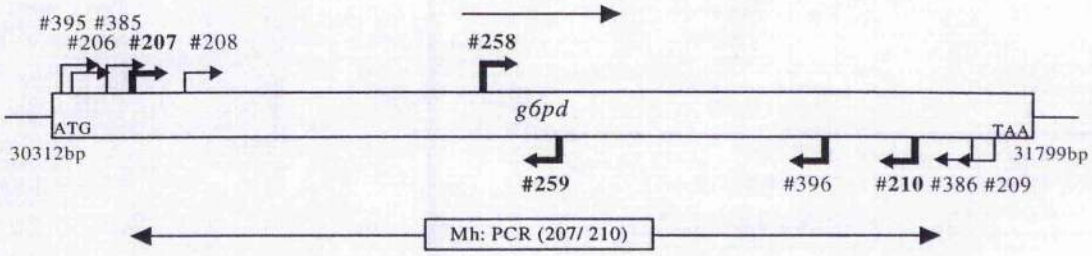


Figure 2.14 Diagrammatic representation of the *g6pd* gene of *M. haemolytica* isolate PHL213 (contig 160 of the genome sequence as of 2/1/04) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the contig. The large arrow above *g6pd* indicates the direction of transcription.

```
ATGAATGCTG AAAATAGCTG TATCGTCATT TTTGGTGCTT CCGGTGACTT AACTTTTCGT AAATTAATTC CTGCTCTTTA TAACTTATAT AAGATTGGCA
                      #395     #206     #385    #207
GATTGGGCGA GCATTTTTCC GTCCTAGGTG TTTCTCGTTC AGAATTAACC GACGAATCTT TCCGCCAAAA AATGCGTGAT GCGTTAGTCA AATTTGAAAA
          #208
AGCAAGCGGT GAAGAATTAG ATAAATTTTG CGAACATCTT TATTACCAAG CTGTAAATAC CTCTGATGCA GTGGATTATG CCAAGTTGCT GCCTCGTCTT
GATGAGCTAC ACGATAAATA TCAAACAGCC GGTAACACGC TTTATTACCT CTCCACGCCA CCAAGTTTAT ATGGCGTGAT TCCGGAATGT TTAGCCGCTC
ACGGCTTAAC TACCGAAGAG TTTGGCTGGA AACGGATTAT TGTTGAAAAA CCATTCGGCT ACGATATCGA AACCGCCAAA AAATTAGACG TACAAATCCA
CAAATGCTTT GAAGAGCATC AAATCTACCG TATCGACCAC TATTTAGGTA AAGAAACGGT ACAAAACTTA CTCGTTTTAC GTTTTTCTAA CGGCTTATTT
GAACCGCTTT GGAACCGTAA TTTTATTGAT TATGTGGAAA TTACCGGGGC AGAATCCATC GGCGTTGAAG ATCGTGGCGG TTACTATGAT GGCTCGGGGG
CAATGCGTGA TATGTTCCAA AACCACTTAT TGCAAGTATT AGCCATGGTT GCAATGGAGC CACCGGCAAT TATTAATGCT AATTCAATGC GTGATGAAGT
#258                                                                                                    #259
CGCAAAAGTG CTATATTGCT TGCATCCGTT AACTGAAGAA GACGTTAAAC ACAATGTGGT GTTAGGTCAG TATGCTCGTG GCACGGTAGA CGGCAAAGAA
GTGCCGGCTT ACTTGGAAGA AAAAGGCGTG CCGGCTGATT CCAACACTGA AACCTTTATG GCGGTAAAAT GTAAAATCGA TAACTGGCGT TGGGCAGGTG
TGCCGTTCTA TGTGCGAACC GGTAAACGCT TACCAACCCG TGTGACGGAA ATCGTAATTC ATTTTAAAAC TACACCACAC CCGGTATTTA GCCAAAATGC
TCCGGAAAAT AAATTGATTA TTCGTGTGCA ACCGGACGAA GGCATTTCAA TGCGTTTCGG CTTGAAAAAA CCGGGAGCAG GCTTTGAAGC GAAAGAAGTC
TCAATGGATT TCCGCTATTC TGATTTAAGT TCATCATCAA GCTTACTCAC CGCTTATGAA CGCTTACTAT TAGACGCACT TAAAGGTGAT GCAACCTTAT
                                                                #396
TTGCCCGTAC TGATGCTGTA CACGCTTGTT GGAAATTTGT GCAACCGATT TTAGCCTACA AAGCTAATCG TGGGCGTGTG TATGAATATG AATCCGGCAC
TTGGGGGCCA ACCGAGGCAG ATAAATTAAT CGCCAGACAC GGCAAAGTAT GGCGAAAACC GTCCGGTACA ATGAAGAAGA AAGTGTAA
                #210                          #386       209
```

Figure 2.15 Nucleotide sequence of the *g6pd* gene of *M. haemolytica* isolate PHL213 (contig 160 of the genome sequence as of 2/1/04) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.9 Details of primers used for PCR amplification and sequencing of the *g6pd* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
	#206	G6pd/F/1		TTTTYGGYGCVTCHGGYG	29 - 46
P, S1	#207	G6pd/F/1	*Mh*	CHGGYGAYYTMACTYATCG	41 - 59
	#208	G6pd/F/1		TGRSTGARVAYTTYTCCG	104 - 121
	#385	G6pd/F/1		TTTTGGYGCDTCHGGKGA	30 - 47
	#395	G6pd/F/1		TATCGTAATTTTCGGCGC	21 - 38
S2	#258	G6pd/F/2	*Mh*	GCAATGCGTGATATGTTC	700 - 717
	#209	G6pd/R/1		GCCACYACTTTRCGCCA	1478 - 1461
P, S1	#210	G6pd/R/1	*Mh*	TTTBGCGATBARTTTRTCRGC	1437 - 1417
	#386	G6pd/R/1		TTWCGCCAYACTTTGCC	1467 - 1451
	#396	G6pd/R/1		TAAVCGYTCATAAGCGG	1266 - 1250
S2	#259	G6pd/R/2	*Mh*	TGCGACTTCATCACGCA	804 - 788

[a] P: PCR amplification, S1: First stage of sequencing, S2:Second stage of sequencing

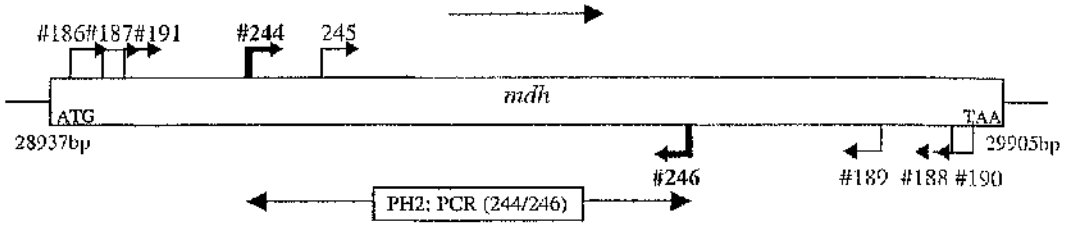
[b] G6pd/F/1: G6pd/Forward/1, G6pd/R/1: G6pd/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706

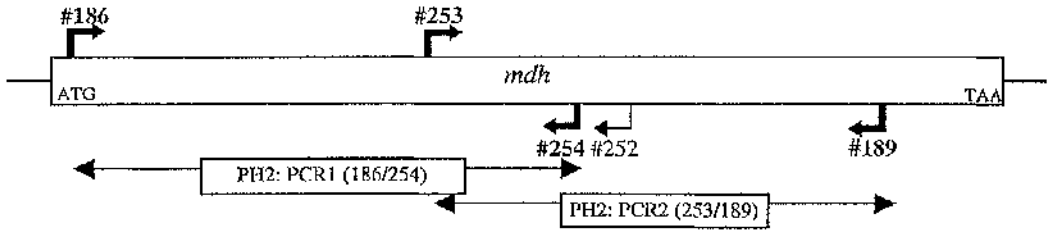
[d] Y: C+T, V: A+C+G, H: C+T+A, M: C+A, R: G+A, S: G+C, D: A+G+T, K: T+G, B: C+T, W: A+T

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *g6pd* of contig 160 of the *M. haemolytica* genome sequence as of 01/02/04 (see Figure 2.15)

(A) *M. haemolytica* PH2



(B) *M. haemolytica* PH2



(C) *M. haemolytica* (*Mh*), *M. glucosida* (*Mg*), and *P. trehalosi* (*Pt*)

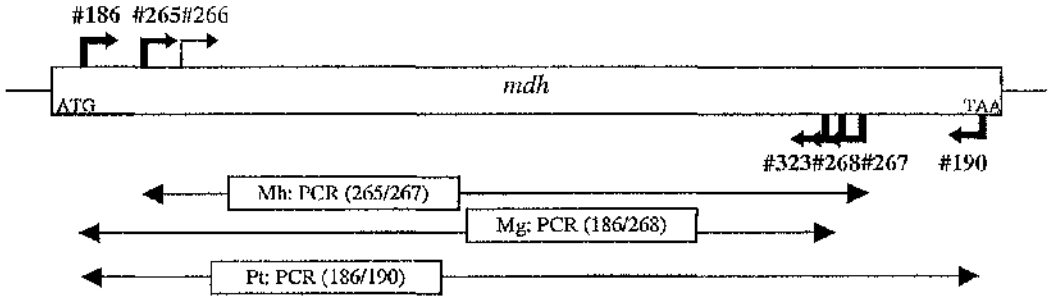


Figure 2.16 Diagrammatic representation of the *mdh* gene of *M. haemolytica* isolate PHL213 (contig 178 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers. (A) PH2, first PCR, (B) PH2, second PCRs, and (C) *Mh*, *Mg* and *Pt*. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the contig. The large arrow above *mdh* indicates the direction of transcription.

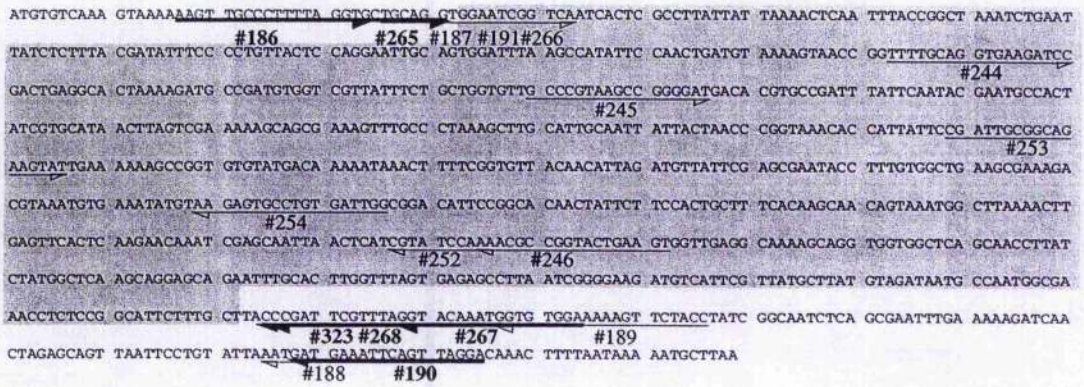


Figure 2.17 Nucleotide sequence of the *mdh* gene of *M. haemolytica* isolate PHL213 (contig 178 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers. The primers selected for final PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.10 Details of primers used for PCR amplification and sequencing of the *mdh* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer [a] use	Primer No.	Primer [b] name	Isolates [c]	Sequence (5′ to 3′) [d]	Position [e]
P, S	#186	Mdh/F/1	*Mh* (PH2), *Mg*, *Pt*	AAGTTGCWGTWYTAGGTG	17 - 34
	#187	Mdh/F/1		CGCAGGYGGTATYGGTCA	36 - 53
	#191	Mdh/F/1		GCAGGTGGKATWGGWCA	37 - 53
P, S	#244	Mdh/F/1	*Mh* (PH2)	TWTKCBGGTGAAGATCC	183 - 200
	#245	Mdh/F/1		GCRCGTAAACCKGGTAT	250 - 266
P, S	#253	Mdh/F/1		CTTGTACCRCCTACTGC	389 - 406
P, S	#265	Mdh/F/1	*Mh*	TGTWYTAGGTGCYGCAGG	24 - 41
	#266	Mdh/F/1		GCAGGYGGWATYGGTCA	37 - 53
	#188	Mdh/R/1		CCTAATTCAATATCYGCACG	944 - 925
P. S	#189	Mdh/R/1		GGTARRATTTCTTCHACRCC	866 - 847
P, S	#190	Mdh/R/1	*Pt*	RCCYAATTCAATATCYGC	945 - 928
P, S	#246	Mdh/R/1	*Mh* (PH2)	ACTTCBGTVCCBGCATTT	662 - 645
P, S	#267	Mdh/R/1	*Mh*	TCCACRCCATTTKTACC	854 - 838
	#252	Mdh/R/1		CCGGCATTTTGGATACG	653 - 637
P, S	#254	Mdh/R/1	*Mh* (PH2)	CCAATCACAGGCACTCTT	
P, S	#268	Mdh/R/1	*Mg*	KTACCTAARCGAAYCGG	842 - 826
S	#323	Mdh/R/1	*Pt*	TTACCTAAACGCACAGGC	842 - 825

[a] P: PCR amplification, S: Sequencing

[b] Mdh/F/1: Mdh/Forward/1, Mdh/R/1: Mdh/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706

Mg: *M. glucosida* isolate PH344

Pt: *P. trehalosi* isolate PH246

[d] W: A+T, Y: C+T, K: T+G, B: C+T, R: G+A, V: A+C+G

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *mdh* of contig 178 of the *M. haemolytica* genome sequence as of 01/02/04 (see Figure 2.17)

primers were tested in nine combinations (#186/#188, #186/#189, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #186/#189 produced a band of the expected size, but also a strong non-specific band. Therefore, additional forward (#244 and #245) and reverse (#246) primers were designed. The primer pair #244/#246 was amplified a small segment of *mdh* for isolate PH2 (Figure 2.16A). The amplified product was sequenced and one forward (#253) and two reverse (#252 and #254) internal primers were designed to amplify a larger segment of the *mdh* gene. Two overlapping regions of *mdh* were amplified without non-specific bands and sequenced for isolate PH2 with primer pairs #186/#254 and #253/#189 (Figure 2.16B). Finally, two forward #265 and #266 and two reverse #267 and #268 internal primers were designed. The primer pair #265/#267 was used for PCR and sequencing of the *mdh* gene in 10 representative isolates of *M. haemolytica*, the primer pair #186/#268 was used for PCR and sequencing of the *mdh* gene in one isolate of *M. glucosida*, and the primer pair #186/#190 was used for PCR and sequencing of the *mdh* gene in one isolate of *P. trehalosi* (Figure 2.16C). The internal primer #323 was designated for sequencing of the *mdh* gene in *P. trehalosi* because the sequence quality with the reverse primer #190 was poor. The optimum annealing temperature of the PCRs was 55 °C for *M. haemolytica* and 50 °C for *M. glucosida* and *P. trehalosi*. The positions of all primers on the nucleotide sequence of the *mdh* gene of *M. haemolytica* are shown in Figure 2.17, and the details of all primers are listed in Table 2.10.

2.3.2.2.8 Mannitol-1-phosphate dehydrogenase (*mtlD*)

Since a GenBank sequence was not available and the genome sequence was not published at the time for the of *M. haemolytica mdh* gene, two forward (#362 and #363) and two reverse (#364 and #365) primers were designed within the conserved

regions of the aligned *mtlD* sequences from closely related species of *M. haemolytica* (see section 2.3.1.2) (Figures 2.18 and 2.19, and Table 2.11). The primers were tested in four combinations (#362/#364, #362/#365, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #363/#364 was used for PCR and for the first stage of sequencing of *mtlD* in 17 isolates of *M. haemolytica*, six isolates of *M. glucosida*, and one isolate of *P. trehalosi* (Figure 2.18). The optimum annealing temperature for the PCRs was 55 °C for *M. haemolytica* and *P. trehalosi* and 58 °C for *P. trehalosi*. For the second stage of sequencing, primers specific to individual isolates were designed as sequence data became available. The positions of all primers on the nucleotide sequence of the *mtlD* gene of *M. haemolytica* are shown in Figure 2.19, and the details of all primers are listed in Table 2.11.

2.3.2.2.9 Phosphomannomutase (*pmm*)

Since a GenBank sequence was not available and the genome sequence was not published at the time for the of *M. haemolytica pmm* gene, two forward (#215 and #216) and two reverse (#217 and #218) primers were designed within the conserved regions of the aligned *pmm* sequences from closely related species of *M. haemolytica* (see section 2.3.1.2) (Figures 2.20 and 2.21, and Table 2.12). The primers were tested in four combinations (#215/#217, #215/#218, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #216/#218 was used for PCR and sequencing of the *pmm* gene in 30 isolates of *M. haemolytica* group *Mh1* and four isolates of *P. trehalosi*, and the primer pair #215/#218 was used for PCR and sequencing of the *pmm* gene in two isolates of *M. haemolytica* group *Mh2* and six isolates of *M. glucosida* (Figure 2.20). The optimum annealing temperatures were 50 °C for isolates PH246 and PH254, 52 °C for

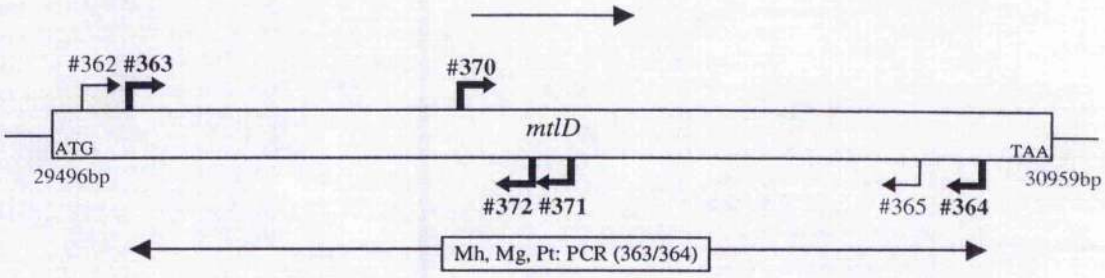


Figure 2.18 Diagrammatic representation of the *mtlD* gene of *M. haemolytica* isolate PHL213 (complementary contig 121 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the contig. The large arrow above *mtlD* indicates the direction of transcription.

```
ATGAACGCAC TCCATTTTGG TGCCGGTAAT ATCGGGCGTG GCTTTATCGG CAAATTGCTT GCTGATGCTC GCATTTTCGT GACTTTTGCC GATATTAACC
                          #362          #363
AAACTCAAAT CGACCAAATC AACCAAAACA AGCAATACGG CGTGAAAATC GTGGGCGATG ACAGCCGTGT GGAAATTGTC AAAAACATTG CGGCAATTAA
CTCGAAAGAC GAAAATGCCG TGATTGAGCA AGTGAAAAAC ACCGATTTAA TTACCACCGC AGTCGGCCCG AATGTGCTTG GCTTTATTGC CCCACTGTTC
GCCAAAGCGT TAGTGGCTCG TGTGGAAAGT GGCAACACCC AACCGCTGAA CATTATCGCC TGTGAAAATA TGGTGCGTGG CACAACTTTC TTTAAAGGTA
AAATTTTTGA GCATTTAACT GCCGAGCAGC AAGCAGAAAT TGAAAAAGTG GTTGGTTTTG TTGATAGTGC GGTGGACAGA ATTGTACCAC CTGCCGAGCC
GAACCCTGCT GATCCGTTGG AAGTCACGGT GGAAGAGTTC AGCGAGTGGA TTGTGGATCA AACCCAGTTC AAAGGTGACA TTCCAAACAT CAAAGGAATG
                                                #370                                #372
GAATTAACCG ATAACCTGAT GGCGTTTGTG GAACGCAAAC TGTTCACCCT GAACACCGGG CATTTAATCT GTGCCTACTT GGGCAAACAA GCGGGCGTGA
          #371
AATGGATTAA AGAGGCGATT GCGATTGAGG AAATCAAAAC TCAGGTGAAA GCCACAATGG AAGAGAGCGG TGCTGTGCTG ATTAAACGTT ACGGCTTCGA
CCCACAAGCC CATTCAGCGT ATATCGAAAA AATCCTCAAA CGTTTTGCCA ACCCGTACTT AAACGATGAC GTAAACCGTG TAGGGCGTGA GCCAATCCGC
AAACTGAGCG AAAACGACCG CTTGATCAAA CCGTTGCGTG GCACGTTGGA ATATGGCTTA CCGTATCAAA ACTTGGTCAA AGGCGTGGTG ATGGCATTGC
                                                #365
AATATCGCAA TGAGGAAGAC CCACAAGCGG TCGAACTTGC CGAATTTATT GCAAATCACG GTGTGGCGGC AGCGGTGGAA AAATACACCG GTTTAACCGA
#364
CCAAGCGGTC ATTGCTCAAG TGGTGGAATT ATATCAATAA
```

Figure 2.19 Nucleotide sequence of the *mtlD* gene of *M. haemolytica* isolate PHL213 (complementary contig 121 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.11 Details of primers used for PCR amplification and sequencing of the *mtlD* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
	#362	MtlD/F/1		TGGTGCCGGTAATATYG	18 - 34
P, S1	#363	MtlD/F/1	*Mh*, *Mg*, *Pt*	GGRCGTGGMTTTATCGG	34 - 50
S2	#370	MtlD/F/2	*Mh*, *Mg*, *Pt*	GGATTGTGGATCAAACCC	548 - 565
P, S1	#364	MtlD/R/1	*Mh*, *Mg*, *Pt*	GRTCWTCCTCATTGCGAT	1021 - 1004
	#365	MtlD/R/1		CCATATTYCAASGTGCCA	956 - 939
S2	#371	MtlD/R/2	*Mh*, *Mg*	GCCATCAGGTTATCGGTT	623 - 606
S2	#372	MtlD/R/2	*Pt*	ATGACGGGAATTTCGCC	590 - 574

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing

[b] MtlD/F/1: MtlD/Forward/1, MtlD/R/1: MtlD/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706 PH540, PH8, PH398, PH232, PH396, PH484, PH598

Mg: *M. glucosida* isolate PH344, PH290, PH240, PH574, PH496, PH498

Pt: *P. trehalosi* isolate PH246

[d] Y: C+T, R: G+A, M: C+A, W: A+T, S: G+C

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *mtlD* of contig 121 of the *M. haemolytica* genome sequence as of 01/02/04 (see Figure 2.19)

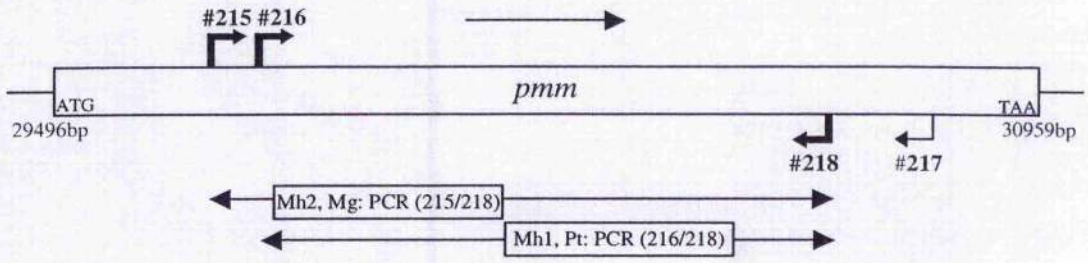


Figure 2.20 Diagrammatic representation of the *pmm* gene of *M. haemolytica* isolate PHL213 (contig 171 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the contig. The large arrow above *pmm* indicates the direction of transcription.

ATGGGTATGA ACCGTGTTTT GGTGGCACAA GCTGCCAGCG GTTTAGCTGG GTTTATTAAA GGTTATGATA AAGAGCCGTC AATTGTTATT GGTTATGACG
GTCGTAAAAA TTCCGATGTG TTTGCCCGTG ACACCGCTGA AATTATGGCA GCGGCAGGCA TTAAAACGTA CCTGCTTCCT CGCAAATTGC CAACACCAGT
GCTTGCCTAT GCGATTCAAT ATTTTGATAC CACAGCCGGT GTGATGGTAA CTGCCAGCCA CAACCCACCG GAAGATAACG GCTATAAAGT TTATTTAGGT
AAAGCCAACG GTGGCGGACA AATTGTTTCG CCGGCAGATA AAGAAATTGC TGCTCTGATT GATAAAGTAG CAGCAGGCGA TATTCGTGAT TTACCTCGTA
GCCAAGATTT CACGGTGTTA GACGATGAAG TGGTAAATGC TTATATTGAG AAAACCGCCT CACTAGCCAA ACGGCCAAAA GCCGAAATAA ACTACGTTTA
CACTGCAATG CACGGTGTGG GCTATGAAGT GTTAAGCAAA ACATTAGAAA AAGCCGGCTT ACCGCAACCG TATTTAGTTA GTGAGCAGAT TCAGCCGGAC
GGTTCGTTCC CAACGGTTAA TTTCCCGAAC CCGGAAGAGA AAGGGGCGTT AGATTTAGCG ATTAAATTGG CGAAAGAGAA AAATGCCGAA TTTATTATCG
CCAACGACCC GGATGCTGAC CGTTTAGCGG TAGCCGTGCC TGATGCTCAA GGCAACTGGA AACCACTGCA TGGCAATGTG ATCGGCTGTT TCTTAGGCTG
GTATTTAGCC AAACAATTCC ACGCACAAGG TAAACAAGGC GTGTTGGCTT GCTCGTTAGT ATCCTCTCCG GCGTTAGCTG AAATTGCGAA AAAATATGGT
TTAAGCTCGG AAGAAACCTT AACCGGCTTT AAATATATCG GAAAAGTTGA AAACCTATTA TTTGGCTTTG AAGAAGCCTT AGGTTATTTG GTTGACCCAG
ACAAAGTGCG TGATAAAGAC GGCATTTCCG CTGCCATTAT GTTCTTAGAT TTAGTGTGCA GCCTCAAACA AGAAGGTAAA ACCCTTGCGG ATTACACCAC
CGAGTTTGTA CAAGAGTTCG GGGCTTATGT AAGCGGTCAA ATTTCGATCA GAGTTTGTGA TTTAGCCGAA ATCGGCAAAT TAATGACGGC GTTACGCAAC
AATCCTCCAA GCGAAATTGG TGGCTTTAAA GTGGCTGAGT TTATTGACCA TACCAAAACC ACACGCCAAA ATGACATTTT AGTGTTTGTG TTGGAAAATG
GCAGCCGTTT GATTGCTCGC CCTTCAGGTA CTGAGCCGAA AATTAAATTC TATTTAGATG CTCGTGGCAC AGACGCTGAA AACGCAGAGC AAGTGCTGGC
TCAGTTTGAC GAAAGTGTGC GTTCGCTGCT TCGCCAAGAA GAATACGGCA AACAAGATTG CTAA

Figure 2.21 Nucleotide sequence of the *pmm* gene of *M. haemolytica* isolate PHL213 (contig 171 of the genome sequence as of 22/4/04) showing locations of the PCR and sequencing primers for *M. haemolytica* group *Mh1*. The primers used for PCR and sequencing are in bold type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.12 Details of primers used for PCR amplification and sequencing of the *pmm* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
P, S	#215	Pmm/F/1	*Mh2*, *Mg*	ACMGCAAGYCATAACCC	Not shown
P, S	#216	Pmm/F/1	*Mh1*, *Pt*	CCCRCCAGAAGATAATG	264 - 281
	#217	Pmm/R/1		CTTTGRCGRTCKGTTTT	Not shown
P, S	#218	Pmm/R/1	*Mh1+2*, *Mg*, *Pt*	GCWGARATRCCGTCTTTAT	1031 - 1013

[a] P: PCR amplification, S: Sequencing

[b] Pmm/F/1: Pmm/Forward/1, Pmm/R/1: Pmm/Reverse/1

[c] *Mh1*: *M. haemolytica* isolates PH2, PH30, PH50, PH56, PH66, PH196, PH202, PH232, PH238, PH278, PH284, PH292, PH296, PH338, PH346, PH372, PH376, PH388, PH392, PH396, PH398, PH470, PH484, PH526, PH540, PH598, PH588, PH706, PH8, PH202,

Mh2: *M. haemolytica* isolates PH494, PH550

Mg : *M. glucosida* isolates PH344, PH240, PH290, PH496, PH498, PH574

Pt : *P. trehalosi* isolates PH246, PH68, PH252, PH254

[d] M: C+A, Y: C+T, R: G+A, K: T+G, W: A+T

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *pmm* of contig 171 of the *M. haemolytica* genome sequence as of 01/02/04 (Figure 2.21)

isolate PH68, 54 °C for isolate PH252, 58 °C for isolates PH56 and PH232, 60 °C for isolate PH290, and 56 °C for the remaining 35 isolates. The positions of all primers on the nucleotide sequence of the *pmm* gene of *M. haemolytica* are shown in Figure 2.21, and details of all primers are listed in Table 2.12.

2.3.2.3 Genes encoding secreted proteins

2.3.2.3.1 Glycoprotease (*gcp*)

Initially, two forward (#196 and #197) and two reverse (#198 and #199) primers were designed within the conserved regions of the aligned *gcp* sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.22 and 2.23, and Table 2.13). The primers were tested in four combinations (#196/#198, #196/#199, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #196/#199 was used for PCR and for the first stage of sequencing of the *gcp* gene in 10 isolates of *M. haemolytica*, one isolates of *M. glucosida*, and one isolate of *P. trehalosi* (Figure 2.22). The optimum annealing temperature for all the PCRs was 55 °C. The internal forward (#230) and reverse (#231 and #264) primers were designed for second and third stage sequencing of the *gcp* gene in *P. trehalosi* because the quality of original sequence data for *P. trehalosi gcp* was poor. The positions of all primers on the nucleotide sequence of the *gcp* gene of *M. haemolytica* are shown in Figure 2.23, and the details of all primers are listed in Table 2.13.

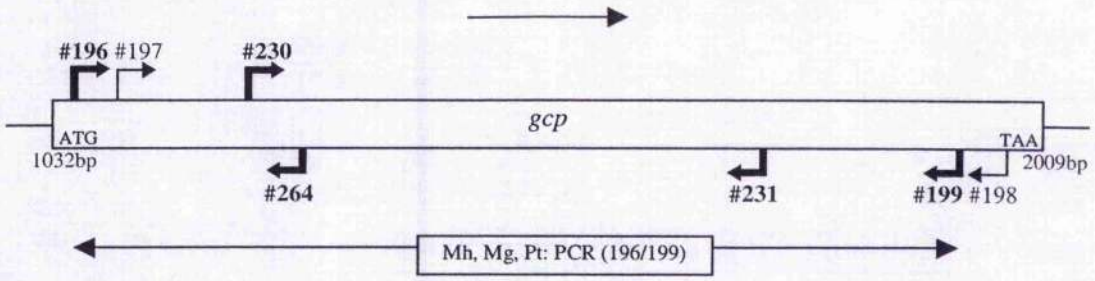


Figure 2.22 Diagrammatic representation of the *gcp* gene of *M. haemolytica* taxon 746 (U15958) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrow above *gcp* indicates the direction of transcription.

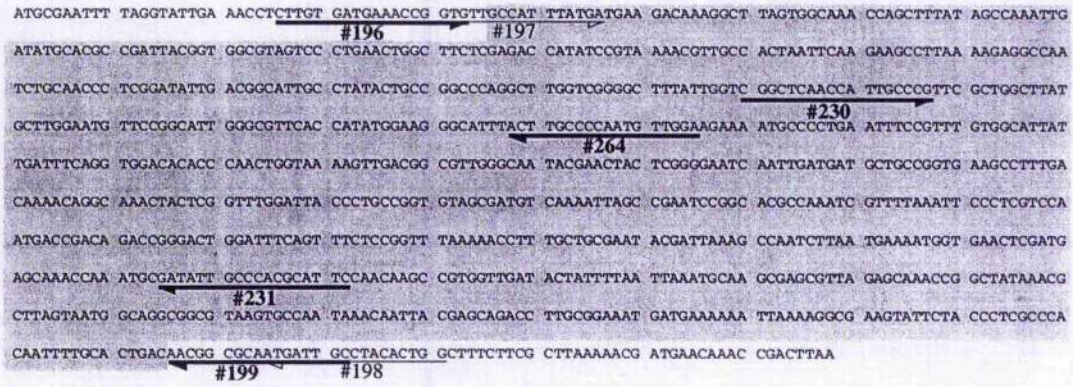


Figure 2.23 Nucleotide sequence of the *gcp* gene of *M. haemolytica* taxon 746 (U15958) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.13 Details of primers used for PCR amplification and sequencing of the *gcp* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
P, S1	#196	Gcp/F/1	*Mh, Mg, Pt*	CHTGTGATGAAACNGGKG	26 - 43
	#197	Gcp/F/1		NGGKGTTGCBATTTATGA	38 - 56
P, S2	#230	Gcp/F/2	*Pt*	AGGATCCACTATTGCTCG	370 - 387
	#198	Gcp/R/1		CCDGTRTARGCAATCAT	941 - 925
P, S1	#199	Gcp/R/1	*Mh, Mg, Pt*	GCAATCATKGCNCCRTT	932 - 916
S2	#231	Gcp/R/2	*Pt*	GAAGGCGTGTGCAATATC	732 - 715
S3	#264	Gcp/R/3	*Pt*	TCTAACATAGGAGCCAGC	365 - 348

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing, S3: Third stage of sequencing

[b] Gcp/F/1: Gcp/Forward/1, Gcp/R/1: Gcp/Reverse/1

[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706

Mg: *M. glucosida* isolate PH344

Pt: *P. trehalosi* isolate PH246

[d] H: C+T+A, N: C+T+G+A, K: T+G, B: C+T, D: A+G+T, R: G+A

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *gcp* of the *M. haemolytica* GenBank sequence (U15958) (see Figure 2.23)

2.3.2.4 Genes encoding periplasm-associated proteins

2.3.2.4.1 Lipoproteins (*plpA*, *plpB*, and *plpC*)

Initially, two forward (#150 and #151) and two reverse (#152 and #153) primers were designed within the conserved regions of the aligned *plpA* sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.24 and 2.25, and Table 2.14). The primers were tested in four combinations (#150/#152, #150/#153, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #150/#152 was used for PCR and sequencing of the *plpA* gene in 10 isolates of *M. haemolytica* and one isolates of *M. glucosida*, and the primer pair #150/#153 was used for PCR and sequencing of the *plpA* gene in one isolate of *P. trehalosi* (Figure 2.24). To amplify the continuous *plpB* and *plpC* genes, two internal forward primers (#247 for *M. haemolytica* and *M. glucosida* and #248 for *P. trehalosi*) were designed as sequence data became available, and three reverse primers (#249 to #251) were designed within the conserved regions of the aligned continuous *plpA*, *plpB*, and *plpC* genes from *M. haemolytica* and closely related species (see section 2.3.1.2). The primers were tested in three combinations (#247/#249, #247/#250, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #247/#250 was used for PCR and for the first stage of sequencing of the *plpB* and *plpC* genes in eight isolates of *M. haemolytica* group *Mh1* and the primer pair #247/#249 was used for PCR and for the first stage of sequencing of the *plpB* and *plpC* genes in two isolates of *M. haemolytica* group *Mh2* and one isolate of *M. glucosida* (Figure 2.24). Attempts to amplify the *plpB* and *plpC* genes in *P. trehalosi* with the three primer combinations (#248/#249, #248/#250, *etc*) were unsuccessful. The optimum annealing temperature for all the PCRs was 55 °C.

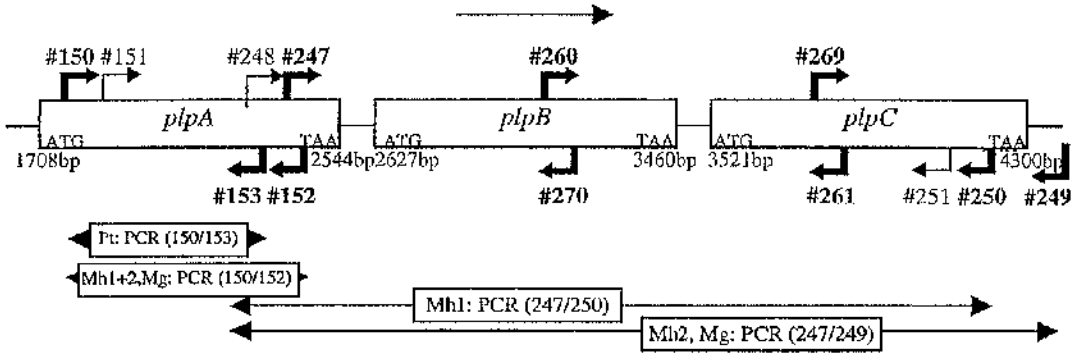


Figure 2.24 Diagrammatic representation of the *plpA*, *plpB*, and *plpC* genes of *M. haemolytica* taxon 746 (L16627) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrow above *plpA*, *plpB*, and *plpC* indicates the direction of transcription.

Figure 2.25 Nucleotide sequence of the *plpA*, *plpB*, and *plpC* genes of *M. haemolytica* taxon 746 (L16627) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type and the unused primers are in thin type. The dashed lines indicate intragenic region (bases are not shown for simplicity) and the shaded area corresponds to the region of sequence that was analysed.

Table 2.14 Details of primers used for PCR amplification and sequencing of the *plpA*, *plpB*, and *plpC* genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
plpA					
P, S	#150	PlpA/F/1	*Mh1*+2, *Mg*, *Pt*	AGTAGGGGTAATGTCAGG	A123 - 140
	#151	PlpA/F/1		ATGTCAGGCCCTGAGCA	A133 - 149
P, S	#152	PlpA/R/1	*Mh1*+2, *Mg*	CCAACCTTTTACTACACC	A834 - 817
P, S	#153	PlpA/R/1	*Pt*	CACTTCTTCGGTTTGGTA	A786 - 769
plpB and *plpC*					
P, S1	#247	PlpBC/F/1	*Mh1*+2, *Mg*	GGCTCGCCAAGACAATAA	A717 - 734
	#248	PlpBC/F/1		CATCATTGTAGCTCGTGA	A708 - 725
S2	#260	PlpBC/F/2	*Mh1*+2, *Mg*	TGTTTGCYACCGAAAACG	B509 - 526
S3	#269	PlpBC/F/3	*Mh1*+2, *Mg*	TTACCGATTATGCCACG	C170 - 186
P, S1	#249	PlpBC/R/1	*Mh1*	GGAGCAGGTCTGTAACAT	Flanking region
P, S1	#250	PlpBC/R/1	*Mh2*, *Mg*	CCAGCCTTTTATCATCGC	C777 - 760
	#251	PlpBC/R/1		GWAKGCTTTCACMAAATC	C714 - 697
S2	#261	PlpBC/R/2		CCGGCACTAATTTATAGC	C289 - 272
S3	#270	PlpBC/R/3	*Mh1*+2, *Mg*	CGGGTTAAAAGTGAGGT	B587 - 571

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing

[b] PlpA/F/1: PlpA/Forward/1, PlpA/R/1: PlpA/Reverse/1,
PlpBC/F/1: PlpBC/Forward/1, PlpBC/R/1: PlpBC/Reverse/1

[c] *Mh1*: *M. haemolytica* isolates PH2, PH66, PH202, PH278, PH296, PH494, PH588, PH706
Mh2: *M. haemolytica* isolates PH196, PH292
Mg: *M. glucosida* isolate PH344
Pt: *P. trehalosi* isolate PH68

[d] Y: C+T, W: A+T, K: T+G, M: C+A

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *plpA* (A), *plpB* (B), and *plpC* (C) of the *M. haemolytica* GenBank sequence (L16627) (see Figure 2.25)

Primers for the second and third stages of sequencing were designed as sequence data became available. The positions of all primers on the nucleotide sequence of the *plpA*, *plpB*, and *plpC* genes of *M. haemolytica* are shown in Figure 2.25, and the details of all primers are listed in Table 2.14.

2.3.2.4.2 Lipoprotein (*plpD*)

The primer pair (#154/#155) was designed within the conserved regions of the aligned *plpD* sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.26 and 2.27, and Table 2.15). This primer pair was used for PCR and sequencing of the *plpD* gene in 11 isolates of *M. haemolytica*, one isolate of *M. glucosida*, and one isolate of *P. trehalosi* (Figure 2.26). The optimum annealing temperature for all the PCRs was 55 °C. The positions of all primers on the nucleotide sequence of the *plpD* gene of *M. haemolytica* are shown in Figure 2.27, and the details of all primers are listed in Table 2.15.

2.3.2.5 Genes encoding outer membrane proteins

2.3.2.5.1 Heat modifiable outer membrane protein (*ompA*)

Initially, two forward (#464 and #465) and two reverse (#466 and #467) primers were designed within the conserved regions of the aligned sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.28 and 2.29, and Table 2.16). The primers were designed in the two flanking genes (Figures 2.28) and preliminary PCR testing was carried out with representative isolates of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Successfully amplified bands were partially sequenced by use of the same primers, and a second set of internal forward

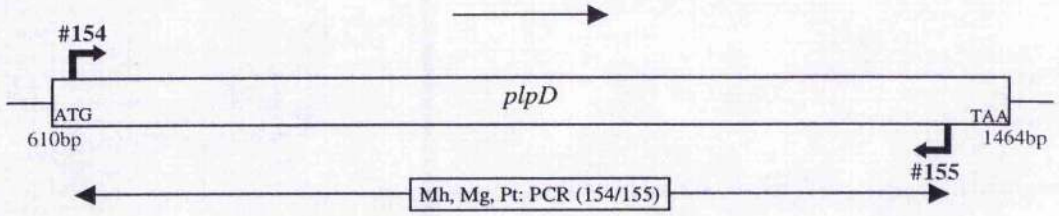


Figure 2.26 Diagrammatic representation of the *plpD* gene of *M. haemolytica* taxon 746 (AF058703) showing locations of the PCR and sequencing primers. The locations of the primers used for PCR and sequencing are indicated by bold arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrow above *plpD* indicates the direction of transcription.

ATGAAAATGG AAAAAAAAGC CCTTTTCCGT GGATTTTTAT TATCGACAGT GGCATTAGCA GTGGCTGCTT GTGGTAACTT AAGTAAAGTG AGTGATGAAG
#154
GAACAACTGA AAACCCGGTA TTCCCGAAAA TCTCAGAATC TGAATTTAAC CACGATGGTT CACAATTTGG TTCATGGCCA AACTGGGAAA ACGTTCGTCA
AATTGAAAAA GGGATGAACA AAGATCAATT ATATAATCTA ATTGGTCGTC CTCATTTTGA AGAAGGCTTA TATGGTGTTC GTGAATGGGA CTATGCATTC
AATTACCGTG AAAATGGCGT ACATAAAATT TGCCAATTTA AAATTTTATT CGATAAAAAT ATGAATGCAC AAAGTTTCTT CTGGTATCCA AATGGATGTA
ATGGAAACTC ATCATTTGCT TTAACTGGCG ATTTCTTATT CGATTTCGAT AAAGCGACTT TAACTACAAA AGGTAAAGAA GTTGTTGATA ATGTTGCTCA
ACAATTAAAA GCATCTAAAG CACAACAAGT TAAAGTAGCA GGCTATACAG ACCGTTTAGG CTCTGCTGCT TATAACTTAG ACTTATCACA GCGTCGTTCT
AATACAGTTA AAGCTCGTTT AGTACAACAA GGCGTTACAG CACAAATTGA AGCGGTAGGT TATGGTAAAG CGAATCAAGT AAAAGCTTGT GATGGTGAAA
CTGGTCAAGC ATTAAAAGAT TGTTTACGCC CTAACCGTCG TGTTGAAATT TCTGCAAATT GTGGTGTAAT GAAACAAAAT GAAGGTGGCA ATGTTGCAGG
#155
TCCTAATGGT CCGGCTCCAC TTTATCAAAC CCCAGCATAT GATGGCAGTA AATAA

Figure 2.27 Nucleotide sequence of the *plpD* gene of *M. haemolytica* taxon 746 (AF058703) showing locations of the PCR and sequencing primers. The primers used for PCR and sequencing are in bold type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.15 Details of primers used for PCR amplification and sequencing of the *plpD* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer[a] use	Primer No.	Primer[b] name	Isolates[c]	Sequence (5′ to 3′)[d]	Position[e]
P, S	#154	PlpD/F/1	*Mh*, *Mg*, *Pt*	AGTGGCTGCTTGTGGTAA	60 - 77
P, S	#155	PlpD/R/1	*Mh*, *Mg*, *Pt*	TTCAACACGACGGTTAGG	747 - 730

[a] P: PCR amplification, S: Sequencing

[b] PlpD/F/1: PlpD/Forward/1, PlpD/R/1: PlpD/Reverse/1

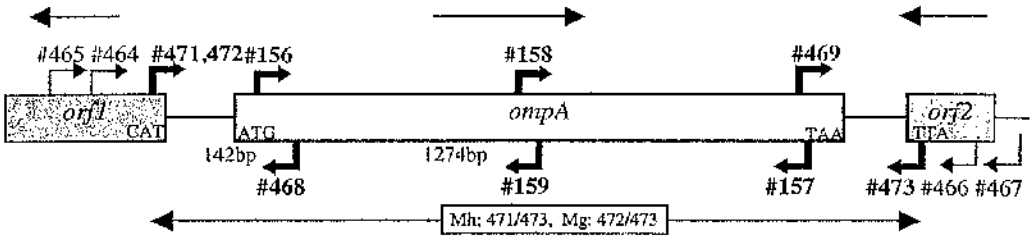
[c] *Mh*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH588, PH706, PH550

Mg: *M. glucosida* isolate PH344

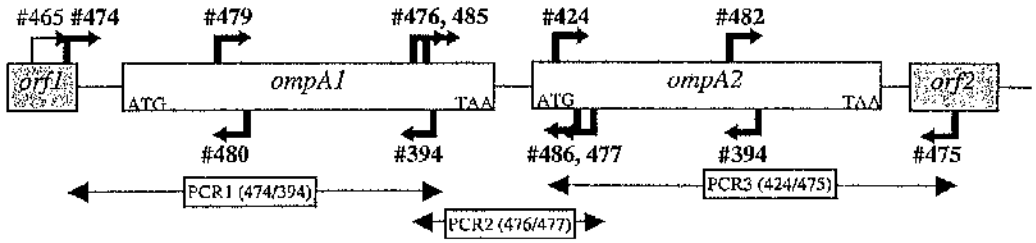
Pt: *P. trehalosi* isolate PH68

[e] Nucleotide position corresponding to the first 5′ bp of the primer within *plpD* of the *M. haemolytica* GenBank sequence (AF058703) (see Figure 2.27)

(A) *M. haemolytica* and *M. glucosida ompA*



(B) *P. trehalosi* group *Pt1 ompA*



(C) *P. trehalosi* group *Pt2 ompA*

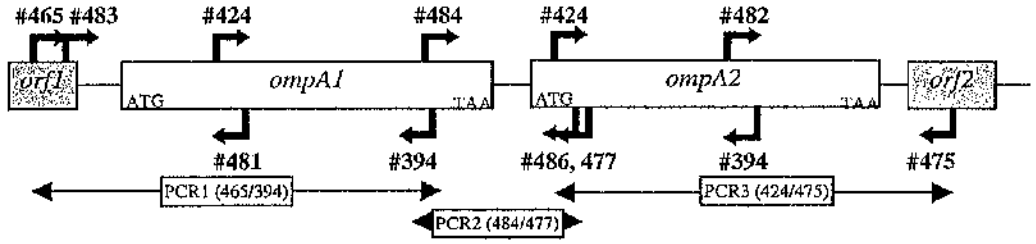


Figure 2.28 Diagrammatic representation of the *ompA* gene of *M. haemolytica* isolate 89010807N (AF133259) showing locations of the PCR and sequencing primers. (A) *M. haemolytica* and *M. glucosida*, (B) *P. trehalosi* group *Pt1*, and (C) *P. trehalosi* group *Pt2*. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrows above *ompA*, *orf1* (FAD-binding protein gene) and *orf2* (hypothetical protein gene) indicates the direction of transcription.

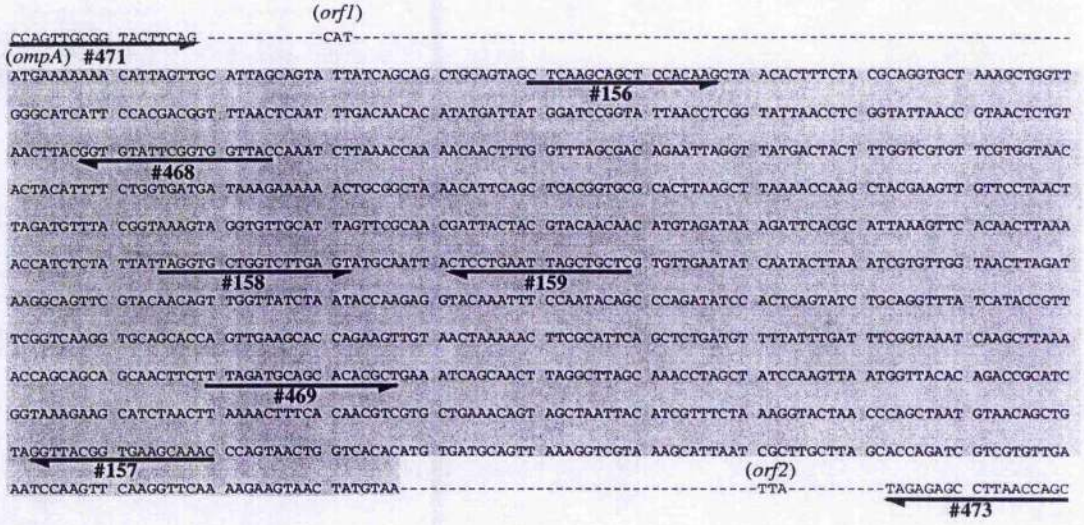


Figure 2.29 Nucleotide sequence of the *ompA* gene of *M. haemolytica* isolate 89010807N (AF133259) showing locations of the PCR and sequencing primers for *M. haemolytica*. The primers used for PCR and sequencing are in bold type. Dashed lines indicate flanking regions of *ompA* gene (nucleotides are not shown for simplicity). The shaded area corresponds to the region of sequence that was analysed.

Table 2.16 Details of primers used for PCR amplification and sequencing of the *ompA* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer [a] use	Primer No.	Primer [b] name	Isolates [c]	Sequence (5′ to 3′) [d]	Position [e]
Mh* and *Mg* *ompA					
	#464	OmpA/F/1		CCRGARGCATCGGTGTT	Not shown
P	#471	OmpA/F/1	*Mh*	CCAGTTGCGGTACTTCAG	Flanking region
P	#472	OmpA/F/1	*Mg*	CCAGTTGAGGTACTTCAG	Not shown
S1	#156	OmpA/F/1	*Mh*, *Mg*	CTCAAGCAGCTCCACAAG	50 - 67
S2	#158	OmpA/F/2	*Mh*, *Mg*	TAGGTGCTGGTCTTGAGT	515 - 532
S3	#469	OmpA/F/3	*Mh*, *Mg*	TTAGATGCAGCACACGCT	820 - 837
	#466	OmpA/R/1		GACGSCTGARCAACGTAA	Not shown
	#467	OmpA/R/1		TGMACATAAYCKCCCCAC	Not shown
P	#473	OmpA/R/1	*Mh*, *Mg*	GCTGGTTAAGGCTCTCTA	Flanking region
S1	#157	OmpA/R/1	*Mh*, *Mg*	GTTTGCTTCACCGTAACC	1020 - 1003
S2	#159	OmpA/R/2	*Mh*, *Mg*	GAGCAGCTAATTCAGGAG	559 - 542
S3	#468	OmpA/R/3	*Mh*, *Mg*	GTAACCACCGAATACACC	225 - 208
Pt1* and *Pt2* segment 1 of *ompA'* and *ompA"					
P	#465	$OmpA_1$/F/1	*Pt2*	ACTCGMACCCAACGYTC	Not shown
P, S1	#474	$OmpA_1$/F/1	*Pt1*	GTATCAGTGGCAAGCGAA	Not shown
S1	#483	$OmpA_1$/F/1	*Pt2*	CTCGGCATAACTATCAGC	Not shown
S2	#479	$OmpA_1$/F/2	*Pt1*	GTGATGGTCCAACTGCTT	Not shown
S2	#424	$OmpA_1$/F/2	*Pt2*	GGTGCTAAAGCTGGTTGG	Not shown
S1	#394	$OmpA_1$/R/1	*Pt1*, *Pt2*	AGCGTGTGCTGCATCTAA	Not shown
S2	#480	$OmpA_1$/R/2	*Pt1*	TGCCACGAACACGACCAA	Not shown
S2	#481	$OmpA_1$/R/2	*Pt2*	GAACGCGACCGAAGTAGT	Not shown
Pt1* and *Pt2* segment 2 of *ompA'* and *ompA"					
P	#476	$OmpA_2$/F/1	*Pt1*	GTGCTGGCAATTACGGAA	Not shown
P, S1	#484	$OmpA_2$/F/1	*Pt2*	TCCAGTAGCAGCTCCTGA	Not shown
S1	#485	$OmpA_2$/F/1	*Pt1*	TCGACTTCGGTAAAGCA	Not shown
P	#477	$OmpA_2$/R/1	*Pt1+Pt2*	CCTGCTTTGGTTTGCTCT	Not shown
S1	#486	$OmpA_2$/R/1	*Pt1+Pt2*	CCGTATTTACCACCGTT	Not shown
Pt1* and *Pt2* segment 3 of *ompA'* and *ompA"					
P, S1	#424	$OmpA_3$/F/1	*Pt1+Pt2*	GGTGCTAAAGCTGGTTGG	Not shown
S2	#482	$OmpA_3$/F/2	*Pt1+Pt2*	CCAGTTGCTGAGCCAGA	Not shown
P, S1	#475	$OmpA_3$/R/1	*Pt1+Pt2*	TATGCAAGCTGGCTAAGG	Not shown
S2	#394	$OmpA_3$/R/2	*Pt1+Pt2*	AGCGTGTGCTGCATCTAA	Not shown

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing, S3: Third stage of sequencing

[b] OmpA/F/1: OmpA/Forward/1, OmpA/R/1: OmpA/Reverse/1

[c] *Mh1*: *M. haemolytica* isolates PH2, PH8, PH66, PH196, PH202, PH232, PH238, PH278, PH284, PH292, PH296, PH338, PH388, PH396, PH398, PH484, PH494, PH526, PH540, PH550, PH588, PH598, PH706

Mh2: *M. haemolytica* isolate PH470

Mh3: *M. haemolytica* isolates PH30, PH376, PH346, PH50, PH56, PH484, PH372, PH392

Mg1: *M. glucosida* isolate PH290

Mg2: *M. glucosida* isolate PH344

Mg3: *M. glucosida* isolates PH240, PH496, PH498

Mg4: *M. glucosida* isolate PH574

Pt: *P. trehalosi* isolates PII246, PII252, PH254, PII68

[c] Positions of the *ompA* primers for *M. haemolytica* only are shown. Nucleotide position corresponding to the first 5′ bp of the primer within *ompA* of contig 160 of the *M. haemolytica* genome sequence as of 01/02/04 (Figure 2.29)

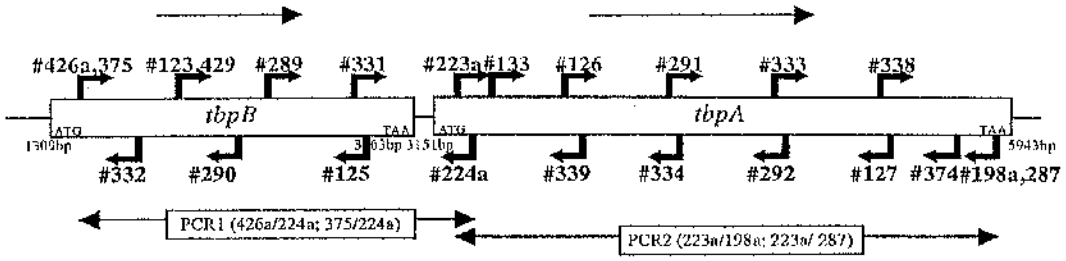
primers (#471 and #472) and a reverse primer (#473) specific for individual isolates were designed (Figure 2.28). The primer pairs #471/#473 and #472/#473 were used to amplify *ompA* in *M. haemolytica* and *M. glucosida*, respectively (Figure 2.28A). During the course of these preliminary experiments, difficulties were encountered in sequencing the *ompA* gene of *P. trehalosi*, despite the fact that definite PCR products were obtained. The results led us to suspect that two tandem *ompA* genes were present. Consequently, it became necessary to adopt a more complex sequencing strategy in which the *ompA* genes were amplified and sequenced as three separate overlapping fragments (Figure 2.28B and C). The primer pairs #474/#394, #476/#477, and #424/#475 were used to amplify *ompA* in *P. trehalosi* group *Pt1* (Figure 2.28B), and the primer pairs #465/#394, #484/#477, and #424/#475 were used to amplify *ompA* in *P. trehalosi* group *Pt2* (Figure 2.28C). The optimum annealing temperature for all PCRs was 56 °C. Internal sequencing primers were designed as sequence data became available. The positions of representative *M. haemolytica* group *Mh1* primers on the nucleotide sequence of the *ompA* gene of *M. haemolytica* are shown in Figure 2.29, and details of all primers are listed in Table 2.16.

2.3.2.5.2 Transferrin binding proteins (*tbpB* and *tbpA*)

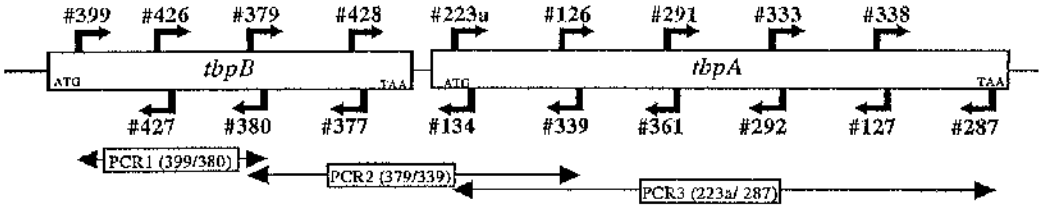
Various primers were used to amplify and sequence the continuous *tbpB* and *tbpA* genes in 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates (Figures 2.30 and 2.31, and Table 2.17). Of these, four forward (#426a, #192a, #193a, #223a) and three reverse (#198a, #224a, and #191a) primers have been shown to amplify transferrin binding protein genes in several members of the *Pasteurellaceae* (Ogunnariwo & Schryvers, 1996). However, the remaining PCR primers were designed within conserved regions of the aligned sequences of *M. haemolytica* and closely related species (see section 2.3.1.2). The primer pairs #426a/#224a,

Figure 2.30 Diagrammatic representation of the *tbpB* and *tbpA* genes of *M. haemolytica* taxon 75985 (PHU73302) showing locations of the PCR and sequencing primers. (A) *M. haemolytica* groups *Mh1* to *Mh4* and *M. glucosida* group *Mg*, (B) *M. haemolytica* group *Mh*5, (C) *M. haemolytica* group *Mh*6, (D) *M. haemolytica* group *Mh*7, and (E) *P. trehalosi* groups *Pt1* and *Pt2*. The locations of the primers used for PCR and sequencing are indicated by bold arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrows above *tbpB* and *tbpA* indicate the direction of transcription.

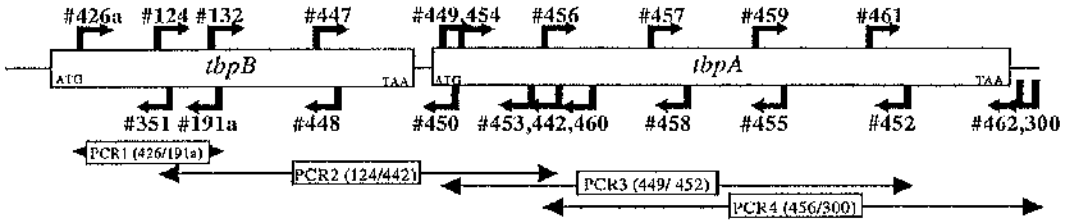
(A) *M. haemolytica* groups *Mh1-Mh4* and *M. glucosida* group *Mg*



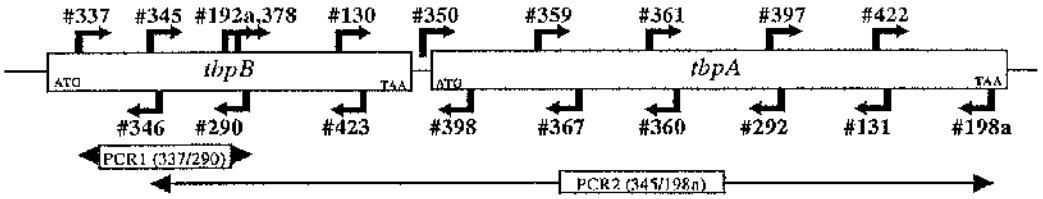
(B) *M. haemolytica* group *Mh5*



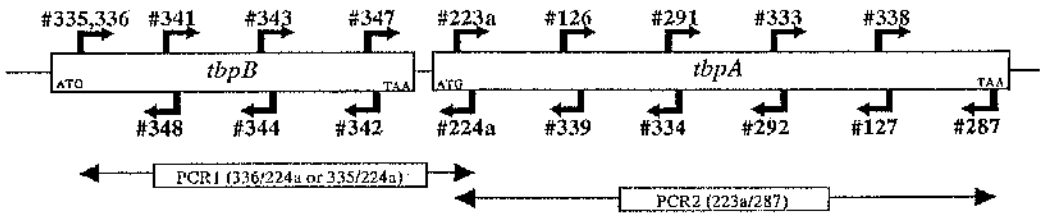
(C) *M. haemolytica* group *Mh6*



(D) *M. haemolytica* group *Mh7*



(E) *P. trehalosi* group *Pt1* and *Pt2*



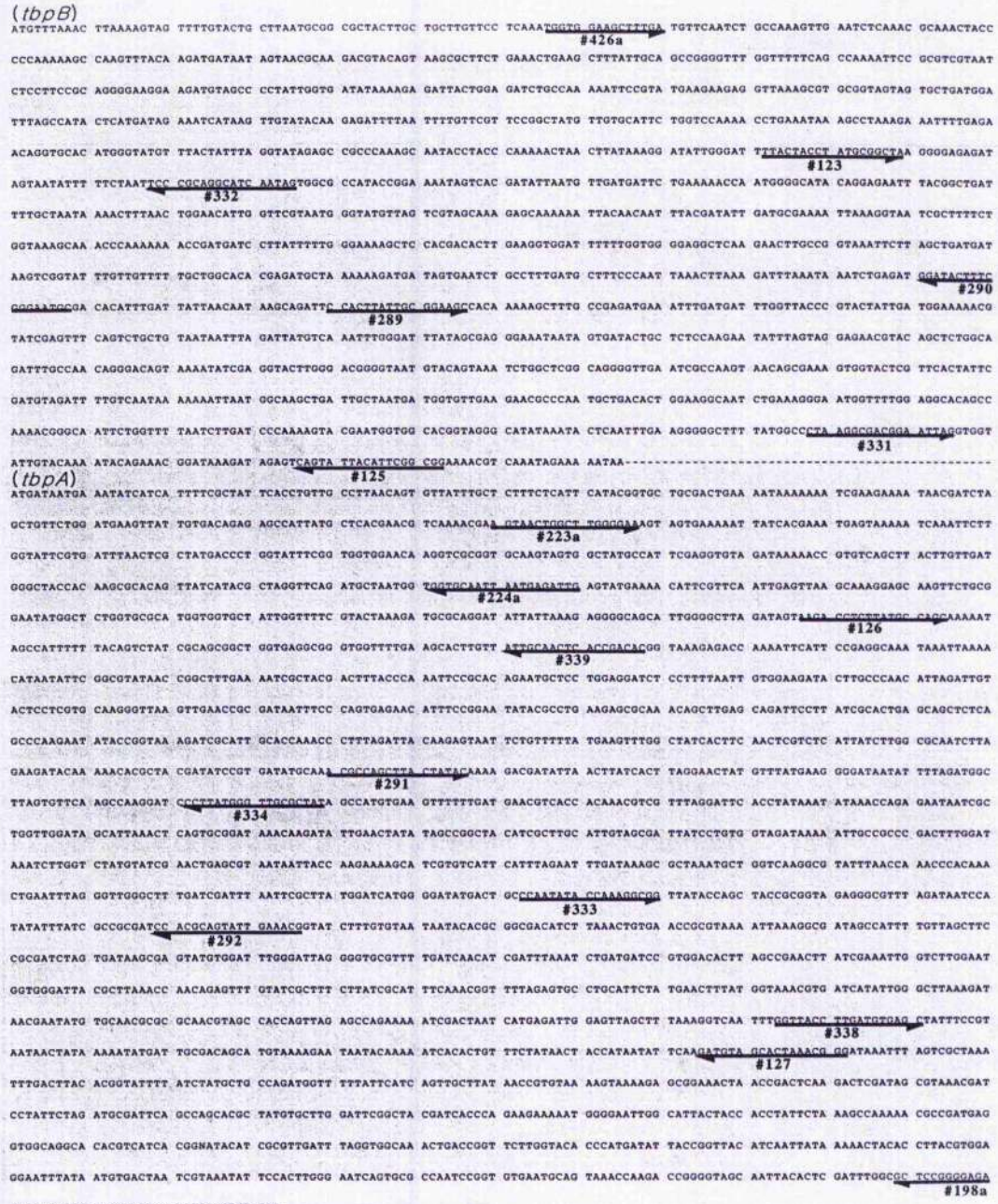


Figure 2.31 Nucleotide sequence of the *tbpB* and *tbpA* genes of *M. haemolytica* taxon 75985 (U73302) showing locations of the PCR and sequencing primers for *M. haemolytica* group *Mh1*. The primers used for PCR and sequencing are in bold type. Dashed lines indicate intergenic region between *tbpB* and *tbpA* (nucleotides are not shown for simplicity). The shaded area corresponds to the region of sequence that was analysed.

Table 2.17 Details of primers used for PCR amplification and sequencing of the *tbpB* and *tbpA* genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer [a] use	Primer No.	Primer [b] name	Isolates [c]	Sequence (5′ to 3′) [d]	Position [e]
***M. haemolytica* groups *Mh1-Mh4* and *M. glucosida* group *Mg* for PCR1 (426a/224; 375/224)**					
P, S1	#426a	TbpB/F/1	*Mh1, Mh4, Mg*	TGGTGGAAGCTTTGA	B 66 - 80
P, S1	#375	TbpB/F/1	*Mh2+Mh3,*	CTGCCAAAGTTGAATCTC	Not shown
S2	#123	TbpB/F/2	*Mh1+2,4, Mg*	TTACTACCGATGCGAGAA	B 572 - 589
S2	#429	TbpB/F/2	*Mh3,*	TCTAATTCCGCAGGCATC	Not shown
S3	#289	TbpB/F/3	*Mh1-4, Mg*	CCACTTATTGCGGAAGG	B 1120 - 1136
S4	#331	TbpB/F/4	*Mh1-4, Mg*	CTAAGGCGACGGAATTAG	B 1658 - 1675
P, S1	#224a	TbpB/R/1	*Mh1-4, Mg*	CAATCTCATTAATTGCACC	A 430 - 412
S2	#125	TbpB/R/2	*Mh1-4, Mg*	CCGCCGAATGTAATACTG	B 1730 - 1716
S3	#290	TbpB/R/3	*Mh1-4, Mg*	GCATTCCCGAAAGTATCC	B 1088 - 1071
S4	#332	TbpB/R/4	*Mh1-4, Mg*	ATATTGATGCCTGCGGA	B 635 - 618
***M. haemolytica* groups *Mh1-Mh4* and *M. glucosida* group *Mg* for PCR2 (223a/198a; 223a/287)**					
P, S1	#223a	TbpA/F/1	*Mh1-3, Mg*	AGTAACTGGCTTGGGGAA	A 180 - 197
S1	#133	TbpA/F/1	*Mh4*	GGTGCAAGTAGTGGCTAT	Not shown
S2	#126	TbpA/F/2	*Mh1-4, Mg*	AAGACCTCTTATGCCAGC	A 577 - 594
S3	#291	TbpA/F/3	*Mh1-4, Mg*	ACGCCAGCTTACTATAC	A 1120 - 1136
S4	#333	TbpA/F/4	*Mh1-4, Mg*	CCAATATACCAAAGGCGG	A 1623 - 1640
S5	#338	TbpA/F/5	*Mh1-4 Mg*	GGTTACCTTGATGTGAGC	A 2134 - 2151
P, S1	#198a	TbpA/R/1	*Mh1-3, Mg*	AAATTTCTCCCCGGAGC	A 2765 - 2749
P	#287	TbpA/R/1	*Mh4*	ACTGAAATTYCKCCCYGG	Not shown
S1	#374	TbpA/R/1	*Mh4*	ACTGAAATTTCTCCCCGG	Not shown
S2	#127	TbpA/R/2	*Mh1-4, Mg*	CCCGTTTAGTGCTACATC	A 2262 - 2245
S3	#292	TbpA/R/3	*Mh1-4, Mg*	CGTTTCAATACTGCGTGG	A 1716 - 1699
S4	#334	TbpA/R/4	*Mh1-4, Mg*	ATAGCGCAACCCATAAGG	A1239 - 1222
S5	#339	TbpA/R/5	*Mh1-4, Mg*	GTGTCGGTGAGTTGCAAT	A 678 - 661
***M. haemolytica* group *Mh5* for PCR1 (399/380)**					
P, S1	#399	TbpB1/F/1	*Mh5*	CTTCAAACGGTGGTAGT	Not shown
S2	#426	TbpB1/F/2	*Mh5*	CAGGTCCTACAGGGTATG	Not shown
P, S1	#380	TbpB1/R/1	*Mh5*	GCCAGACCAAAAGTATCC	Not shown
S2	#427	TbpB1/R/2	*Mh5*	GCCAGCATTACTTCCTTC	Not shown
***M. haemolytica* group *Mh5* for PCR2 (379/339)**					
P, S1	#379	TbpB2/F/1	*Mh5*	GGTGAAGACAAGGTCGAA	Not shown
S2	#428	TbpB2/F/2	*Mh5*	CCGATGCTGACATTAGA	Not shown
P	#339	TbpB2/R/1	*Mh5*	GTGTCGGTGAGTTGCAAT	Not shown
S1	#134	TbpB2/R/1	*Mh5*	ATAGCCACTACTTGCACC	Not shown
S2	#377	TbpB2/R/2	*Mh5*	CAGCTTGCGGACCATAA	Not shown
***M. haemolytica* group *Mh5* for PCR3 (223a/287)**					
Primers are identical to *Mh1-Mh4* and *Mg* PCR2 primers except for #361					
S4	#361	TbpA/R/4	*Mh5*	CACGGATATCGTAGCGTT	Not shown

Table 2.17 (continued)

Primer [a] use	Primer No.	Primer [b] name	Isolates [c]	Sequence (5′ to 3′) [d]	Position [e]
***M. haemolytica* group *Mh6* for PCR1 (426a/191a)**					
P, S1	#426a	TbpB1/F/1	*Mh6*	TGGTGGAAGCTTTGA	Not shown
S2	#124	TbpB1/F/2	*Mh6*	ACTGATGCAAGAAAAGGG	Not shown
P, S1	#191a	TbpB1/R/1	*Mh6*	CCAAAAAATCCACCTTC	Not shown
S2	#351	TbpB1/R/2	*Mh6*	ACATACCCATGTGCACCT	Not shown
***M. haemolytica* group *Mh6* for PCR2 (124/442)**					
P	#124	TbpB2/F/1	*Mh6*	ACTGATGCAAGAAAAGGG	Not shown
S1	#132	TbpB2/F/1	*Mh6*	GTCGTACTACAGAGCAAG	Not shown
S2	#447	TbpB2/F/2	*Mh6*	GTGCAATATCGCGGAACT	Not shown
S3	#449	TbpB2/F/3	*Mh6*	CGGCGGCTGAAAATGAAA	Not shown
P	#442	TbpB2/R/1	*Mh6*	CSSRKAAAGATCATAGCG	Not shown
S1	#453	TbpB2/R/1	*Mh6*	GCAGCGGCAATAGATTGT	Not shown
S2	#450	TbpB2/R/2	*Mh6*	TTCCTAAGCCGGTTACCT	Not shown
S3	#448	TbpB2/R/3	*Mh6*	AAGAGTGAGCGAGTACCA	Not shown
***M. haemolytica* group *Mh6* for PCR3 (449/ 452)**					
P	#449	TbpA1/F/1	*Mh6*	CGGCGGCTGAAAATGAAA	Not shown
S1	#454	TbpA1/F/1	*Mh6*	TACGCTCACGAACGTCAA	Not shown
S2	#456	TbpA1/F/2	*Mh6*	CTCACTCTGAAGCCAAT	Not shown
S3	#457	TbpA1/F/3	*Mh6*	CTGCGTTATAGCCGTGTA	Not shown
S4	#459	TbpA1/F/4	*Mh6*	TGGACACTTAGTCGAAC	Not shown
P, S1	#452	TbpA1/R/1	*Mh6*	ACGATGTATACCGTGGTG	Not shown
S2	#455	TbpA1/R/2	*Mh6*	GCCGGTACTCTAAATCCA	Not shown
S3	#458	TbpA1/R/3	*Mh6*	CTACAATGCAAGCGGTG	Not shown
S4	#460	TbpA1/R4	*Mh6*	AATGATCCTCCAGGAGCA	Not shown
***M. haemolytica* group *Mh6* for PCR4 (456/300)**					
P	#456	TbpA2/F/1	*Mh6*	CTCACTCTGAAGCCAAT	Not shown
S1	#461	TbpA2/F/1	*Mh6*	GATCGCAACGGCTTGTAA	Not shown
P	#300	TbpA2/R/1	*Mh6*	ACTCGTGCAGCAACAATG	Not shown
S1	#462	TbpA2/R/1	*Mh6*	CCGTGGTACATTACGCAA	Not shown
***M. haemolytica* group *Mh7* for PCR1 (337/290)**					
P, S1	#337	TbpB/F/1	*Mh7*	CTTAATGCGGCGCTACT	Not shown
S2	#345	TbpB/F/2	*Mh7*	GGTGCACATGGGTATGTT	Not shown
P, S1	#290	TbpB/R/1	*Mh7*	GCATTCCCGAAAGTATCC	Not shown
S2	#346	TbpB/R/2	*Mh7*	GATGCCTGCTGAATCAGT	Not shown
***M. haemolytica* group *Mh7* for PCR2 (345/198a)**					
P	#345	TbpB/F/1	*Mh7*	GGTGCACATGGGTATGTT	Not shown
S1	#378	TbpA/F/1	*Mh7*	CTCAAGAACTTGCCGGTA	Not shown
S2	#130	TbpA/F/2	*Mh7*	ACGTACCGAATTTCAGTC	Not shown
S3	#350	TbpA/F/3	*Mh7*	TAGATAGCAAGCGGGCAA	Not shown
S4	#359	TbpA/F/4	*Mh7*	ATATGGTTCTGGTGCGCT	Not shown
S5	#361	TbpA/F/5	*Mh7*	CACGGATATCGTAGCGTT	Not shown
S6	#397	TbpA/F/6	*Mh7*	AGGGTTGGGCTTTGATCG	Not shown

Table 2.17 (continued)

Primer [a] use	Primer No.	Primer [b] name	Isolates [c]	Sequence (5' to 3') [d]	Position [e]
S7	#422	TbpA/F/7	*Mh7*	GGTTACCTTGATGTGAG	Not shown
S1	#198a	TbpA/R/1	*Mh7*	AAATTTCTCCCCGGAGC	Not shown
S2	#131	TbpA/R/2	*Mh7*	CCGTTTAACACCACATCT	Not shown
S3	#292	TbpA/R/3	*Mh7*	CGTTTCAATACTGCGTGG	Not shown
S4	#360	TbpA/R/4	*Mh7*	TAGTAAGCCTGCGTTTGC	Not shown
S5	#367	TbpA/R/5	*Mh7*	GCTGGCATAGGCAGTTTT	Not shown
S6	#398	TbpA/R/6	*Mh7*	AGGCAACAGGTGAATAGC	Not shown
S7	#423	TbpA/R/7	*Mh7*	GTTACTTGGCGATTCAGC	Not shown
***P. trehalosi* group *Pt1* and *Pt2* for PCR1 (335/224; 336/224)**					
P, S1	#335	TbpB/F/1	*Pt2*	GGYGGWAGYTTTGATGT	Not shown
P, S1	#336	TbpB/F/1	*Pt1*	CGGCGGTAGTTTTGATGT	Not shown
S2	#341	TbpB/F/2	*Pt1+Pt2*	TACGACCGATGCAAGAA	Not shown
S3	#343	TbpB/F/3	*Pt1+Pt2*	CCTCTGATTGCAGAAGGT	Not shown
S4	#347	TbpB/F/4	*Pt1+Pt2*	GGAGGCACCGTCATACAT	Not shown
P, S1	#224	TbpB/R/1	*Pt1+Pt2*	CAATCTCATTAATTGCACC	Not shown
S2	#342	TbpB/R/2	*Pt1+Pt2*	TGCACAATTCCACCAAGC	Not shown
S3	#344	TbpB/R/3	*Pt1+Pt2*	ACCAACGGTGGGCATAA	Not shown
S4	#348	TbpB/R/4	*Pt1+Pt2*	AGATTCACGTCTCGACTG	Not shown
***P. trehalosi* group *PtI* and *Pt2* for PCR2 (223a/287)** Primers are identical to *Mh1-Mh4* and *Mg* PCR2 primers					

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing, S3: Third stage of sequencing, S4: Forth stage of sequencing

[b] TbpB/F/1: TbpB/Forward/1, TbpB/R/1: TbpB/Reverse/1, TbpA/F/1: TbpA/Forward/1, TbpA/R/1: TbpA/Reverse/1

[c] *Mh1*: *M. haemolytica* isolates PH2, PH66, PH232, PH296, PH588, PH30, PH376, PH346, PH8, PH396, PII484, PII50, PII388

Mh2: M. haemolytica isolates PH540, PH338, PH238, PH398

: *Mh3: M. haemolytica* isolates PH56, PH284

Mh4: M. haemolytica isolates PH706, PH278, PH372

Mh5: M. haemolytica isolates PII494, PII292, PII392, PH526, PH598

Mh6: *M. haemolytica* isolates PH550, PH202, PH470

Mh7: M. haemolytica isolates PH196, PH786

Mg: M. glucosida isolates PH344, PH290, PH240, PH496, PH498, PH574

Pt1: *P. trehalosi* isolates PH246, PII68

Pt2: *P. trehalosi* isolates PH252, PH254

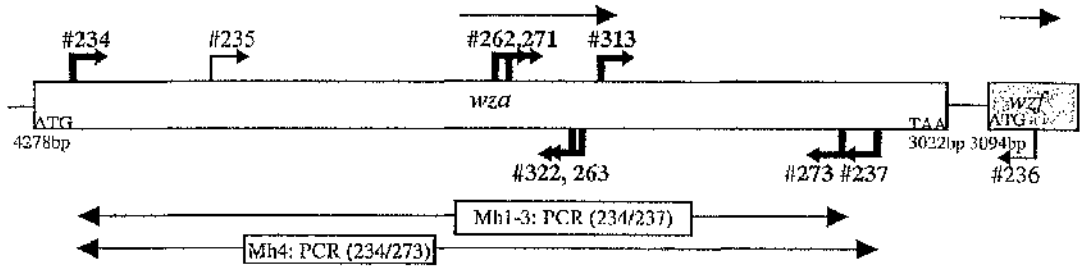
[e] Positions of the *tbpB* and tbpA primers for *M. haemolytica* group *Mh1* only are shown. Nucleotide position corresponding to the first 5' bp of the primer within the *tbpB* (B) and *tbpA* (A) gene of the *M. haemolytica* GenBank sequence (PHU73302)

#375/#224a, and #223a/#198a were used for PCRs and for the first stage of sequencing of *tbpB* and *tbpA* in *M. haemolytica* group *Mh1* to *Mh4* and *M. glucosida* (Figure 2.30A). The primer pairs #399/#380, #379/#339, and #223a/#287 were used for PCR and for the first stage of sequencing of *tbpB* and *tbpA* in *M. haemolytica* group *Mh5* (Figure 2.30B). The primer pairs #426a/#191a, #124/#442, #449/#452, and #456/#300 were used for PCR and for the first stage of sequencing of *tbpB* and *tbpA* in *M. haemolytica* group *Mh6* (Figure 2.30C). The primer pairs #337/#290 and #345/#198a were used for PCR and for the first stage of sequencing of *tbpB* and *tbpA* in *M. haemolytica* group *Mh7* (Figure 2.30D). The primer pairs #336/#224a and #223a/#287 were used for PCR and for the first stage of sequencing of *tbpB* and *tbpA* in *P. trehalosi* groups *Pt1* and *Pt2* (Figure 2.30E). The optimum annealing temperature of the PCRs was 55 °C in most cases, but 50 °C (PCR 223/287 for PH246, PH252, PH254, PH68), 52 °C (PCR 399/380 and 379/339 for PH494, PH292, PH392, PH526, PH598), 56 °C (PCR 345/198 for PH196 and PH786), 57 °C (PCR 335/224 for PH252, PH254), and 58 °C (PCR 456/300 for PH550, PH202, PH470 and PCR 223/287 for PH706, PH278, PH372, PH494, PH292, PH392) were also used. Internal sequencing primers were designed as sequence data became available. Also, first stage sequencing primers were redesigned when the sequences were poor. The positions of *M. haemolytica* group *Mh1* primers on the nucleotide sequence of the *tbpB* and *tbpA* genes of *M. haemolytica* are shown in Figure 2.31, and details of all primers for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* are listed in Table 2.17.

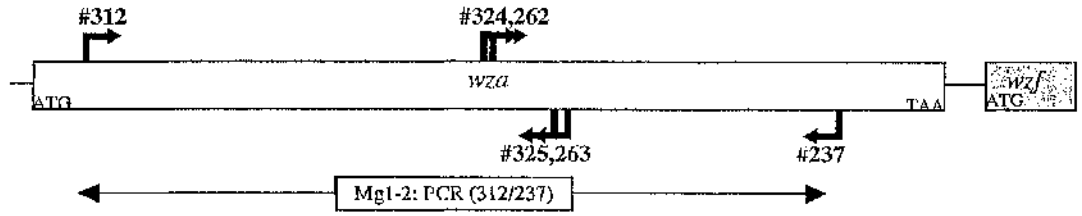
2.3.2.5.3 Capsule transport protein (*wza*)

Initially, two forward (#234 and #235) and two reverse (#236 and #237) primers were designed within the conserved regions of the aligned *wza* sequences from *M. haemolytica* and closely related species (see section 2.3.1.2) (Figures 2.32 and 2.33

(A) *M. haemolytica* groups *Mh1-Mh4*



(B) *M. glucosida* groups *Mg1-Mg2*



(C) *P. trehalosi* group *Pt*

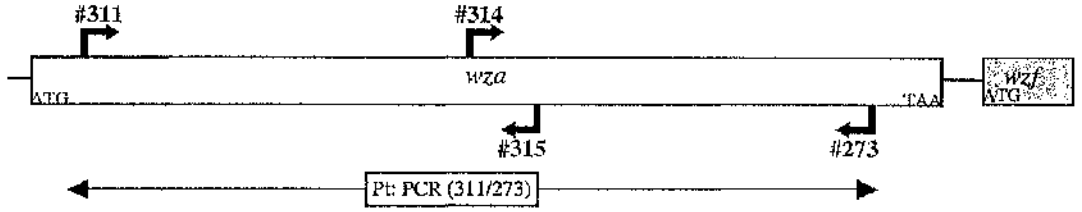


Figure 2.32 Diagrammatic representation of the *wza* gene of *M. haemolytica* taxon 746 (complementary GenBank sequence AF170495) showing locations of the PCR and sequencing primers. (A) *M. haemolytica* groups *Mh1* to *Mh4*, (B) *M. glucosida* groups *Mg1* and *Mg2*, and (C) *P. trehalosi* group *Pt*. The locations of the primers used for PCR and sequencing are indicated by bold arrows and the locations of the unused primers are indicated by thin arrows. The numbers under the start and stop codons indicate the positions of the first base of the start codon and third base of the stop codon within the GenBank sequence. The large arrow above *wza* indicates the direction of transcription.

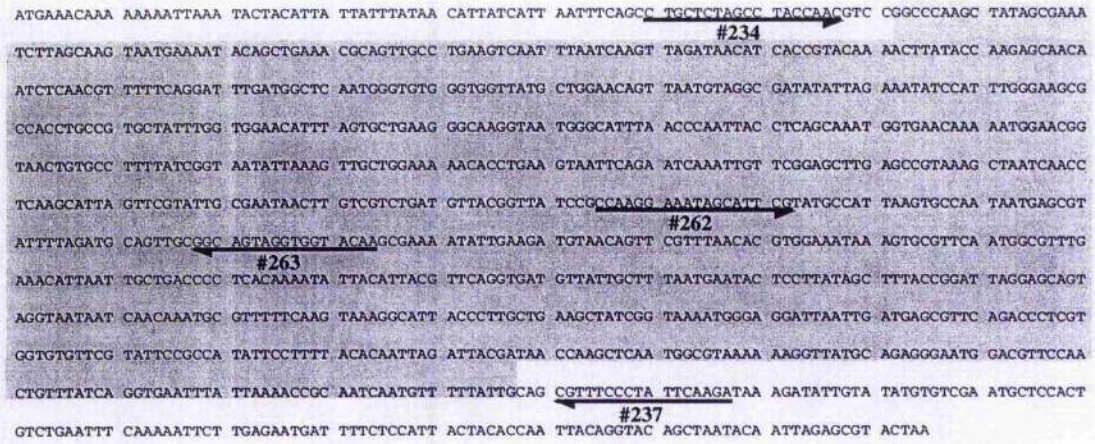


Figure 2.33 Nucleotide sequence of the *wza* gene of *M. haemolytica* taxon 746 (complementary GenBank sequence AF170495) showing locations of the PCR and sequencing primers for *M. haemolytica* group *Mh1*. The primers used for PCR and sequencing are in bold type. The shaded area corresponds to the region of sequence that was analysed.

Table 2.18 Details of primers used for PCR amplification and sequencing of the *wza* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Primer [a] use	Primer No.	Primer [b] name	Isolates [c]	Sequence (5′ to 3′) [d]	Position [e]
***M. haemolytica* groups *Mh1-Mh3* (234/237) and *Mh4* (234/273)**					
P, S1	#234	Wza/F/1	*Mh1-Mh4*	CTGCTCTAGCCTACCAAC	60-77
	#235	Wza/F/1		GAACAGTTAATGTAGGCG	Not shown
S2	#262	Wza/F/2	*Mh1*	CCAAGGAAATAGCATTCG	555-572
S2	#271	Wza/F/2	*Mh2+Mh3*	CTGATGTTACGGTTATCC	Not shown
S2	#313	Wza/F/2	*Mh4*	CAGTTGGTGGTGCAAGTG	Not shown
	#236	Wza/R/1		GTTAGCCACAACAAAGGG	Not shown
P, S1	#237	Wza/R/1	*Mh1-Mh3*,	TCTTGAATAGGGAAACG	1067-1051
P, S1	#273	Wza/R/1	*Mh4*	GGMGCATTCGACACATA	Not shown
S2	#263	Wza/R/2	*Mh1*, *Mh3*, *Mh4*	CTTGTACCRCCTACTGC	625-619
S2	#322	Wza/R/2	*Mh2*	TTGTGCCGCCTACTGCT	Not shown
***M. glucosida* groups *Mg1* and *Mg2* (312/237)**					
P, S1	#312	Wza/F/1	*Mg1+Mg2*	TGTTCTAGCCTACCTAC	Not shown
S2	#262	Wza/F/2	*Mg1*	CCAAGGAAATAGCATTCG	Not shown
S2	#324	Wza/F/2	*Mg2*	ACAGTGGTTCGTGAAGGT	Not shown
P, S1	#237	Wza/R/1	*Mg1+Mg2*	TCTTGAATAGGGAAACG	Not shown
S2	#263	Wza/R/2	*Mg1*	CTTGTACCRCCTACTGC	Not shown
S2	#325	Wza/R/2	*Mg2*	TACCGCCTACAGCTGCAA	Not shown
***P. trehalosi* group *Pt* (311/273)**					
P, S1	#311	Wza/F/1	*Pt*	TCTAGCCTACCTACCTCG	Not shown
S2	#314	Wza/F/2	*Pt*	GCTGCAGTAGGTGGTTC	Not shown
P, S1	#273	Wza/R/1	*Pt*	GGMGCATTCGACACATA	Not shown
S2	#315	Wza/R/2	*Pt*	GATACGGCTCTGCGATTA	Not shown

[a] P: PCR amplification, S1: First stage of sequencing, S2: Second stage of sequencing

[b] Wza/F/1: Wza/Forward/1, Wza/R/1: Wza/Reverse/1

[c] *Mh1*: *M. haemolytica* isolates PH2, PH66, PH196, PH202, PH278, PH292, PH296, PH494, PH706, PH8, PH30, PH238, PH346, PH372, PH376, PH388, PH392, PH396, PH398, PH470, PH484, PH526, PH540, PH550, PH598

Mh2: *M. haemolytica* isolates PH56, PH238, PH338

Mh3: *M. haemolytica* isolates PH50, PH232, PH284 *Mh4*: *M. haemolytica* isolates PH588

Mg1: *M. glucosida* isolates PH344, PH498, PH240 *Mg2*: *M. glucosida* isolate PH496

Pt: *P. trehalosi* isolates PH246, PH252, PH254, PH68

[d] M: C+A, R: G+A

[e] Positions of the *wza* primers for *M. haemolytica* group *Mh1* only are shown. Nucleotide position corresponding to the first 5′ bp of the primer within the *wza* gene of the *M. haemolytica* GenBank sequence (AF170495) (see Figure 2.33)

and Table 2.18). The primers were tested in four combinations (#234/#236, #234/#237, *etc*) by PCR with *M. haemolytica* isolate PH2 to determine the optimum primer pair and annealing temperature. The primer pair #234/#237 was used for PCR and for the first stage of sequencing of the *wza* gene in *M. haemolytica* groups *Mh1-Mh3* (Figure 2.32A). Attempts to amplify *wza* in *M. haemolytica* groups *Mh4*, *M. glucosida*, and *P. trehalosi* with the four primer combinations failed. Therefore, additional forward (#311 and #312) and reverse (#273) primers were designed. The primer pairs #234/#273, #312/#237, and #311/#273 were used for PCR and for the first stage of sequencing of *wza* in *M. haemolytica* group *Mh4* (Figure 2.32A), *M. glucosida* (Figure 2.32B), and *P. trehalosi* (Figure 2.32C), respectively. Attempts to amplify *wza* in *M. glucosida* isolates PH574 and PH290 with the primers were unsuccessful. The optimum annealing temperatures for the PCRs were 45 °C for *M. glucosida* group *Mg2*, 51 °C for *P. trehalosi*, 52 °C for *M. haemolytica* groups *Mh1-Mh3*, and 55 °C for *M. haemolytica* group *Mh4* and *M. glucosida* group *Mg1*. Internal sequencing primers were designed as sequence data became available. When the sequence data were poor, the first stage sequencing primers were redesigned. The positions of representative *M. haemolytica* group *Mh1* primers on the nucleotide sequence of the *wza* gene of *M. haemolytica* are shown in Figure 2.33, and details of all primers for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* are listed in Table 2.18.

2.4 SDS-PAGE analysis of *P. trehalosi* OmpA proteins

P. trehalosi outer membrane proteins (OMPs) were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using the SDS discontinuous system (Laemmli, 1970) as described previously (Davies *et al.*, 1992). After stacking in a 4% (w/v) gel, OMPs were separated in 12% (w/v) resolving gels. A total of 25 µg protein was loaded per lane (25 µl). Electrophoresis was performed

in buffer comprising 25 mM Tris/HCl, 192 mM-glycine and 0.1% SDS (pH 8.3), at a constant current of 20 mA per gel through the stacking gel and 30 mA per gel through the resolving gel. Proteins were visualized by staining with Coomassie blue.

2.5 Characterization of temperate phages

2.5.1 Optimum concentration of mitomycin C for bacteriophage induction

To determine the optimum concentration of mitomycin C for the induction of temperate bacteriophages in *M. haemolytica*, *M. glucosida* and *P. trehalosi*, different concentrations of mitomycin C (final concentration of 0, 0.01, 0.05, 0.1, 0.2, 1.0, and 2.0 μg/ml) were tested with four isolates. *M. haemolytica* serotype A1 isolates PH2, PH280, PH342, and PH372 were selected because isolates of this serotype were previously shown to produce temperate phages (Froshauer *et al.*, 1996; Richards *et al.*, 1985). For each isolate, 200 μl of overnight BHIB culture were inoculated into 20 ml of prewarmed BHIB in seven conical flasks. The cultures were mixed by swirling and 1 ml was removed from each flask to measure the optical density (OD) at 660 nm in a spectrophotometer (Helios Thermo Spectronic). The flasks were incubated at 37 °C with shaking at 120 rpm and further OD readings were measured at 60 min intervals. When the OD value was over 0.2, increasing volumes (0, 2, 10, 20, 40, 200, and 400 μl) of mitomycin C solution (0.1mg/ml) were added to each group of seven flasks and incubation was continued. To monitor the extent of bacterial lysis, OD_{660} readings were taken at 30 min intervals during the incubation period.

2.5.2 Induction of bacteriophages in *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates

Thirty two *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates (Table 2.1) selected for previous comparative nucleotide sequence analysis were tested for the presence of temperate phages by mitomycin C induction (0.2 µg/ml final concentration). For the first 12 isolates (PH2, PH66, PH196, PH202, PH246, PH278, PH292, PH296, PH344, PH494, PH588, and PH706), 0.2 ml of an overnight BHIB culture were inoculated into 20 ml of prewarmed BHIB in each of two conical flasks, but 2.0 ml of overnight culture was used for the remaining 30 isolates so that the mitomycin C could be added sooner. The flasks were incubated until the OD_{660} value of the culture reached a value of 0.2 or greater at which point 40 µl of mitomycin C solution (0.1 mg/ml) were added to one of each pair of flasks. The mitomycin C treated and untreated (control) cultures were incubated for 6 to 7 h and OD_{660} readings were measured at 30 min intervals. If there was no sign of induction with 0.2 µg/ml of mitomycin C for a particular isolate, the experiment was repeated with a higher concentration of mitomycin C.

2.5.3 Preparation of phage samples

After phage induction by mitomycin C treatment, the lysates were centrifuged at 4,000 × g for 20 min at 4°C and the supernatants were filtered through 0.2 µm membrane filters (Millipore) to remove bacterial cell debris. These phage preparations were immediately used for plaque assay (Section 2.5.4.2) or stored for short periods (less than five days) at 4 °C before electron microscopy (section 2.5.5) and DNA isolation (section 2.5.6).

2.5.4 Host range of bacteriophages

The host range of induced bacteriophages was examined by plaque assay using appropriate indicator strains.

2.5.4.1 Preparation of overlay indicator plates

A base layer of BHIA (1 %) was poured into a standard petri dish and allowed to solidify. The plate was dried in a 37 °C constant temperature room for 20 min. One hundred microlitres of an overnight culture of each indicator isolate were inoculated into 3 ml of molten soft agar (0.7 % w/v agar) at 56 °C, containing 2 mM $CaCl_2$, mixed, and immediately poured onto the base layer. The overlay plates were dried in a laminar-flow hood for 20 min before use.

2.5.4.2 Plaque assays

Five microlitres of each phage sample (filtered supernatants) were spotted onto each plate seeded with each of the indicator isolates. The plates were incubated at 37 °C overnight and examined for plaques. The plaque assays were carried out on the same day as phage induction to avoid the possible reduction of phage activity with time.

2.5.5 Electron microscopy by negative staining

Five to ten millilitres of phage sample (filtered supernatants) were centrifuged at 20,000 × g for 3 h at 4 °C. The supernatant was carefully removed and the pellet was slowly resuspended in 0.5 to 1.0 ml of 0.1 M ammonium acetate buffer (pH 7.3)

at 4°C overnight. Five microlitres of phage suspension were dropped onto a glow-discharged carbon-coated 200 mesh copper grid. After 10 s of adsorption time, excess fluid was removed by touching the rim of the grid with a piece of filter paper and a drop of the staining solution, Nanovan (Nanoprobes), was placed on the grid and excess fluid was immediately removed by touching the rim of the grid with a piece of filter paper. The grid was air-dried and imaged using a LEO 902 electron microscope (Zeiss) at 80 kV using different magnifications (20,000 to 250,000 ×).

2.5.6 Isolation of bacteriophage DNA

Bacteriophage DNA was prepared with a Wizard lambda DNA purification kit (Promega). Ten millilitres of phage samples (filtered supernatants) were treated with DNase (Promega) and RNAse (Promega) (both at 1 µg/ml final concentration) for 30 min at 37 °C. After adding 4 ml of phage precipitant (33 % polyethylene glycol and 3.3 M NaCl), the samples were mixed gently, placed on ice for 2 to 3 h, and centrifuged at 10,000 × g for 10 min at 4 °C. The pellets were resuspended in 500 µl of phage buffer (150 mM NaCl, 40 mM Tris-HCl [pH 7.4], 10 mM $MgSO_4$), and treated with proteinase K (Promega, lyophilized powder) (50 µg/ml final concentration) for 1 h at 56 °C. The samples were transferred to 1.5 ml microcentrifuge tubes, centrifuged at 10,000 × g for 10 s to remove any insoluble particles, and the supernatants were transferred to clean microcentrifuge tubes. The supernatants were mixed with 1 ml of purification resin and the mixtures were transferred to syringe barrels attached to minicolumns. The resin/lysate mixtures were forced through the minicolumns, and the minicolumns were washed with 2 ml of 80 % isopropanol and centrifuged at 10,000 × g for 2 min. Phage DNA was eluted by adding 100 µl of distilled water preheated to 80°C and centrifuging at 10,000 × g for 20 s. The phage DNA was examined by agarose gel electrophoresis and stored at

4 °C.

In some experiments, larger volumes (40 ml) of phage sample were used. The phage samples were treated with 9 ml of phage precipitant and centrifuged as described above. The pellets were resuspended in 700 µl of phage buffer and phage suspensions were treated with 6 µl of nuclease mixture (0.25 mg/ml of DNase I and RNAse A) for 30 min at 37 °C. The phage samples were treated with SDS (0.5 % final concentration) and EDTA (20 mM final concentration) for 15 min at 70 °C. The phage DNA was purified with Promega minicolumns as described above.

2.5.7 Restriction endonuclease digestion

Bacteriophage DNA was double digested with 1 µl of each of the restriction enzymes HindIII and ClaI (Promega). The final 20 µl reaction mixture contained 5 µl of phage DNA and 9 µl distilled water (PH2, PH376, PH786) or 14 µl of phage DNA and no distilled water (PH56, PH8, PH398, PH284, PH66, PH484, PH344, PH252, PH396, PH196, PH292, PH296, PH392, PH598, PH388, and PH238) together with 2 µl of 10× buffer E, 2 µl of 1× bovine serum albumin (1 µg/µl), and 1 µl of each restriction enzyme (1 U/µl). The reactions were incubated at 37 °C for 3 h and the digested DNA was assessed by agarose gel electrophoresis. Five (PH2, PH376, PH56, PH8, PH398, PH284, PH66, PH484, PH344, and PH252) or nine (PH196, PH786, PH396, PH292, PH296, PH392, PH598, PH388, PH238) microlitres of digested DNA were run in 0.8 % (w/v) agarose gels containing ethidium bromide (0.5 µg/ml) at 100 V for 1 h and visualised and photographed under UV light.

2.5.8 Genomic analysis of bacteriophages from *M. haemolytica* bovine A1 isolates

2.5.8.1 Restriction endonuclease assays

Six different restriction endonuclease (RE) assays were carried out with bacteriophage DNA obtained from *M. haemolytica* isolate PH2. Five RE assays were single digestions with restriction enzymes HindIII, ClaI, HpaII, PstI, and RsaI (all from Promega) using buffers E, C, A, H, and C, respectively. One RE assay was a double digestion with restriction enzymes HindIII and Cla I using buffer E. The 20 µl reaction mixture contained 4 µl of phage DNA, 11 µl of distilled water, 2 µl of 10× buffer, 2 µl of bovine serum albumin (1 µg/µl), and 1 µl of restriction enzyme (1 U/µl). The reactions were incubated at 37 °C for 3 h. The reannealing of cohesive ends of digested fragments was avoided by heating the samples for 10 min at 70 °C and cooling immediately on ice. Five microlitres of each sample were run on a 0.8 % agarose gel containing ethidium bromide (0.5 µg/ml) at 100 V for 1 h and visualised and photographed under UV light.

2.5.8.2 Construction of bacteriophage genome from *M. haemolytica* isolate PHL213

Using six different digested fragment patterns of PH2Φ representing phages of bovine serotype A1 and A6 *M. haemolytica* isolates, the prophage genome was searched within the unfinished genome sequence of the bovine serotype A1 *M. haemolytica* isolate PHL213 (www.hgsc.bcm.tmc.edu/projects/microbial/mhaemolytica). Twelve contigs of PHL213 (as of 10/04/01) containing phage related genes were kindly provided by Dr. S. K. Highlander (Baylor College of Medicine, Houston, Texas) . Each of them was examined for putative digested fragments of six different RE

enzyme by using a restriction map generator program, NEB Cutter (http://tools.neb.com/NEBcutter2) and one contig, Contig C243 (as of 10/04/01), was suspected to be associated with the prophage genome. Blast analyses of the contig C243 (4978 bp) against the phage sequence databanks (http://salmonella.utmem.edu/phage) identified that the contig C243 contain genes similar to those of phage P2. Since the genetic organization of contig C243 genes was similar to that of the corresponding region of phage P2, the flanking genes of the regions of phage P2 were thought to be similar to the flanking region of contig C243 of *M. haemolytica* isolate PHL213. As expected, three additional contigs C19, C358, and C359 (as of 10/04/01) that contained genes similar to phage P2 genes were identified by a blast analysis with the flanking genes of phage P2 against *M. haemolytica* database. These contigs have been updated and the most recent contigs (as of 01/02/04), C91 (updated from C243), C137 (updated from C19), and C150 (updated from C358 and C359), were used for the construction of phage genome. The order of C150, C91, and C137 was similar to the genetic organization of the complete P2 phage genome.

To confirm that these contigs are associated with phages from *M. haemolytica* PH2 (bovine A1), and to determine the size of the gaps between C150 and C91, and between C91 and C137, two forward #504 and #506 and two reverse #503 and #505 primers were designed to amplify the gaps as shown in Figure 2.34. The primer pair #505/#506 was used to amplify the segment A region and the primer pair #503/#504 was used to amplify the segment B region in the phage DNA of *M. haemolytica* isolate PH2 (as described previously - see section 2.3.1.3). The optimum annealing temperature of the two PCRs was 55 °C and details of the primers are listed in Table 2.19.

2.5.8.3 Presence of the PH2Φ phage genome among *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates

The existence of the PH2Φ phage genome among isolates of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* was examined by PCR amplification of segment B, using the primer pair #505/#506 and an annealing temperature of 55 °C, in bacterial DNA from 42 isolates.

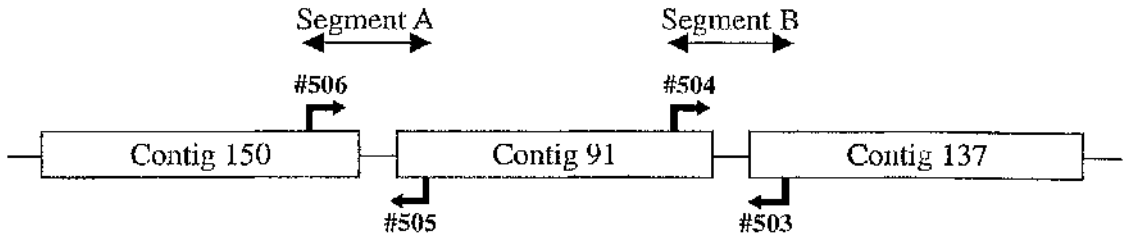


Figure 2.34 Diagrammatic representation of the contigs 137, 91 and 150 of *M. haemolytica* isolate PHL213 showing positions of PCR primers.

Table 2.19 Sequence of primers used to amplify segments A and B in phage DNA from *M. haemolytica* isolate PH2 (see Figure 2.34)

Primer No.	Sequence (5′ to 3′)
503	AGATCCAGCGTGCAACTT
504	CTAGCACCAGCACAATCA
505	AGCAAGAGCAAGATCGTC
506	CAATAGTGCCAAGTCGGA

CHAPTER 3: RESULTS

3.1 Comparative sequencing analysis

Details of the 19 genes studied are given in Table 3.1. Two genes (*plpB* and *ompA*) were completely sequenced whereas the remaining 17 genes (*recA*, *aroA*, *asd*, *galE*, *gap*, *gnd*, *g6pd*, *mdh*, *mtlD*, *pmm*, *gcp*, *plpA*, *plpC*, *plpD*, *tbpB*, *tbpA*, and *wza*) were partially sequenced. Seven genes (*recA*, *asd*, *mdh*, *gnd*, *gcp*, *galE*, and *plpA*) were sequenced in 12 representative isolates of *M. haemolytica* (10), *M. glucosida* (1), and *P. trehalosi* (1). Nine genes (*aroA*, *gap*, *mtlD*, *pmm*, *plpD*, *ompA*, *tbpB*, *tbpA*, and *wza*) were sequenced in more than 12 representative isolates of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* because the preliminary analysis of 12 isolates of these genes showed evidence of recombination or other interesting points (discussed in each gene section). Three genes (*g6pd*, *plpB* and *plpC*) were sequenced in less than 12 representative isolates because of unsuccessful amplifications.

The amino acid sequence of each gene of *M. haemolytica* isolate PH2 was compared with the amino acid sequence of the corresponding gene in seven members of the *Pasteurellaceae* including *M. glucosida* isolate PH344, *P. trehalosi* isolate PH246, *A. pleuropneumoniae* isolate 4074, *H. ducreyi* isolate 35000hp, *H. somnus* isolate 2336, *P. multocida* isolate PM70, and *H. influenzae* isolate KW20 (Table 3.2). The sequence data for *M. glucosida* and *P. trehalosi* are from this study, whereas those of the other five species are from the GenBank database (Table 3.2). Each of the genes from *M. haemolytica* isolate PH2 showed amino acid identity with the corresponding genes of *M. glucosida* (93 to 100 %), *P. trehalosi* (63 to 94 %), *A. pleuropneumoniae* (38 to 93 %), *H. ducreyi* (62 to 92 %), *H. somnus* (31 to 91 %), *P. multocida* (50 to 90 %), and *H. influenzae* (34 to 91 %). However, there was no Gap homologue in *A. pleuropneumoniae*, no G6pd homologues in *M. glucosida* and *P. trehalosi* (unsuccessful amplification in this study), no MtlD homologues in *H. ducreyi* and

Table 3.1 Size of gene and sequenced fragment and number of isolates and alleles analyzed for 19 genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Gene name	Size of gene [a]	Size of sequenced fragment	No. of isolates / No. of alleles		
	No. of nucleotides	No. of nucleotides [b] (%)	*M. haemolytica*	*M. glucosida*	*P. trehalosi*
Genes encoding DNA repair and recombination enzymes					
recA	1107	810 (73)	10 / 3	1 / 1	1 / 1
Genes encoding metabolic enzymes					
aroA	1305	1239 (95)	32 / 11	6 / 5	4 / 4
asd	1116	846 (76)	10 / 5	1 / 1	1 / 1
galE	1017	933 (92)	10 / 4	1 / 1	1 / 1
gap	1005	870 (87)	32 / 10	6 / 5	4 / 4
gnd	1455	1281 (88)	10 / 3	1 / 1	1 / 1
g6pd	1488	1353 (91)	10 / 4	ND	ND
mdh	969	780 (80)	10 / 6	1 / 1	1 / 1
mtlD	1140	951 (83)	17 / 6	6 / 5	1 / 1
pmm	1464	729 (50)	32 / 10	6 / 5	4 / 4
Genes encoding secreted proteins					
gcp	978	870 (89)	10 / 3	1 / 1	1 / 1

Table 3.1 (continued)

Gene name	Size of gene [a] No. of nucleotides	Size of sequenced fragment No. of nucleotides [b] (%)	No. of isolates / No. of alleles *M. haemolytica*	*M. glucosida*	*P. trehalosi*
Genes encoding periplasm-associated proteins					
plpA	837	624 (75)	10 / 2	1 / 1	1 / 1
plpB	831	831 (100)	10 / 4	1 / 1	ND
plpC	780	771 (99)	10 / 6	1 / 1	ND
plpD	855	651 (76)	11 / 6	1 / 1	1 / 1
Genes encoding outer membrane proteins					
ompA	1137	1137 (100)	31 / 11	6 / 3	4 / 4 + 4*
tbpB	1755	1677 (95)	32 / 20	6 / 5	4 / 4
tbpA	2793	2748 (98)	32 / 12	6 / 5	4 / 3
wza	1185	966 (82)	32 / 12	4 / 3	4 / 3

[a] The source of the complete gene is indicated in Materials and Methods.

[b] The size of the sequenced fragment is based on *M. haemolytica* isolate PH2

ND: not determined (unable to amplify)

* *ompA'* + *ompA"*

Table 3.2 Comparison of the proteins encoded by *M. haemolytica* isolate PH2 with the corresponding proteins of seven members of the *Pasteurellaceae*

Protein and organism	Protein [a] identity (%)	GenBank accession No.
DNA repair and recombination proteins		
RecA		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH246	88	This study
A. pleuropneumoniae isolate 4074	90	ZP_00134135
H. ducreyi isolate 35000hp	89	NP_872986
H. somnus isolate 2336	79	ZP_00132641
P. multocida isolate PM70	77	NP_246756
H. influenzae isolate KW20	77	NP_438757
Metabolic enzymes		
AroA		
M. glucosida isolate PH344	96	This study
P. trehalosi isolate PH246	85	This study
A. pleuropneumoniae isolate 4074	81	ZP_00134045
H. ducreyi isolate 35000hp	71	NP_873807
H. somnus isolate 2336	80	P_00132008
P. multocida isolate PM70	74	NP_245776
H. influenzae isolate KW20	78	NP_439734
Asd		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH246	93	This study
A. pleuropneumoniae isolate 4074	93	ZP_00134096
H. ducreyi isolate 35000hp	81	NP_873349
H. somnus isolate 2336	91	ZP_00131937
P. multocida isolate PM70	86	NP_246571
H. influenzae isolate KW20	91	NP_438806
GalE		
M. glucosida isolate PH344	99	This study
P. trehalosi isolate PH246	88	This study
A. pleuropneumoniae isolate 4074	81	ZP_00204476
H. ducreyi isolate 35000hp	80	NP_873335
H. somnus isolate 2336	80	ZP_00133678
P. multocida isolate PM70	77	Q9CNY5
H. influenzae isolate KW20	76	NP_438515
Gap		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH246	91	This study
A. pleuropneumoniae isolate 4074	-	-
H. ducreyi isolate 35000hp	92	NP_873724

Table 3.2 (continued)

Protein and organism	Protein [a] identity (%)	GenBank accession No.
H. somnus isolate 2336	90	ZP_00131873
P. multocida isolate PM70	90	NP_245861
H. influenzae isolate KW20	85	NP_438174
Gnd		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH246	90	This study
A. pleuropneumoniae isolate 4074	89	ZP_00135245
H. ducreyi isolate 35000hp	86	NP_873339
H. somnus isolate 2336	90	ZP_00131777
P. multocida isolate PM70	87	NP_246493
H. influenzae isolate KW20	91	NP_438711
G6pd		
M. glucosida isolate PH344	-	This study
P. trehalosi isolate PH246	-	This study
A. pleuropneumoniae isolate 4074	88	ZP_00135250
H. ducreyi isolate 35000hp	84	NP_873342
H. somnus isolate 2336	82	ZP_00131780
P. multocida isolate PM70	83	NP_246488
H. influenzae isolate KW20	82	NP_438715
Mdh		
M. glucosida isolate PH344	98	This study
P. trehalosi isolate PH246	72	This study
A. pleuropneumoniae isolate 4074	65	ZP_00135237
H. ducreyi isolate 35000hp	63	NP_872857
H. somnus isolate 2336	31	ZP_00122604
P. multocida isolate PM70	60	Q9CN86
H. influenzae isolate KW20	60	NP_439366
MtlD		
M. glucosida isolate PH344	96	This study
P. trehalosi isolate PH246	89	This study
A. pleuropneumoniae isolate 4074	82	ZP_00348359
H. ducreyi isolate 35000hp	-	-
H. somnus isolate 2336	62	ZP_00133150
P. multocida isolate PM70	64	Q9CLY7
H. influenzae isolate KW20	-	-
Pmm		
M. glucosida isolate PH344	99	This study
P. trehalosi isolate PH246	90	This study
A. pleuropneumoniae isolate 4074	92	ZP_00134557
H. ducreyi isolate 35000hp	83	NP_873912
H. somnus isolate 2336	90	ZP_00131762
P. multocida isolate PM70	88	NP_246011

Table 3.2 (continued)

Protein and organism	Protein [a] identity (%)	GenBank accession No.
H. influenzae isolate KW20	86	NP_438900
Secreted proteins		
Gcp		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH246	88	This study
A. pleuropneumoniae isolate 4074	86	ZP_00135231
H. ducreyi isolate 35000hp	76	NP_873041
H. somnus isolate 2336	83	ZP_00132762
P. multocida isolate PM70	77	NP_246175
H. influenzae isolate KW20	76	NP_438688
Periplasm-associated proteins		
PlpA		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH68	85	This study
A. pleuropneumoniae isolate 4074	81	ZP_00133876
H. ducreyi isolate 35000hp	72	NP_873577
H. somnus isolate 2336	79	ZP_00204652
P. multocida isolate PM70	59	NP_246668
H. influenzae isolate KW20	69	NP_438778
PlpB		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH68	-	
A. pleuropneumoniae isolate 4074	59	ZP_00133875
H. ducreyi isolate 35000hp	-	-
H. somnus isolate 2336	-	-
P. multocida isolate PM70	-	-
H. influenzae isolate KW20	-	-
PlpC		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH68	-	-
A. pleuropneumoniae isolate 4074	59	ZP_00204591
H. ducreyi isolate 35000hp	-	-
H. somnus isolate 2336	-	-
P. multocida isolate PM70	-	-
H. influenzae isolate KW20	-	-
PlpD		
M. glucosida isolate PH344	100	This study
P. trehalosi isolate PH68	85	This study
A. pleuropneumoniae isolate 4074	89	ZP_00134888
H. ducreyi isolate 35000hp	-	-
H. somnus isolate 2336	42	ZP_00204642

Table 3.2 (continued)

Protein and organism	Protein [a] identity (%)	GenBank accession No.
P. multocida isolate PM70	78	NP 245523
H. influenzae isolate KW20	-	-
Outer membrane proteins		
OmpA		
M. glucosida isolate PH344	93	This study
P. trehalosi isolate PH246	69 + 63*	This study
A. pleuropneumoniae isolate 4074	65	ZP_00134486
H. ducreyi isolate 35000hp	62 + 63*	NP_872672 +
H. somnus isolate 2336	63	ZP_00132943
P. multocida isolate PM70	51	NP_245723
H. influenzae isolate KW20	52	NP_439322
TbpB		
M. glucosida isolate PH344	94	This study
P. trehalosi isolate PH246	78	This study
A. pleuropneumoniae isolate 4074	38	ZP_00135181
H. ducreyi isolate 35000hp	-	-
H. somnus isolate 2336	34	ZP_00122045
P. multocida isolate PM70	-	-
H. influenzae isolate KW20	34	NP_439158
TbpA		
M. glucosida isolate PH344	98	This study
P. trehalosi isolate PH246	94	This study
A. pleuropneumoniae isolate 4074	45	ZP_00135182
H. ducreyi isolate 35000hp	-	-
H. somnus isolate 2336	50	ZP_00122044
P. multocida isolate PM70	-	-
H. influenzae isolate KW20	40	NP_439157
Wza		
M. glucosida isolate PH344	96	This study
P. trehalosi isolate PH246	70	This study
A. pleuropneumoniae isolate 4074	79	ZP_00133758
H. ducreyi isolate 35000hp	-	-
H. somnus isolate 2336	-	-
P. multocida isolate PM70	50	NP_245715
H. influenzae isolate KW20	-	-

[a] Protein identity (%) was calculated against protein of *M. haemolytica* isolate PH2

H. influenzae, no TbpB and TbpA homologues in *H. ducreyi* and *P. multocida*, and no Wza homologues in *H. ducreyi*, *H. somnus*, and *H. influenzae*.

The nucleotide and amino acid sequence diversities of the 19 genes were determined separately for *M. haemolytica* isolates and also for all *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates (Table 3.3). The nucleotide and amino acid diversities of *recA*, *asd*, *galE*, *gnd*, *g6pd*, *mtlD*, *gcp*, and *plpD* within *M. haemolytica* (0.2 to 0.8 % and 0.0 to 1.4 %, respectively) were relatively low, while the corresponding diversities of *aroA*, *gap*, *mdh*, *pmm*, *plpA*, *plpB*, *plpC*, *ompA*, *tbpB*, *tbpA*, and *wza* within *M. haemolytica* (2.3 to 45.2 % and 0.5 to 49.8 %, respectively) were relatively high. In the case of the *aroA*, *gap*, *pmm*, *ompA*, *tbpB*, *tbpA* and *wza* genes, which were sequenced in a larger number of *M. glucosida* and *P. trehalosi* isolates, the nucleotide and amino acid sequence diversities were also determined separately for isolates of *M. glucosida* and *P. trehalosi*. With the exception of *ompA'* (12.3 % and 10.8 %, respectively), the nucleotide and amino acid diversity was relatively low in *P. trehalosi* (0.7 to 2.3 % and 0.0 to 1.6 %, respectively). The higher diversity of *ompA'* in *P. trehalosi* was due to variation in one isolate (PH254), since the other three isolates were highly conserved (0.9 and 0.2 %, respectively). In contrast, the corresponding diversities were higher in *M. glucosida* (0.8 to 16.0 % and 0.0 to 7.5 %) and were even higher in *M. haemolytica* (2.3 to 45.2 % and 1.0 to 49.8 %, respectively). The reasons for this will be discussed later.

The numbers of synonymous substitutions per 100 synonymous sites (d_S) and the number of nonsynonymous substitutions per 100 nonsynonymous sites (d_N) (Nei & Gojobori, 1986) were estimated for the 19 genes (Table 3.3). When the nucleotide diversity was > 1 %, the d_S/d_N ratios were also calculated. A high d_S/d_N ratio indicates that natural selection at the molecular level is purifying (conservative),

Table 3.3 Nucleotide and amino acid sequence diversity, d_S and d_N values, and d_S/d_N ratios for conserved genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Gene and organism [a] (No. of strains)	No of polymorphic sites		Mean ± SD [b]		d_S/d_N [c]
	Nucleotides (%)	Amino acids (%)	d_S	d_N	
Genes encoding DNA repair and recombination enzymes					
recA					
Mh (10)	2 (0.2)	0 (0.0)	0.39 ± 0.02	0.00 ± 0.00	-
Mh (10) + *Mg* (1) + *Pt* (1)	157 (19.4)	33 (12.2)	19.10 ± 3.12	1.30 ± 0.22	14.69
Genes encoding metabolic enzymes					
aroA					
Mh (32)	46 (3.7)	11 (2.7)	4.54 ± 0.79	0.36 ± 0.11	12.61
Mg (6)	94 (7.6)	15 (3.6)	12.22 ± 1.40	0.67 ± 0.18	18.24
Pt (4)	28 (2.3)	5 (1.2)	4.69 ± 0.96	0.26 ± 0.11	18.04
Mh (32) + *Mg* (6) + *Pt* (4)	360 (29.1)	80 (19.4)	32.68 ± 3.45	2.90 ± 0.34	11.27
asd					
Mh (10)	5 (0.6)	3 (1.1)	0.45 ± 0.31	0.13 ± 0.08	-
Mh (10) + *Mg* (1) + *Pt* (1)	158 (18.7)	23 (8.2)	22.25 ± 4.27	1.03 ± 0.20	21.60
galE					
Mh (10)	4 (0.4)	3 (1.0)	0.09 ± 0.09	0.13 ± 0.08	-
Mh (10) + *Mg* (1) + *Pt* (1)	214 (22.9)	40 (12.9)	28.60 ± 5.48	1.53 ± 0.25	18.69
gap					
Mh (32)	20 (2.3)	3 (1.0)	2.93 ± 0.71	0.05 ± 0.03	58.60
Mg (6)	60 (6.9)	0 (0.0)	13.42 ± 1.77	0.00 ± 0.00	-
Pt (4)	6 (0.7)	0 (0.0)	1.50 ± 0.59	0.00 ± 0.00	-
Mh (32) + *Mg* (6) + *Pt* (4)	170 (19.5)	28 (9.7)	16.52 ± 1.78	1.20 ± 0.24	13.77
gnd					
Mh (10)	2 (0.2)	0 (0.0)	0.34 ± 0.23	0.00 ± 0.00	-
Mh (10) + *Mg* (1) + *Pt* (1)	234 (18.3)	41 (9.6)	17.96 ± 2.04	1.10 ± 0.17	16.33
g6pd					
Mh (10)	5 (0.4)	2 (0.4)	0.38 ± 0.24	0.12 ± 0.08	-
mdh					
Mh (10)	36 (4.6)	4 (1.5)	4.05 ± 0.74	0.14 ± 0.07	28.93
Mh (10) + *Mg* (1) + *Pt* (1)	261 (33.5)	76 (29.2)	29.39 ± 4.38	3.76 ± 0.42	7.82
mtlD					
Mh (17)	7 (0.7)	4 (1.3)	0.26 ± 0.15	0.14 ± 0.07	-
Mg (6)					
Mh (17) + *Mg* (6) + *Pt* (1)	259 (27.2)	53 (16.7)	11.36 ± 7.67	1.60 ± 0.28	7.10
pmm					
Mh (32)	94 (12.9)	15 (6.2)	9.44 ± 1.47	0.60 ± 0.21	15.73
Mg (6)	39 (5.3)	0 (0.0)	10.74 ± 1.75	0.00 ± 0.00	-
Pt (4)	5 (0.7)	1 (0.4)	1.20 ± 0.56	0.09 ± 0.09	13.33
Mh (32) + *Mg* (6) + *Pt* (4)	193 (26.5)	34 (14.0)	34.13 ± 5.27	1.74 ± 0.34	19.61
Genes encoding secreted proteins					
gcp					
Mh (10)	2 (0.2)	2 (0.7)	0.00 ± 0.00	0.10 ± 0.07	-
Mh (10) + *Mg* (1) + *Pt* (1)	228 (26.2)	37 (12.8)	8.42 ± 1.90	1.79 ± 0.26	4.70

Table 3.3 (Continued)

Gene and organism [a] (No. of strains)	No of polymorphic sites		Mean ± SD [b]		d_s/d_N [c]
	Nucleotides (%)	Amino acids (%)	d_s	d_N	
Genes encoding periplasm-associated proteins					
plpA					
Mh (10)	25 (4.0)	1 (0.5)	3.89 ± 0.88	0.04 ± 0.04	97.25
Mh (10) + *Mg* (1) + *Pt* (1)	154 (24.7)	30 (14.4)	27.89 ± 6.93	1.95 ± 0.36	14.30
plpB					
Mh (10)	57 (6.9)	9 (3.3)	6.31 ± 0.91	0.31 ± 0.10	20.35
Mh (10) + *Mg* (1)	61 (7.3)	10 (3.6)	6.76 ± 0.98	0.31 ± 0.10	21.81
*plpC**					
Mh (9)	41 (5.4)	6 (2.4)	5.13 ± 0.90	0.26 ± 0.10	19.73
Mh (9) + *Mg* (1)	44 (5.8)	7 (2.8)	8.04 ± 1.34	0.32 ± 0.11	25.13
plpD[§]					
Mh (10)	5 (0.8)	3 (1.4)	0.16 ± 0.16	0.35 ± 0.19	-
Mh (10) + *Mg* (1) + *Pt* (1)	135 (20.7)	36 (16.6)	20.82 ± 4.35	2.18 ± 0.38	9.55
Genes encoding outer membrane proteins					
*ompA**					
Mh (31)	49 (4.3)	28 (7.4)	0.84 ± 0.31	1.33 ± 0.30	0.63
Mg (6)	26 (2.3)	7 (1.9)	3.28 ± 0.80	0.65 ± 0.24	5.05
Pt' (4)	140 (12.3)	41 (10.8)	17.86 ± 2.13	4.09 ± 0.62	4.37
Pt" (6)	9 (0.8)	0 (0.0)	1.84 ± 0.57	0.00 ± 0.00	-
Mh (31) + *Mg* (6) + *Pt* (10)	438 (38.5)	150 (39.7)	19.71 ± 1.68	7.62 ± 0.70	2.59
tbpB					
Mh (32)	758 (45.2)	278 (49.8)	57.56 ± 7.98	12.71 ± 0.77	4.53
Mg (6)	54 (3.2)	18 (3.2)	4.41 ± 0.73	0.73 ± 0.18	6.04
Pt (4)	14 (0.8)	9 (1.6)	0.80 ± 0.35	0.40 ± 0.12	2.00
Mh (32) + *Mg* (6) + *Pt* (4)	850 (50.7)	299 (53.6)	68.16 ± 7.07	12.09 ± 0.67	5.64
tbpA					
Mh (32)	687 (25.0)	168 (18.3)	26.01 ± 1.68	2.69 ± 0.21	9.67
Mg (6)	23 (0.8)	10 (1.1)	0.68 ± 0.18	0.18 ± 0.06	3.78
Pt (4)	34 (1.2)	7 (0.8)	2.39 ± 0.45	0.20 ± 0.07	11.95
Mh (32) + *Mg* (6) + *Pt* (4)	727 (26.5)	185 (20.2)	21.60 ± 1.32	2.74 ± 0.23	7.88
wza					
Mh (32)	241 (24.9)	45 (14.0)	33.57 ± 3.20	1.76 ± 0.30	19.07
Mg (4)	155 (16.0)	24 (7.5)	46.95 ± 6.60	2.29 ± 0.44	20.50
Pt (4)	14 (1.4)	4 (1.2)	1.97 ± 0.67	0.37 ± 0.20	5.32
Mh (32) + *Mg* (4) + *Pt* (4)	414 (42.9)	118 (36.6)	56.18 ± 4.96	5.87 ± 0.56	9.57

[a] *Mh*: *M. haemolytica*, *Mg*: *M. glucosida*, *Pt*: *P. trehalosi*

[b] d_s and d_N represent the number of synonymous substitutions per 100 synonymous sites and the number of nonsynonymous substitutions per 100 nonsynonymous sites, respectively.

[c] d_s/d_N was calculated when nucleotide diversity was > 1 %.

* Isolate PH202 has been removed from the *Mh* calculation because this isolate has a nucleotide insertion in the middle of *plpC*, which disrupts the reading frame.

[§] Isolates PH494 and PH550 have been removed from the *Mh* calculation because these isolates have a 13 nucleotides deletion in the middle of *plpD*, which disrupts the reading frame.

* *Pt'* indicates *P. trehalosi ompA'* and *Pt"* indicates *P. trehalosi ompA"*.

acting against mutations resulting in amino acid replacements. Conversely, a d_S/d_N ratio of < 1 indicates that selection is diversifying and favours amino acid replacement. Four metabolic enzyme genes, *aroA*, *gap*, *mdh*, and *pmm*, had relatively high d_S/d_N ratios in *M. haemolytica* (12.61 to 58.60) whereas two outer membrane protein genes, *ompA* and *tbpB* had relatively low d_S/d_N ratios in *M. haemolytica* (0.63 and 4.53, respectively). The d_S/d_N ratios of the other outer membrane protein genes, *tbpA*, and *wza*, were 9.67 and 19.07, respectively. The reasons for this will be discussed later.

3.1.1 Details of nucleotide analyses for specific genes

3.1.1.1 Genes encoding DNA repair and recombination enzymes

3.1.1.1.1 Recombinase A (*recA*)

The *recA* gene, encoding recombinase A, was selected because it plays an important role in DNA repair and recombination and because it has been used previously for population studies (see section 1.6.1.1). The partial sequence of *recA* (810 nucleotides [73%]) was determined in 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate (Table 3.4). Five different *recA* nucleotide sequences, each representing a distinct allele, were identified, and the alleles were assigned to three subclasses, *recA1* to *recA3*, based on their overall sequence similarities (Table 3.4).

The polymorphic nucleotide and inferred amino acid sites within the five *recA* alleles, with respect to *recA1.1*, are shown in Figure 3.1. Nucleotide sequence analysis indicates that the *M. haemolytica recA1*-type alleles, *M. glucosida recA2.1* allele, and *P. trehalosi recA3.1* allele have 2, 9, and 157 polymorphic nucleotide sites,

Table 3.4 Distribution of *recA* alleles among 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	*recA* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*recA1.1*	AY837586
PH66	10	A14	Ovine	*recA1.2*	AY837587
PH706	11	A16	Ovine	*recA1.2*	-
PH296	12	A7	Ovine	*recA1.3*	AY837588
PH588	15	A13	Ovine	*recA1.3*	-
PH494	16	A2	Ovine	*recA1.2*	-
PH196	18	A2	Bovine	*recA1.3*	-
PH202	21	A2	Bovine	*recA1.3*	-
PH278	21	A2	Ovine	*recA1.3*	-
PH292	22	A2	Ovine	*recA1.3*	-
M. glucosida					
PH344	1	A11	Ovine	*recA2.1*	AY837589
P. trehalosi					
PH246	2	T4	Ovine	*recA3.1*	AY837590

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

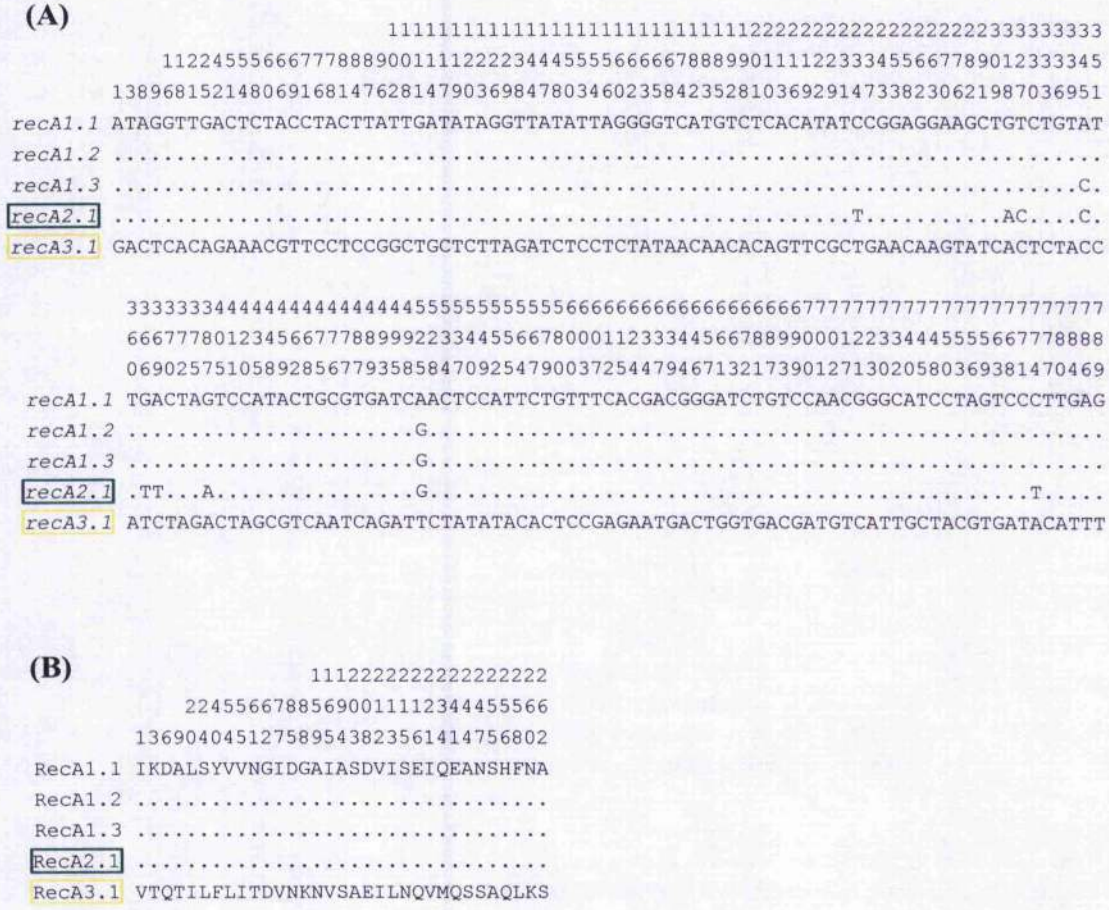


Figure 3.1 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *recA* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *recA1.1*).

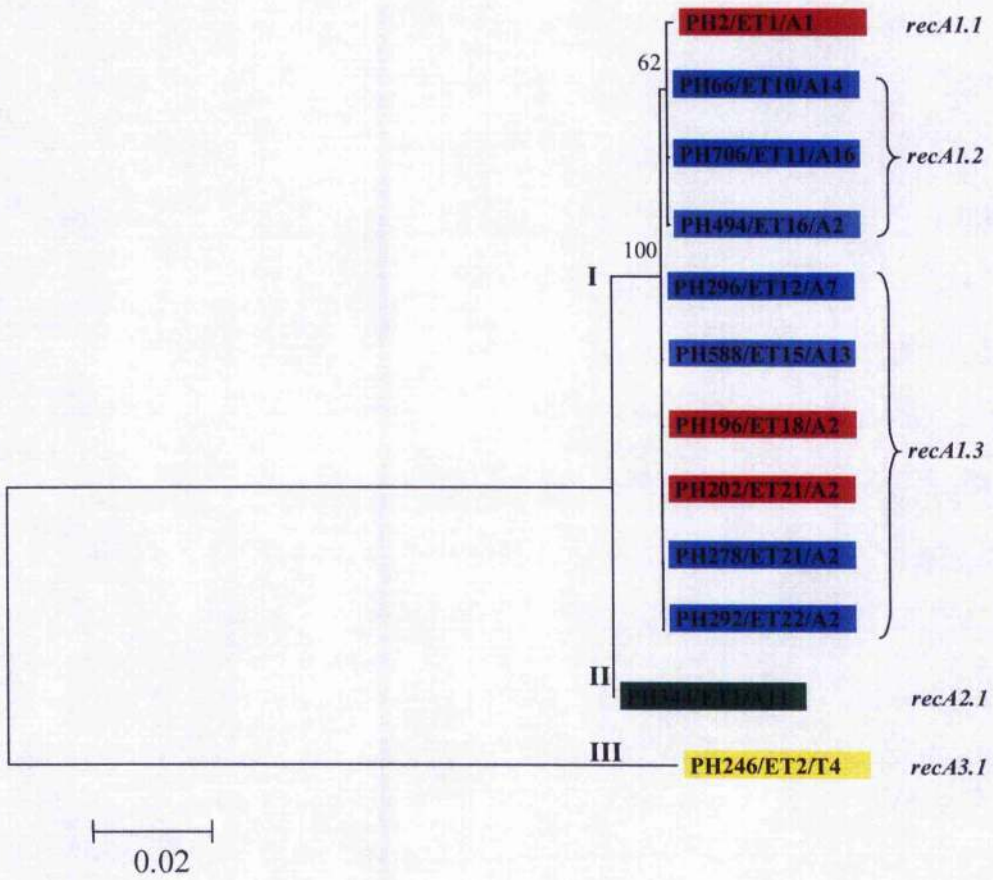


Figure 3.2 Minimum evolution (ME) tree for the *recA* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

respectively (Figure 3.1A). Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.1B). Two amino acid variants were identified among the five *recA* alleles; the *M. haemolytica* (*recA1*-types) and *M. glucosida* (*recA2.1*) alleles have identical amino acid sequences while the *P. trehalosi recA3.1* allele has 33 variable amino acid positions compared to the *M. haemolytica* (*recA1*-type) and *M. glucosida* (*recA2.1*) alleles. There was no evidence of intragenic recombination within any of the *recA* alleles.

Pairwise differences in nucleotide sequences indicated that the *M. haemolytica recA1*-type alleles are highly conserved (1 to 2 polymorphic sites), whereas the *M. glucosida recA2.1* allele shows moderate divergence from the *M. haemolytica recA1*-type alleles (7 to 9 polymorphic sites), and the *P. trehalosi recA3.1* allele exhibits substantial divergence from both the *M. haemolytica recA1*-type alleles (156 to 157 polymorphic sites) and the *M. glucosida recA2.1* allele (151 polymorphic sites).

The *recA* tree topology (Figure 3.2) shows three distinct phylogenetic lineages, I to III, corresponding to the three species, *M. haemolytica* (*recA1*), *M. glucosida* (*recA2*), and *P. trehalosi* (*recA3*), respectively. This phylogeny is in good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) and indicates that recombination has not affected the evolution of *recA*.

3.1.1.2 Genes encoding metabolic enzymes

3.1.1.2.1 5-enolpyruvylshikimate 3-phosphate synthase (*aroA*)

The *aroA* gene, encoding 5-enolpyruvylshikimate 3-phosphate synthase, was selected because it is an important housekeeping gene for bacterial growth (see section 1.6.2.1).

The partial sequence of *aroA* (1239 nucleotides [95 %]) was determined in 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates (Table 3.5). Twenty different *aroA* sequences, each representing a distinct allele, were identified, and the alleles were assigned to five subclasses, *aroA1* to *aroA5*, based on their overall sequence similarities (Table 3.5). *M. haemolytica* was represented by *aroA1* to *aroA3* type alleles, *M. glucosida* by *aroA4* type alleles, *P. trehalosi* by *aroA5* type alleles

The polymorphic nucleotide and inferred amino acid sites within the 16 *aroA1*- to *aroA4*-type alleles of *M. haemolytica* and *M. glucosida* are shown in Figure 3.3. The *P. trehalosi aroA5*-type alleles were excluded because they are highly divergent from both *M. haemolytica* and *M. glucosida* (Table 3.6) and there was no visible evidence of recombination. Visual inspection of the nucleotide sequences of the *aroA1*- to *aroA4*-type alleles indicated that the *M. haemolytica aroA1*-type alleles are highly conserved, whereas the *M. haemolytica aroA2.1* and *aroA3*-type alleles have diverged from the *aroA1*-type alleles due to intragenic recombination (Figure 3.3A). The *aroA2.1* allele (ovine serotype A2 isolates of ETs 19 to 22) contains a recombinant segment (green, nucleotides 330 to 565) that is identical to the corresponding region of the *aroA3*-type alleles. The *aroA3*-type alleles (serotype A14, A16, A7, and A2 isolates of ETs 10 to 14 and ET18) contain divergent recombinant segments (yellow, nucleotides 42 to 565) that have possibly been derived from an external source.

The nucleotide substitutions in the recombinant segments were mostly synonymous because only four and seven amino acid changes occurred within the green and yellow regions, respectively (Figure 3.3B), suggesting strong selective constraint for amino acid replacement within the *aroA* gene of *M. haemolytica*.

Table 3.5 Distribution of *aroA* alleles among 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Isolate	ET [a]	Capsular serotype	Host species	*aroA* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*aroA1.1*	AY847793
PH30	1	A1	Bovine	*aroA1.1*	-
PH376	1	A6	Bovine	*aroA1.1*	-
PH346	1	A12	Ovine	*aroA1.3*	AY847794
PH540	2	A1	Bovine	*aroA1.2*	AY847795
PH338	3	A9	Ovine	*aroA1.3*	-
PH388	4	A7	Ovine	*aroA1.3*	-
PH50	5	A5	Ovine	*aroA1.3*	-
PH56	5	A8	Ovine	*aroA1.3*	-
PH238	5	A9	Ovine	*aroA1.3*	-
PH8	6	A1	Ovine	*aroA1.6*	AY847796
PH398	7	A1	Ovine	*aroA1.6*	-
PH284	8	A6	Ovine	*aroA1.7*	AY847797
PH232	9	A6	Ovine	*aroA1.7*	-
PH66	10	A14	Ovine	*aroA3.1*	AY847798
PH706	11	A16	Ovine	*aroA3.1*	-
PH296	12	A7	Ovine	*aroA3.3*	AY847799
PH396	13	A7	Ovine	*aroA3.3*	-
PH484	14	A7	Ovine	*aroA3.3*	-
PH588	15	A13	Ovine	*aroA1.3*	-
PH494	16	A2	Ovine	*aroA1.5*	AY847800
PH550	17	A2	Bovine	*aroA1.5*	-
PH196	18	A2	Bovine	*aroA3.2*	AY847801
PH786	18	A2	Bovine	*aroA3.2*	-
PH526	19	A2	Ovine	*aroA2.1*	-
PH598	20	A2	Ovine	*aroA2.1*	-
PH202	21	A2	Bovine	*aroA1.4*	AY847802
PH470	21	A2	Bovine	*aroA1.4*	-
PH278	21	A2	Ovine	*aroA2.1*	AY847803
PH372	21	A2	Ovine	*aroA2.1*	-
PH292	22	A2	Ovine	*aroA2.1*	-
PH392	22	A2	Ovine	*aroA2.1*	-
M. glucosida					
PH344	1	A11	Ovine	*aroA4.1*	AY847804
PH498	3	A11	Ovine	*aroA4.1*	-
PH240	5	A11	Ovine	*aroA4.2*	AY847805
PH496	7	UG3	Ovine	*aroA4.3*	AY847806
PH574	10	UG3	Ovine	*aroA4.5*	AY847807
PH290	16	UG3	Ovine	*aroA4.4*	AY847808
P. trehalosi					
PH246	2	T4	Ovine	*aroA5.1*	AY847809
PH252	4	T10	Ovine	*aroA5.2*	AY847810
PH254	15	T15	Ovine	*aroA5.3*	AY847811
PH68	19	T3	Ovine	*aroA5.4*	AY847812

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

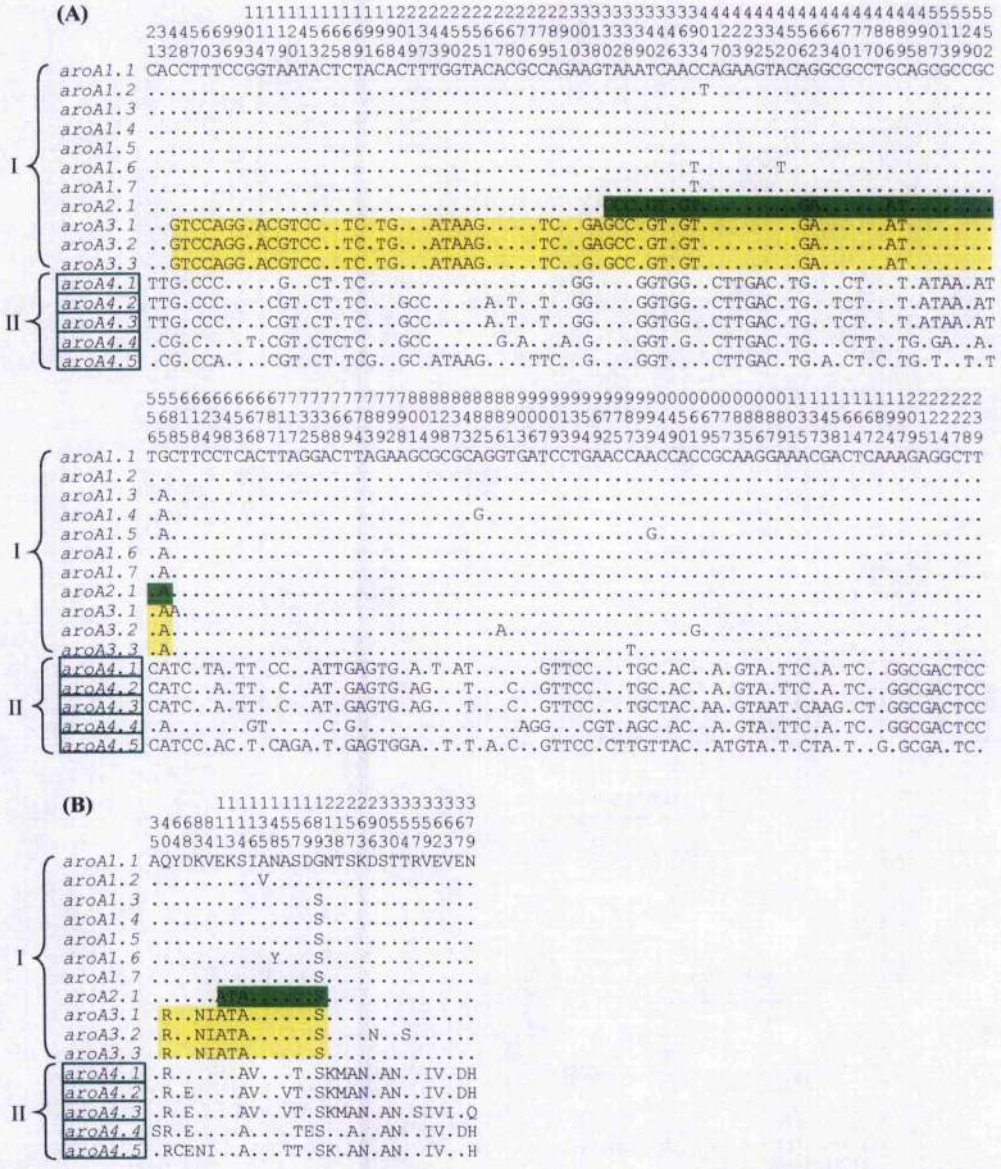


Figure 3.3 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *aroA* alleles of *M. haemolytica* and *M. glucosida*. Allele designations are shown to the left of each sequence. The Roman numerals I and II correspond to the major lineages of Figure 3.4. The *M. glucosida* alleles are highlighted within the green boxes. The coloured shading indicates sequence identity and putative recombinant segments in alleles that are otherwise different. The vertical numbers above the sequences represent the positions of polymorphic nucleotide or amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *aroA1.1*)

Table 3.6 Nucleotide and amino acid differences between representative pairs of the five *aroA* allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Allele	Pairwise differences in nucleotide and amino acid sequences (%) [a]				
	aroA1.1	*aroA2.1*	*aroA3.1*	*aroA4.1*	*aroA5.1*
aroA1.1		4 (1.0)	7 (1.7)	15 (3.6)	63 (15.3)
aroA2.1	12 (1.0)		3 (0.7)	15 (3.6)	63 (15.3)
aroA3.1	39 (3.1)	27 (2.2)		16 (3.9)	63 (15.3)
aroA4.1	86 (6.9)	86 (6.9)	95 (7.7)		63 (15.3)
aroA5.1	291 (23.5)	287 (23.2)	288 (23.2)	295 (23.8)	

[a] Values in the lower left represent pairwise differences in nucleotide sequences (number of polymorphic nucleotide sites and percentage of nucleotide variation); values in the upper right represent pairwise differences in inferred amino acid sequences (number of polymorphic amino acid sites and percentage of amino acid variation).

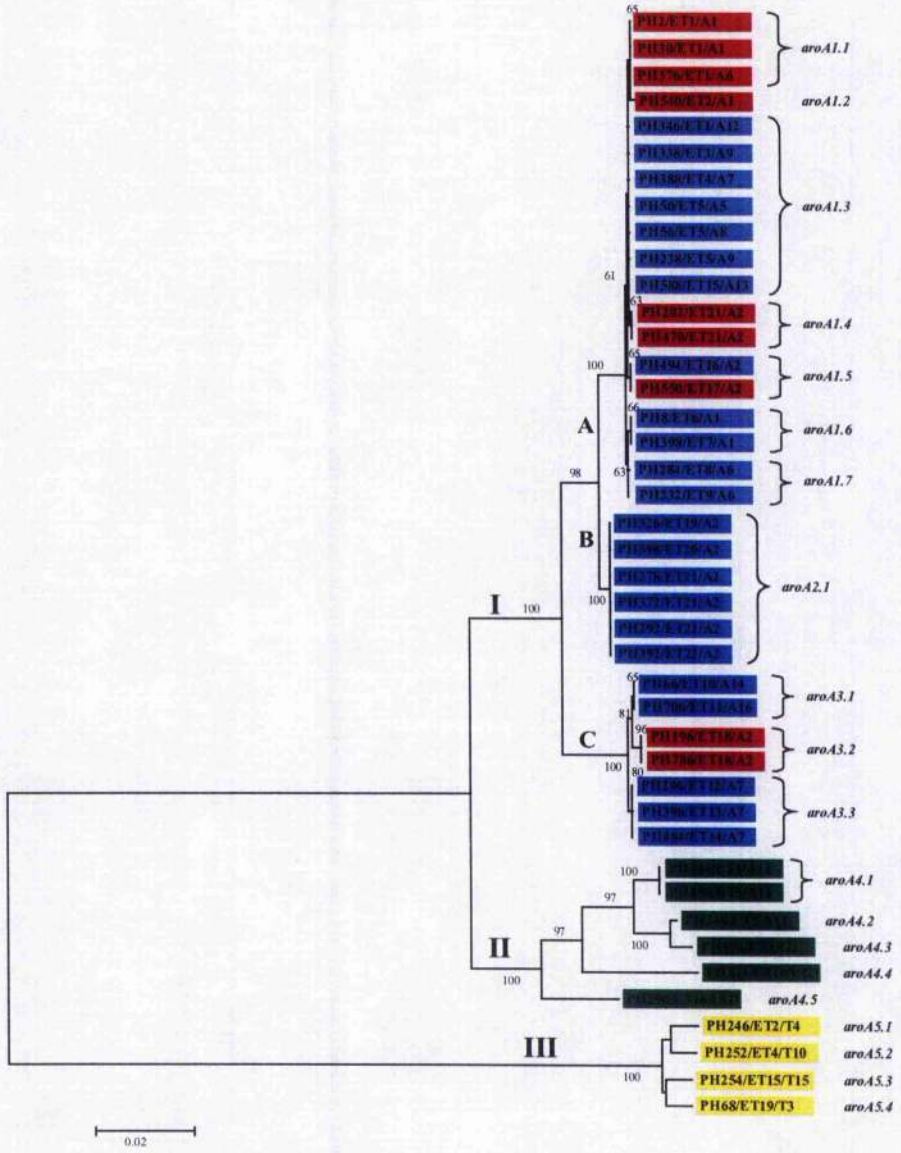


Figure 3.4 Minimum evolution (ME) tree for the *aroA* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

Pairwise differences in nucleotide and amino acid sequences between representative pairs of the five allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* ranged from 12 to 295 (1.0 to 17.0 %) nucleotide sites and 3 to 63 (0.7 to 15.3 %) amino acid positions (Table 3.6).

The *aroA* tree topology (Figure 3.4) shows three distinct phylogenetic lineages, I to III, corresponding to the three species, *M. haemolytica* (*aroA1* to *aroA3*), *M. glucosida* (*aroA4*), and *P. trehalosi* (*aroA5*), respectively. This phylogeny is in good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE data (Davies *et al.*, 1997a). However, the *M. haemolytica* isolates are represented by three clusters, A, B, and C, which correspond to *aroA1*-type, *aroA2.1* type, and *aroA3*-type alleles, respectively.

3.1.1.2.2 Aspartate-semialdehyde dehydrogenase (*asd*)

The *asd* gene, encoding aspartate semialdehyde dehydrogenase, was selected because the Asd protein is essential in the biosynthesis of cell wall material and some amino acids (see section 1.6.2.2). The partial sequence of *asd* (846 nucleotides [76 %]) was determined in 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate (Table 3.7). Seven different *asd* nucleotide sequences, each representing a distinct allele, were identified, and the alleles were assigned to three subclasses, *asd1* to *asd3*, based on their overall sequence similarities (Table 3.7).

The polymorphic nucleotide and inferred amino acid sites within the seven *asd* alleles, with respect to allele *asd1.1*, are shown in Figure 3.5. Nucleotide sequence analysis indicates that the *M. haemolytica asd1*-type, *M. glucosida asd2.1*, and *P. trehalosi asd3.1* alleles have 7, 14, and 150 polymorphic nucleotide sties, respectively

Table 3.7 Distribution of *asd* alleles among 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	*asd* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*asd1.1*	AY837573
PH66	10	A14	Ovine	*asd1.1*	-
PH706	11	A16	Ovine	*asd1.1*	-
PH296	12	A7	Ovine	*asd1.2*	AY837574
PH588	15	A13	Ovine	*asd1.2*	-
PH494	16	A2	Ovine	*asd1.3*	AY837575
PH196	18	A2	Bovine	*asd1.4*	AY837576
PH202	21	A2	Bovine	*asd1.5*	AY837577
PH278	21	A2	Ovine	*asd1.5*	-
PH292	22	A2	Ovine	*asd1.5*	-
M. glucosida					
PH344	1	A11	Ovine	*asd2.1*	AY837578
P. trehalosi					
PH246	2	T4	Ovine	*asd3.1*	AY837579

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

(A)

```
                         1111111111111111111222222222222222222222333333333333344444444
          2223347778889990001233444555667778000111234555667788999011334566778900111222
       3491474652340143780281706347036251473047369840035146929189058098719281928147039
asd1.1 TGACTCGTTTGTGCACGATACCTATTTCATCTTTATCACTCCTCTAAGCTTCGGCTTCACCTACACTCTACGTTAATCC
asd1.2 ...............................................................................
asd1.3 .........................C.....................................................
asd1.4 ...............................................................................
asd1.5 ...........................................T...................................
asd2.1 ..............GT........C...G..CCC.C...C...T...........C....T..................
asd3.1 ACGTCTACCAACCGCTACCCTTCC..CTTCTC.CCCTCA.TTCTAGGAAGCTCCA.GATT.CTATTATCGTAAAGGCTA

       4444444444444444455555555555555555555666666666677777777777777777888888888888888
       3344778888899999901111355556677788899011334456900113455566788890000111123333444
       2517140123902345840345725896703657847668792510302475103928703622478023921467036
asd1.1 TCCGTATGAAAGTAGCATAAAATTGTAGTGACCAGCTGTATAAGCTTTATCTCTAACCACATGATGTACGCGTTGCCTA
asd1.2 ......................G...................G....................................
asd1.3 ......................G........................................................
asd1.4 ......................G....A...................................................
asd1.5 ......................G........................................................
asd2.1 ......................G.T......................................................
asd3.1 ATGAAGCATTGCATCAGAGGGCGGTAC.AACTTCATGACTCT.ATCCCGCTCTCTGTTGTCCCTAACTAATACCCGTCG
```

(B)

```
                  111111111222
        1223446889066778899677
       22573487457014297916819
Asd1.1 EVVGEKIEKDHSEDNDKREEIAC
Asd1.2 ...............E.......
Asd1.3 ......T........E.......
Asd1.4 ...............E.H.....
Asd1.5 ...............E.......
Asd2.1 ...............E.......
Asd3.1 QITATQ.AEKNDIQAEQ.DALES
```

Figure 3.5 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *asd* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *asd1.1*).

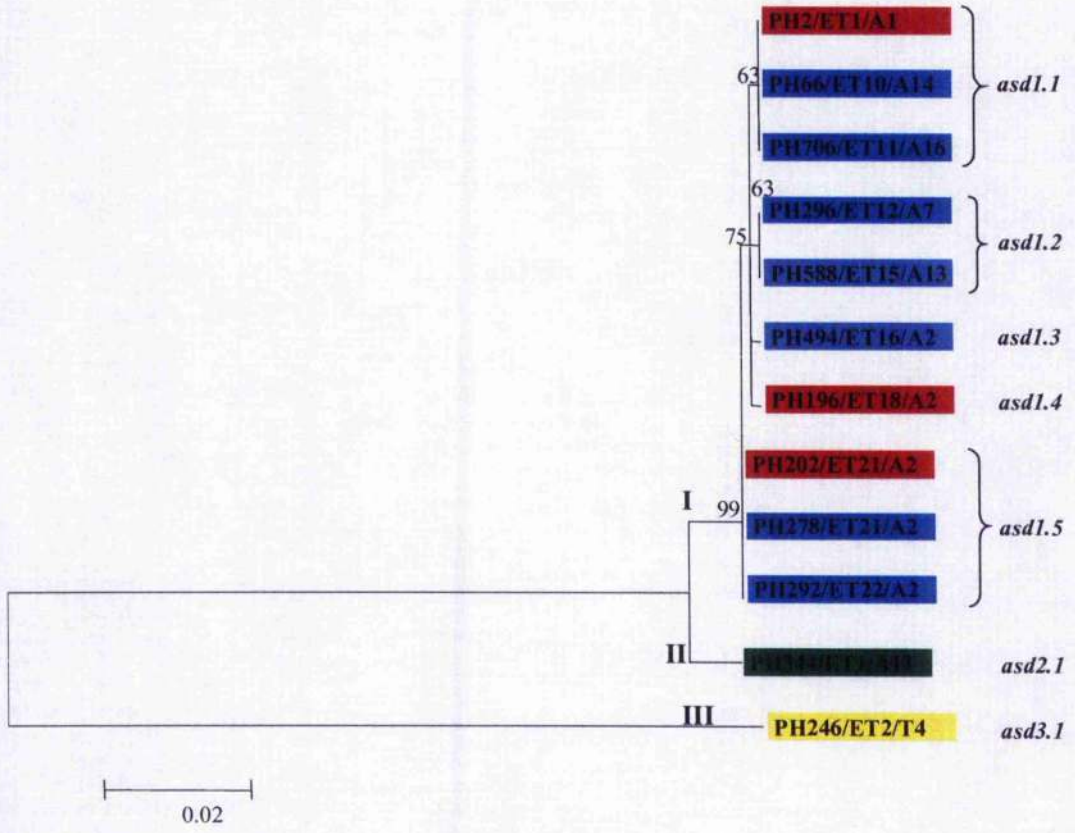


Figure 3.6 Minimum evolution (ME) tree for the *asd* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

(Figure 3.5A). Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.5B). Five amino acid variants were identified among the seven *asd* alleles; 3, 1, and 21 variable amino acid positions were detected in the *M. haemolytica asd1*-type, *M. glucosida asd2.1*, and *P. trehalosi asd3.1* alleles, with respect to the *M. haemolytica asd1.1* allele, respectively. There was no evidence of intragenic recombination within any of the *asd* alleles.

Pairwise differences in nucleotide sequences indicated that the *M. haemolytica asd1*-type alleles are highly conserved (two polymorphic sites), whereas the *M. glucosida asd2.1* allele shows moderate divergence from *M. haemolytica asd1*-type alleles (12 to 14 polymorphic sites), and the *P. trehalosi asd3.1* allele exhibits substantial divergence from both the *M. haemolytica asd1*-type alleles (148 to 150 polymorphic sites) and the *M. glucosida asd2.1* allele (148 polymorphic sites).

The *asd* tree topology (Figure 3.6) shows three distinct phylogenetic lineages, I to III, corresponding to the three species, *M. haemolytica* (*asd1*), *M. glucosida* (*asd2*), and *P. trehalosi* (*asd3*), respectively. This phylogeny is in good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) and indicates that recombination has not affected the evolution of *asd*.

3.1.1.2.3 UDP-galactose 4-epimerase (*galE*)

The *galE* gene, encoding UDP-galactose 4 epimerase, was selected because the enzyme is involved in carbohydrate metabolism as well as LPS biosynthesis (see section 1.6.2.3). The partial sequence of *galE* (933 nucleotides [92%]) was determined in 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate (Table 3.8). Six different *galE* sequences, each representing a

Table 3.8 Distribution of *galE* alleles among 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	*galE* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*galE1.1*	AY837580
PH66	10	A14	Ovine	*galE1.3*	AY837581
PH706	11	A16	Ovine	*galE1.1*	-
PH296	12	A7	Ovine	*galE1.3*	-
PH588	15	A13	Ovine	*galE1.1*	-
PH494	16	A2	Ovine	*galE1.1*	-
PH196	18	A2	Bovine	*galE1.4*	AY837582
PH202	21	A2	Bovine	*galE1.3*	-
PH278	21	A2	Ovine	*galE1.3*	-
PH292	22	A2	Ovine	*galE1.2*	AY837583
M. glucosida					
PH344	1	A11	Ovine	*galE2.1*	AY837584
P. trehalosi					
PH246	2	T4	Ovine	*galE3.1*	AY837585

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

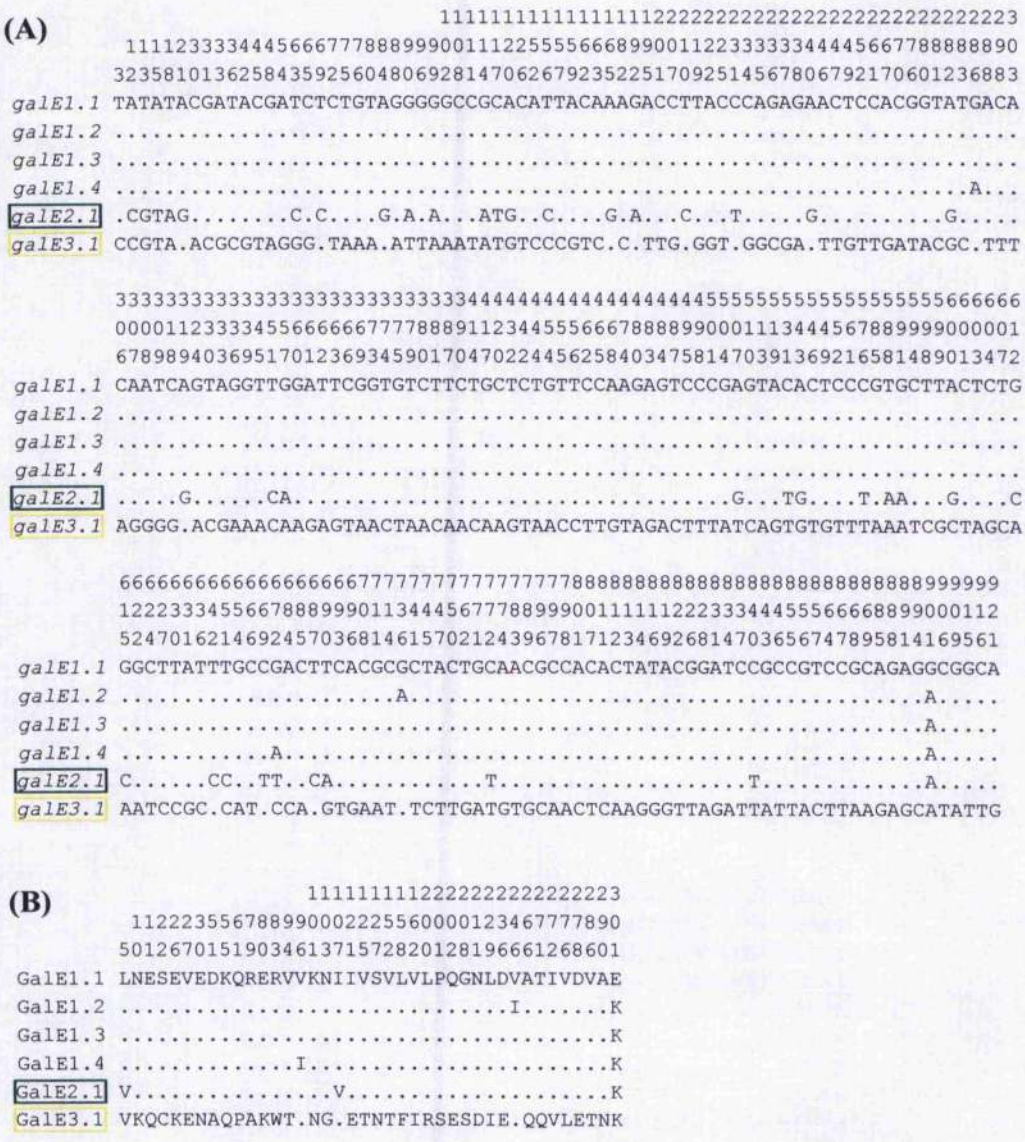


Figure 3.7 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *galE* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *galE1.1*).

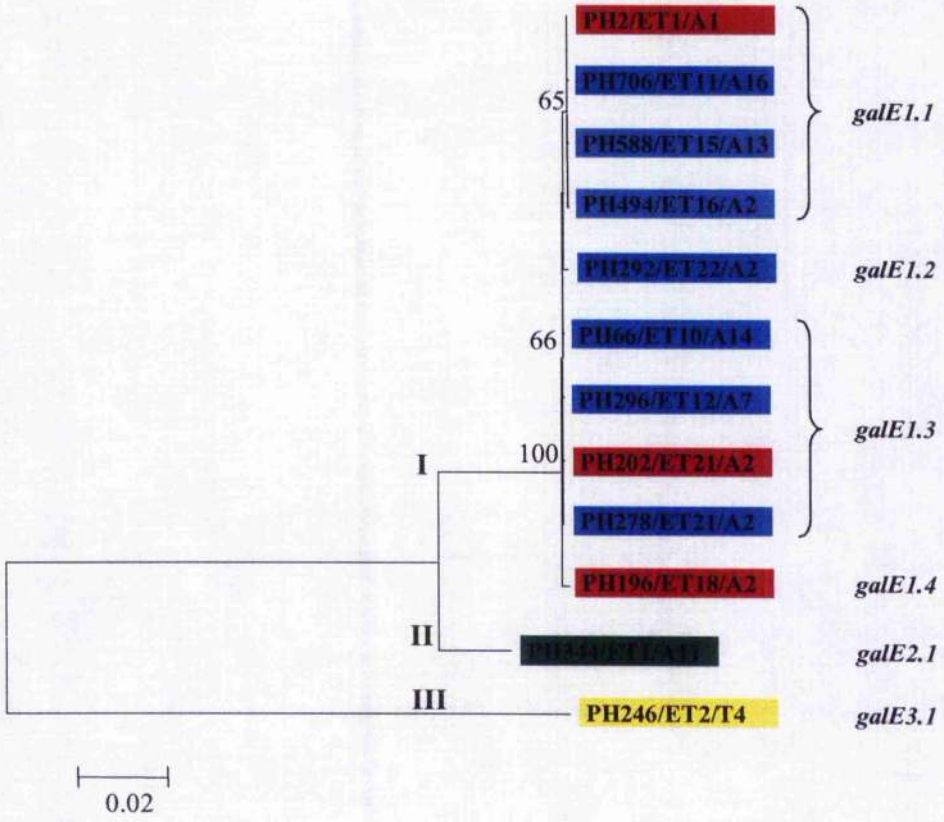


Figure 3.8 Minimum evolution (ME) tree for the *galE* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

distinct allele, were identified, and the alleles were assigned to three subclasses, *galE1* to *galE3*, based on their overall sequence similarities (Table 3.8).

The polymorphic nucleotide and inferred amino acid sites within the six *galE* alleles, with respect to *galE1.1*, are shown in Figure 3.7. Nucleotide sequence analysis indicates that the *M. haemolytica galE1*-type, *M. glucosida galE2.1*, and *P. trehalosi galE3.1* alleles have 4, 41, and 200 polymorphic nucleotide sites, respectively (Figure 3.7A). Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.7B). Six amino acid variants were identified among the six *galE* alleles; 3, 3, and 37 variable amino acid positions were detected in the *M. haemolytica galE1*-type, *M. glucosida galE2.1*, and *P. trehalosi galE3.1* alleles, with respect to the *M. haemolytica galE1.1* allele, respectively. There was no evidence of intragenic recombination within any of the *galE* alleles.

Pairwise differences in nucleotide sequences indicated that the *M. haemolytica galE1*-type alleles are highly conserved (one to three polymorphic sites), whereas the *M. glucosida galE2.1* allele shows moderate divergence from the *M. haemolytica galE1*-type alleles (40 to 41 polymorphic sites), and the *P. trehalosi galE3.1* allele exhibits substantial divergence from both the *M. haemolytica galE1*-type alleles (199 to 200 polymorphic sites) and the *M. glucosida galE2.1* allele (191 polymorphic sites).

The *galE* tree topology (Figure 3.8) shows three distinct phylogenetic lineages, I to III, corresponding to the three species, *M. haemolytica* (*galE1*), *M. glucosida* (*galE2*), and *P. trehalosi* (*galE3*), respectively. This phylogeny is in good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) and indicates that recombination has not affected the evolution of *galE*.

3.1.1.2.4 Glyceraldehyde-3-phosphate dehydrogenase (*gap*)

The *gap* gene, encoding glyceraldehyde-3-phosphate dehydrogenase, was selected because it is a housekeeping gene that has undergone gene transfer between eubacteria and eukaryotes as well as between different species of eubacteria (see section 1.6.2.4). The partial sequence of *gap* (870 nucleotides [87 %]) was determined in 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates (Table 3.9). Nineteen different *gap* sequences, each representing a distinct allele, were identified, and the alleles were assigned to six subclasses, *gap1* to *gap6*, based on their overall sequence similarities (Table 3.9)

The polymorphic nucleotide and inferred amino acid sites within the *gap1*- to *gap5*-type alleles of *M. haemolytica* and *M. glucosida* are shown in Figure 3.9. *P. trehalosi gap6*-type alleles were excluded because they are highly divergent from both *M. haemolytica* and *M. glucosida* (Table 3.10) and there was no visible evidence of recombination. Visual inspection of the nucleotide sequences of the *gap1*- to *gap5*-type alleles (Figure 3.9A) indicates that *M. haemolytica gap1* to *gap2*-type alleles are relatively conserved, whereas the *M. haemolytica gap3.1* allele has diverged from the *M. haemolytica gap1*- to *gap2*-type alleles due to intragenic recombination. The *gap3.1* allele (A2 isolates of ETs 19 to 22) contains a recombinant segment (yellow, nucleotides 339 to 612) that is very similar to the corresponding region of the *M. glucosida gap3.2* and *gap3.3* alleles. The *M. glucosida* isolates contain five allele types that have complex mosaic structures: *gap2.4*, *gap3.2* and *gap3.3*, *gap4.1*, and *gap5.1*. The *gap2.4* allele (UG3 isolate of ET 16) is very similar to the *M. haemolytica gap2*-type alleles, indicating that assortative (entire gene) recombination has occurred. The *gap3.3* allele (A11 isolate of ET 5) contains a recombinant segment (red, nucleotides 417 to 768) that is very similar to the corresponding region

Table 3.9 Distribution of *gap* alleles among 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Isolate	ET [a]	Capsular serotype	Host species	*gap* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*gap1.1*	AY839682
PH30	1	A1	Bovine	*gap1.2*	AY839683
PH376	1	A6	Bovine	*gap1.1*	-
PH346	1	A12	Ovine	*gap1.1*	-
PH540	2	A1	Bovine	*gap1.1*	-
PH338	3	A9	Ovine	*gap1.1*	-
PH388	4	A7	Ovine	*gap1.3*	AY839684
PH50	5	A5	Ovine	*gap1.1*	-
PH56	5	A8	Ovine	*gap1.1*	-
PH238	5	A9	Ovine	*gap1.1*	-
PH8	6	A1	Ovine	*gap1.1*	-
PH398	7	A1	Ovine	*gap1.1*	-
PH284	8	A6	Ovine	*gap1.1*	-
PH232	9	A6	Ovine	*gap1.1*	-
PH66	10	A14	Ovine	*gap1.4*	AY839685
PH706	11	A16	Ovine	*gap1.1*	-
PH296	12	A7	Ovine	*gap2.1*	AY839686
PH396	13	A7	Ovine	*gap2.1*	-
PH484	14	A7	Ovine	*gap2.1*	-
PH588	15	A13	Ovine	*gap2.5*	AY839687
PH494	16	A2	Ovine	*gap2.6*	AY839688
PH550	17	A2	Bovine	*gap2.6*	-
PH196	18	A2	Bovine	*gap2.3*	AY839689
PH786	18	A2	Bovine	*gap2.2*	AY839690
PH526	19	A2	Ovine	*gap3.1*	AY839691
PH598	20	A2	Ovine	*gap3.1*	-
PH202	21	A2	Bovine	*gap3.1*	-
PH470	21	A2	Bovine	*gap3.1*	-
PH278	21	A2	Ovine	*gap3.1*	-
PH372	21	A2	Ovine	*gap3.1*	-
PH292	22	A2	Ovine	*gap3.1*	-
PH392	22	A2	Ovine	*gap3.1*	-
M. glucosida					
PH344	1	A11	Ovine	*gap3.2*	AY839692
PH498	3	A11	Ovine	*gap3.2*	-
PH240	5	A11	Ovine	*gap3.3*	AY839693
PH496	7	UG3	Ovine	*gap5.1*	AY839694
PH574	10	UG3	Ovine	*gap4.1*	AY839695
PH290	16	UG3	Ovine	*gap2.4*	AY839696
P. trehalosi					
PH246	2	T4	Ovine	*gap6.1*	AY839697
PH252	4	T10	Ovine	*gap6.2*	AY839698
PH254	15	T15	Ovine	*gap6.4*	AY839699
PH68	19	T3	Ovine	*gap6.3*	AY839700

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

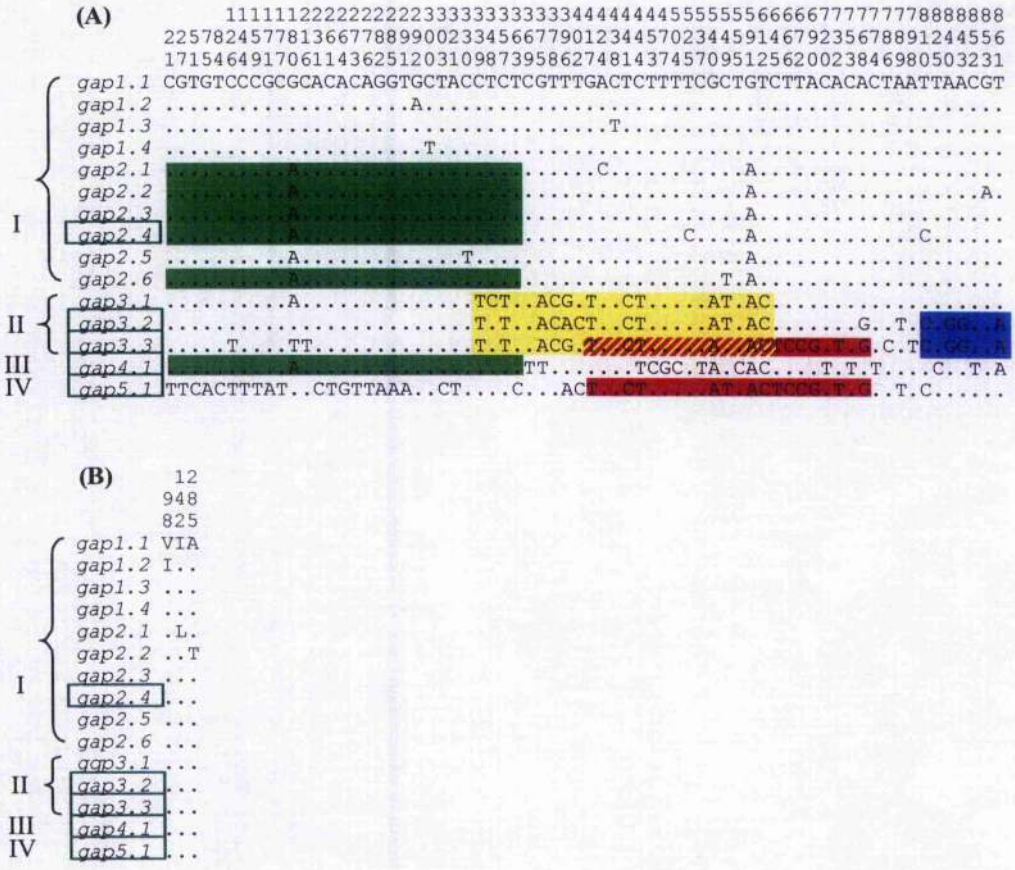


Figure 3.9 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *gap* alleles of *M. haemolytica* and *M. glucosida*. Allele designations are shown to the left of each sequence. The Roman numerals I to IV correspond to the major lineages of Figure 3.10. The *M. glucosida* alleles are highlighted within the green boxes. The different colors indicate nucleotide sequence identity and putative recombinant segments in alleles that are otherwise different. The vertical numbers above the sequences represent the positions of polymorphic nucleotide or amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *gap1.1*)

Table 3.10 Nucleotide and amino acid differences between representative pairs of the six *gap* allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Allele	Pairwise differences in nucleotide and amino acid sequences (%) [a]					
	gap1.1	***gap2.1***	***gap3.1***	***gap4.1***	***gap5.1***	***gap6.1***
gap1.1		1 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	25 (6.4)
gap2.1	3 (0.3)		1 (0.3)	1 (0.3)	1 (0.3)	26 (6.7)
gap3.1	14 (1.6)	13 (1.5)		0 (0.0)	0 (0.0)	25 (6.4)
gap4.1	18 (2.1)	17 (2.0)	21 (2.4)		0 (0.0)	25 (6.4)
gap5.1	38 (4.4)	39 (4.5)	37 (4.3)	48 (5.5)		25 (6.4)
gap6.1	137 (15.7)	137 (15.7)	136 (15.6)	135 (15.5)	140 (16.1)	

[a] Values in the lower left represent pairwise differences in nucleotide sequences (number of polymorphic nucleotide sites and percentage of nucleotide variation); values in the upper right represent pairwise differences in inferred amino acid sequences (number of polymorphic amino acid sites and percentage of amino acid variation).

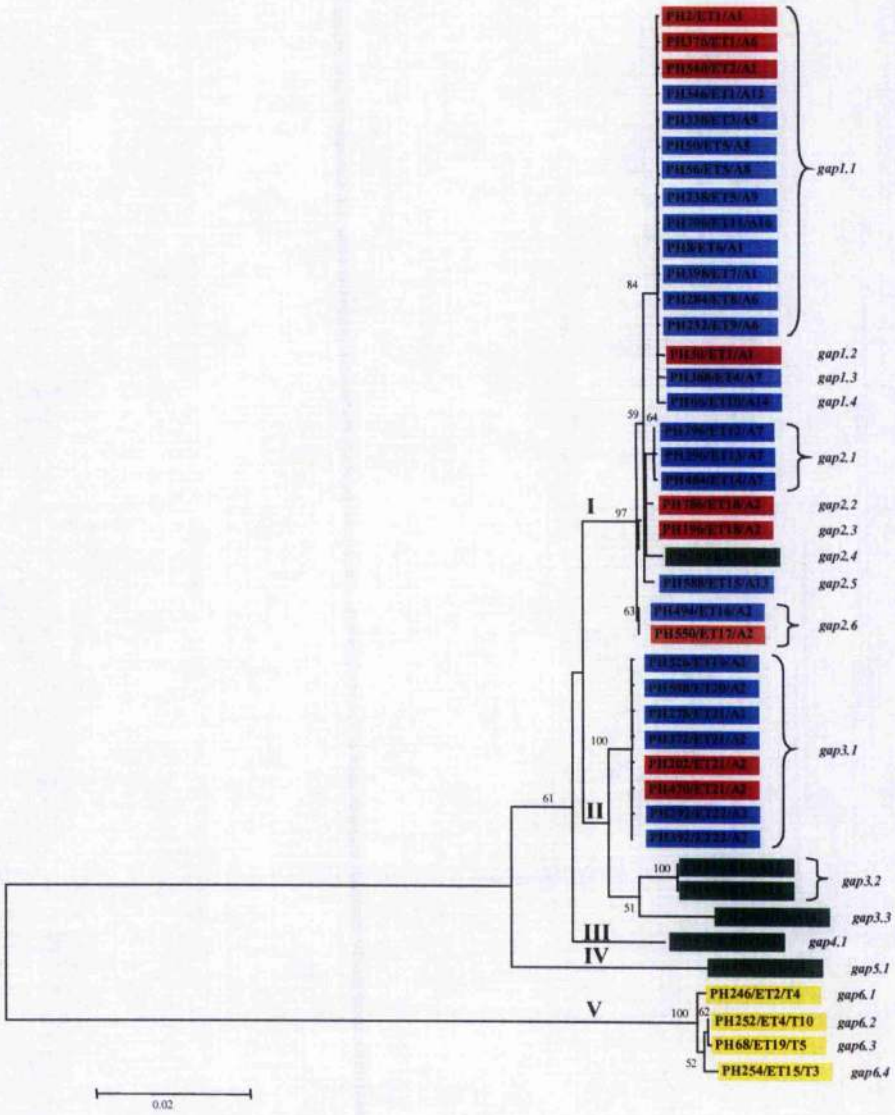


Figure 3.10 Minimum evolution (ME) tree for the *gap* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

of the *gap5.1* allele (UG3 isolate of ET 7) and a recombinant segment (blue, nucleotides 810 to 861) that is identical to the corresponding region of the *gap3.2* allele. The *gap4.1* allele (UG3 isolate of ET 10) has a recombinant segment (green, nucleotides 1 to 363) that is identical to the corresponding region of the *M. glucosida gap2.4* allele and the *M. haemolytica gap2*-type alleles.

The nucleotide substitutions within the recombinant segments of *M. haemolytica* and *M. glucosida* isolates are all synonymous changes and the recombination events do not lead to any amino acid changes (Figure 3.9B). This suggests that there is strong selective constraint against amino acid replacement acting on the *gap* gene of *M. haemolytica* and *M. glucosida*.

Pairwise differences in nucleotide and amino acid sequences between representative pairs of the six allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* ranged from 3 to 140 (0.3 to 16.1 %) nucleotide sites and from 0 to 25 (0 to 6.4 %)amino acid positions (Table 3.10).

The *gap* tree topology (Figure 3.10) indicates that lineages I and II each includes isolates of both *M. haemolytica* and *M. glucosida*, whereas lineages III to V are associated exclusively with *M. glucosida* (III and IV) or *P. trehalosi* (V). For example, the *M. haemolytica gap1*- and *gap2*-type alleles and the *M. glucosida gap2.4* allele cluster together in lineage I, and the *M. haemolytica gap3.1* allele and the *M. glucosida gap3.2* and *gap3.3* alleles cluster together in lineage II. This topology is clearly different from that based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE data (Davies *et al.*, 1997a) and provides strong evidence that horizontal DNA transfer and recombination have disrupted the phylogeny of the *gap* gene in *M. haemolytica* and *M. glucosida*.

3.1.1.2.5 Phosphogluconate dehydrogenase (*gnd*)

The *gnd* gene, encoding 6-phosphogluconate dehydrogenase, was selected because it exhibits high levels of variation in *E. coli* and *S. enterica* in spite of its housekeeping nature (see section 1.6.2.5). However, Gnd was shown to be monomorphic in the MLEE study of *M. haemolytica* (Davies *et al.*, 1997a). The partial sequence of *gnd* (1281 nucleotides [88 %]) was determined in 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate (Table 3.11). Five different *gnd* sequences, each representing a distinct allele, were identified, and the alleles were assigned to three subclasses, *gnd1* to *gnd3*, based on their overall sequence similarities (Table 3.11).

The polymorphic nucleotide and inferred amino acid sites within the five *gnd* alleles, with respect to *gnd1.1*, are shown in Figure 3.11. Nucleotide sequence analysis indicated that the *M. haemolytica gnd1*-type, *M. glucosida gnd2.1*, and *P. trehalosi gnd3.1* alleles have 2, 30, and 218 polymorphic nucleotide sites, respectively (Figure 3.11A). Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.11B). Three amino acid variants (A to C) were identified among the five *gnd* alleles. There were no amino acid variations among the *M. haemolytica gnd1*-type alleles, which corresponds to the MLEE data at the Gnd locus (Table 3.11). In contrast, 1 and 41 variable amino acid sites were detected for *M. glucosida gnd2.1* and *P. trehalosi gnd3.1* respectively. There was no evidence of intragenic recombination within any of the *gnd* alleles.

Pairwise differences in nucleotide sequences indicated that the *M. haemolytica gnd1*-type alleles are highly conserved (one to two polymorphic sites), whereas the *M. glucosida gnd2.1* allele shows moderate divergence from *M. haemolytica gnd1*-type

Table 3.11 Distribution of *gnd* alleles among 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	Gnd [b] allele	*gnd* allele (amino acid type [c])	GenBank accession no.
M. haemolytica						
PH2	1	A1	Bovine	3	*gnd1.1* (A)	AY839648
PH66	10	A14	Ovine	3	*gnd1.1* (A)	-
PH706	11	A16	Ovine	3	*gnd1.1* (A)	-
PH296	12	A7	Ovine	3	*gnd1.1* (A)	-
PH588	15	A13	Ovine	3	*gnd1.1* (A)	-
PH494	16	A2	Ovine	3	*gnd1.2* (A)	AY839649
PH196	18	A2	Bovine	3	*gnd1.1* (A)	-
PH202	21	A2	Bovine	3	*gnd1.3* (A)	AY839650
PH278	21	A2	Ovine	3	*gnd1.3* (A)	-
PH292	22	A2	Ovine	3	*gnd1.3* (A)	-
M. glucosida						
PH344	1	A11	Ovine	3	*gnd2.1* (B)	AY839651
P. trehalosi						
PH246	2	T4	Ovine	3	*gnd3.1* (C)	AY839652

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

[b] Gnd alleles as defined by MLEE (Davies *et al.*, 1997a; 1997b).

[c] Based on the inferred amino acid sequence (Figure 3.11B).

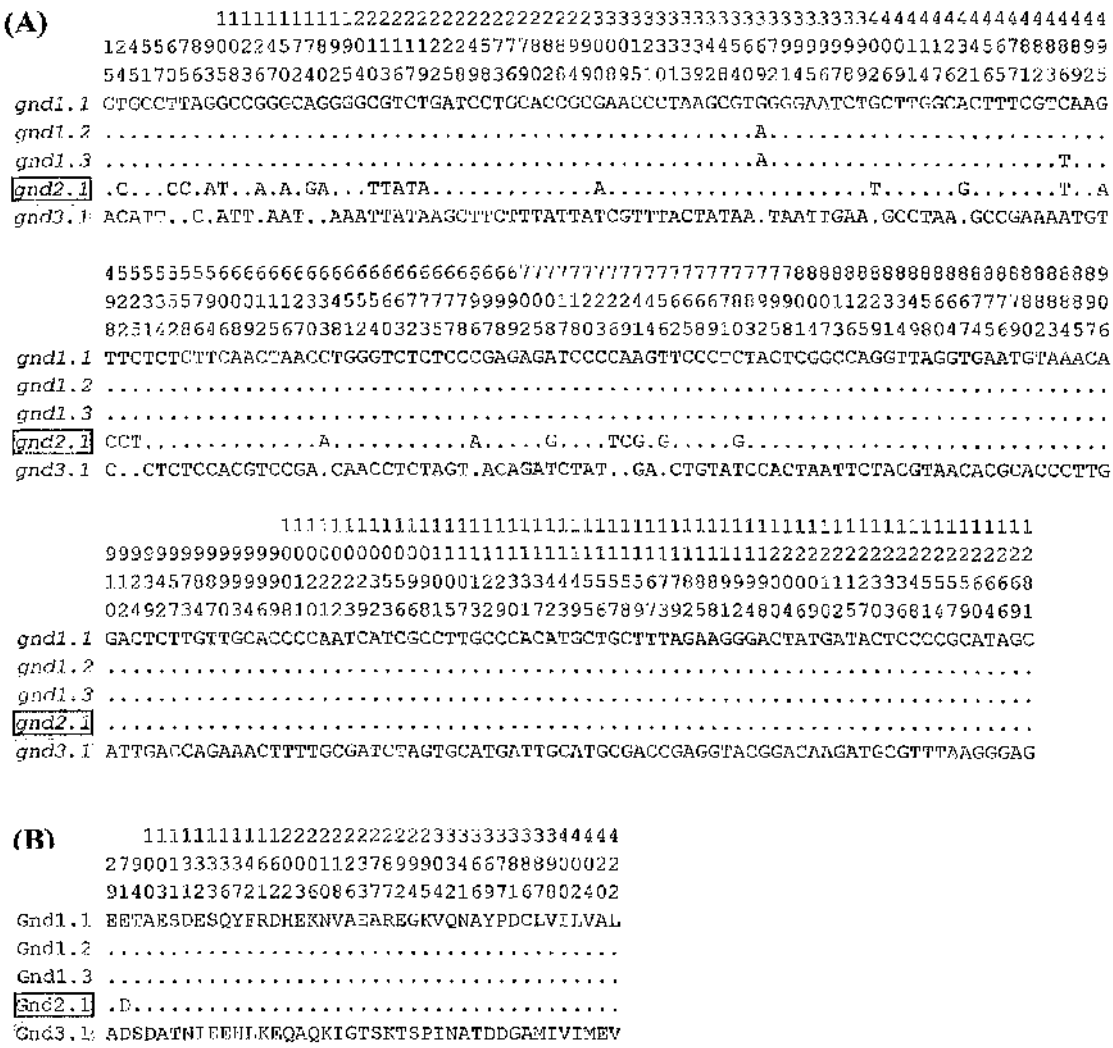


Figure 3.11 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *gnd* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *gnd1.1*).

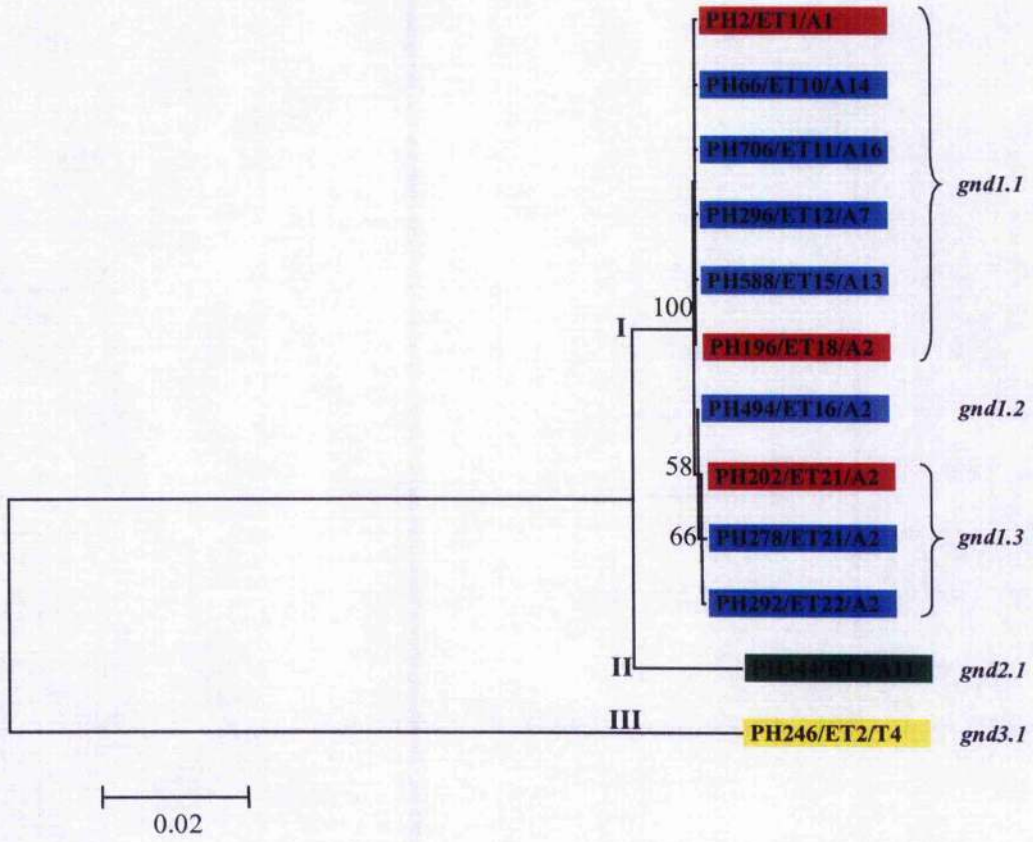


Figure 3.12 Minimum Evolution (ME) tree for the *gnd* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

alleles (30 to 31 polymorphic sites), and the *P. trehalosi gnd3.1* allele exhibits substantial divergence from both the *M. haemolytica gnd1*-type alleles (218 to 219 polymorphic sites) and the *M. glucosida gnd2.1* allele (224 polymorphic sites).

The *gnd* tree topology (Figure 3.12) shows three distinct phylogenetic lineages, I to III, corresponding to the three species, *M. haemolytica* (*gnd1*), *M. glucosida* (*gnd2*), and *P. trehalosi* (*gnd3*), respectively. This phylogeny is in good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) and indicates that recombination has not affected the evolution of *gnd*.

3.1.1.2.6 Glucose-6-phosphate-1-dehydrogenase (*g6pd*)

The *g6pd* gene, encoding glucose-6-phosphate 1-dehydrogenase, was selected because G6pd is an important metabolic enzyme (see section 1.6.2.6) and has been shown to be variable in the previous MLEE study of *M. haemolytica* (Davies *et al.*, 1997a). The partial sequence of *g6pd* (1353 nucleotides [91 %]) was determined only in 10 *M. haemolytica* isolates because all attempts to amplify *g6pd* in *M. glucosida* and *P. trehalosi* were unsuccessful (Table 3.12). Four different *g6pd* nucleotide sequences, each representing a distinct allele, were identified, and the alleles were assigned to the same subclass, *g6pd1*, because their sequences are very similar (Table 3.12).

The polymorphic nucleotide and inferred amino acid sites within the four *g6pd* alleles, with respect to *g6pd1.1*, are shown in Figure 3.13. Nucleotide sequence analysis indicates that the *M. haemolytica g6pd1*-type alleles have five polymorphic nucleotide sites (Figure 3.13A). Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.13B). Three amino acid variants were

Table 3.12 Distribution of *g6pd* alleles among 10 *M. haemolytica* isolates

Isolate	ET [a]	Capsular serotype	Host species	G6pd [b] allele	*g6pd* allele (amino acid type [c])	GenBank accession no.
M. haemolytica						
PH2	1	A1	Bovine	3	*g6pd1.1* (A)	AY839653
PH66	10	A14	Ovine	3	*g6pd1.1* (A)	-
PH706	11	A16	Ovine	3	*g6pd1.1* (A)	-
PH296	12	A7	Ovine	3	*g6pd1.2* (B)	AY839654
PH588	15	A13	Ovine	3	*g6pd1.2* (B)	-
PH494	16	A2	Ovine	2.5	*g6pd1.4* (C)	AY839655
PH196	18	A2	Bovine	3	*g6pd1.2* (B)	-
PH202	21	A2	Bovine	2.5	*g6pd1.3* (C)	AY839656
PH278	21	A2	Ovine	2.5	*g6pd1.3* (C)	-
PH292	22	A2	Ovine	2.5	*g6pd1.3* (C)	-

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a).

[b] G6pd alleles as defined by MLEE (Davies *et al.*, 1997a).

[c] Based on the inferred amino acid sequence (Figure 3.13B).

(A)

```
            111
          79222
          61089
          93975
g6pd1.1   TAACC
g6pd1.2   CG..A
g6pd1.3   CGC.G
g6pd1.4   CGCAG
```

(B)

```
          34
          03
          52
G6pd1.1   KA
G6pd1.2   ED
G6pd1.3   EG
G6pd1.4   EG
```

Figure 3.13 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *g6pd* alleles of *M. haemolytica*. Allele designations are shown to the left of each sequence. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5'end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *g6pdA1.1*).

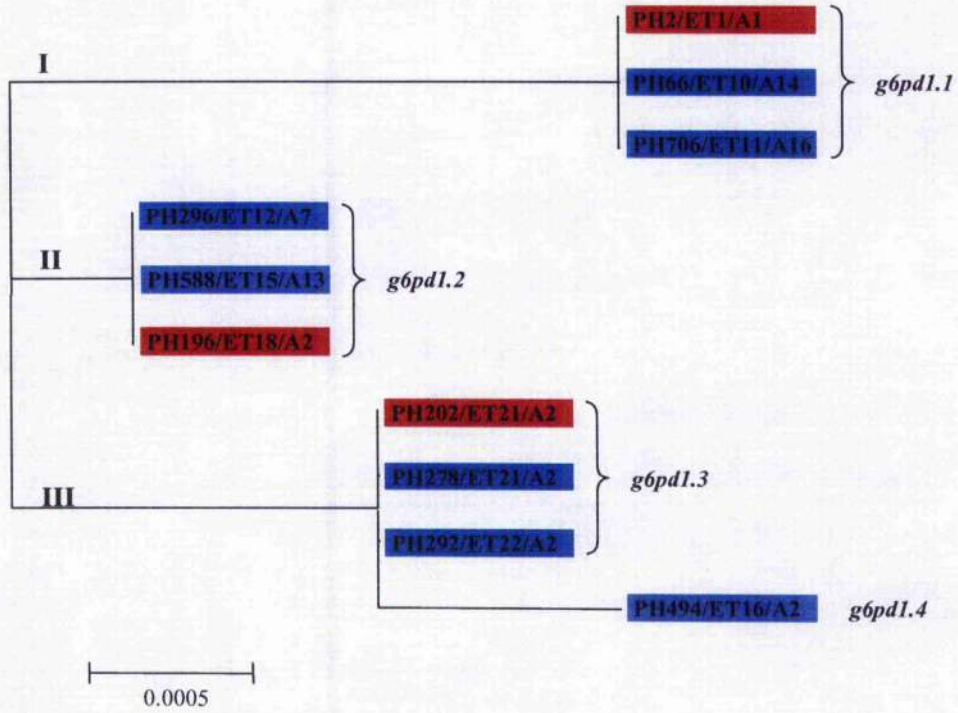


Figure 3.14 Minimum evolution (ME) tree for the *g6pd* gene of *M. haemolytica*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bovine *M. haemolytica* isolates are highlighted in red, and ovine *M. haemolytica* isolates in blue. Electrophoretic types (ETs) have been described for *M. haemolytica* (Davies *et al.*, 1997a)

identified among the four *g6pd* alleles and two variable amino acid positions were detected. There was no evidence of intragenic recombination within any of the *g6pd* alleles.

The relationship between G6pd (MLEE) allele types and inferred amino acid types of the *g6pd* alleles of *M. haemolytica* is shown in Table 3.12. As expected, there was a correlation between the G6pd (MLEE) allele types and the inferred amino acid types of the *g6pd* alleles of *M. haemolytica*. However, nucleotide sequence analysis (four alleles) and amino acid sequence analysis (three protein types, A to C) were more discriminating than MLEE analysis (two alleles).

The *g6pd* tree topology is shown in Figure 3.14. Isolates of ETs 1, 10, and 11 cluster together in lineage I, isolates of ETs 12, 15, 18 cluster together in lineage II, and isolates of ETs 16, 21, and 22 cluster together in lineage III. With the exception of isolate PH494 of ET 16, lineages I, II, and III of the *g6pd* gene tree correspond to lineages A, B, and C, respectively of the MLEE tree (Davies *et al.*, 1997a) and indicates that recombination has not affected the evolution of *g6pd*. In the MLEE tree (see Figure 2.1) isolate PH494 of ET 16 appears to be more closely related to isolates of ETs 12, 15, and 18 (lineage B), whereas in the gene tree (Figure 3.14) isolate PH494 of ET 16 is more closely related to isolates of ETs 21 and 22 (lineage III)

3.1.1.2.7 Malate dehydrogenase (*mdh*)

The *mdh* gene, encoding malate dehydrogenase, was selected because it has been used in other molecular evolutionary studies (see section 1.6.2.7) and Mdh has shown to be polymorphic in the previous MLEE studies of *M. haemolytica*, *M. glucosida*, and *P.*

trehalosi (Davies *et al.*, 1997a; 1997b). The partial sequence of *mdh* (780 nucleotides [80 %]) was determined in 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate (Table 3.13). Eight different *mdh* nucleotide sequences, each representing a distinct allele, were identified, and the alleles were assigned to three subclasses, *mdh1* to *mdh3*, based on their overall sequence similarities (Table 3.13).

The polymorphic nucleotide and inferred amino acid sites within the eight *mdh* alleles, with respect to *mdh1.1*, are shown in Figure 3.15. Nucleotide sequence analysis indicates that the *M. haemolytica mdh1*-type and *mdh2.1* alleles, *M. glucosida mdh2.2* allele, and *P. trehalosi mdh3.1* alleles have seven, 31, 49, and 229 polymorphic nucleotide sites, respectively (Figure 3.15A). However, the *M. haemolytica mdh2.1* allele contains a recombinant segment (green, 507 to 748bp) that is almost identical to the corresponding region of the *M. glucosida mdh2.2* allele, although the rest of these two alleles are different. Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.15B). Five amino acid variants were identified among the eight *mdh* alleles; 2, 2, 4, and 70 variable amino acid sites were detected in the *M. haemolytica mdh1*-type and *mdh2.1* alleles, *M. glucosida mdh2.2* allele, and *P. trehalosi mdh3.1* alleles, respectively.

The relationship between Mdh (MLEE) allele types and inferred amino acid types of the *mdh* alleles of *M. haemolytica* is shown in Table 3.13. As expected, there was a correlation between the Mdh (MLEE) allele types and the inferred amino acid types of the *mdh* alleles of *M. haemolytica*. However, nucleotide sequence analysis (six alleles) and amino acid sequence analysis (three protein types, A to C) were more discriminating than MLEE analysis (two alleles). Surprisingly, isolate PH588 was not differentiated by MLEE.

Table 3.13 Distribution of *mdh* alleles among 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	Mdh [b] allele	*mdh* allele (amino acid type [c])	GenBank accession no.
M. haemolytica						
PH2	1	A1	Bovine	3	*mdh1.1* (A)	AY839657
PH66	10	A14	Ovine	3	*mdh1.4* (A)	AY839658
PH706	11	A16	Ovine	3	*mdh1.5* (A)	AY839659
PH296	12	A7	Ovine	3	*mdh1.4* (A)	-
PH588	15	A13	Ovine	3	*mdh2.1* (C)	AY839660
PH494	16	A2	Ovine	3	*mdh1.2* (A)	AY839661
PH196	18	A2	Bovine	4	*mdh1.3* (B)	AY839662
PH202	21	A2	Bovine	3	*mdh1.2* (A)	-
PH278	21	A2	Ovine	3	*mdh1.2* (A)	-
PH292	22	A2	Ovine	3	*mdh1.2* (A)	-
M. glucosida						
PH344	1	A11	Ovine	1.8	*mdh2.2* (D)	AY839663
P. trehalosi						
PH246	2	T4	Ovine	1	*mdh3.1* (E)	AY839664

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

[b] Mdh alleles as defined by MLEE (Davies *et al.*, 1997a; 1997b).

[c] Based on the inferred amino acid sequence (Figure 3.15B).

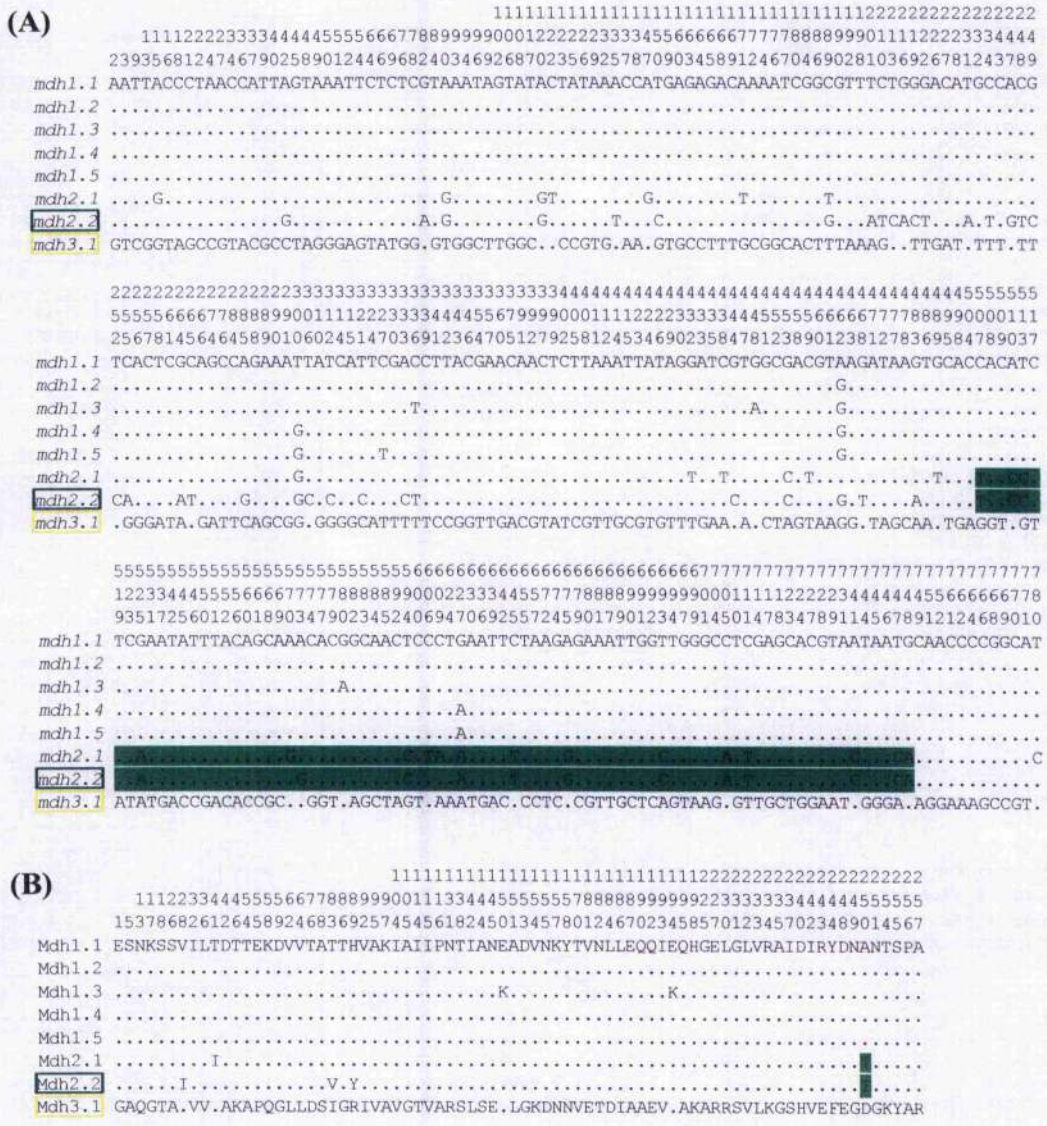


Figure 3.15 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *mdh* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The green shading indicates sequence identity and putative recombinant segments in alleles that are otherwise different. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *mdh1.1*).

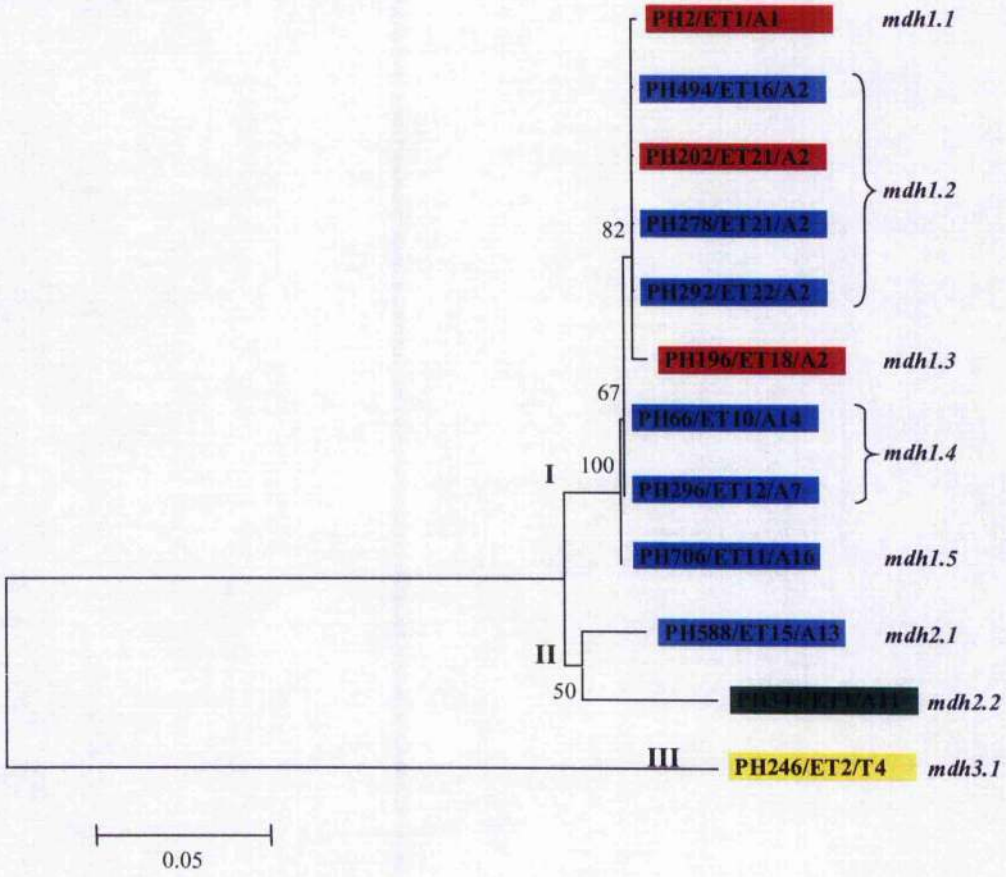


Figure 3.16 Minimum evolution (ME) tree for the *mdh* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

Pairwise differences in nucleotide sequences indicated that the *M. haemolytica mdh1*-type alleles are highly conserved (one to five polymorphic sites), whereas the *M. haemolytica mdh2.1* allele (30 to 35 polymorphic sites) and the *M. glucosida mdh2.2* allele (46 to 49 polymorphic sites) have diverged form the *M. haemolytica mdh1*-type alleles, and the *P. trehalosi mdh3.1* allele exhibits substantial divergence from the *M. haemolytica mdh1*-type (225 to 229 polymorphic sites) and *mdh2.1* (233 polymorphic sites) alleles, and from the *M. glucosida mdh2.2* allele (235 polymorphic sites).

The *mdh* tree topology (Figure 3.16) shows three distinct phylogenetic lineages, I to III. With the exception of *M. haemolytica* isolate PH588 (*mdh2.1*) which clustered with *M. glucosida* (*mdh2.2*) in lineage II, the three lineages corresponded to the three species *M. haemolytica* (*mdh1*), *M. glucosida* (*mdh2*), and *P. trehalosi* (*mdh3*), respectively. This phylogeny is, with the exception of isolate PH588, is good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a).

3.1.1.2.8 Mannitol-1-phosphate dehydrogenase (*mtlD*)

The *mtlD* gene, encoding mannitol-1-phosphate, was selected because MtlD has been shown to be polymorphic in the previous MLEE studies of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b). The partial sequence of *mtlD* (951 nucleotides [83 %]) was initially determined in the 'standard' set of 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate. However, an additional seven *M. haemolytica* and five *M. glucosida* isolates, which showed variation in this gene in the MLEE study (Davies *et al.*, 1997a), were examined (Table 3.14). Twelve different *mtlD* nucleotide sequences, each

Table 3.14 Distribution of *mtlD* alleles among 17 *M. haemolytica* isolates, six *M. glucosida* isolates, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	MtlD [b] allele	*mtlD* allele (amino acid type [c])	GenBank accession no.
M. haemolytica						
PH2	1	A1	Bovine	3	*mtlD1.1* (A)	AY839665
PH540*	2	A1	Bovine	2.5	*mtlD1.2* (B)	AY839666
PH8*	6	A1	Ovine	4	*mtlD1.6* (E)	AY839667
PH398*	7	A1	Ovine	4	*mtlD1.6* (E)	-
PH232*	9	A6	Ovine	4	*mtlD1.6* (E)	-
PH66	10	A14	Ovine	3	*mtlD1.1* (A)	-
PH706	11	A16	Ovine	3	*mtlD1.1* (A)	-
PH296	12	A7	Ovine	2.5	*mtlD1.4* (D)	AY839668
PH396*	13	A7	Ovine	2.5	*mtlD1.4* (D)	-
PH484*	14	A7	Ovine	2.5	*mtlD1.4* (D)	-
PH588	15	A13	Ovine	3	*mtlD1.5* (A)	AY839669
PH494	16	A2	Ovine	3	*mtlD1.1* (A)	-
PH196	18	A2	Bovine	3	*mtlD1.1* (A)	-
PH598*	20	A2	Ovine	4	*mtlD1.3* (C)	AY839670
PH202	21	A2	Bovine	3	*mtlD1.1* (A)	-
PH278	21	A2	Ovine	3	*mtlD1.1* (A)	-
PH292	22	A2	Ovine	3	*mtlD1.1* (A)	-
M. glucosida						
PH344	1	A11	Ovine	1	*mtlD2.1* (F)	AY839671
PH498*	3	A11	Ovine	1	*mtlD2.1* (F)	-
PH240*	5	A11	Ovine	2	*mtlD2.3* (H)	AY839672
PH496*	7	UG3	Ovine	2	*mtlD2.5* (J)	AY839673
PH574*	10	UG3	Ovine	2	*mtlD2.2* (G)	AY839674
PH290*	16	UG3	Ovine	1.5	*mtlD2.4* (I)	AY839675
P. trehalosi						
PH246	2	T4	Ovine	1	*mtlD3.1* (K)	AY839676

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

[b] MtlD alleles as defined by MLEE (Davies *et al.*, 1997a; 1997b).

[c] Based on the inferred amino acid sequence (Figure 3.17B).

* additional isolates

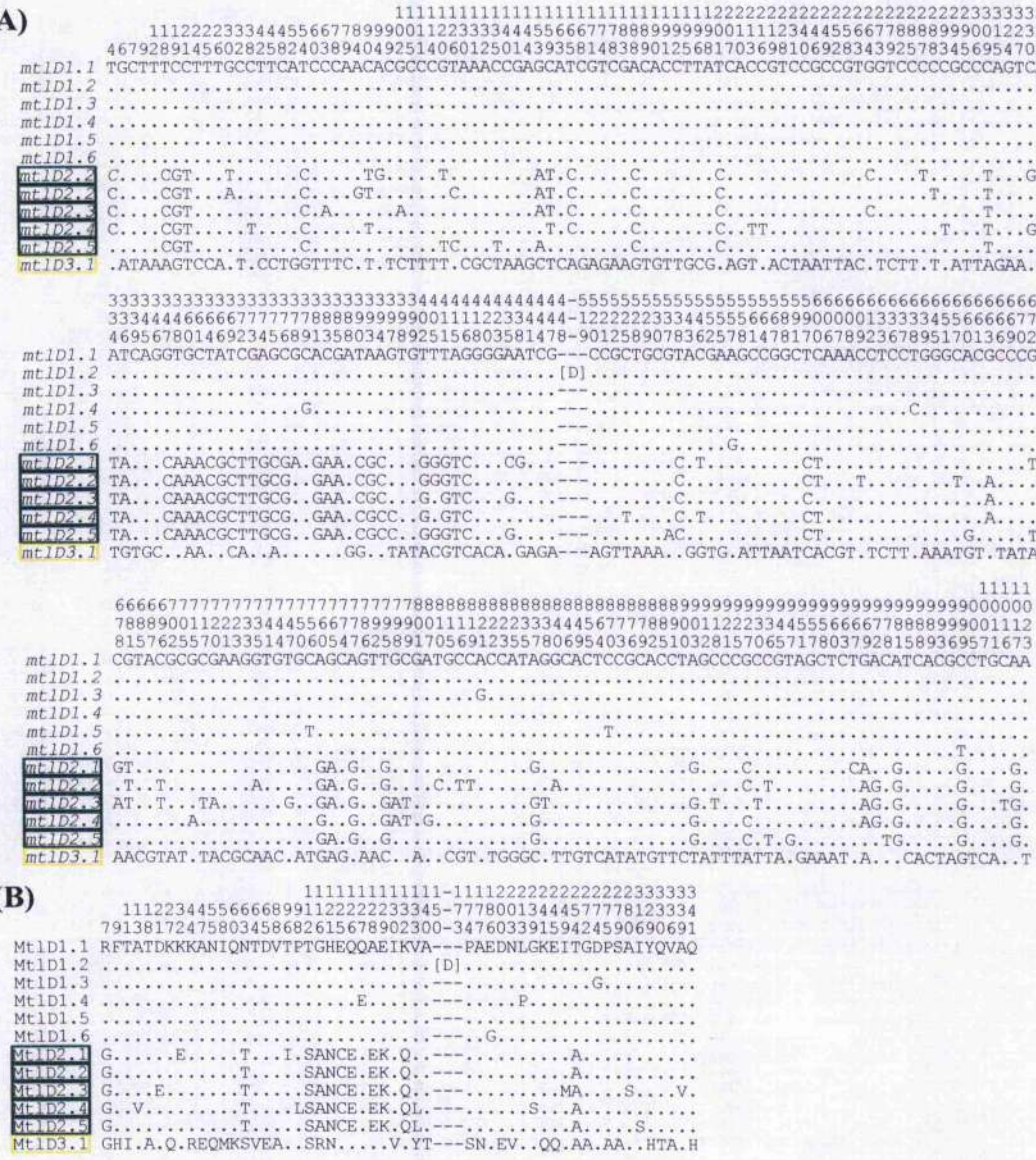


Figure 3.17 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *mtlD* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *mtlDA1.1*). Note: Nucleotides 448 to 519 (segment D) of *mtlD1.2* represents a duplicated DNA segment that is identical to nucleotides 376-447 (CAGCAAGCAGAAATTGAAAAAGTGGTTGGTTTTGTTGATAGTGCGGTGGACAGA ATTGTACCACCTGCCGAG). Similarly, amino acids 150 to 173 (segment D) of MtlD1.2 represents a duplicated amino acid segment that is identical to amino acids 126 to 149 (QQAEIEKVVGFVDSAVDRIVPPAE).

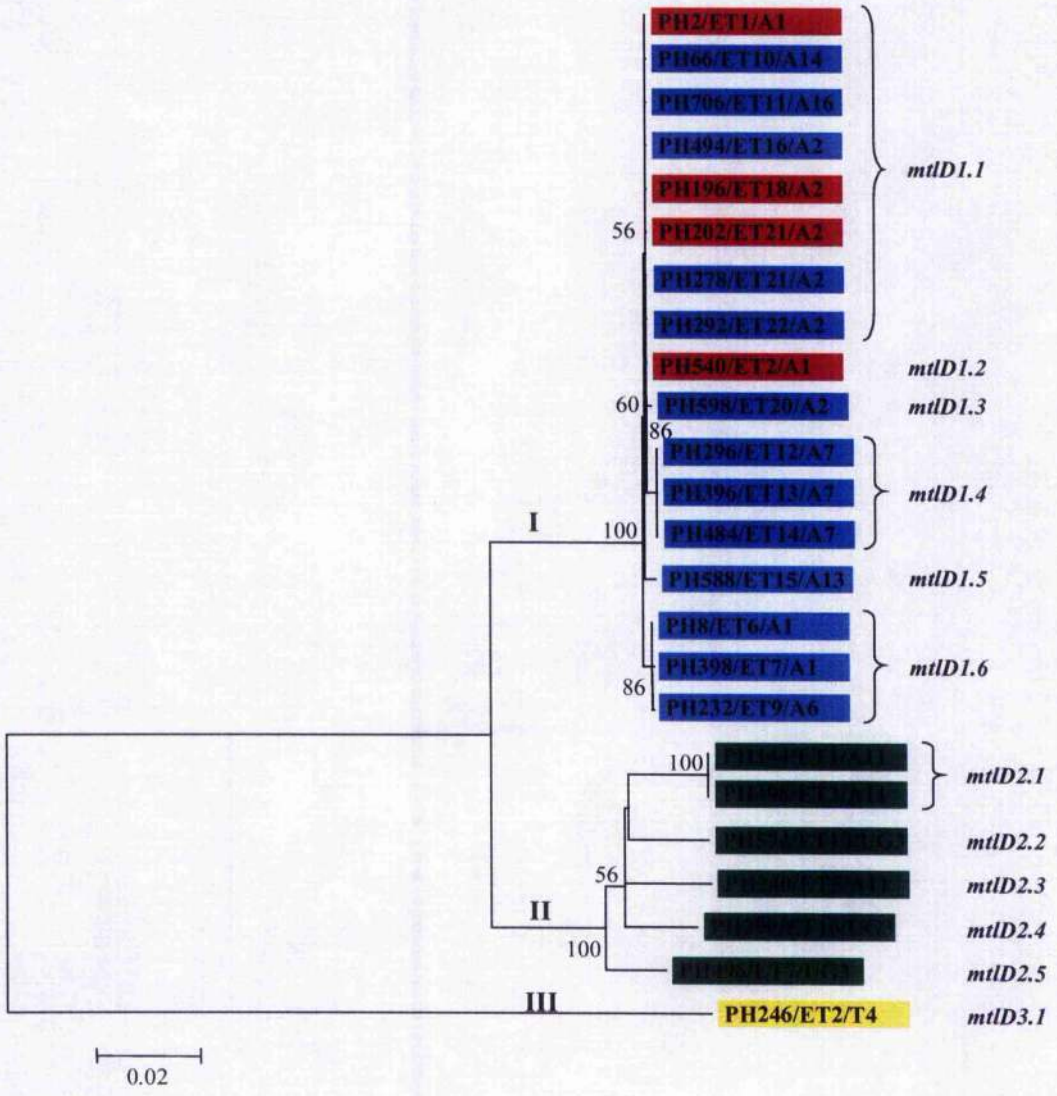


Figure 3.18 Minimum evolution (ME) tree for the *mtlD* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option was used for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

representing a distinct allele, were identified, and the alleles were assigned to three subclasses, *mtlD1* to *mtlD3*, based on their overall sequence similarities (Table 3.14).

The polymorphic nucleotide and inferred amino acid sites within the twelve *mtlD* alleles, with respect to *mtlD1.1* are shown in Figure 3.17. Nucleotide sequence analysis indicates that the *M. haemolytica mtlD1*-type, *M. glucosida mtlD2*-type, and *P. trehalosi mtlD3.1* alleles have 7, 104, and 210 polymorphic nucleotide sites, respectively (Figure 3.17A). Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.17B). Eleven amino acid variants were identified among the twelve *mtlD* alleles; 4, 22, and 34 variable amino acid sites were detected in the *M. haemolytica mtlD1*-type, *M. glucosida mtlD2*-type, and *P. trehalosi mtlD3.1* alleles, respectively. There was no evidence of intragenic recombination within any of the *mtlD* alleles.

The relationship between MtlD (MLEE) allele types and inferred amino acid types (A to K) of the *mtlD* alleles of *M. haemolytica* is shown in Table 3.14. As expected, there was a correlation between MtlD (MLEE) allele types and the inferred amino acid types of the *mtlD* alleles of *M. haemolytica* and *M. glucosida*. However, nucleotide sequence analysis (6 alleles for *M. haemolytica* and 5 alleles for *M. glucosida*) and amino acid sequence analysis (5 protein types for both *M. haemolytica* and *M. glucosida*) were more discriminating than MLEE analysis (3 alleles for both *M. haemolytica* and *M. glucosida*).

Excluding the duplicated segment of *mtlD1.2*, pairwise differences in nucleotide sequences indicate that *M. haemolytica mtlD1*-type alleles are highly conserved (1 to 4 polymorphic sites), whereas the *M. glucosida mtlD2*-type alleles, which are themselves relatively divergent (25 to 34 polymorphic sites), shows moderate

divergence from the *M. haemolytica mtlD1*-type alleles (58 to 68 polymorphic sites), and the *P. trehalosi mtlD3.1* allele exhibits substantial divergence from both the *M. haemolytica mtlD1*-type alleles (209 to 212 polymorphic sites) and the *M. glucosida mtlD2*-type alleles (213 to 221 polymorphic sites).

The *mtlD* tree topology (Figure 3.18) shows three distinct phylogenetic lineages, I to III, corresponding to the three species, *M. haemolytica* (*mtlD1*), *M. glucosida* (*mtlD2*), and *P. trehalosi* (*mtlD3*), respectively. This phylogeny is in good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) and indicates that recombination has not affected the evolution of *mtlD*.

3.1.1.2.9 Phosphomannomutase (*pmm*)

The *pmm* gene, encoding phosphomannomutase, was selected because it is a housekeeping gene but is also involved in LPS biosynthesis (see section in 1.6.2.9). The partial sequence of *pmm* (729 nucleotides [50 %]) was determined in 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates (Table 3.15). Nineteen different *pmm* sequences, each representing a distinct allele, were identified, and the alleles were assigned to six subclasses, *pmm1* to *pmm6*, based on their overall sequence similarities (Table 3.15).

The polymorphic nucleotide and inferred amino acid sites within *pmm1*- to *pmm5*-type alleles of *M. haemolytica* and *M. glucosida* are shown in Figure 3.19. *P. trehalosi pmm6* type alleles were excluded because they are highly divergent from both *M. haemolytica* and *M. glucosida* (Table 3.16) and there was no visible evidence of recombination. Visual inspection of the nucleotide sequences of the *pmm1*- to *pmm5*-type alleles (Figure 3.19A) indicates that *M. haemolytica* isolates contain three

Table 3.15 Distribution of *pmm* alleles among 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Isolate	ET [a]	Capsular serotype	Host species	*pmm* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*pmm1.1*	AY847831
PH30	1	A1	Bovine	*pmm1.1*	-
PH376	1	A6	Bovine	*pmm1.1*	-
PH346	1	A12	Ovine	*pmm1.3*	AY847832
PH540	2	A1	Bovine	*pmm1.1*	-
PH338	3	A9	Ovine	*pmm1.3*	-
PH388	4	A7	Ovine	*pmm1.1*	-
PH50	5	A5	Ovine	*pmm4.1*	AY847833
PH56	5	A8	Ovine	*pmm4.3*	AY847834
PH238	5	A9	Ovine	*pmm4.1*	-
PH8	6	A1	Ovine	*pmm4.1*	-
PH398	7	A1	Ovine	*pmm4.1*	-
PH284	8	A6	Ovine	*pmm4.3*	-
PH232	9	A6	Ovine	*pmm4.3*	-
PH166	10	A14	Ovine	*pmm4.1*	-
PH706	11	A16	Ovine	*pmm4.2*	AY847835
PH296	12	A7	Ovine	*pmm1.1*	-
PH396	13	A7	Ovine	*pmm1.1*	-
PH484	14	A7	Ovine	*pmm1.1*	-
PH588	15	A13	Ovine	*pmm5.1*	AY847836
PH494	16	A2	Ovine	*pmm4.4*	AY847837
PH550	17	A2	Bovine	*pmm4.5*	AY847838
PH196	18	A2	Bovine	*pmm1.4*	AY847839
PH786	18	A2	Bovine	*pmm1.4*	-
PH526	19	A2	Ovine	*pmm1.2*	AY847840
PH598	20	A2	Ovine	*pmm1.2*	-
PH202	21	A2	Bovine	*pmm1.2*	-
PH470	21	A2	Bovine	*pmm1.2*	-
PH278	21	A2	Ovine	*pmm1.2*	-
PH372	21	A2	Ovine	*pmm1.2*	-
PH292	22	A2	Ovine	*pmm1.2*	-
PH392	22	A2	Ovine	*pmm1.2*	-
M. glucosida					
PH344	1	A11	Ovine	*pmm3.2*	AY847841
PH498	3	A11	Ovine	*pmm3.2*	-
PH240	5	A11	Ovine	*pmm3.3*	AY847842
PH496	7	UG3	Ovine	*pmm3.1*	AY847843
PH574	10	UG3	Ovine	*pmm2.1*	AY847844
PH290	16	UG3	Ovine	*pmm3.4*	AY847845
P. trehalosi					
PH246	2	T4	Ovine	*pmm6.1*	AY847846
PH252	4	T10	Ovine	*pmm6.2*	AY847847
PH254	15	T15	Ovine	*pmm6.3*	AY847848
PH68	19	T3	Ovine	*pmm6.4*	AY847849

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

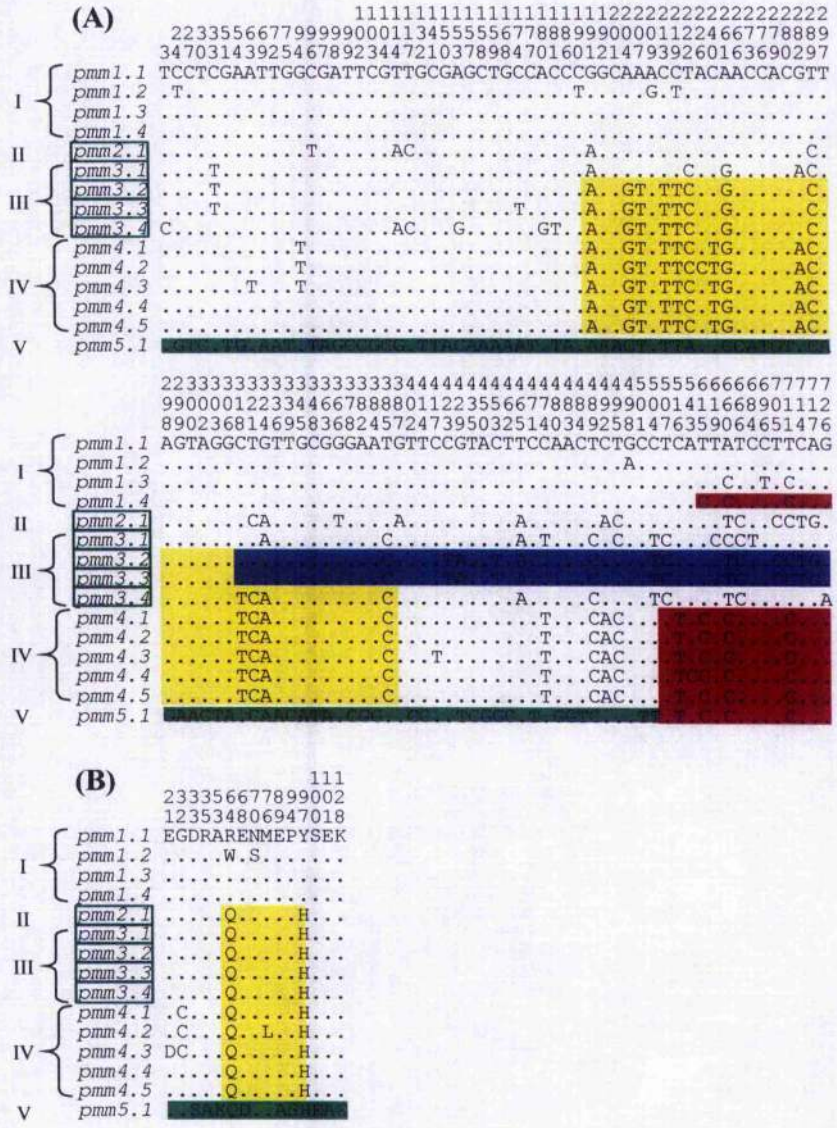


Figure 3.19 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *pmm* alleles of *M. haemolytica* and *M. glucosida*. Allele designations are shown to the left of each sequence. The Roman numerals correspond to the major lineages of Figure 3.20. The *M. glucosida* alleles are highlighted within the green boxes. The coloured shading indicates nucleotide sequence identity and putative recombinant segments in alleles that are otherwise different. The vertical numbers above the sequences represent the positions of polymorphic nucleotide sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *pmm1.1*)

Table 3.16 Nucleotide and amino acid differences between representative pairs of the six *pmm* allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Allele	Pairwise differences in nucleotide and amino acid sequences (%) [a]					
	pmm1.1	***pmm2.1***	***pmm3.1***	***pmm4.1***	***pmm5.1***	***pmm6.1***
pmm1.1		2 (0.8)	2 (0.8)	3 (1.2)	11 (4.5)	25 (10.3)
pmm2.1	18 (7.4)		0 (0.0)	1 (0.4)	9 (3.7)	24 (9.9)
pmm3.1	18 (7.4)	23 (9.5)		1 (0.4)	9 (3.7)	24 (9.9)
pmm4.1	23 (9.5)	26 (10.7)	19 (7.8)		10 (4.1)	25 (10.3)
pmm5.1	78 (32.1)	82 (33.7)	82 (33.7)	74 (30.5)		28 (11.5)
pmm6.1	149 (61.3)	145 (59.7)	150 (61.7)	144 (59.3)	149 (61.3)	

[a] Values in the lower left represent pairwise differences in nucleotide sequences (number of polymorphic nucleotide sites and percentage of nucleotide variation); values in the upper right represent pairwise differences in inferred amino acid sequences (number of polymorphic amino acid sites and percentage of amino acid variation).

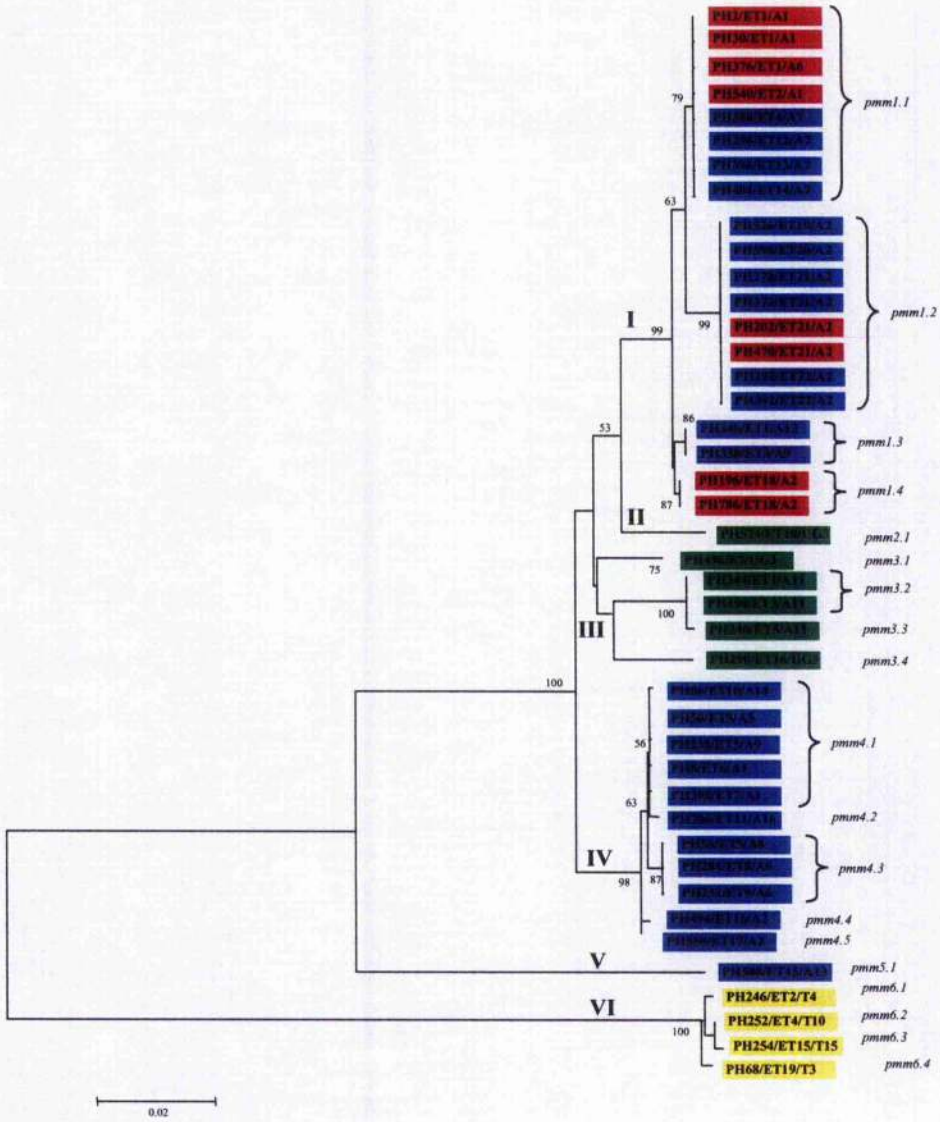


Figure 3.20 Minimum evolution (ME) tree for the *pmm* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option was used for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

major allele types, *pmm1*, *pmm4*, and *pmm5*. The *M. haemolytica pmm1*-type alleles are relatively conserved, whereas the *pmm4*-type alleles and *pmm5.1* allele have diverged from the *M. haemolytica pmm1*-type alleles, due mainly to intragenic recombination. In particular, allele 5.1 isolate PH588 has diverged substantially from *M. haemolytica* alleles. The *pmm4*-type alleles (ovine A1, A5, A6, A8, A9, A14, A16 isolates of ETs 5 to 11 and bovine A2 isolates of ETs 16 and 17) contain a recombinant segment (yellow, nucleotides 191 to 385) that is very similar to the corresponding region of the *M. glucosida pmm3.4* allele. The *pmm5.1* allele (ovine A13 isolate of ET 15) contains a highly divergent segment (green; nucleotides 24 to 504) that has possibly been acquired from an external source by recombination. Interestingly, the *pmm4*-type alleles and *pmm5.1* allele contain segments (red, nucleotides 507 to 726) that are identical, or nearly identical, in all alleles. In addition, allele *pmm1.4* contains a segment (red, nucleotides 615 to 726) that is identical to the corresponding region of *pmm4*-type and *pmm5.1* alleles. *M. glucosida* isolates contain two major allele types, *pmm2* and *pmm3*. *M. glucosida* alleles *pmm3.2* to *pmm3.4* have identical segments (yellow, nucleotides 191 to 306) but alleles *pmm3.2* and *pmm3.3* differ from allele *pmm3.4* between nucleotides 318 to 726 (blue segment).

The nucleotide substitutions within the recombinant segments of the *M. haemolytica* and *M. glucosida* isolates are mostly synonymous changes (Figure 3.19B) and result in three and eleven amino acid changes within the yellow and green recombinant segments, respectively. The conserved amino acid sequence within the yellow recombinant region of the *M. haemolytica pmm4*-type alleles and the *M. glucosida pmm3*-type alleles also suggest that the segment originated from the same source (possibly *M. glucosida*).

Pairwise differences in nucleotide and amino acid sequences between representative pairs of the six allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Table 3.16) ranged from 18 to 150 (7.4 to 61.7 %) nucleotide sites and from 0 to 28 (0 to 11.5 %) amino acid positions.

The *pmm* tree topology (Figure 3.20) indicates that *M. glucosida pmm2*- (lineage II) and *pmm3*- (lineage III) type alleles are more closely related to *M. haemolytica pmm1*-type (lineage I) alleles than are the *M. haemolytica pmm4*- (lineage IV) and *pmm5*- (lineage V) type alleles. *P. trehalosi pmm6*-type alleles represent a more distantly related lineage, VI. This gene tree topology is clearly different from that based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) and provides strong evidence that horizontal DNA transfer and recombination have disrupted the phylogeny of the *pmm* gene in *M. haemolytica* and *M. glucosida*. .

3.1.1.3 Genes encoding secreted proteins

3.1.1.3.1 Glycoprotease (*gcp*)

The *gcp* gene, encoding glycoprotease, was selected because glycoprotease is one of the secreted putative virulence factors of *M. haemolytica* (see section 1.3.1.2). The partial sequence of *gcp* (870 nucleotides [89 %]) was determined in 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate (Table 3.17). Five different *gcp* sequences, each representing a distinct allele, were identified, and the alleles were assigned to three subclasses, *gcp1* to *gcp3*, based on their overall sequence similarities (Table 3.17).

Table 3.17 Distribution of *gcp* alleles among 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	*gcp* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*gcp1.1*	AY839677
PH166	10	A14	Ovine	*gcp1.1*	-
PH1706	11	A16	Ovine	*gcp1.1*	-
PH296	12	A7	Ovine	*gcp1.1*	-
PH588	15	A13	Ovine	*gcp1.1*	-
PH494	16	A2	Ovine	*gcp1.3*	AY839678
PH196	18	A2	Bovine	*gcp1.2*	AY839679
PH202	21	A2	Bovine	*gcp1.3*	-
PH278	21	A2	Ovine	*gcp1.3*	-
PH292	22	A2	Ovine	*gcp1.1*	-
M. glucosida					
PH344	1	A11	Ovine	*gcp2.1*	AY839680
P. trehalosi					
PH246	2	T4	Ovine	*gcp3.1*	AY839681

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

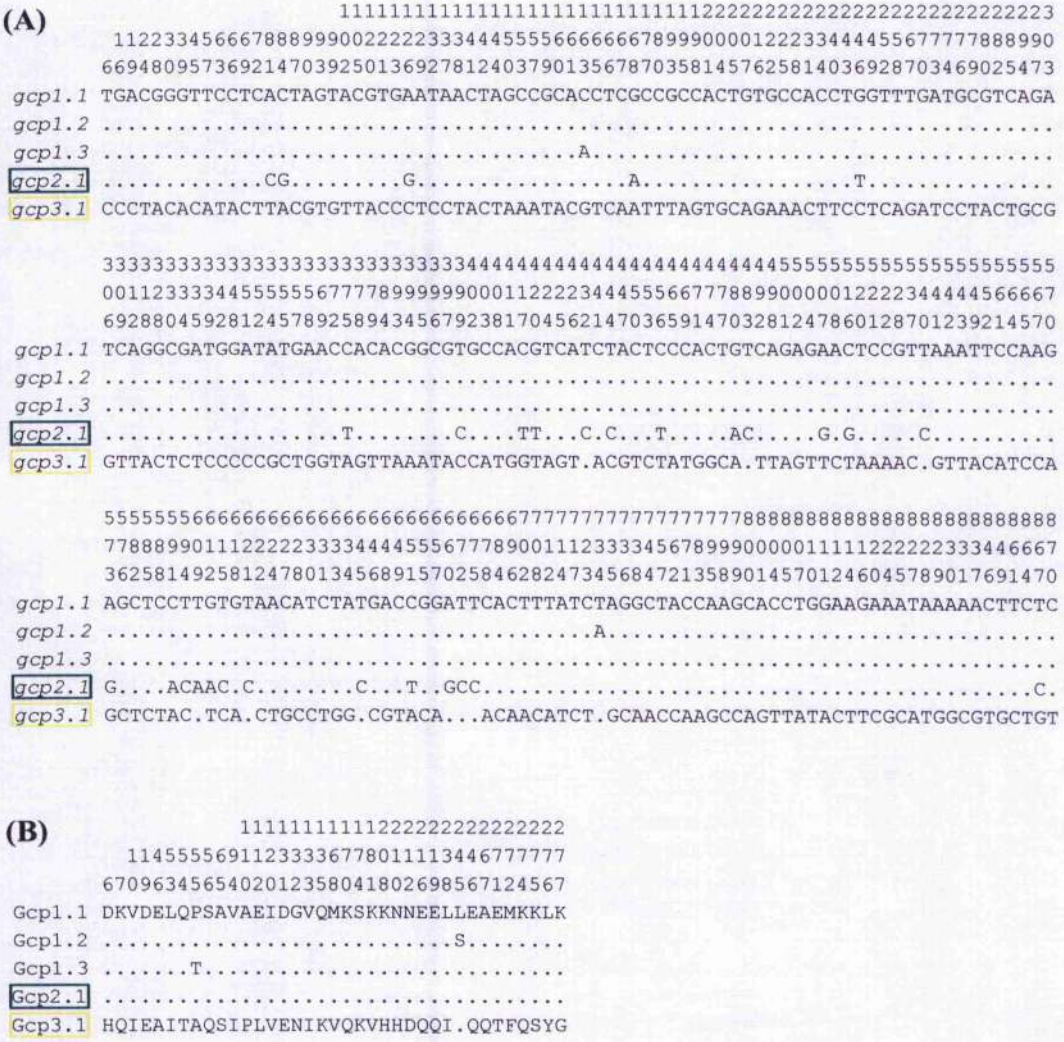


Figure 3.21 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *gcp* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *gcp1.1*).

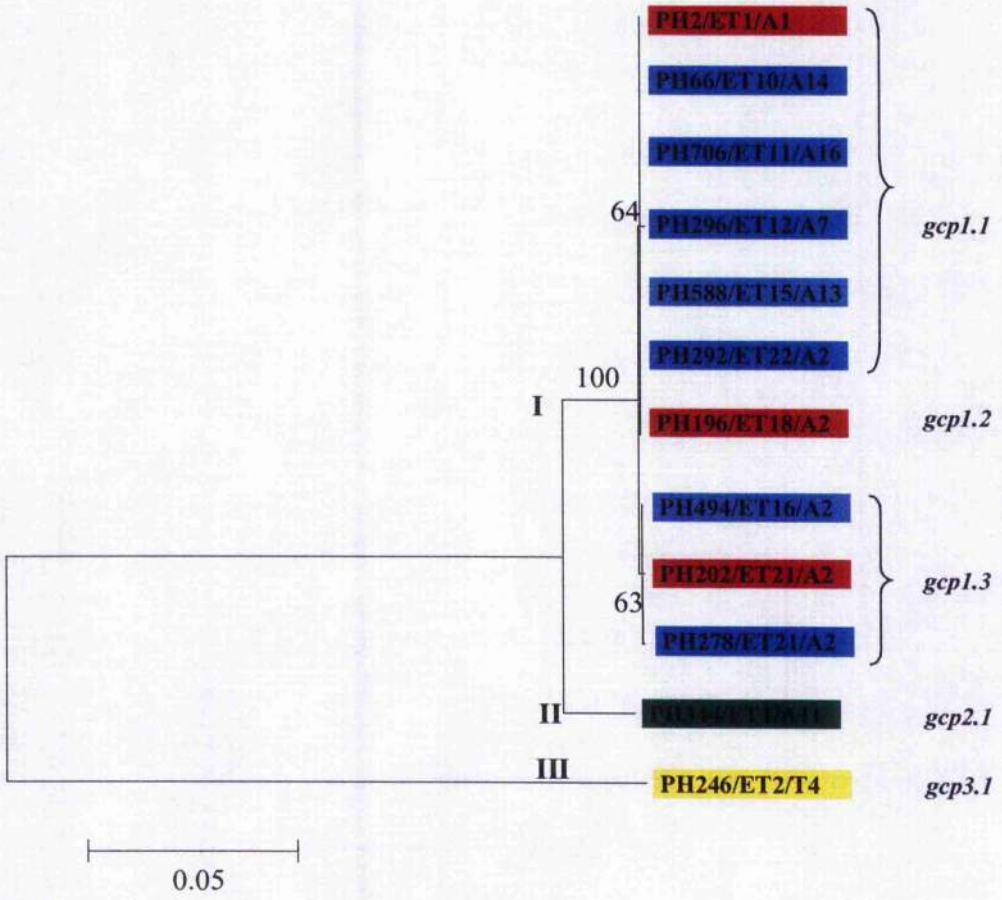


Figure 3.22 Minimum evolution (ME) tree for the *gcp* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option was used for handling gaps. The complete deletion option was used for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

The polymorphic nucleotide and inferred amino acid sites within the five *gcp* alleles, with respect to *gcp1.1* are shown in Figure 3.21. Nucleotide sequence analysis indicates that the *M. haemolytica gcp1*-type, *M. glucosida gcp2.1*, and *P. trehalosi mdh3.1* alleles have 2, 31, 219 polymorphic nucleotide sites, respectively (Figure 3.21A). Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.21B). Four amino acid variants were identified among the five *gcp* alleles; 2, 1, and 37 variable amino acid sites were detected in *M. haemolytica gcp1*-type, *M. glucosida gcp2.1*, and *P. trehalosi gcp3.1* alleles, respectively. There was no evidence of intragenic recombination within any of the *gcp* alleles.

The pairwise differences in nucleotide sequences indicated that the *M. haemolytica gcp1*-type alleles are highly conserved (1 to 2 polymorphic sites), whereas the *M. glucosida gcp2.1* allele shows moderate divergence from *M. haemolytica gcp1*-type alleles (30 to 31 polymorphic sites), and the *P. trehalosi gcp3.1* allele exhibits substantial divergence from both *M. haemolytica* (218 to 219 polymorphic sites) and *M. glucosida* (217 polymorphic sites).

The *gcp* tree topology (Figure 3.22) shows three distinct phylogenetic lineages, I to III, corresponding to the three species, *M. haemolytica* (*gcp1*), *M. glucosida* (*gcp2*), and *P. trehalosi* (*gcp3*), respectively. This phylogeny is in good agreement with those based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) and indicates that recombination has not affected the evolution of *gcp*.

3.1.1.4 Genes encoding periplasm-associated proteins

3.1.1.4.1 Lipoproteins (*plpA*, *plpB*, and *plpC*)

The contiguous *plpA*, *plpB*, and *plpC* genes, encoding lipoproteins PlpA, PlpB, and PlpC, were selected because the lipoproteins are highly immunogenic and are presumably subject to immune selection (section 1.3.2.3.1). The partial sequence of *plpA* (624 nucleotides [75 %]) was determined for 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate, whereas the complete sequence of *plpB* (831 nucleotides) and the partial sequence of *plpC* (771 nucleotides [99 %]) were determined for only 10 *M. haemolytica* isolates and one *M. glucosida* isolate (Table 3.18). Attempts to amplify *plpB* and *plpC* in *P. trehalosi* were unsuccessful. Four, five, and seven different nucleotide sequences, each representing a distinct allele, were identified for *plpA*, *plpB*, and *plpC*, respectively (Table 3.18). The contiguous *plpA*, *plpB*, and *plpC* genes (*plpABC*) of *M. haemolytica* and *M. glucosida* have nine different sequences which were designated alleles *plpABC1.1*, etc (Table 3.18).

The polymorphic nucleotide and inferred amino acid sites within the nine *plpABC* alleles, with respect to *plpABC1.1*, are shown in Figure 3.23. The number of polymorphic nucleotide sites within the *M. haemolytica plpABC1*-type, *M. glucosida plpABC2.1*, and *M. haemolytica plpABC3.1* alleles are 0, 3, and 29 in *plpA*, 2, 13, and 55 in *plpB*, and 4, 37, and 39 in *plpC*, respectively (Figure 3.23A). The *plpABC2.1* allele of *M. glucosida*, which is divergent from *M. haemolytica* in 16S rRNA and MLEE studies, has a recombinant segment (yellow, nucleotides1 to 1292) that is almost identical to the corresponding region of the *M. haemolytica plpABC1*-type alleles. The *M. haemolytica plpABC3.1* allele has diverged substantially from the *M. haemolytica plpABC1*-type alleles presumably due to the acquisition of a large DNA

Table 3.18 Distribution of *plpA*, *plpB*, and *plpC* alleles among 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	*plpA* allele	*plpB* allele	*plpC* allele	*plpABC* allele	GenBank accession no.
M. haemolytica								
PH2	1	A1	Bovine	*plpA1.1*	*plpB1.1*	*plpC1.1*	*plpABC1.1*	AY847813
PH66	10	A14	Ovine	*plpA1.1*	*plpB1.2*	*plpC1.4*	*plpABC1.3*	AY847814
PH706	11	A16	Ovine	*plpA1.1*	*plpB1.1*	*plpC1.2*	*plpABC1.2*	AY847815
PH296	12	A7	Ovine	*plpA1.1*	*plpB1.1*	*plpC1.4*	*plpABC1.6*	AY847816
PH588	15	A13	Ovine	*plpA1.1*	*plpB1.1*	*plpC1.4*	*plpABC1.6*	-
PH494	16	A2	Ovine	*plpA1.1*	*plpB1.1*	*plpC1.3*	*plpABC1.5*	AY847817
PH196	18	A2	Bovine	*plpA2.1*	*plpB3.1*	*plpC2.2*	*plpABC3.1*	AY847818
PH202	21	A2	Bovine	*plpA1.1*	*plpB1.1*	*plpC1.5*	*plpABC1.7*	AY847819
PH278	21	A2	Ovine	*plpA1.1*	*plpB1.1*	*plpC1.4*	*plpABC1.6*	-
PH292	22	A2	Ovine	*plpA1.1*	*plpB1.3*	*plpC1.4*	*plpABC1.4*	AY847820
M. glucosida								
PH344	1	A11	Ovine	*plpA1.2*	*plpB2.1*	*plpC2.1*	*plpABC2.1*	AY847821
P. trehalosi								
PH68	19	T3	Ovine	*plpA3.1*	-	-		AY847822

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

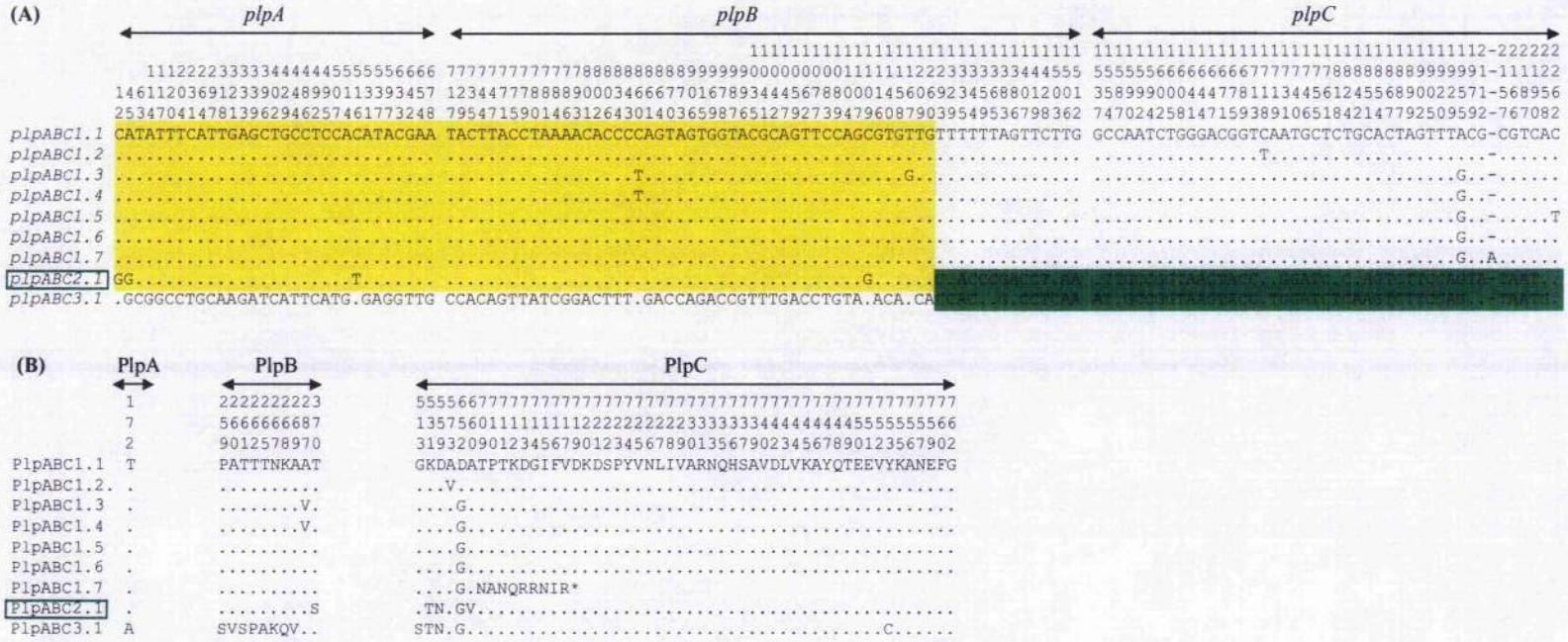


Figure 3.23 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *plpABC* alleles of *M. haemolytica* and *M. glucosida*. Allele designations are shown to the left of each sequence. The *M. glucosida* allele is highlighted within the green box. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial *plpA* gene. The coloured shading indicates nucleotide sequence similarity and putative recombinant segments in alleles that are otherwise different. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *plpABC1.1*). The * in the amino acid sequence of PlpABC1.7 in PlpC indicates a stop codon.

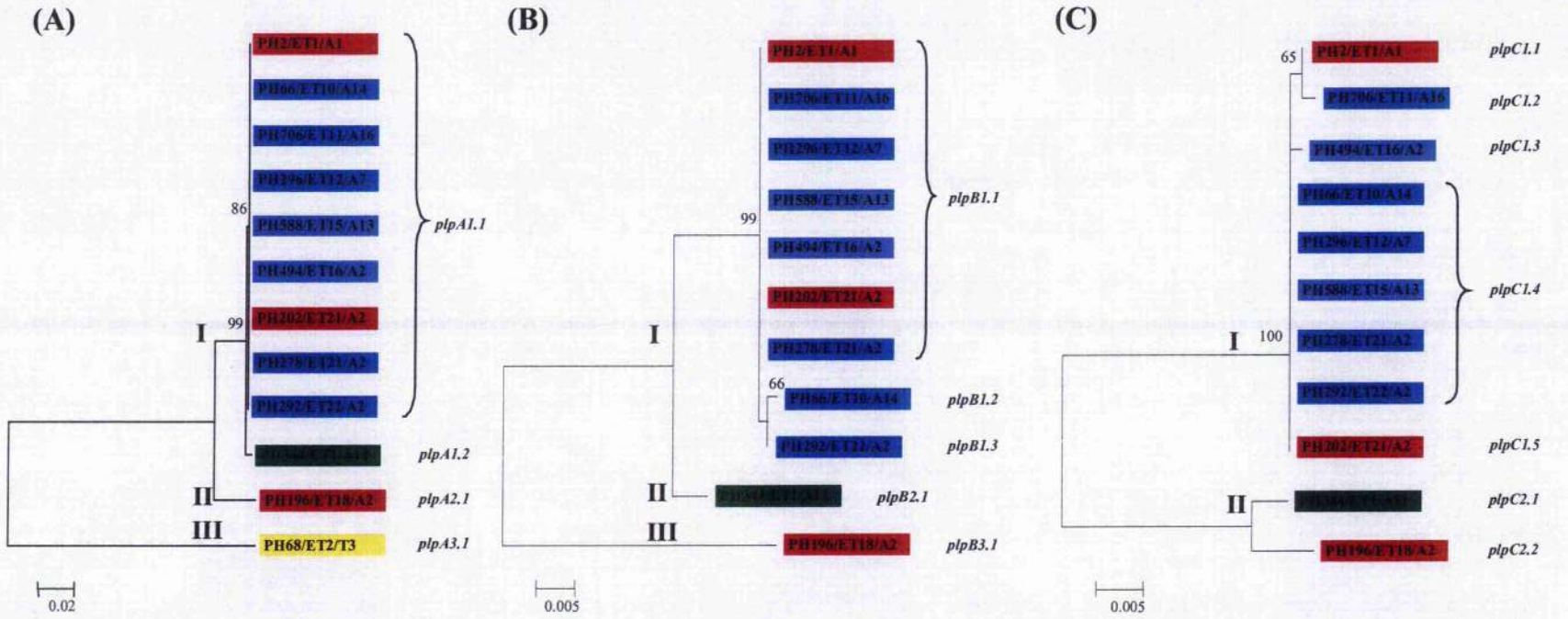


Figure 3.24 Minimum evolution (ME) trees for the (A) *plpA*, (B) *plpB*, and (C) *plpC* genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

segment by assortative recombination that includes all three genes. The recombinant segment of the *M. haemolytica* allele *plpABC3.1* has probably been derived from *M. glucosida*, since the green region of the *M. haemolytica plpABC3.1* allele (nucleotides 1293 to 2262) is very similar to the corresponding region of the *M. glucosida plpABC2.1* allele (Figure 3.23A). In addition, the *plpABC1.7* allele (isolate PH202) has a nucleotide insertion at position 2125 in *plpC*, which disrupts the reading frame. Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.23B). Six amino acid variants were identified among the nine *plpABC* alleles; *M. haemolytica plpABC1*-type, *M. glucosida plpABC2.1*, and *M. haemolytica plpABC3.1* alleles have 0, 0, and 1 variable amino acid sites in *plpA*, 1, 1, and 8 variable amino acid sites in *plpB*, and 2, 4, and 5 variable amino acid sites in *plpC* (excluding allele *plpABC1.7*).

The *plpA*, *plpB*, and *plpC* tree topologies are shown in Figure 3.24. The previously constructed phylogenies based on 16S rRNA sequence analysis (Davies *et al.*, 1996) and MLEE (Davies *et al.*, 1997a) indicate that *M. haemolytica* isolates are closely related, but have diverged from *M. glucosida* and *P. trehalosi*. In *plpA*, *plpB*, and *plpC*, this is true for all of the *M. haemolytica* isolates with the exception of isolate PH196. The *plpA*, *plpB*, and *plpC* genes of *M. haemolytica* isolate PH196 have diverged from those of the other *M. haemolytica* isolates (Figure 3.24 A to C). The *plpA* and *plpB* genes of *M. glucosida* isolate PH344 are more closely related to the majority of the *M. haemolytica* isolates than are the *plpA* and *plpB* genes of *M. haemolytica* isolate PH196 (Figure 3.24A and B). The *plpC* gene of *M. glucosida* isolate PH344 and *M. haemolytica* isolate PH196 cluster together in lineage II (Figure 3.24C). These findings suggest that horizontal DNA transfer and recombination have occurred in the *plpA*, *plpB*, and *plpC* genes of *M. haemolytica* isolate PH196 and *M. glucosida* isolate PH344, and disrupted the phylogenies of these genes.

3.1.1.4.2 Lipoprotein (*plpD*)

The *plpD* gene, encoding lipoprotein PlpD, was selected because the lipoprotein has a highly immunogenic nature (see section 1.3.2.3.2) and is presumably subject to immune selection. The partial sequence of *plpD* (651 nucleotides [76 %]) was determined in 10 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate (Table 3.19). An additional *M. haemolytica* isolate, PH550, was examined because the inferred amino acid sequence of the *plpD* gene in the closely related isolate PH494 has an abnormal termination codon due to nucleotide deletions (see below). Eight different *plpD* nucleotide sequences, each representing a distinct allele, were identified, and the alleles were assigned to two subclasses, *plpD1* and *plpD2*, based on their overall sequence similarities (Table 3.19).

The polymorphic nucleotide and inferred amino acid sites within the eight *plpD* alleles, with respect to *plpD1.1* are shown in Figure 3.25. Nucleotide sequence analysis indicated that the *M. haemolytica plpD1*-type, *M. glucosida plpD1.7*, and *P. trehalosi plpD2.1* alleles have 5, 5, and 131 polymorphic nucleotide sties, respectively (Figure 3.25A). In addition, the *plpD1.4* allele (isolates PH494 and PH550) both have 13 nucleotide deletions, between nucleotides 507 to 519, which disrupts the reading frame. Amino acid sequence analysis detected less variation than nucleotide sequence analysis (Figure 3.25B). Five amino acid variants were identified among the seven of eight *plpD* alleles; 3, 1, and 32 variable amino acid positions were detected in the *M. haemolytica plpD1*-type, *M. glucosida plpD1.7*, and *P. trehalosi plpD2.1* alleles, respectively. The 13 nucleotide deletion in *plpD1.4* caused a sudden change in amino acid sequence from position 169 and a stop codon at position 178.

Pairwise differences in nucleotide sequences indicated that the *M. haemolytica plpD1*-

Table 3.19 Distribution of *plpD* alleles among 11 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate

Isolate	ET [a]	Capsular serotype	Host species	*plpD* allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*plpD1.1*	AY847823
PH66	10	A14	Ovine	*plpD1.1*	-
PH706	11	A16	Ovine	*plpD1.1*	-
PH296	12	A7	Ovine	*plpD1.2*	AY847824
PH588	15	A13	Ovine	*plpD1.6*	AY847825
PH494	16	A2	Ovine	*plpD1.4*	AY847826
PH550*	17	A2	Bovine	*plpD1.4*	-
PH196	18	A2	Bovine	*plpD1.5*	AY847827
PH202	21	A2	Bovine	*plpD1.3*	AY847828
PH278	21	A2	Ovine	*plpD1.3*	-
PH292	22	A2	Ovine	*plpD1.3*	-
M. glucosida					
PH344	1	A11	Ovine	*plpD1.7*	AY847829
P. trehalosi					
PH68	19	T3	Ovine	*plpD2.1*	AY847830

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

* Additional isolate (see text)

```
                               11111111111111111111222222222333333333333333333
           11122333344556667789113333455556677788992344789991233333344444455
        16723404036928470695846470125136790534709785736681478701456801267814
plpD1.1 TTAGAGAATACGAGCAAATCTAACTAAAGCATATCTCTTTCGTTTATTTAATAACTTCATGCTACTCT
plpD1.2 ......................................................T.............
plpD1.3 ....................................T.......................A.......
plpD1.4 ....................................T.......................A.......
plpD1.5 ............................................................A.......
plpD1.6 ............................................................AA......
plpD1.7 .......................................C....C...............A.......
plpD2.1 CCGATCCTATTTTATTGGCTCTTTACGCTTGCTCTC.AC.TCCCCTCCCTGCTT.GAGCAAACGAATC
```

```
        3333334444444444444444444445555555555555555555555555566666666666666
        6778890222333344445556667788000011222233446667777889901122233334455
        3787802012358901473791561738145736027817451570123455735857801890114
plpD1.1 CCGTAAATCAGATAAAAATGACCTATCCACTAGTTTCATTGTACATGAAATATATTGAATGCATTTA
plpD1.2 ...................................................................
plpD1.3 .............................................................A.....
plpD1.4 ...............................---...........................A.....
plpD1.5 ...................................................................
plpD1.6 ...................................................................
plpD1.7 ......................T.......C....................................
plpD2.1 TACAGTGAGCTTAGGTTGAATATCTCTTGTCTACGGTGGACACGGGACTTACCTACCTGAA.TGACG
```

```
               1111111111111111111111111111111111111111222222222222222
          14556111123444555667777777778888888999999999900000000001111
        3714386134660157346390123456890124789012345678901234567890134
PlpD1.1 KEEKNPGSFATATQAKVAYALSQRRSNTKARLVQTAQIEAVGYGKANQVKACDG--ETGAL
PlpD1.2 ......................................................--.....
PlpD1.3 .....L...T............................................--...E.
PlpD1.4 .....L...T.........FLIQLKLV*
PlpD1.5 .........T............................................--.....
PlpD1.6 .........N............................................--.....
PlpD1.7 .........T............................................--.....
PlpD2.1 EADRY.AAYNEDAASGIEHV.....A.M....Q...EMT...L...H.....E.YAHAS.E
```

Figure 3.25 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *plpD* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The vertical numbers above the sequences represent the positions of polymorphic nucleotide and amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *plpD1.1*). The * in the amino acid sequence of *plpD1.4* indicates a stop codon.

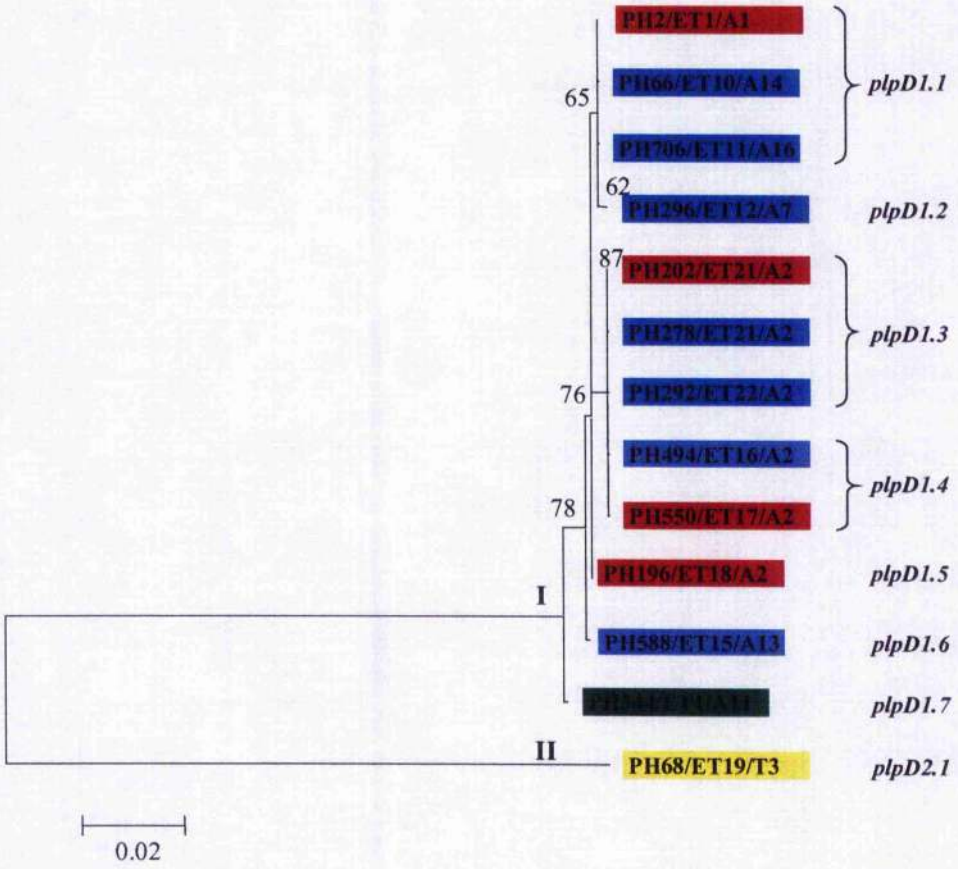


Figure 3.26 Minimum evolution (ME) tree for the *plpD* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

type and *M. glucosida plpD1.7* alleles are highly conserved (1 to 8 polymorphic sites), whereas the *P. trehalosi plpD2.1* allele exhibits substantial divergence from both species (128 to 132 polymorphic sites).

The *plpD* tree topology (Figure 3.26) shows two distinct phylogenetic lineages, I and II. *M. haemolytica plpD1*-type and *M. glucosida plpD1.7* alleles cluster together in lineage I, whereas *P. trehalosi plpD2.1* represents the more distantly related lineage II.

3.1.1.5 Genes encoding outer membrane proteins

3.1.1.5.1 Heat modifiable outer membrane protein (*ompA*)

The *ompA* gene, encoding the heat modifiable outer membrane protein OmpA, was selected because of its possible role in bacterial pathogenesis and host specificity (see section 1.3.2.4.1. The complete sequence of *ompA* (1,104 to 1,137 nucleotides) was determined in 31 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates (Table 3.20). *M. haemolytica* and *M. glucosida* isolates contain a single *ompA* gene, but *P. trehalosi* was shown to possess two tandemly-arranged *ompA* genes, designated *ompA'* and *ompA"* (Figure 2.27). Twenty two different *ompA* sequences, each representing a distinct allele, were identified, and the alleles were assigned to ten subclasses, *ompA1* to *ompA10*, based on their overall sequence similarity (Table 3.20). Of these, four subclasses, *ompA1* to *ompA4*, occur in *M. haemolytica*, three subclasses, *ompA5* to *ompA7*, are present in *M. glucosida*, and three subclasses, *ompA8* to *ompA10*, are associated with *P. trehalosi*. The molecular masses of the inferred amino acid OmpA sequences of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* varied from 39,135 to 40, 528 Da (Table 3.20). Since these proteins contain a putative signal sequence of 19 amino acids (1,855 Da) (Zeng *et al.*, 1999), the predicted

Table 3.20 Distribution of *ompA* alleles among 31 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Isolate	ET [a]	Capsular serotype	Host species	*ompA* [b] allele	Molecular mass of protein (Da)	GenBank accession no.
M. haemolytica						
PH2	1	A1	Bovine	*ompA1.1*	40,460	AY244653
PH30	1	A1	Bovine	*ompA1.1*	40,460	-
PH376	1	A6	Bovine	*ompA1.1*	40,460	-
PH346	1	A12	Ovine	*ompA2.1*	39,778	AY244658
PH540	2	A1	Bovine	*ompA1.2*	40,472	AY244654
PH338	3	A9	Ovine	*ompA2.1*	39,778	-
PH388	4	A7	Ovine	*ompA2.1*	39,778	-
PH50	5	A5	Ovine	*ompA2.1*	39,778	-
PH56	5	A8	Ovine	*ompA2.2*	39,794	AY244659
PH238	5	A9	Ovine	*ompA2.1*	39,778	-
PH8	6	A1	Ovine	*ompA2.1*	39,778	-
PH398	7	A1	Ovine	*ompA2.1*	39,778	-
PH284	8	A6	Ovine	*ompA2.1*	39,778	-
PH232	9	A6	Ovine	*ompA2.1*	39,778	-
PH66	10	A14	Ovine	*ompA2.1*	39,778	-
PH706	11	A16	Ovine	*ompA2.1*	39,778	-
PH296	12	A7	Ovine	*ompA4.1*	39,151	AY244662
PH396	13	A7	Ovine	*ompA4.1*	39,151	-
PH484	14	A7	Ovine	*ompA4.1*	39,151	-
PH588	15	A13	Ovine	*ompA4.2*	39,179	AY244663
PH494	16	A2	Ovine	*ompA1.4*	40,528	AY244656
PH550	17	A2	Bovine	*ompA1.5*	40,062	AY244657
PH196	18	A2	Bovine	*ompA3.1*	39,911	AY244661
PH786	18	A2	Bovine	-		-
PH526	19	A2	Ovine	*ompA2.3*	39,798	-
PH598	20	A2	Ovine	*ompA2.3*	39,798	-
PH202	21	A2	Bovine	*ompA1.3*	40,460	AY244655
PH470	21	A2	Bovine	*ompA1.3*	40,460	-
PH278	21	A2	Ovine	*ompA2.3*	39,798	AY244660
PH372	21	A2	Ovine	*ompA2.3*	39,798	-
PH292	22	A2	Ovine	*ompA2.3*	39,798	-
PH392	22	A2	Ovine	*ompA2.3*	39,798	-
M. glucosida						
PH344	1	A11	Ovine	*ompA5.1*	39,135	AY244664
PH498	3	A11	Ovine	*ompA5.1*	39,135	-
PH240	5	A11	Ovine	*ompA7.1*	39,233	AY244666
PH496	7	UG3	Ovine	*ompA6.1*	39,233	AY244665
PH574	10	UG3	Ovine	*ompA7.1*	39,233	-
PH290	16	UG3	Ovine	*ompA6.1*	39,233	-
P. trehalosi						
PH246	2	T4	Ovine	*ompA8.1*	38,633	AY582755
				ompA10.1	37,978	
PH252	4	T10	Ovine	*ompA8.2*	38,633	AY582756
				ompA10.2	37,978	
PH254	15	T15	Ovine	*ompA9.1*	38,662	AY582757
				ompA10.4	37,978	
PH68	19	T3	Ovine	*ompA8.3*	38,844	AY582758
				ompA10.3	37,978	

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

molecular masses of the putative mature proteins varied from 37,280 to 38,673 Da.

The polymorphic nucleotide sites within the *ompA1*- to *ompA7*-type alleles of *M. haemolytica* and *M. glucosida*, with respect to *ompA1.1*, are shown in Figure 3.27. The *P. trehalosi ompA8- to ompA10*-type alleles were excluded because they are highly divergent from both *M. haemolytica* and *M. glucosida* and there was no visible evidence of intragenic recombination. There was a relatively high degree of nucleotide variation between *ompA1*-type and *ompA2*-type alleles, but alleles representing each group contain only one to three polymorphic nucleotide sites. Visual inspection of the nucleotide sequences of the *ompA1*- to *ompA7*-type alleles indicates that there was no evidence of intragenic recombination within *M. haemolytica* alleles *ompA1.1* to *ompA3.1*, but recombinant segments were present in *M. haemolytica ompA4*- and *M. glucosida ompA5*-type alleles. The *M. haemolytica ompA4.1* and *ompA4.2* alleles (serotype A7 and A13 isolates of ETs 12-14) contain recombinant segments (green and yellow) that are identical to the corresponding regions of the *M. glucosida ompA5.1*, *ompA6.1*, and/or *ompA7.1* alleles. The *M. glucosida* alleles *ompA5.1* to *ompA7.1* are similar to each other but contain regions (yellow and blue segments) that differ due to recombinational exchanges.

Pairwise differences in nucleotide and amino acid sequences between representative pairs of the 10 allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Table 3.21) ranged from 8 to 304 (0.7 to 26.7 %) nucleotide sites and from 0 to 122 (0.0 to 32.3 %) amino acid positions.

The *ompA* tree topology (Figure 3.28) indicates that the *M. haemolytica* lineages I to III and *M. glucosida* lineage IV are closely related, whereas *P. trehalosi* lineages V and VI, representing *ompA'* and *ompA"*, respectively, are highly divergent from

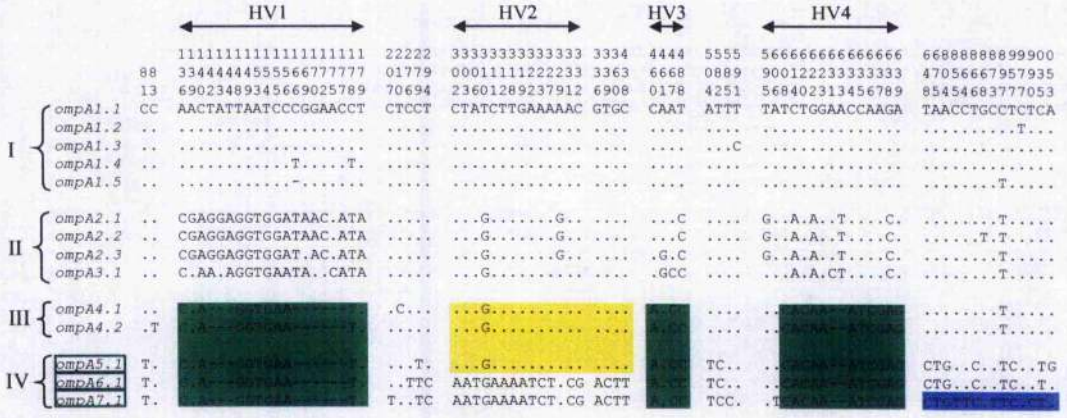


Figure 3.27 Distribution of polymorphic nucleotide sites among the 14 *ompA* alleles of *M. haemolytica* and *M. glucosida*. Allele designations are shown to the left of each sequence. Roman numerals I to IV represent the major allele classes and lineages in Figure 3.28. The *M. glucosida* alleles are highlighted within the green box. The numbers above the sequences represent the positions of polymorphic nucleotide sites. The dots represent sites where the nucleotides match those of the first (topmost) sequence. Gaps are indicated by dashes. Green, yellow, and blue shaded boxes highlight identical, or nearly identical, segments of DNA in class III and IV alleles. HV1 to HV4 represent the hypervariable domains.

Table 3.21 Nucleotide and amino acid differences between representative pairs of the 10 *ompA* allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Allele	Pairwise differences in nucleotide and amino acid sequences (%)[a]									
	ompA1.1	***ompA2.1***	***ompA3.1***	***ompA4.1***	***ompA5.1***	***ompA6.1***	***ompA7.1***	***ompA8.1***	***ompA9.1***	***ompA10.1***
ompA1.1		18 (4.8)	17 (4.5)	16 (4.2)	17 (4.5)	23 (6.1)	23 (6.1)	101 (26.7)	98 (25.9)	122 (32.3)
ompA2.1	29 (2.6)		12 (3.2)	15 (4.0)	16 (4.2)	22 (5.8)	22 (5.8)	96 (25.4)	94 (24.9)	120 (31.7)
ompA3.1	26 (2.3)	13 (1.1)		9 2.4()	10 (2.6)	17 (4.5)	17 (4.5)	95 (25.1)	92 (24.3)	118 (31.2)
ompA4.1	26 (2.3)	21 (1.8)	16 (1.4)		1 (0.3)	8 (2.1)	8 (2.1)	90 (23.8)	88 (23.3)	118 (31.2)
ompA5.1	36 (3.2)	31 (2.7)	26 (2.3)	12 (1.1)		7 (1.9)	7 (1.9)	89 (23.5)	87 (23.0)	117 (31.0)
ompA6.1	53 (4.7)	47 (4.1)	43 (3.8)	29 (2.6)	19 (1.7)		0 (0.0)	88 (23.3)	86 (22.8)	117 (31.0)
ompA7.1	59 (5.2)	53 (4.7)	49 (4.3)	35 (3.1)	25 (2.2)	8 (0.7)		88 (23.3)	86 (22.8)	117 (31.0)
ompA8.1	264 (23.2)	256 (22.5)	252 (22.2)	239 (21.0)	237 (20.8)	236 (20.8)	238 (20.9)		40 (10.6)	83 (22.0)
ompA9.1	244 (21.5)	233 (20.5)	231 (20.3)	218 (19.2)	216 (19.0)	215 (18.9)	217 (19.1)	135 (11.9)		86 (22.8)
ompA10.1	304 (26.7)	298 (26.2)	294 (25.9)	297 (26.1)	298 (26.2)	293 (25.8)	293 (25.8)	223 (19.6)	221 (19.4)	

[a] Values in the lower left represent pairwise differences in nucleotide sequences (number of polymorphic nucleotide sites and percentage of nucleotide variation); values in the upper right represent pairwise differences in inferred amino acid sequences (number of polymorphic amino acid sites and percentage of amino acid variation).

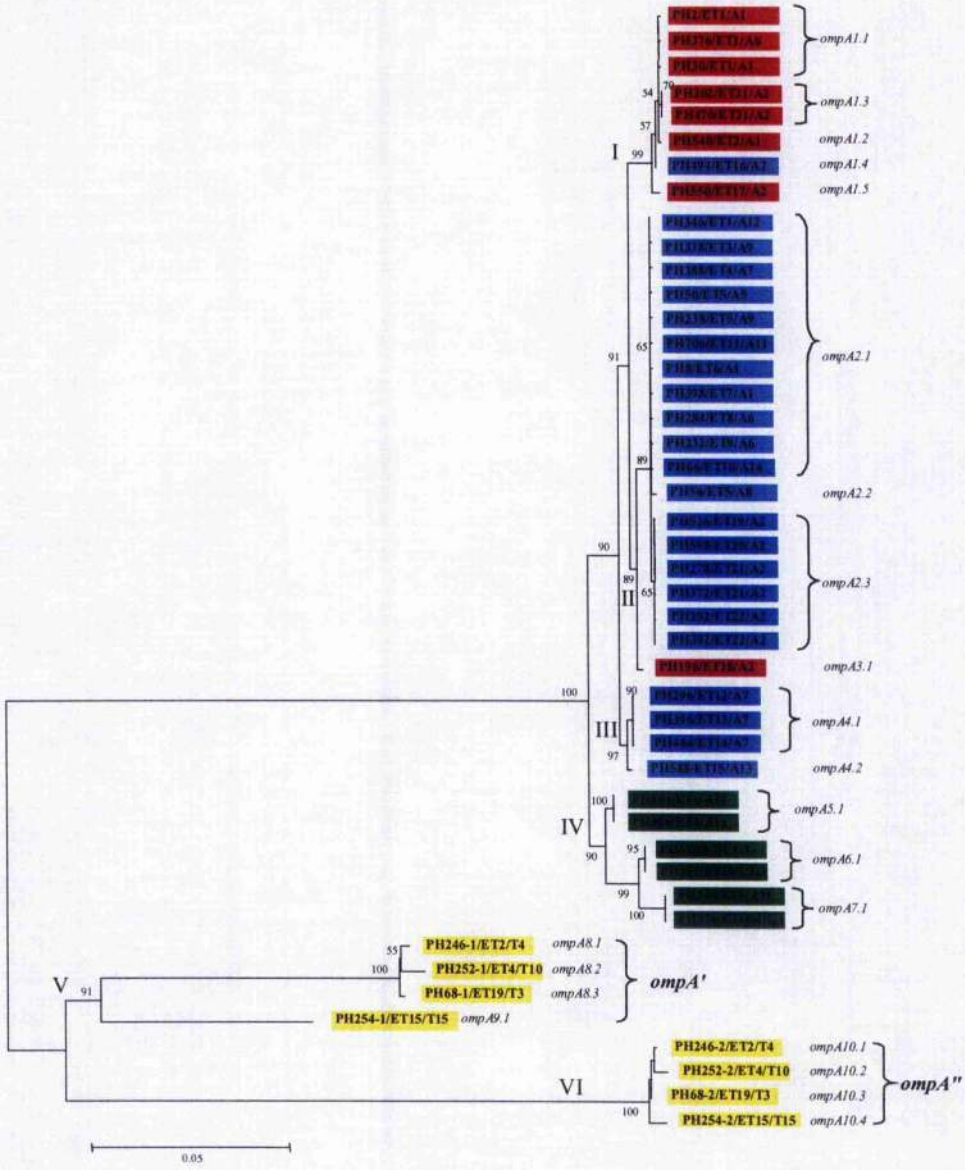


Figure 3.28 Minimum evolution (ME) tree for the *ompA* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

M. haemolytica and *M. glucosida* and are as divergent from each other as they are from the *M. haemolytica* and *M. glucosida* lineages. Lineage I is associated exclusively with bovine *M. haemolytica* isolates, with the exception of the atypical ovine isolate PH494 (ET 16), whereas lineages II and III are associated exclusively with ovine *M. haemolytica* isolates, with the exception of the atypical bovine isolate PH196 (ET 18).

The nucleotide and amino acid sequence analysis of *M. haemolytica* and *M. glucosida* indicates that the majority of polymorphic sites occur within four hypervariable regions, HV1 to HV4 (Figures 3.27 and 3.29). The hypervariable domains contain higher nucleotide and amino acid substitutions (35 and 57 %, respectively), whereas the remaining regions are highly conserved and contain a low percentage of nucleotide and amino acid substitutions (3 and 1 %, respectively). Domains HV1, HV2, and HV4 were characterized by amino acid deletions and/or insertions (Figure 3.29) that accounted for the molecular mass variation of OmpA described in Table 3.20.

The locations of the four hypervariable regions within the *M. haemolytica* and *M. glucosida* OmpA proteins, in relation to the β-strands and surface-exposed loops of the transmembrane domain, were identified by secondary structure prediction (Jones, 1999; McGuffin *et al.*, 2000), alignment with the three-dimensional structural models of the *E.coli* OmpA protein (Arora *et al.*, 2001; Pautsch & Schulz, 1998), and comparison with the proposed secondary structure model of the P5 (OmpA) protein of *H. influenzae* (Webb & Cripps, 1998). The OmpA protein of *M. haemolytica* isolate PH2 contains four surface exposed loops and eight β-strands regions. The four hypervariable regions are located at the distal ends of the surface exposed loops (Figure 3.30). The d_s/d_N ratios for the hypervariable domains HV1 to HV4 ranged

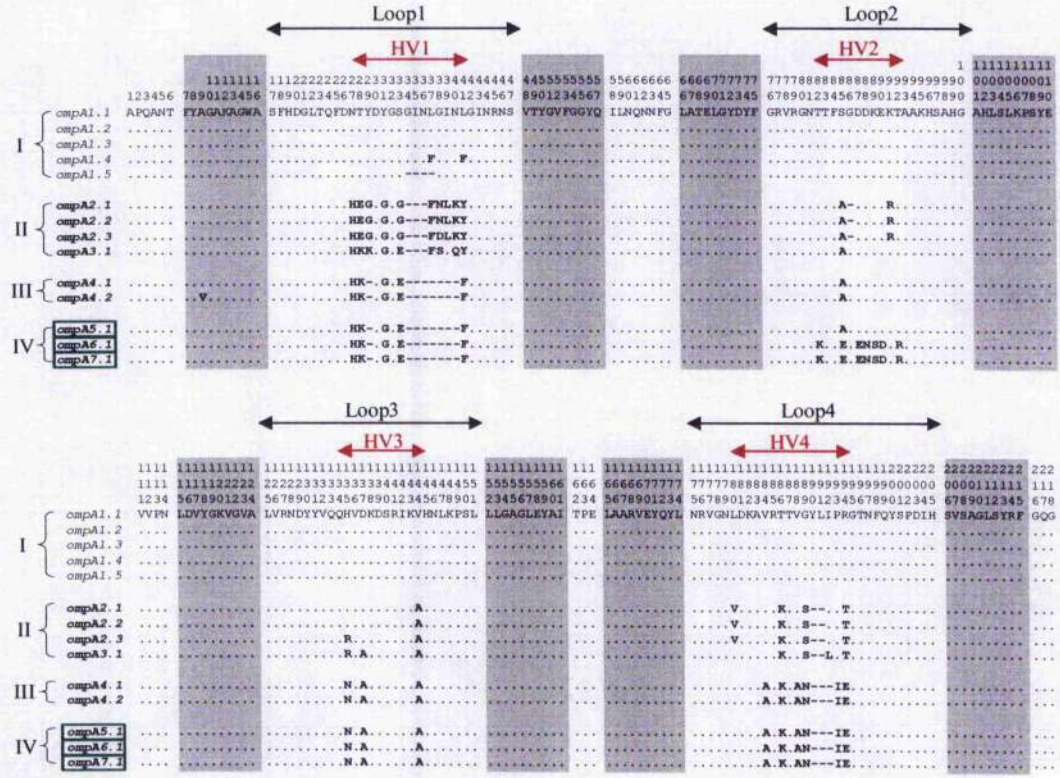


Figure 3.29 Distribution of polymorphic amino acid sites among the 14 *ompA* alleles within the transmembrane domain of the OmpA proteins of *M. haemolytica* and *M. glucosida*. Protein designations are shown to the left of each sequence. Roman numerals I to IV represent the major allele classes. The *M. glucosida* alleles are highlighted within the green box. The vertical numbers above the sequences represent amino acid positions. The dots represent sites where the amino acids match those of the first (topmost) sequence (i.e. *ompA1.1*). Gaps are indicated by dashes. HV1 to HV4 represent the hypervariable domains within the surface exposed loops 1 to 4. Shaded regions represent predicted membrane-spanning β-strand structures.

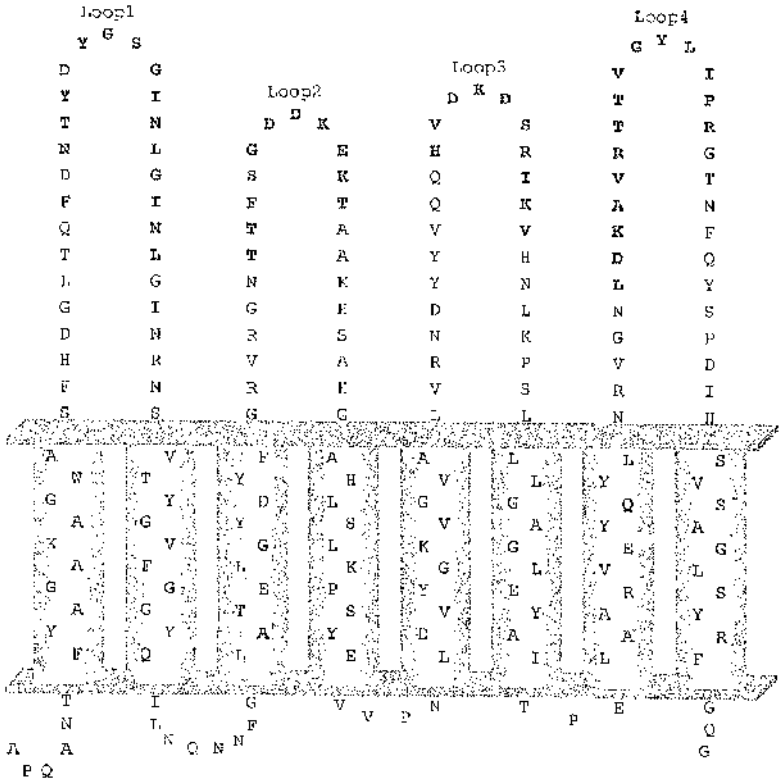


Figure 3.30 Proposed secondary structure of the N-terminal transmembrane domain of the OmpA proteins of *M. haemolytica* and *M. glucosida*. The sequence is based on OmpA1.1 of isolate PH2. The hypervariable domains are shown in red.

from 0.34 to 0.78 (i.e < 1), whereas the corresponding value for the combined conserved regions was relatively high at 35.17 (Table 3.22). These data provide strong evidence of selective constraint against amino acid replacement in the conserved parts of the gene and of diversifying selection in the hypervariable regions.

The comparison of the MLEE and *ompA* trees (Figure 3.31) indicate that distantly related isolates have identical or almost identical alleles. This suggests that these alleles have undergone horizontal transfer and assortative (entire gene) recombination. Distantly related bovine isolates (PH2, PH540, PH550, and PH202; ETs 1, 2, 17, and 21) and one ovine isolate (PH494; ET 16, which has previously been shown to have bovine properties (Davies & Donachie, 1996; Davies *et al.*, 2001), in the MLEE tree clustered together in the *ompA* tree to form a single major bovine lineage, I. Similarly, distantly related ovine isolates of lineages A and C (PH338, PH388, PH56, PH8, PH398, PH284, PH232, PH66, PH706,PH56; ETs 3 to 11 and PH526, PH598, PH278, PH292; ETs 19 to 22) in the MLEE tree clustered together in the *ompA* tree to form the major ovine lineage, II. The atypical bovine isolate PH196 represents a sub branch of lineage II. Ovine *M. haemolytica* isolates of lineage B together with certain *M. glucosida* isolates (PH344 and PH498; ETs 1 and 3) in the MLEE tree clustered together in the *ompA* tree to form lineage III. Other *M. glucosida* isolates clustered together in the *ompA* tree to form lineage IV.

For the two tandemly arranged *ompA'* and *ompA"* genes of *P. trehalosi*, putative ribosome-binding and promoter sites were identified upstream of the start codons of *ompA'* and *ompA"*, and potential inverted repeat terminator sequences were present downstream of the stop codons of *ompA'* and *ompA"* (Figure 3.32). The *ompA'* alleles were assigned to two subclasses, *ompA8* (*ompA8.1* to *ompA8.3*) and *ompA9* (*ompA9.1*), whereas the *ompA"* alleles consisted of a single subclass, *ompA10*

Table 3.22 Sequence diversity and substitution rates for the hypervariable domains HV1 to HV4 and for the conserved regions of the *ompA* genes of 31 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Domain	Sequence diversity (%)		d_S [a]	d_N [a]	d_S/d_N
	Nucleotide	Amino acid			
HV1	44.4	66.7	13.96 ± 9.92	40.64 ± 14.70	0.34
HV2	42.4	72.7	8.31 ± 6.01	15.86 ± 5.64	0.52
HV3	13.3	30.0	0.00 ± 0.00	6.91 ± 3.98	-
HV4	33.3	53.3	15.18 ± 11.03	19.54 ± 8.23	0.78
Conserved regions [b]	3.0	1.2	2.11 ± 0.43	0.06 ± 0.04	35.17

[a] d_S is the number of synonymous substitution per 100 synonymous sites; d_N is the number of nonsynonymous substitution per 100 nonsynonymous sites. Values are means ± standard deviations.

[b] Values are for the individual conserved regions combined.

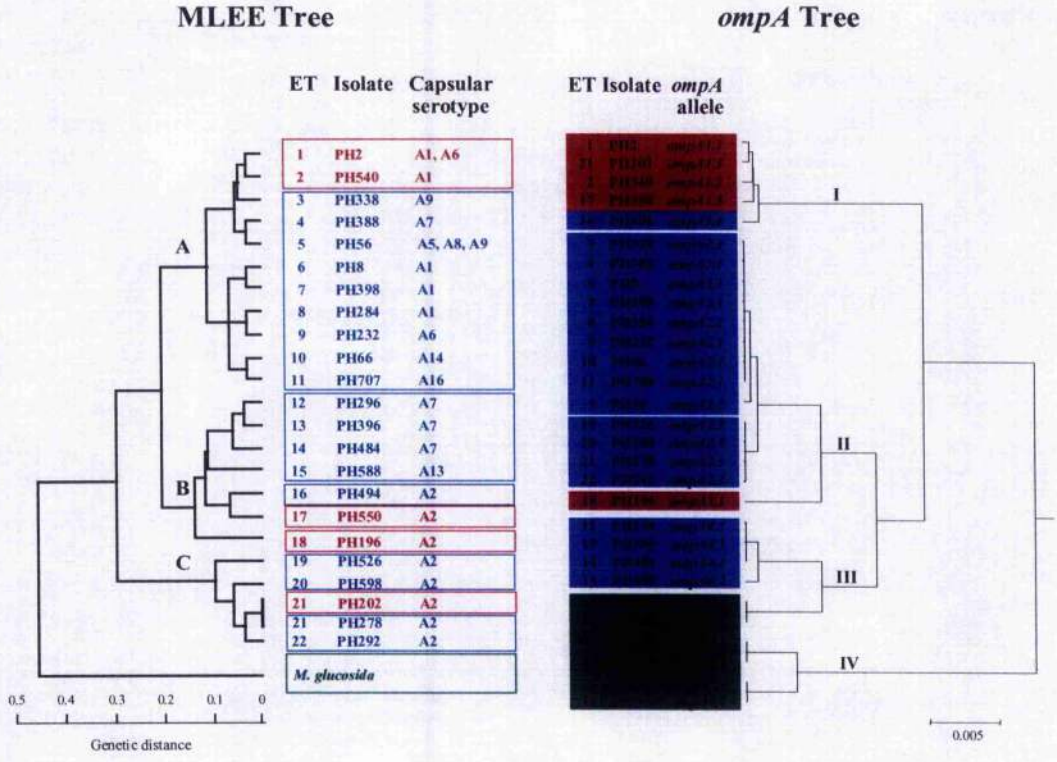


Figure 3.31 Comparison of MLEE tree (left) and *ompA* tree (right). (Left) Phylogenetic relationships of 178 *M. haemolytica* and 16 *M. glucosida* isolates based on electrophoretically demonstrable variation of 18 housekeeping enzymes (Davies *et al.*, 1997a). (Right) Phylogenetic relationships of *ompA* genes from 31 *M. haemolytica* and six *M. glucosida* strains. Both trees were constructed by the unweighted pair group method using arithmetic average (UPGMA). Red boxes represent the bovine isolates. Blue boxes represent the ovine isolates. Green boxes represent the *M. glucosida* isolates.

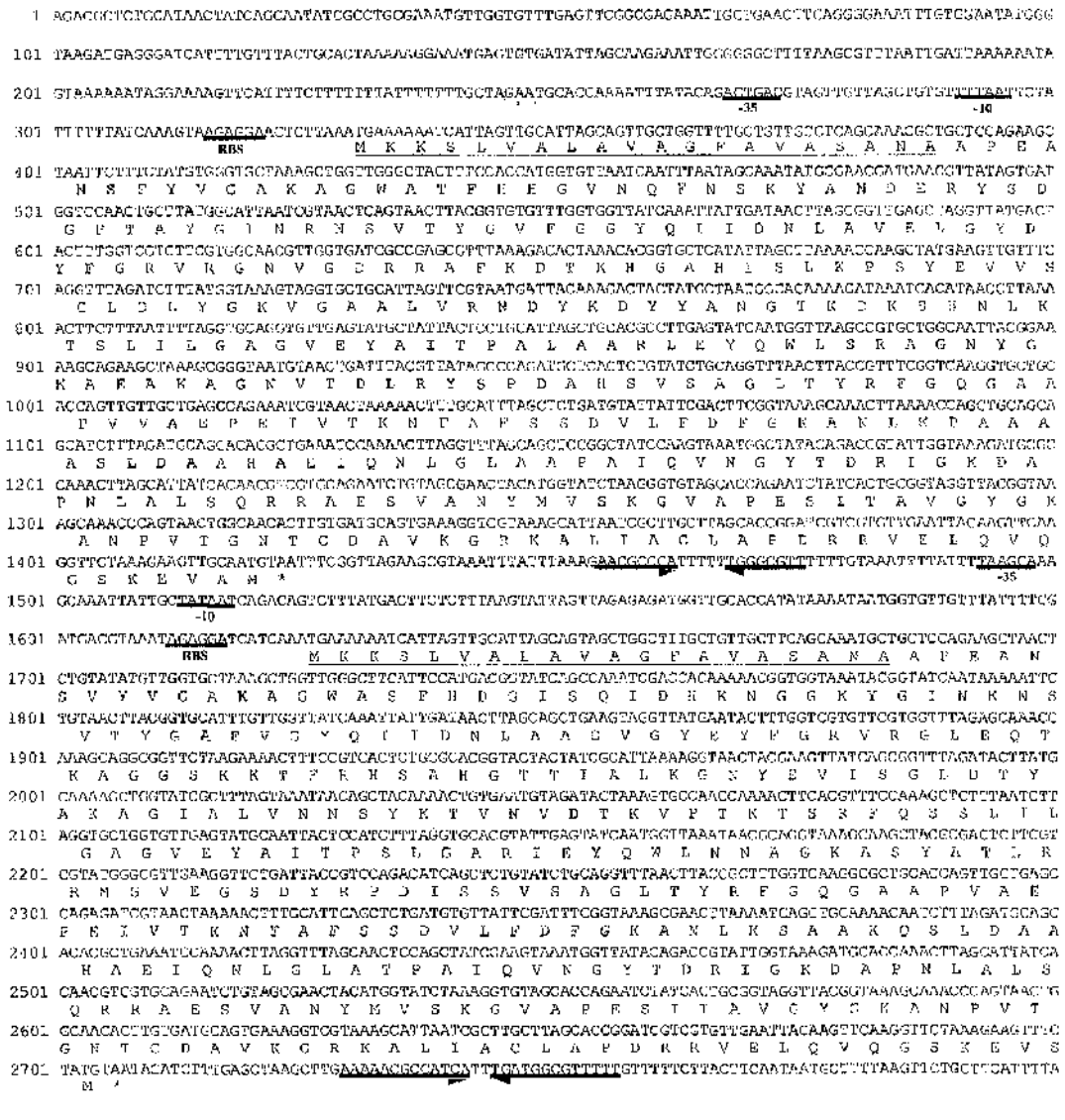


Figure 3.32 Nucleotide sequence of the *P. trehalosi* isolate PH246 *ompA'* and *ompA"* genes, together with the deduced amino acid sequences of the OmpA' and OmpA" proteins, respectively, encoded by these genes. The numbers (left) correspond to nucleotide positions within the GenBank sequence (AY582755). The * in the amino acid sequences indicate stop codons. The proposed promoters (-35 and -10 regions) and the putative ribosomal binding sites (RBS) are underlined in bold. The putative N-terminal signal sequences of the OmpA' and OmpA" proteins are underlined. Inverted nucleotide repeats are indicated by opposing arrows.

(*ompA10.1* to *ompA10.4*) (Figure 3.28). Nucleotide and amino acid analyses of *ompA'* indicate that allele *ompA9.1* is highly divergent, whereas alleles *ompA8.1* to *ompA8.3* are highly conserved (0.9 and 0.2 % nucleotide and amino acid diversity, respectively). The corresponding analyses of *ompA"* show that all four *ompA10*-type alleles are highly conserved (0.8 and 0.0 % nucleotide and amino acid diversity, respectively). The sequence comparison of the two OmpA homologs of *P. trehalosi* indicates that the major three subclasses, OmpA8, OmpA9, and OmpA10 contain significant amino acid variation in their N- terminal regions, particularly in the surface-exposed regions, whereas the C- terminal regions are highly conserved (Figure 3.33).

The expression of two OmpA proteins, OmpA' and OmpA", by *P. trehalosi* was confirmed by SDS-PAGE analysis of OMPs after heating the samples at 80, 90, and 100°C (Figure 3.34). The OmpA protein undergoes a characteristic shift from low- to high-molecular-mass forms after heating at 100°C. Two low-molecular-mass proteins (29 to 30 kDa) were clearly visible for isolates PH68, PH246, and PH252 at 80 and 90°C, whereas these were transformed to two high-molecular-mass proteins (37 to 39 kDa) at 100°C (Figure 3.34, lanes 1 to 9, arrows). However, only one low-molecular-mass band was present in isolate PH254 at 80 and 90°C (Figure 3.34, lanes 10 and 11, arrow), although two bands, of 29 and 39 kDa, were present at 100°C (Figure 3.34, lane 12, arrows). The most probable explanation for this difference is that the single 29-kDa band present at 80 and 90°C consists of both proteins (OmpA' and OmpA"), which in this isolate differ in their heat modification properties. Thus, at 100°C the 29-kDa band presumably corresponds to the OmpA' protein OmpA9.1 (since this band and protein were not present in any of the other isolates), whereas the 39-kDa band represents the OmpA" protein OmpA10.4.

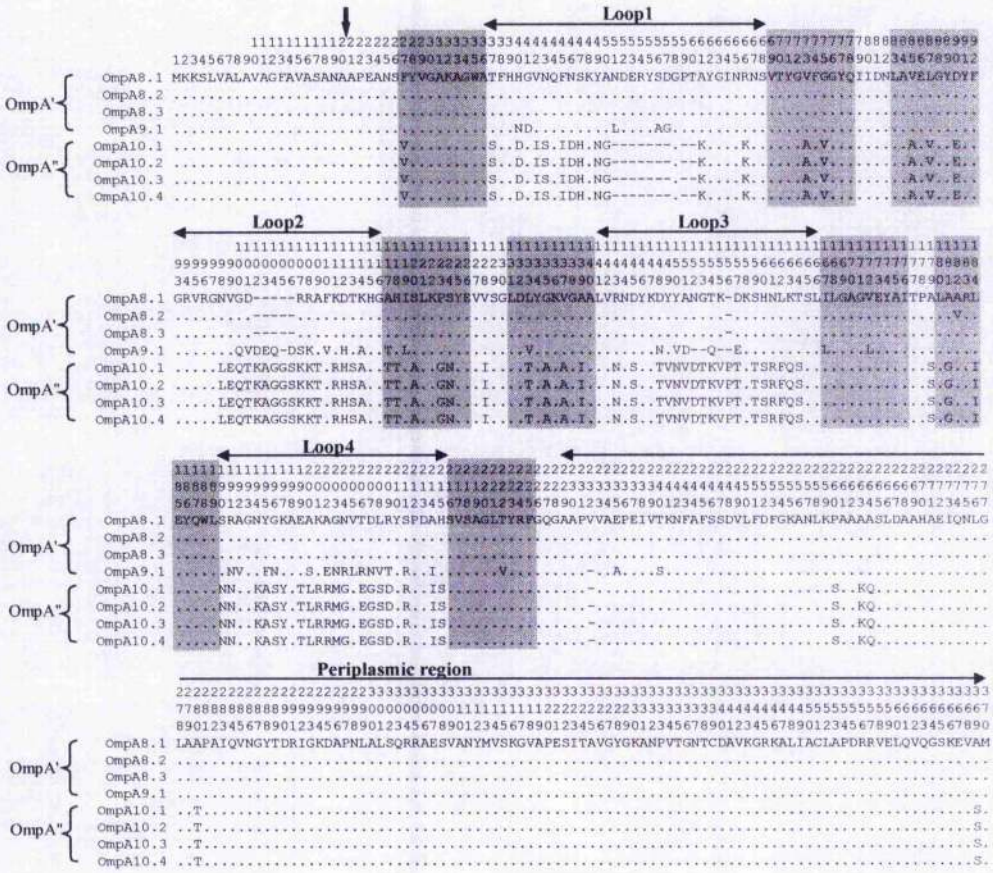


Figure 3.33 Distribution of polymorphic amino acid sites among the eight *ompA* alleles of the OmpA' and OmpA" proteins of *P. trehalosi*. Protein designations are shown to the left of each sequence. The vertical numbers above the sequences represent amino acid positions. The dots represent sites where the amino acids match those of the first (topmost) sequence (i.e. *ompA1.1*). Gaps are indicated by dashes. The locations of the surface exposed loops 1 to 4 and periplasmic region are indicated by arrows. Shaded regions represent predicted membrane-spanning β-strand structures.

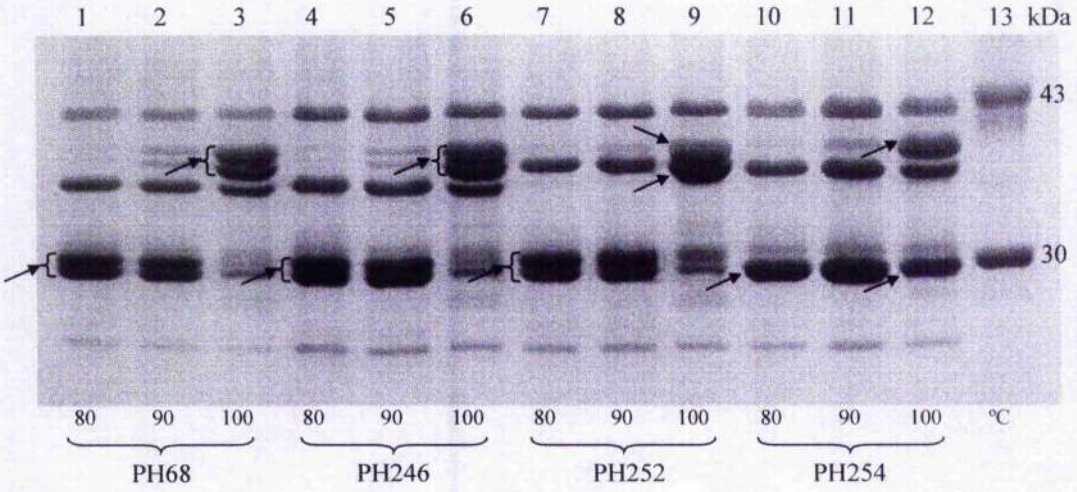


Figure 3.34 Coomassie blue-stained SDS-PAGE OMP profiles of *P. trehalosi* isolates. The isolates include PH68 (lanes 1 to 3), PH246 (lanes 4 to 6), PH252 (lanes 7 to 9), and PH254 (lanes 10 to 12) after heating at 80 °C (lanes 1, 4, 7, and 10), 90 °C (lanes 2, 5, 8, and 11), and 100 °C (lanes 3, 6, 9, and 12). Molecular mass standards (ovalbumin, 43 kDa; carbonic anhydrase, 30 kDa) are shown in lane 13. Only the relevant part of the gel is shown. Arrows indicate the low- and high-molecular-mass forms of the two OmpA proteins in each isolate. The transition is clearly seen between 90 and 100 °C. Only one band is present in isolate PH254 at 80 and 90 °C, whereas two bands occur at 100 °C (see the text for further explanation).

3.1.1.5.2 Transferrin binding proteins (*tbpB* and *tbpA*)

The *tbpB* and *tbpA* genes, encoding the transferrin binding proteins TbpB and TbpA, were selected because of the specificity of the transferrin receptor for ruminant transferrin and their utility as vaccine antigens (see section 1.3.2.4.2). The partial sequences of *tbpB* (1677 nucleotides [95 %]) and *tbpA* (2748 nucleotide [98 %]) were determined in 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates (Table 3.23). Nucleotide sequence analysis of the *tbpB* and *tbpA* genes identified 28 unique *tbpB* sequences, 20 unique *tbpA* sequences, and 29 unique continuous sequences for the *tbpB* and *tbpA* genes. These unique sequences were designated as individual alleles (Table 3.23).

Transferrin binding protein B (TbpB). The 28 *tbpB* alleles could be classified into six subclasses, *tbpB1* to *tbpB6*, based on their overall sequence similarities (Table 3.23). Of these subclasses, *tbpB1*, *tbpB3*, *tbpB4*, and *tbpB6* occur in *M. haemolytica*, *tbpB2* is present in both *M. haemolytica* and *M. glucosida*, and *tbpB5* is associated with *P. trehalosi*.

The polymorphic nucleotide sequences within the *tbpB1*- to *tbpB3*-type alleles of *M. haemolytica* and *M. glucosida* are shown in Figure 3.35. *M. haemolytica tbpB4*- and *tbpB6*-type alleles and *P. trehalosi tbpB5*-type alleles were excluded because they are highly divergent from the *tbpB1*- to *tbpB3*-type alleles of *M. haemolytica* and *M. glucosida* (see below) and there was no visible evidence of intragenic recombination. Visual inspection of the distribution of polymorphic nucleotide sites among the *tbpB* alleles revealed that intragenic recombinational exchanges have occurred within *tbpB1*-, *tbpB2*-, and *tbpB3*-type alleles because runs of nucleotides representing recombinant segments were present (Smith, 1999). The aligned *tbpB1*-, *tbpB2*-, and

Table 3.23 Distribution of *tbpB* and *tbpA* alleles among 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Isolate	ET [a]	Capsular serotype	Host species	Alleles [b]			GenBank accession no.
				tbpB	*tbpA*	*tbpBA*	
M. haemolytica							
PH2	1	A1	Bovine	*tbpB1.1*	*tbpA1.1*	*tbpBA1*	AY850230
PH30	1	A1	Bovine	*tbpB1.2*	*tbpA1.1*	*tbpBA2*	AY850231
PH376	1	A6	Bovine	*tbpB1.1*	*tbpA1.1*	*tbpBA1*	-
PH346	1	A12	Ovine	*tbpB2.1*	*tbpA1.5*	*tbpBA3*	AY850232
PH540	2	A1	Bovine	*tbpB1.3*	*tbpA1.1*	*tbpBA4*	AY850233
PH338	3	A9	Ovine	*tbpB2.1*	*tbpA1.5*	*tbpBA3*	-
PH388	4	A7	Ovine	*tbpB2.2*	*tbpA1.5*	*tbpBA5*	AY850234
PH50	5	A5	Ovine	*tbpB2.2*	*tbpA1.5*	*tbpBA5*	-
PH56	5	A8	Ovine	*tbpB1.5*	*tbpA1.3*	*tbpBA6*	AY850235
PH238	5	A9	Ovine	*tbpB1.4*	*tbpA1.2*	*tbpBA7*	AY850236
PH8	6	A1	Ovine	*tbpB2.2*	*tbpA1.4*	*tbpBA8*	AY850237
PH398	7	A1	Ovine	*tbpB2.2*	*tbpA1.4*	*tbpBA8*	-
PH284	8	A6	Ovine	*tbpB2.2*	*tbpA1.4*	*tbpBA8*	-
PH232	9	A6	Ovine	*tbpB2.2*	*tbpA1.4*	*tbpBA8*	-
PH66	10	A14	Ovine	*tbpB1.6*	*tbpA1.5*	*tbpBA9*	AY850238
PH706	11	A16	Ovine	*tbpB1.7*	*tbpA1.7*	*tbpBA10*	AY850239
PH296	12	A7	Ovine	*tbpB2.4*	*tbpA1.6*	*tbpBA11*	AY850240
PH396	13	A7	Ovine	*tbpB2.4*	*tbpA1.6*	*tbpBA11*	-
PH484	14	A7	Ovine	*tbpB2.4*	*tbpA1.6*	*tbpBA11*	-
PH588	15	A13	Ovine	*tbpB2.3*	*tbpA1.4*	*tbpBA12*	AY850241
PH494	16	A2	Ovine	*tbpB6.1*	*tbpA4.1*	*tbpBA13*	AY850242
PH550	17	A2	Bovine	*tbpB4.1*	*tbpA6.1*	*tbpBA14*	AY850243
PH196	18	A2	Bovine	*tbpB3.1*	*tbpA5.1*	*tbpBA15*	AY850244
PH786	18	A2	Bovine	*tbpB3.2*	*tbpA5.1*	*tbpBA16*	AY850245
PH526	19	A2	Ovine	*tbpB6.2*	*tbpA4.2*	*tbpBA17*	AY850246
PH598	20	A2	Ovine	*tbpB6.3*	*tbpA4.2*	*tbpBA18*	AY850247
PH202	21	A2	Bovine	*tbpB4.2*	*tbpA6.1*	*tbpBA19*	AY850248
PH470	21	A2	Bovine	*tbpB4.3*	*tbpA6.1*	*tbpBA20*	AY850249
PH278	21	A2	Ovine	*tbpB1.8*	*tbpA1.8*	*tbpBA21*	AY850250
PH372	21	A2	Ovine	*tbpB1.8*	*tbpA1.8*	*tbpBA21*	-
PH292	22	A2	Ovine	*tbpB6.2*	*tbpA4.2*	*tbpBA17*	-
PH392	22	A2	Ovine	*tbpB6.2*	*tbpA4.2*	*tbpBA17*	-
M. glucosida							
PH344	1	A11	Ovine	*tbpB2.9*	*tbpA2.4*	*tbpBA22*	AY850251
PH498	3	A11	Ovine	*tbpB2.9*	*tbpA2.4*	*tbpBA22*	-
PH240	5	A11	Ovine	*tbpB2.6*	*tbpA2.1*	*tbpBA23*	AY850252
PH496	7	UG3	Ovine	*tbpB2.5*	*tbpA2.3*	*tbpBA24*	AY850253
PH574	10	UG3	Ovine	*tbpB2.7*	*tbpA2.5*	*tbpBA25*	AY850254
PH290	16	UG3	Ovine	*tbpB2.8*	*tbpA2.2*	*tbpBA26*	AY850255
P. trehalosi							
PH246	2	T4	Ovine	*tbpB5.1*	*tbpA3.2*	*tbpBA27*	AY850256
PH252	4	T10	Ovine	*tbpB5.2*	*tbpA3.1*	*tbpBA28*	AY850257
PH254	15	T15	Ovine	*tbpB5.3*	*tbpA3.3*	*tbpBA29*	AY850258
PH68	19	T3	Ovine	*tbpB5.1*	*tbpA3.2*	*tbpBA27*	-

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

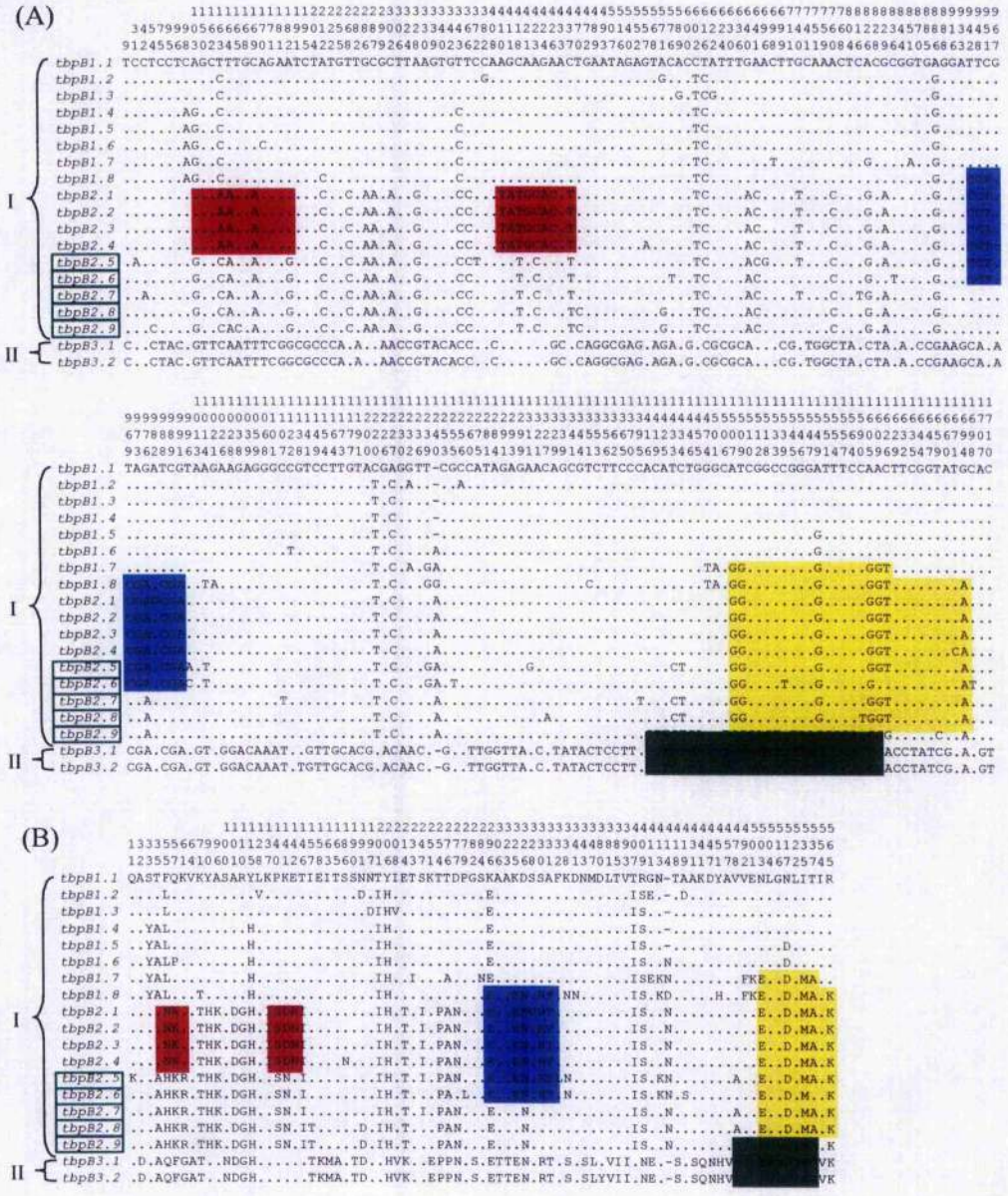


Figure 3.35 Distribution of polymorphic nucleotide (A) and amino acid (B) sites among the *tbpB* alleles of *M. haemolytica* and *M. glucosida* isolates of lineages I and II. Allele designations are shown to the left of each sequence. The Roman numerals correspond to the major lineages of Figure 3.36. The *M. glucosida* alleles are highlighted within the green box. The coloured shading indicates nucleotide sequence identity and putative recombinant segments. The vertical numbers above the sequences represent the positions of polymorphic nucleotide or amino acid sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *tbpB1.1*).

tbpB3-type allele sequences (Figure 3.35A) clearly show visible evidence of mosaicism. There was no evidence of intragenic recombination within *M. haemolytica* alleles *tbpB1.1* to *tbpB1.6*, but recombinant segments were present in *tbpB1.7* (yellow, nucleotides 1507 to 1626) and *tbpB1.8* (blue, nucleotides 942 to 991 and yellow, nucleotides 1507 to 1694). The *M. haemolytica tbpB2.1* to *tbpB2.4* alleles and the *M. glucosida tbpB2.5* to *tbpB2.9* alleles had similar overall sequences but also possessed small putative recombinant segments (red, nucleotides 103 to 182 and 410 to 426; blue, nucleotides 942 to 991; and green, nucleotides 1416 to 1609). In particular, the *M. glucosida tbpB2.9* allele contains a recombinant segment (green, nucleotides 1416 to 1609) that is very similar to the corresponding region of the *M. haemolytica tbpB3*-type alleles.

Amino acid sequence analyses of the *tbpB1*- to *tbpB3*- type alleles of *M. haemolytica* and *M. glucosida* (Figure 3.35B) indicated that the majority of nucleotide substitutions that occur in the recombinant segments result in amino acid changes. The relatively low d_N/d_N ratio for *M. haemolytica* (4.53) and *M. glucosida* (6.04) *tbpB* alleles indicates that constraint against amino acid replacement is, overall, relaxed in the TbpB protein (Table 3.3).

The distribution of the *tbpB* alleles among *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates indicates that, in general, the same subclass, i.e., *tbpB1* to *tbpB6*, is associated with isolates representing closely related ETs. However, certain subclasses are associated with distantly related ETs (Table 3.23) suggesting that assortative (entire gene) recombinational exchanges between unrelated isolates have frequently occurred in the evolution of the *tbpB* gene. For example, subclass *tbpB1* is associated with *M. haemolytica* isolates of ETs 1, 2, 5, 10, 11, and 21; subclass *tbpB2* is associated with *M. haemolytica* isolates of ETs 1, 3 to 9, and 12 to 15 but

Table 3.24 Nucleotide and amino acid differences between representative pairs of the six *thpB* allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Allele	Pairwise differences in nucleotide and amino acid sequences (%) [a]					
	tbpB1.1	***tbpB2.1***	***tbpB3.1***	***tbpB4.1***	***tbpB5.1***	***tbpB6.1***
tbpB1.1		33 (5.9)	56 (10.0)	92 (16.5)	115 (20.6)	233 (41.8)
tbpB2.1	47 (2.8)		50 (9.0)	88 (15.8)	118 (21.1)	236 (42.3)
tbpB3.1	156 (9.3)	147 (8.8)		72 (12.9)	109 (19.5)	232 (41.6)
tbpB4.1	253 (15.1)	250 (14.9)	222 (13.2)		119 (21.3)	233 (41.8)
tbpB5.1	394 (23.5)	407 (24.3)	392 (23.4)	404 (24.1)		231 (41.4)
tbpB6.1	629 (37.5)	629 (37.5)	607 (36.2)	604 (36.0)	628 (37.4)	

[a] Values in the lower left represent pairwise differences in nucleotide sequences (number of polymorphic nucleotide sites and percentage of nucleotide variation); values in the upper right represent pairwise differences in inferred amino acid sequences (number of polymorphic amino acid sites and percentage of amino acid variation).

also with *M. glucosida* isolates; subclass *tbpB4* is associated with *M. haemolytica* isolates of ETs 17 and 21; and *tbpB6* is associated with *M. haemolytica* isolates of ETs 16, 19, 20, and 22. Subclasses such as *tbpB3* (bovine A2 isolates of ET 18), *tbpB4* (bovine A2 isolates of ETs 17 and 21) and *tbpB6* (ovine A2 isolates of ETs 19, 20, and 22) contain highly divergent host-specific sequences (Table 3.23) that have possibly been acquired from external sources by assortative (entire gene) recombination.

Pairwise differences in nucleotide and amino acid sequences between representative pairs of the six allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* ranged from 47 to 629 (2.8 to 37.5 %) nucleotide sites and from 33 to 233 (5.9 to 41.8 %) amino acid positions (Table 3.24).

The *tbpB* tree topology (Figure 3.36) indicates that the *M. haemolytica tbpB1*- and *tbpB2*-type alleles form two clusters, A and B, in lineage I. However, cluster B also contains the *M. glucosida* alleles *tbpB2.5* to *tbpB2.9*. The *M. haemolytica tbpB3*- and *tbpB4*-type alleles represent lineages II and III, respectively, and are more distantly related to the *M. haemolytica tbpB1* and *tbpB2*-type alleles of lineage I than are the *M. glucosida tbpB2*-type alleles. The *P. trehalosi tbpB5*-type alleles represent lineage IV but the *M. haemolytica tbpB6*-type alleles represent lineage V which, surprisingly, is more divergent than the *P. trehalosi* alleles. This is clearly very different from the tree topologies based on 16S rRNA sequence analysis and MLEE (Davies *et al.*, 1996; 1997a) which show that *M. glucosida* has diverged from *M. haemolytica* and that *P. trehalosi* has further diverged from both species. The data provide strong evidence that horizontal DNA transfer and recombination have played important roles in the evolution of *tbpB*.

Alignment of the N-terminal (amino acid sites 1 to 339) and C-terminal (amino acid

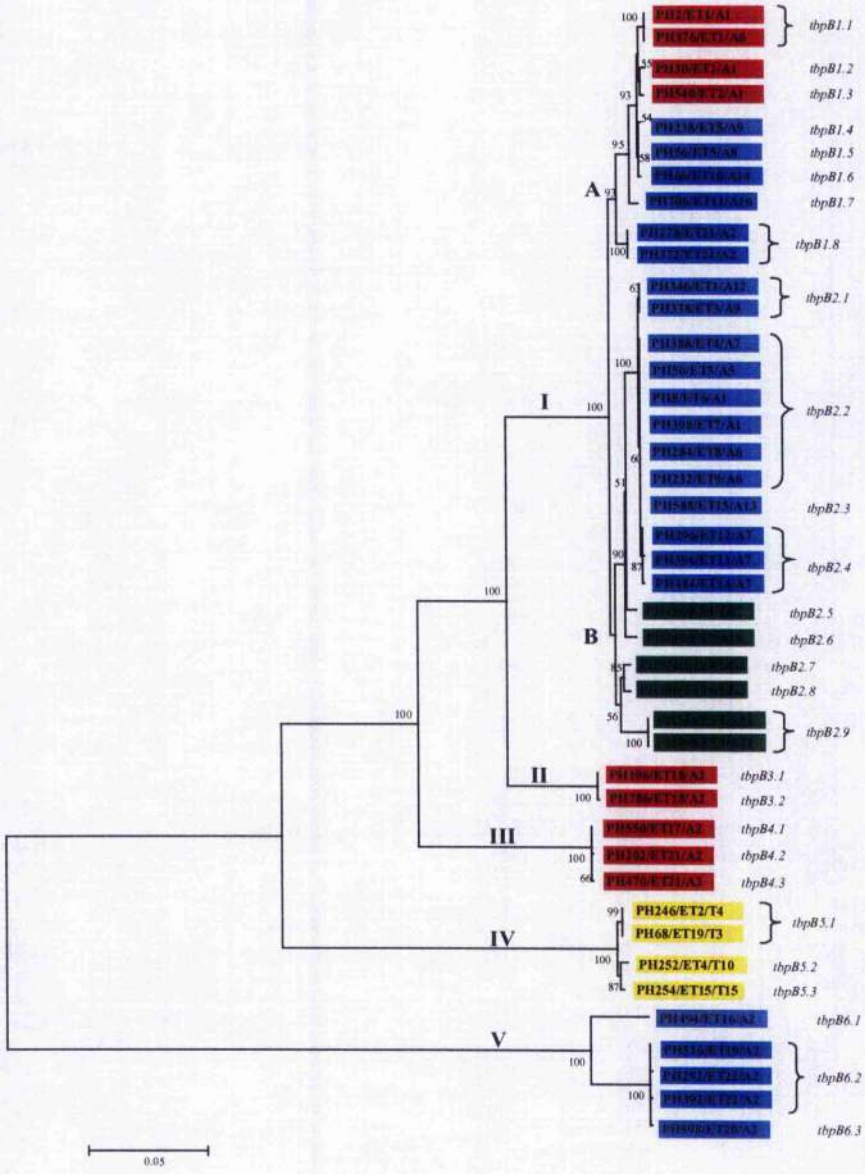


Figure 3.36 Minimum evolution (ME) tree for the *tbpB* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

sites 340 to 678) regions of the inferred amino acid sequences of six representative TbpB proteins of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* revealed that several identical or similar amino acid sequences were present in the N- and C-terminal halves of the same protein (Figure 3.37).

Recently, two transferrin binding sites were identified in the N- terminal half of TbpB in meningococcal isolate B16B6 by a sequence-based prediction method based on hydrophobicity/hydrophilicity profiles and confirmed by site-directed mutagenesis (Renauld-Mongenie *et al.*, 2004). The corresponding transferrin binding sites were located in the conserved regions of the inferred amino acid sequences of six TbpB proteins of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Figure 3.37). In particular, there was no variation in the two amino acid sites corresponding to the sites used for site-directed mutagenesis in *N. meningitidis* TbpB (i.e. R in site 1 and Y in site 2)

Transferrin binding protein A (*tbpA*) The twenty *tbpA* alleles could also be classified into six subclasses, *tbpA1* to *tbpA6*, based on their overall sequence similarities (Table 3.23). Of these subclasses, *tbpA1*, *tbpA4*, *tbp5*, and *tbpA6* occur in *M. haemolytica*, *tbpA2* is present in *M. glucosida*, and *tbpA3* is associated with *P. trehalosi* (Table 3.23).

The polymorphic nucleotide sequences of the alleles representing subclasses *tbpA1* to *tbpA4* of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*, with respect to *tbpA1.1*, are shown in Figure 3.38. The *M. haemolytica* subclasses *tbpA5* and *thpB6* were excluded because they are highly divergent from the subclasses *tbpA1* to *tbpA4* (see below) and there was no visible evidence of intragenic recombination. Visual *tbpA2*-, *tbpB3*-, and *tbpA4*-type alleles because runs of nucleotides representing

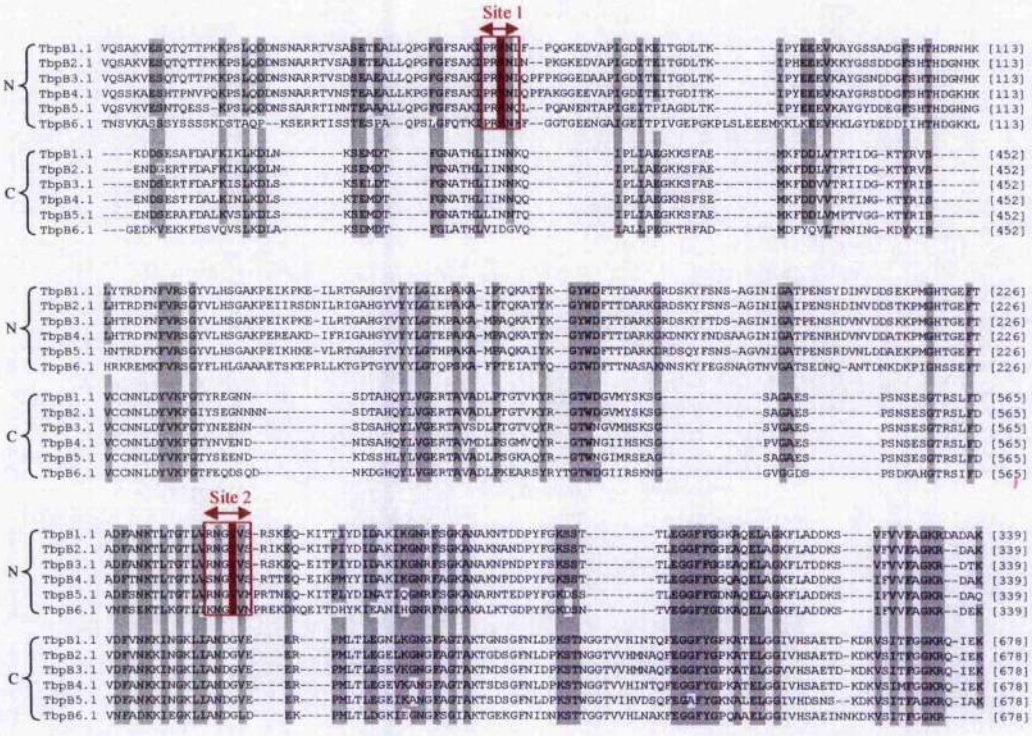


Figure 3.37 Comparison of the inferred amino acid sequences corresponding to the N-terminal regions (N) and the C-terminal regions (C) of the six TbpB proteins of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Numbers (left) represent the amino acid positions from the 5' end of the protein. The shaded regions indicate the positions where amino acid sites are identical or similar. Similar residues are defined as Y, F and W; V, L, I and M; K and R; D and E; S and T; P and A (Renauld-Mongenie *et al.*, 2004). Red boxes in the N-terminal region indicate putative transferrin binding sites similar to the corresponding sites of the *N. meningitidis* TbpB protein (Renauld-Mongenie *et al.*, 2004). Red lines indicate amino acid sites used for site-directed mutagenesis of the *N. meningitidis* TbpB protein.

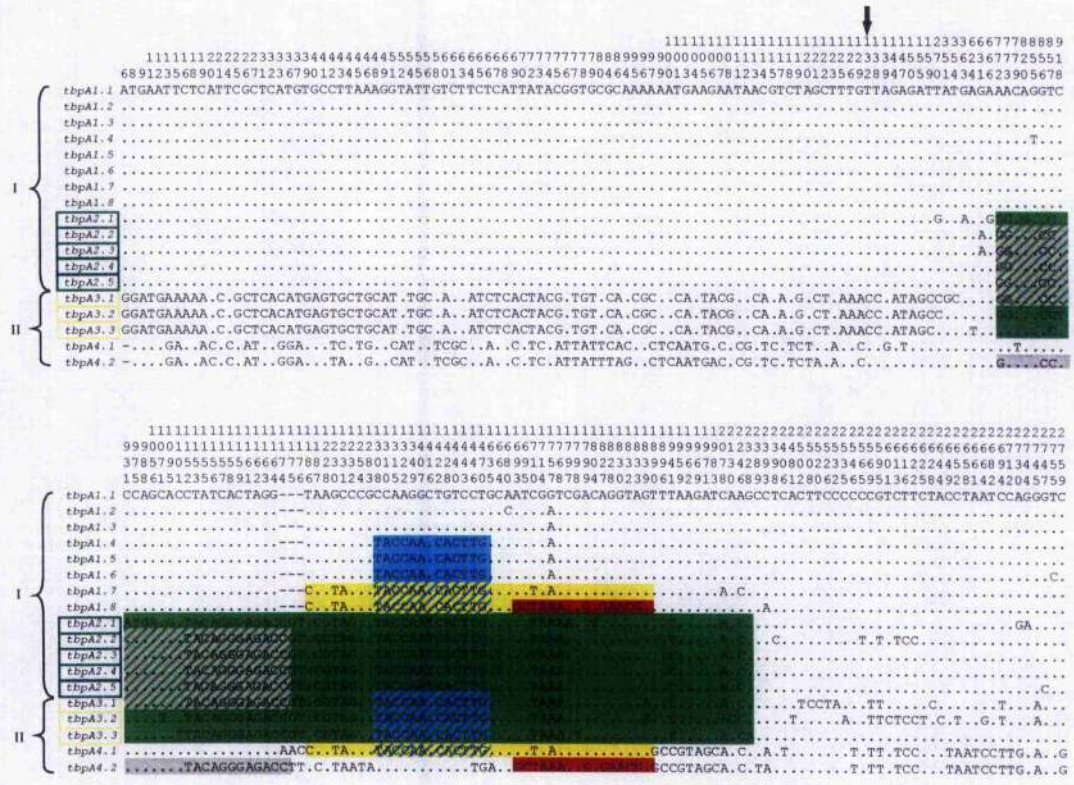


Figure 3.38 Distribution of polymorphic nucleotide sites among the *tbpA* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. Allele designations are shown to the left of each sequence. The Roman numerals correspond to the major lineages of Figure 3.39. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. The black vertical arrow represents the mature amino acid terminus of TbpA identified in *M. haemolytica* h196 (U73302). The coloured shading indicates nucleotide sequence identity and putative recombinant segments. The vertical numbers above the sequences represent the positions of polymorphic nucleotide sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides match those of the first (topmost) sequence (i.e. *tbpA1.1*).

recombinant segments were present (Smith, 1999). The *M. haemolytica tbpA1.1* to *tbpA1.3* alleles are conserved, but the *M. haemolytica tbpA1.4* to *tbpA1.6* alleles contain a recombinant segment (blue, nucleotides 1308 to 1470) that is identical to the corresponding region in various other isolates of *M. haemolytica* and *P. trehalosi*. The *M. haemolytica tbpA1.7* allele contains a recombinant segment (yellow; nucleotides 1186 to 1839) that is identical to the corresponding region of allele *tbpA4.1*. The *M. haemolytica* allele *tbpA1.8* contains a recombinant segment (yellow, nucleotides 1186 and 1680) that is identical to the corresponding region of alleles *tbpA1.7* and *tbpA4.1*, but also a short recombinant segment (red, nucleotides 1693 to 1839) that is identical to the corresponding region of allele *tbpA*4.2. The *M. glucosida tbpA2*-type alleles contain a recombinant segment (green, nucleotides 772 to 2246) that is almost identical to the corresponding region of the *P. trehalosi tbpA3*-type alleles. Finally, the *M. haemolytica tbpA4.2* allele contains a recombinant segment (grey, nucleotides 772 to 1174) that is almost identical to the corresponding region of the *tbpA2*- and *tbpA3*-type alleles of *M. glucosida* and *P. trehalosi*, respectively.

The distribution of the *tbpA* alleles among *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Table 3.23) indicates that *tbpA1*-type alleles are associated with *M. haemolytica* isolates of ETs 1 to 15 whereas *tbpA4*-, *tbpA5*-, and *tbpA6*-type alleles are associated with *M. haemolytica* isolates of ETs 16 to 22. The *tbpA5.1* and *tbpA6.1* alleles are associated with bovine *M. haemolytica* isolates of ETs 18, and 17 and 21, respectively, and these differ at 10.2 % and 19 % of nucleotide sites from the *M. haemolytica tbpA1.1* allele, respectively (Table 3.25). The isolates also have distinctive *tbpB* genes that differ at 9.3 % (*tbpB3.1*) and 15.1 % (*tbpB4.1*) of nucleotide sites from *tbpB1.1*, respectively (Table 3.24). The *tbpA3.1* and *tbpA4.1* alleles of *P. trehalosi* and ovine *M. haemolytica* A2 isolates differ at 4.7 % and 3.6 %

Table 3.25 Nucleotide and amino acid differences between representative pairs of the six *tbpA* allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Allele	Pairwise differences in nucleotide and amino acid sequences (%) [a]					
	tbpA1.1	***tbpA2.1***	***tbpA3.1***	***tbpA4.1***	***tbpA5.1***	***tbpA6.1***
tbpA1.1		23 (2.5)	53 (5.8)	37 (4.0)	64 (7.0)	123 (13.4)
tbpA2.1	52 (1.9)		41 (4.5)	44 (4.8)	66 (7.2)	122 (13.3)
tbpA3.1	128 (4.7)	103 (3.7)		51 (5.6)	95 (10.4)	142 (15.5)
tbpA4.1	100 (3.6)	115 (4.2)	134 (4.9)		88 (9.6)	137 (15.0)
tbpA5.1	281 (10.2)	280 (10.2)	338 (12.3)	329 (12.0)		108 (11.8)
tbpA6.1	521 (19.0)	520 (18.9)	564 (20.5)	570 (20.7)	491 (17.9)	

[a] Values in the lower left represent pairwise differences in nucleotide sequences (number of polymorphic nucleotide sites and percentage of nucleotide variation); values in the upper right represent pairwise differences in inferred amino acid sequences (number of polymorphic amino acid sites and percentage of amino acid variation).

of nucleotide sites from *tbpA1.1*, respectively. However, these isolates contain *tbpB5.1* and *tbpB6.1* alleles that differ at 23.5 % and 37.5 % of nucleotide sites from *tbpB1.1*, respectively. These different levels of diversity between *tbpB* and *tbpA* of *P. trehalosi* and *M. haemolytica* isolates of ETs 19, 20, and 22, with respect to other *M. haemolytica* isolates, indicates they are derived from different external sources.

Pairwise differences in nucleotide and amino acid sequences between representative pairs of the six allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Table 3.25) ranged from 52 to 570 (1.9 to 20.7 %) nucleotide sites and from 23 to 142 (2.5 to 15.5 %) amino acid positions.

The *tbpA* tree topology (Figure 3.39) indicates that the *M. glucosida tbpA2*-type alleles (cluster B) are closely related to the *M. haemolytica tbpA1*-type alleles (cluster A) and together comprise lineage I. The *P. trehalosi tbpA3*-type alleles and the *M. haemolytica tbpA4*-type alleles together represent the diverse lineage II. The *M. haemolytica* alleles *tbpA5.1* and *tbpA6.1* represent two lineages, III and IV, that are both highly divergent from the *tbpA* alleles of other *M. haemolytica* isolates, as well as from *M. glucosida* and *P. trehalosi* isolates. The *tbpA* topology is clearly different from those based on 16S rRNA sequence analysis and MLEE (Davies *et al.*, 1996; 1997a) which show that *M. glucosida* has diverged from *M. haemolytica* and that *P. trehalosi* has further diverged from both species. The data provide strong evidence that horizontal DNA transfer and recombination have played important roles in the evolution of the *tbpA* gene.

The locations of the tonB box, 21 β-strands and eleven surface-exposed loop regions (L1 to L11) were identified in the inferred amino acid sequences of the *tbpA* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Figure 3.40) based on alignment with

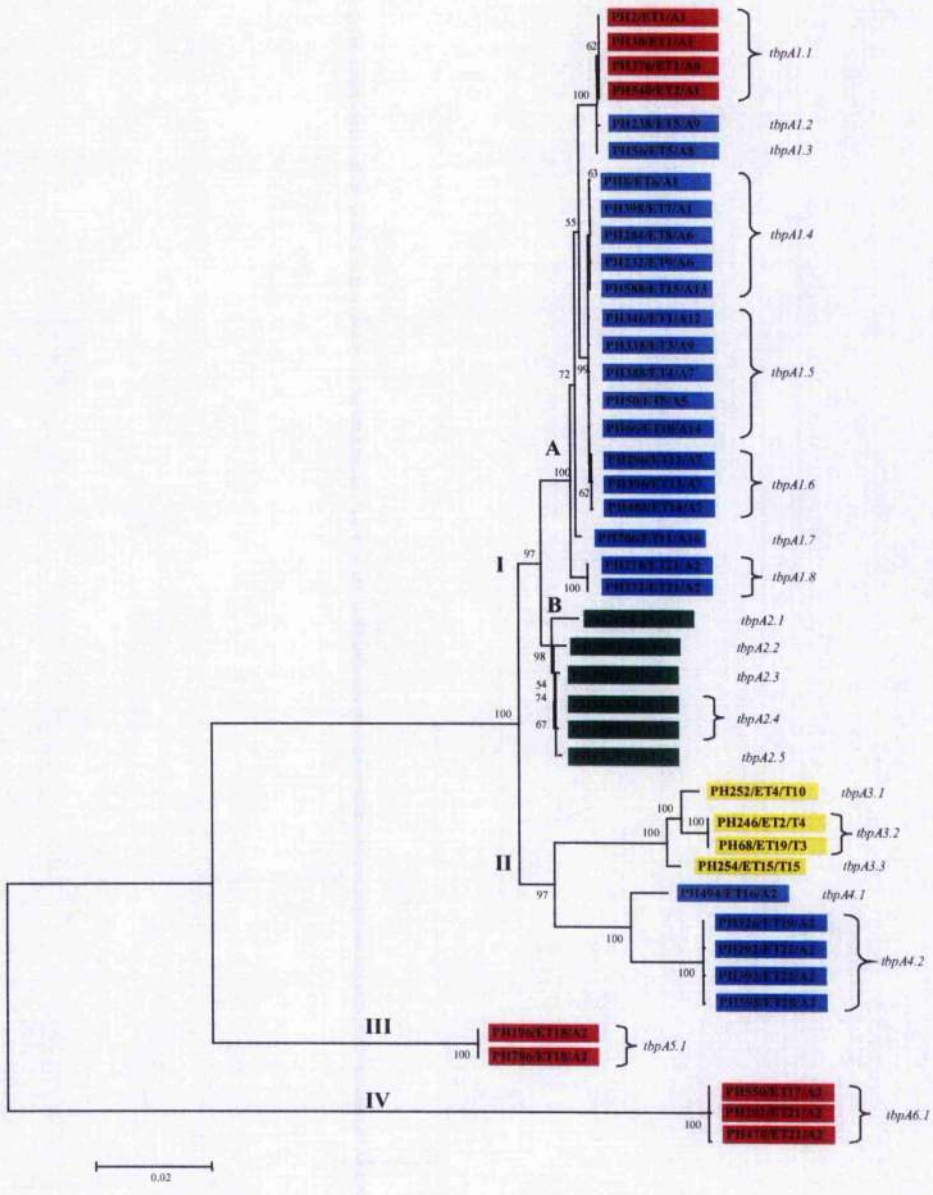
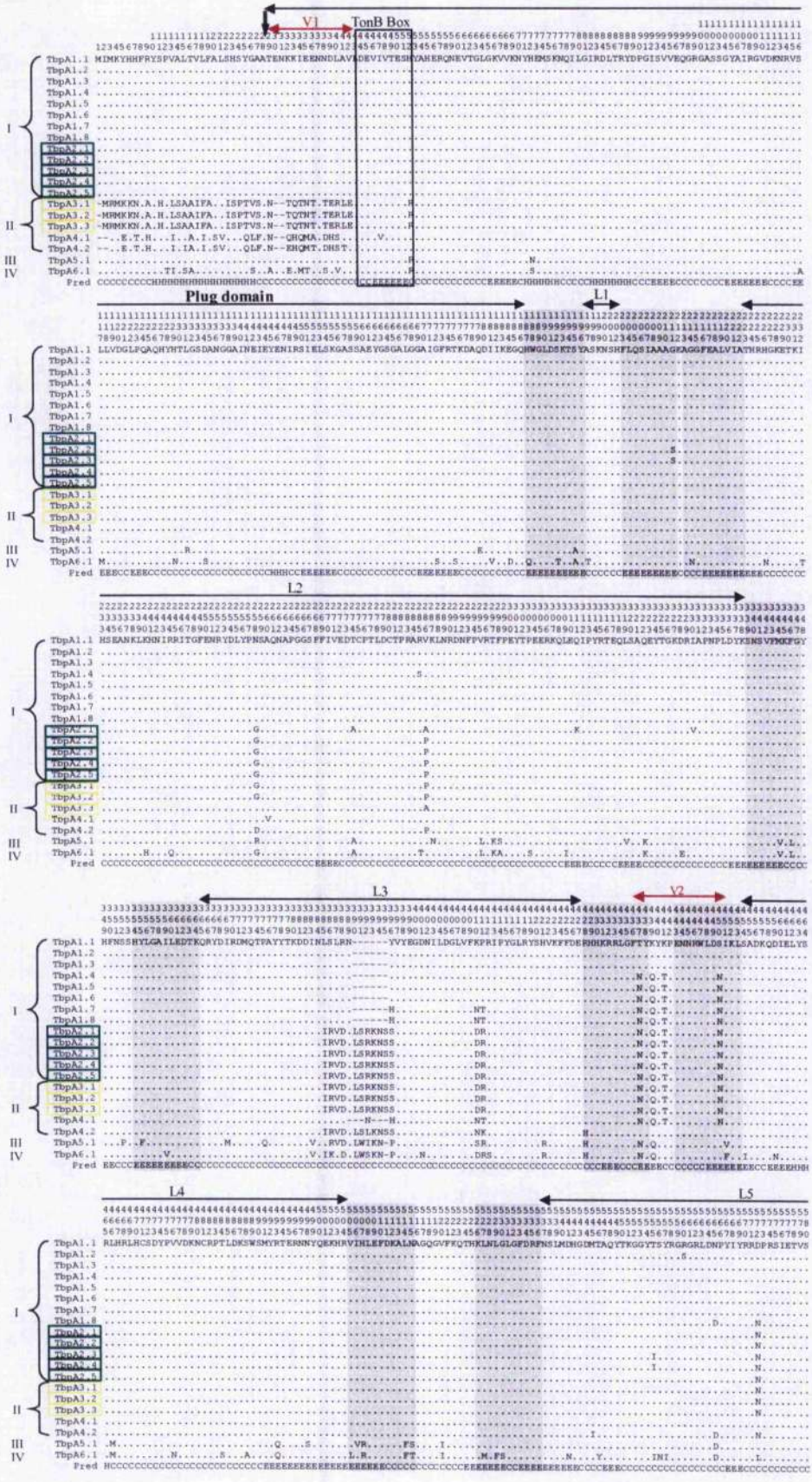


Figure 3.39 Minimum evolution (ME) tree for the *tbpA* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of >50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).


V1
TonB Box
Plug domain
L1
L2
L3
V2
L4
L5
I
II
III
IV

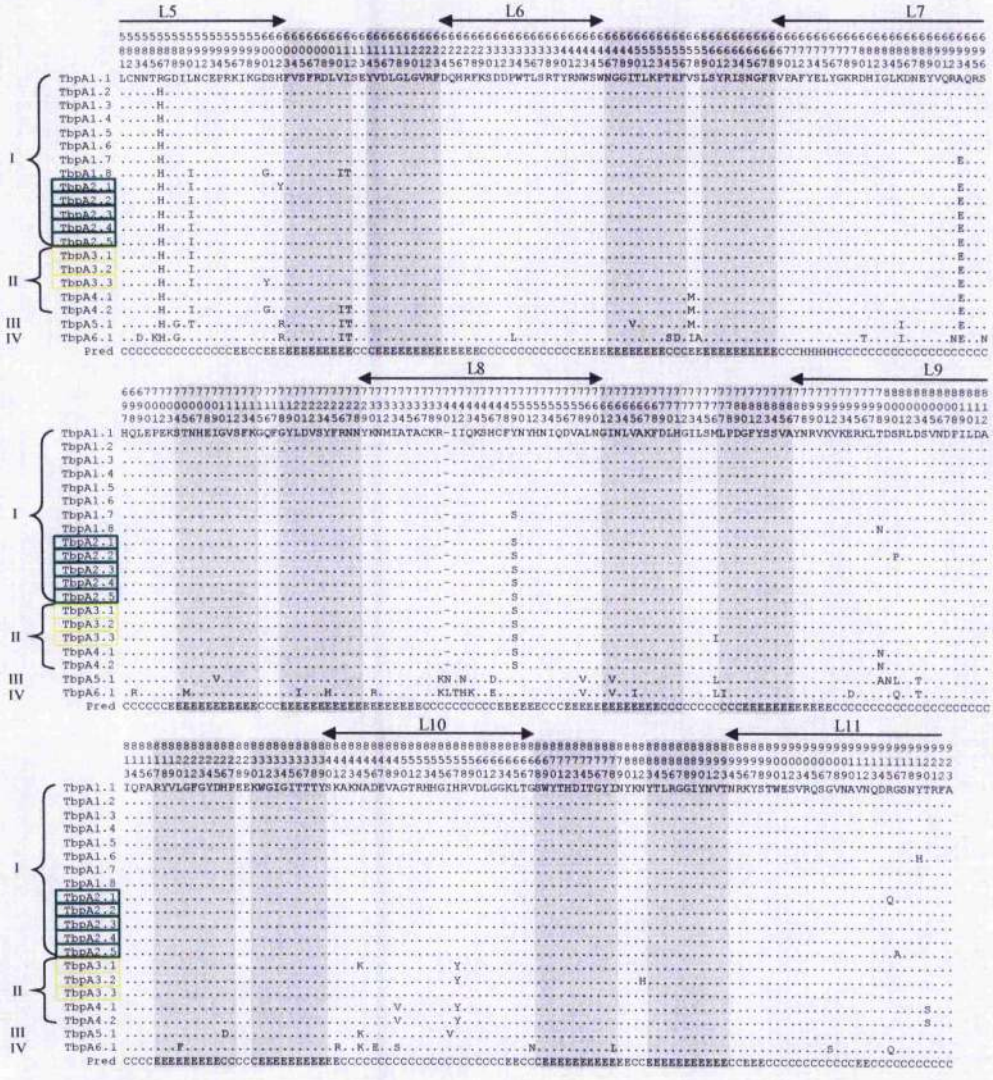


Figure 3.40 Amino acid sequence alignments of the *tbpA* alleles of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The Roman numerals I to IV represent the major allele classes and lineages of Figure 3.38. The green boxes indicate *M. glucosida* alleles and the yellow boxes indicate *P. trehalosi* alleles. The vertical numbers above the sequences represent the positions of amino acid sites from the 5′end of the gene. The dots represent sites where the amino acids match those of the first (topmost) sequence (i.e. *tbpA1.1*). The black vertical arrow represents the mature amino acid terminus of TbpA identified in *M. haemolytica* h196 (U73302). The white box represents the conserved TonB box. Gaps are indicated by dashes. L1 to L11 represent the surface-exposed loops 1 to 11. V1 and V2 represent variable domains that are not located within the loop regions. The shaded regions represent predicted membrane-spanning β-strand structures.

Table 3.26 Sequence diversity and substitution rates for the 11 loop regions L1 to L11, the variable domains, V1 and V2, and the conserved regions of the *thpA* genes of 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

Domain	Sequence diversity (%)		d_S [a]	d_N [a]	d_S/d_N
	Nucleotide	Amino acid			
L1	5.6	16.7	0.00 ± 0.00	0.93 ± 1.06	-
L2	24.7	17.2	22.34 ± 5.50	2.02 ± 0.61	11.1
L3	27.0	23.8	21.08 ± 6.30	6.09 ± 2.17	3.5
L4	23.5	15.7	34.83 ± 7.27	1.31 ± 0.51	26.6
L5	25.3	24.2	18.32 ± 5.96	2.88 ± 0.79	6.4
L6	10.6	4.5	8.32 ± 4.90	0.27 ± 0.27	30.8
L7	22.9	17.1	14.17 ± 5.31	2.40 ± 0.94	5.9
L8	30.2	28.1	21.37 ± 8.44	3.42 ± 1.32	6.2
L9	27.8	16.7	23.85 ± 8.82	2.44 ± 1.17	9.8
L10	29.8	25.0	21.85 ± 7.22	2.60 ± 0.96	8.4
L11	21.1	16.7	18.00 ± 5.26	1.06 ± 0.49	17.0
V1	73.8	78.6	52.92 ± 14.98	37.37 ± 9.14	1.4
V2	28.9	33.3	7.73 ± 9.81	5.63 ± 2.39	1.4
Conserved regions	23.3	13.4	22.76 ± 2.50	1.26 ± 0.18	18.1

[a] d_S is the number of synonymous substitution per 100 synonymous sites; d_N is the number of nonsynonymous substitution per 100 nonsynonymous sites. Values are means ± standard deviations.

the hypothetical topology model of gonococcal TbpA (Boulton *et al.*, 2000), and secondary structure prediction (Jones, 1999; McGuffin *et al.*, 2000). With the exception of the variable regions V1 and V2, polymorphic amino acid sites occurred more frequently in the surface-exposed loop regions than in the N-terminal plug region or β-strand regions. The d_S/d_N ratios were calculated for the loop regions, L1 to L11, the two variable regions, V1 and V2, and the remaining conserved regions of the sequences (Table 3.26). The d_S/d_N ratio for L3 (3.5) and for the variable regions V1 (1.4) and V2 (1.4) were relatively low whereas those for L2 (11.1), L4 (26.6), L6 (30.8), and L11 (17.0), together with the remaining conserved regions (18.1) were relatively high, suggesting that different regions of the TbpA protein are subject to different degrees of evolutionary constraint.

Nucleotide variation and mosaic structure within the *tbpB* and *tbpA* genes. The *tbpB* and *tbpA* genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates were classified into 13 groups based on sequence similarity (Table 3.27). Group 1 is associated with both bovine and ovine *M. haemolytica* isolates, whereas groups 2 to 5, 12, and 13 are associated exclusively with ovine *M. haemolytica* isolates and groups 9 and 10 are associated exclusively with bovine *M. haemolytica* isolates. The *tbpB* and *tbpA* genes of ovine *M. haemolytica* isolates PH278 and PH372 of ET 21 (group 4) are very different from those of closely related bovine *M. haemolytica* isolates PH202 and PH470 of ET 21 (group 10) and ovine *M. haemolytica* isolates PH292 and PH392 of ET 22 (group 13). The distribution of polymorphic nucleotide sites within the *tbpB* and *tbpA* genes of 13 representative isolates of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* is shown in Figure 3.41. The mosaic structure of the *tbpB* and *tbpA* genes in these isolates is shown schematically in Figure 3.42. The data show that intragenic recombination has been frequent in the *tbpB* and *tbpA* genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. *M. haemolytica* isolates PH196 and

Table 3.27 Thirteen groups of *tbpB* and *tbpA* genes of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* based on sequence similarity

Group	*tbpB* and *tbpA* alleles	ET [a]	Serotype	Host species	Strain
Group 1	*tbpB1.1-1.5 & tbpA1.1-1.3*	ETs 1, 2/ET 5	A1, A6/A8, A9	Bovine/Ovine	PH2, PH30, PH376, PH540, PH56, PH238
Group 2	*tbpB1.6 & tbpA1.5*	ET 10	A14	Ovine	PH66
Group 3	*tbpB1.7 & tbpA1.7*	ET 11	A16	Ovine	PH706
Group 4	*tbpB1.8 & tbpA1.8*	ET 21	A2	Ovine	PH278, PH372
Group 5	*tbpB2.1-2.4 & tbpA1.4-1.6*	ETs 1, 3-9, 12-15	A1, A5-7, A9, A12, A13	Ovine	PH346, PH338, PH388, PH50, PH8, PH398, PH284, PH232, PH296, PH396, PH484, PH588
Group 6	*tbpB2.5, 2.6 & tbpA2.1, 2.3*	ETs 5, 7	A11, UG3	Ovine	PH240, PH496
Group 7	*tbpB2.7, 2.8 & tbpA2.2, 2.5*	ETs 10, 16	UG3	Ovine	PH574, PH290
Group 8	*tbpB2.9 & tbpA2.4*	ETs 1, 3	A11	Ovine	PH344, PH498
Group 9	*tbpB3.1, 3.2 & tbpA5.1*	ET18	A2	Bovine	PH196, PH786
Group 10	*tbpB4.1-4.3 & tbpA6.1*	ETs 17, 21	A2	Bovine	PH550, PH202, PH470
Group 11	*tbpB5.1-5.3 & tbpA3.1-3.3*	ETs 2, 4, 15, 19	T3, T4, T10, T15	Ovine	PH246, PH252, PH254, PH68
Group 12	*tbpB6.1 & tbpA4.1*	ET 16	A2	Ovine	PH494
Group 13	*tbpB6.2, 6.3 & tbpA4.2*	ETs 19, 20, 22	A2	Ovine	PH526, PH598, PH292, PH392

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively.

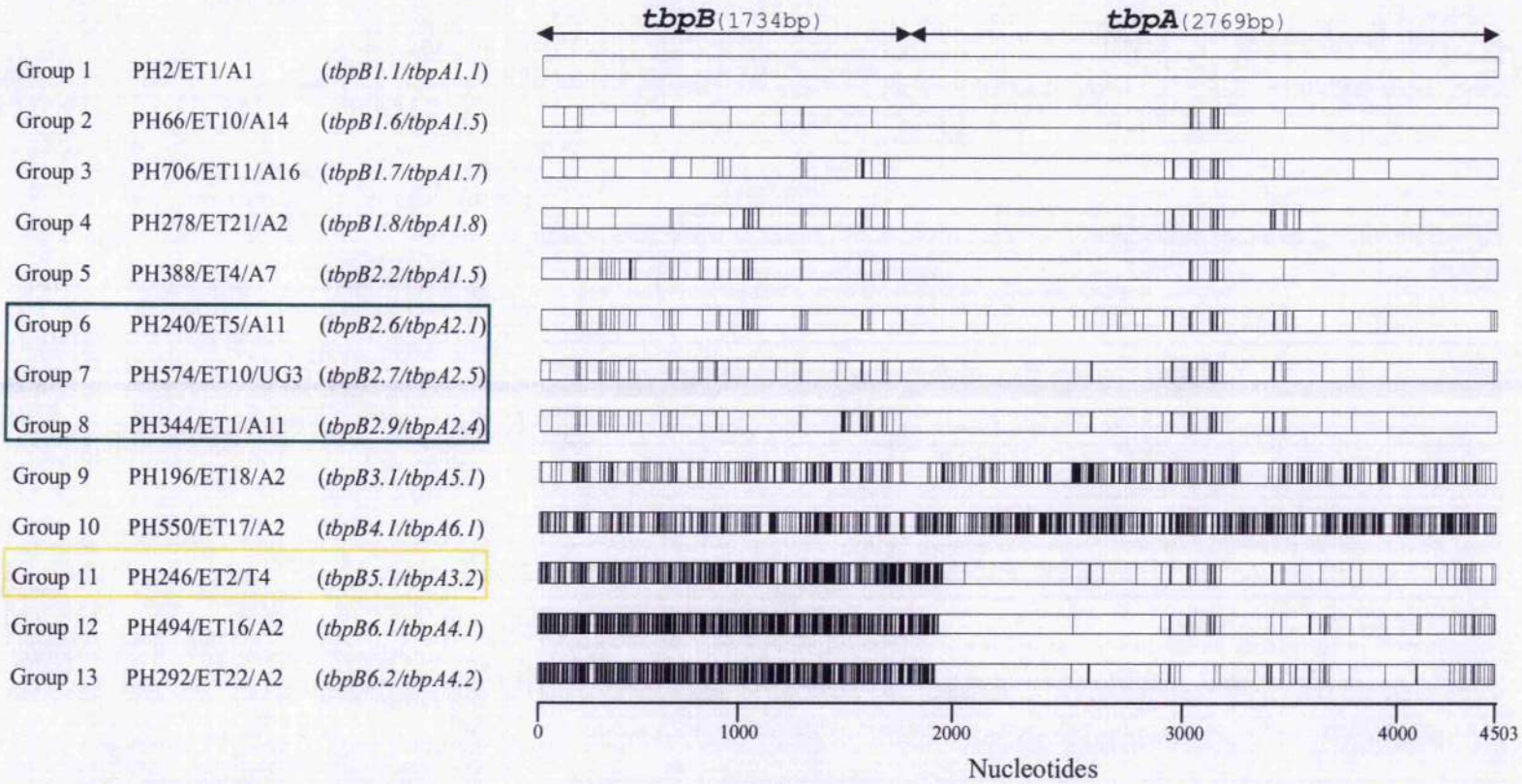


Figure 3.41 Distribution of polymorphic nucleotide sites within the *tbpB* and *tbpA* genes of representative *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b). Vertical lines represent polymorphic nucleotide sites with respect to isolate PH2.

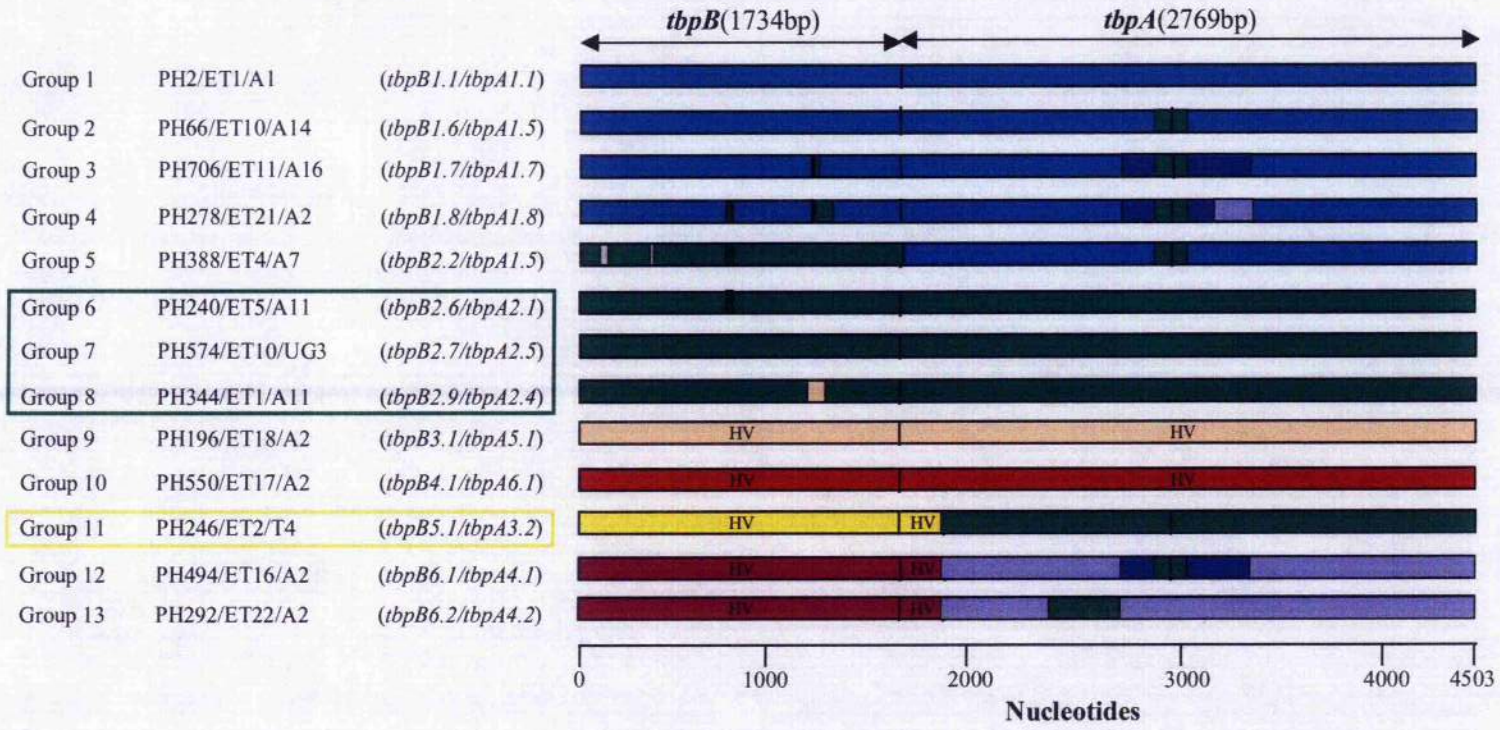


Figure 3.42 Schematic representation of the mosaic structures of the *tbpB* and *tbpA* genes of representative *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates. The *M. glucosida* and *P. trehalosi* alleles are highlighted within the green and yellow boxes, respectively. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b). The coloured shading indicates sequence similarity and the likely origins of recombinant segments.

Table 3.28 Distribution of *wza* alleles among 32 *M. haemolytica*, four *M. glucosida*, and four *P. trehalosi* isolates

Isolate	ET [a]	Capsular serotype	Host species	*wza* [b] allele	GenBank accession no.
M. haemolytica					
PH2	1	A1	Bovine	*wza1.1*	AY847850
PH30	1	A1	Bovine	*wza1.1*	-
PH376	1	A6	Bovine	*wza1.1*	-
PH346	1	A12	Ovine	*wza3.3*	AY847851
PH540	2	A1	Bovine	*wza1.1*	-
PH338	3	A9	Ovine	*wza9.2*	AY847852
PH388	4	A7	Ovine	*wza3.1*	AY847853
PH50	5	A5	Ovine	*wza4.1*	AY847854
PH56	5	A8	Ovine	*wza9.1*	AY847855
PH238	5	A9	Ovine	*wza9.2*	-
PH8	6	A1	Ovine	*wza2.1*	AY847856
PH398	7	A1	Ovine	*wza2.1*	-
PH284	8	A6	Ovine	*wza2.2*	AY847857
PH232	9	A6	Ovine	*wza2.3*	AY847858
PH66	10	A14	Ovine	*wza5.1*	AY847859
PH706	11	A16	Ovine	*wza5.1*	-
PH296	12	A7	Ovine	*wza3.1*	-
PH396	13	A7	Ovine	*wza3.1*	-
PH484	14	A7	Ovine	*wza3.1*	-
PH588	15	A13	Ovine	*wza7.1*	AY847860
PH494	16	A2	Ovine	*wza3.4*	AY847861
PH550	17	A2	Bovine	*wza3.4*	-
PH196	18	A2	Bovine	*wza3.2*	AY847862
PH786	18	A2	Bovine	*wza3.2*	-
PH526	19	A2	Ovine	*wza3.4*	-
PH598	20	A2	Ovine	*wza3.4*	-
PH202	21	A2	Bovine	*wza3.4*	-
PH470	21	A2	Bovine	*wza3.4*	-
PH278	21	A2	Ovine	*wza3.4*	-
PH372	21	A2	Ovine	*wza3.4*	-
PH292	22	A2	Ovine	*wza3.4*	-
PH392	22	A2	Ovine	*wza3.4*	-
M. glucosida					
PH344	1	A11	Ovine	*wza6.1*	AY847863
PH498	3	A11	Ovine	*wza6.1*	-
PH240	5	A11	Ovine	*wza6.2*	AY847864
PH496	7	UG3	Ovine	*wza8.1*	AY847865
P. trehalosi					
PH246	2	T4	Ovine	*wza10.1*	AY847866
PH252	4	T10	Ovine	*wza10.3*	AY847867
PH254	15	T15	Ovine	*wza10.2*	AY847868
PH68	19	T3	Ovine	*wza10.4*	AY847869

[a] Each species has a different MLEE scheme (Davies *et al.*, 1997a; 1997b).

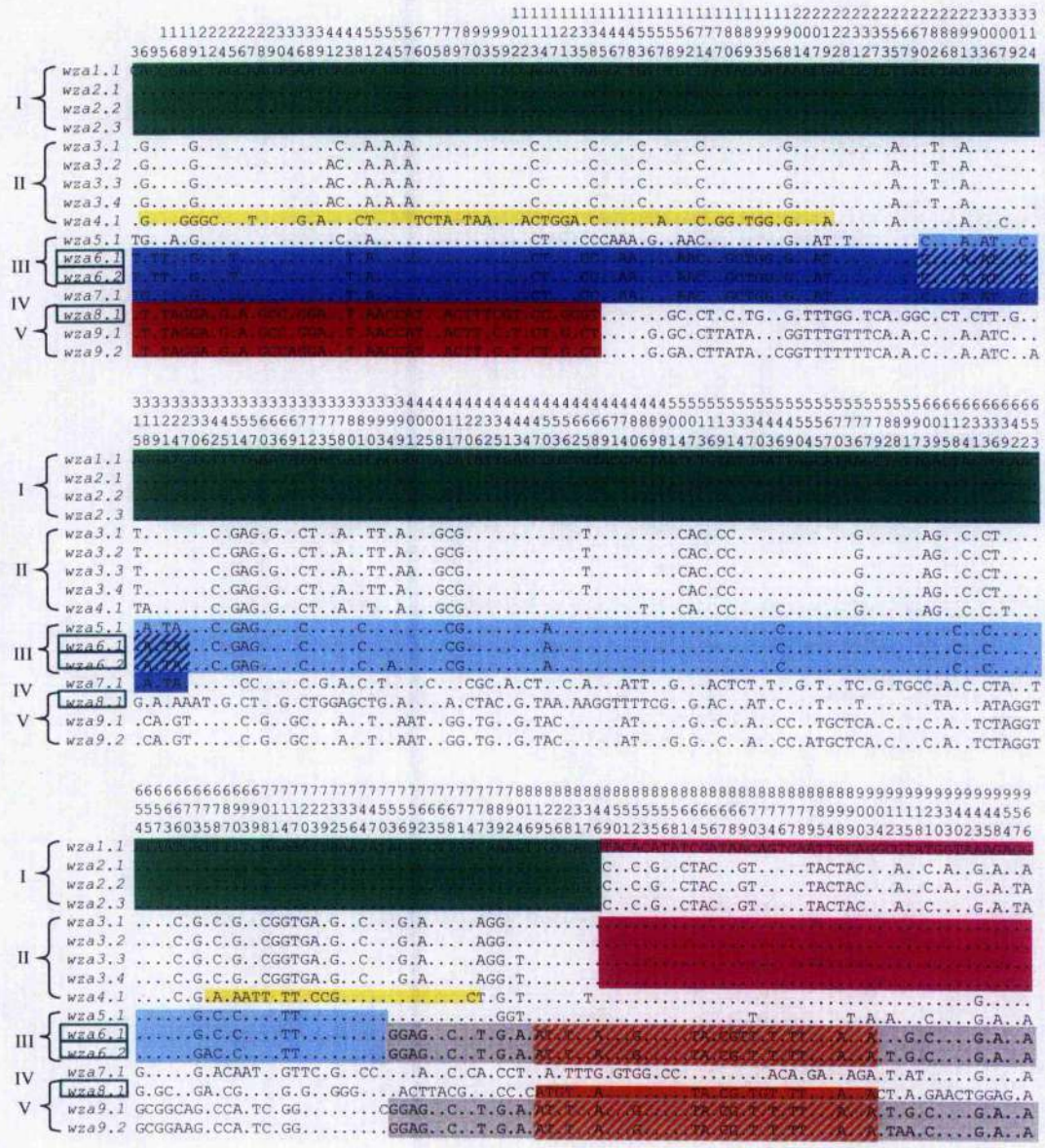


Figure 3.43 Distribution of polymorphic nucleotide sites among the *wza* alleles of *M. haemolytica* and *M. glucosida*. Allele designations are shown to the left of each sequence. The Roman numerals correspond to the major lineages of Figure 3.44. The *M. glucosida* alleles are highlighted within the green boxes. The coloured shading indicates nucleotide sequence identity and putative recombinant segments. The vertical numbers above the sequences represent the positions of polymorphic nucleotide sites from the 5′end of the partial gene sequence. The dots represent sites where the nucleotides or amino acids match those of the first (topmost) sequence (i.e. *wza1.1*).

Table 3.29 Nucleotide and amino acid differences between representative pairs of the 10 *wza* allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Allele	Pairwise differences in nucleotide and amino acid sequences (%) [a]									
	wza1.1	*wza2.1*	*wza3.1*	*wza4.1*	*wza5.1*	*wza6.1*	*wza7.1*	*wza8.1*	*wza9.1*	*wza10.1*
wza1.1		6 (1.9)	5 (1.6)	12 (3.7)	11 (3.4)	14 (4.3)	17 (5.3)	33 (10..2)	28 (8.7)	96 (29.8)
wza2.1	21 (2.2)		11 (3.4)	16 (5.0)	14 (4.3)	14 (4.3)	19 (5.9)	33 (10.2)	28 (8.7)	96 (29.8)
wza3.1	58 (6.0)	79 (8.2)		10 (3.1)	8 (2.5)	11 (3.4)	14 (4.3)	30 (9.3)	25 (7.8)	93 (28.9)
wza4.1	81 (8.4)	100 (10.4)	60 (6.2)		12 (3.7)	15 (4.7)	16 (5.0)	30 (9.3)	25 (7.8)	93 (28.9)
wza5.1	58 (6.0)	68 (7.0)	72 (7.5)	87 (9.0)		12 (3.7)	13 (4.0)	31 (9.6)	26 (8.1)	93 (28.9)
wza6.1	78 (8.1)	82 (8.5)	92 (9.5)	100 (10.4)	41 (4.2)		12 (3.7)	23 (7.1)	20 (6.2)	90 (28.0)
wza7.1	116 (12.0)	126 (13.0)	119 (12.3)	124 (12.8)	99 (10.2)	100 (10.4)		30 (9.3)	29 (9.0)	92 (28.6)
wza8.1	166 (17.2)	171 (17.7)	172 (17.8)	177 (18.3)	166 (17.2)	153 (15.8)	173 (17.9)		15 (4.7)	98 (30.4)
wza9.1	146 (15.1)	151 (15.6)	148 (15.3)	148 (15.3)	133 (13.8)	113 (11.7)	155 (16.0)	109 (11.3)		97 (30.1)
wza10.1	287 (29.7)	292 (30.2)	282 (29.2)	279 (28.9)	289 (29.9)	289 (29.9)	294 (30.4)	297 (30.7)	288 (29.8)	

[a] Values in the lower left represent pairwise differences in nucleotide sequences (number of polymorphic nucleotide sites and percentage of nucleotide variation); values in the upper right represent pairwise differences in inferred amino acid sequences (number of polymorphic amino acid sites and percentage of amino acid variation).

alleles. Since the 5′ region of the *wza* gene is close to the variable capsular biosynthesis region and the *wza3*-type alleles are most common in *M. haemolytica* isolates, the upstream segments of *wza1.1* and the *wza2*-type alleles (green) have possibly been acquired by horizontal transfer. The *wza2*-type alleles contain unique sequences in the 3′ region of the gene which may be also have been acquired by horizontal transfer. The *wza3*-type alleles are highly conserved and only single nucleotide substitutions occur. The *wza4.1* allele differs from the *wza*3-type alleles mainly due to two recombinant segments (yellow, nucleotides 1 to 219, and 678 to 777) although single site substitutions are also present. The *M. haemolytica wza5.1* allele is similar to the *M. glucosida* alleles *wza6.1* and *wza6.2* due to a recombinant segment (light blue, nucleotides 270 to 750). The *M. haemolytica wza7.1* allele contains a recombinant segment (blue, nucleotides 1 to 324) that is almost identical to the corresponding region of the *M. glucosida* alleles *wza6.1* and *wza6.2*, but the remainder differs from any *wza* alleles. The *M. glucosida* allele *wza8.1* contains a recombinant segment (red, nucleotides 1 to 138) that is almost identical to the corresponding region of the *M. haemolytica wza9*-type alleles, and also contains a recombinant segment (orange, nucleotides 819 to 903) that is almost identical to the corresponding region of the *M. glucosida wza6*-type alleles as well as the *M. haemolytica wza9*-type alleles. The *M. haemolytica wza9*-type alleles contain a recombinant segment (grey, nucleotides 753 to 966) that is almost identical to the corresponding region of the *M. glucosida wza6*-type alleles.

The distribution of *wza* allele types among isolates (Table 3.28) indicates that the *wza3*-type alleles occur in a broad range of *M. haemolytica* isolates of serotypes A2, A7, and A12 and ETs 1, 4, 12-14, and 16-22, whereas the *wza1.1* allele occurs in bovine *M. haemolytica* isolates of serotypes A1 and A6 and ETs 1 and 2, the *wza2*-type alleles occur in ovine *M. haemolytica* isolates of serotypes A1 and A6 and ETs 6-

9, the *wza4.1* allele occurs in an ovine *M. haemolytica* A5 isolate of ET 5, the *wza5.1* allele occurs in ovine *M. haemolytica* A14 and A16 isolates of ETs 10 and 11, the *wza7.1* allele occurs in an ovine *M. haemolytica* A13 isolate of ET 15, and the *wza9*-type alleles occur in ovine *M. haemolytica* isolates of serotypes A8 and A9 of ETs 8 and 9. The *wza6*-type alleles occur in *M. glucosida* serotype A11 isolates of ETs 1, 3, and 5, whereas the *wza8.1* allele is present in an untypeable (UG3) *M. glucosida* isolate of ET 5. These alleles (*wza6.1*, *wza6.2*, and *wza8.1*) are divergent except for a recombinant segment (orange) that is also similar to the corresponding region of the *M. haemolytica wza9*-type alleles.

Pairwise differences in nucleotide and amino acid sequences between representative pairs of the 10 allele types of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* ranged from 21 to 297 (2.2 to 30.7 %) nucleotide sites and from 5 to 98 (1.6 to 30.4 %) amino acid positions (Table 3.29).

The *wza* tree topology (Figure 3.44) shows that lineages I, II, and IV are associated exclusively with *M. haemolytica* and lineage VI is associated with *P. trehalosi*, whereas lineages III and V consist of isolates of both *M. haemolytica* and *M. glucosida*. Lineage I is associated with *M. haemolytica wza1*- and *wza2*-type alleles, lineage II with *M. haemolytica wza3*- and *wza 4*-type alleles, lineage IV with *M. haemolytica wza7*-type alleles, and lineage VI with *P. trehalosi wza10*-type alleles. Lineage III is associated with the *M. haemolytica wza5.1* allele and *M. glucosida wza6.1* and *wza6.2* alleles and lineage V is associated with the *M. glucosida wza8.1* allele and *M. haemolytica wza9.1* and *wza9.2* alleles.

The positions of variable inferred amino acid sites within the Wza protein of M. haemolytica and M. glucosida correspond to the locations of three hypervariable

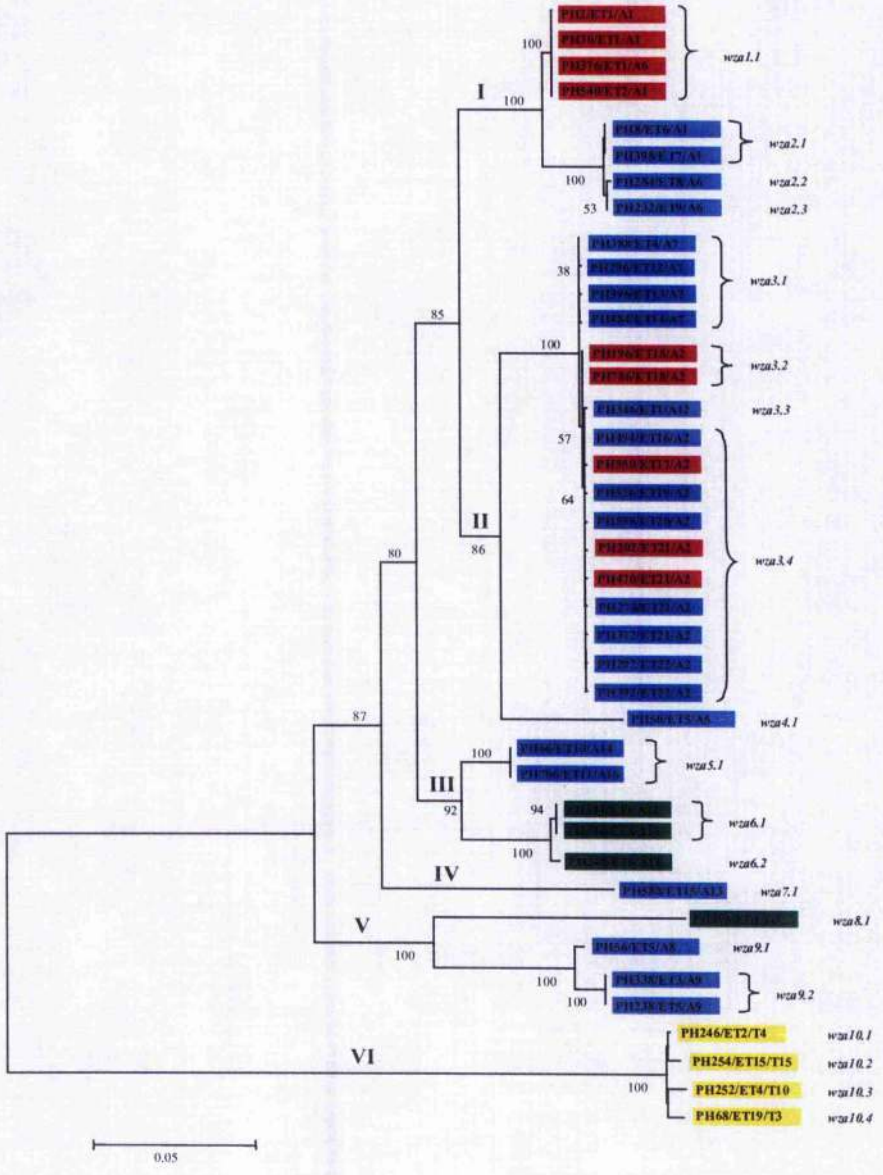


Figure 3.44 Minimum evolution (ME) tree for the *wza* gene of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, ovine *M. haemolytica* isolates in blue, *M. glucosida* in green, and *P. trehalosi* in yellow. Electrophoretic types (ETs) have been described in separate schemes for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b).

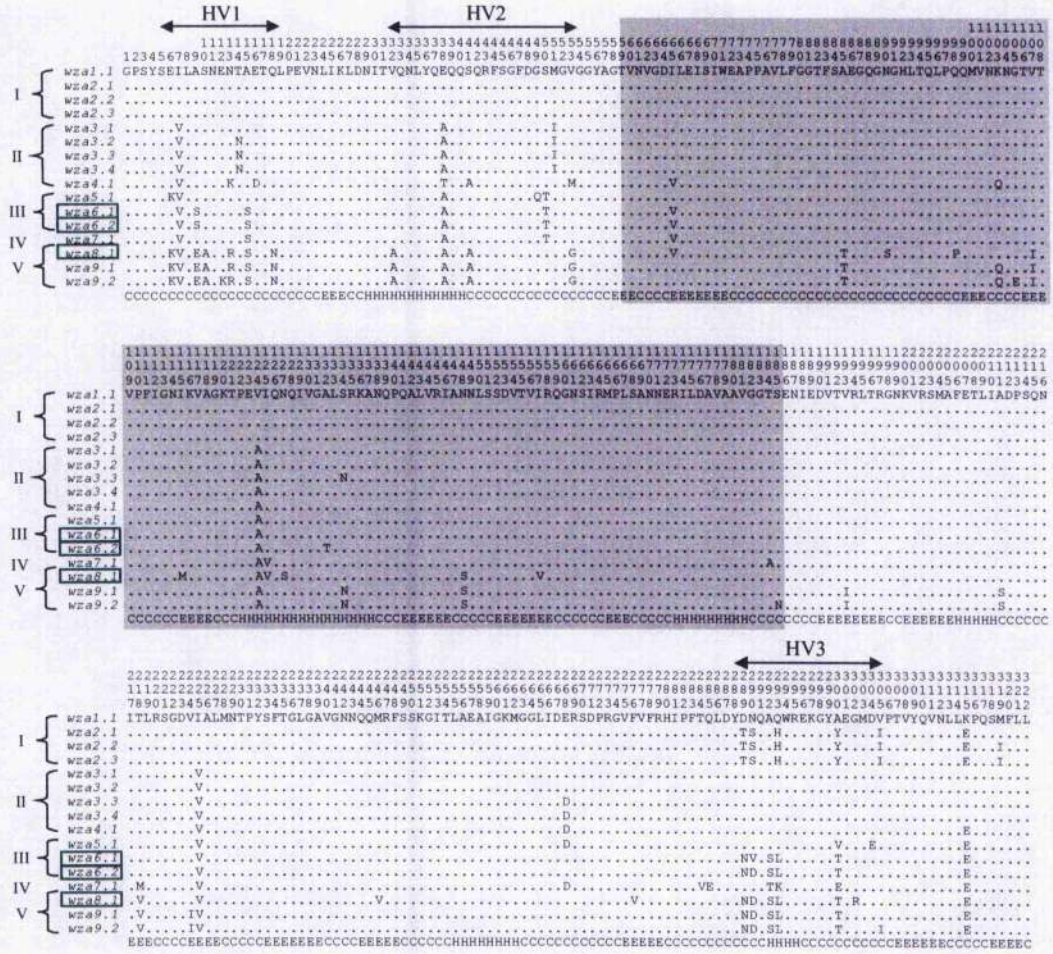


Figure 3.45 Distribution of polymorphic amino acid sites among the *wza* alleles of *M. haemolytica* and *M. glucosida*. Allele designations are shown to the left of each sequence. The Roman numerals correspond to the major lineages of Figure 3.44 The *M. glucosida* alleles are highlighted within the green boxes. The numbers above the sequences represent amino acid positions. The dots represent sites where the amino acids match those of the first (topmost) sequence (i.e. *wza1.1*). HV1 to HV3 represent the hypervariable domains. Shaded region represents a common domain of the polysaccharide biosynthesis/export protein.

regions, HV1 to HV3 (Figure 3.45). The secondary structure prediction (Jones, 1999; McGuffin et al., 2000) indicates that the three hypervariable regions are located within the helices and coils of the 5' and 3' ends of the Wza protein sequence. In contrast, the remainder of the Wza protein sequence, particularly the common domains of polysaccharide biosynthesis/export proteins (grey region), are highly conserved. The relatively low d_S/d_N ratios for the HV1 (3.93), HV2 (5.26), and HV3 (5.91), in contrast to the high d_S/d_N ratio for the conserved regions (43.84) indicate that the variable regions are subject to different degrees of evolutionary constraint compared to the conserved regions (Table 3.30). Amino acid divergence in the conserved regions is subject to a higher degree of selective constraint in comparison to that in the hypervariable regions.

3.1.2 Phylogenetic relationships among 10 *M. haemolytica* isolates based on the concatenated sequences of seven conserved genes

The conserved genes *recA*, *asd*, *mtlD*, *galE*, *gnd*, *g6pd*, and *gcp* were selected for analysis of the concatenated nucleotide sequences because the variation that occurred in each gene was due mainly to point mutation and there was no evidence of horizontal DNA transfer and recombination. The seven sequences of each gene for the 10 *M. haemolytica* isolates were concatenated to provide an in-frame sequence of 7044 nucleotides. There were 24 polymorphic nucleotide sites and pairwise differences raged from 1 to 12 nucleotide sites. A minimum-evolution tree was constructed from the concatenated sequences (Figure 3.46). The tree shows that isolates PII2, PH706, and PH66 cluster together in lineage I, PH296, PH588, and PH196 cluster together in lineage II, and PH494, PH202, PH278, and PH292 cluster together in lineage III. The tree topology (Figure 3.46) is very similar to that based on MLEE (Figure 2.1). With the exception of isolate PH494, lineages I, II, and III of

Table 3.30 Sequence diversity and substitution rates for the hypervariable domains, HV1 to HV3, and conserved regions of the *wza* genes of 32 *M. haemolytica* and four *M. glucosida* isolates

Domain	Sequence diversity (%)		d_S [a]	d_N [a]	d_S/d_N
	Nucleotide	Amino acid			
HV1	59.0	76.9	53.30 ± 18.54	13.55 ± 3.79	3.93
HV2	30.3	31.8	24.32 ± 9.35	4.62 ± 1.75	5.26
HV3	43.1	58.8	58.08 ± 14.63	9.83 ± 3.90	5.91
Conserved regions	26.3	10.7	39.46 ± 3.88	0.90 ± 0.19	43.84

[a] d_S is the number of synonymous substitution per 100 synonymous sites; d_N is the number of nonsynonymous substitution per 100 nonsynonymous sites. Values are means ± standard deviations.

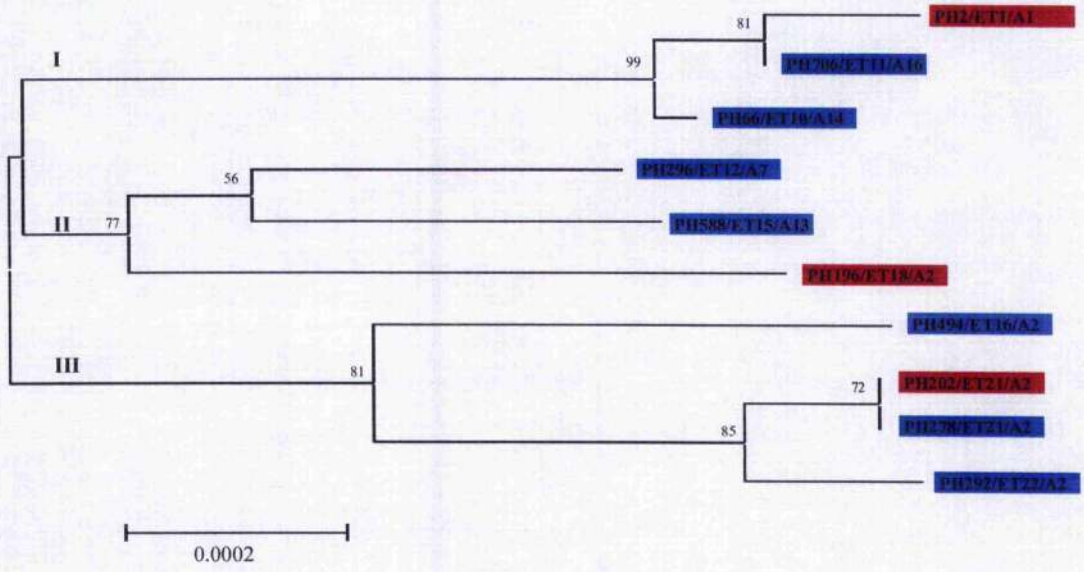


Figure 3.46 Minimum evolution (ME) tree for the concatenated sequences (7044 bp) of the seven conserved genes *recA*, *asd*, *mtlD*, *galE*, *gnd*, *g6pd*, and *gcp* of *M. haemolytica*. The tree was constructed with Jukes-Cantor correction for nucleotide substitutions and the complete deletion option for handling gaps. Bootstrap values (1000 replications) of > 50 only are shown. Bovine *M. haemolytica* isolates are highlighted in red, and ovine *M. haemolytica* isolates in blue. Electrophoretic types (ETs) have been previously described in *M. haemolytica* (Davies *et al.*, 1997a; 1997b).

the concatenated sequence tree correspond to lineages A, B, and C, respectively, of the MLEE tree. In the MLEE tree (Figure 2.1) isolate PH494 appears to be more closely related to isolates PH296, PH1588, and PH196 (lineage B), whereas in the sequence tree (Figure 3.46) isolate PH494 is more closely related to isolates PH202, PH278, and PH292 (lineage III). In particular, the bovine and ovine A2 isolates PH202 and PH278 have identical sequences which suggests that they have common origins.

3.2 Characterization of temperate phages

3.2.1 Optimum concentration of mitomycin C

The optimum concentration of mitomycin C required to induce bacteriophages was determined in four serotype A1 *M. haemolytica* isolates (PH2, PH280, PH342, PH370) by adding increasing volumes of mitomycin C (final concentration of 0, 0.01, 0.05, 0.1, 0.2, 1.0, and 2.0 μg/ml) to early logarithmic phage cultures and measuring the OD_{660} at regular intervals. The data for isolate PH342 (Figure 3.47) indicates that the minimum concentration of mitomycin C required to give complete lysis was 0.1 μg/ml. The results for isolates PH2, PH280, and PH370 were similar to those for isolate PH342 (data not shown). A final mitomycin C concentration of 0.2 μg/ml was selected for subsequent experiments because this concentration resulted in more rapid lysis than a concentration of 0.1 μg/ml.

3.2.2 Induction of bacteriophages in *M. haemolytica*, *M. glucosida*, and *P. trehalosi* isolates

Thirty two *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates

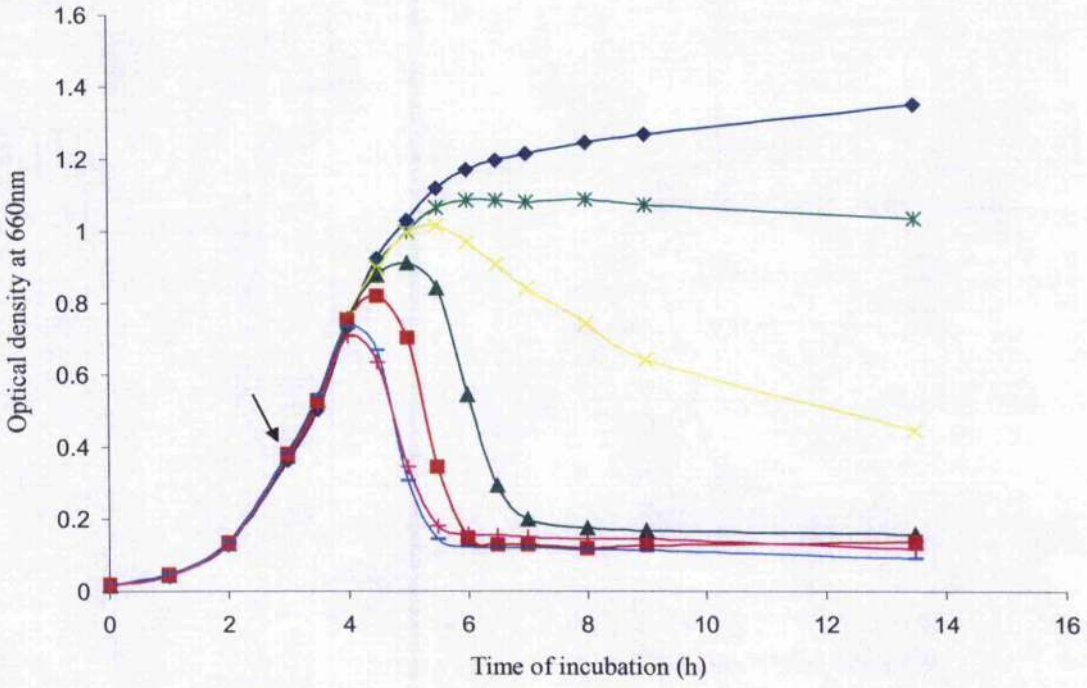


Figure 3.47 Influence of mitomycin C concentration on the lysis kinetics (induction of bacteriophages) of bovine *M. haemolytica* isolate PH342 (serotype A1). The cultures were treated with 0.00 µg/ml (—◆—)(control), 0.01 µg/ml (—✱—), 0.05 µg/ml (—×—), 0.1 µg/ml (—▲—), 0.2 µg/ml (—■—), 1.0 µg/ml (—–—), and 2.0 µg/ml (—+—) final concentration of mitomycin C. The arrow indicates the point at which the mitomycin C was added.

representing various ETs, serotypes, and host species, were tested for phage induction with a final mitomycin C concentration of 0.2 µg/ml. The addition of mitomycin C induced lysis in 26 of the 32 *M. haemolytica* isolates, in one of the six *M. glucosida* isolates, and in two of the four *P. trehalosi* isolates. Three different lysis patterns were identified in the 42 cultures of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* based on the final OD_{660} values 6-7 h after addition of mitomycin C (Figure 3.48). Complete lysis was represented by a final OD_{660} value of 0.4 or less (Figure 3.48A), partial lysis by a final OD_{660} value in the range 0.4 to 1.0 (Figure 3.48B), and no lysis was represented by no reduction of the final OD_{660} value (Figure 3.48C). Complete lysis occurred in 20 *M. haemolytica* isolates, partial lysis in six *M. haemolytica* isolates, one *M. glucosida* isolate, and three *P. trehalosi* isolates, whereas no lysis was detected in six *M. haemolytica* isolates, five *M. glucosida* isolates, and one *P. trehalosi* isolate (Table 3.31).

The isolates that were not induced by 0.2µg/ml mitomycin C treatment were subsequently tested with higher concentration of mitomycin C (0.5, 1.0, and 5.0 µg/ml). Cultures of *M. haemolytica* isolates PH202, PH494, and PH550 were partially lysed with 1.0 µg/ml mitomycin C and cultures of *M. glucosida* isolate PH574 and *P. trehalosi* isolate PH254 were partially lysed with 0.5 µg/ml mitomycin C (Table 3.31).

3.2.3 Phage morphology

The 30 cultures that were lysed with 0.2 µg/ml mitomycin C were examined by electron microscopy (EM) for the presence of bacteriophage particles. Intact phage particles were detected in 26 cultures, but no phages were detected in the cultures of *M. haemolytica* isolates PH232 and PH706, and *P. trehalosi* isolates PH246 and PH68.

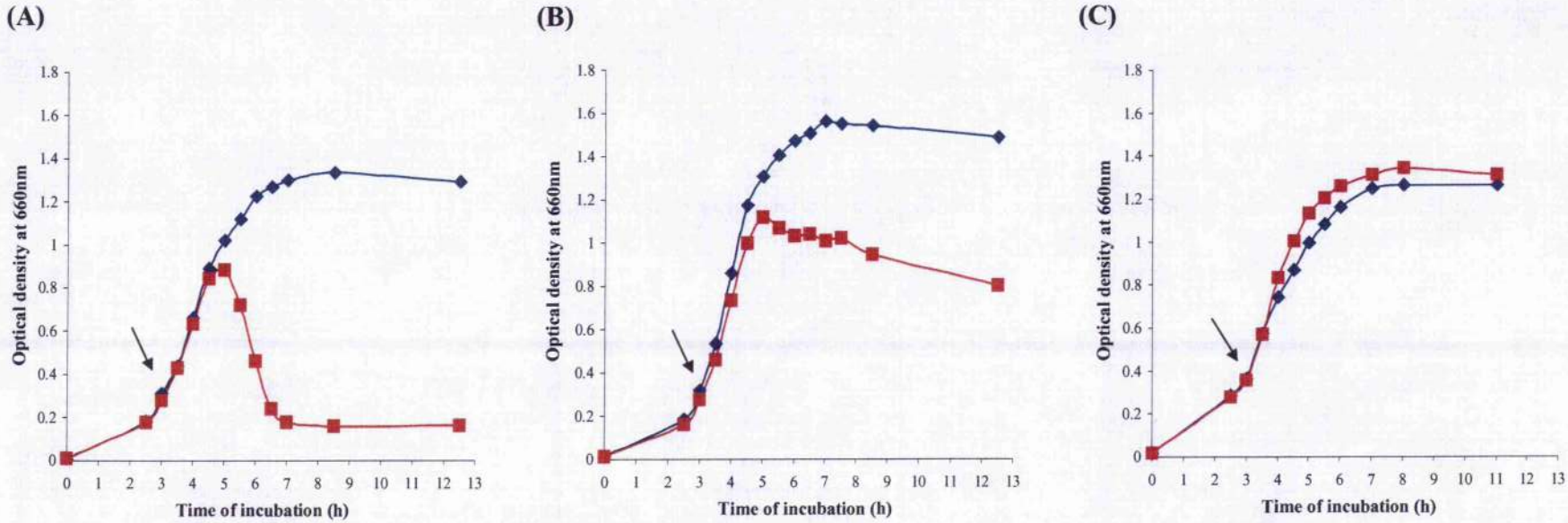


Figure 3.48 Phage induction profiles showing (A) complete lysis (isolate PH2), (B) partial lysis (PH66), and (C) no lysis (PH494).

The cultures were treated with 0.2 µg/ml of mitomycin C (—■—). The untreated culture (—◆—) is shown as control. The arrow indicates the point at which the mitomycin C was added.

The cultures of isolates PH484, PH598, and PH372 contained two types of phages, but the remaining 23 cultures contained only one type of phage. All phages had icosahedral heads as evidenced by the presence of capsids with hexagonal outlines, but three different tail morphologies occurred (Figure 3.49). The differences in tail morphologies were used to classify the phages into one of three phage families, namely *Myoviridae*, *Siphoviridae*, and an unrecognised family. The classification of the phages, including head and tail dimensions, from *M. haemolytica*, *M. glucosida*, and *P. trehalosi* are shown in Table 3.31.

Bacteriophages from 11 *M. haemolytica* isolates had long, flexible, and non-contractile tails and were classified as *Siphoviridae* (Figure 3.50 and Table 3.31). The phages had three different head types; an elongated head (Figure 3.50A-C), a small-sized isometric head (Figure 3.50D), and a normal-sized isometric head (Figure3.50E-F).

Bacteriophages from 14 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate had contractile tails and were classified as *Myoviridae* (Figure 3.51 and Table 3.31). These phages varied in the size of their head and tail shapes resulting in four distinctive types: a normal sized isometric head (Figure 3.51A-E), a large-sized isometric head (Figure 3.51F), a small-sized isometric head (Figure 3.51G), and a large, well defined base plate (Figure3.51H-I).

Bacteriophages from two *M. haemolytica* isolates, PH196 and PH786, had heads similar to those of the *Siphoviridae*, *Myoviridae*, and *Tectiviridae*, but had unique noncontractile, short, thick tails that are unique to these phages and represent an unrecognized phage family type (Figure 3.52). Rough spherical particles were also detected in the lysates of PH196 and PH786 and, in most cases, intact phages were

Table 3.31 Properties of bacteriophages induced in 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* isolates representing various ETs, serotypes, and host species.

Strain	ET	Serotype	Host species	Type of [a] lysis	No of phage types	Phage [b] family	Head size [c] (nm)	Tail size [d] (nm)	DNA isolation	RE type	PCR [e] band
M. haemolytica											
PH2	1	A1	Bovine	Complete	1	*Siphoviridae*	53 × 63	175 × 7	Yes	A	+
PH30	1	A1	Bovine	Complete	1	*Myoviridae* *	55 × 55	156 × 18	No	ND	+§
PH376	1	A6	Bovine	Complete	1	*Myoviridae*	55 × 55	151 × 15	Yes	A	+
PH346	1	A12	Ovine	No	ND	ND	ND	ND	ND	ND	+
PH540	2	A1	Bovine	Complete	1	*Myoviridae* *	55 × 55	151 × 18	No		-
PH338	3	A9	Ovine	No	ND	ND	ND	ND	ND	ND	+
PH388	4	A7	Ovine	Complete	1	*Myoviridae*	60 × 60	157 × 17	Low yield	ND	+
PH150	5	A5	Ovine	Complete	1	*Siphoviridae*	48 × 48	166 × 9	No	ND	+
PH56	5	A8	Ovine	Complete	1	*Siphoviridae*	58 × 58	170 × 9	Yes	B	-
PH238	5	A9	Ovine	Complete	1	*Siphoviridae*	56 × 56	233 × 8	Low yield	ND	-
PH8	6	A1	Ovine	Complete	1	*Siphoviridae*	58 × 58	207 × 8	Yes	C	-
PH398	7	A1	Ovine	Complete	1	*Siphoviridae*	58 × 58	207 × 8	Yes	C	-
PH284	8	A6	Ovine	Complete	1	*Siphoviridae*	58 × 58	199 × 8	Yes	D	-
PH232	9	A6	Ovine	Partial	0	ND	ND	ND	ND	ND	+
PH66	10	A14	Ovine	Partial	1	*Myoviridae*	61 × 61	162 × 19	Yes	E	+
PH706	11	A16	Ovine	Partial	0	ND	ND	ND	ND	ND	-
PH296	12	A7	Ovine	Complete	1	*Myoviridae*	64 × 64	144 × 17	Yes	F	+
PH396	13	A7	Ovine	Partial	1	*Myoviridae*	57 × 57	133 × 17	Yes	G	+
PH484	14	A7	Ovine	Complete	2	*Myoviridae*	57 × 57	133 × 17	Yes	G	+
						Myoviridae *	43 × 43	150 × 20	No	-	+
PH588	15	A13	Ovine	No	ND	ND	ND	ND	ND	ND	-
PH494	16	A2	Ovine	No [$]	0	ND	ND	ND	ND	ND	+
PH550	17	A2	Bovine	No [$]	0	ND	ND	ND	ND	ND	+
PH196	18	A2	Bovine	Complete	1	Unknown	60 × 60	40 × 12	Yes	H	-
PH786	18	A2	Bovine	Complete	1	Unknown	60 × 60	40 × 12	Yes	H	-
PH526	19	A2	Ovine	Partial	1	*Myoviridae* *	43 × 43	150 × 20	No	ND	-
PH598	20	A2	Ovine	Complete	2	*Myoviridae* *	43 × 43	150 × 20	No	-	-
						Siphoviridae	58 × 58	196 × 8	Yes	I	
PH202	21	A2	Bovine	No [$]	0	ND	ND	ND	ND	ND	+
PH470	21	A2	Bovine	Partial	1	*Myoviridae* *	53 × 53	144 × 20	No	ND	-
PH278	21	A2	Ovine	Complete	1	*Myoviridae* *	53 × 53	157 × 19	No	ND	+
PH372	21	A2	Ovine	Complete	2	*Myoviridae* *	68 × 68	168 × 14	No	ND	+
						Siphoviridae	58 × 58	166 × 6	No	ND	
PH292	22	A2	Ovine	Complete	1	*Siphoviridae*	58 × 58	196 × 8	Yes	I	+
PH392	22	A2	Ovine	Complete	1	*Siphoviridae*	58 × 58	196 × 8	Yes	I	+

Table 3.31 (continued)

Strain	ET	Serotype	Host species	Type of [a] lysis	No of phage types	Phage [b] family	Head size [c] (nm)	Tail size [d] (nm)	DNA isolation	RE type	PCR [e] band
M. glucosida											
PH344	1	A11	Ovine	Partial	1	*Myoviridae*	53 × 53	127 × 15	Yes	J	+
PH498	3	A11	Ovine	No	ND	ND	ND	ND	ND	ND	-
PH240	5	A11	Ovine	No	ND	ND	ND	ND	ND	ND	-
PH496	7	UG3	Ovine	No	ND	ND	ND	ND	ND	ND	-
PH574	10	UG3	Ovine	No [§]	0	ND	ND	ND	ND	ND	-
PH290	16	UG3	Ovine	No	ND	ND	ND	ND	ND	ND	-
P. trehalosi											
PH246	2	T4	Ovine	Partial	0	ND	ND	ND	ND	ND	-
PH252	4	T10	Ovine	Partial	1	*Myoviridae* *	62 × 62	139 × 20	Yes	K	-
PH254	15	T15	Ovine	No [§]	0	ND	ND	ND	ND	ND	-
PH68	19	T3	Ovine	Partial	0	ND	ND	ND	ND	ND	-

[a] 0.2 µg/ml mitomycin C was used. Complete lysis, final OD_{660} of 0.4 or less; partial lysis, final OD_{660} of 0.4 - 1.2; No lysis, final OD_{660} is the same as the control.

[§] At higher concentrations of mitomycin C (1.0 µg/ml - PH202, PH494, and PH550 and 0.5 µg/ml - PH574 and PH254) partial lysis was observed.

[b], [c], and [d] Based on electron microscopy (see Figure 3.49).

* Only contracted tail forms were found and the measurements are based on these.

[e] PCRs were carried out for all 42 isolates with primer pair 505/506. This primer pair amplifies tail genes of phage from bovine *M. haemolytica* A1 isolate PHL213 (see Appendix I).

§ PCR band of larger size (approximately 1980 bp) than expected (approximately 700 bp) (see section 3.2.9.6)

ND: Not determined

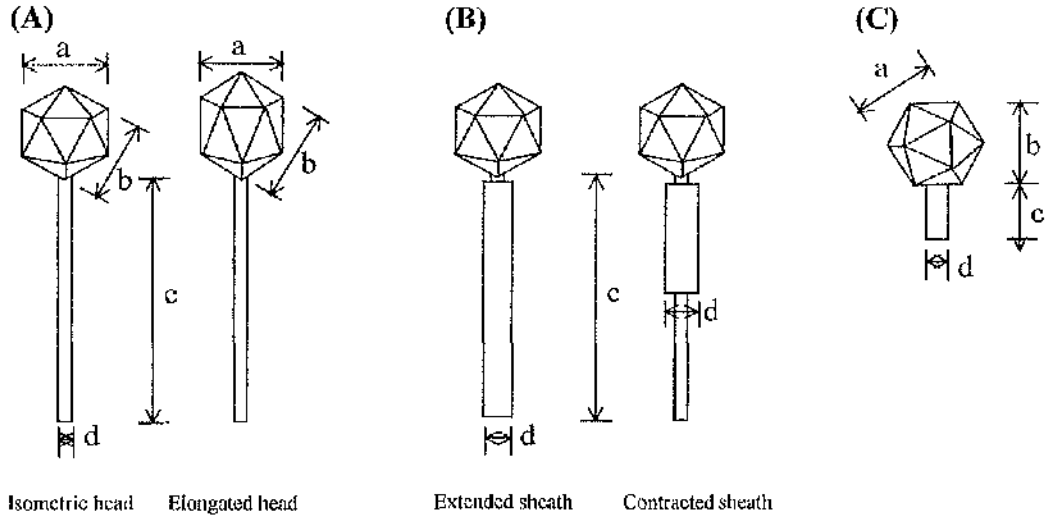


Figure 3.49 Diagrammatic representation of morphological types of temperate phages from *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. (A) *Siphoviridae* type (long noncontractile tail), (B) *Myoviridae* type (contractile tail), and (C) unrecognized family type (short tail). Measurement details are also indicated (a, head width; b, head height; c, tail length; d, tail width).

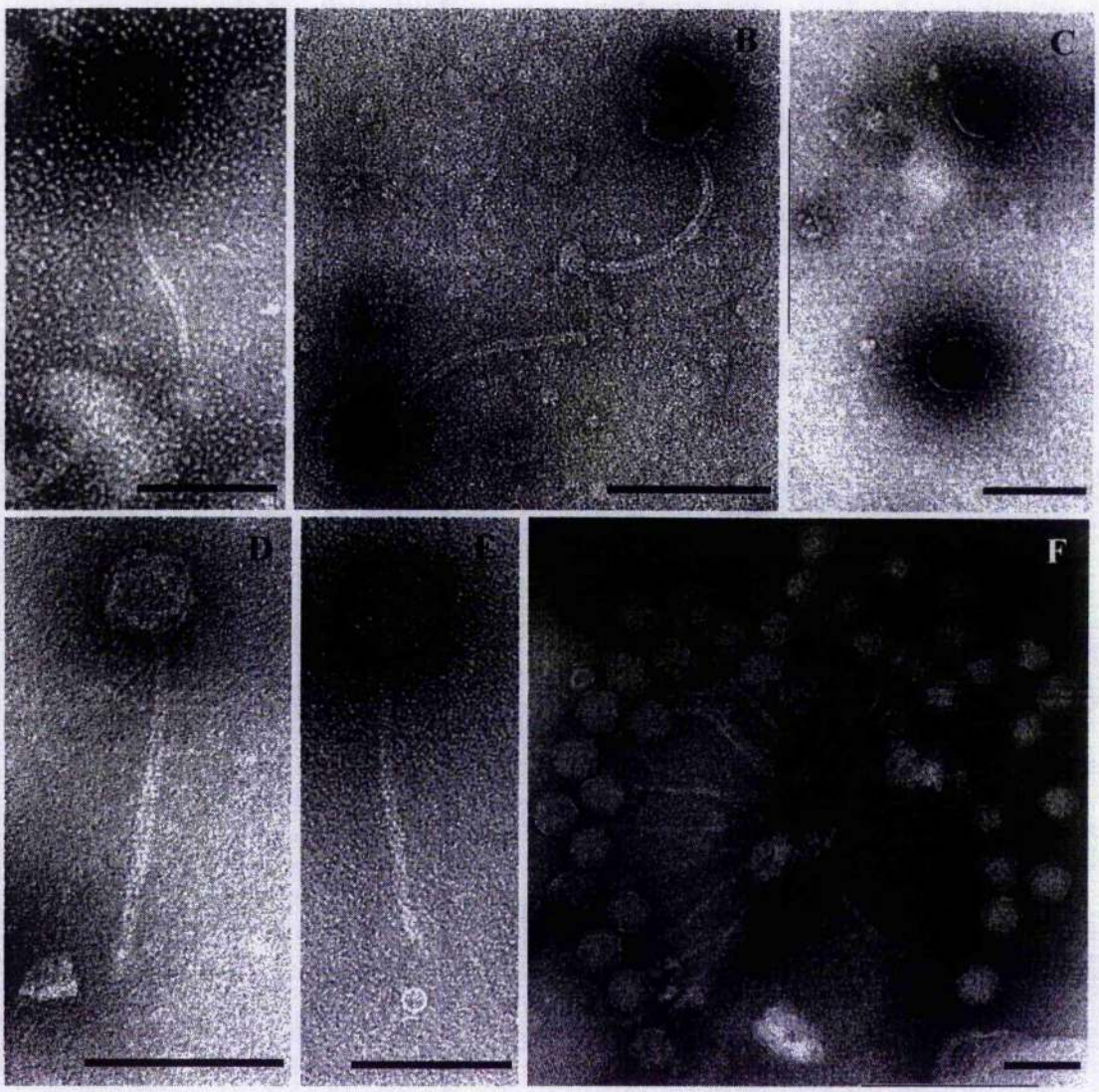


Figure 3.50 Negatively-stained electron micrographs of *Siphoviridae*-type phages of *M. haemolytica*. Phages are from *M. haemolytica* isolate PH2 (elongated head) (A to C), PH50 (small isometric head, 48 nm) (D), PH392 (normal isometric head, 58 nm) (E), and PH292 (normal isometric head, 58 nm) (F). Bar = 100 nm.

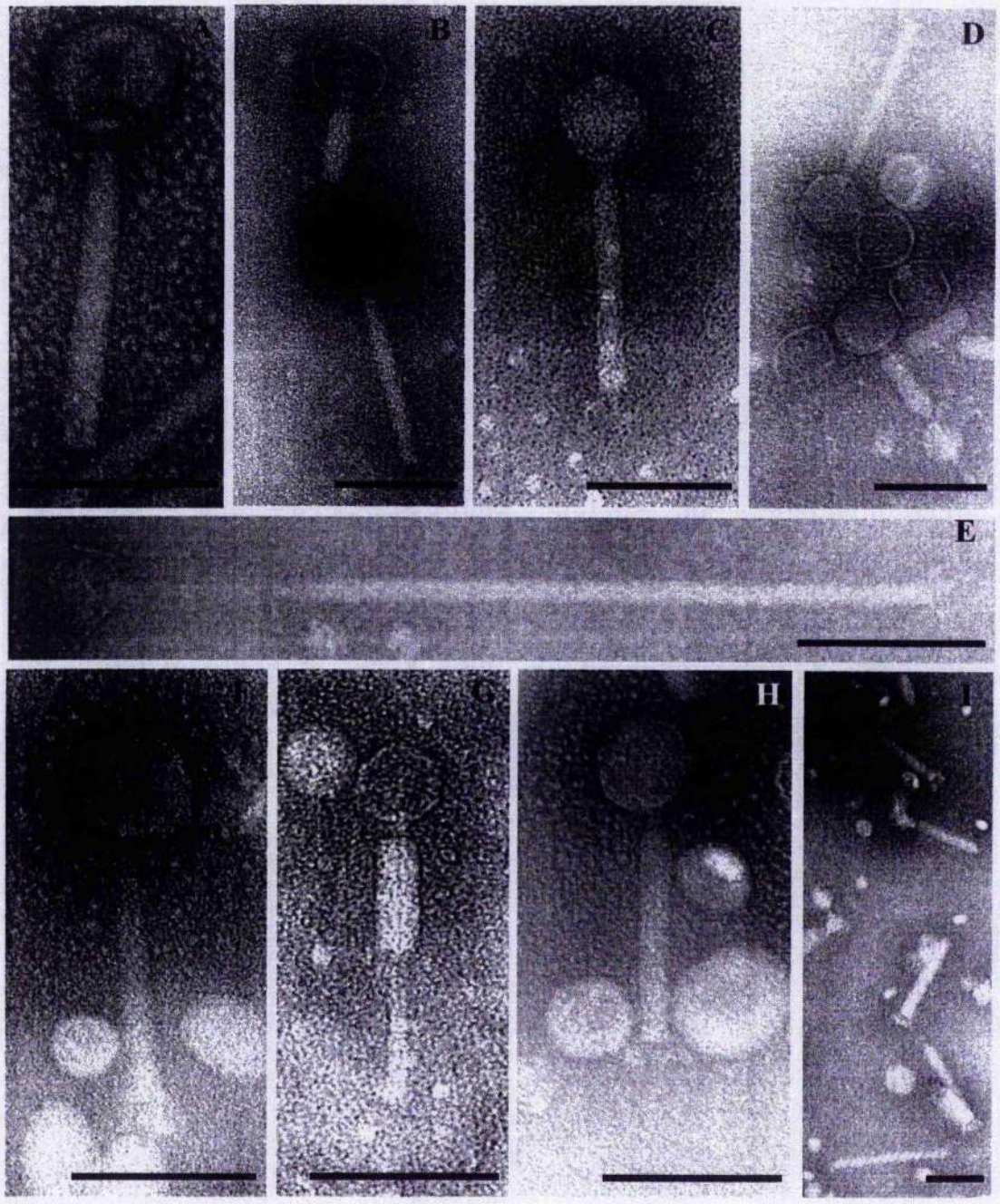


Figure 3.51 Negatively-stained electron micrographs of *Myoviridae*-type phages of *M. haemolytica*. Phages are from *M. haemolytica* isolate PH66 (head, 61 nm) (A and B), PH388 (head, 60 nm) (C), PH376 (head, 56 nm) (D), PH376 (abnormal phage) (E), PH372 (head, 68 nm) (F), PH598 (head, 43 nm) (G), PH484 (prolonged base plate) (H), and phage tails from isolate PH484 (I). Bar = 100 nm.

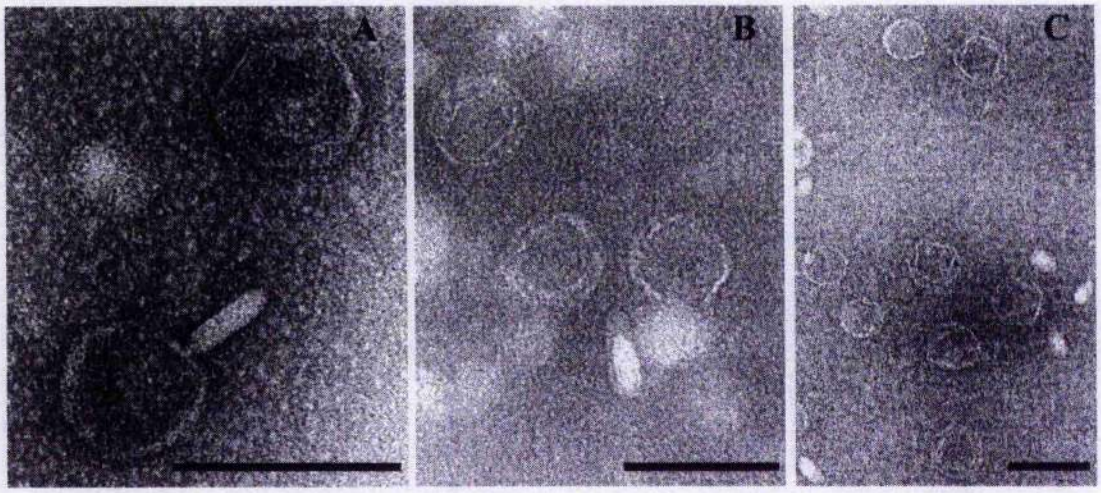


Figure 3.52 Negatively-stained electron micrographs of an unrecognized family type of phages of *M. haemolytica*. Phages are from *M. haemolytica* isolate PH196 (A and B) and PH786 (C). Bar = 100 nm.

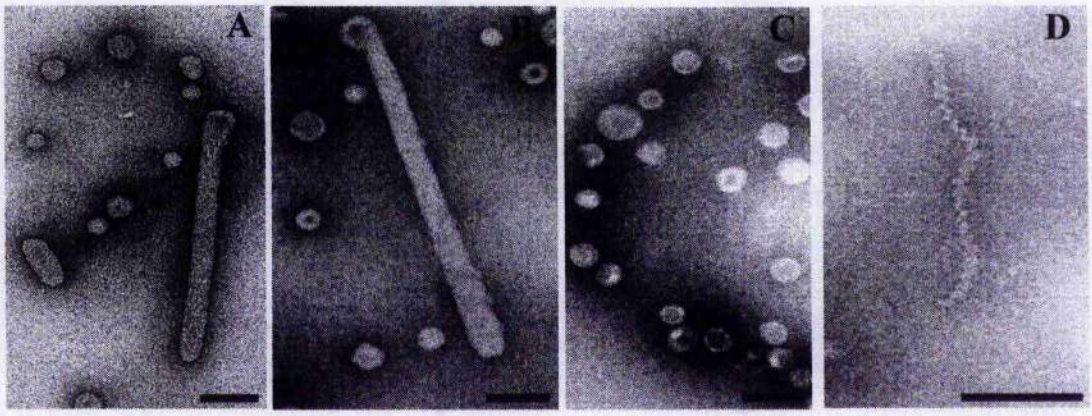


Figure 3.53 Negatively-stained electron micrographs of phage head-like structures of *M. haemolytica*. They are from isolate *M. haemolytica* isolate PH494 (A), *M. haemolytica* isolate PH202 (B), *P. trehalosi* isolate PH254 (C), and spiral structure from *M. haemolytica* isolate PH398. Bar = 100 nm.

relatively rare because the heads and tails were separate structures. In this respect, the hexagonal heads and rough spherical particles are similar to colicin K (Bradley, 1967) and the tail like objects are similar to *Plectrovirus* virions of *Inoviridae*.

The isolates that were lysed with higher concentrations of mitomycin C (*M. haemolytica* isolates PH202, PH494, and PH550, *M. glucosida* isolate PH574, and *P. trehalosi* isolate PH254) were also examined by electron microscopy. Typical phage particles were not observed in these isolates, but many spherical particles were present (Figure 3.53A to C). In addition, spiral structures (Figure 3.53D) were commonly seen in cultures of *M. haemolytica* isolates PH30, PH50, PH66, PH238, PH292, PH396, and PH398 and *P. trehalosi* isolate PH246.

3.2.4 Phage DNA isolation

Electron microscopy unambiguously identified lysogenic phages from 24 *M. haemolytica* isolates, one *M. glucosida* isolate, and one *P. trehalosi* isolate. However, phage DNA was successfully isolated from only 17 (65 %) of the corresponding lysates (Table 3.31). These included 15 *M. haemolytica* isolates (PH2, PH376, PH56, PH8, PH398, PH284, PH66, PH296, PH396, PH484, PH196, PH786, PH598, PH292, and PH392), one *M. glucosida* isolate (PH344), and one *P. trehalosi* isolate (PH254). Of these, isolates PH484 and PH598 produced two phage types, whereas the remaining isolates produced one phage type (Table 3.31). However, restriction endonuclease analysis of isolates PH484 and PH598 suggested that DNA representing only one phage type was isolated (see section 3.2.5). The phages were designated as PH2Φ, PH376Φ, PH56Φ, PH8Φ, PH398Φ, PH284Φ, PH66Φ, PH296Φ, PH396Φ, PH484Φ, PH196Φ, PH786Φ, PH598Φ, PH292Φ, PH392Φ, PH344Φ, and PH254Φ, respectively.

The yield of phage DNA varied for different isolates. For example, 10 ml of phage lysate were used to recover DNA from 11 phages (PH2Φ, PH376Φ, PH56Φ, PH8Φ, PH398Φ, PH284Φ, PH66Φ, PH396Φ, PH484Φ, PH344Φ, and PH246Φ), but 50 ml of phage lysates were required to obtain a similar amount of DNA from six phages (PH296Φ, PH196Φ, PH786Φ, PH292Φ, PH392Φ, and PH598Φ). Attempts to isolate phage DNA from the remaining nine lysates (PH30, PH540, PH388, PH50, PH238, PH526, PH470, PH278, and PH372) were unsuccessful. Yields of DNA were extremely low from 50 ml of lysate for two isolates (PH388 and PH238) whereas no DNA was recovered from the remaining seven lysates.

3.2.5 Restriction endonuclease analysis

The 17 samples of bacteriophage DNA were characterized by restriction endonuclease analysis using two different restriction enzymes, HindIII and ClaI. Eleven distinct restriction enzyme (RE) patterns (A to K) were observed among the 17 phages (Figure 3.54 and Table 3.31). RE types A to I were represented by *M. haemolytica* isolates, RE type J was represented by *M. glucosida* isolate PH344, and RE type K was represented by *P. trehalosi* isolate PH252.

Phage of the same RE types occurred only in closely related *M. haemolytica* isolates (Table 3.31). For example, RE type A phages were associated with isolates PH2 and PH376 (bovine A1 and A6 isolates of ET 1), RE type C phages were associated with isolates PH8 and PH398 (ovine A6 isolates of ETs 8 and 9), RE type G phages were associated with isolates PH396 and PH484 (ovine A7 isolates of ET 13 and 14), RE type H phages were associated with isolates PH196 and PH786 (bovine A2 isolates of ET 18), and RE type I phages were associated with isolates PH598, PH292, and PH392 (ovine A2 isolates of ETs 20 and 22).

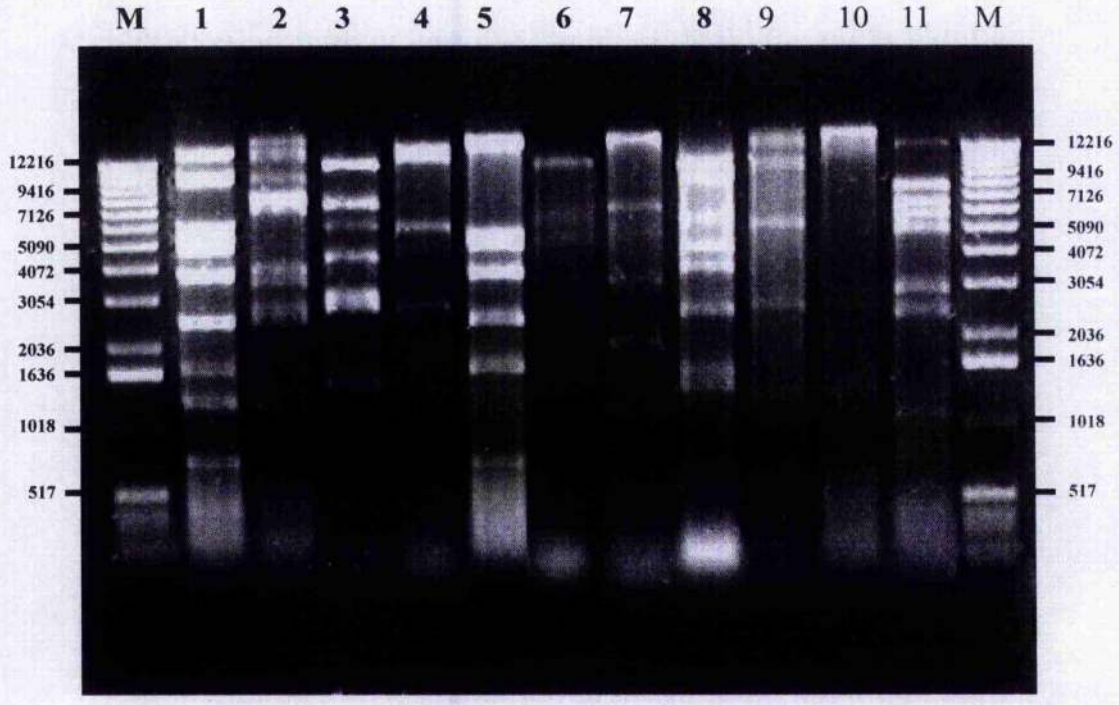


Figure 3.54 Restriction enzyme (RE) patterns of phage DNA from *M. haemolytica* (lane 1– 9), *M. glucosida* (lane 10), and *P. trehalosi* (lane 11) isolates digested with HindIII and ClaI. The phages and their RE types in lanes 1-11 are as follows: 1, PH2Φ (type A); 2, PH56Φ (type B); 3, PH8Φ (type C); 4, PH284Φ (type D); 5, PH66Φ (type E); 6, PH296Φ (type F); 7, PH484Φ (type G); 8, PH786Φ (type H); 9, PH392Φ (type I); 10, PH344Φ (type J); 11, PH252Φ (type K). One Kb DNA markers are shown in the end lanes (M).

Although *M. haemolytica* isolates PH484 and PH598 produced two types of phages, the RE types of their phage DNA samples (G and I, respectively) were identical to those of phage DNA samples from isolates that produced only one phage type. These results indicate that only one type of phage DNA was recovered from isolates PH484 and PH598. The same RE types C, G, H, and I were associated with morphologically identical phages, whereas RE type A occurred in heterogeneous phage families (*Siphoviridae* and *Myoviridae*) (Table 3.31).

3.2.6 Determination of host range of bacteriophages

The host range of induced bacteriophages and non-induced cultures representing 42 isolates was examined by plaque assay using the same 42 indicator isolates (Table 3.31). Lysates from 13 of 32 *M. haemolytica*, one of six *M. glucosida*, and one of four *P. trehalosi* isolates caused zones of lysis (plaques) on certain sensitive indicator isolates (Table 3.32). The plaques caused by lysate of isolate PH376 on indicator isolates PH30, PH540, PH8 and PH398 were clear but all other plaques were faint (Figure 3.55). There was a correlation between the lysates and sensitive indicator strains. For example, *M. haemolytica*, *M. glucosida*, and *P. trehalosi* lysates caused plaques only on indicator isolates of their own species. Furthermore, *M. haemolytica* isolates PH30 (bovine A1) and PH376 (bovine A6) caused plaques mostly on serotype A1 and A6 strains, whereas *M. haemolytica* isolates PH598, PH470, PH278, and PH392 (ovine A2) caused plaques mostly on serotype A2 isolates (Table 3.32).

3.2.7 Comparison of *M. haemolytica* bovine serotype A1 and A6 strains

In the first series of experiments described above, *M. haemolytica* isolates PH2, PH30,

Table 3.32 Lytic spectra of lysates of *M. haemolytica* (*Mh*), *M. glucosida* (*Mg*), and *P. trehalosi* (*Pt*) isolates against indicator isolates

Indicator[a] \ Lysate	PH30 (A1)	PH376 (A6)	PH388 (A7)	PH50 (A5)	PH56 (A8)	PH284 (A6)	PH296 (A7)	PH396 (A7)	PH598 (A2)	PH470 (A2)	PH278 (A2)	PH292 (A2)	PH392 (A2)	PH344 (A11)	PH252 (T10)
	Mh lineage A							*Mh* lineage B		*Mh* lineage C				*Mg*	*Pt*
PH2/ET1/A1	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-
PH30/ET1/A1	-	++	-	-	+	-	-	-	-	-	-	-	-	-	-
PH346/ET1/A12	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
PH540/ET2/A1	-	++	-	-	+	-	-	-	-	-	-	-	-	-	-
PH338/ET3/A9	-	-	-	-	+	+	-	-	-	-	-	+	-	-	-
PH50/ET5/A5	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
PH238/ET5/A9	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
PH8/ET6/A1	+	++	-	-	-	-	-	-	-	-	-	-	-	-	-
PH398/ET7/A1	+	++	-	-	-	-	-	-	-	-	-	-	-	-	-
PH706/ET11/A16	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-
PH494/ET15/A2	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
PH550/ET17/A2	-	-	+	-	-	-	-	-	-	+	-	-	-		-
PH196/ET18/A2	-	-	+	-	-	-	-	+	-	+	+	+	-	-	-
PH786/ET18/A2	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
PH526/ET19/A2	-	+	+	+	-	+	-	-	+	-	-	+	+	-	-
PH598/ET20/A2	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
PH202/ET21/A2	-	-	+	-	-	-	+	-	-	+	+	-	-	-	-
PH278/ET21/A2	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
PH372/ET21/A2	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
PH292/ET22/A2	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
PH392/ET22/A2	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
PH344/ET1/A11	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
PH498/ET3/A11	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
PH252/ET4/T10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Mh lineage A; *Mh* lineage B; *Mh* lineage C; *Mg*; *Pt*

[a] The indicator isolates are shown with their ETs and serotypes. The ETs have been described separately for *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (see Davies *et al.*, 1996a; 1996b).

++, clear plaque zone; +, faint plaque zone; -, no plaque zone

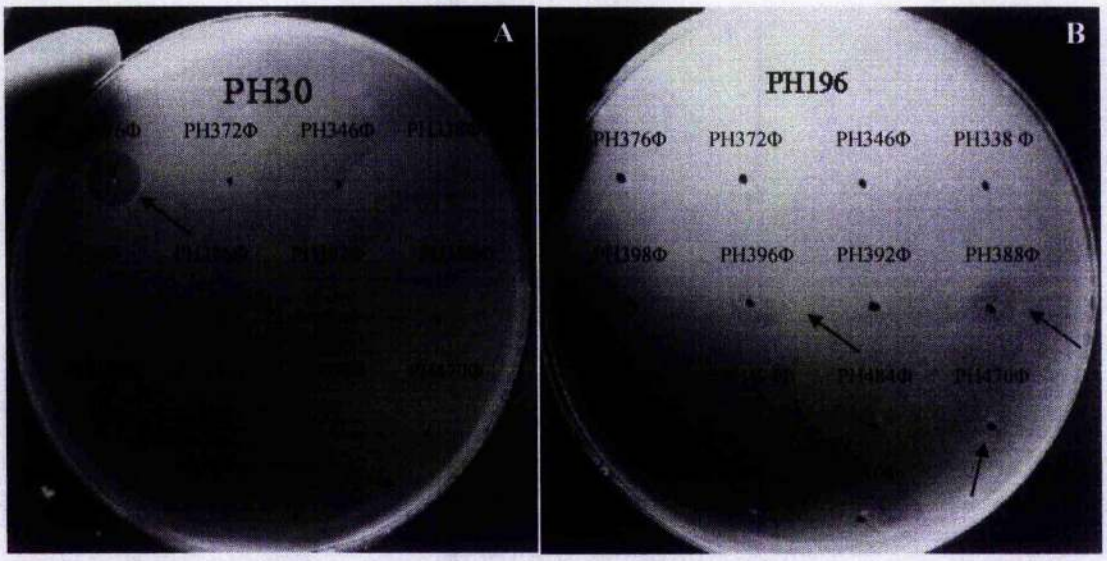


Figure 3.55 Plaque assays showing (A) a clear plaque (lysate of isolate PH376 on indicator isolate PH30) and (B) faint plaques (lysates of isolates PH396, PH388, and PH470 on indicator isolate PH196).

PH376, and PH540, which represent bovine A1 and A6 isolates of ETs 1 and 2, showed significant variation in phage morphology, DNA yield, and host range (Tables 3.31 and 3.32). Since isolates of ET 1 are the major causes of bovine pasteurellosis (Davies *et al.*, 1997a), seven additional isolates were characterized in further detail. The properties of 11 *M. haemolytica* isolates representing serotypes A1 and A6 of ETs 1 and 2 are shown in Table 3.33.

3.2.7.1 Phage morphology

Electron microscopy was carried out for two additional isolates, PH280 and PH560, that underwent lysis after mitomycin C treatment. Bacteriophages from isolates PH280 and PH560 had contractile tails and were classified as *Myoviridae* (Table 3.33). These phages were, morphologically, indistinguishable from those of isolates of PH30, PH540, and PH376.

3.2.7.2 Host range

The host range of phages from the 11 isolates was examined by plaque assay using the same range of indicator strains. Four different patterns of lysis, A -D, were observed for the lysates (phages), whereas three different patterns of lysis, I to III, were observed for the indicator isolates (Table 3.34). When the lysis patterns of phage and indicator isolates were combined, six different combinations (lytic types) were apparent (Table 3.35). Interestingly, there were strong correlations between the lytic types and cell-surface characteristics (capsule and OMP-type) of the isolates (Table 3.35). For example, lytic types 1, 2, and 3 are associated with isolates of serotype A1 and OMP type 1.1.1; lytic type 4 is associated with isolates of serotype A1 and OMP type 1.1.3; and lytic types 5 and 6 are associated with isolates of serotype A6

Table 3.33 Properties of bacteriophages induced in seven serotype A1 and four serotype A6 isolates of *M. haemolytica*

Strain	ET	Serotype	Host species	LPS type	OMP type	Geographic origin	Type of [a] lysis	No of phage types	Phage [b] Family	Head size [c] (nm)	Tail size [d] (nm)	Phage DNA Isolation
PH2	**1**	**A1**	**Bovine**	**1A**	**1.1.1**	**Glasgow**	**Complete**	**1**	***Siphoviridae***	**53 × 63**	**175 × 7**	**Yes**
PH280	1	A1	Bovine	1A	1.1.1	Glasgow	Complete	1	*Myoviridae**	55 × 55	151 × 18	Yes
PH26	1	A1	Bovine	1A	1.1.1	Glasgow	Complete	1	*Myoviridae*	55 × 55	151 × 18	Yes
PH1560	1	A1	Bovine	2A	1.1.1	Germany	Complete	1	*Myoviridae**	55 × 55	151 × 18	Yes
PH30	**1**	**A1**	**Bovine**	**2A**	**1.1.1**	**Glasgow**	**Complete**	**1**	***Myoviridae****	**55 × 55**	**151 × 18**	**No**
PH540	**2**	**A1**	**Bovine**	**2A**	**1.1.3**	**Germany**	**Complete**	**1**	***Myoviridae****	**55 × 55**	**151 × 18**	**No**
PH554	2	A1	Bovine	2A	1.1.3	Germany	Complete	1	*Myoviridae*	55 × 55	151 × 18	No
PH376	**1**	**A6**	**Bovine**	**1A**	**1.1.4**	**Dumfries**	**Complete**	**1**	***Myoviridae****	**55 × 55**	**151 × 18**	**Yes**
PH524	1	A6	Bovine	1A	1.1.4	Newcastle	Complete	1	*Myoviridae*	55 × 55	151 × 18	Yes
PH564	1	A6	Bovine	1A	1.1.4	Germany	Complete	1	*Myoviridae*	55 × 55	151 × 18	Yes
PH812	1	A6	Bovine	2A	1.1.4	Aberdeen	Complete	1	*Myoviridae*	55 × 55	151 × 18	Yes

The previously studied four isolates (PH2, PH30, PH540, PH376) are indicated in bold type.

[a] Complete lysis, final OD_{650} of 0.4 or less

[b, c] and [d] Based on electron microscopy (see Figure 3.49).

* The measurements are based on contracted tail forms.

ND: Not determined

Table 3.34 Lytic spectra of lysates from seven serotype A1 and four serotype A6 isolates of *M. haemolytica* against the same indicator isolates. A - D represent the lytic patterns of the lysates (phages); I - III represent the lytic patterns of the indicator isolates.

		A	A	B	B	C	A	A	D	B	B	B
	Lysate / Indicator	**PH2**	PH280	PH26	PH560	**PH30**	**PH540**	PH554	**PH376**	PH524	PH564	PH812
I	**PH2**	-	-	-	-	+	-	-	+	-	-	-
I	PH280	-	-	-	-	+	-	-	+	-	-	-
I	PH26	-	-	-	-	+	-	-	+	-	-	-
I	PH560	-	-	-	-	+	-	-	+	-	-	-
II	**PH30**	-	-	+	+	-	-	-	++	+	+	+
II	**PH540**	-	-	+	+	-	-	-	++	+	+	+
II	PH554	-	-	+	+	-	-	-	++	+	+	+
III	PH376	-	-	-	-	-	-	-	-	-	-	-
III	PH524	-	-	-	-	-	-	-	-	-	-	-
III	PH564	-	-	-	-	-	-	-	-	-	-	-
III	PH812	-	-	-	-	-	-	-	-	-	-	-

+, faint plaque zone; ++, clear plaque zone; -, no plaque zone

Table 3.35 Cell surface characteristics and lysis patterns of serotype A1 and A6 isolates of *M. haemolytica*

Strain	ET	Serotype	LPS type	OMP type	Lysate (Phage)	Indicator strain	Combined lytic type
PH2	**1**	A1	**1A**	**1.1.1**	**A**	**I**	**1**
PH280	1	A1	1A	1.1.1	A	I	1
PH26	1	A1	1A	1.1.1	B	I	2
PH560	1	A1	2A	1.1.1	B	I	2
PH30	**1**	**A1**	**2A**	**1.1.1**	**C**	**II**	**3**
PH540	**2**	**A1**	**2A**	**1.1.3**	**A**	**II**	**4**
PH554	2	A1	2A	1.1.3	A	II	4
PH376	**1**	**A6**	**1A**	**1.1.4**	**D**	**III**	**5**
PH524	1	A6	1A	1.1.4	B	III	6
PH564	1	A6	1A	1.1.4	B	III	6
PH812	1	A6	2A	1.1.4	B	III	6

and OMP type 1.1.4.

3.2.7.3 Phage DNA isolation and RE assay

Phage DNA was successfully isolated from lysates of isolates PH280, PH26, PH560, PH524, PH564, and PH812 in addition to isolates PH2 and PH376 (Table 3.32). However, attempts to isolate phage DNA from isolate PH554 were unsuccessful as was the case for isolates PH30 and PH540 (Table 3.33). Interestingly, these three isolates are associated with the same indicator isolate pattern II (Table 3.34). The six additional phage DNA preparations were analysed by RE assay using the two restriction enzymes Hind IIII and ClaI. The RE assay showed that all phages from bovine A1 and A6 isolates have identical RE types, i.e. RE type A (Figure 3.56).

3.2.8 Comparison of *M. haemolytica* serotype A2 isolates

In the first series of experiments described above, *M. haemolytica* strains, PH202, PH470, PH278, and PH372, which represent serotype A2 isolates of ET 21, showed significant variation in phage induction, morphology, and host range (Tables 3.31 and 3.32). Since ovine A2 isolates of ET 21 represent a large proportion of isolates associated with ovine pasteurellosis, four additional isolates (PH204, PH208, PH714, and PH776) were characterized in further detail. The properties of bacteriophages induced in eight *M. haemolytica* isolates representing serotype A2 of ET 21 are shown in Table 3.36.

3.2.8.1 Phage morphology

Electron microscopy was carried out for three additional isolates, PH208, PH714, and

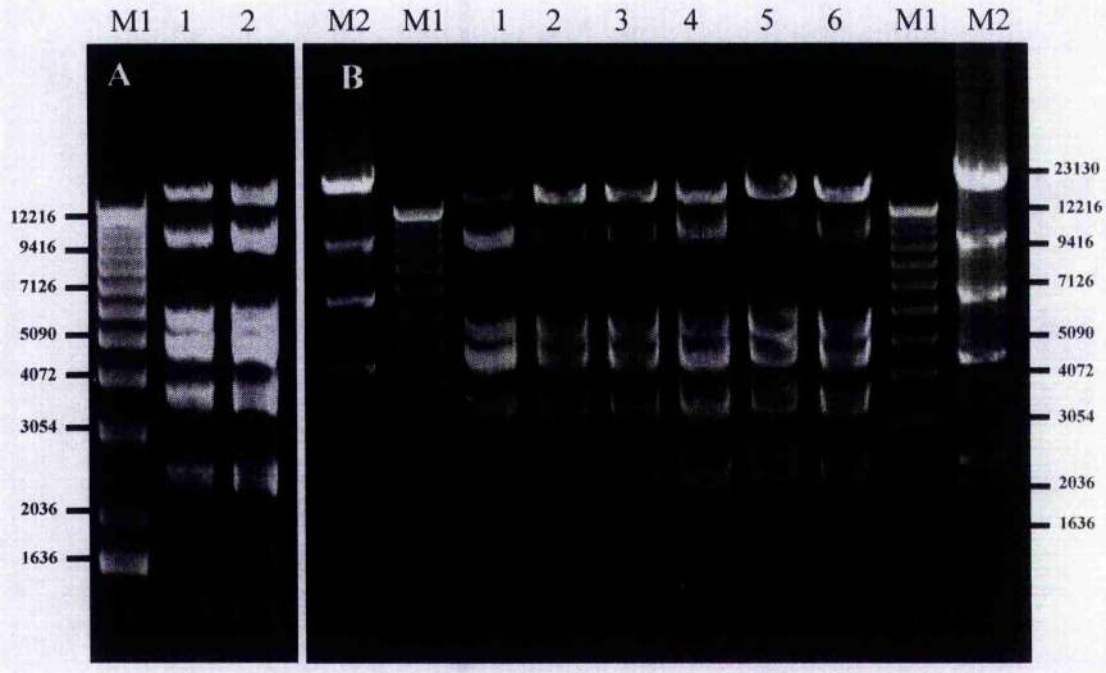


Figure 3.56 Restriction enzyme (RE) patterns of phage DNA from serotype A1 and A6 isolates of *M. haemolytica* digested with HindIII and ClaI. (A) Preliminary examination of the phages PH2Φ (lane 1) and PH376 (lane 2). (B) Subsequent examination of the additional phages PH26Φ (lane 1), PH280Φ (lane 2); PH560Φ (lane 3), PH524Φ (lane 4), PH564Φ (lane 5), and PH812Φ (lane 6). 1 Kb DNA markers and phage λ markers are shown in lanes labelled M1 and M2, respectively.

Table 3.36 Properties of bacteriophages induced in eight serotype A2 isolates of *M. haemolytica*

Strain	ET	Serotype	Host species	LPS type	OMP type	Geographic origin	Type of [a] lysis	No of phage types	Phage [b] family	Head size [c] (nm)	Tail size [d] (nm)	Phage DNA Isolation	Plaque [e] ability
PH202	**21**	**A2**	**Bovine**	**3B**	**2.2.2**	**Glasgow**	**No**	**ND**	**ND**	**ND**	**ND**	**No**	**No**
PH470	**21**	**A2**	**Bovine**	3B	**2.2.2**	**Aberdeen**	**Partial**	**1**	***Myoviridae****	**53 x 53**	**144 x 20**	**No**	**Yes**
PH204	21	A2	Bovine	3B	2.2.2	Glasgow	No	ND	ND	ND	ND	No	No
PH208	21	A2	Bovine	3B	2.2.3	Glasgow	Complete	1	*Myoviridae**	55 x 55	153 x 24	No	Yes
PH278	**21**	**A2**	**Ovine**	**3B**	**2.1.2**	**Penrith**	**Complete**	**1**	***Myoviridae****	**53 x 53**	**157 x 19**	**No**	**Yes**
PH372	**21**	**A2**	**Ovine**	**3B**	**2.1.2**	**Edinburgh**	**Complete**	**2**	***Myoviridae****	**68 x 68**	**168 x 14**	**No**	**No**
									Siphoviridae	**58 x 58**	**166 x 6**	**No**	**No**
PH714	21	A2	Ovine	3B	2.1.2	Carmarthen	Complete	1	*Myoviridae**	53 x 53	143 x 17	No	No
PH776	21	A2	Ovine	3B	2.1.2	Carmarthen	Partial	1	*Myoviridae**	55 x 55	145 x 19	No	Yes

The previously studied four isolates (PH202, PH470, PH278, PH372) are indicated in bold type.

[a] Complete lysis, final OD_{600} of 0.4 or less; partial lysis, final OD_{660} of 0.4 - 1.2; No lysis, final OD_{660} is the same as the control.

[b, c,] and [d] Based on electron microscopy (see Figure 3.49).

[e] PH196 was used as the indicator strain.

* Only contracted tail forms were found and they were used for measurement.

ND: Not determined

PH776, that underwent lysis after mitomycin C treatment. The isolate PH204 was not induced by 0.2 μg/ml mitomycin C treatment, as was the case for isolate PH202. Bacteriophages from isolates PH208, PH714, and PH776 had contractile tails and were classified as *Myoviridae*. However, the head and tail sizes of these phages were slightly different (Table 3.36).

3.2.8.2 Host range

The host range of bacteriophages from four additional isolates was examined by plaque assay using the indicator isolate PH196, which had previously been shown to be sensitive for PH470 and PH278. Zones of lysis (plaques) were observed for the lysates of isolates PH208 and PH776 (as with PH470 and PH278), but no plaques were visible for isolates PH204 and PH714 (as with isolates PH202 and PH372).

3.2.8.3 Phage DNA isolation

Attempts to isolate phage DNA from the four additional ovine serotype A2 isolates were unsuccessful, as they were with isolates PH202, PH470, PH278, and PH372.

3.2.9 The genome of RE type A phages

3.2.9.1 Prophage genome from *M. haemolytica* PHL213

Of the 11 different RE types (Figure 3.54), the RE type A prophage genome was identified within the unfinished genome sequence of the bovine serotype A1 *M. haemolytica* isolate PHL213 (www.hgsc.bcm.tmc.edu/microbial). Using RE cutting sites and Blast analysis (see section 2.5.8.2), three contigs joined in the order C150-

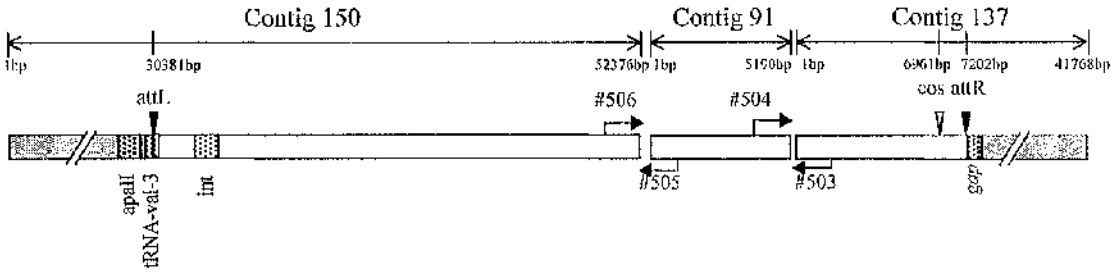


Figure 3.57 Diagrammatic representation showing position of the proposed prophage genome in *M. haemolytica* isolate PHL213 genomic data (contigs 150, 91, and 137 of the genome sequence as of 01/02/04). The range of the contigs are indicated by arrows and the numbers indicate the positions of the first and last bases of the contigs and the positions of the *attL*, *cos*, and *attR* sites. Bacterial genomic DNA is shown in grey, and prophage DNA in white. The locations of the primers used for PCR are indicated by small arrows and the locations of the phage integrase gene, *int*, and bacterial genes, *apaII*, $tRNA_{val}$, and *gap*, are shown by stippled lined boxes.

C91-C137 (as of 01/02/04) were identified to contain a segment that represents a P2-like prophage genome of PHL213. Positive PCR amplification of PH2Φ and PH376Φ with primer pairs (#506/#505 and #504/#503) designed in the end of each contig of *M. haemolytica* PHL213 (Figure 3.57) confirmed this arrangement and provided an approximate size of the unsequenced gaps between C150 and C91 (300 bp), and between C91 and C137 (250 bp).

3.2.9.2 Identification of attachment sites (*attL* and *attR*)

To identify the homologous attachment sites *attL* and *attR* within the bacterial genome, the sequences of the two end contigs C150 and C137 were aligned and compared. Two 197 bp homologous regions were identified in each contig (Figure 3.58). The homologous region in contig C150 starts from 14 bp upstream of the 3' end of the $tRNA_{val}$ gene and finishes 28 bp downstream of the start codon of the phage *int* gene. Based on the crossover points of other phages that always occur at the upstream end of the duplication (Campbell, 2003), the first 48 bp of identical residues (one mismatch) of the homologous regions of C150 and C137 were proposed as the attL and attR core sequences, respectively. The sequence of the proposed prophage attL site of *M. haemolytica* isolate PHL213 is similar to that of the *P. multocida* isolate LFB3 (Pullinger *et al.*, 2004) and that of the *Streptococcus thermophilus* isolate Sfi21 (Bruttin *et al.*, 1997) (Figure 3.59).

3.2.9.3 Cohesive end (cos) sequences

Examination of the *M. haemolytica* PHL213 prophage sequence revealed that it contains a 19-bp segment that is homologues to the corresponding cos sequence of *E. coli* P2 phages. The cos site indicates that the PHL213 prophage DNA is able to be

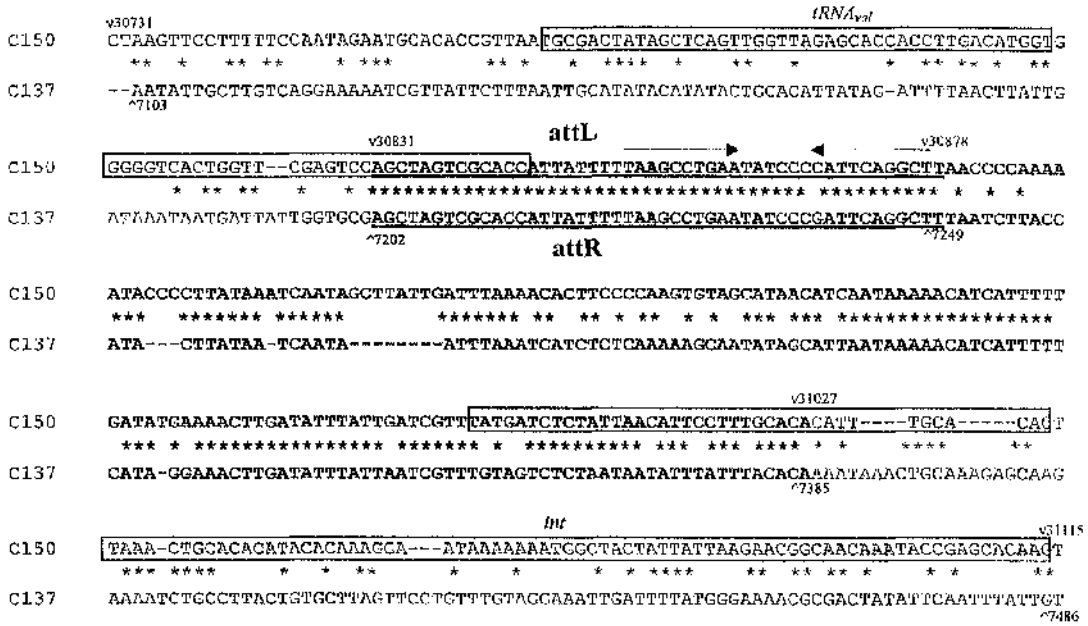


Figure 3.58 Comparison of the nucleotide sequences of contigs C150 and C137 of *M. haemolytica* isolate PHL213 in the vicinity of the attachment sites (attL and attR). The homologous regions are shown in bold type. The numbers above and below the sequences indicate the bases of each contig. The asterisks indicate residues that are identical in the two sequences. The underlined sequences of contigs C150 and C137 represent the attL and attR sites, respectively. The $tRNA_{val}$ gene and the 5' end of the phage integrase (*int*) gene are represented by boxes. The inverted repeats are indicated by arrows.

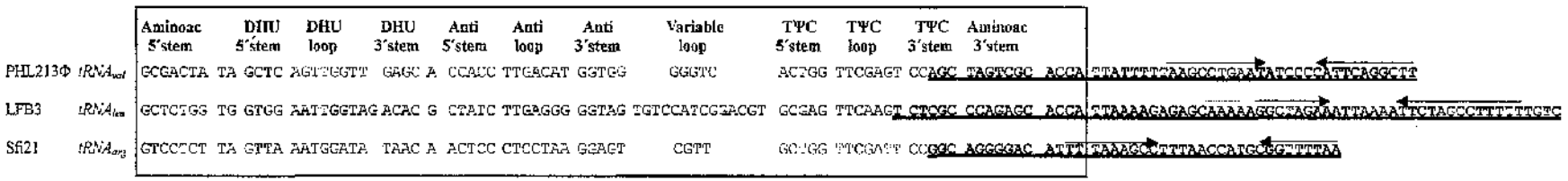


Figure 3.59 Comparison of prophage attachment sites (attL) in the various *tRNA* genes of *M. haemolytica* isolate PHL213, *P. multocida* isolate LFB3, and *S. thermophilus* isolate Sfi21. The box indicates the *tRNA* region and the core sequences of each phages are underlined. The inverted repeats are indicated by arrows.

circularized by pairing complementary cohesive ends (cos) after entering the host cell (see Figure 1.7). The cos sequences of PHL213Φ and P2 phage are compared in Figure 3.60.

3.2.9.4 Proposed RE type A phage genome

A putative PHL213Φ phage genome (Figure 3.61) was constructed based on the positions of the attL, attR, and cos regions (Figure 3.59 and 3.60). The three contigs C137, C91, and C150 were reverse complemented due to the nature of phage integration (see Figure 1.7). The proposed genome size of PHL213Φ phages was approximately 34,505 bp. The hypothetical RE profile of the proposed PHL213Φ genome (Figure 3.62A and Table 3.37) corresponds to the RE assays of PH2Φ (Figure 3.62B) and confirmed that RE type A phages represent the P2-like phage PHL213Φ.

3.2.9.5 Bioinformatic analysis and identification of phage encoding sequences

Both strands of the phage genome were analyzed to identify open reading frames (ORFs) and details are summarized in Table 3.38. The amino acid sequences deduced from the phage ORFs were compared to the contents of the phage database. Thirty three of 47 ORFs showed homology to published phage genes and functions could be provisionally assigned. Of these, 25 ORFs resembled proteins encoded by the P2 phage of the *Myoviridae*. Alignment of the *M. haemolytica* PHL213Φ and the E. coli P2 phage genome showed very similar genetic organisation (Figure 3.63). Thus, *M. haemolytica* RE type A phages represent new members of the P2 phage family.

```
PHL213Φ cos  :  GGCGAGGTGAGGGATAGAC
                ******* * ** *    **
P2 cos       :  GGCGAGGCGGGGAAAGCAC
```

Figure 3.60 Comparison of the cos sequences from *M. haemolytica* prophage PHL213Φ and *E. coli* prophage P2. The asterisks indicate residues that are identical in the two sequences.

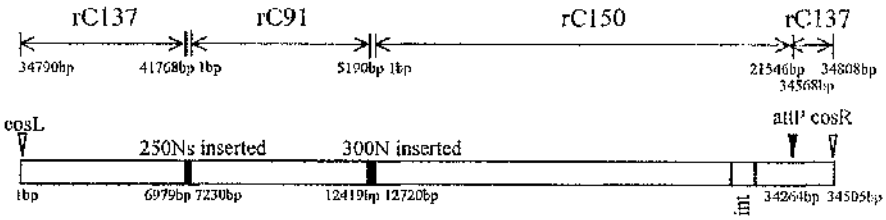


Figure 3.61 Diagrammatic representation showing the proposed PHL213Φ phage genome. The three contigs were reverse complemented. Based on the size of the PCR products the gaps between reverse complemented contigs rC137 and rC91 and rC91 and rC150 are approximately 250 and 300 nucleotides respectively (see section 3.2.9.1). The range of the contigs are indicated by the arrows and the numbers indicate the positions of the first and last base of each contig and the first base of cosL, attP (attachment sits of phage), and cosR sites. The phage genome starts at the first base of the cosL site and ends at the last base of the cosR site (Appendix I). The numbers below the proposed phage genome indicate the position of the first and last bases of the contigs in the proposed phage genome.

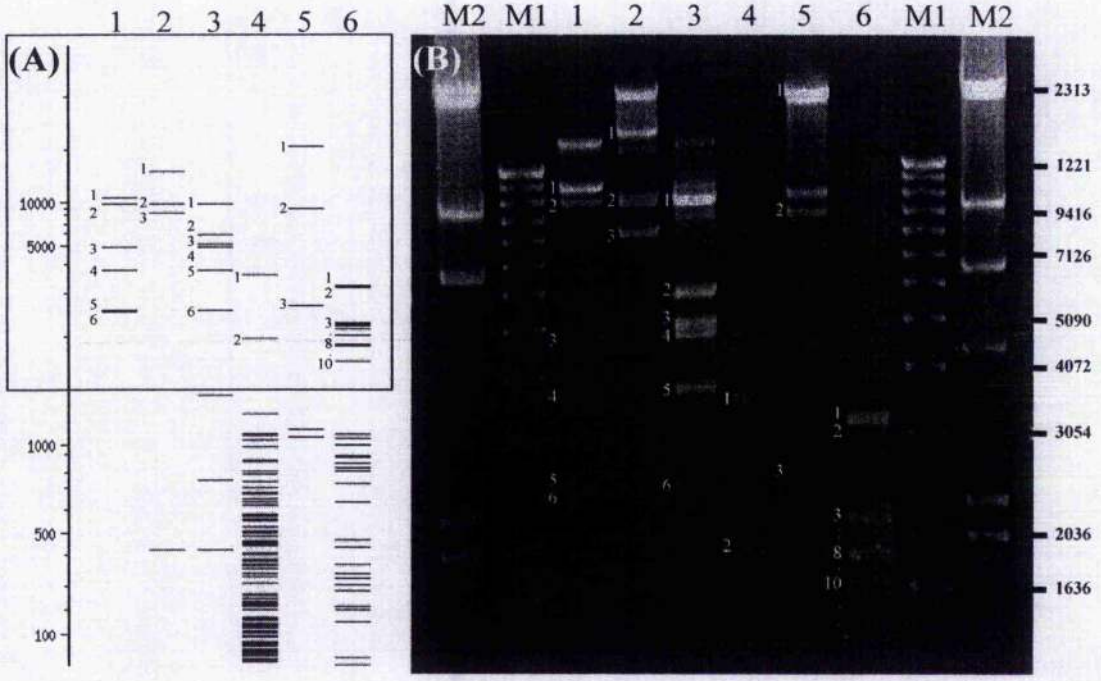


Figure 3.62 Predicted fragment patterns of the putative PHL213Φ in a virtual gel generated by NEB cutter (A) and restriction fragment patterns of PH2Φ DNA in a 0.7% agarose gel (B). The restriction enzymes used in lanes 1-6 were as follows: 1, HindIII; 2, ClaI; 3, HindIII & ClaI; 4, HpaII; 5, PstI; 6, RsaI. 1 Kb DNA markers and phage λ markers are shown in lanes M1 and M2 lanes respectively. Corresponding predicted and actual bands for each enzyme of PHL213Φ and PH2Φ, respectively, are numbered.

Table 3.37 The size and position of restriction fragments of putative PHL213Φ.

	HindIII (a/agctt)	ClaI (at/cgat)	HindIII & ClaI	HpaII (cc/gg)	PstI (ctgca/g)	RsaI (gt/ac)
1	10925 (9989-20913)	15843 (1-15843)	9988 (1-9988)	3591 (8387-11977)	20651 (12737-33387)	3235 (6319-9553)
2	9928 (1-9989)	9918 (24588-34505)	5855 (9989-15843)	1970 (16310-18279)	9033 (1-9033)	3212 (20330-23541)
3	4838 (29668-34505)	8328 (15844-24171)	5070 (15844-20913)	1205 (6553-7757)	2639 (9034-11672)	2234 (30411-32644)
4	3736 (25932-29667)	416 (24172-24587)	4838 (29668-34505)	1086 (15224-16309)	1118 (33388-34505)	2211 (10898-13108)
5	2514 (20914-23427)		3736 (25932-29667)	1083 (22590-23672)	1064 (11673-12736)	2155 (4164-6318)
6	2504 (23428-25931)		2514 (20914-23427)	1042 (13105-14146)		2120 (26613-28732)
7			1344 (24588-25931)	993 (30709-31701)		2020 (16787-18806)
8			744 (23428-24171)	868 (5456-6323)		1885 (24695-26579)
9			416 (24172-24587)	857 (12248-13104)		1861 (32645-34505)
10				805 (32254-33058)		1678 (28733-30410)
				⋮ ⋮		⋮ ⋮

Table 3.38 Bioinformatic analysis of ORFs of the putative PHL213Φ of *M. haemolytica*

orf	Coding region	Length (aa)	Protein description	Homologous gene (aa length)	% Identities/Range (aa)
Reverse complementary sequence of Contig 137 (34790bp-41768bp)					
cosL	34790-34808		Cohesive end	P2 cos	
orf1	34995-36035 (-)	346	Portal protein	P2 phage Q (344)	56/316
orf2	36044-37882 (-)	612	Terminase	P2 phage P (590)	55/578
orf3	37996-38823	275	Capsid scaffolding protein	P2 phage O (284)	43/273
orf4	38837-39865	342	Major capsid precursor	P2 phage N (357)	55/344
orf5	39875-40564	229	Terminase	P2 phage M (247)	40/220
orf6	40676-41191	171	Capsid completion protein	P2 phage L (169)	38/160
orf7	41188-41400	70	Essential tail gene	P2 phage X (67)	52/70
Reverse complementary sequence of Contig 91 (1bp-5190bp)					
orf1	1-86 (Partial)	28 (Partial)	Lysin	HP1 Phage Lys (186)	73/26
orf2	87-542	151	Affects timing of lysis	P2 phage LysB (141)	28/139
orf3	373-666	97	Hypothetical protein	P2 phage 81 (100)	25/48
orf4	691-912	73	Hypothetical protein	P2 phage 82 (74)	34/73
orf5	909-1394	161	Tail completion	P2 phage R (155)	41/158
orf6	1387-1845	152	Tail completion	P2 phage S (150)	36/146
orf7	2213-5125	970	Putative tail length	P2 phage T (815)	26/706
Reverse complementary sequence of Contig 150 (1bp-21546bp)					
orf1	234-839	201	Baseplate assembly protein	P2 phage V (211)	40/204
orf2	839-1174	111	Baseplate assembly protein	P2 phage W (115)	50/75
orf3	1171-2088	305	Baseplate assembly protein	P2 phage J (302)	53/300
orf4	2075-2707	210	Baseplate assembly protein	P2 phage I (176)	45/172
orf5	2710-4989	759	Tail fiber protein	P2 phage H (669)	39/218
orf6	5489-6871	460	Tail sheath	P2 phage FI (396)	44/395
orf7	6880-7386	168	Tail tube	P2 phage FII (172)	40/172
orf8	7465-7779	104	Tail protein	P2 phage E (91)	35/80
orf9	7983-8330	115		No homology	
orf10	8332-8769	145	Tail protein	P2 phage U (159)	47/129
orf11	8769-10007	412	Tail protein	P2 phage D (387)	50/374
orf12	10190-10999 (-)	269		No homology	
orf13	11031-11303 (-)	90		No homology	
orf14	11391-11801 (-)	135		No homology	
orf15	11819-12337 (-)	172		No homology	
orf16	12341-13084 (-)	247	CI repressor protein	D3 phage cI (223)	29/222
orf17	13151-13363	70	Cro protein	D3 phage cro (73)	32/49
orf18	13347-13631	91	Recombination protein	c2 phage e15 (179)	32/52
orf19	13462-13707 (-)	81	Not known	D3 phage orf56 (38)	38/54
orf20	13727-13897 (-)	56		No homology	
orf21	13840-14112	90		No homology	
orf22	14195-14527	110	Transcription regulator	Mu phage ner (75)	31/58
orf23	14540-14833	97		No homology	
orf24	14985-15227	80		No homology	
orf25	15224-15556	110		No homology	
orf26	15553-17913	786	DNA replication	P2 phage A (586)	37/424
orf27	17926-18258	77		No homology	
orf28	18258-18710	150	SSB protein	A118 phage ssb (160)	33/159
orf29	18721-19053	110	predicted 17.3 kDa protein	D29 phage gp55 (151)	40/70
orf30	19043-19312	89		No homology	
orf31	19287-19577	96		No homology	
orf32	19905-20057	50		No homology	
orf33	20307-21377 (-)	356	Integrase	P2 phage int (337)	71/251
attP	21546-21499		Phage attachment site		
Reverse complementary sequence of Contig 137 (34568bp-34808bp)					
cosR	34789-34808	19		P2 cos	

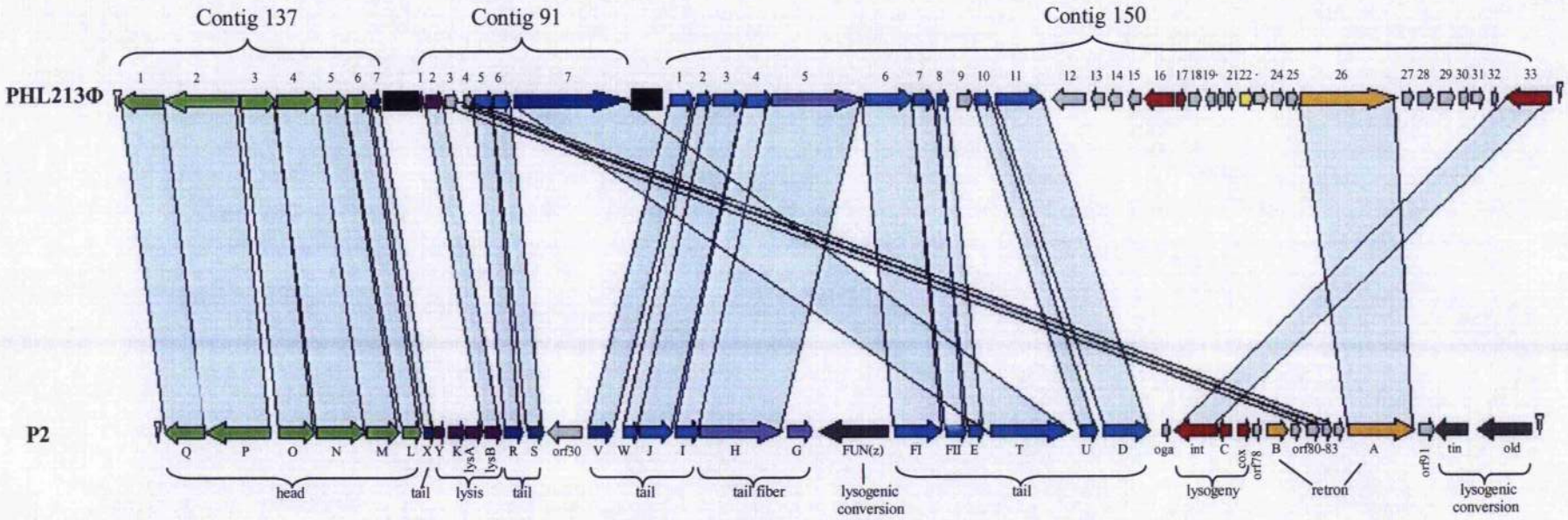


Figure 3.63 Alignment of the putative *M. haemolytica* PHL213Φ phage genome with the *E.coli* P2 phage genome. The phage genes are colour coded according to their attribution to phage modules. Lysogeny, red; DNA replication, orange; transcriptional regulation, yellow; DNA packaging and head, green; head-to-tail joining, brown; tail, blue; tail fiber, mauve; lysis, violet; lysogenic conversion, black; unattributed genes, grey. Genes linked by sequence similarity are connected by shading. Numbers above phage PHL213Φ represent ORFs in each contig and letters below phage P2 represent annotated genes. Details of the ORFs are shown in Table 3.38.

3.2.9.6 Prevalence of RE type A phages in *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

To determine whether the RE type A phages are present in other isolates of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*, PCRs were carried out with 32 *M. haemolytica*, six *M. glucosida*, and four *P. trehalosi* strains, using the primer pair #505/#506 (see Figure 2.34 and Table 2.19). A 700 bp (approximate size) DNA fragment was amplified not only in the two *M. haemolytica* isolates, PH2 and PH376, in which RE type A phages has been induced, but also in a further 19 *M. haemolytica* isolates and in one *M. glucosida* isolate (Figure 3.64, lane 1 and Table 3.31). However, a 1980 bp (approximate size) of DNA fragment was amplified in one *M. haemolytica* isolate, PH30 (Figure 3.64, lane 2) and no DNA fragments were amplified in the remaining 10 *M. haemolytica*, five *M. glucosida*, and four *P. trehalosi* isolates (Figure 3.64, lane 3 and Table 3.31).

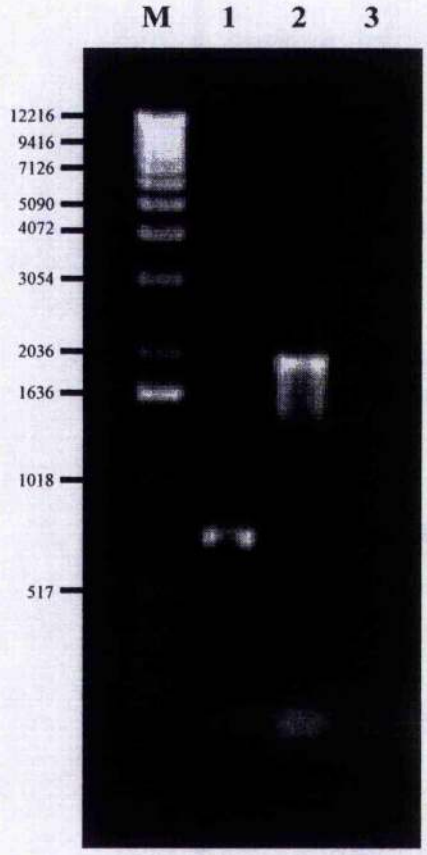


Figure 3.64 PCR assay with primer pair #505/#506. The *M. haemolytica* isolates in lanes 1 to 3 are as follows: 1, PH2; 2, PH30; 3, PH540. The one Kb DNA marker is in lane M.

CHAPTER 4: DISCUSSION

4.1 Comparative sequence analysis of nineteen genes of various functions of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Previous population genetic studies of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*, based on MLEE, provide an evolutionary framework with which to compare the evolution of selected genes (Davies *et al.*, 1997a; 1997b). The previous study (Davies *et al.*, 1997a) indicates that distinct evolutionary lineages of *M. haemolytica* are differentially adapted to cattle and sheep. Comparative sequence analysis has been shown to be an effective method of understanding the evolution of bacterial species (Field *et al.*, 1999; Kondrashov, 1999; Spratt & Maiden, 1999). Comparative sequence analysis of the leukotoxin operon of *M. haemolytica* has shown that host switching, together with horizontal DNA transfer and recombination, have played important roles in the evolution of the leukotoxin and associated proteins (Davies *et al.*, 2001; 2002).

In the present study, comparative sequence analyses of nineteen genes of various functions were carried out in *M. haemolytica* to investigate in further detail the roles of host switching, horizontal DNA transfer, and recombination in the evolution of *M. haemolytica*. The corresponding genes were also analysed in the closely related species *M. glucosida* and *P. trehalosi*, previously shown to be the origin of leukotoxin DNA in isolates of *M. haemolytica* (Davies *et al.*, 2001; 2002).

4.1.1 Comparative sequence analysis of genes encoding DNA repair and recombination enzymes (*recA*)

The nucleotide and inferred amino acid sequence analysis of *recA* (Figure 3.1) indicated that this gene is highly conserved in *M. haemolytica* and there was no

evidence of recombinational exchange. This is not surprising because the RecA protein is involved in many important functions related to cell viability including recombination and DNA repair (Kowalczykowski *et al.*, 1994; Radman *et al.*, 2000) and is highly conserved within diverse bacterial species (Karlin *et al.*, 1995). The *recA* tree topology agrees with 16S rRNA sequence analysis which shows that *M. glucosida* has diverged from *M. haemolytica,* and *P. trehalosi* is further diverged from both *M. haemolytica* and *M. glucosida* (Davies *et al.*, 1996).

4.1.2 Comparative sequence analysis of genes encoding metabolic enzymes (*aroA*, *asd*, *galE*, *gap*, *gnd*, *g6pd*, *mdh*, *mtlD*, and *pmm*)

Genes encoding metabolic enzymes are subject to either neutral or purifying selection since variations in these genes have no advantage or even lead to fatal consequences (Musser, 1996). This was also the case for metabolic enzyme genes of *M. haemolytica*. The *aroA*, *asd*, *galE*, *gap*, *gnd*, *g6pd*, *mdh*, *mtlD*, and *pmm* genes of *M. haemolytica* have low d_N values (0.00 to 0.60) due to purifying selection based on functional constraint. The *asd*, *gnd*, *g6pd*, and *mtlD* genes also have very low d_S values (0.26 to 0.45). This supports the view that synonymous substitutions are also subject to purifying selection due, for example, to codon bias (Jordan *et al.*, 2002). The *aroA*, *gap*, *mdh*, and *pmm* genes of *M. haemolytica* have high d_S values (2.93 to 9.44) but nucleotide sequence analysis of these genes indicates that this variation is due to horizontal DNA transfer and recombination. The high d_S/d_N values of *aroA* (12.61), *gap* (58.69), *mdh* (28.9), and *pmm* (15.73), in particular, indicate that purifying selection is acting on these genes and constraining amino acid changes.

Because the *aroA*, *gap*, and *pmm* genes showed evidence of recombination in the preliminary analysis of 12 isolates, a total 32 *M. haemolytica*, six *M. glucosida*, and

four *P. trehalosi* isolates were examined. The *mdh* gene showed limited evidence of recombination in the preliminary analysis of 12 isolates and additional isolates were not examined.

In the case of *aroA*, isolates of ETs 10 to 14 and ET 18 (*aroA3*-type alleles) contain an almost identical recombinant segment that has been derived from an unknown source (Figure 3.3). The donor could not be identified, but it must be a closely related species because the recombinant segment is less divergent than the corresponding region of *P. trehalosi*. Since isolates of ETs 10 and 11 (MLEE lineage A2), 12 to 14 (MLEE lineage B), and 18 (MLEE lineage B) are distantly related, the recombinant segment probably indicates a 'hot spot' region which is susceptible to horizontal DNA transfer and recombination. This recombinant segment acts as a donor for a second recombinant event, which occurred in isolates of ETs 19 to 22 (*aroA2.1*).

In the case of *gap*, the identical recombinant segments in *M. haemolytica* isolates of ETs 19 to 22 (*gap3.1*) have most likely been derived from the same source, *M. glucosida* (Figure 3.9). Since isolates of ETs 19 to 22 are closely related, the common recombinant segment probably reflects a common origin rather than a 'hot spot' region. In addition, *M. glucosida* isolates of ETs 10 (PH574, *gap4.1*) and 16 (PH290, *gap2.4*) contain recombinant segments or entire genes that are probably derived from *M. haemolytica*, suggesting that interspecies recombination has also occurred in this gene.

In the case of *pmm*, distantly related isolates of ETs 5 (MLEE lineage A1), 6 to 11 (MLEE lineage A2), and 18 (MLEE lineage B) have very similar sequences (*pmm4*-type alleles) that have numerous nucleotide sites in common with the *pmm3*-type

alleles of *M. glucosida* (Figure 3.19). This suggests that *pmm4*-type alleles may originally have been derived, by assortative and/or intragenic recombination, from *M. glucosida* isolates. In addition, the unusual divergent sequences of *pmm* (*pmm5.1*) as well as *mdh* (*mdh2.1*) in isolate PH588 suggests that this isolate is more susceptible to horizontal DNA transfer and recombination than other *M. haemolytica* isolates.

Since the Gnd, G6pd, Mdh, and MtlD enzymes have been used in the previous MLEE study of *M. haemolytica* (Davies *et al.*, 1997a), their MLEE allele types were compared with the nucleotide and inferred amino acid sequence types of the corresponding genes. In particular, the *mtlD* gene of additional isolates of *M. haemolytica* and *M. glucosida* that showed variation in the MLEE study were sequenced. The MLEE allele types of each enzyme reflect the nucleotide and amino acid sequence variations, but MLEE analysis was less discriminating than nucleotide and amino acid sequence analysis.

In conclusion, intra- and interspecies recombinational exchanges have not occurred frequently in the metabolic enzyme genes (as they have in surface-exposed protein genes). Nevertheless, they are important mechanisms for promoting allelic diversity in certain metabolic enzyme genes of *M. haemolytica*. However, amino acid replacement within these genes is highly constrained due to strong purifying selection. The low rate of recombination within the majority of housekeeping enzyme genes suggests that the dendrogram generated by the MLEE data provides an accurate indication of the phylogenetic relationships of *M. haemolytica* isolates. This is also indicated by the fact that the tree topology based on the concatenated sequences of seven highly conserved genes (Figure 3.46) is very similar to that based on MLEE (Figure 2.1). Low rates of recombinational exchanges in metabolic enzyme genes are also seen in other clonal bacterial populations such as *Salmonella* and *E. coli*

(Boyd *et al.*, 1994; Nelson & Selander, 1992). However, frequent recombination in housekeeping enzyme genes has been described in non-clonal bacterial populations such as *Neisseria* species (Feil *et al.*, 1995; Feil *et al.*, 1996; Zhou *et al.*, 1997).

4.1.3 Comparative sequence analysis of genes encoding secreted proteins (*gcp*)

The *gcp* gene was shown to be highly conserved in *M. haemolytica* because only two nucleotide and two amino acid polymorphic sites were found (Figure 3.21). This finding is clearly very different from another secreted protein, the leukotoxin, which is highly divergent and has a complex mosaic structures due to extensive horizontal DNA transfer and intragenic recombination (Davies *et al.*, 2001; 2002). In addition, the *M. haemolytica* bovine serotype A1 isolate PH2 and the *M. glucosida* isolate PH344 have identical Gcp amino acid sequences and the *P. trehalosi* isolate PH246 has a similar Gcp amino acid sequence (88 % identity) (Figure 3.21B). Previous studies have shown that *M. glucosida* possesses the *gcp* gene but has no Gcp activity, whereas *P. trehalosi* has no *gcp* gene or Gcp activity (Abdullah *et al.*, 1990; Lee *et al.*, 1994; Watt *et al.*, 1997). However, a high percentage of amino acid identity (76 to 100 %) was also detected in Gcp in various members of the *Pasteurellaceae* (Table 3.2) and may reflect an important common role for this protein in bacterial pathogenesis.

4.1.4 Comparative sequence analysis of genes encoding periplasm-associated proteins (*plpA*, *plpB*, *plpC*, and *plpD*)

The three lipoprotein genes, *plpA*, *plpB*, and *plpC*, are highly conserved among nine *M. haemolytica* isolates. In contrast, the *plpA*, *plpB*, and *plpC* genes of

M. haemolytica isolate PH196 have diverged from the corresponding genes of other *M. haemolytica* isolates due to a large recombinant segment derived from *M. glucosida*. However, the high d_s/d_N values of *plpA* (97.3), *plpB* (20.4), and *plpC* (19.7) of *M. haemolytica* suggest that amino acid replacements are strongly constrained by purifying selection.

Although the *plpA*, *plpB*, and *plpC* genes of *M. haemolytica* and *M. glucosida* were successfully amplified and sequenced, only the *plpA* gene was amplified in *P. trehalosi*. The other two genes could not be amplified in *P. trehalosi* despite many attempts. This suggests that *P. trehalosi* only contains the *plpA* gene. This situation is common in other members of the *Pasteurellaceae* which also do not contain *plpB* and *plpC* (Table 3.2). Of seven selected members (Table 3.2) of the *Pasteurellaceae*, only *M. haemolytica* and *A. pleuropneumoniae* contain contiguous *plpA*, *plpB* and *plpC* lipoprotein genes whereas the remaining species contain only a single lipoprotein gene that is similar to *plpA*.

Sequence analysis of the fourth lipoprotein gene, *plpD*, indicates that this gene is also highly conserved in *M. haemolytica* with low d_s (0.16) and d_N (0.35) values. The location of PlpD in the cell wall is not known, but it is probably located in the periplasmic space since the C-terminal end of PlpD has significant identity to the corresponding periplasm-spanning region of OmpA (Pautsch & Schulz, 1998). In addition, the N-terminal end of PlpD contains typical lipoprotein sequences (Nardini *et al.*, 1998).

Cooney & Lo (1993) examined the inferred amino acid sequences of PlpA, PlpB, and PlpC and suggested that all three lipoproteins should be localized to the inner membrane in spite of a high antibody response to these proteins. They explained

that the three lipoproteins may be released as a result of cell lysis or membrane turnover and stimulated an antibody response due to their highly immunogenic nature (Cooney & Lo, 1993). Lipoproteins are extremely immunogenic because ester-linked fatty acids are located at the amino-terminal end of the lipoproteins (Melchers *et al.*, 1975). Based on these views, it seems reasonable that the PlpA, PlpB, PlpC, and PlpD proteins are highly conserved in *M. haemolytica* because there is no selective pressure driving variation of these proteins from, for example, the immune system.

4.1.5 Comparative sequence analysis of genes encoding outer membrane proteins (*ompA*, *tbpB*, *tbpA*, and *wza*)

In contrast to the various genes described above the surface-exposed protein genes, *ompA*, *tbpB*, *tbpA*, and *wza*, are highly divergent due to extensive recombinational exchanges. The patterns of amino acid variation in different regions of these proteins vary, indicating that different regions are under different selection pressures. Presumably, this variation in related to different structures and functions. Therefore, the amino acid diversity of different proteins depends not only on the degree of nucleotide diversity generated by recombination but also on evolutionary constraints imposed by the function of the encoded gene product. In general, conserved amino acid regions correspond to important structures that are maintained by functional constraint (purifying selection), whereas highly variable amino acid regions reflect diversification that is in some way advantageous to the organism.

4.1.5.1 Heat modifiable outer membrane protein gene (*ompA*)

Nucleotide variation of *ompA*. The *ompA1*- and *ompA2*-type alleles are associated

exclusively with bovine and ovine isolates of *M. haemolytica*, respectively, and have very different nucleotide sequences (Figure 3.27). There was no evidence that intragenic recombination has occurred between these two subclasses. However, the association of identical, or nearly identical, *ompA1*- and *ompA2*-type alleles with divergent lineages of *M. haemolytica* clearly suggests that assortative (entire gene) recombination has been an important factor in the evolution of *ompA* in these isolates (Table 3.20). In contrast, the *ompA4*-type alleles associated with *M. haemolytica* isolates of ETs 12 to 15 (serotypes A7 and A13) have recombinant segments (green) that have probably been derived from *M. glucosida.*

Molecular evolution of OmpA in relation to structure and function. Comparison with secondary structure prediction models of the *H. influenzae* OmpA protein (Webb & Cripps, 1998) and with 3D structural models of the *E. coli* OmpA protein (Arora *et al.*, 2001; Pautsch & Schulz, 1998; Pautsch & Schulz, 2000; Webb & Cripps, 1998) indicate that the transmembrane domain of OmpA in *M. haemolytica* and *M. glucosida* consists of eight membrane-spanning regions and four relatively long, mobile, hydrophilic surface-exposed loop regions (Figure 3.29). Sequence analysis of the *ompA* gene of *M. haemolytica* and *M. glucosida* indicates that different regions of the gene have different nucleotide and amino acid diversity. The four surface-exposed loop regions are highly variable, whereas non-loop regions are highly conserved. Different selection pressures acting on the loop and non-loop regions account for this variation. The d_S/d_N values of < 1 (0.34 to 0.77) for the surface-exposed loop regions indicates that diversifying selection is acting on these parts of the molecule. In contrast, amino acid replacement is highly constrained within the non-loops regions ($d_S/d_N = 35.17$) because these parts of the molecule correspond to the membrane-spanning and periplasmic domains and cannot tolerate excessive amino acid change. Similar patterns of nucleotide and amino acid diversity have also been

described in other OmpA proteins (Duim *et al.*, 1997; Yuan *et al.*, 1989).

However, the OmpA protein of *M. haemolytica* is the first in which a host-specific pattern of sequence variation has been described. The exclusive association of the OmpA1- and OmpA2-type proteins with bovine and ovine isolates of *M. haemolytica*, respectively, together with evidence that the *ompA1*- and *ompA2*-type genes have undergone horizontal transfer between isolates of divergent phylogenetic lineages, indicates that OmpA is under strong selective pressure from the host and plays an important role in host-pathogen relationships. It has been shown that the OmpA proteins of other pathogens function as ligands, are involved in binding to specific host cell receptor molecules, and play a role in adherence and colonization (Dabo *et al.*, 2003; Hill *et al.*, 2001; Millman *et al.*, 2001; Prasadarao, 2002; Reddy *et al.*, 1996; Torres & Kaper, 2003). In view of this, it is proposed that the four surface-exposed loops of the *M. haemolytica* OmpA protein are involved in binding to host-specific receptors. This would account for the variation in the surface-exposed loop regions between the bovine OmpA1-type and ovine OmpA2-type proteins and also for the amino acid conservation within each class. Clearly, the surface-exposed loops of OmpA need to be different in bovine and ovine isolates if cattle and sheep have different receptor molecules, but they would also need to be conserved within each class if they are involved in binding to specific bovine (OmpA1-type) or ovine (OmpA2-type) receptor molecules. The locations of the four hypervariable domains, HV1 to HV4, at the distal ends of the corresponding loops (Figure 3.30) provide further evidence to support the hypothesis that these regions are involved in receptor recognition and binding. Confirmation of this hypothesis will require the production of genetically modified isolates and the development of appropriate *in vitro* adherence assays. In a similar way, receptor binding of the variable loop regions of the related Opa protein determines tissue tropism in *Neisseria* (Bos *et al.*, 1999; Popp *et al.*,

1999; Virji *et al.*, 1999).

***P. trehalosi* produces two OmpA homologs.** The present study has demonstrated that *P. trehalosi* produces two OmpA homologs which are encoded by different tandemly arranged genes, *ompA'* and *ompA"* (Figure 3.32). In contrast, *M. haemolytica* and *M. glucosida* produce only one OmpA protein. A retrospective examination of the OMP profiles of a wide range of *P. trehalosi* isolates (Davies & Quirie, 1996) indicated that the expression of two OmpA homologs is common in this species. However, it is not unusual in Gram-negative organisms since the presence of two OmpA homologs has been described in *Aeromonas salmonicida* (Costello *et al.*, 1996) and *Haemophilus ducreyi* (Klesney-Tait *et al.*, 1997). It seems likely that a common underlying mechanism is responsible for generating tandem *ompA* genes in certain bacterial species and that the expression of two OmpA homologs provides a selective advantage to these organisms. The two OmpA homologs of *P. trehalosi* have significant amino acid variation in their N- terminal regions, particularly in the surface-exposed regions, whereas the C- terminal regions that are located in the periplasmic membrane are highly conserved (Figure 3.33). This is similar to the two OmpA proteins of *A. salmonicida* and *H. ducreyi* (Costello *et al.*, 1996; Klesney-Tait *et al.*, 1997). Whether the differences in the surface-exposed loops of the two OmpA proteins reflect functional differences during pathogenesis remains to be determined.

It has been suggested that the two *ompA* genes of *A. salmonicida* and *H. ducreyi* have arisen by gene duplication (Costello *et al.*, 1996; Klesney-Tait *et al.*, 1997). Based on the relatively low level of homology between the two OmpA proteins of *A. salmonicida*, Costello *et al.* (1996) concluded that such a gene duplication event occurred in the distant evolutionary past. However, the data presented here suggest an alternative possibility. The presence of very different *ompA'*-type alleles in

isolates PH68, PH246, and PH252 (*ompA8.1* to *8.3*) compared to that in isolate PH254 (*ompA9.1*), and the occurrence of very similar *ompA"*-type alleles in the same four isolates (*ompA10.1* to *10.4*), indicate that one or other, or both, of the *ompA8*- or *ompA9*-type alleles has been acquired by horizontal transfer (Figure 3.28). The low frequency of occurrence of the *ompA9.1* allele suggests that this allele has probably replaced an *ompA8*-type allele, not *vice versa*, by horizontal gene transfer. In addition, the low level of similarity between the *ompA'* and *ompA"* alleles, in contrast to the high degree of similarity among the *ompA8*- and *ompA10*-type alleles, is consistent with acquisition by horizontal transfer rather than by gene duplication. If one of the *ompA* genes had arisen by duplication in the distant past, we would also expect to see more divergence among alleles representing each of the *ompA* types, but this is not the case. Therefore, horizontal DNA transfer, rather than gene duplication, might account for the second *ompA* gene in this, and probably other, bacterial species.

4.1.5.2 Transferrin binding protein genes (*tbpB* and *tbpA*)

Nucleotide variation of *tbpB* and *tbpA*. The contiguous transferrin-binding protein genes, *tbpB* and *tbpA*, are highly divergent in *M. haemolytica* due to extensive horizontal DNA transfer and recombination (Figures 3.41 and 3.42). The diversity of these two genes varies among the different groups of *M. haemolytica* (Figure 3.41). For example, *M. haemolytica* isolates of groups 1 to 4 have similar *tbpB* and *tbpA* genes except for short recombinant segments due to intragenic recombination. *M. haemolytica* isolates of group 5 have different *tbpB* genes from *M. haemolytica* isolates of groups 1 to 4 because they have been derived by assortative recombination from *M. glucosida* isolates of groups 6 to 8. Bovine *M. haemolytica* A2 isolates of groups 9 (ET 18) and 10 (ETs 17 and 21) have divergent *tbpB* genes that have been derived from unknown sources (Figures 3.41 and 3.42). The *tbpB* tree topology

suggests that the group 9 and 10 donors are less diverged from *M. haemolytica* isolates of group 1 than are those of *P. trehalosi* isolates (Figures 3.36). Ovine *M. haemolytica* isolates of groups 12 and 13 (ETs 16, 19, 20, and 22) have extremely divergent *tbpB* genes (*tbpB6.1* to *tbpB6.3*) that have been derived from unknown sources and are, surprisingly, more divergent than *P. trehalosi* isolates (Figure 3.36). However, these isolates have much less nucleotide diversity in *tbpA* (3.6 %) than in *tbpB* (37.5 %), with respect to isolates of group 1, suggesting that the *tbpA* gene has been derived from isolates that are more closely related to *M. haemolytica* isolates of group 1 than to the donors of the *tbpB* gene (Figures 3.41 and 3.42). The *tbpA* gene of *P. trehalosi* isolates were also shown to contain recombinant segments that have been derived from *M. glucosida* (Figure 3.38).

Molecular evolution of TbpB and TbpA in relation to structure and function. The TbpB protein of *M. haemolytica*, which is thought to be fully surface-exposed (Gray-Owen & Schryvers, 1996), has extremely high amino acid diversity (49.8%) and a relatively low d_S/d_N ratio (4.53) throughout the protein. This suggests that heterogeneity of TbpB has a selective advantage to the organism, such as evasion of the host immune system. The TbpB protein of *M. haemolytica* serotype A1 isolates appears to be readily accessible to antibody at the cell surface (Potter *et al.*, 1999). The comparative sequence data indicate that TbpB heterogeneity in *M. haemolytica* is due to extensive horizontal DNA transfer and recombination (Figure 3.35). In particular, the *tbpB* genes of bovine and ovine A2 isolates (except for the ovine isolates of ET 21, group 4) (Figure 3.41) have been acquired from different unknown sources, likely to be species more divergent than *M. glucosida* (groups 9 and 10) or *P. trehalosi* (groups 12 and 13). Recombinational exchange is thought to confer a selective advantage on pathogens by generating antigenic variation in cell surface antigens (Hobbs *et al.*, 1994; Li *et al.*, 1994; Mooi & Bik, 1997; Rich *et al.*, 2001;

Spratt *et al.*, 1992). Surface-exposed protein genes of *M. haemolytica* such as *ompA* (this study), *tbpA* (this study), and *wza* (this study), and secreted protein genes of *M. haemolytica* such as *lktA* (Davies *et al.*, 2001), show evidence of recombinational exchange. However, *M. haemolytica* TbpB (49.8%) has much higher amino acid variation than OmpA (7.4%), TbpA (18.3%), Wza (14%), and LktA (16%). In contrast to TbpA, TbpB is not essential for the internalisation of transferrin-bound iron (Anderson *et al.*, 1994). In bovine isolates of *P. multocida* it has been shown that TbpA alone is capable of mediating iron acquisition from bovine transferrin without the involvement of TbpB (Ogunnariwo & Schryvers, 2001). Nevertheless, most pathogens produce TbpB as well as TbpA because TbpB plays a role in increasing the efficiency and specificity of the transferrin receptor (Anderson *et al.*, 1994; Boulton *et al.*, 1998; Gray-Owen *et al.*, 1995; Irwin *et al.*, 1993).

As described previously (Ogunnariwo *et al.*, 1997), the TbpB protein of *M. haemolytica* has several identical amino acid sequences in both the N- and C-terminal halves of the molecule (Figure 3.37). This indicates that TbpB has a symmetrical structure, as has transferrin (Tf). Tf is composed of two homologous lobes (C and N lobes) and each lobe can bind a single Fe molecule (Baker & Lindley, 1992). Mszarin *et al.* (1995) proposed that the TbpB receptor probably evolved in parallel to Tf in order to bind both Tf lobes (Mazarin *et al.*, 1995). This hypothesis has been reinforced by the binding of both lobes of Tf by TbpB in *M. haemolytica* serotype A1 (Ogunnariwo *et al.*, 1997) as well as other bacteria (Alcantara & Schryvers, 1996; Boulton *et al.*, 1999; Renauld-Mongenie *et al.*, 1997). Recently, two domains involved in transferrin binding were identified in the N- terminal half of TbpB from the meningococcal isolate B16B6 (Renauld-Mongenie *et al.*, 2004). The corresponding domains were also identified in the conserved regions of the inferred amino acid sequences of TbpB in *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

suggesting a general functional role for these domains. Furthermore, single amino acids in each domain, previously identified by site-directed mutagenesis in *N. meningitidis* (Renauld-Mongenie *et al.*, 2004), were conserved in all three species.

Consistent with other bacterial TbpA proteins, *M. haemolytica* TbpA proteins are less divergent than the TbpB proteins and have similar amino acid sequences to the TbpA proteins of *M. glucosida* and *P. trehalosi* (Figure 3.40). This is because TbpA is a transmembrane protein and is less surface-exposed than TbpB (Gray-Owen & Schryvers, 1996). In addition, TbpA is more important for transferrin-iron acquisition than TbpB (Cornelissen & Sparling, 1996; Cornelissen *et al.*, 1997; Rokbi *et al.*, 1993). To understand the structure of TbpA in *M. haemolytica*, *M. glucosida*, *P. trehalosi*, sequence alignments and secondary structure predictions were carried out in relation to the previously proposed gonococcal TbpA model (Boulton *et al.*, 2000). The N-terminal plug domain and the C-terminal β-barrel domain were identified in the TbpA proteins of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Figure 3.40). The C-terminal β-barrel domain is composed of 11 surface-exposed loop regions and 21 membrane-spanning regions (the last and 22nd β strand is missing because the sequence was partial). The plug occludes the pore formed by the barrel until transferrin binds the TbpA from outside the bacterial cell (Ferguson & Deisenhofer, 2002; Usher *et al.*, 2001).

As expected, most amino acid diversity occurred in the surface-exposed loop regions, but two additional variable segments were identified in the plug and membrane-spanning regions, respectively. The first additional variable segment is located within the plug region associated with TbpA of ovine *M. haemolytica* isolates of groups 12 (*tbpA4.1*) and 13 (*tbpA4.2*) and *P. trehalosi* isolates of group 11 (*tbpA3.1* to *tbpA3.3*) and is due to its location adjunct to TbpB (Figure 3.40). These

hypervariable sequences in the N-terminal plug region were continuations from TbpB and interestingly stopped just before the TonB box. This suggests that amino acid replacement is highly constrained within the TonB box that is required for interaction with TonB. The second additional variable region (V2) is located in a membrane-spanning domain, but the reason for this is not clear.

Recently, several functional domains were identified among eleven surface loops in the gonococcal TbpA protein (Boulton *et al.*, 2000). Loops 3, 4, 5, and 8 were associated with transferrin binding, and loops 2, 9, and 11 were important for TbpA-TbpB interaction. In the case of the *M. haemolytica* TbpA protein, the transferrin-binding domains corresponding to loops 3, 5, and 8 have high amino acid diversity (23.8% to 28.1%) and low d_S/d_N ratios (3.5 to 6.2), whereas loop 4 has relatively low amino acid diversity (15.7%) and high d_S/d_N ratio (26.6). Loop 4 is homologous to the corresponding regions of other bacterial TbpA proteins (Cornelissen *et al.*, 2000) suggesting that loop 4 is likely to be constrained by the common function of transferrin binding. Bovine *M. haemolytica* isolates of group 1 and ovine *M. haemolytica* isolates of groups 2 to 5 have identical, or almost identical, sequences in these transferrin binding domains (Figures 3.41 and 3.42). This suggests that the transferrin binding receptors of bovine *M. haemolytica* serotype A1 isolates of group 1 and ovine *M. haemolytica* isolates of groups 2 to 5 do not discriminate between ovine and bovine transferrin (Yu *et al.*, 1992), although they are able to discriminate between ruminant transferrin and transferrin from other species (Gray-Owen & Schryvers, 1993; Ogunnariwo & Schryvers, 1990; Schryvers & Gonzalez, 1990; Yu & Schryvers, 1994). However, bovine and ovine *M. haemolytica* serotype A2 isolates may still be able to discriminate between ovine and bovine transferrin because they do have sequence variation in the transferrin binding domains.

In contrast to these transferrin binding domains, TbpA-TbpB interaction domains (loops 2, 9, and 11) are relatively conserved (16.7 to 17.2%) and have relatively high d_S/d_N ratios (9.8 to 17.0) among *M. haemolytica*, *M. glucosida*, and *P. trehalosi*. These loop regions are also more homologous with the corresponding loop regions of other bacterial TbpA proteins than are other loops (Cornelissen *et al.*, 2000), indicating conserved TbpA-TbpB interactions. Indeed, there is evidence for conserved interactions between heterologous receptors from different bacterial species such as the human pathogens *N. meningitidis* and *H. influenzae*, and the porcine pathogen *A. pleuropneumoniae* (Fuller *et al.*, 1998; Gray-Owen *et al.*, 1995; Irwin *et al.*, 1993). The most conserved loop among *M. haemolytica*, *M. glucosida*, and *P. trehalosi* TbpA was loop 6. The low amino acid diversity (4.5%) and high *ds/dn* ratio (30.8) clearly suggests that amino acid replacement is highly constrained within this loop, but its function has yet to be identified.

Vaccine candidates. Recombinant TbpB from bovine *M. haemolytica* serotype A1 has previously been shown to be capable of inducing an effective immune response in cattle (Potter *et al.*, 1999). In addition, the antibodies produced in response to this recombinant TbpB appeared to cross-react with the TbpB from several different serotypes, suggesting that recombinant TbpB bovine *M. haemolytica* serotype A1 may serve as a broadly crossreactive vaccine (Potter *et al.*, 1999). Nucleotide sequence analysis of *tbpB* from different isolates of *M. haemolytica* indicates that the inferred amino acid sequence of TbpB in isolates of groups 1 and 4, including the major bovine A1 (ET 1) and ovine A2 (ET 21) lineages, respectively, is relatively conserved (0.2 to 5.9 % amino acid diversity). However, the corresponding sequence of TbpB in *M. haemolytica* isolates of group 13, including another major ovine A2 lineage (ET 22), is highly diverged from TbpB in *M. haemolytica* isolates of groups 1 and 4 (41.8 to 42.8% amino acid diversity). Therefore, a vaccine based on TbpB from a single

isolate is highly unlikely to protect the host from all other *M. haemolytica* isolates. In addition, continuous recombinational exchanges in *tbpB* may limit TbpB utility as a vaccine antigen for long periods.

4.1.5.3 Capsule transport protein gene (*wza*)

Nucleotide variation of *wza*. The *wza* gene of *M. haemolytica* was selected because it is located within the conserved capsular transport region but is adjacent to the capsular biosynthesis region (Figure 1.2) (Roberts, 1996). The homologous gene (*ctrA*) of *N. meningitidis* was shown to be conserved among different meningococcal serogroups (Frosch *et al.*, 1992). However, nucleotide sequence comparison of the *wza* gene from *M. haemolytica* isolates representing different lineages indicates that the gene is highly divergent in *M. haemolytica* due to frequent intragenic and assortative (entire gene) recombinational exchanges

The sequence diversity pattern correlates strongly with the capsular types of the isolates (Figures 3.43 and 3.44). For example, class I sequences are associated exclusively with *M. haemolytica* serotypes A1 and A6, class II with *M. haemolytica* serotypes A2, A5, A7, and A12, class III with *M. haemolytica* serotypes A14 and A16 and *M. glucosida* serotype A11, class IV with *M. haemolytica* serotype A13, and class V with *M. haemolytica* serotype A8 and A9 and untypeable *M. glucosida* (UG3). These findings suggest that *wza* diversity in *M. haemolytica* is related to the location of *wza* next to the highly polymorphic capsule biosynthetic region (Figure 1.2) that is responsible for the different capsular types (Lo *et al.*, 2001). Similarly *kpsS* and *kpsT*, which are located in the conserved capsule transport (region 1) and phospholipid substitution (region 3) regions of group II *E. coli* capsule clusters, respectively, showed a marked divergence at the C -termini (Pavelka *et al.*, 1994; Petit

et al., 1995). Since the 3′ ends of *kpsS* and *kpsT* (complementary gene) are located at the junctions between regions 1 and 2 and 2 and 3, respectively, the divergence of these genes is thought to have arisen through recombination involving the polymorphic capsule biosynthetic region. In addition, a housekeeping gene, *gnd*, has also been shown to be highly variable due to its location next to the highly polymorphic *rfb* genes that are involved in LPS biosynthesis and are subject to diversifying selection (Bisercic *et al.*, 1991; Nelson & Selander, 1994; Thampapillai *et al.*, 1994).

Molecular evolution of Wza in relation to structure and function. In Gram-negative bacteria, a complex capsule transport system exists to allow the transport of high molecular weight capsular polysaccharide products across both the inner and outer membranes of the cell wall (Paulsen *et al.*, 1997; Whitfeld & Roberts, 1999). As a member protein of this system, the *wza* gene was highly divergent in *M. haemolytica* and *M. glucosida* due to a large number of recombinational segments derived from each other or from unknown sources (Figure 3.43). However, most amino acid replacement was highly constrained (d_s/d_N of 19.07 in *M. haemolytica* and 20.50 in *M. glucosida*, respectively), suggesting an important function for Wza in the translocation of capsular polysaccharide molecules across the outer membrane (DeRosa *et al.*, 2000). These data also suggest that the high degree of nucleotide diversity observed in the *wza* gene of *M. haemolytica* and *M. glucosida* is primarily due to its close proximity to the capsule biosynthetic region and is not necessarily because of positive selection acting on Wza. In fact, comparative sequence analysis of Wza in other Gram-negative bacteria indicates that the Wza protein is highly conserved (Paulsen *et al.*, 1997; Rahn *et al.*, 1999; Whitfeld & Roberts, 1999). Therefore, different capsule types share a common mechanism of export.

The amino acid sequence of Wza in *M. haemolytica* and *M. glucosida* was particularly conserved in the central region (d_S/d_N ratio of 43.84) because this region of the protein is involved in capsular polysaccharide transport (Figure 3.45). In contrast, amino acid replacement is less constrained in the three variable regions at the N- and C terminal ends of the protein (d_S/d_N ratios of 3.93 to 5.91), indicating that these parts of the molecule have less structural importance. Electron microscopy of the *E. coli* Wza protein has shown that the protein is morphologically similar to the secretin protein family which is composed of the C-terminal domain embedded in the outer membrane and the N-terminal domain located in periplasm (Genin & Boucher, 1994; Russel, 1998). Although it is not clear whether the Wza protein of *M. haemolytica* and *M. glucosida* has surface-exposed loop regions, the pattern of diversity in Wza is clearly different from that of OmpA and TbpA which contain long, surface-exposed loop regions (Figures 3.29 and 3.40).

4.1.6 Emergence of new *M. haemolytica* pathogenic isolates

The bovine A1/A6 isolates of ET 1 and the ovine A2 isolates of ETs 21 and 22, which are responsible for the majority of bovine and ovine disease, respectively (Davies *et al.*, 1997a), may have emerged relatively recently. Insufficient time has elapsed for equilibrium to have been achieved between pathogen and host. It is proposed that host switching and acquisition of specific genes necessary for adaptation to a new host has led to the emergence of these pathogens (Figure 4.1). For example, bovine A1/A6 isolates of ET 1 possibly evolved from an ovine ancestral serotype A12 isolate of ET 1, since they have identical electrophoretic types (Davies *et al.*, 1997a) and similar *aroA*, *gap*, and *pmm* alleles to those of the A12 isolate. An ovine ancestral isolate may have acquired the serotype A1 or A6 capsular genes from an ovine A1 or A6 isolate. A common origin for bovine and ovine A1 and A6 isolates is also

suggested by the possessions of very similar *lktA* (Davies *et al.*, 2001), *tbpA*, and *wza* alleles. This isolate may then have crossed into cattle and adapted to the bovine environment by acquiring bovine-specific genes, such as *ompA*, from bovine serotype A2 isolates. This may be a relatively recent evolutionary event, since the bovine A1/A6 isolates of ETs 1 and 2 have recombinant segments in *wza* that are identical to the corresponding regions of the ovine A1 and A6 isolates of ETs 6 to 9 (Figure 3.43, *wza1.1* and *wza2.1* to *wza2.3*) and almost identical *ompA* genes to those of bovine A2 isolates of ETs 21 and 17 (Figures 3.27 and 3.28, *ompA1.1*, *ompA1.3*, and *ompA1.5*)

On the other hand, the major ovine A2 pathogens (ETs 21 to 22) may have emerged from a bovine ancestor similar to serotype A2 isolate of ET 21. Ancestral bovine isolates may have crossed into sheep and adapted to the ovine environment by acquiring ovine-specific genes such as *ompA*. Subsequently, they evolved into two different ovine serotype A2 pathogen types by acquiring different recombinant segments in other virulence genes. As a result, ovine A2 isolates of ETs 21 and 22 have closely related electrophoretic types (Davies *et al.*, 1997a) and identical sequences in the genes *aroA2.1*, *gap3.1*, *pmm1.2*, and *wza3.4*, but they have very different sequences in the transferrin binding protein genes (this study) and the leukotoxin genes (Davies *et al.*, 2001; 2002). For example, ovine serotype A2 isolates of ET 21 have *tbpB1.8*, *tbpA1.8*, *lktA10.1*, and *lktB1.2* alleles, whereas ovine A2 isolates of ET 22 isolates have *tbpB6.2*, *tbpA4.2*, *lktA8.1* or *lktA8.2*, and *lktB5.3* type alleles. Ovine serotype A2 isolates of ET 21 have very different *ompA*, *tbpB*, *tbpA*, *lktA*, *lktB*, and *aroA* alleles from bovine serotype A2 isolates of ET 21 although they are on the same MLEE lineage.

These data support the previously proposed view that *M. haemolytica* diversity and pathogenicity have been triggered by the transmission of isolates from cattle to sheep

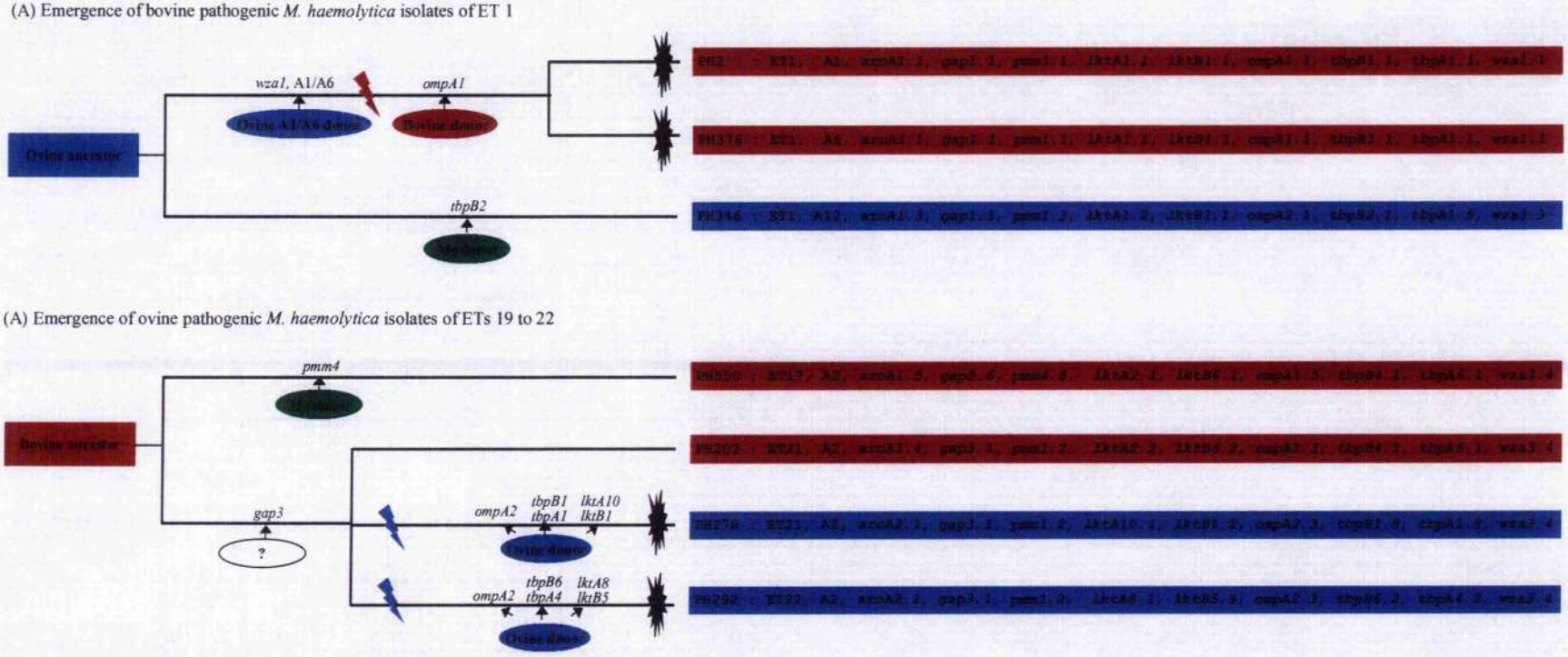


Figure 4.1 Proposed evolution of (A) pathogenic bovine *M. haemolytica* isolates of ET 1 and (B) pathogenic ovine *M. haemolytica* isolates of ETs 21 to 22. Red and blue colors indicate bovine and ovine specific isolates, respectively. Transmission event from cattle to sheep and *vice versa* are shown by ↯ and ↯, respectively. Possible donors for horizontally derived genes are indicated by squares. Bovine and ovine pathogenic isolates are indicated by ✸.

and *vice versa*, which is probably linked to the domestication of these species (Davies *et al.*, 2001; 2002).

4.1.7 Genetic diversity of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Calculations of the mean genetic diversity per locus (H) of *M. haemolytica*, *M. glucosida*, and *P. trehalosi* (Davies *et al.*, 1997a; 1997b) indicate that *M. glucosida* ($H = 0.485$) is more divergent than *M. haemolytica* ($H = 0.297$) and *P. trehalosi* ($H = 0.289$). Consistent with these data, *P. trehalosi* showed the lowest level of diversity in most of the genes examined compared to *M. haemolytica* and *M. glucosida* (Table 3.3). Davies *et al.* (1997b) suggested that this could be due to a recent evolutionary origin although another possibility was a low effective population size.

However, the four outer membrane protein genes *ompA*, *tbpB*, *tbpA*, and *wza* of *M. haemolytica* were more divergent than those of *M. glucosida* and *P. trehalosi* (Table 3.3). The higher diversity of these genes in *M. haemolytica* could be explained by the ability of this pathogen to infect multiple host species. *M. haemolytica* is frequently recovered from cattle and sheep, whereas *M. glucosida* occurs mostly in sheep, and *P. trehalosi* occurs only in sheep (Biberstein & Thompson, 1966; Quirie *et al.*, 1986). The transmission of *M. haemolytica* isolates into new hosts has led to the acquisition of new genes, such as *ompA*, *tbpB*, *tbpA*, *lktA*, and *lktB*, by horizontal gene transfer and resulted in rapid diversification of the species. However, the small sample size of *M. glucosida* (six isolates representing ETs 1, 3, 5, 7, 10, and 16) and *P. trehalosi* (four isolates representing ETs 2, 4, 15, and 19), in contrast to the larger sample size of *M. haemolytica* (42 isolates representing different MLEE lineages), might also account for lower diversity of these genes in *M. glucosida* and *P. trehalosi*, although the isolates were selected to represent divergent lineages in each species.

In particular, the *wza* gene was sequenced in only four *M. glucosida* isolates because amplification failed in two *M. glucosida* isolates of (ETs 10 and 16) presumably due to primer incompatibility (sequence diversity or absence of the primer site). Thus, the *wza* gene of *M. glucosida* may be more divergent than that of *M. haemolytica*.

4.2 Temperate bacteriophages of *M. haemolytica*, *M. glucosida*, and *P. trehalosi*

Temperate bacteriophages were studied in *M. haemolytica*, *M. glucosida* and *P. trehalosi* because they might be associated with the horizontal DNA transfer and recombination events described in the first part of this thesis.

4.2.1 Examination of bacteriophages

Electron microscopy. Temperate phages were shown to be prevalent in *M. haemolytica*. After induction with mitomycin C, electron microscopy indicated that 24 of 32 *M. haemolytica* isolates (75 %) produced temperate bacteriophages. In contrast, a relatively low rate of lysogenic isolates were associated with *M. glucosida* (one of six) and *P. trehalosi* (one of four). These findings at least suggests that bacteriophages could be responsible for the high levels of recombinational exchanges that have occurred in some genes of *M. haemolytica* rather than those of *M. glucosida* and *P. trehalosi*. It is well documented that bacteriophages play important roles in horizontal DNA transfer in bacteria (Brussow *et al.*, 2004; Canchaya *et al.*, 2003b).

With the exception of phages from *M. haemolytica* isolates PH196 and PH786, temperate phages from *M. haemolytica*, *M. glucosida*, and *P. trehalosi* belonged to common virus families (Van Regenmortel *et al.*, 2000). Phages with long,

contractile tails are members of the *Myoviridae*, and phages with long, noncontractile tails are members of the *Siphoviridae*. Within the same family, size differences in head and tail assemblies further discriminate between the phages (Table 3.31) and suggest that identical phages occur only in closely related isolates, thereby providing further support to the evolutionary relationships based on MLEE data. PH8 (ET 6) and PH398 (ET 7) phages have identical *Siphoviridae* phages; PH396 (ET 13) and PH484 (ET 14) phages have identical *Myoviridae* phages. Isolate PH598 (ET 20) contains both *Siphoviridae* and *Myoviridae* phages; the *Siphoviridae* phages are identical to those of isolates PH292 and PH392 (ET 22), whereas the *Myoviridae* phages are identical to those of isolate PH526 (ET 19)

Phages from *M. haemolytica* isolates PH196 and PH786 have short thick tails and could not be classified. Their head and tail sizes are similar to phage PRD1 of the *Tectiviridae* family (Grahn *et al.*, 2002), but they differ from this phage by being lysogenic, having a much larger DNA size, and having detachable tails. Indeed, separate forms of heads and tails were more commonly seen by electron microscopy. The tail-less heads of this phage are similar to phage-like colicin K (Bradley, 1967) but they differ from colicin K by having DNA.

Host range of bacteriophages. Thirteen of 32 *M. haemolytica*, one of six *M. glucosida*, and one of four *P. trehalosi* lysates were able to produce a zone of lysis on a sensitive indicator isolate (Table 3.32). For *M. haemolytica*, the plaque assay was a less sensitive method for the detection of temperate phages than electron microscopy which detected 24 temperate phages. The lower detection rate by plaque assay could be explained by a lack of sensitive indicator isolates for temperate phages since phages can be detected only if an appropriate host for the phage is available or by host cell immunity against similar or identical phages. Appropriate hosts are

isolates that have compatible phage receptors. The plaque patterns (Table 3.32) indicate that sensitive indicator isolates are closely related to the original host isolates, and are likely to contain identical phage receptors. In particular, phage lysates from isolates PH30 and PH376 produced lysis zones exclusively on serotype A1 and A6 isolates, and phage lysates from isolates PH598, PH470, PH278, and PH392 produced lysis zones exclusively on serotype A2 isolates. These findings suggest that capsule may act as a phage receptor for these phages. In general, bacteriophages infecting encapsulated bacteria produce capsule depolymerase to degrade the capsular polysaccharide layer and to reach the cell surface (Hughes *et al.*, 1998; Pelkonen *et al.*, 1992). Phage depolymerase enzymes have shown to be highly specific and subtle changes in capsular polysaccharide composition prevent them from degrading the polymer (Hughes *et al.*, 1998).

Surprisingly, phages induced from closely related isolates and having identical RE types (types A [PH2Φ and PH376Φ], H [PH396Φ and PH484Φ], and I [PH598Φ, PH292Φ and PH392Φ]) were associated with different indicator isolates (Table 3.32). This suggests that phages of the same RE type do not have identical genomes. Variation in the phage genes encoding the proteins Cro and CI, which control immunity (see section 1.8.6), or phage attachment proteins, may cause the different lysis patterns. Alternatively, the lysis zones may be due to other factors such as bacteriocins encoded by each lysate. Some bacteriocins are very small and cannot be identified by electron microscopy, although they can be induced by mitomycin C and produce lysis zones (Bradley, 1967). There is also a possibility that lysis zones were caused by mitomycin C-treated cultures. Some indicator isolates may be very sensitive to mitomycin C and grow slowly on the spot where the mitomycin C treated lysate was dropped since most lysis zones were very faint (Figure 3.55B).

Restriction endonuclease assays. Although electron microscopy detected 24 *M. haemolytica*, one *M. glucosida*, and one *P. trehalosi* lysates containing temperate phages, phage DNA was successfully isolated from only 15 *M. haemolytica* lysates, one *M. glucosida* lysate, and one *P. trehalosi* lysate. The failure of phage DNA isolation in the other isolates could be explained by lower yields of DNA, since there was variation in the amount of DNA obtained from the 17 positive samples. However, another possibility is that some of the phages detected by EM might be bacteriocins rather than phages. Bacteriocins are thought to be derived from prophage genomes (Bradley, 1967). They are inducible by mitomycin C and have typical phage morphology but they lack DNA (Bradley, 1967).

The RE assay detected 11 RE types (RE types A to K) based on restriction enzyme patterns among the 17 phage DNA samples (Figure 3.54). There was a correlation between RE pattern and MLEE electrophoretic type (ET) because, of five RE types (types A, C, G, H, and I) associated with more than one *M. haemolytica* isolate, the same RE type was present in closely related isolates. For example, RE type A was associated with bovine serotype A1 and A6 isolates of ET 1, RE type C with ovine serotype A1 isolates of ETs 6 and 7, RE type G with ovine serotype A7 isolates of ETs 13 and 14, RE type H with bovine serotype A2 isolates of ET 18, and RE type I with ovine serotype A2 isolates of ETs 20 and 22.

4.2.2 RE type A phages

M. haemolytica serotype A1 and A6 isolates of ET 1 are responsible for the majority of bovine disease and their virulence factors have been well studied (section 1.3). However, phages of *M. haemolytica* have received little attention and only two studies have previously reported the induction of temperate bacteriophages from bovine *M.*

haemolytica serotype A1 isolates (Froshauer *et al.*, 1996; Richards *et al.*, 1985). Richards *et al.* (1985) reported that phages were inducible from 15 bovine *M. haemolytica* serotype A1 isolates by UV light. The phages were of two morphological types (contracted and noncontracted forms), and none of the phages produced lysis zones on the same range of indicator isolates (Richards *et al.*, 1985). Froshauer (1996) reported that phages were inducible from 14 bovine *M. haemolytica* serotype A1 isolates by mitomycin C or danofloxacin and that phage DNA gave the same restriction endonuclease pattern. In the present study, phages induced from seven representative bovine *M. haemolytica* serotype A1 isolates were compared to each other. They showed significant variation in morphology, host range and DNA content. The phages of bovine serotype A6 *M. haemolytica* isolates were also compared with those of bovine serotype A1 *M. haemolytica* isolates (discussed below).

As described previously (Richards *et al.*, 1985), two different morphological types, contracted and noncontracted forms which are thought to be *Myoviridae* and *Siphoviridae*, respectively, were identified in five bovine *M. haemolytica* serotype A1 isolates and one bovine *M. haemolytica* serotype A6 isolate. Richards *et al.* (1985) reported that both phage types were seen together in lysates of the same isolates and suggested that the two morphological forms may represent the same phage. The addition of 5 mM EDTA increased the concentration of the contracted forms. This is clearly different from the present study since two phage types were seen separately in different isolates without the addition of 5 mM EDTA (Table 3.33). Only noncontracted forms (*Siphoviridae*) occurred in the lysate of isolate PH2, whereas only contracted forms (*Myoviridae*) occurred in the lysates of isolates PH280, PH560, PH30, PH540, PH554, and PH376.

As suggested previously (Richards *et al.*, 1985), phage lysates from bovine *M.*

haemolytica serotype A1 isolates PH2, PH280, PH540, and PH554 did not produce lysis zones on the other A1 indicator isolates, but phage lysates from bovine *M. haemolytica* serotype A1 isolates PH26, PH560, and PH30 produced lysis zones on several A1 indicator isolates (Table 3.34). This suggests that the seven *M. haemolytica* serotype A1 isolates selected for phage study represent a more diverse range of *M. haemolytica* isolates than the 14 and 15 isolates previously selected in two studies (Froshauer *et al.*, 1996; Richards *et al.*, 1985). The seven *M. haemolytica* isolates (this study) were chosen because they had different LPS or OMP types (Table 3.35). Plaque assays with phage lysates from 11 *M. haemolytica* serotype A1 and A6 isolates of ETs 1 to 2 on the same range of indicator isolates showed six different lysis patterns (Table 3.35). There was a correlation between lysis patterns and OMP types suggesting that OMPs might act as receptors for these phages. For *M. haemolytica* isolates of ETs 1 and 2, plaque assays were more discriminating than other typing methods since six variants were identified by plaque assay whereas only two variants were identified by each of serotyping, LPS typing, and MLEE analysis and three variants by OMP typing (Table 3.35).

As suggested previously (Froshauer *et al.*, 1996), restriction endonuclease analysis with phage DNA isolated from eight bovine *M. haemolytica* serotype A1 and A6 isolates revealed that all of the phages had identical RE types (Figure 3.56). This suggests that they have very similar, if not identical, phage genomes. This was unexpected because variation was detected in their host range and morphologies. In particular, the contracted and noncontracted phage types represent different phage families, *Myoviridae* and *Siphoviridae*, respectively (Van Regenmortel *et al.*, 2000). However, there is substantial evidence that phage taxonomy based on morphology does not reflect evolutionary relationships because phage genome diversity has been influenced by frequent horizontal DNA transfer (Lawrence *et al.*, 2002).

4.2.2.1 Bacteriophages of bovine *M. haemolytica* serotype A1 isolate PHL213

In view of the likely importance of phages in bovine *M. haemolytica* serotype A1 and A6 isolates an attempt was made to identify the genome sequence of the type A phages from the published unannotated genome sequence of *M. haemolytica* isolate PHL213. A complete prophage genome sequence was identified from the bacterial genome sequence (in three different contigs) and sequence analysis of the prophage genome identified putative cos, attL, and attR regions (Figures 3.58 and 3.60). This allowed the construction of a hypothetical PHL213Φ phage genome (Figure 3.61). In addition, the six restriction endonuclease profiles of the putative PHL213Φ genome were identical to the corresponding enzyme profiles of PH2Φ and provided strong evidence that the PHL213Φ phage corresponds to the RE type A phage.

The genetic organization (ORFs) of phage PHL213Φ was very similar to that of the P2 phage of *E. coli* (Table 3.38 and Figure 3.63). Therefore, the RE type A phages of *M. haemolytica* represent new members of the P2-like phage family. In addition to the sequence homology, RE type A phages have hexagonal heads and long contractile tails that are very similar to those of P2 phages (Bertani & Six, 1988). The genome size is also in a similar range to that of the P2 phage family. For example, the phage DNA size is about 34.5 kb in PHL213Φ, 33.5 kb in P2 (GenBank accession number NC_001895), 35.6 kb in ΦCTX (GenBank accession number NC_003278), and 32.4 kb in HP1 (GenBank accession number NC_001697).

This study indicates that temperate phages are more common in *M. haemolytica* isolates than they are in *M. glucosida* and *P. trehalosi*. This suggests that phages may confer a selective advantage to *M. haemolytica* by carrying virulence factors.

However, no virulence factors were detected within the prophage genome of PHL213Φ, indicating that this phage does not appear to be involved in the horizontal DNA transfer of virulence genes. However, this phage may be involved in the transfer of genes that flank the phage genome although this was not investigated. The remaining 10 RE type phages were not examined. Whether they are involved in specialized transduction or generalized transduction remains to be determined.

4.3 General conclusions and further work

Comparative sequence analysis of 19 genes of various functions in *M. haemolytica*, *M. glucosida*, and *P. trehalosi* indicates that each gene has varying degree of nucleotide sequence diversity. There was no evidence of recombinational exchanges and only single site nucleotide substitutions occurred in the *recA*, *asd*, *galE*, *gnd*, *g6pd*, *mtlD*, and *gcp* genes of *M. haemolytica* and the tree topologies were similar to that based on 16S rRNA. In contrast, there was evidence of varying degrees of recombinational exchanges in *aroA*, *gap*, *mdh*, *pmm*, *plpA*, *plpB*, *plpC*, *plpD*, *ompA*, *tbpB*, *tbpA*, and *wza* of *M. haemolytica*. In particular, the outer membrane protein genes *ompA*, *tbpB*, *tbpA*, and *wza* have undergone recombinational exchanges because they have different tree topologies which indicate different evolutionary histories. The transmission of *M. haemolytica* isolates form cattle to sheep and *vice versa* may have triggered many of these recombinational exchanges. Although there was no direct evidence that *M. haemolytica* bacteriophages carry foreign gene segments, the phage induction rate of *M. haemolytica* was much higher than that of *M. glucosida* and *P. trehalosi*. This corresponds to the fact that horizontal DNA transfer and recombination have occurred more frequently in the evolution of *M. haemolytica* than in *M. glucosida* and *P. trehalosi*, and strongly suggests that prophages play an important role in the evolution of *M. haemolytica*.

Further studies will be necessary to determine whether the eleven different RE type phages identified in this study are directly involved in the evolution of *M. haemolytica.* Comparative sequence analysis of additional genes such as the capsule biosynthetic genes will also contribute to a better understanding of the evolution of *M. haemolytica.* The use of micro-array could also contribute to our understanding of the evolution of *M. haemolytica.*

Inferred amino acid sequence analysis of *M. haemolytica* OmpA, TbpB, and TbpA provides evidence that these proteins are involved in host specificity and virulence. Structure and function studies of these proteins will be important to confirm and understand virulence and host specificity in these proteins.

M. haemolytica Gcp, PlpA, PlpB, and PlpC, OmpA, and TbpB proteins have previously been shown to be protective antigens (Dabo *et al.*, 1994; Potter *et al.*, 1999; Shewen *et al.*, 2003; Zeng *et al.*, 1999), but little attention has been paid to the diversity of these genes. Further studies are required to examine the significance of antigenic variation of these proteins in relation to their potential use in vaccines.

REFERENCES

Abdullah, K. M., Lo, R. Y. C. & Mellors, A. (1990). Distribution of glycoprotease activity and the glycoprotease gene among serotypes of *Pasteurella haemolytica*. *Biochemical Society Transactions* 18, 901-903.

Abdullah, K. M., Lo, R. Y. C. & Mellors, A. (1991). Cloning, nucleotide sequence, and expression of the *Pasteurella haemolytica* A1 glycoprotease gene. *Journal of Bacteriology* 173, 5597-5603.

Abdullah, K. M., Udoh, E. A., Shewen, P. E. & Mellors, A. (1992). A neutral glycoprotease of *Pasteurella haemolytica* A1 specifically cleaves O-sialoglycoproteins. *Infection and Immunity* 60, 56-62.

Ackermann, H. W. (2003). Bacteriophage observations and evolution. *Research in Microbiology* 154, 245-251.

Ackermann, M. R. & Brogden, K. A. (2000). Response of the ruminant respiratory tract to *Mannheimia (Pasteurella) haemolytica*. *Microbes and Infection* 2, 1079-1088.

Adlam, C. (1989). The structure, function and properties of cellular and extracellular components of *Pasteurella haemolytica*. In *Pasteurella and pasteurellosis*, pp. 75-96. Edited by C. A. a. J. M. Rutter. London, England: Academic Press.

Adlam, C., Knights, J. M., Mugridge, A. (1984). Purification, characterization and immunological properties of the serotype-specific capsular polysaccharide of *Pasteurella haemolytica* (serotype A1) organisms. *Journal of General Microbiology* 130, 2415-2426.

Adlam, C., Knights, J. M., Mugridge, A. (1985a). Purification, characterization and immunological properties of the capsular polysaccharide of *Pasteurella haemolytica* serotype T15: Its identity with the K62 (K2ab) capsular polysaccharide of Escherichia coli and the capsular polysaccharide of Neisseria meningitidis serogroup H. *Journal of General Microbiology* 131, 1963-1972.

Adlam, C., Knights, J. M., Mugridge, A. (1985b). Purification, characterization, and immunological properties of the serotype-specific capsular polysaccharide of *Pasteurella haemolytica* (serotype T4) organisms. *Journal of General Microbiology* 131, 387-394.

Adlam, C., Knights, J. M., Mugridge, A. (1986). Purification, characterization and immunological properties of the serotype-specific capsular polysaccharide of *Pasteurella haemolytica* serotype A7 organisms. *Journal of General Microbiology*

132, 1079-1087.

Adlam, C., Knights, J. M., Mugridge, A. (1987). Production of colominic acid by *Pasteurella haemolytica* serotype A2 organisms. *FEMS Microbiology Letters* 42, 23-25.

Albert, M. J., Siddique, A. K., Islam, M. S., Faruque, A. S., Ansaruzzaman, M., Faruque, S. M. & Sack, R. B. (1993). Large outbreak of clinical cholera due to *Vibrio cholerae* non-O1 in Bangladesh. *Lancet* 341, 704.

Alcantara, J. & Schryvers, A. B. (1996). Transferrin binding protein 2 interacts with both the N-lobe and C-lobe of ovotransferrin. *Microbial Pathogenesis* 20, 73-85.

Ali, Q., Davies, R. L., Parton, R., Coote, J. G. & Gibbs, H. A. (1992). Lipopolysaccharide heterogeneity in *Pasteurella haemolytica* isolates from cattle and sheep. *Journal of General Microbiology* 138, 2185-2195.

Allen, C. A., Adams, L. G. & Ficht, T. A. (1998). Transposon-derived *Brucella abortus* rough mutants are attenuated and exhibit reduced intracellular survival. *Infection and Immunity* 66, 1008-1016.

Anderson, J. E., Sparling, P. F. & Cornelissen, C. N. (1994). Gonococcal transferrin-binding protein 2 facilitates but is not essential for transferrin utilization. *Journal of Bacteriology* 176, 3162-3170.

Angen, O., Quirie, M., Donachie, W. & Bisgaard, M. (1999a). Investigations on the species specificity of *Mannheimia (Pasteurella) haemolytica* serotyping. *Veterinary Microbiology* 65, 283-290.

Angen, O., Mutters, R., Caugant, D. A., Olsen, J. E. & Bisgaard, M. (1999b). Taxonomic relationships of the *[Pasteurella] haemolytica* complex as evaluated by DNA-DNA hybridizations and 16S rRNA sequencing with proposal of *Mannheimia haemolytica* gen. nov., comb. nov., *Mannheimia granulomatis* comb. nov., *Mannheimia glucosida* sp. nov., *Mannheimia ruminalis* sp. nov. and *Mannheimia varigena* sp. nov. *International Journal of Systematic Bacteriology* 49, 67-86.

Angen, O., Ahrens, P., Kuhnert, P., Christensen, H. & Mutters, R. (2003). Proposal of *Histophilus somni* gen. nov., sp. nov. for the three species incertae sedis '*Haemophilus somnus*', '*Haemophilus agni*' and '*Histophilus ovis*'. *International Journal of Systematic Evolutionary Microbiology* 53, 1449-1456.

Arber, W. (2000). Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiology Reviews* 24, 1-7.

Arber, W. (2003). Elements for a theory of molecular evolution. *Gene* 317, 3-11.

Arora, A., Abildgaard, F., Bushweller, J. H. & Tamm, L. K. (2001). Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nature Structural Biology* 8, 334-338.

Baker, E. N. & Lindley, P. F. (1992). New perspectives on the structure and function of transferrins. *Journal of Inorganic Biochemistry* 47, 147-160.

Barksdale, L. & Arden, S. B. (1974). Persisting bacteriophage infections, lysogeny, and phage conversions. *Annual Review of Microbiology* 28, 265-299.

Barondess, J. J. & Beckwith, J. (1990). A bacterial virulence determinant encoded by lysogenic coliphage lambda. *Nature* 346, 871-874.

Barondess, J. J. & Beckwith, J. (1995). *bor* gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *Journal of Bacteriology* 177, 1247-1253.

Begley, T. J., Haas, B. J., Noel, J., Shekhtman, A., Williams, W. A. & Cunningham, R. P. (1999). A new member of the endonuclease III family of DNA repair enzymes that removes methylated purines from DNA. *Current Biology* 9, 653-656.

Beher, M. G., Schnaitman, C. A. & Pugsley, A. P. (1980). Major heat-modifiable outer membrane protein in gram-negative bacteria: comparison with the OmpA protien of *Escherichia coli*. *Journal of Bacteriology* 143, 906-913.

Beres, S. B., Sylva, G. L., Barbian, K. D., Lei, B., Hoff, J. S., Mammarella, N. D., Liu, M. Y., Smoot, J. C., Porcella, S. F., Parkins, L. D., Campbell, D. S., Smith, T. M., McCormick, J. K., Leung, D. Y., Schlievert, P. M. & Musser, J. M. (2002). Genome sequence of a serotype M3 strain of group A Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proceedings of the National Academy of Sciences of the United States of America* 99, 10078-10083.

Bertani, L. E. & Six, E. W. (1988). The P2-like phages and their parasite, P4. In *The Bacteriophages*, pp. 73-143. Edited by R. Calendar. New York: Plenum Press.

Beynon, L. M., Richards, J. C. & Perry, M. B. (1992). Nuclear-magnetic-resonance

analysis of the capsular antigen of *Actinobacillus pleuropneumoniae* serotype 9. Its identity with the capsular antigen of *Escherichia coli* K62 (K2ab), *Neisseria meningitidis* serogroup H and *Pasteurella haemolytica* serotype T15. *European Journal of Biochemistry* 210, 119-124.

Biberstein, E. L. (1990). Our understanding of the *Pasteurellaceae*. *Canadian Journal of Veterinary Research* 54, S78-82.

Biberstein, E. L. & Gills, M. G. (1962). The relationship of the antigenic types of the A and T types of *Pasteurella haemolytica*. *Journal of comparative pathology* 72, 316-320.

Biberstein, E. L. & Thompson, D. A. (1966). Epidemiological studies on *Pasteurella Haemolytica* in sheep. *Journal of comparative pathology* 76, 83-94.

Bilge, S. S., Vary, J. C., Jr., Dowell, S. F. & Tarr, P. I. (1996). Role of the *Escherichia coli* O157:H7 O side chain in adherence and analysis of an rfb locus. *Infection and Immunity* 64, 4795-4801.

Birge, E. A. (1994a). T4 bacteriohage as a model genetic system. In *Bacterial and bacteriophage genetics*, pp. 137-173. London: Spring-Verlag.

Birge, E. A. (1994b). Genetics of other intemperate bacteriophages. In *Bacterial and bacteriophage genetics*, pp. 174-205. London: Spring-Verlag.

Birge, E. A. (1994c). Genetics of temperate bacteriophages. In *Bacterial and bacteriophage genetics*, pp. 206-239. London: Spring-Verlag.

Birge, E. A. (1994d). Transduction. In *Bacterial and bacteriophage genetics*, pp. 240-256. London: Spring-Verlag.

Bisercic, M., Feutrier, J. Y. & Reeves, P. R. (1991). Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli*: Evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. *Journal of Bacteriology* 1991.

Boling, M. E. & Randolph, M. L. (1977). Effects of X irradiation on a temperate bacteriophage of *Haemophilus influenzae*. *Journal of Viology* 22, 47-53.

Bos, M. P., Hogan, D. & Belland, R. J. (1999). Homologue Scanning Mutagenesis Reveals CD66 Receptor Residues Required for Neisserial Opa Protein Binding.

Journal of Experimental Medicine 190, 331-340.

Boulton, I. C., Yost, M. K., Anderson, J. E. & Cornelissen, C. N. (2000). Identification of discrete domains within gonococcal transferrin-binding protein A that are necessary for ligand binding and iron uptake functions. *Infection and Immunity* 68, 6988-6996.

Boulton, I. C., Gorringe, A. R., Allison, N., Robinson, A., Gorinsky, B., Joannou, C. L. & Evans, R. W. (1998). Transferrin-binding protein B isolated from *Neisseria meningitidis* discriminates between apo and diferric human transferrin. *Biochemical Journal* 334, 269-273.

Boulton, I. C., Gorringe, A. R., Gorinsky, B., Retzer, M. D., Schryvers, A. B., Joannou, C. L. & Evans, R. W. (1999). Purified meningococcal transferrin-binding protein B interacts with a secondary, strain-specific, binding site in the N-terminal lobe of human transferrin. *Biochemical Journal* 339, 143-149.

Boyce, J. D. & Adler, B. (2000). The capsule is a virulence determinant in the pathogenesis of *Pasteurella multocida* M1404 (B:2). *Infection and Immunity* 68, 3463-3468.

Boyd, E. F. & Brussow, H. (2002). Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in Microbiology* 10, 512-529.

Boyd, E. F., Nelson, K., Wang, F. S., Whittam, T. S. & Selander, R. K. (1994). Molecular genetic basis of allelic polymorphism in malate dehydrogenase (mdh) in natural populations of *Escherichia coli* and *Salmonella enterica*. *Proceedings of the National Academy of Sciences of the United States of America* 91, 1280-1284.

Bradley, D. E. (1967). Ultrastructure of bacteriophages and bacteriocins. *Bacteriological review* 31, 230-314.

Breider, M. A., Kumar, S. & Corstvet, R. E. (1990). Bovine pulmonary endothelial cell damage mediated by *Pasteurella haemolytica* pathogenic factors. *Infection and Immunity* 58, 1671-1677.

Brogden, K. A., Lehmkuhl, H. D. & Cutlip, R. C. (1998). *Pasteurella haemolytica* complicated respiratory infections in sheep and goats. *Veterinary Research* 29, 233-254.

Brogden, K. A., Adlam, C., Lehmkuhl, H. D., Cutlip, R. C., Knights, J. M. & Engen, R. L. (1989). Effect of *Pasteurella haemolytica* (A1) capsular polysaccharide on sheep lung in vivo and on pulmonary surfactant in vitro. *American Journal of Veterinary Research* 50, 555-559.

Brok, R., Van Gelder, P., Winterhalter, M., Ziese, U., Koster, A. J., de Cock, H., Koster, M., Tommassen, J. & Bitter, W. (1999). The C-terminal domain of the *Pseudomonas* secretin XcpQ forms oligomeric rings with pore activity. *Journal of Molecular Biology* 294, 1169-1179.

Brussow, H., Canchaya, C. & Hardt, W. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews* 68, 560-602.

Bruttin, A., Foley, S. & Brussow, H. (1997). The site-specific integration system of the temperate *Streptococcus thermophilus* bacteriophage *fSfi21*. *Virology* 237, 148-158.

Buchanan, S. K., Smith, B. S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D. & Deisenhofer, J. (1999). Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. *Nature streucrtural biology* 6, 56-63.

Budzik, J. M., Rosche, W. A., Rietsch, A. & O'Toole, G. A. (2004). Isolation and characterization of a generalized transducing phage for *Pseudomonas aeruginosa* strains PAO1 and PA14. *Journal of Bacteriology* 186, 3270-3273.

Burke, J., Schneider, D. & Westpheling, J. (2001). Generalized transduction in *Streptomyces coelicolor*. *PNAS* 98, 6289-6294.

Byun, R., Elbourne, L. D., Lan, R. & Reeves, P. R. (1999). Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. *Infection and Immunity* 67, 1116-1124.

Campbell, A. (2003). Prophage insertion sites. *Research in Microbiology* 154, 277-282.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. (2003a). Prophage genomics. *Microbiology and Molecular Biology Reviews* 67, 238-276.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. & Brussow, H.

(2003b). Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* 6, 417-424.

Caspar, D. L. (1980). Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophysical Journal* 32, 103-138.

Chae, C. H., Gentry, M. J., Confer, A. W. & Anderson, G. A. (1990). Resistance to host immune defense mechanisms afforded by capsular material of *Pasteurella haemolytica*, serotype 1. *Veterinary Microbiology* 25, 241-251.

Chanyangam, M., Smith, A. L., Moseley, S. L., Kuehn, M. & Jenny, P. (1991). Contribution of a 28-kilodalton membrane protein to the virulence of *Haemophilus influenzae*. *Infection and Immunity* 59, 600-608.

Christensen, H., Bisgaard, M., Aalbaek, B. & Olsen, J. E. (2004a). Reclassification of Bisgaard taxon 33, with proposal of *Volucribacter psittacicida* gen. nov., sp. nov. and *Volucribacter amazonae* sp. nov. as new members of the *Pasteurellaceae*. *International journal of systematic and evolutionary microbiology* 54, 813-818.

Christensen, H., Kuhnert, P., Olsen, J. E. & Bisgaard, M. (2004b). Comparative phylogenies of the housekeeping genes *atpD, infB* and *rpoB* and the 16S rRNA gene within the *Pasteurellaceae*. *International journal of systematic and evolutionary microbiology* 54, 1601-1609.

Christensen, H., Bisgaard, M., Bojesen, A. M., Mutters, R. & Olsen, J. E. (2003). Genetic relationships among avian isolates classified as *Pasteurella haemolytica*, '*Actinobacillus salpingitidis*' or *Pasteurella anatis* with proposal of *Gallibacterium anatis* gen. nov., comb. nov. and description of additional genomospecies within *Gallibacterium* gen. nov. *International Journal of Systematic Evolutionary Microbiology* 53, 275-287.

Chung, J. Y., Zhang, Y. & Adler, B. (1998). The capsule biosynthetic locus of *Pasteurella multocida* A:1. *FEMS Microbiology Letters* 166, 289-296.

Comstock, L. E., Johnson, J. A., Michalski, J. M., Morris, J. G., Jr. & Kaper, J. B. (1996). Cloning and sequence of a region encoding a surface polysaccharide of Vibrio cholerae O139 and characterization of the insertion site in the chromosome of *Vibrio cholerae* O1. *Molecular Microbiology* 19, 815-826.

Comstock, L. E., Maneval, D., Jr., Panigrahi, P., Joseph, A., Levine, M. M., Kaper, J. B., Morris, J. G., Jr. & Johnson, J. A. (1995). The capsule and O antigen in Vibrio

cholerae O139 Bengal are associated with a genetic region not present in *Vibrio cholerae* O1. *Infection and Immunity* 63, 317-323.

Confer, A. W., Panciera, R. J., Clinkenbeard, K. D. & Mosier, D. A. (1990). Molecular aspects of virulence of *Pasteurella haemolytica*. *Canadian Journal of Veterinary Research* 54, S48-52.

Cooney, B. J. & Lo, R. Y. (1993). Three contiguous lipoprotein genes in *Pasteurella haemolytica* A1 which are homologous to a lipoprotein gene in Haemophilus influenzae type b. *Infection and Immunity* 61, 4682-4688.

Cornelissen, C. N. & Sparling, P. F. (1994). Iron piracy: acquisition of transferrin-bound iron by bacterial pathogens. *Molecular Microbiology* 14, 843-850.

Cornelissen, C. N. & Sparling, P. F. (1996). Binding and surface exposure characteristics of the gonococcal transferrin receptor are dependent on both transferrin-binding proteins. *Journal of Bacteriology* 178, 1437-1444.

Cornelissen, C. N., Anderson, J. E. & Sparling, P. F. (1997). Energy-dependent changes in the gonococcal transferrin receptor. *Molecular Microbiology* 26, 25-35.

Cornelissen, C. N., Anderson, J. E., Boulton, I. C. & Sparling, P. F. (2000). Antigenic and sequence diversity in gonococcal transferrin-binding protein A. *Infection and Immunity* 68, 4725-4735.

Costello, G. M., Vipond, R. & MacIntyre, S. (1996). *Aeromonas salmonicida* possesses two genes encoding homologs of the major outer membrane protein, OmpA. *Journal of Bacteriology* 1996, 1623-1630.

Czuprynski, C. J., Nocl, E. J. & Adlam, C. (1989). Modulation of bovine neutrophil antibacterial activities by *Pasteurella haemolytica* A1 purified capsular polysaccharide. *Microbial Pathogenesis* 6, 133-141.

Dabo, S. M., Confer, A. W. & Quijano-Blas, R. A. (2003). Molecular and immunological characterization of *Pasteurella multocida* serotype A:3 OmpA: evidence of its role in P. multocida interaction with extracellular matrix molecules. *Microbial Pathogenesis* 35, 147-157.

Dabo, S. M., Confer, A. W., Styre, D. & Murphy, G. L. (1994). Expression, purification and immunologic analysis of three *Pasteurella haemolytica* A1 28-30 kDa lipoproteins. *Microbial Pathogenesis* 17, 149-158.

Davies, R. L. & Donachie, W. (1996). Intra-specific diversity and host specificity within *Pasteurella haemolytica* based on variation of capsular polysaccharide, lipopolysaccharide and outer-membrane proteins. *Microbiology* 142, 1895-1907.

Davies, R. L. & Quirie, M. (1996). Intra-specific diversity within *Pasteurella trehalosi* based on variation of capsular polysaccharide, lipopolysaccharide and outer-membrane proteins. *Microbiology* 142, 551-560.

Davies, R. L., Paster, B. J. & Dewhirst, F. E. (1996). Phylogenetic relationships and diversity within the *Pasteurella haemolytica* complex based on 16S rRNA sequence comparison and outer membrane protein and lipopolysaccharide analysis. *International Journal of Systematic Bacteriology* 46, 736-744.

Davies, R. L., Arkinsaw, S. & Selander, R. K. (1997a). Evolutionary genetics of *Pasteurella haemolytica* isolates recovered from cattle and sheep. *Infection and Immunity* 65, 3585-3593.

Davies, R. L., Arkinsaw, S. & Selander, R. K. (1997b). Genetic relationship among *Pasteurella trehalosi* isolates based on multilocus enzyme. *Microbiology* 1997, 2841-2849.

Davies, R. L., Whittam, T. S. & Selander, R. K. (2001). Sequence diversity and molecular evolution of the leukotoxin (lktA) gene in bovine and ovine strains of *Mannheimia (Pasteurella) haemolytica*. *Journal of Bacteriology* 183, 1394-1404.

Davies, R. L., Campbell, S. & Whittam, T. S. (2002). Mosaic structure and molecular evolution of the leukotoxin operon (*lktCABD*) in *Mannheimia* (*Pasteurella*) *haemolytica*, *Mannheimia glucosida*, and *Pasteurella trehalosi*. *Journal of Bacteriology* 184, 266-277.

Davies, R. L., Parton, R., Coote, J. G, Gibbs, H. A. & Freer, J. H. (1992). Outer-membrane protein and lipopolysaccharide variation in *Pasteurella haemolytica* serotype A1 under different growth conditions. *Journal of General Microbiology* 138, 909-922.

De Alwis, M. C. L. (1993). Pasteurellosis in production animals: A review. In *Pasteurellosis in production animals*, pp. 11-22. Edited by B. E. Patten, T. L. Spencer, R. B. Johnson, D. Hoffmann & L. Lehane. Canberra: ACIAR proceedings.

Deneer, H. G. & Potter, A. A. (1989). Iron-repressible outer membrane proteins of

Pasteurella haemolytica. *Journal of General Microbiology* 135, 435-443.

DeRosa, D. C., Mechor, G. D., Staats, J. J., Chengappa, M. M. & Shryock, T. R. (2000). Comparison of *Pasteurella* spp. simultaneously isolated from nasal and transtracheal swabs from cattle with clinical signs of bovine respiratory disease. *Journal of Clinical Microbiology* 38, 327-332.

Dewhirst, F. E., Paster, B. J., Olsen, I. & Fraser, G. J. (1992). Phylogeny of 54 representative strains of species in the family *Pasteurellaceae* as determined by comparison of 16S rRNA sequences. *Journal of Bacteriology* 174, 2002-2013.

Donachie, W., Fraser, J., Quirie, M. & Gilmour, N. J. L. (1984). Studies on strains of *Pasteurella haemolytica* not typable by the indirect haemagglutination test. *Research in Veterinary Science* 37, 188-193.

Donnenberg, M. S. & Whittam, T. S. (2001). Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. *Journal of Clinical Investigation* 107, 539-548.

Drummelsmith, J. & Whitfield, C. (2000). Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. *EMBO Journal* 19, 57-66.

Duim, B., Bowler, L. D., Eijk, P. P., Jansen, H. M., Dankert, J. & van Alphen, L. (1997). Molecular variation in the major outer membrane protein P5 gene of nonencapsulated *Haemophilus influenzae* during chronic infections. 1351-1356.

Edwards, R. A., Helm, R. A. & Maloy, S. R. (1999). Increasing DNA transfer efficiency by temporary inactivation of host restriction. *Bio Techniques* 26, 892-898.

Endo, Y., Yamada, T., Matsunaga, K., Hayakawa, Y., Kaidoh, T. & Takeuchi, S. (2003). Phage conversion of exfoliative toxin A in Staphylococcus aureus isolated from cows with mastitis. *Veterinary Microbiology* 96, 81-90.

Etz, H., Minh, D. B., Schellack, C., Nagy, E. & Meinke, A. (2001). Bacterial phage receptors, versatile tools for display of polypeptides on the cell surface. *Journal of Bacteriology* 183, 6924-6935.

Fay, J. C. & Wu, C. I. (2001). The neutral theory in the genomic era. *Current Opinion in Genetics and Development* 11, 642.646.

Feil, E., Carpenter, G. & Spratt, B. G. (1995). Electrophoretic variation in adenylate kinase of *Neisseria meningitidis* is due to inter- and intraspecies recombination. *Proceedings of the National Academy of Sciences of the United States of America* 92, 10535-10539.

Feil, E., Zhou, J., Maynard Smith, J. & Spratt, B. G. (1996). A comparison of the nucleotide sequences of the *adk* and *recA* genes of pathogenic and commensal *Neisseria* species: evidence for extensive interspecies recombination within *adk*. *Journal of Molecular Evolution* 43, 631-640.

Feil, E. J. & Spratt, B. G. (2001). Recombination and the population structures of bacterial pathogens. *Annual Review of Microbiology* 55, 561-590.

Ferguson, A. D. & Deisenhofer, J. (2002). TonB-dependent receptors structural perspectives. *Biochimica et Biophysica Acta* 1565, 318-332.

Fidelma Boyd, E. & Waldor, M. K. (1999). Alternative mechanism of cholera toxin acquisition by *Vibrio cholerae*: Generalized transduction of CTXPhi by bacteriophage CP-T1. *Infection and Immunity* 67, 5898-5905.

Field, D., Hood, D. & Moxon, R. (1999). Contribution of genomics to bacterial pathogenesis. *Current Opinion in Genetics and Development* 9, 700-703.

Figueroa-Bossi, N. & Bossi, L. (1999). Inducible prophages contribute to *Salmonella* virulence in mice. *Molecular Microbiology* 33, 167-176.

Fodor, L., Varga, J., Hajtos, J., Donachie, W. & Gilmour, N. J. L. (1988). Characterisation of a new serotype of *P. haemolytica* isolated in Hungary. *Research in Veterinary Science* 44, 399.

Foster, G., Ross, H. M., Malnick, H., Willems, A., Hutson, R. A., Reid, R. J. & Collins, M. D. (2000). *Phocoenobacter uteri* gen. nov., sp. nov., a new member of the family *Pasteurellaceae* Pohl (1979) 1981 isolated from a harbour porpoise (*Phocoena phocoena*). *International Journal of Systematic Evolutionary Microbiology* 50, 135-139.

Fothergill-Gilmore, L. A. & Michels, P. A. M. (1993). Evolution of glycolysis. *Progress in Biophysics and Molecular Biology* 59, 105.

Foulongne, V., Bourg, G., Cazevieille, C., Michaux-Charachon, S. & O'Callaghan, D. (2000). Identification of *Brucella suis* genes affecting intracellular survival in an in

vitro human macrophage infection model by signature-tagged transposon mutagenesis. *Infection and Immunity* 68, 1297-1303.

Frank, G. H. (1989). Pasteurellosis of cattle. In *Pasteurella and pasteurellosis*, pp. 197-222. Edited by C. F. A. a. J. M. Rutter. London, England: Academic Press.

Fraser, J., Laird, S. & Gilmour, N. J. L. (1982). A new serotype (biotype T) of *Pasteurella haemolytica*. *Research in Veterinary Science* 32, 127-128.

Fraser, J., Donachie, W., Quirie, M. & Gilmour, N. J. L. (1983). Rapid indirect hemagglutination test for serotyping *Pasteurelle haemolytica*. *Journal of Clinical Microbiology* 18, 206-207.

Frosch, M., Muller, D., Bousset, K. & Muller, A. (1992). Conserved outer membrane protein of *Neisseria meningitidis* involved in capsule expression. *Infection and Immunity* 60, 798-803.

Froshauer, S., Silvia, A. M., Chidambaram, M., Sharma, B. & Weinstock, G. M. (1996). Sensitization of bacteria to danofloxacin by temperate prophages. *Antimicrobial agents and chemotheraphy* 40, 1561-1563.

Fuller, C. A., Yu, R., Irwin, S. W. & Schryvers, A. B. (1998). Biochemical evidence for a conserved interaction between bacterial transferrin binding protein A and transferrin binding protein B. *Microbial Pathgenesis* 24, 75-87.

Garcia-Vallve, S., Romeu, A. & Palau, J. (2000). Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Research* 10, 1719-1725.

Genin, B. & Boucher, C. A. (1994). A superfamily of proteins involved in different secretion pathways in Gram-negative bacteria: modular structure and specificity of the N-terminal domain. *Molecular and general genetics* 243, 112-118.

Gentry, M. J., Confer, A. W. & Holland, S. G. (1988). Comparison of the toxic and antigenic properties of single bovine isolates of *Pasteurella haemolytica* representing five serotypes and an untypable strain. *Veterinary Microbiology* 16, 351-367.

Gerlach, G. F., Klashinsky, S., Anderson, C., Potter, A. A. & Willson, P. J. (1992). Characterization of two genes encoding distinct transferrin-binding proteins in different *Actinobacillus pleuropneumoniae* isolates. *Infection and Immunity* 60, 3253-3261.

Gilmour, N. J. L. & Gilmour, J. S. (1989). Pasteurellosis of sheep. In *Pasteurella and pasteurellosis*, pp. 223-261. Edited by C. F. A. a. J. M. Rutter. London, England: Academic Press.

Gogarten, J. P. & Olendzenski, L. (1999).Orthologs, paralogs and genome comparisons. In *Current Opinion in Genetics and Development*, pp. 630-636.

Goldberg, J. B., Hatano, K. & Pier, G. B. (1993). Synthesis of lipopolysaccharide O side chains by *Pseudomonas aeruginosa* PAO1 requires the enzyme phosphomannomutase. *Journal of Bacteriology* 175, 1605-1611.

Gonzalez, C. T. & Maheswaran, S. K. (1993). The role of induced virulence factors produced by *Pasteurella haemolytica* in the pathogenesis of bovine pneumonic pasteurellosis: review and hypotheses. *British Veterinary Journal* 149, 183-193.

Gottschalk, A. (1960). Correlation between composition, structure, shape, and function of a salivary mucoprotein. *Nature* 186, 949-951.

Govan, J. R. & Deretic, V. (1996). Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiological Reviews* 60, 539-574.

Grahn, A. M., Daugelavicius, R. & Bamford, D. H. (2002). Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. *Molecular Microbiology* 46, 1199-1209.

Gray-Owen, S. D. & Schryvers, A. B. (1993). The interaction of primate transferrins with receptors on bacteria pathogenic to humans. *Microbial Pathgenesis* 14, 389-398.

Gray-Owen, S. D. & Schryvers, A. B. (1996). Bacterial transferrin and lactoferrin receptors. *Trends in Microbiology* 4, 185-191.

Gray-Owen, S. D., Loosmore, S. & Schryvers, A. B. (1995). Identification and characterization of genes encoding the human transferrin-binding proteins from *Haemophilus influenzae*. *Infection and Immunity* 63, 1201-1210.

Guiney, D. G. (1997). Regulation of bacterial virulence gene expression by the host environment. *Journal of Clinical Investigation* 99, 565-569.

Gupta, S. & Maiden, M. C. J. (2001). Exploring the evolution of diversity in pathogen population. *Trends in Microbiology* 9, 181-185.

Harb, O. S. & Abu Kwaik, Y. (1998). Identification of the aspartate-beta-semialdehyde dehydrogenase gene of *Legionella pneumophila* and characterization of a null mutant. *Infection and Immunity* 66, 1898-1903.

Hatten, L. A., Schweizer, H. P., Averill, N., Wang, L. & Schryvers, A. B. (1993). Cloning and characterization of the *Neisseria meningitidis* asd gene. *Gene* 129, 123-128.

Hava, D. L. & Camilli, A. (2001). Isolation and characterization of a temperature-sensitive generalized transducing bacteriophage for *Vibrio cholerae*. *Journal of Microbiological Methods* 46, 217-225.

Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America* 96, 2192-2197.

Hernanz Moral, C., Flano del Castillo, E., Lopez Fierro, P., Villena Cortes, A., Anguita Castillo, J., Cascon Soriano, A., Sanchez Salazar, M., Razquin Peralta, B. & Naharro Carrasco, G. (1998). Molecular characterization of the *Aeromonas hydrophila aroA* gene and potential use of an auxotrophic *aroA* mutant as a live attenuated vaccine. *Infection and Immunity* 66, 1813-1821.

Highlander, S. K. (2001). Molecular genetic analysis of virulence in *Mannheimia (pasteurella) haemolytica*. *Frontiers in Bioscience* 6, D1128-1150.

Hill, D. J., Toleman, M. A., Evans, D. J., Villullas, S., Van Alphen, L. & Virji, M. (2001). The variable P5 proteins of typeable and non-typeable *Haemophilus influenzae* target human CEACAM1. *Molecular Microbiology*, 850-862.

Hitchcock, P. J., Leive, L., Makela, P. H., Rietschel, E. T., Strittmatter, W. & Morrison, D. C. (1986). Lipopolysaccharide nomenclature--past, present, and future. *Journal of Bacteriology* 166, 699-705.

Hoang, T. T., Williams, S., Schweizer, H. P. & Lam, J. S. (1997). Molecular genetic analysis of the region containing the essential *Pseudomonas aeruginosa* asd gene encoding aspartate-beta-semialdehyde dehydrogenase. *Microbiology* 143, 899-907.

Hobbs, M. M., Seiler, A., Achtman, M. & Cannon, J. G. (1994). Microevolution within a clonal population of pathogenic bacteria: Recombination, gene duplication

and horizontal genetic exchange in the *opa* gene family of *Neisseria meningitidis*. *Molecular Microbiology* 12, 171-180.

Holland, J., Parsons, T. R., Hasan, A. A., Cook, S. M., Stevenson, P., Griffith, E. & Williams, P. (1996). Conserved and antigenic cross-reactivity of the transferrin-binding proteins of *Haemophilus influenzae*, *Actinobacillus pleuropneumoniae* and *Neisseria meningitidis*. *Microbiology* 142, 3505-3513.

Hughes, K., Sutherland, I. & Jones, M. (1998). Biofilm susceptibility to bacteriophage attack: the role of phage-borne polysaccharide depolymerase. *Microbiology* 144, 3039-3047.

Humphrey, S. B., Stanton, T. B. & Jensen, N. S. (1995). Mitomycin C induction of bacteriophages from *Serpulina hyodysenteriae* and *Serpulina innocens*. *FEMS Microbiology Letters*. 134, 97-101.

Inoue, K. & Iida, H. (1968). Bacteriophages of *Clostridium botulinum*. *Journal of Viology* 2, 537-540.

Irwin, S. W., Averil, N., Cheng, C. Y. & Schryvers, A. B. (1993). Preparation and analysis of isogenic mutants in the transferrin receptor protein genes, *tbpA* and *tbpB*, from *Neisseria meningitidis*. *Molecular Microbiology* 8, 1125-1133.

Jacques, M. & Paradis, S. E. (1998). Adhesin-receptor interactions in *Pasteurellaceae*. *FEMS Microbiology Reviews* 22, 45-59.

Jain, R., Rivera, M. C. & Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 96, 3801-3806.

Jericho, K. W. F., Lejeune, A. & Tiffin, G. B. (1986). Bovine herpesvirus-1 and *Pasteurella haemolytica* acrobiology in experimentally infected calves. *American Journal of Veterinary Research* 47, 205-209.

Jiang, W. H., Patterson, J. A. & Steenson, L. R. (1995). Isolation and characterization of a temperate bacteriophage from a ruminal acetogen. *Current Microbiology*. 31, 336-339.

Johnson, J. A., Salles, C. A., Panigrahi, P., Albert, M. J., Wright, A. C., Johnson, R. J. & Morris, J. G., Jr. (1994). *Vibrio cholerae* O139 synonym bengal is closely related to Vibrio cholerae El Tor but has important differences. *Infection and Immunity* 62,

2108-2110.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* 292, 195-202.

Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. (2002). Microevolutionary Genomics of Bacteria. *Theoretical Population Biology* 61, 435-447.

Karaolis, D. K., Somara, S., Maneval, D. R., Jr., Johnson, J. A. & Kaper, J. B. (1999). A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* 399, 375-379.

Karlin, S. & Brocchieri, L. (1996). Evolutionary conservation of *recA* genes in relation to protein structure and function. *Journal of Bacteriology* 178, 1881-1894.

Karlin, S., Weinstock, G. M. & Brendel, V. (1995). Bacterial classifications derived from RecA protein sequence comparisons. *Journal of Bacteriology* 177, 6881-6893.

Kelly, R. T. & D., G. (1970). Toxicity of pneumococcal neuraminidase. *Infection and Immunity* 2, 115-117.

Keppel, F. & Georgopoulos, C. (1988). Stratehies of bacteriophage DNA replication. In *The Bacteriophages*, pp. 145-262. Edited by R. Calendar. New York: Plenum Press.

Khan, S. A., Everest, P., Servos, S., Foxwell, N., Zahringer, U., Brade, H., Rietschel, E. T., Dougan, G., Charles, I. G. & Maskell, D. J. (1998). A lethal role for lipid A in Salmonella infections. *Molecular Microbiology* 29, 571-579.

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267, 275-276.

Klesney-Tait, J., Hiltke, T. J., Maciver, I., Spinola, S. M., Radolf, J. D. & Hansen, E. J. (1997). The major outer membrane protein of Haemophilus ducreyi consists of two OmpA homologs. *Journal of Bacteriology* 179, 1764-1773.

Koebnik, R., Locher, K. P. & Van Gelder, P. (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell,. *Molecular Microbiology* 37, 239-253.

Kondrashov, A. S. (1999). Comparative genomics and evolutionary biology. *Current Opinion in Genetics & Development* 9, 624-629.

Korczak, B., Christensen, H., Emler, S., Frey, J. & Kuhnert, P. (2004). Phylogeny of the family *Pasteurellaceae* based on *rpoB* sequences. *International journal of systematic and evolutionary microbiology* 54, 1393-1399.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiological Reviews* 58, 401-465.

Kroll, J. S. & Booy, R. (1996). *Haemophilus influenzae*: capsule vaccine and capsulation genetics. *Molecular Medicine Today* 2, 160-165.

Kroll, J. S., Wilks, K. E., Farrant, J. L. & Langford, P. R. (1998). Natural genetic exchange between *Haemophilus* and *Neisseria*: intergeneric transfer of chromosomal genes between major human pathogens. *Proceedings of the National Academy of Sciences of the United States of America* 95, 12381-12385.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 17, 1244-1245.

Lacroix, R. P., Duncan, J. R., Jenkins, R. P., Leitch, R. A., Perry, J. A. & Richards, J. C. (1993). Structural and serological specificities of *Pasteurella haemolytica* lipopolysaccharides. *Infection and Immunity* 61, 170-181.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Lan, R. & Reeves, P. R. (1996). Gene transfer is a major factor in bacterial evolution. *Molecular biology and evolution* 13, 47-55.

Lan, R. & Reeves, P. R. (2000). Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends in Microbiology* 8, 396-401.

Lang, H. (2000). Outer membrane proteins as surface display systems. *International Journal of Medical Microbiology* 290, 579-585.

Lawrence, J. G. (1999). Gene transfer, speciation, and the evolution of bacterial genomes. *Current Opinion in Microbiology* 2, 519-523.

Lawrence, J. G. & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences of the United States of*

America 95, 9413-9417.

Lawrence, J. G., Hendrix, R. W. & Casjens, S. (2001). Where are the pseudogenes in bacterial genomes? *Trends in Microbiology* 9, 535-540.

Lawrence, J. G., Hatfull, G. F. & Hendrix, R. W. (2002). Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *Journal of Bacteriology* 184, 4891-4905.

Le, Y., He, J. & Vining, L. C. (1996). *Streptomyces akiyoshiensis* differs from other gram-positive bacteria in the organization of a core biosynthetic pathway gene for aspartate family amino acids. *Microbiology* 142, 791-798.

Lee, C. W., Lo, R. Y., Shewen, P. E. & Mellors, A. (1994). The detection of the sialoglycoprotease gene and assay for sialoglycoprotease activity among isolates of *Pasteurella haemolytica* A1 strains, serotypes A13, A14, T15 and A16. *FEMS Microbiology Letters* 121, 199-205.

Legrain, M., Mazarin, V., Irwin, S. W., Bouchon, B., Quentin-Millet, M. J., Jacobs, E. & Schryvers, A. B. (1993). Cloning and characterization of *Neisseria meningitidis* genes encoding the transferrin-binding proteins Tbp1 and Tbp2. *Gene* 130, 73-80.

Li, J., Nelson, K., McWhorter, A. C., Whittam, T. S. & Selander, R. K. (1994). Recombinational basis of serovar diversity in *Salmonella enterica*. *Proceedings of the National Academy of Sciences of the United States of America* 91, 2552-2556.

Lindberg, A. A. (1973). Bacteriophage receptors. *Annual Review of Microbiology* 205-241.

Linz, B., Schenker, M., Zhu, P. & Achtman, M. (2000). Frequent interspecific genetic exchange between commensal *Neisseriae* and *Neisseria meningitidis*. *Molecular Microbiology* 36, 1049-1058.

Lo, R. Y. (2001). Genetic analysis of virulence factors of *Mannheimia (Pasteurella) haemolytica* A1. *Veterinary Microbiology* 83, 23-35.

Lo, R. Y., McKerral, L. J., Hills, T. L. & Kostrzynska, M. (2001). Analysis of the capsule biosynthetic locus of *Mannheimia (Pasteurella) haemolytica* A1 and proposal of a nomenclature system. *Infection and Immunity* 69, 4458-4464.

Loosmore, S. M., Yang, Y. P., Coleman, D. C., Shortreed, J. M., England, D. M.,

Harkness, R. E., Chong, P. S. & Klein, M. H. (1996). Cloning and expression of the *Haemophilus influenzae* transferrin receptor genes. *Molecular Microbiology* 19, 575-586.

Lubkowski, J., Hennecke, F., Pluckthun, A. & Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure* 7, 711-722.

Luderitz, O., Galanos, C., Lehmann, V., Nurminen, M., Rietschel, E. T., Rosenfelder, G., Simon, M. & Westphal, O. (1973). Lipid A: chemical structure and biological activity. *Journal of Infectious Diseases* 128, Suppl:17-29.

Lugtenberg, B. & Van Alphen, L. (1983). Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other Gram-negative bacteria. *Biochemica et Biophysica Acta* 737, 51-115.

Mahasreshti, P. J., Murphy, G. L., Wyckoff, J. H., 3rd, Farmer, S., Hancock, R. E. & Confer, A. W. (1997). Purification and partial characterization of the OmpA family of proteins of *Pasteurella haemolytica*. *Infection and Immunity* 65, 211-218.

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

Malmborg, A. C., Soderlind, E., Frost, L. & Borrebaeck, C. A. (1997). Selective phage infection mediated by epitope expression on F pilus. *Journal of Molecular Biology* 273, 544-551.

Maniloff, J. & Ackermann, H. W. (1998). Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. *Archives of Virology* 143, 2051-2063.

Martin, W. B. (1996). Respiratory infections of sheep. *Comparative immunology, microbiology and infectious diseases* 19, 171-179.

Maskell, D. J., Szabo, M. J., Deadman, M. E. & Moxon, E. R. (1992). The gal locus from *Haemophilus influenzae*: cloning, sequencing and the use of gal mutants to study lipopolysaccharide. *Molecular Microbiology* 6, 3051-3063.

Mazarin, V., Rokbi, B. & Quentin-Millet, M. J. (1995). Diversity of the transferrin-binding protein Tbp2 of *Neisseria meningitidis*. *Gene* 158, 145-146.

McBride, J. W., Wozniak, E. J., Brewer, A. W., Naydan, D. K. & Osburn, B. I. (1999). Evidence of *Pasteurella haemolytica* linked immune complex disease in natural and experimental models. *Microbial Pathogenesis* 26, 183-193.

McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404-405.

McKerral, L. J. & Lo, R. Y. (2002). Construction and characterization of an acapsular mutant of *Mannheimia haemolytica* A1. *Infection and Immunity* 70, 2622-2629.

Melchers, F., Braun, V. & Galanos, C. (1975). The lipoprotein of the outer membrane of *Escherichia coli*: a B-lympocyte mitogen. *The Journal of Experimental Medicine* 142, 473-481.

Miao, E. A. & Miller, S. I. (1999). Bacteriophages in the evolution of pathogen-host interactions. *Proceedings of the National Academy of Sciences of the United States of America* 96, 9452-9454.

Milkman, R., , Raleigh, E. A., Mchane, M., Cryderman, D., Ryderman, P., Bilodeau, P. & Mcweeny, K. (1999). Molecular evolution of the *Escherichia coli* chromosome. V. Recombination patterns among strains of diverse origin. *Genetics* 153, 539-554.

Millman, K. L., Tavare, S. & Dean, D. (2001). Recombination in the *ompA* gene but not the *omcB* gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. *Journal of Bacteriology*, 5997-6008.

Mooi, F. R. & Bik, E. M. (1997). The evolution of epidemic *Vibrio cholerae* strains. *Trends in Microbiology* 5, 161-165.

Morona, R., Klose, M. & Henning, U. (1984). *Escherichia coli* K-12 outer membrane protein (OmpA) as a bacteriophage receptor: analysis of mutant genes expressing altered proteins. *Journal of Bacteriology* 159, 570-578.

Morona, R., Kramer, C. & Henning, U. (1985). Bacteriophage receptor area of outer membrane protein OmpA of *Escherichia coli* K12. *Journal of Bacteriology* 164, 539-543.

Morschhauser, J. K., G. Ziebuhr, W. Blum-Oehler, G. Dobrindt, U. Hacker, J. (2000). Evolution of microbial pathogens. *Philosophical transactions of the Royal Society of London Series B: Biological sciences* 29, 695-704.

Murphy, G. L. & Whitworth, L. C. (1993). Analysis of tandem, multiple genes encoding 30-kDa membrane proteins in *Pasteurella haemolytica* A1. *Gene* 129, 107-111.

Musser, J. M. (1995). Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clinical Microbiology Reviews* 8, 496-514.

Musser, J. M. (1996). Molecular population genetic analysis of emerged bacterial pathogens-selected insights. *Emerging Infectious Diseases* 2, 1-17.

Mutters, R., Mannheim, W. & Bisgaard, M. (1989). Taxonoomy of the group. In *Pasteurella and pasteurellosis*, pp. 3-34. Edited by C. F. A. a. J. M. Rutter. London, England: Academic Press.

Myers, L. E., Yang, Y. P., Du, R. P., Wang, Q., Harkness, R. E., Schryvers, A. B., Klein, M. H. & Loosmore, S. M. (1998). The transferrin binding protein B of *Moraxella catarrhalis* elicits bactericidal antibodies and is a potential vaccine antigen. *Infection and Immunity* 66, 4183-4192.

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y. & Hayashi, T. (1999). The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa:* implications for phage evolution and horizontal gene transfer via bacteriophages. *Molecular Microbiology* 31, 399-419.

Nardini, P. M., Mellors, A. & Lo, R. Y. (1998). Characterization of a fourth lipoprotein from *Pasteurella haemolytica* A1 and its homology to the OmpA family of outer membrane proteins. *FEMS Microbiology Letters* 165, 71-77.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution* 3, 418-426.

Nei, M. & Kumar, S. (2000a). Evolutionary change of DNA squences. In *Molecular evolution and phylogenetics*, pp. 33-50. Oxford: Oxford University Press.

Nei, M. & Kumar, S. (2000b). Synonymous and nonsynonymous nucleotide substitution. In *Molecular evolution and phylogenetics*. Oxford: Oxford University Press.

Nei, M. & Kumar, S. (2000c). Phylogenetic inference: Distance methods. In

Molecular evolution and phylogenetics, pp. 87-114. Oxford: Oxford University Press.

Nei, M. & Kumar, S. (2000d). Accuracies and statistical tests of phylogenetic trees. In *Molecular evolution and phylogenetics*, pp. 165-186. Oxford: Oxford University Press.

Nelson, K. & Selander, R. K. (1992). Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. *Journal of Bacteriology* 174, 6886-6895.

Nelson, K. & Selander, R. K. (1994). Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 91, 10227-10231.

Nelson, K., Whittam, T. S. & Selander, R. K. (1991). Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 88, 6667-6671.

Nesper, J., Kapfhammer, D., Klose, K. E., Merkert, H. & Reidl, J. (2000). Characterization of *Vibrio cholerae* O1 antigen as the bacteriophage K139 receptor and identification of IS1004 insertions aborting O1 antigen biosynthesis. *Journal of Bacteriology* 182, 5097-5104.

Nesper, J., Hill, C. M., Paiment, A., Harauz, G., Beis, K., Naismith, J. H. & Whitfield, C. (2003). Translocation of group 1 capsular polysaccharide in *Escherichia coli* serotype K30. Structural and functional analysis of the outer membrane lipoprotein Wza. *Journal of Biological Chemistry* 278, 49763-49772.

Newsom, I. E. & Cross, F. (1932). Some bipolar organisms found in pneumonia in sheep. *Journal of American veterinary medical association* 80, 711-719.

Nordeen, R. O. & Currier, T. C. (1983). Generalized transduction in the phytopathogen *Peudomonas syringae*. *Applied and Environmental Microbiology* 45, 1884-1889.

Nyarko, K. A., Coomber, B. L., Mellors, A. & Gentry, P. A. (1998). Bovine platelet adhesion is enhanced by leukotoxin and sialoglycoprotease isolated from *Pasteurella haemolytica* A1 cultures. *Veterinary Microbiology* 61, 81-91.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299-304.

Odugbo, M. O., Odamab, L. E., Umohc, J. U. & Makinde, A. A. (2003). Serotypes of *Pasteurella haemolytica* from pneumonic lungs of sheep in northern Nigeria. *Small Ruminant Research* 2003, 239-243.

Ogg, J. E., Timme, T. L. & Alemohammad, M. M. (1981). General Transduction in *Vibrio cholerae*. *Infection and Immunity* 31, 737-741.

Ogunnariwo, J. A. & Schryvers, A. B. (1990). Iron acquisition in *Pasteurella haemolytica*: Expression and identification of a bovine-specific transferrin receptor. *Infection and Immunity* 58, 2091-2097.

Ogunnariwo, J. A. & Schryvers, A. B. (1996). Rapid identification and cloning of bacterial transferrin and lactoferrin receptor protein genes. *Journal of Bacteriology* 178, 7326-7328.

Ogunnariwo, J. A. & Schryvers, A. B. (2001). Characterization of a novel transferrin receptor in bovine strains of *Pasteurella multocida*. *Journal of Bacteriology* 183, 890-896.

Ogunnariwo, J. A., Woo, T. K., Lo, R. Y., Gonzalez, G. C. & Schryvers, A. B. (1997). Characterization of the *Pasteurella haemolytica* transferrin receptor genes and the recombinant receptor proteins. *Microbial Pathogenesis* 23, 273-284.

Ohnishi, M., Kurokawa, K. & Hayashi, T. (2001). Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends in Microbiology* 9, 481-485.

Osawa, R., Rainey, F. A., Fujisawa, T., Lang, E., Busse, H. J., Walsh, T. & Stackebrandt, E. (1995). *Lonepinella koalarum* gen. nov., sp. nov., a new tannin-protein complex degrading bacterium. *Systematic and Applied Microbiology* 18, 368-373.

Paulsen, I., Beness, A. & Saier, M., Jr (1997). Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. *Microbiology* 143, 2685-2699.

Pautsch, A. & Schulz, G. E. (1998). Structure of the outer membrane protein A transmembrane domain. *Nature Structural Biology* 5, 1013-1017.

Pautsch, A. & Schulz, G. E. (2000). High-resolution structure of the OmpA membrane

domain. *Journal of Molecular Biology* 298, 273-282.

Pavelka, M. S., Jr., Hayes, S. F. & Silver, R. P. (1994). Characterization of KpsT, the ATP-binding component of the ABC-transporter involved with the export of capsular polysialic acid in *Escherichia coli* K1. *Journal of Biological Chemistry* 269, 20149-20158.

Pegram, R. G., Roeder, P. L. & Scott, J. M. (1979). Two new serotypes of *Pasteurella haemolytica* from sheep in Ethiopia. *Tropical Animal Health and Production* 11, 29-30.

Pelkonen, S., Aalto, J. & Finne, J. (1992). Differential activities of bacteriophage depolymerase on bacterial polysaccharide: binding is essential but degradation is inhibitory in phage infection of K1-defective *Escherichia coli*. *Journal of Bacteriology* 174, 7757-7761.

Perna, N. T., Plunkett 3rd, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J. & Kirkpatrick, H. A. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529-533.

Petit, C., Rigg, G. P., Pazzani, C., Smith, A., Sieberth, V., Stevens, M., Boulnois, G., Jann, K. & Roberts, I. S. (1995). Region 2 of the *Escherichia coli* K5 capsule gene cluster encoding proteins for the biosynthesis of the K5 polysaccharide. *Molecular Microbiology* 17, 611-620.

Pettersson, A., Klarenbeek, V., Van Deurzen, J., Poolman, J. T. & Tommassen, J. (1994). Molecular characterization of the structural gene ofr the lactoferrin receptors of the meningococcal strain H44/76. *Microbial Pathgenesis* 17, 395-408.

Pluschke, G., Mayden, J., Achtman, M. & Levine, R. P. (1983). Role of the capsule and the O antigen in resistance of O18:K1 *Escherichia coli* to complement-mediated killing. *Infection and Immunity* 42, 907-913.

Popp, A., Dehio, C., Grunert, F., Meyer, T. F. & Gray-Owen, S. D. (1999). Molecular analysis of neisserial Opa protein interactions with the CEA family of receptors: identification of determinants contributing to the differential specificities of binding. *Cell Microbiol* 1, 169-181.

Potter, A. A., Ready, K. & Gilchrist, J. (1988). Purification of fimbriae from *Pasteurella haemolytica* A-1. *Microbial Pathogenesis* 4, 311-316.

Potter, A. A., Schryvers, A. B., Ogunnariwo, J. A., Hutchins, W. A., Lo, R. Y. & Watts, T. (1999). Protective capacity of the *Pasteurella haemolytica* transferrin-binding proteins TbpA and TbpB in cattle. *Microbial Pathogenesis* 27, 197-206.

Potter, M. D. & Lo, R. Y. (1996). Cloning and characterization of the *galE* locus of *Pasteurella haemolytica* A1. *Infection and Immunity* 64, 855-860.

Prasadarao, N. V. (2002). Identification of *Escherichia coli* outer membrane protein A receptor on human brain microvascular endothelial cells. *Infection and Immunity* 70, 4556-4563,.

Preston, A., Mandrell, R. E., Gibson, B. W. & Apicella, M. A. (1996). The lipooligosaccharides of pathogenic gram-negative bacteria. *Critical Reviews in Microbiology* 22, 139-180.

Pullinger, G. D., Bevir, T. & Lax, A. J. (2004). The *pasteurella multocida* toxin is encoded within a lysogenic bacteriophage. *Molecular Microbiology* 51, 255-269.

Pupo, G. M., Karaolis, D. K., Lan, R. & Reeves, P. R. (1997). Evolutionary relationships among pathogenic and nonpathogenic Escherichia coli strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. *Infection and Immunity* 65, 2685-2692.

Purdy, C. W., Richards, A. B. & Foster, G. S. (1989). Blood bactericidal assay (*Pasteurella haemolytica*) comparison of morbidity in marketed feeder calves. *American Journal of Veterinary Research* 50, 221-225.

Quirie, M., Donachie, W. & Gilmour, N. J. (1986). Serotypes of *Pasteurella haemolytica* from cattle. *Veterinary Record* 119, 93-94.

Quirk, A. V., Sletten, A. & Hignett, R. C. (1976). Properties of phage-receptor lipopolysaccharide from *Pseudomonas morsprunorum*. *Journal of General Microbiology* 96, 375-381.

Radman, M., Taddei, F. & Matic, I. (2000). Evolution-driving genes. *Research in Microbiology* 151, 91-95.

Rahn, A., Drummelsmith, J. & Whitfield, C. (1999). Conserved organization in the cps gene clusters for expression of *Escherichia coli* group 1 K antigens: Relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *Journal of Bacteriology* 181, 2307-2313.

Ramamurthy, T., Garg, S., Sharma, R., Bhattacharya, S. K., Nair, G. B., Shimada, T., Takeda, T., Karasawa, T., Kurazano, H. & Pal, A. (1993). Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. *Lancet* 341, 703-704.

Reddy, M. S., Bernstein, J. M., Murphy, T. F. & Faden, H. S. (1996). Binding between outer membrane proteins of nontypeable *Haemophilus influenzae* and human nasopharyngeal mucin. *Infection and Immunity* 64, 1477-1479.

Rees, W. D. & Hay, S. M. (1995). The biosynthesis of threonine by mammalian cells: expression of a complete bacterial biosynthetic pathway in an animal cell. *Biochemical Journal* 309, 999-1007.

Reeves, P. R. (1992). Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiology Letters* 79, 509-516.

Regni, C. A., Tipton, P. A. & Beamer, L. J. (2000). Crystallization and initial crystallographic analysis of phosphomannomutase/phosphoglucomutase from Pseudomonas aeruginosa. *Acta Crystallographica Section D-Biological Crystallography* 56, 761-762.

Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. (2000). Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406, 64-67.

Renauld-Mongenie, G., Poncet, D., von Olleschik-Elbheim, L., Cournez, T., Mignon, M., Schmidt, M. A. & Quentin-Millet, M. J. (1997). Identification of human transferrin-binding sites within meningococcal transferrin-binding protein B. *Journal of Bacteriology* 179, 6400-6407.

Renauld-Mongenie, G., Lins, L., Krell, T., Laffly, L., Mignon, M., Dupuy, M., Delrue, R. M., Guinet-Morlot, F., Brasseur, R. & Lissolo, L. (2004). Transferrin-binding protein B of Neisseria meningitidis: sequence-based identification of the transferrin-Binding site confirmed by site-directed mutagenesis. *Journal of Bacteriology* 186, 850-857.

Rich, S. M., Sawyer, S. A. & Barbour, A. G. (2001). Antigen polymorphism in *Borrelia hermsii*, a clonal pathogenic bacterium. *Proceedings of the National Academy of Sciences of the United States of America* 98, 15038-15043.

Richards, A. B., Renshaw, H. W. & Sneed, L. W. (1985). *Pasteurella haemolytica* bacteriophage: identification, partial characterization, and relationship of temperate bacteriophages from isolates of *Pasteurella haemolytica* (biotype A, serotype 1). *American Journal of Veterinary Research* 46, 1215-1220.

Riley, L. W., Remis, R. S., Helgerson, S. D., McGee, H. B., Wells, J. G., Davis, B. R., Hebert, R. J., Olcott, E. S., Johnson, L. M., Hargrett, N. T., Blake, P. A. & Cohen, M. L. (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *New England Journal of Medicine* 308, 681-685.

Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences of the United States of America* 95, 6239-6244.

Roberts, I. S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annual Review of Microbiology* 50, 285-315.

Robertson, B. D. & Meyer, T. F. (1992). Genetic variation in pathogenic bacteria. *Trends in Genetics* 8, 422-427.

Rohwer, F. & Edwards, R. (2002). The phage proteomic tree: a genome-based taxonomy for phage. *Journal of Bacteriology* 184, 4529-4535.

Rokbi, B., Mazarin, V., Maitre-Wilmotte, G. & Quentin-Millet, M. J. (1993). Identification of two major families of transferrin receptors among *Neisseria meningitidis* strains based on antigenic and genomic features. *FEMS Microbiology letters* 110, 51-58.

Rossi-Campos, A., Anderson, C., Gerlach, G. F., Klashinsky, S., Potter, A. A. & Willson, P. J. (1992). Immunization of pigs against *Actinobacillus pleuropneumoniae* with two recombinant protein preparations. *Vaccine* 10, 512-518.

Russel, M. (1998). Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. *Journal of Molecular Biology* 279, 485-499.

Sadowski, P. (1986). Site-specific recombinases: Changing partners and doing the twist. *Journal of Bacteriology* 165, 341-347.

Samuel, A. D., Pitta, T. P., Ryu, W. S., Danese, P. N., Leung, E. C. & Berg, H. C. (1999). Flagellar determinants of bacterial sensitivity to chi-phage. *Proceedings of the*

National Academy of Sciences of the United States of America 96, 9863-9866.

Saunders, N. J., Hood, D. W. & Moxon, E. R. (1999). Bacterial evolution: bacteria play pass the gene. *Current Biology* 9, R180-183.

Schicklmaier, P., Moser, E., Wieland, T., Rabsch, W. & Schmieger, H. (1998). A comparative study on the frequency of prophages among natural isolates of *Salmonella* and *Escherichia coli* with emphasis on generalized transducers. *Antonie van Leeuwenhoek* 73, 49-54.

Schryvers, A. B. & Gonzalez, G. C. (1990). Receptors for transferrin in pathogenic bacteria are specific for the host's protein. *Canadian Journal of Microbiology* 36, 145-147.

Schweizer, M. & Henning, U. (1977). Action of a major outer cell envelope membrane protein in conjugation of *Escherichia coli* K-12. *Journal of Bacteriology* 129, 1651-1652.

Selander, R. K. & Smith, N. H. (1990). Molecular population genetics of *Salmonella*. *Reviews in medical microbiology* 1, 219-228.

Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology* 51, 873-884.

Shewen, P. E., Lee, C. W., Perets, A., Hodgins, D. C., Baldwin, K. & Lo, R. Y. (2003). Efficacy of recombinant sialoglycoprotease in protection of cattle against pneumonic challenge with *Mannheimia (Pasteurella) haemolytica* A1. *Vaccine* 21, 1901-1906.

Sisay, T. & Zerihun, A. (2003). Diversity of *Mannheimia haemolytica* and *Pasteurella trehalosi* serotypes from apparently healthy sheep and abattoir specimens in the highlands of Wollo, North East Ethiopia. *Veterinary Research communications* 27, 3-14.

Smith, G. R. (1961). Isolation of two types of *Pasteurella haemolytica* from sheep. *Nature* 183, 1132-1133.

Smith, J. M. (1999). The detection and measurement of recombination from sequence data. *Genetics* 153, 1021-1027.

Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America* 90, 4384-4388.

Sneath, P. H. & Stevens, M. (1985). A numerical taxonomic study of *Actinobacillus, Pasteurella*, and *Yersinia*. *Journal of General Microbiology* 131, 2711-2738.

Sneath, P. H. & Stevens, M. (1990). *Actinobacillus rossii* sp. nov., *Actinobacillus seminis* sp. nov., nom. rev., *Pasteurella bettii* sp. nov., *Pasteurella lymphangitidis* sp. nov., *Pasteurella mairi* sp. nov., and *Pasteurella trehalosi* sp. nov. *International Journal of Systematic Bacteriology* 40, 148-153.

Sonntag, I., Schwartz, H., Hirota, Y. & Henning, U. (1978). Cell envelope and shape of Escherichia coli: multiple mutants missing the outer membrane lipoprotein and other major outer membrane proteins. *Journal of Bacteriology* 136, 280-285.

Spratt, B. G. & Maiden, M. C. (1999). Bacterial population genetics, evolution and epidemiology. *Philosophical transactions of the Royal Society of London Series B: Biological sciences* 354, 701-710.

Spratt, B. G., Bowler, L. D., Zhang, Q. Y., Zhou, J. & Smith, J. M. (1992). Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *Journal of Molecular Evolution* 34, 115-125.

Sprenger, G. A. (1995). Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12. *Archives of Microbiology* 164, 324-330.

Stevenson, P., Williams, P. & Griffith, E. (1992). Common antigenic domains in transferrin-binding protein 2 of *Neisseria meningitidis, Neisseria gonorrhoeae* and *Haemophilus influenzae* type b. *Infection & Immunity* 60, 2391-2396.

Stine, O. C., Sozhamannan, S., Gou, Q., Zheng, S., Morris, J. G., Jr. & Johnson, J. A. (2000). Phylogeny of *Vibrio cholerae* based on *recA* sequence. *Infection and Immunity* 68, 7180-7185.

Strathdee, C. A. & Lo, R. Y. C. (1987). Extensive homology between the leukotoxin of *Pasteurella haemolytica* A1 and the alpha-hemolysin of *Escherichia coli*. *Infection and Immunity* 55, 3233-3236.

Strauch, E., Schaudinn, C. & Beutin, L. (2004). First-time isolation and

characterization of a bacteriophage encoding the Shiga toxin 2c variant, which is globally spread in strains of *Escherichia coli* O157. *Infection & Immunity* 72, 7030-7039.

Straus, D. C., Jolley, W. L. & Purdy, C. W. (1993). Characterization of neuraminidases produced by various serotypes of *Pasteurella haemolytica*. *Infection and Immunity* 61, 4669-4674.

Straus, D. C., Purdy, C. W., Loan, R. W., Briggs, R. F. & Frank, G. H. (1998). In vivo production of neuraminidase by *Pasteurella haemolytica* in market stressed cattle after natural infection. *Current Microbiology* 37, 240-244.

Tabatabaei, M., Liu, Z., Finucane, A., Parton, R. & Coote, J. (2002). Protective immunity conferred by attenuated *aroA* derivatives of *Pasteurella multocida* B:2 strains in a mouse model of hemorrhagic septicemia. *Infection and Immunity* 70, 3355-3362.

Tatum, F. M., Briggs, R. E. & Halling, S. M. (1994). Molecular gene cloning and nucleotide sequencing and construction of an *aroA* mutant of *Pasteurella haemolytica* serotype A1. *Applied and Environmental Microbiology* 60, 2011-2016.

Teichmann, S. A. & Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature Genetics* 36, 492 - 496.

Teschner, W., Serre, M. C. & Garel, J. R. (1990). Enzymatic properties, renaturation and metabolic role of mannitol-1-phosphate dehydrogenase from *Escherichia coli*. *Biochimie* 72, 33-40.

Thampapillai, G., Lan, R. & Reeves, P. R. (1994). Molecular evolution in the *gnd* locus of *Salmonella enterica*. *Molecular biology and evolution* 11, 813-828.

Torres, A. G. & Kaper, J. B. (2003). Multiple elements controlling adherence of enterohemorrhagic *Escherichia coli* O157:H7 to HeLa cells. *Infection and Immunity* 71, 4985-4995.

Ugalde, J. E., Czibener, C., Feldman, M. F. & Ugalde, R. A. (2000). Identification and characterization of the *Brucella abortus* phosphoglucomutase gene: role of lipopolysaccharide in virulence and intracellular multiplication. *Infection and Immunity* 68, 5716-5723.

Unkmeir, A., Kammerer, U., Stade, A., Hubner, C., Haller, S., Kolb-Maurer, A.,

Frosch, M. & Dietrich, G. (2002). Lipooligosaccharide and polysaccharide capsule: virulence factors of *Neisseria meningitidis* that determine meningococcal interaction with human dendritic cells. *Infection and Immunity* 70, 2454-2462.

Usher, K. C., Ozkan, E., Gardner, K. H. & Deisenhofer, J. (2001). The plug domain of FepA, a TonB-dependent transport protein from *Escherichia coli*, binds its siderophore in the absence of the transmembrane barrel domain. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10676-10681.

Van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D., Pringle, C. R. & Wickner R, B. (2000). *Virus Taxonomy: Classification and Nomenclature of Viruses. Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego: Academic Press.

Vander Byl, C. & Kropinski, A. M. (2000). Sequence of the genome of *Salmonella* bacteriophage *P22*. *Journal of Bacteriology* 182, 6472-6481.

Videira, P. A., Cortes, L. L., Fialho, A. M. & Sa-Correia, I. (2000). Identification of the *pgmG* gene, encoding a bifunctional protein with phosphoglucomutase and phosphomannomutase activities, in the gellan gum-producing strain *Sphingomonas paucimobilis* ATCC 31461. *Applied and Environmental Microbiology* 66, 2252-2258.

Virji, M., Evans, D. J., Hadfield, A., Grunert, F., Teixeira, A. M. & Watt, S. M. (1999). Critical determinants of host receptor targeting by *Neisseria meningitidis* and *Neisseria gonorrhoeae*: identification of Opa adhesiotopes on the N-domain of CD66 molecules. *Molecular Microbiology* 34, 538-551.

Voyles, B. (2002a). Assembly, maturation, and release of virions. In *The biology of virus*, pp. 201-234. Boston: Mc Graw Hill.

Voyles, B. (2002b). Effects of viral infection on host cells:integreted viruses and persistent infection. In *The biology of virus*, pp. 201-234. Boston: Mc Graw Hill.

Waldor, M. K. & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910-1914.

Walker, J. C. (1984). Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiological Reviews* 48, 60-93.

Ward, C. K. & Inzana, T. J. (1997). Identification and characterization of a DNA

region involved in the export of capsular polysaccharide by *Actinobacillus pleuropneumoniae* serotype 5a. *Infection and Immunity* 65, 2491-2496.

Watt, M. A., Mellors, A. & Lo, R. Y. (1997). Comparison of the recombinant and authentic forms of the *Pasteurella haemolytica* A1 glycoprotease. *FEMS Microbiology Letters* 147, 37-43.

Webb, D. C. & Cripps, A. W. (1998). Secondary structure and molecular anaysis of interstrain variability in the P5 outer membrnae protein of non-typeable *Haemophilus influenzae* isolated from diverse anatomical sites. *Journal of Medical Microbiology* 47, 1059-1067.

Weinbauer, M. G. & Rassoulzadegan, F. (2004). Are viruses driving microbial diversification and diversity? *Environmental Microbiology* 6, 1-11.

Weiner, R., Langille, S. & Quintero, E. (1995). Structure, function and immunochemistry of bacterial exopolysaccharides. *J Ind Microbiol* 15, 339-346.

Weiser, J. N. & C., G. E. (1991). Outer membrane protein A (OmpA) contributes to serum resistance and pathogenicity of *Escherichia coli* K-1. *Infection and Immunity* 59, 2252-2258.

Whitfeld, C. & Roberts, I. S. (1999). Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Molecular Microbiology* 31, 1307-1319.

Whittam, T. S., Wolfe, M. L., Wachsmuth, I. K., Orskov, F., Orskov, I. & Wilson, R. A. (1993). Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infection and Immunity* 61, 1619-1629.

Willi, K., Sandmeier, H., Kulik, E. M. & Meyer, J. (1997). Transduction of antibiotic resistance markers among *Actinobacillus actinomycetemcomitans* strains by temperate bacteriophages Aa [diameter]23. *Cellular and Molecular Life Sciences* 53, 904-910.

Wooldridge, K. G. & Williams, P. H. (1993). Iron uptake mechanisms of pathogenic bacteria. *FEMS Microbiology Reviews* 12, 325-348.

Ye, R. W., Zielinski, N. A. & Chakrabarty, A. M. (1994). Purification and characterization of phosphomannomutase/phosphoglucomutase from Pseudomonas aeruginosa involved in biosynthesis of both alginate and lipopolysaccharide. *Journal of Bacteriology* 176, 4851-4857.

Younan, M. & Fodar, L. (1995). Characterisation of a new *Pasteurella haemolytica* serotype (A17). *Research in Veterinary Science* 58, 98.

Yu, F., Inouye, S. & Inouye, M. (1986). Lipoprotein-28, a cytoplasmic membrane lipoprotein from *Escherichia coli*. *Journal of Biological Chemistry* 261, 2284-2288.

Yu, R. & Schryvers, A. B. (1994). Transferrin receptors on ruminant pathogens vary in their interaction with the C-lobe and N-lobe of ruminant transferrins. *Canadian Journal of Microbiology* 40, 532-540.

Yu, R. H., Gray-Owen, S. D., Ogunnariwo, J. & Schryvers, A. B. (1992). Interaction of ruminant transferrins with transferrin receptors in bovine isolates of *Pasteurella haemolytica* and *Haemophilus somnus*. *Infection and Immunity* 60, 2992-2994.

Yuan, Y., Zhang, Y. X., Watkins, N. G. & Caldwell, H. D. (1989). Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 *Chlamydia trachomatis* serovars. *Infection and Immunity* 57, 1040-1049.

Zeng, H., Pandher, K. & Murphy, G. L. (1999). Molecular cloning of the *Pasteurella haemolytica pomA* gene and identification of bovine antibodies against PomA surface domains. *Infection and Immunity* 67, 4968-4973.

Zhang, S., Santos, R. L., Tsolis, R. M., Mirold, S., Hardt, W., Adams, L. G. & Baumler, A. J. (2002). Phage mediated horizontal transfer of the *sopE1* gene increases enteropathogenicity of *Salmonella enterica* serotype typhimurium for calves. *FEMS Microbiology Letters* 217, 243-247.

Zhou, J. & Spratt, B. G. (1992). Sequence diversity within the *argF, fbp* and *recA* genes of natural isolates of *Neisseria meningitidis*: interspecies recombination within the *argF* gene. *Molecular Microbiology* 6, 2135-2146.

Zhou, J., Bowler, L. D. & Spratt, B. G. (1997). Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species. *Molecular Microbiology* 23, 799-812.

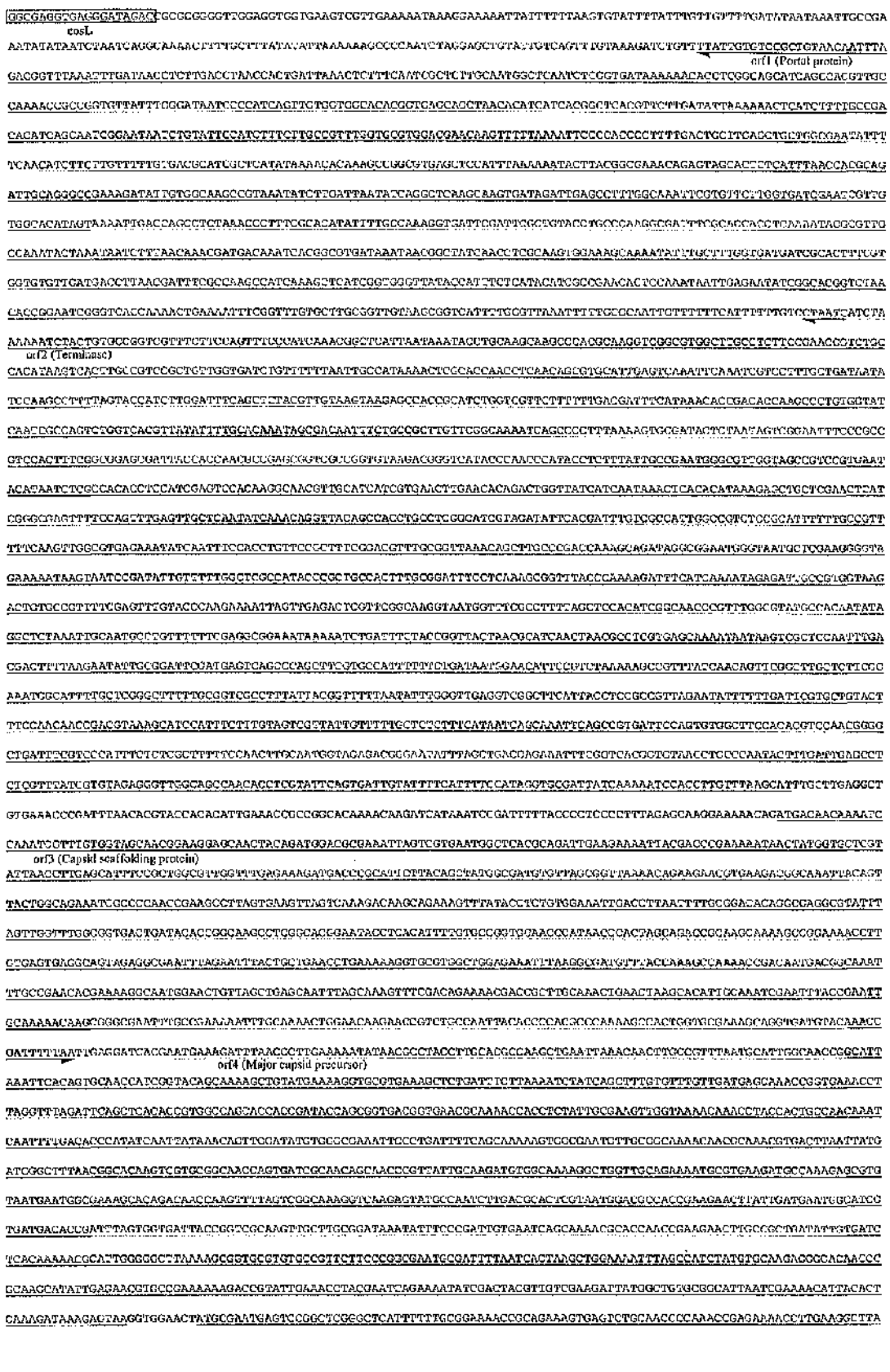
Ziebuhr, W., Ohlsen, K., Karch, H., Korhonen, T. & Hacker, J. (1999). Evolution of bacterial pathogenesis. *Cellular and Molecular Life Sciences* 56, 719-728.

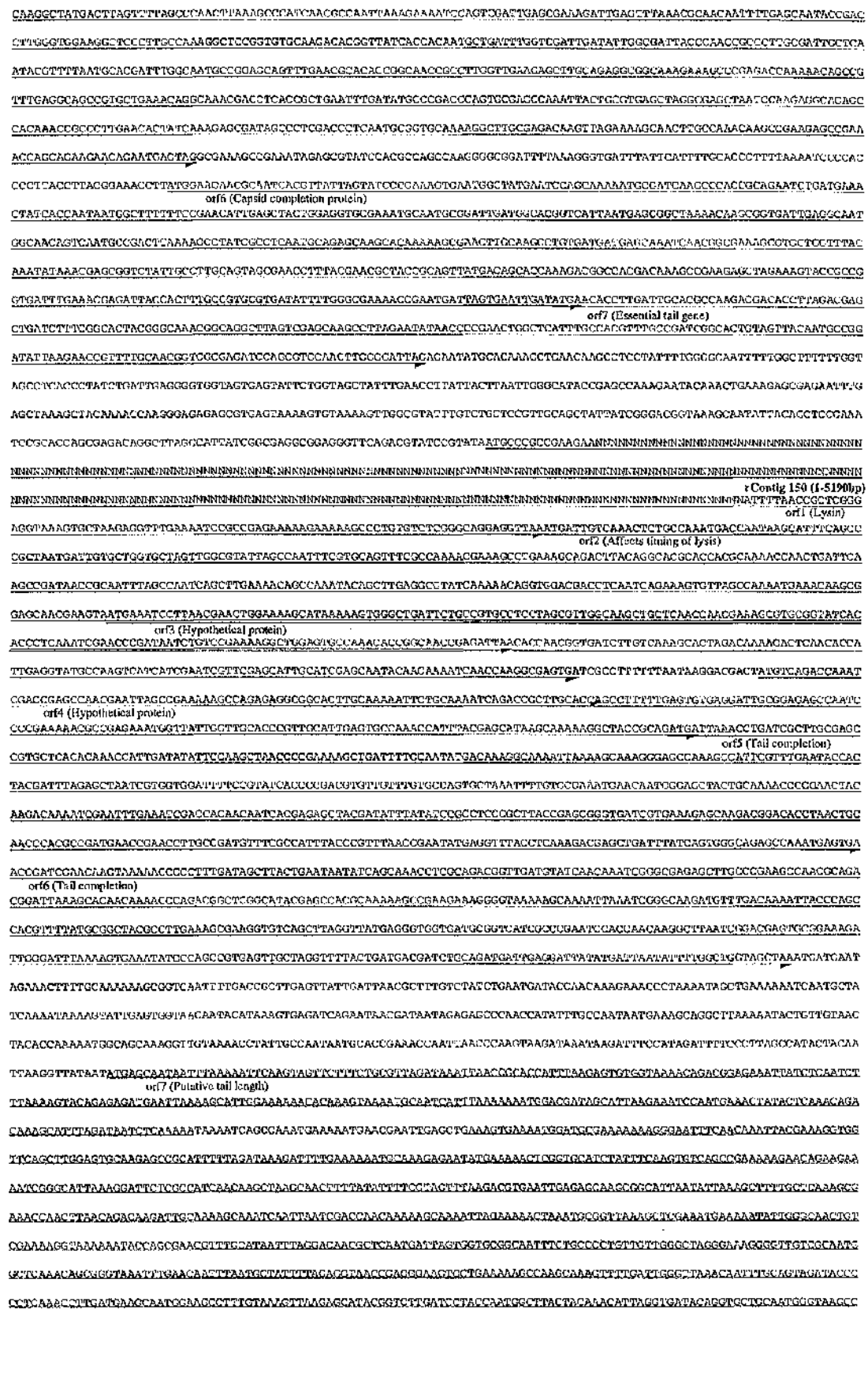
Zinser, E. R., Schneider, D., Blot, M. & Kolter, R. (2003). Bacterial evolution through

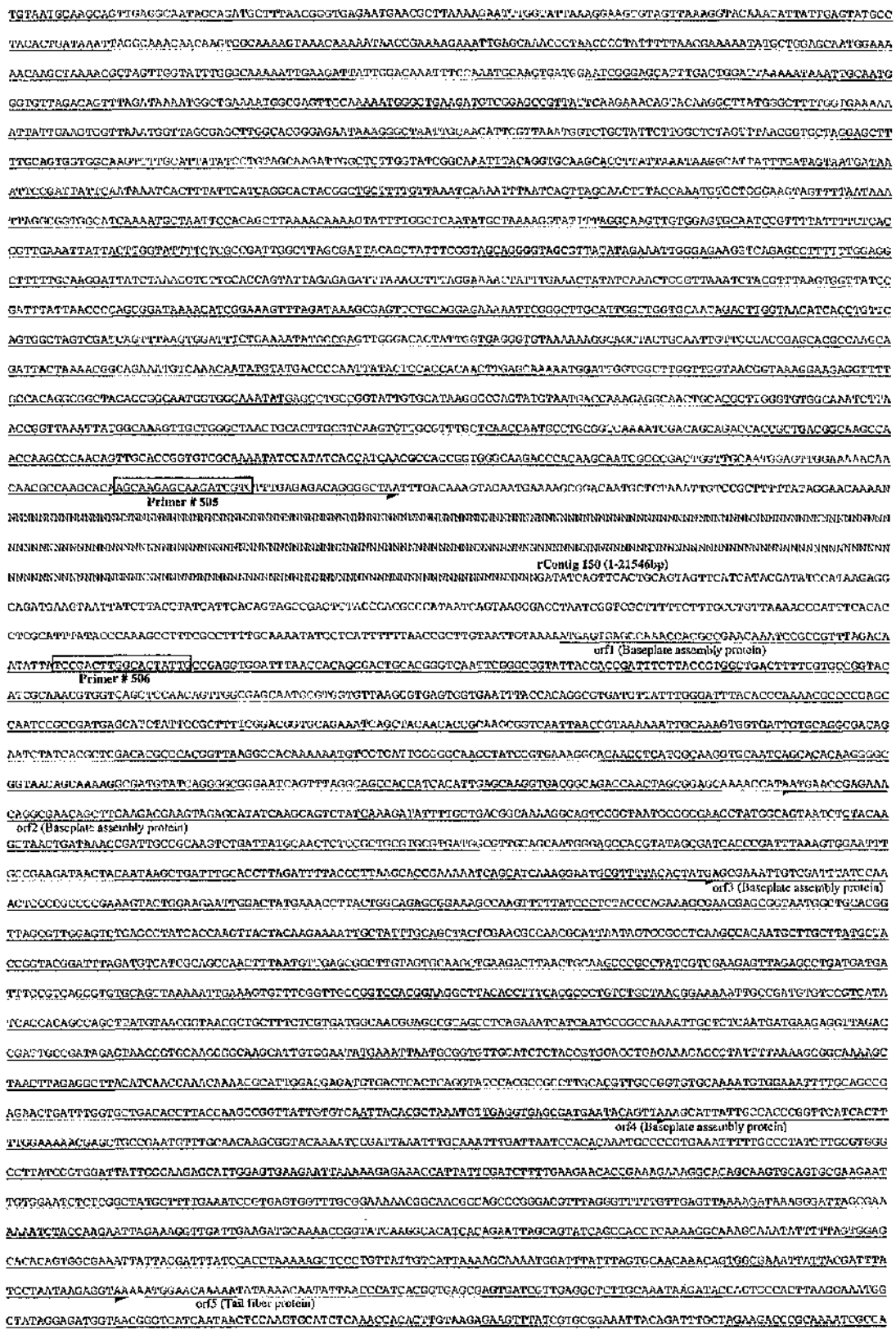
the selective loss of beneficial genes: Trade-Offs in expression involving two loci. *Genetics* 164, 1271-1277.

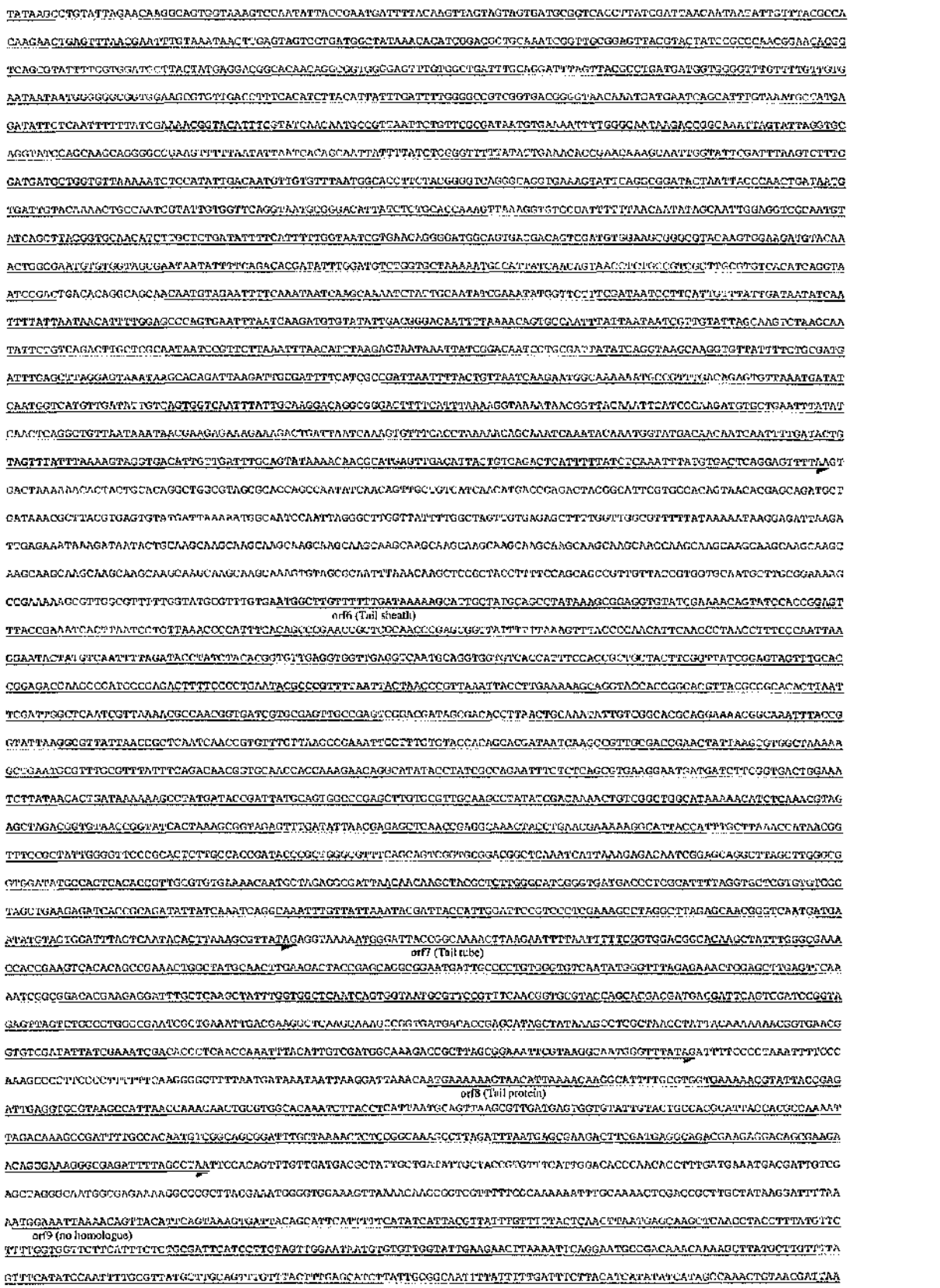
APPENDIX

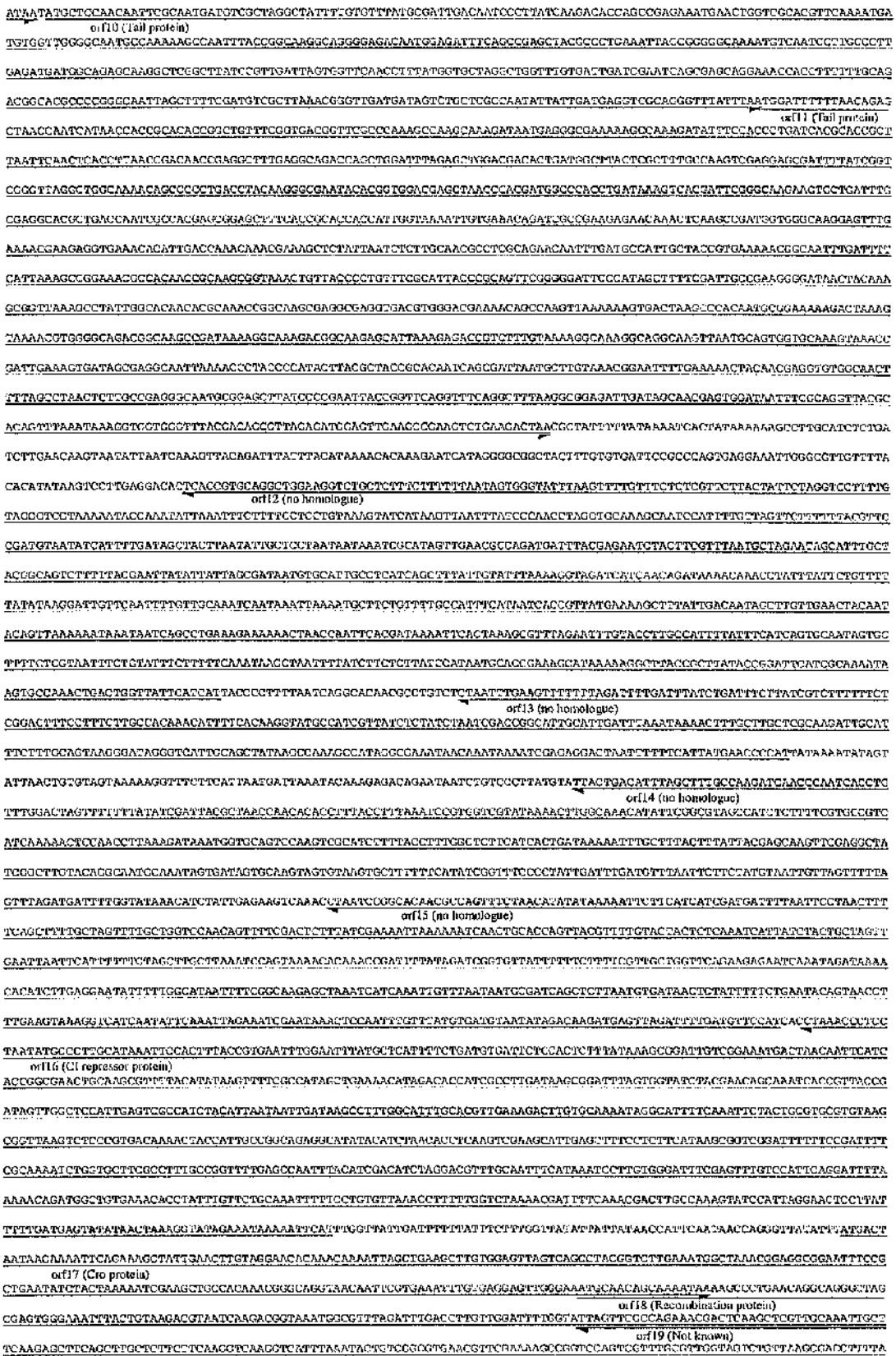
rContig 137 (34790bp-41768bp)



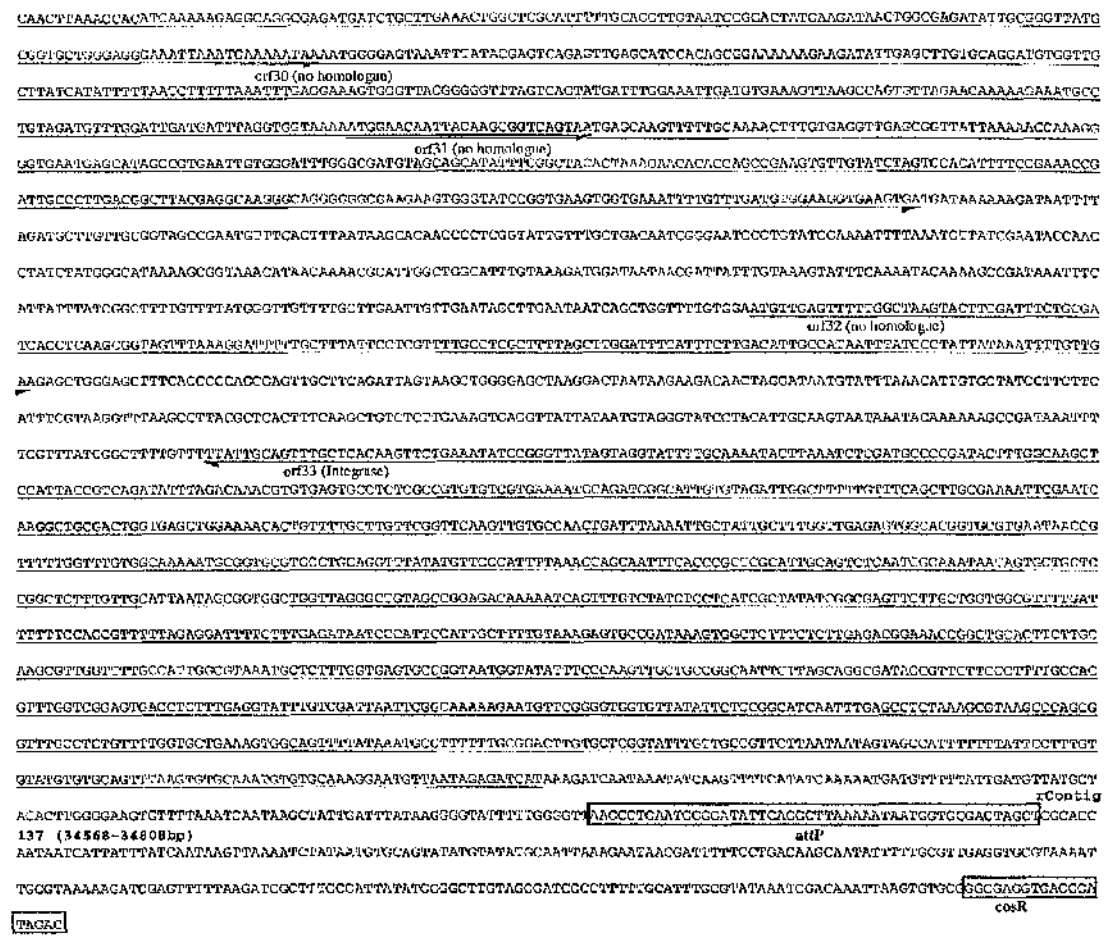

orf6 (Capsid completion protein)
orf7 (Essential tail gene)
rContig 150 (1-5190bp)
orf1 (Lysin)
orf2 (Affects timing of lysis)
orf3 (Hypothetical protein)
orf4 (Hypothetical protein)
orf5 (Tail completion)
orf6 (Tail completion)
orf7 (Putative tail length)


Primer # 505
rContig 150 (1-21546bp)
orf1 (Baseplate assembly protein)
Primer # 506
orf2 (Baseplate assembly protein)
orf3 (Baseplate assembly protein)
orf4 (Baseplate assembly protein)
orf5 (Tail fiber protein)


orf6 (Tail sheath)
orf7 (Tail tube)
orf8 (Tail protein)
orf9 (no homologue)


orf10 (Tail protein)
orf11 (Tail protein)
orf12 (no homologue)
orf13 (no homologue)
orf14 (no homologue)
orf15 (no homologue)
orf16 (CI repressor protein)
orf17 (Cro protein)
orf18 (Recombination protein)
orf19 (Not known)


orf20 (no homologue)
orf21 (no homologue)
orf22 (Transcription regulator)
orf23 (no homologue)
orf24 (no homologue)
orf25 (no homologue)
orf26 (DNA replication)
orf27 (no homologue)
orf28 (SSB protein)
orf29 (predicted 17.3 kDa protein)



Appendix I Genome sequence of PHL213Φ of *M. haemolytica* showing locations of ORFs and attP, cosL cosR, and primer sites. The ORFs are underlined with arrows which indicate the direction of transcription (see Table 3.38). The attP, cosL, cosR, and primer sites are shown in boxes.