In silico Analysis of *Treponema* and *Brachyspira* genomes:

Assembly, Annotation and Phylogeny

Mamoona Mushtaq
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2015

Acta Universitatis agriculturae Sueciae
2015:21

Cover photo: Mamoona Mushtaq

ISSN 1652-6880
ISBN (print version) 978-91-576-8240-6
ISBN (electronic version) 978-91-576-8241-3

Print: SLU Service/Repro, Uppsala 2015

In Silico Analysis of *Treponema* and *Brachyspira* genomes: Assembly, Annotation and Phylogeny

Abstract

Spirochaetal bacteria are highly diverse in terms of life style, growth requirements and virulence, but they all belong to a monophyletic ancient lineage. Many spirochaetes are difficult to culture or have fastidious growth requirements. Genus *Brachyspira* and genus *Treponema*, are two important members of phylum spirochetes that include well known pathogens, such as *T.pallidum* subsp. *pallidum* and *B. hyodysenteriae*.

Other members of the phylum include *T.pedis*, associated with both bovine digital dermatitis (BDD) in cattle and necrotic skin lesions in pigs, *T.phagedenis* associated with BDD, and "*B.suanatina"*, which has been reported in connection with swine dysentery-like enteric disease in pigs. All these bacteria are thus associated with significant health problems in farm animals leading to economic losses. The aims of this thesis were to identify and characterize potential pathogenic factors in *T.pedis* and *T. phagedenis* and to study the genomic characteristics of *B. suanatina* for a complete species validation and description.

Assembly, annotation and bioinformatics analysis of the genome sequence of *T. phagedenis* strain V1 and *T. pedis* strain TA4 were performed. The complete genome of *T. pedis* strain TA4 was compared to that of *T. denticola*, a species associated with human periodontal disease. Results of the analysis showed close relatedness between these two bacterial species. Some of the pathogenicity related genes that are already known in *T. denticola* were also found in *T. pedis* strain TA4. In the genome of *T. phagedenis* strain V1, homologous proteins to known pathogenicity factors in *T. denticola* and *T. pallidum* were found and a locus encoding antigenic lipoproteins with potential for antigenic variation was characterized

To identify the taxonomical position of *B. suanatina*, we performed assembly, annotation and phylogenetic analysis of the *B. suanatina* strain AN4859/03 genome along with its comparison with *B. hyodysenteriae* and other *Brachyspira* species. Comparative analysis suggested that *B. suanatina* is a novel species, though phenotypically it showed no difference from *B. hyodysenteriae.*

Keywords: Spirochaetes, *Treponema*, *Brachyspira*, Bovine digital dermatitis, Porcine skin ulcers, Swine dysentery, Next generation sequencing, Genome comparison, Bioinformatics, Core genome, Housekeeping genes, Phylogeny

Author's address: Mamoona Mushtaq, SLU, Department of Animal Breeding and Genetics,

P.O. Box 7023, 75007 Uppsala, Sweden
E-mail: Mamoona.mushtaq@hgen.slu.se

Dedication

To my parents, my husband and my beloved son.

"Keep your face to the sunshine and you cannot see the shadow."
Helen Keller

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Svartström O, Mushtaq M, Pringle M, Segerman B (2013). Genome-wide relatedness of *Treponema pedis*, from gingiva and necrotic skin lesions of pigs, with the human oral pathogen *Treponema denticola*. *PLoS One* 8(8), e71281.

II Mushtaq M, Manzoor S, Pringle M. Rosander A, Bongcam-Rudloff E (2014). Draft genome sequence of *'Treponema phagedenis'* strain V1, isolated from bovine digital dermatitis. (Submitted to *Standards in Genomic Sciences*).

III Mushtaq M, Loftsdottir L, Pringle M, Segerman B, Rosander A. Genetic analysis of a *Treponema phagedenis* locus encoding antigenic lipoproteins with potential for antigenic variation. (Manuscript).

IV Mushtaq M, Zubair S, Råsbäck T, Bongcam-Rudloff E, Jansson D.S. *Brachyspira suanatina* sp. nov., an enteropathogenic intestinal spirochaete isolated from pigs and mallards: genomic and phenotypic characteristics. (Manuscript).

The contribution of Mamoona Mushtaq to the papers included in this thesis was as follows:

I Partly performed the data analysis and helped Olov Svartström with writing.

II Majorly planned the study, performed the data analysis, and wrote the manuscript with comments and suggestions from other authors.

III Partly planned the study, performed the bioinformatics analysis, and partly wrote the manuscript.

IV Majorly planned and performed the bioinformatics analysis with help from other authors and wrote the bioinformatics part of the manuscript with comments and suggestions from other authors.

Abbreviations

NGS	Next generation sequencing
BDD	Bovine digital dermatitis
Msp	Major surface sheath protein
HIS	Human intestinal spirochaetosis
T.	*Treponema*
B.	*Brachyspira*
rRNA	Ribosomal RNA
tRNA	Transfer RNA
bp	Base pair
CDS	Coding sequence
ORF	Open reading frame
MGE	Mobile genetic element
VSH-1	Virus of *Serpulina hyodysenteriae*
nox	NADH oxidase gene
ANI	Average nucleotide identity
DDH	DNA-DNA hybridization
ANI	Average nucleotide identity
NJ	Neighbour joining
ML	Maximum likelihood
COG	Clusters of orthologous groups

1 Introduction

1.1 Phylum Spirochaetes

The phylum *Spirochaetes* (etymology Gr. speira ‘coil’ and chaite ‘hair’) consists of a large group of Gram stain-negative, spiral or helical shaped microorganisms (Figure 1) that represent an ancient monophyletic lineage within the domain bacteria (Gupta *et al.*, 2013; Paster & Dewhirst, 2000; Paster *et al.*, 1991; Woese, 1987). A distinct morphology of spirochaetes readily distinguishes them from other bacteria. Different species of spirochaetes range in size from 0.1–3.0 μm in diameter and 2.0–180 μm in length (Charon & Goldstein, 2002; Margulis *et al.*, 1993). An important morphological feature of these bacteria is the presence of an outer lipid bilayered membrane in addition to the plasma membrane. This lipid bilayered membrane is known as the outer membrane sheath. The periplasmic space between the cell membrane and outer membrane sheath contains the periplasmic flagella attached sub-terminally to each end and they run lengthwise towards the middle of the cell within the periplasmic space. Asymmetrical rotation of the periplasmic flagella allows spirochetes to move through viscous media where many other bacteria become immobilized. The number of periplasmic flagella varies from 1–100 depending upon the species (Charon *etal*., 2009; Charon & Goldstein, 2002; Charon *etal*., 1992).

Spirochaetes are highly diverse in terms of oxygen requirements, life style, host range and pathogenicity. Some species of spirochaetes are free living in marine sediments, deep within soil and intertidal microbial mat communities (Margulis *et al.*, 1993) while others live as commensal or obligate parasites in a wide range of hosts. Depending upon the species, they could be aerobic, facultative anaerobic or anaerobic. Spirochaetes are also highly variable in their genomic characteristics. Most of the spirochaetes have a circular chromosome like most bacteria, with the exception of *Borrelia* spp. that have

linear chromosomes. In *B. burgdorferi* there are 12 linear plasmids and 10 circular plasmids (Casjens *et al.*, 2000; Fraser *et al.*, 1997; Baril *et al.*, 1989). Some species of genus *Leptospira* contain two chromosomes that is another rare feature (Picardeau *et al.*, 2008; Bulach *et al.*, 2006). According to the publicly available information at ncbi, (http://www.ncbi.nlm.nih.gov) the genome sequence of spirochaetes ranges in size from 1Mb to 4.5Mb with a GC content of 25 to 60%.

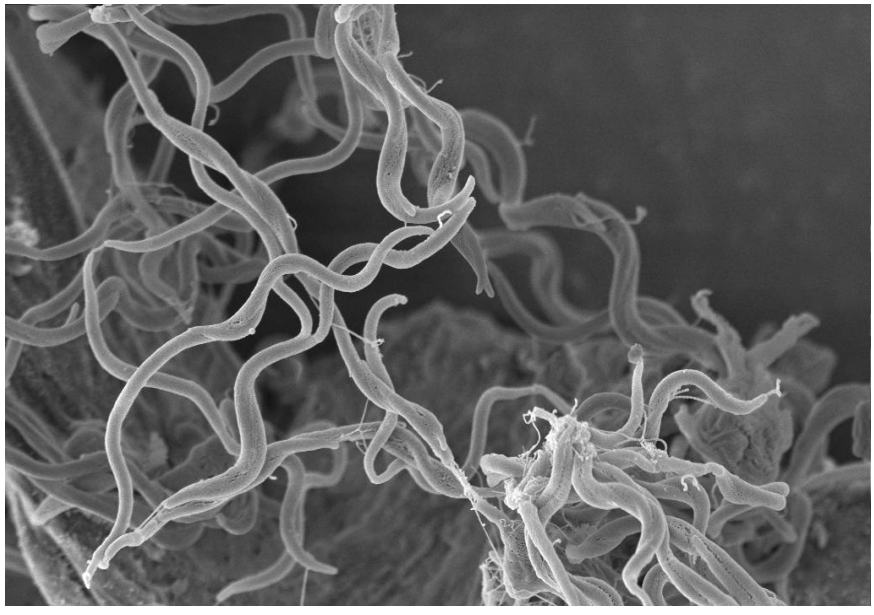


Figure 1: Scanning electron microscopy picture of *Brachyspira suanatina* strain AN4859/03, a spirochaete isolated from a pig with swine dysentery-like enteric disease (photo: L. Ljung and D. Jansson).

1.1.1 Taxonomy

Phylum *Spirochaetes* is considered to have emerged from a single free-living, anaerobic proto-spirochaete (Paster & Dewhirst, 2000; Paster *et al.*, 1991; Canale-Parola, 1977). Spirochaetes are classified into a single class *Spirochaetes*, which consists of a single order *Spirochaetales*. The order *Spriochaetales* is currently comprised of 14 genera contained in four families and one unclassified genus (Table 1). Based upon the comparative and phylogenomics analysis of 48 sequenced genomes of spirochaetes, there is a proposal for reclassification of four families to the order level taxonomic rank and transfer of genera *Borrelia* and *Cristispira* to a new family *Borreliaceae* fam. nov. (Gupta *et al.*, 2013).

Table 1. Taxonomic outline of order Spirochaetales. (Parte, 2013)

Family	Genera
Brachyspiraceae	*Brachyspira*
Brevinemataceae	*Brevinema*
Leptospiraceae	*Leptonema, Leptospira, Turneriella*
Spirochaetaceae	*Borrelia, Clevelandina, Cristispira, Diplocalyx,Hollandina, Pillotina, Sphaerochaeta, Spirochaeta, Treponema*
Unclassified	*Exilispira*

1.2 Genus *Treponema*

Genus *Treponema* currently comprises 26 valid species with *T. pallidum* subsp. *pallidum* as type species, and it is one of the major genera of phylum *Spirochaetes* (Parte, 2013). It consists of both pathogenic and non-pathogenic members that are generally found in the digestive tract, oral cavity and genital tract of humans and animals (Norris *et al.*, 2011).

1.2.1 Pathogenicity

Pathogenic bacteria of genus *Treponema* are associated with different skin diseases in mammals, a very well known example of disease causing treponemes is *T. pallidum* subsp. *pallidum*, the causative agent of syphilis in humans. Other examples include a human oral pathogen, *T. denticola* and *T. pedis* and *T. phagedenis* that are potential pathogens, found in bovine digital dermatitis (BDD) and porcine skin ulcers.

Bovine Digital Dermatitis

Bovine digital dermatitis, also known as papillomatous digital dermatitis (Walker *et al.*, 1995), inter digital papillomatous and foot wart (Read *et al.*, 1992) is a painful skin infection of the bovine foot (Figure 2). Since its first report in Italy (Cheli, 1974) the disease has been reported worldwide (Yano *et al.*, 2009; Koenig *et al.*, 2005; el-Ghoul & Shaheed, 2001) and is considered as a major cause of lameness in dairy cattle. Besides being an animal welfare concern, BDD has been associated with significant economic losses because of decreased milk production, early culling and treatment expenses (Bruijnis *et al.*, 2010). Despite the efforts made to elucidate the disease etiology, it still remains unclear. However, resolution of lesions after treatment with antibiotics

and successful isolation of bacteria, suggest these as causative agents. Bacteria from several different genera have been isolated from BDD lesions; among them spirochaetes of genus *Treponema* have been detected most frequently (Klitgaard *et al.*, 2013; Nordhoff *et al.*, 2008; Collighan & Woodward, 1997). Repeated detection of a *Treponema* phylotype recently suggested as being the same species as the human commensal *T. phagedenis* (Wilson-Welder *et al.*, 2013) allows it to be considered as a potential key agent in the pathogenesis of BDD (Yano *et al.*, 2009; Pringle *et al.*, 2008; Trott *et al.*, 2003).

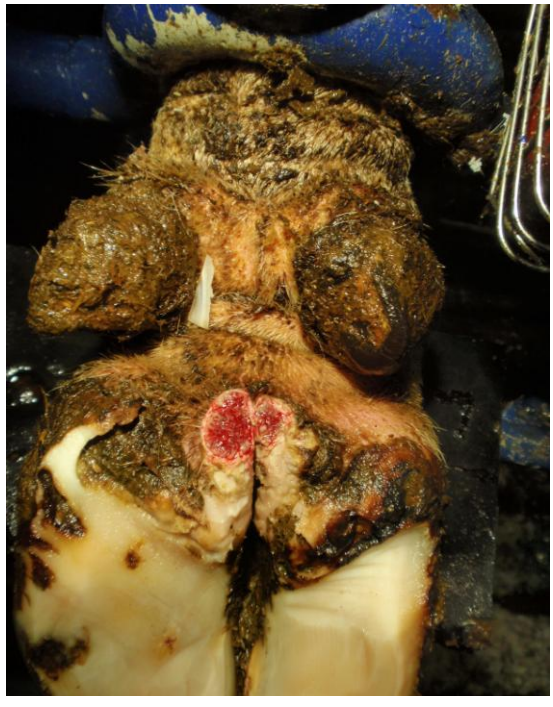


Figure 2: Cattle claw with digital dermatitis (photo: E. Hultman).

Necrotic skin ulcers in pigs

Necrotic skin ulcers in pigs refer to the development of non-healing chronic lesions on the skin. These lesions could be situated anywhere on the skin but they are most often found on the ear and on shoulders known as ear necrosis and shoulder ulcers, respectively.

Porcine ear necrosis is a condition mostly affecting young pigs and it is characterized by the occurrence of large erosive lesions at the margin of the pinna (Richardson *et al.*, 1984) (Figure 3). Clinical signs include the appearance of open wounds, crusts, and bleeding from one or both ear wounds (Petersen *et al.*, 2008). Risk factors involved in the severity and prevalence of ear necrosis are not very well understood and in its early stages it is considered to have little effect on pig performance (Busch *et al.*, 2010). However in the later stages it could lower the sale value of pigs causing economic losses. This syndrome appears to be infectious with ear biting and humid environment considered to be the predisposing factors (Park *et al.*, 2013).

Another kind of porcine skin ulcers are shoulder ulcers. They are pressure ulcers that develop on the skin that overlies the spine of the scapula (shoulder blade), which often develop during the early days of lactation in young sows. (Davies *et al.*, 1996). Shoulder ulcers are associated with economic losses because affected sows are often culled earlier than normal (Zurbrigg, 2006).

The first spirochaete isolated and characterized from shoulder ulcers, ear necrosis and gingiva of pigs was *T. pedis* (Pringle & Fellstrom, 2010; Pringle *et al.*, 2009). The type strain T3552B^{T} of *T. pedis* was originally isolated from a BDD lesion (Evans *et al.*, 2009). Later, other strains of *T. pedis* along with other *Treponema* spp. were obtained from ear necrosis, shoulder ulcers and the gingiva of pigs (Karlsson *et al.*, 2014; Svartstrom *et al.*, 2013). According to the 16S rRNA gene, 16S rRNA-tRNA(Ile) intergenic spacer region and *flaB2* phylogeny, *T. pedis* strains form a coherent taxonomic group with almost identical 16S rRNAgene sequences and very similar *flaB2* gene sequences sharing ancestry with *T. denticola* and *T. putidum* (Svartstrom *et al.*, 2013; Pringle & Fellstrom, 2010; Evans *et al.*, 2009)

Figure 3: A pig with ear necrosis (photo: F. Karlsson)

1.2.2 Genomic features of treponemes

The first *Treponema* genome to be sequenced and published was that of *T. pallidum* subsp. *pallidum* strain Nichols (Fraser *et al.*, 1998), thus laying foundation for the sequencing and comparison of other *Treponema* genomes. To date, genomes of 19 different species of *Treponema* are available publicly at ncbi (http://www.ncbi.nlm.nih.gov/genome/?term=treponema*).* All of the sequenced genomes of *Treponema* possess a single circular chromosome, and some species also contain an extra chromosomal plasmid sequence (Han *et al.*, 2011; Chauhan & Kuramitsu, 2004). The genome sizes of sequenced treponemes range from 1 to 4.5 Mb with a GC content varying between 37 and 54%.

1.2.3 Potential pathogenicity factors

Availability of genome sequences of different *Treponema* species permits a thorough search to find potential pathogenicity factors. Most of the attention in this regard has been given to genomes of *T. pallidum* subsp. *pallidum* and *T.*

denticola. An overview of some of the pathogenicity related factors defined in their genomes follows.

Motility and chemotaxis related genes

The ability of bacteria to move is called motility whereas chemotaxis enables bacteria to monitor their environment and move towards perceived stimuli. These two factors thus hold importance in the pathogenesis of bacteria. In spirochaetal bacteria they are considered to be of major importance because of their unique cellular structure that make them swim efficiently through viscous media where other bacteria become immobilized (Canale-Parola, 1977). A large number of motility and chemotaxis related genes (~5% of the whole genome) in the genomes of *T. denticola* and *T. pallidum* subsp. *pallidum* (Seshadri *et al.*, 2004; Fraser *et al.*, 1998) have been identified, that also indicates the importance of these factors in the pathogenesis of these bacteria.

Cell surface proteins

Cell surface proteins are other important factors for the pathogenicity of bacteria, because they may mediate binding to receptor molecules on the surface of the host cell and thereby help establish an infection. Different cell surface proteins that have been identified in *T. denticola* and *T. pallidum* subsp. *pallidum* using genomic information include putative antigens, adhesins, YD repeat proteins, peptidases, proteases and hydrolases (Seshadri *et al.*, 2004). Two kinds of surface proteins that have been well studied in *T. denticola* are the dentilisin protein complex and the major surface sheath protein (Msp). Dentilisin is located on the surface of *T. denticola* and is encoded from an operon with three open reading frames (ORFs) named *prcB*, *prcA* and *prtP* (Godovikova *et al.*, 2010). They have been shown to be involved in protease activity and abscess formation (Ishihara *et al.*, 1998). The Msp is a highly immunogenic protein that forms a dense hexagonal array on the surface of the bacterium. Msp is involved in binding to host cells and it possesses porin-like activity (Fenno *et al.*, 1998).

In *T. pallidum* subsp. *pallidum*, a family of 12 related genes named *tpr* (A-L) encode similar products as Msp, has been identified. The presence of multiple versions of these genes suggests their possible role in antigenic variation (Fraser *et al.*, 1998).

Lipoproteins

Lipoproteins are considered to be of special attention in spirochaetes because of their abundance in different spirochaetal genera including *Treponema* (Haake, 2000). Several of them localize to the bacterial surface and may serve

as important vaccine targets. In the genomes of *T. denticola* and *T. pallidum* subsp. *pallidum*, 166 and 22 lipoproteins have been identified respectively (Seshadri *et al.*, 2004; Fraser *et al.*, 1998).

1.3 Genus *Brachyspira*

Genus *Brachyspira* is the only genus assigned to the family *Brachyspiraceae*. Bacteria of genus *Brachyspira* colonize the intestines of mammals and birds. *Brachyspira* species are distinguished from species of other genera based on 16SrRNA gene sequence data, however these species have low interspecies 16S rRNA gene variability, which means that it may be difficult to identify and differ between some species (Hovind-Hougen *et al.*, 2011)

1.3.1 Taxonomy

Genus *Brachyspira* encompasses seven valid species, with *Brachyspira aalborgi* as the type species (Hovind-Hougen *et al.*, 1982), and several other unrecognized species. Members of the genus *Brachyspira* have undergone several taxonomic changes in the past few decades. Initially the two known species colonizing pigs were allocated to genus *Treponema* as *T. hyodysenteriae* (Harris *et al.*, 1972) as the pathogenic member, and *T. innocens* (Kinyon & Harris, 1979) as the nonpathogenic member. Based on 16S rRNA gene sequences and DNA-DNA homology analysis, these two species were transferred to a new genus *Serpula* (Paster *et al.*, 1991; Stanton *et al.*, 1991) that was later discovered to be an illegitimate name as there was a fungal genus existing with the same name. The name of the genus was then changed to *Serpulina* (Stanton, 1992) and both species were designated as *Serpulina hyodysenteriae* and *Serpulina innocens*. A new member *Serpulina pilosicoli* that may colonize many different species of animals was added to the genus in 1996 (Trott *et al.*, 1996). In 1997 genus *Serpulina* and genus *Brachyspira* were unified and all the three species in the genus were then transferred to genus *Brachyspira* as *Brachypira hyodysenteriae* comb. nov., *Brachyspira innocens* comb. nov. and *Brachyspira pilosicoli* comb. nov. (Ochiai *et al.*, 1997). The species *Serpulina intermedia* and *Serpulina murdochii* were described and validly named as two new species within the genus *Serpulina* before the publication of the unification of genera *Serpulina* and *Brachyspira* (Stanton *et al.*, 1997). Reassessment of characteristics of these two species was done and they were reclassified as *B. murdochii* and *B. intermedia* (Hampson & La, 2006). Thus genus *Brachyspira* now contain *B. hyodysenteriae, B. innocens, B. pilosicoli, B. aalborgi, B. alvinipulli, B. intermedia and B. murdochii* (Parte, 2013).

1.3.2 Disease causing members

The disease causing members of genus *Brachyspira* are:

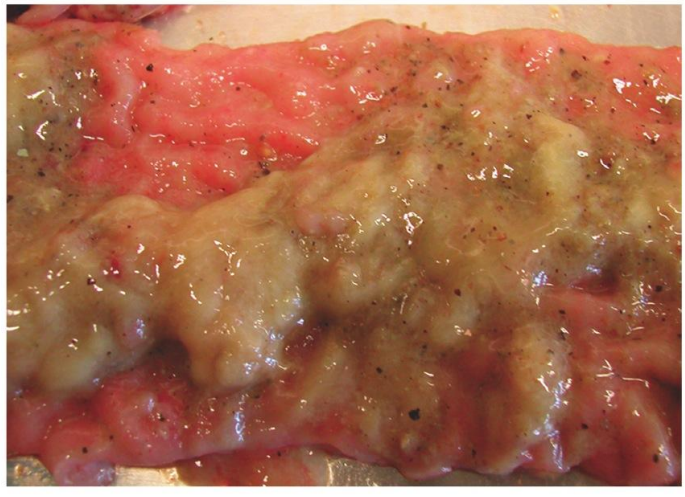
- *Brachyspira hyodysenteriae*, the most well-known member of genus *Brachyspira* that causes swine dysentery (Harris *et al.*, 1972; Taylor & Alexander, 1971).
- *Brachyspira pilosicoli* is associated with gastrointestinal disease in pigs as well as in poultry (Trott *et al.*, 2003; McLaren *et al.*, 1997) and it can also infect humans leading to human intestinal spirochaetosis (HIS) (Oxberry *et al.*, 1998)
- *Brachyspira intermedia* and *B. alvinipulli* cause production losses and diarrhoea in chickens (Stanton *et al.*, 1998; Stanton *et al.*, 1997).
- *Brachyspira aalborgi* is considered as a cause of HIS (Hovind-Hougen *et al.*, 1982). Note that the involvement of *B. aalborgi*, *B. pilosicoli* and as-yet-uncharacterized *Brachyspira* species in the pathogenesis of colitis in humans remains unclear.

1.3.3 *Brachyspira suanatina*

"*Brachyspira suanatina*" is a provisionally described new species within the genus *Brachyspira* (Rasback *et al.*, 2007). *Brachyspira suanatina* is found in pigs and ducks and an isolate originating from a diseased pig was found to possess enteropathogenic properties in pigs but not in mallards (Rasback *et al.*, 2007). Additionally, in a challenge study performed in pigs, an isolate of *B. suanatina* from mallards was shown to cause similar symptoms as swine dysentery, whereas *B. suanatina* isolated from pigs colonized mallards without any clinical symptoms (Jansson *et al.*, 2009; Rasback *et al.*, 2007).

Brachyspira suanatina isolates were phenotypically similar to *B. hyodysenteriae*, but they showed negative results when a species-specific PCR targeting the *tlyA* gene of *B. hyodysenteriae* was applied (Rasback *et al*., 2007). Figure 4 shows the colonic mucosa from pigs infected with *B. hyodysenteriae* and *B. suanatina* . Phylogenetic analysis based on the 16S rRNA and partial *nox* genes, showed that *B. suanatina* isolates formed a separate phylogenetic clade distinct from all currently recognized *Brachyspira* species and sharing ancestry with *B. hyodysenteriae* (Rasback *et al.*, 2007). Further phenotypic, molecular and phylogenetic characterization of *B. suanatina* is required to assign a taxonomic position to the proposed species.

a)



b)

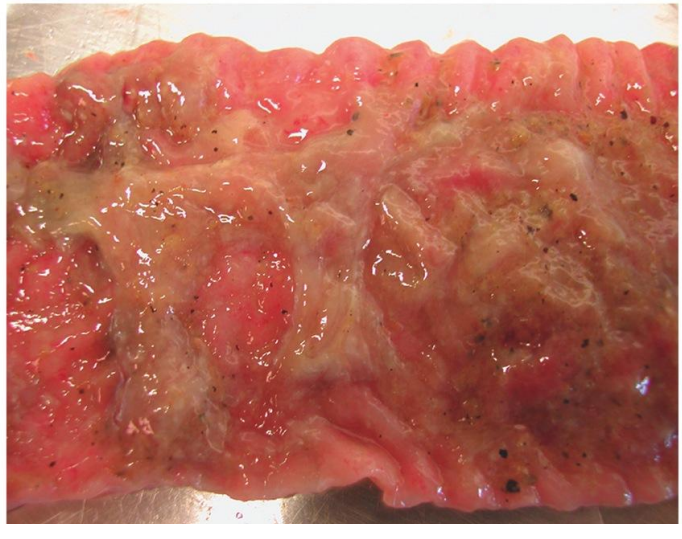


Figure 4: Colonic mucosa from pigs infected with B. *hyodysenteriae* (a) and *B. suanatina* (b). The mucosa is hyperaemic and mucous is seen adhering to the mucosal surface (photo D. Jansson)

1.3.4 Genomic features

Genome sequences of different strains of six *Brachyspira* species are available on ncbi (http://www.ncbi.nlm.nih.gov/genome/?term=brachyspira). The genome sequence of *B. aalborgi* is not available on ncbi, however it is available on the MetaHit website and can be downloaded from there (http://www.sanger.ac.uk/resources/downloads/bacteria/metahit/). According to the general genomic features, all *Brachyspira* species contain a single circular chromosome. Extra chromosomal plasmid sequences have also been reported in *B. hyodysenteriae* strain WA1 and *B. intermedia* strain PWS/A^T (Hafstrom *et al.*, 2011; Bellgard *et al.*, 2009).

In the published genomes of different *Brachyspira* species, bacteriophage and mobile genetic elements (MGEs) have been predicted. These elements are important for the inter and intra species transfer of genetic material (Hafstrom *et al.*, 2011) Different MGEs including insertion sequence elements, integrases, recombinases and transposases have been identified in *Brachyspira* genomes. Some of them also appear to be associated with major genomic rearrangements and reductive evolution events (Mappley *et al.*, 2012). Besides MGEs, putative bacteriophage regions are also identified in *Brachyspira* genomes. However, it is not clear if these putative bacteriophages are functional and capable of transferring genetic material or not. Some *Brachyspira* species share their components with other *Brachyspira* species and also with other bacteria, which points towards their role in horizontal gene transfer events (Mappley *et al.*, 2012; Hafstrom *et al.*, 2011; Wanchanthuek *et al.*, 2010).

VSH-1

Virus of *Serpulina hyodysenteriae* (VSH-1) is an unusual phage-like gene transfer agent produced by *B. hyodysenteriae*. It is involved in natural gene transfer and recombination within the species. This gene transfer agent (GTA) is in a state of permanent lysogeny and does not self-propagate, rather it assembles and transfers 7.5 kb random fragments of host DNA, including genes for antibiotic resistance, between different strains of *Brachyspira* (Matson *et al.*, 2007; Humphrey *et al.*, 1997). The VSH-1 genes span a 16.3 kb region of the *B. hyodysenteriae* strain B204^R genome and contains 11 genes encoding for structural proteins and 7 unidentified ORFs arranged in clusters of head (seven genes), tail (seven genes) and lysis (four genes) genes (Matson *et al.*, 2005). Since the first report describing gene organization of VSH-1, genomic region similar to VSH-1 has been identified in the following strains: *B. hyodysenteriae* WA1, *B. intermedia* HB60, *B. intermedia* PWS/A^T, *B. pilosicoli* 95/1000, *B. pilosicoli* WesB, *B. pilosicoli* B2904 and *B. murdochii* 56-150^T. The GTA region identified in all these species contains 11 late

function genes, described in *B. hyodysenteriae* strain B204^{R} with different gene rearrangements and insertions found in those regions. They are not yet reported to be functional in any of the species except *B. hyodysenteriae*, but gene rearrangements and insertions identified in the GTA regions of these species has been found and they may be able to transfer genetic materials among different *Brachyspira* species (Mappley *et al.*, 2012; Hafstrom *et al.*, 2011; Wanchanthuek *et al.*, 2010; Motro *et al.*, 2009).

1.4 Next generation sequencing

Next generation sequencing (NGS), selected by the journal Nature as "method of the year" (Schuster, 2008), with its low cost and high throughput has revolutionized the field of genomics. Before the advent of NGS, the Maxam and Gilbert chemical degradation method (Maxam & Gilbert, 1977) and Sanger enzymatic dideoxy technique (Sanger *et al.*, 1977) were used for sequencing purposes. These techniques were initially used to decipher complete genes, and later, complete genomes. The first complete genome to be sequenced was that of a virus (Fiers *et al.*, 1978) performed using Sanger sequencing. The technique prevailed in the sequencing community until 2001 when the first human genome was sequenced (Lander *et al.*, 2001). Completion of the human genome project took a long time and considerable resources were used. It was necessary to develop faster, cheaper and high throughput technologies. Next generation sequencing technologies were thus introduced with the goal to overcome the shortcomings of Sanger sequencing. The first NGS technology was 454 pyrosequencing (Roche) introduced in 2005 (Margulies *et al.*, 2005). Later Solexa, Solid (Valouev *et al.*, 2008) and Ion Torrent technologies were also released. The introduction of NGS technologies has lowered the cost of sequencing of a human genome from $100,000,000 in 2001 to $1000 in 2014 (van Dijk *et al.*, 2014), however these technologies have their own pros and cons. Some of the features of 454, Solexa/Illumina, solid and Ion Torrent have been summarized in Table 2.

Table 2: Comparison of different features of 454, Solexa, Solid and Ion Torrent methods/technologies (van Dijk et al., 2014)

Technology	Maximum read length	Maximum throughput (GB)	Runtime, bacterial genome (hours)	Disadvantage
454	1000	0.7	10	Homopolymer errors
Solexa/Illumina	300	1800	240	Long run time
Solid	75	320	336	Short reads and long run time
Ion Torrent	400	10	3	Homopolymer errors

The advent of NGS did not just revolutionize human genomic research, but it also had major impact on other genomic fields, including bacterial genomics. Large numbers of bacterial genomes have been sequenced since 2005 with the intention to understand the mechanisms involved in their role in nature. For pathogenic bacteria, the main aim of sequencing is usually to find potential virulence genes. The availability of large numbers of bacterial genomes has also increased the use of whole genome sequencing for bacterial classification purposes. This thesis provides insights into both the use of genomic sequences for screening of potential virulence genes in *Treponema* genomes and the classification of a species of *Brachyspira*.

1.5 Background for the thesis

1.5.1 Paper I, II and III

Spriochaetes of the genus *Treponema* are fastidious bacteria that require an anaerobic environment. These bacteria are difficult to grow which makes it difficult to perform *in vitro* characterization and manipulation of these bacteria. Since the first successful isolation of *T. pedis* from porcine ear necrosis (Pringle *et al*. 2009), different isolates of *T. pedis* have been obtained from ear necrosis, shoulder ulcers and gingiva of pigs (Svartstrom *et al*. 2013). Despite

being isolated from porcine necrotic lesions, the role of *T. pedis* in these skin ulcers is yet to be understood. Similarly, the pathogenic potential of *T. phagedenis* in BDD has not been determined. *Treponema phagedenis*is one of a few cultivable treponemes from BDD lesions, but investigations of underlying pathogenicity factors have only just begun. In Sweden, *T. phagedenis* strain V1 and several other isolates have been obtained from BDD lesions in dairy cattle (Rosander *et al.*, 2011; Pringle *et al.*, 2008)(Paper 3, this thesis). The idea behind the current study was that genomes of these bacteria could contain information on potential pathogenicity factors, and that bioinformatics analysis of their genome sequences will provide insights into pathogenic mechanisms.

1.5.2 Paper IV

Brachyspira suanatina which causes swine dysentery-like enteric disease in pigs, was recently suggested a new species within genus *Brachyspira*. Based on phenotypic characterization, *B. suanatina* is similar to *B. hyodysenteriae*, the agent of swine dysentery, but they differ genetically based on 16S rRNA and *nox* gene phylogeny. We therefore hypothesised that *B. suanatina* is a separate species, and that the taxonomic rank would be elucidated by performing genomic comparisons of *B. suanatina* with other *Brachyspira* species.

2 Aims of the thesis

The main aim of the thesis was to produce complete and annotated genome sequences that allows in depth bioinformatics analysis that explain the biological characteristics of *T. pedis*, *T. phagedenis* and *B. suanatina* and the specific aims were to:

- Get the complete and annotated genomes of *T. pedis* TA4, *T. phagedenis* V1 and *B. suanatina* AN4859/03.

- Obtain the draft genome assemblies of additional *T. pedis* and *T. phagedenis* isolates.

- Perform whole genome comparisons of *Treponema* and *Brachsyspira* genomes.

- Identify potential pathogenicity related factors in the genomes of *T. pedis* and *T. phagedenis*.

- Describe the genomic characteristics of *B. suanatina* for a complete species validation and species recognition.

3 Considerations on Materials and Methods

In order to meet the aims of the thesis, we performed whole genome sequencing and analyses of *T. pedis* (paper I), *T. phagedenis* (paper II, III) and *B. suanatina* (paper IV) genomes. Below are the steps involved in the analysis of these genomes.

3.1 DNA sequencing

The first step in any whole genome analysis project is to get the genomes sequenced, this could be done using different commercial NGS technological platforms. Some important considerations while choosing the platform are availability of a reference genome and the purpose of sequencing. For projects concerning *de novo* genome sequencing, it is important to get longer reads with high throughput, high accuracy and paired end information to get high quality, and long contig/scaffold assemblies. All the genomes used in this thesis were *de novo* and it was not possible to get all of these attributes in a single platform, especially at the beginning of the study. We therefore used different combinations of Roche's FLX 454, Illumina Hiseq and Miseq and Ion torrent platforms.

For sequencing of *T. pedis* strain TA4 (paper I) and *T. phagedenis* strain V1 (papers II and III) genomes, we first used 454 sequencing platform to obtain long reads (200–600 bp). Long reads allow maximum overlap of reads in the assembly process producing more reliable assemblies as compared to short reads. But, because of homopolymer errors (detection of wrong number of similar bases) generated by 454 sequencing and single end data, it was not possible to use 454 sequencing alone. Therefore, additional sequencing generating paired-end reads was performed using Illumina Hiseq platform for these genomes. Paired end reads for additional isolates of *T. pedis* (paper I) and

T. phagedenis (paper III) were obtained using the Illumina Miseq platform that generated reads of approximately 300 bp length.

The genome sequence of *B. suanatina* AN4859/03 (paper IV) was produced in 2013 and by then, Ion Torrent was already introduced. Ion Torrent with its long read length and low cost was an ideal alternative to 454, therefore we obtained single end reads from Ion Torrent. Ion Torrent uses a similar sequencing chemistry as 454, which generates homopolymer errors. The data was therefore complemented with the paired end reads from Illumina Miseq platform. For paper I, the Illumina reads (2×100 bp) that were used to assemble draft genome sequences from 12 *T. denticola* strains were downloaded from GenBank, SRA (http://www.ncbi.nlm.nih.gov/sra) and converted to FASTQ format using the SRA toolkit (NCBI).

3.2 Genome Assembly

Genome assembly is a process of turning a jigsaw puzzle of millions of raw sequencing reads into a full chromosome. However, the process is hampered by the presence of repeat sequences, missing sequences and low quality sequences leading to long continuous sequences of various lengths instead of a complete chromosome. Dozens of assembly programs are available using different assembly algorithms with the aim to minimize the number of contigs and maximize the length of each contig. An important thing therefore, is to choose the right assembly program that could fit one's needs and resources.

For all the genomes used in this thesis, *de novo* assemblies were made because of unavailability of reference genomes. For *T. phagedenis* (studies II and III), two already sequenced genomes of *T. phagedenis* strains F0421 and 4A were available, but due to their draft nature they could not be used as a reference. The choice of assembler depended majorly on the platform used for sequencing. For the *T. phagedenis* V1 genome (paper II), we used Newbler (Roche), which was developed specifically for 454 data and for *de novo* genome assemblies, for making assemblies of 454 data as well as hybrid assemblies of 454 and Illumina data in combination. For making hybrid assemblies of 454 and Illumina data the MIRA assembler (Chevreux *et al.*, 1999) could also be used but, because of its extensive memory requirements and long run times, Newbler was the preferred choice. For the *T. pedis* TA4 genome (paper I), we used Newbler for 454 data assembly. Since a single contig was obtained with 454 data and PCR amplification of some areas, Illumina reads were mapped to that contig to remove misassemblies, specifically homopolymer errors. For all the additional isolates of *T. pedis*, *T. denticola* and *T. phagedenis* (papers I and III) we used MIRA assembler.

For the *B. suanatina* AN4859/03 genome (paper IV), Ion Torrent reads were assembled using Newbler assembler because of the similarity of data produced by Ion Torrent and 454. For the data produced by Illumina sequencing of *B. suanatina* AN4859/03, we used the MIRA assembler. Assemblies produced by Newbler and MIRA were compared in MAUVE (Darling *et al.*, 2004) genome aligner.

For visualization and manual editing of the assembled contigs obtained from all assemblies, we used Consed (Gordon, 2003). Consed supports visualization of assemblies from both Newbler and MIRA and also allows mapping of reads to the contigs. This feature proved very useful for manual joining of contigs; after joining different contigs, reads were aligned to the new set of contigs and all the wrong joinings were split again based on the reads covering the joined contigs.

Scaffolding of the *T. phagedenis* V1 (paper II) and *B. suanatina* AN4859/03 (paper IV) genomes was performed with SSPACE (Boetzer *et al.*, 2011), using Illumina paired end reads. For scaffolding, we used all the reads that were not used in the assembly process due to data filtration done for coverage reduction.

3.3 Annotation

The process to deduce biological information from sequenced and assembled genomes is called annotation. Usually, in a genome analysis project, the first thing after the assembly is to predict the ORFs and then assign functions to the predicted ORFs. Different tools as well as online servers are available for annotation purposes. In this thesis, we have used different sets of tools and pipelines for different studies. The ORF predictions in *T. pedis* TA4 and all additional *T. pedis* and *T. denticola* isolates were performed using Glimmer 3 (Delcher *et al.*, 1999) (paper I). Transfer RNA genes were predicted using tRNA scan (Lowe & Eddy, 1997) and 16S, 5S and 23S rRNA genes were predicted on the basis of their homology with the corresponding genes in *T. denticola*. Function prediction of the putative coding sequences was performed using BLASTP (Altschul *et al.*, 1990) searches against all-bacterial genome database. Functional assignment to a CDS was done on the basis of best BLASTP hit where >30% amino acid identity was found in the alignment and length difference was <25% and e-value was $<1x10^{-6}$. CDS's overlapping at least 50% with a tRNA, rRNA or another CDS was removed from the annotation. Comparative analysis of *T. pedis* and *T. denticola* genomes was performed using BLASTP comparisons of the predicted CDSs in their genomes. For the *T. phagedenis* V1 (paper II) and *B. suanatina* AN4859/03

(paper IV) genomes, automated annotation platforms MAGE (Vallenet *et al.*, 2009) and GenDB (Meyer *et al.*, 2003) were used respectively for genes prediction and functional classification of genes. For comparative analysis, the EDGAR (Blom *et al.*, 2009) platform was used in paper IV. For the *T. phagedenis* isolates used in paper III, whole genome annotation was not performed. Only a specific locus was predicted in the draft genome assemblies of all isolates. Local BLAST searches against their genome assemblies were performed using *T. phagedenis* V1's genes sequences present in the locus as a query. For prediction of bacteriophage regions and lipoproteins, PHAST (Zhou *et al.*, 2011) and SpLip (Setubal *et al.*, 2006) were used respectively.

3.4 Phylogenetic analysis and ANI

Phylogenetic analysis using the 16S rRNA gene has been performed since 1977 (Woese & Fox, 1977) for the classification of bacteria. However, the degree of 16SrRNA gene sequence similarity in some species is very high, that is why it cannot always be used as a reliable phylogenetic marker (Fox *et al.*, 1992). With the advent of NGS technologies, the cost and time of sequencing a bacterial genome has been reduced considerably, enabling taxonomists to use whole genome information for the classification of bacteria (den Bakker *et al.*, 2013). In this thesis, phylogenetic analyses were performed in paper I, II and IV. Phylogenetic analysis of the intergenic spacer region between the 16S rRNA and $tRNA_{Ile}$ genes was performed in paper I to find the intrastrain variability in *T. pedis* and *T. denticola.* In paper II, 16S rRNA gene phylogeny was performed to show the taxonomical position of *T. phagedenis* V1with respect to other *Treponema* species.

In paper IV, the aim was to perform phylogenetic analysis for species validation. Therefore, 25 housekeeping genes and the core genomes of all valid *Brachyspira* species were used. The general procedure used during the phylogenetic analysis was to perform the alignment using CLUSTAL (Larkin *et al.*, 2007) and/or MUSCLE (Edgar, 2004) algorithms. Conserved blocks in the alignments were then selected using Gblocks (Castresana, 2000) and a phylogenetic tree was constructed using ML (Felsenstein, 1981) or NJ (Saitou & Nei, 1987) methods. Moreover, ANI was calculated between *B. suanatina* and type strains of all the seven valid *Brachyspira* species using Jspecies v1.2.1 (Richter & Rosselló-Móra, 2009).

4 Results and Discussion

4.1 Whole genome sequence and comparative analysis of *Treponema pedis* (Paper I)

The complete genome sequence of *T. pedis* strain TA4 isolated from a case of pig ear necrosis was obtained. According to the general genomic features, the *T. pedis* TA4 genome consisted of 2,889,325 bp and the GC content was 37.9%. There were 2086 putative CDSs, 45 putative tRNA genes and 6 rRNA genes. The *T. pedis* TA4 genome was most closely related to the *T. denticola* ATCC 35405 genome, sharing 2077 (~74%) CDSs.

In order to investigate the relatedness of different strains of *T. pedis* with *T. denticola*, and to find potential pathogenicity factors in the *T. pedis* genome, draft genome assemblies of 6 additional *T. pedis* isolates and 12 additional *T. denticola* strains were obtained using the *T. pedis* TA4 genome and the *T. denticola* ATCC 35405 genome as their respective reference. Additional *T. pedis* isolates were obtained from the gingiva and necrotic lesions in pigs (Svartstrom *et al.*, 2013). Illumina reads for genomes of 12 *T. denticola* strain were downloaded from the GenBank Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra). Draft genome assemblies of *T. pedis* isolates ranged in size from 2.95 to 3.47 Mbp with a GC content varying between 36.9 and 37.3%, and those of the *T. denticola* strains ranged in size from 2.76 to 3.03 Mbp with a GC content varying between 37.7 and 38.0%.

Pan and core genomes of T. pedis and T. denticola

Clusters of intraspecies homologous genes in *T. pedis* and *T. denticola* were produced by collapsing CDSs sharing >80% amino acid identity with <30% deviation in length. In *T. pedis*, a total of 8244 gene clusters were produced. Of these, 988 clusters formed a conserved core, 576 clusters were strain-specific and the remaining clusters showed intermediate representation of genes.

Similarly, in *T. denticola* 7269 gene clusters were obtained of which, 1115 were core gene clusters, 224 strain-specific gene clusters and the remaining clusters represented intermediate representation. Signs of lateral gene transfer events were also identified in the genome of *T. pedis* by the presence of genes homologous to genes from other species.

Putative Pathogenicity related factors

Different pathogenicity related genes in the genome of *T. pedis* TA4 were predicted by their similarity to homologous genes in *T. denticola*. These included:

- Surface antigen (TDE2258)
- Motility genes
- Proteases
- Dentilisin operon
- PtrB oligopeptidases
- IgG-specific protease dentipain
- Major outer sheath protein

4.2 Whole genome sequence of *T. phagedenis* and putative pathogenicity related factors (Papers II and III)

A high quality draft genome assembly of *T. phagedenis* strain V1 was obtained. The assembly consisted of 51 scaffolds comprising 3,129,551 bp and a GC content of 39.9%. In the draft genome assembly of *T. phagedenis* V1, 3157 protein-coding genes were predicted. Also 45 tRNA and 6 rRNA genes were found.

4.2.1 Putative pathogenicity related factors

In order to find potential pathogenicity factors, we searched the genome for some already suggested pathogenicity factors in *T. denticola* and *T. pallidum* subsp. *pallidum*. The amino acid sequences of these proteins were obtained from the complete genomes of *T. denticola* strain ATCC 35404 (accession number: NC_002967) and *T. pallidum* subsp. *pallidum* strain Nichols (accession number: NC_000919). These sequences were then blasted against protein sequences in the *T. phagedenis* V1's genome. Sequences showing more than 30% amino acid identity with e-value < 0.00005 are shown in Table 3.

Additionally, the genome of *T. phagedenis* strain V1 contained 3 putative prophage regions, 17 CDSs encoding for putative transposases, 22 CDSs encoding for motility and chemotaxis related genes and 155 lipoproteins.

Table 3: Putative pathogenicity related proteins in T. denticola *strain ATCC 35405 and* T. pallidum *subsp*. pallidum *strain Nichols with homologues in* T. phagedenis *V1*

***T. phagedenis* V1 protein locus_tag**	***Treponema* spp locus_tag[1]**	**Gene product**	**Amino acid identity (%)**
TPHV1 _10302	TP0326	Antigen	56
TPHV1_20066	TP0453	Antigen	40
TPHV1_40181	TP0751	Laminin-binding protein	42
TPHV1_510060	TP0155	Fibronectin-binding protein	58
TPHV1_290003	TP0136	Fibronectin binding protein	37
TPHV1_40016	TP0487	Antigen	59
TPHV1_10302	TP0971	Membrane antigen, pathogen-specific Tpd	58
TPHV1_190050	TP0257	Glycerophosphodiester phosphodiesterase (Gpd)	60
TPHV1_100034	TDE_0405	Major outer sheath protein	38
TPHV1_130036	TDE_2258	Surface antigen BspA	55
TPHV1_60100	TDE_2056	Hemin Binding Protein A (HbpA)	49
TPHV1_60100	TDE_2055	Hemin Binding Protein B (HbpB)	63
TPHV1_30021	TDE_0842	Cytoplasmic filament protein A (CfpA)	82

[1] Locus_tag starting with TDE refers to *T. denticola* protein, locus tag starting with TP refers to *T. pallidum* subsp. *pallidum* protein.

4.2.2 Locus encoding putative lipoproteins with potential for antigenic and phase variation

A genomic locus encoding for the probable lipoproteins VpsA, PrrA and VpsB, with potential for phase and antigenic variation, was identified in the genome of *T. phagedenis* strain V1. The identification was performed by manual curation of the genome. One of the proteins, PrrA, is an already described immunogenic protein in *T. phagedenis* V1 (Rosander *et al.*, 2011). The amino acid sequence of the PrrA protein contained a putative signal peptide followed by several amino acid repeat motifs. Also, the promoter spacer region between the -10 and -35 elements of the putative promoter sequence of the *prrA* gene contained dinucleotide $(TA)_6$ repeats.

In close proximity of the *prrA* gene, two more genes with highly similar promoter sequences were found. These genes were designated as v*psA* and *vpsB*. The encoded proteins, VpsA and VpsB, were also experimentally shown to be immunogenic in enzyme-linked immunosorbent assays (ELISAs). Three additional ORFs were detected within this locus, two between *vpsA* and *prrA* and one between *prrA* and *vpsB.* Translated sequences of two of them shared significant similarity to putative transposase domain containing proteins in *T. denticola*, thus designated as putative transposases 1 and 2.

In order to further analyse this particular locus, draft genome assemblies of 12 additional *T. phagedenis* isolates from BDD lesions (Rosander *et al.*, 2011; Pringle *et al.*, 2008) (paper III) were also obtained. Draft genome assemblies of *T. phagedenis* strains F0421 (GCF_000187105) and 4A (GCF_000513775) were downloaded from Genbank . Local BLAST searches of the predicted genes in the locus were performed in all assemblies to investigate the presence and organization of the genes. The results obtained from BLAST searches showed that all isolates contained the *vpsA* and *vpsB* genes, whereas the *prrA* gene was missing in four isolates. Isolates lacking the *prrA* gene also lacked the putative transposase 2 gene, suggesting its potential involvement in genetic transfer of *prrA.* None of the genes in the locus were present in *T. phagedenis* F0421. *Treponema phagedenis* F0421 is of human origin and considered a harmless commensal. The absence of the locus from this strain suggests that these genes may have a role in the pathogenesis of BDD. However, it cannot be completely ruled out that the failure to detect these genes in F0421 could have been caused by sequencing problems.

Promoter analysis

Dinucleotide TA repeats in the promoter spacer region is a feature that has been shown to regulate the expression of genes in *Mycoplasma mycoides* subsp. *mycoides* (Persson *et al.*, 2002) in a phase variable manner. Different

numbers of TA repeats in the promoter spacers of *vpsA*, *prrA* and *vpsB* were found in different isolates. The promoter spacers most commonly consisted of 16 nucleotides [TAAA$(TA)_6$ or $(TA)_8$] and resulted in protein expression in all cases. Promoter spacers with 18 nucleotides [TAAA$(TA)_7$ or $(TA)_9$] also resulted in expression of PrrA and VpsB while it was not possible to detect any protein expression from promoters with 14 nucleotide spacers [$(TA)_7$ in *vpsB* promoters], where the promoter sequences had been clearly defined. However, in two isolates (V1 and T 551B, both Western blot positive) where the number of TA repeats in the *vpsB* promoter could not be determined, it is possible that 14 nucleotide spacers [$(TA)_7$] also allowed expression of the gene.

Amino acid sequence analysis

In the amino acid sequences of PrrA and VpsB, different repeat motifs varying in copy number in different isolates, were identified. The motifs, KAEEKKPE, PGKEE and PGTEKPVA were found in PrrA in all isolates in varying numbers, except in isolate T 2378 where the PGKEE motif could not be identified. In VpsB, the motif CSGLTSIDLSACTKLTSI was present in different numbers in different isolates, flanked by a TLPDGLTSIG motif. Also, a part of a motif common with PrrA, KAEEKK, was present. There was no obvious repeat motif found in VpsA. The presence of different numbers of repeat motifs has also been reported in *Mycoplasma bovis*, being utilized by the bacteria for antigenic variation (Lysnyansky *et al.*, 1999).

4.3 Draft genome assembly of *B. suanatina* strain AN4859/03 and its comparison with *B. hyodysenteriae* and *B. intermedia* (Paper IV)

In this study we produced the draft genome sequence of *B. suanatina* strain AN4859/03. The draft genome assembly of *B. suanatina* consisted of 35 scaffolds comprising 3,263,337 bp with a GC content of 27%. One of the scaffolds in the *B. suanatina* genome assembly contained a putative plasmid sequence of 30,236 bp sharing 88% identity over 51% of its length with the *B. hyodysenteriae* strain WA1 plasmid (pBHWA1) sequence. The ANI values calculated using the draft genome of *B. suanatina* and the genomes of type strains of all valid species of genus *Brachyspira*, were always less than 95%, which is the suggested threshold for species demarcation (Goris *et al.*, 2007). ANI values correlated well with their corresponding DNA-DNA hybridization values that are considered to be the gold standard of prokaryotic classification. Based on the values obtained from ANI and DNA-DNA hybridization, we suggest that *B. suanatina* is a novel bacterial species. Further, we performed

phylogenetic analyses using, 25 housekeeping genes and the core genome of all available currently recognized *Brachyspira* species. According to the phylograms obtained, *B. suanatina* formed a clade with *B. intermedia*, distinct from the *B. hyodysenteriae* clade but sharing a common ancestor that strengthens our hypothesis that *B. suanatina* should be regarded as a novel species.

Genomic analyses of *B. suanatina* AN4859/03, *B. hyodysenteriae* WA1 and B78^T and *B. intermedia* PWS/A^T showed that the genomes of these three species are very similar in terms of GC content, number of genes, presence of homologous genes and distribution of genes in (clusters of orthologous groups) COG categories. However, the genome size of *B. hyodysenteriae* WA1 was slightly smaller as compared to *B. suanatina* and *B. intermedia*. The reason for the smaller genome size of *B. hyodysenteriae* could be that this species has undergone reductive evolution; a process of reduction in genome size of a host associated bacteria by the loss of genes rendered non-essential (Wixon, 2001). A bacteriophage region, BSP1 was found in the genome of *B. suanatina*. The BSP1 region was partly conserved in *B. hyodysenteriae* B78^T but it was not found in *B. hyodysenteriae* strain WA1 or *B. Intermedia* strain PWS/A^T. The presence of this bacteriophage in two strains of different but closely related species, and absence in the other two is also possibly due to reductive evolutionary events. Putative horizontal gene transfer events were also evident by the presence of genes homologous to genes in *Clostridium* spp. and *Bacillus* spp.

5 Conclusions

By performing whole genome sequencing and bioinformatics analyses of the genomes of *T. pedis*, *T. phagedenis* and *B. suanatina* we can conclude that:

- A complete genome sequence of *T. pedis* TA4, and high quality draft genome assemblies of *T. phagedenis* V1 and *B. suanatina* AN4859/03 were generated.

- Several putative pathogenicity factors were identified in the genomes of *T. pedis* and *T. phagedenis*.

- Extensive interspecies genomic similarities between *T. pedis* and *T. denticola* as well as large intraspecies genomic variability within each species were found.

- A locus containing putative lipoproteins with potential for antigenic variation has been found in the genomes of *T. phagedenis* BDD isolates. Variations in occurrence, sequence, and expression of three genes within this locus exist between isolates.

- We have provided *in silico* support for the classification of *B. suanatina* as a novel species within the genus *Brachyspira* based on its genomic characteristics.

6 Future perspectives

Treponema phagedenis V1 (paper II) and *B. suanatina* AN4859/03 (paper IV) genomes were left unfinished due to the use of short reads NGS technologies that do not always allow completion of genomes without extensive manual work. However, in the future this problem could be addressed using third generation sequencing technologies like Pacific Biosciences: (PACBIO) RS that could generate longer reads in less time as compared to previous NGS technologies. Complete genomes of *T. phagedenis* V1 and *B. suanatina* AN4859/03 could be obtained and used as references for genome mapping and comparisons of other strains of these species.

Putative pathogenicity factors identified in the genomes of *T. pedis* (paper I) and *T. phagedenis* (paper II) need to be further characterized *in vitro* in order to understand their role in the pathogenesis of skin lesions in pigs and cattle.

Results obtained from paper III could be used as a basis for further *in silico* and *in vitro* studies on immunogenicity and antigenic- and phase variation in *T. phagedenis*. Results from protein expression analysis by Western blot should be supplemented with mRNA transcription data by cDNA synthesis and qPCR analysis of *prrA*, *vpsA* and *vpsB* transcripts. This is important since cross-reactivity was detected for the anti-vpsB antibody, which may give false positive results. Additionally, surface exposure of the proteins could be determined using e.g. fluorescent-labelled antibodies against the three proteins. Investigations on which, if any, of the repeats that are associated with antigenicity can be performed. This could be done by performing the *in silico* structure prediction and molecular docking of these proteins with their specific antibodies. Results could be later verified *in vitro* where synthesized overlapping peptides can be used in ELISAs to see the regions of the proteins that are recognized by sera from infected animals. Finally, results from the promoter analysis were not very reliable since, for most of the isolates, reads covering the particular regions were of low quality and coverage. This

prevented the use of a custom -Perl script on the reads data for predicting the accurate number of TA repeats in the *prrA*, *vpsA* and *vpsB* promoter spacers in the different isolates except for V1. Sequencing with high throughput can be performed to provide greater read coverage for promoter analysis and improved phase variability prediction at strain level.

In paper IV, we have analysed the genome sequence of one strain, AN4859/03, which was isolated from pig faeces for a complete species validation of *B. suanatina*. There are other isolates of *B. suanatina* from pigs and mallards (Jansson *et al.*, 2009; Rasback *et al.*, 2007). Sequencing just one genome may not give us a complete genomic picture of the species. Therefore, in order to achieve a better understanding of the species, whole genome sequencing of additional isolates could be performed and used for comparative analysis. Also another phylogenetic study could be performed using whole genome data of these isolates along with whole genome data of more strains of other *Brachyspira* species. Using more than one strain of a species will provide a better taxonomical resolution of the genus. In this study, we have also compared genomic synteny between *B. suanatina*, *B. hyodysenteriae* and *B. intermedia*. Due to the draft nature of *B. suanatina* genome, the results obtained from synteny analysis are not fully accurate and completely reliable, this problem could also be overcome in the future by the use of complete genome of *B. suanatina* AN4859/03.

This thesis provides insights into the putative pathogenicity related factors in the genomes of *T. pedis* and *T. phagedenis*. *In vitro* assessment of these factors will aid in understanding the role of these bacteria in the pathogenesis of porcine skin ulcers and bovine digital dermatitis. Results obtained from the thesis, may also be of help in improved disease diagnostics and treatment of these diseases.

Further, results of this thesis suggest that *B. suanatina* should be regarded as a novel species. This study highlights the importance of integrating genomic information in the taxonomy of bacteria.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol,* 215(3), pp. 403-10.

Baril, C., Richaud, C., Baranton, G. & Girons, I.S. (1989). Linear chromosome of *Borrelia burgdorferi*. *Research in microbiology,* 140(7), pp. 507-516.

Bellgard, M.I., Wanchanthuek, P., La, T., Ryan, K., Moolhuijzen, P., Albertyn, Z., Shaban, B., Motro, Y., Dunn, D.S., Schibeci, D., Hunter, A., Barrero, R., Phillips, N.D. & Hampson, D.J. (2009). Genome sequence of the pathogenic intestinal spirochete *Brachyspira hyodysenteriae* reveals adaptations to its lifestyle in the porcine large intestine. *PLoS One,* 4(3), p. e4641.

Blom, J., Albaum, S.P., Doppmeier, D., Pühler, A., Vorhölter, F.-J., Zakrzewski, M. & Goesmann, A. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *Bmc Bioinformatics,* 10(1), p. 154.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics,* 27(4), pp. 578-579.

Bruijnis, M.R.N., Hogeveen, H. & Stassen, E.N. (2010). Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. *J Dairy Sci,* 93(6), pp. 2419-2432.

Bulach, D.M., Zuerner, R.L., Wilson, P., Seemann, T., McGrath, A., Cullen, P.A., Davis, J., Johnson, M., Kuczek, E. & Alt, D.P. (2006). Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proceedings of the National Academy of Sciences,* 103(39), pp. 14560-14565.

Busch, M., Jensen, I. & Korsgaard, J. (2010). The development and consequences of ear necrosis in a weaner herd and two growing-finishing herds. *Proc Inter Pig Vet Soc Cong,* 45.

Canale-Parola, E. (1977). Physiology and evolution of spirochetes. *Bacteriological reviews,* 41(1), p. 181.

Casjens, S., Palmer, N., Van Vugt, R., Mun Huang, W., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J. & Dodson, R.J. (2000). A bacterial

genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Molecular microbiology,* 35(3), pp. 490-516.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution,* 17(4), pp. 540-552.

Charon, N., Greenberg, E., Koopman, M. & Limberger, R. (1992). Spirochete chemotaxis, motility, and the structure of the spirochetal periplasmic flagella. *Research in microbiology,* 143(6), pp. 597-603.

Charon, N.W. & Goldstein, S.F. (2002). Genetics of motility and chemotaxis of a fascinating group of bacteria: the spirochetes. *Annual Review of Genetics,* 36(1), pp. 47-73.

Charon, N.W., Goldstein, S.F., Marko, M., Hsieh, C., Gebhardt, L.L., Motaleb, M.A., Wolgemuth, C.W., Limberger, R.J. & Rowe, N. (2009). The flat-ribbon configuration of the periplasmic flagella of *Borrelia burgdorferi* and its relationship to motility and morphology. *Journal of Bacteriology,* 191(2), pp. 600-607.

Chauhan, S. & Kuramitsu, H.K. (2004). Sequence analysis of plasmid pTS1 isolated from oral spirochetes. *Plasmid,* 51(1), pp. 61-65.

Cheli, R.a.M., C. (1974). La dermatite digitale del bovino. *Proceedings of 8th International Conference on Diseases of Cattle, Piacenza, Milan, Italy*, p. 6.

Chevreux, B., Wetter, T. & Suhai, S. Genome sequence assembly using trace signals and additional sequence information. In: *Proceedings of German Conference on Bioinformatics*1999, pp. 45-56.

Collighan, R.J. & Woodward, M.J. (1997). Spirochaetes and other bacterial species associated with bovine digital dermatitis. *FEMS microbiology letters,* 156(1), pp. 37-41.

Darling, A.C., Mau, B., Blattner, F.R. & Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res,* 14(7), pp. 1394-1403.

Davies, P., Morrow, W., Miller, D. & Deen, J. (1996). Epidemiologic study of decubital ulcers in sows. *Journal of the American Veterinary Medical Association,* 208(7), pp. 1058-1062.

Delcher, A.L., Harmon, D., Kasif, S., White, O. & Salzberg, S.L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res,* 27(23), pp. 4636-41.

den Bakker, H.C., Manuel, C.S., Fortes, E.D., Wiedmann, M. & Nightingale, K.K. (2013). Genome sequencing identifies *Listeria fleischmannii* subsp. *coloradonensis* subsp. nov., isolated from a ranch. *Int J Syst Evol Microbiol,* 63(Pt 9), pp. 3257-68.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* 32(5), pp. 1792-1797.

el-Ghoul, W. & Shaheed, B.I. (2001). Ulcerative and papillomatous digital dermatitis of the pastern region in dairy cattle: clinical and

histopathological studies. *Dtsch Tierarztl Wochenschr*, 108(5), pp. 216-22.
Evans, N.J., Brown, J.M., Demirkan, I., Murray, R.D., Birtles, R.J., Hart, C.A. & Carter, S.D. (2009). *Treponema pedis* sp. nov., a spirochaete isolated from bovine digital dermatitis lesions. *International journal of systematic and evolutionary microbiology*, 59(5), pp. 987-991.
Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*, 17(6), pp. 368-76.
Fenno, J.C., Hannam, P.M., Leung, W.K., Tamura, M., Uitto, V.J. & McBride, B.C. (1998). Cytopathic effects of the major surface protein and the chymotrypsinlike protease of *Treponema denticola*. *Infect Immun*, 66(5), pp. 1869-77.
Fiers, W., Contreras, R., Haegemann, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. & Ysebaert, M. (1978). Complete nucleotide sequence of SV40 DNA. *Nature*, 273(5658), pp. 113-20.
Fox, G.E., Wisotzkey, J.D. & Jurtshuk, P. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol*, 42(1), pp. 166-170.
Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R. & Hickey, E.K. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature*, 390(6660), pp. 580-586.
Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G., Dodson, R., Gwinn, M., Hickey, E.K., Clayton, R. & Ketchum, K.A. (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science*, 281(5375), pp. 375-388.
Godovikova, V., Wang, H.-T., Goetting-Minesky, M.P., Ning, Y., Capone, R.F., Slater, C.K. & Fenno, J.C. (2010). *Treponema denticola* PrcB is required for expression and activity of the PrcA-PrtP (dentilisin) complex. *Journal of Bacteriology*, 192(13), pp. 3337-3344.
Gordon, D. (2003). Viewing and editing assembled sequences using Consed. *Current protocols in bioinformatics*, pp. 11.2. 1-11.2. 43.
Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P. & Tiedje, J.M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*, 57(1), pp. 81-91.
Gupta, R.S., Mahmood, S. & Adeolu, M. (2013). A phylogenomic and molecular signature based approach for characterization of the phylum *Spirochaetes* and its major clades: proposal for a taxonomic revision of the phylum. *Frontiers in microbiology*, 4.
Haake, D.A. (2000). Spirochaetal lipoproteins and pathogenesis. *Microbiology*, 146 (Pt 7), pp. 1491-504.
Hafstrom, T., Jansson, D.S. & Segerman, B. (2011). Complete genome sequence of *Brachyspira intermedia* reveals unique genomic features in *Brachyspira*

species and phage-mediated horizontal gene transfer. *BMC Genomics,* 12, p. 395.

Håfström, T., Jansson, D.S. & Segerman, B. (2011). Complete genome sequence of *Brachyspira intermedia* reveals unique genomic features in *Brachyspira* species and phage-mediated horizontal gene transfer. *BMC Genomics,* 12(1), p. 395.

Hampson, D.J. & La, T. (2006). Reclassification of *Serpulina intermedia* and *Serpulina murdochii* in the genus *Brachyspira* as *Brachyspira intermedia* comb. nov. and *Brachyspira murdochii* comb. nov. *International journal of systematic and evolutionary microbiology,* 56(5), pp. 1009-1012.

Han, C., Gronow, S., Teshima, H., Lapidus, A., Nolan, M., Lucas, S., Hammon, N., Deshpande, S., Cheng, J.-F. & Zeytun, A. (2011). Complete genome sequence of *Treponema succinifaciens* type strain (6091T). *Stand Genomic Sci,* 4(3), p. 361.

Harris, D., Glock, R., Christensen, C. & Kinyon, J.M. (1972). Inoculation of pigs with *Treponema hyodysenteriae* (new species) and reproduction f the disease. *Veterinary medicine, small animal clinician: VM, SAC,* 67(1), pp. 61-64.

Hovind-Hougen, K., Birch-Andersen, A., Henrik-Nielsen, R., Orholm, M., Pedersen, J., Teglbjaerg, P. & Thaysen, E. (1982). Intestinal spirochetosis: morphological characterization and cultivation of the spirochete *Brachyspira aalborgi* gen. nov., sp. nov. *J Clin Microbiol,* 16(6), pp. 1127-1136.

Hovind-Hougen, K., Birch-Andersen, A., Henrik-Nielsen, R., Orholm, M., Pedersen, J., Teglbjaerg, P. & Thaysen, E. (2011). Genus I. *Brachyspira*. In: Krieg NR, S.J., Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (eds) (ed. *Bergey's Manual of Systematic Bacteriology*4). New York: Springer, pp. 531-544.

Humphrey, S.B., Stanton, T.B., Jensen, N.S. & Zuerner, R.L. (1997). Purification and characterization of VSH-1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*. *J Bacteriol,* 179(2), pp. 323-9.

Ishihara, K., Kuramitsu, H.K., Miura, T. & Okuda, K. (1998). Dentilisin activity affects the organization of the outer sheath of *Treponema denticola*. *J Bacteriol,* 180(15), pp. 3837-44.

Jansson, D., Råsbäck, T., Fellström, C. & Feinstein, R. (2009). Experimental Challenge of Mallards (*Anas platyrhynchos*) with *Brachyspira hyodysenteriae* and "*Brachyspira suanatina*" Isolated from Pigs and Mallards. *Journal of comparative pathology,* 141(4), pp. 211-222.

Karlsson, F., Klitgaard, K. & Jensen, T.K. (2014). Identification of *Treponema pedis* as the predominant *Treponema* species in porcine skin ulcers by fluorescence *in situ* hybridization and high-throughput sequencing. *Veterinary microbiology,* 171(1), pp. 122-131.

Kinyon, J. & Harris, D. (1979). *Treponema innocens*, a new species of intestinal bacteria, and emended description of the type strain of *Treponema hyodysenteriae* Harris et al. *Int J Syst Bacteriol,* 29(2), pp. 102-109.

Klitgaard, K., Bretó, A.F., Boye, M. & Jensen, T.K. (2013). Targeting the treponemal microbiome of digital dermatitis infections by high-resolution phylogenetic analyses and comparison with fluorescent in situ hybridization. *J Clin Microbiol*, 51(7), pp. 2212-2219.

Koenig, S., Sharifi, A.R., Wentrot, H., Landmann, D., Eise, M. & Simianer, H. (2005). Genetic parameters of claw and foot disorders estimated with logistic models. *J Dairy Sci*, 88(9), pp. 3316-25.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D.,

Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S. & Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. *Nature,* 409(6822), pp. 860-921.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics,* 23(21), pp. 2947-8.

Lowe, T.M. & Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res,* 25(5), pp. 955-64.

Lysnyansky, I., Sachse, K., Rosenbusch, R., Levisohn, S. & Yogev, D. (1999). The vsp locus of Mycoplasma bovis: gene organization and structural features. *J Bacteriol,* 181(18), pp. 5734-41.

Mappley, L.J., Black, M.L., AbuOun, M., Darby, A.C., Woodward, M.J., Parkhill, J., Turner, A.K., Bellgard, M.I., La, T., Phillips, N.D., La Ragione, R.M. & Hampson, D.J. (2012). Comparative genomics of *Brachyspira pilosicoli* strains: genome rearrangements, reductions and correlation of genetic compliment with phenotypic diversity. *BMC Genomics,* 13, p. 454.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437(7057), pp. 376-80.

Margulis, L., Ashen, J.B., Sole, M. & Guerrero, R. (1993). Composite, large spirochetes from microbial mats: spirochete structure review. *Proceedings of the National Academy of Sciences,* 90(15), pp. 6966-6970.

Matson, E.G., Thompson, M.G., Humphrey, S.B., Zuerner, R.L. & Stanton, T.B. (2005). Identification of genes of VSH-1, a prophage-like gene transfer

agent of *Brachyspira hyodysenteriae*. *Journal of Bacteriology,* 187(17), pp. 5885-5892.
Matson, E.G., Zuerner, R.L. & Stanton, T.B. (2007). Induction and transcription of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *Anaerobe,* 13(2), pp. 89-97.
Maxam, A.M. & Gilbert, W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci U S A,* 74(2), pp. 560-4.
McLaren, A.J., Trott, D.J., Swayne, D.E., Oxberry, S.L. & Hampson, D.J. (1997). Genetic and phenotypic characterization of intestinal spirochetes colonizing chickens and allocation of known pathogenic isolates to three distinct genetic groups. *J Clin Microbiol,* 35(2), pp. 412-417.
Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O. & Giegerich, R. (2003). GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic acids research,* 31(8), pp. 2187-2195.
Motro, Y., La, T., Bellgard, M.I., Dunn, D.S., Phillips, N.D. & Hampson, D.J. (2009). Identification of genes associated with prophage-like gene transfer agents in the pathogenic intestinal spirochaetes *Brachyspira hyodysenteriae*, *Brachyspira pilosicoli* and *Brachyspira intermedia*. *Veterinary microbiology,* 134(3), pp. 340-345.
Nordhoff, M., Moter, A., Schrank, K. & Wieler, L.H. (2008). High prevalence of treponemes in bovine digital dermatitis-A molecular epidemiology. *Veterinary microbiology,* 131(3), pp. 293-300.
Norris SJ, Paster BJ & RM, S. (2011). Genus IV. *Treponema*. In: Krieg NR, S.J., Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (eds) (ed. *Bergey's Manual of Systematic Bacteriology*4). New York: Springer, pp. 501-531.
Ochiai, S., Adachi, Y. & Mori, K. (1997). Unification of the genera *Serpulina* and *Brachyspira*, and proposals of *Brachyspira hyodysenteriae* comb. nov., *Brachyspira innocens* comb. nov. and *Brachyspira pilosicoli* comb. nov. *Microbiology and immunology,* 41(6), pp. 445-452.
Oxberry, S., Trott, D. & Hampson, D. (1998). *Serpulina pilosicoli*, waterbirds and water: potential sources of infection for humans and other animals. *Epidemiology and infection,* 121(01), pp. 219-225.
Park, J., Friendship, R.M., Poljak, Z., DeLay, J., Slavic, D. & Dewey, C.E. (2013). An investigation of ear necrosis in pigs. *The Canadian Veterinary Journal,* 54(5), p. 491.
Parte, A.C. (2013). LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic acids research,* p. gkt1111.
Paster, B., Dewhirst, F., Weisburg, W., Tordoff, L., Fraser, G., Hespell, R., Stanton, T., Zablen, L., Mandelco, L. & Woese, C. (1991). Phylogenetic analysis of the spirochetes. *Journal of Bacteriology,* 173(19), pp. 6101-6109.
Paster, B.J. & Dewhirst, F.E. (2000). Phylogenetic foundation of spirochetes. *J Mol Microbiol Biotechnol,* 2(4), pp. 341-4.

Persson, A., Jacobsson, K., Frykberg, L., Johansson, K.E. & Poumarat, F. (2002). Variable surface protein Vmm of *Mycoplasma mycoides* subsp. *mycoides* small colony type. *J Bacteriol,* 184(13), pp. 3712-22.

Petersen, H.H., Nielsen, E., Hassing, A., Ersbøll, A.K. & Nielsen, J.P. (2008). Prevalence of clinical signs of disease in Danish finisher pigs. *The Veterinary Record,* 162(12), pp. 377-382.

Picardeau, M., Bulach, D.M., Bouchier, C., Zuerner, R.L., Zidane, N., Wilson, P.J., Creno, S., Kuczek, E.S., Bommezzadri, S. & Davis, J.C. (2008). Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PLoS One,* 3(2), p. e1607.

Pringle, M., Backhans, A., Otman, F., Sjölund, M. & Fellström, C. (2009). Isolation of spirochetes of genus *Treponema* from pigs with ear necrosis. *Veterinary microbiology,* 139(3), pp. 279-283.

Pringle, M., Bergsten, C., Fernström, L.-L., Höök, H. & Johansson, K.-E. (2008). Isolation and characterization of *Treponema phagedenis*-like spirochetes from digital dermatitis lesions in Swedish dairy cattle. *Acta Veterinaria Scandinavica,* 50(1), p. 40.

Pringle, M. & Fellström, C. (2010.). *Treponema pedis* isolated from a sow shoulder ulcer. *Vet. Microbiol,* 142, pp. 461-463.

Rasback, T., Jansson, D.S., Johansson, K.E. & Fellstrom, C. (2007). A novel enteropathogenic, strongly haemolytic spirochaete isolated from pig and mallard, provisionally designated '*Brachyspira suanatina*' sp. nov. *Environ Microbiol,* 9(4), pp. 983-91.

Read, D.H., Walker, R.L., Castro, A.E., Sundberg, J.P. & Thurmond, M.C. (1992). An invasive spirochaete associated with interdigital papillomatosis of dairy cattle. *Vet Rec,* 130(3), pp. 59-60.

Richardson, J., Morter, R., Rebar, A. & Olander, H. (1984). Lesions of porcine necrotic ear syndrome. *Veterinary Pathology Online,* 21(2), pp. 152-157.

Richter, M. & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences,* 106(45), pp. 19126-19131.

Rosander, A., Guss, B., Frykberg, L., Bjorkman, C., Naslund, K. & Pringle, M. (2011). Identification of immunogenic proteins in *Treponema phagedenis*-like strain V1 from digital dermatitis lesions by phage display. *Vet Microbiol,* 153(3-4), pp. 315-22.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution,* 4(4), pp. 406-425.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74(12), pp. 5463-7.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. *Nat Methods,* 5(1), pp. 16-8.

Seshadri, R., Myers, G.S., Tettelin, H., Eisen, J.A., Heidelberg, J.F., Dodson, R.J., Davidsen, T.M., DeBoy, R.T., Fouts, D.E., Haft, D.H., Selengut, J., Ren,

Q., Brinkac, L.M., Madupu, R., Kolonay, J., Durkin, S.A., Daugherty, S.C., Shetty, J., Shvartsbeyn, A., Gebregeorgis, E., Geer, K., Tsegaye, G., Malek, J., Ayodeji, B., Shatsman, S., McLeod, M.P., Smajs, D., Howell, J.K., Pal, S., Amin, A., Vashisth, P., McNeill, T.Z., Xiang, Q., Sodergren, E., Baca, E., Weinstock, G.M., Norris, S.J., Fraser, C.M. & Paulsen, I.T. (2004). Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc Natl Acad Sci U S A,* 101(15), pp. 5646-51.

Setubal, J.C., Reis, M., Matsunaga, J. & Haake, D.A. (2006). Lipoprotein computational prediction in spirochaetal genomes. *Microbiology,* 152(Pt 1), pp. 113-21.

Stanton, T., Fournie-Amazouz, E., Postic, D., Trott, D., Grimont, P., Baranton, G., Hampson, D. & Saint Girons, I. (1997). Recognition of Two New Species of Intestinal Spirochetes: *Serpulina intermedia* sp. nov. and *Serpulina murdochii* sp. nov. *Int J Syst Bacteriol,* 47(4), pp. 1007-1012.

Stanton, T., Jensen, N., Casey, T., Tordoff, L., Dewhirst, F. & Paster, B. (1991). Reclassification of *Treponema hyodysenteriae* and *Treponema innocens* in a New Genus, *Serpula* gen. nov., as *Serpula hyodysenteriae* comb. nov. and *Serpula innocens* comb. nov. *Int J Syst Bacteriol,* 41(1), pp. 50-58.

Stanton, T., Postic, D. & Jensen, N. (1998). *Serpulina alvinipulli* sp. nov., a new *Serpulina* species that is enteropathogenic for chickens. *Int J Syst Bacteriol,* 48(3), pp. 669-676.

Stanton, T.B. (1992). Proposal To Change the Genus Designation *Serpula* to *Serpulina* gen. nov. Containing the Species *Serpulina hyodysenteriae* comb. nov. and *Serpulina innocens* comb. nov. *Int J Syst Bacteriol,* 42(1), pp. 189-190.

Svartstrom, O., Karlsson, F., Fellstrom, C. & Pringle, M. (2013). Characterization of *Treponema spp*. isolates from pigs with ear necrosis and shoulder ulcers. *Vet Microbiol,* 166(3-4), pp. 617-23.

Taylor, D.J. & Alexander, T.J. (1971). The production of dysentery in swine by feeding cultures containing a spirochaete. *Br Vet J,* 127(11), pp. 58-61.

Trott, D.J., Moeller, M.R., Zuerner, R.L., Goff, J.P., Waters, W.R., Alt, D.P., Walker, R.L. & Wannemuehler, M.J. (2003). Characterization of *Treponema phagedenis*-like spirochetes isolated from papillomatous digital dermatitis lesions in dairy cattle. *J Clin Microbiol,* 41(6), pp. 2522-2529.

Trott, D.J., Stanton, T.B., Jensen, N.S., Duhamel, G.E., Johnson, J.L. & Hampson, D.J. (1996). *Serpulina pilosicoli* sp. nov., the agent of porcine intestinal spirochetosis. *Int J Syst Bacteriol,* 46(1), pp. 206-15.

Vallenet, D., Engelen, S., Mornico, D., Cruveiller, S., Fleury, L., Lajus, A., Rouy, Z., Roche, D., Salvignol, G., Scarpelli, C. & Medigue, C. (2009). MicroScope: a platform for microbial genome annotation and comparative genomics. *Database-the Journal of Biological Databases and Curation*.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., Sidow, A., Fire, A. & Johnson,

S.M. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res,* 18(7), pp. 1051-63.
van Dijk, E.L., Auger, H., Jaszczyszyn, Y. & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet,* 30(9), pp. 418-26.
Walker, R.L., Read, D.H., Loretz, K.J. & Nordhausen, R.W. (1995). Spirochetes isolated from dairy cattle with papillomatous digital dermatitis and interdigital dermatitis. *Vet Microbiol,* 47(3-4), pp. 343-55.
Wanchanthuek, P., Bellgard, M.I., La, T., Ryan, K., Moolhuijzen, P., Chapman, B., Black, M., Schibeci, D., Hunter, A., Barrero, R., Phillips, N.D. & Hampson, D.J. (2010). The complete genome sequence of the pathogenic intestinal spirochete *Brachyspira pilosicoli* and comparison with other *Brachyspira* genomes. *PLoS One,* 5(7), p. e11455.
Wilson-Welder, J.H., Elliott, M.K., Zuerner, R.L., Bayles, D.O., Alt, D.P. & Stanton, T.B. (2013). Biochemical and molecular characterization of *Treponema phagedenis*-like spirochetes isolated from a bovine digital dermatitis lesion. *BMC Microbiol,* 13, p. 280.
Wixon, J. (2001). Reductive evolution in bacteria: *Buchnera* sp., *Rickettsia prowazekii* and *Mycobacterium leprae*. *Comparative and Functional Genomics,* 2(1), pp. 44-48.
Woese, C.R. (1987). Bacterial evolution. *Microbiological reviews,* 51(2), p. 221.
Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences,* 74(11), pp. 5088-5090.
Yano, T., Yamagami, R., Misumi, K., Kubota, C., Moe, K.K., Hayashi, T., Yoshitani, K., Ohtake, O. & Misawa, N. (2009). Genetic heterogeneity among strains of *Treponema phagedenis*-like spirochetes isolated from dairy cattle with papillomatous digital dermatitis in Japan. *J Clin Microbiol,* 47(3), pp. 727-33.
Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J. & Wishart, D.S. (2011). PHAST: a fast phage search tool. *Nucleic acids research*, p. gkr485.
Zurbrigg, K. (2006). Sow shoulder lesions: Risk factors and treatment effects on an Ontario farm. *Journal of animal science,* 84(9), pp. 2509-2514.

Acknowledgements

The studies in this thesis were carried out at the Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala funded by Higher Education Commission of Pakistan. I am grateful to acknowledge all of you at HGEN who have helped and supported me during my time as a PhD student. There are some people, whom I would like to say special thanks.

First of all I would like to thank my supervisors:

Erik Bongcam-Rudloff, my main supervisor, Thank you for believing in me and for all the support and encouragement that I have received from you during these years. Thank you for your sincere love that you always show towards all of your students, treating them as your own children and for creating an independent, friendly and relaxed work environment.

Anna Rosander and Desiree Jansson, my present co-supervisors. I would like to say many many thanks to you both, without your constant moral and scientific support especially in the last few months, it would never have been possible. Anna's humble and polite nature with sound scientific knowledge and Desiree's command over so many things at a time always had been a motivation to ask them for help anytime.

Märit Pringle and Richard Zuerner, my former co-supervisors. I would like to express my deepest appreciation to both of you, Thanks, Märit for giving an energetic start to my PhD, introducing me to lab and research environment and always trying to create learning opportunities for me and Rich for always being there to answer questions even after going back to USA.

I am grateful to **Higher Education Commission of Pakistan** for providing me funding to accomplish my PhD studies and **UPPMAX** for providing computing resources.

I would like to say thanks to all my fellow PhD students, who helped me make this long journey interesting and easy with their constant care and support in all the matters.
Olov, Thanks for all your help during all these years, you always had been there whenever I need your help.
Frida, its always nice talking to you, thanks for providing me pictures to use in my thesis.

All the Pakistani friends and their families, thanks for the time that we shared, all the weekend parties with delicious food and fun hours of chit chat made you all like a family to me.

I would like to express my special gratitude to some people who took a very good care of my son Arsalan, **Saima** and **Imran bhai** thanks for being there for us always, **Arooj Asad** for taking care of Arsalan when I was working and **Awaisa** for picking him from School and taking good care of him when it was really hard for me to manage.

I would like to say thanks to my **family** and my **in-laws**, there is indeed no substitute to love and care I have always got from you. Thanks for the patience that you all showed waiting to meet us.

To my parents, I always have been deficient of words to express my feelings of gratitude for your immense love, support, prayers and encouragement that I have received throughout my life. There is indeed no pay back for all the hardships and difficulties that you have gone through in your life just for our future.
My parents in law, for your love and always making me feel like your own daughter. I would especially like to express my deep regard for my father in law (late) for being so caring towards me.

Last, but not the least my husband **Shahzad** and my son **Arsalan,** you two have indeed suffered the most from my studies, Thank you both for everything.