Molecular characterization of the penile microbiome of Dorper rams (*Ovis aries*)

by

Auberi Marie Madeleine Courchay

Thesis presented in fulfilment of the requirements for the degree of Masters in the Faculty of Science at Stellenbosch University

Supervisor: Dr H. Lambrecths

Co-supervisor: Prof J.T. Burger

March 2017

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2017

ABSTRACT

Characterizing the bacterial community inhabiting the penile environment of Dorper rams may provide insight into the aetiology of ulcerative balanoposthitis (UB), a venereal disease of unknown aetiology, which occurs predominantly in this breed, and that is of great economic importance to the mutton industry in South Africa. Profiling of the bacteria present in the penile environment of healthy and diseased rams was carried out previously using culture-based methods. These culture-based studies identified *Mycoplasma mycoides mycoides* large colony (MMMLC) and *Trueperella pyogenes* as potentially contributing to the incidence of UB. Although cultured-based methods have been described as the gold standard for identification of bacteria, many bacteria are uncultivable or grow slowly and poorly *in-vitro*. The advent of high throughput next generation sequencing (HT-NGS), a culture-independent bacterial identification approach, has offered high coverage and depths in determining the bacterial penile community. Thus, this study aimed to characterize and compare the penile bacterial microbiome of Dorper rams that were healthy and diseased by means of 16S amplicon sequencing, a method of HT-NGS. Swab samples of the preputial and penile mucosa were collected from 113 rams, of which 40 Dorper ram samples (20 infected, 20 healthy) were chosen for further analysis in this study. Genomic DNA was extracted and amplified based on the V3V4 hypervariable region of the 16S rRNA bacterial gene. Bioinformatics analysis was performed using UPARSE and the ecological and statistical analyses such as Principal Coordinate Analysis (PCoA) was performed in QIIME and XLSTAT. Additional analyses comparing the predicted bacterial microbiota in healthy and diseased populations was carried out using LEfSe. A total of 789 OTUs from 9 964 842 sequences of high quality were obtained from the healthy and diseased communities indicating a high bacterial diversity in the penile environment, higher than previously reported and isolated using culture-based bacterial identification methods. The genus *Corynebacterium* was the most dominant genus identified (20.9%), irrespective of health status. A high inter-sample variation in microbiota was revealed. There were no significant differences in bacterial diversity or community composition between the healthy and diseased ram groups. The microbiota population was thus similar, with a few OTUs of high biological relevance belonging to genera *Fusobacterium* and *Porphyromonas*, as well as uncharacterized genera within *Aerococcaceae* and *Bacteroidale*s that were enriched in the diseased community. MMMLC and *Trueperella pyogenes* were not associated with UB in this study, which is contradictory to previous reports. However, a new *Mycoplasma* species, *Mycoplasma hyopharyngis*, not previously isolated in sheep and in cases of UB in Dorper rams, was identified and although not significantly different, had a higher abundance in the diseased population. The prevalence and predominance of *Corynebacterium* across all samples suggests this genus forms part of the core microbiome of the penile environment. High inter-sample variation in microbiota may depict true biological representation, however, future studies using homogenous ram populations (i.e. same type, same age) under the same environmental factors (i.e. management, feeding regime, etc.) should be conducted to validate these findings. Ulcerative balanoposthitis is not caused as a result of a change in bacterial diversity or community composition but OTUs enriched in the diseased ram population may be disease-specific/disease-associated and their role in UB warrants further investigation. No definitive aetiological agent was thus identified but the OTUs enriched in diseased rams can help direct future studies towards the identification of an aetiological agent of UB.

OPSOMMING

Die karakterisering van die bakteriële gemeenskap wat die penis skede omgewing van Dorper ramme kan insig oor die etiologie van ulseratiewe balanoposthitis (UB), 'n geslagsiekte van onbekende etiologie wat hoofsaaklik in dié ras voorkom, verskaf. Die siekte is van groot ekonomiese belang vir die skaapvleis bedryf in Suid-Afrika. Vorige studies oor die karakterisering van die bakterieë in die penis skede omgewing van gesonde en siek ramme was gebaseer op kultuur-gebaseerde metodes. Hierdie kultuur-gebaseerde studies het *Mycoplasma mycoides mycoides Large Colony* (MMMLC) en *Trueperella pyogenes* as organismes geïdentifiseer wat moontlik bydra tot die voorkoms van UB. Hoewel kultuur-gebaseerde metodes beskryf word as die goue standaard vir die identifisering van bakterieë, kan baie bakterieë nie *in vitro* gekweek word nie of swak groei word *in vitro* waargeneem. Die bekendstelling van 'n hoë deurset volgende generasie volgorde bepalingstegniek (HT-NGS), 'n kultuur-onafhanklike benadering, bied 'n wye dekking asook besonderse diepte waarmee die bakteriële gemeenskap in die penis skede omgewing bepaal kan word. Dié studie was daarop gemik om die bakteriële mikrobioom van die skede omgewing van siek en gesonde Dorper ramme met behulp van 16S amplikon volgordebepaling, 'n HT-NGS metode, te karakteriseer en te vergelyk. Depper monsters van die voorvel en penis mukosa omgewing is van 113 ramme versamel, waarvan 40 Dorper ram monsters (20 besmet, 20 gesonde) vir verdere analise in hierdie studie gekies is. Die genomiese DNS is geëkstraheer en versterk gebaseer op die V3V4 hiper-variërende streek van die 16S rRNS bakteriële geen. Bio-informatiese analise is met behulp van UPARSE uitgevoer en die ekologiese en statistiese ontledings soos Hoof Koördinaat Ontleding (PCoA) is uitgevoer in QIIME en XLSTAT. Bykomende ontledings wat die voorspelde bakteriële mikrobiota in gesonde en siek bevolkings vergelyk het, is uitgevoer met behulp van LEfSe. 'n Totaal van 789 operasionele taksonomiese eenhede (OTE'e) van 9 964 842 sekwense van 'n hoë gehalte is verkry uit die gesonde en siek gemeenskappe, wat dui op 'n hoë bakteriële diversiteit in die penis omgewing, hoër as wat voorheen berig en geïsoleer was deur kultuur-gebaseerde bakteriële identifikasie metodes. Die genus *Corynebacterium* was die mees dominante genus geïdentifiseer (20.9%), ongeag die gesondheidstatus van 'n ram. 'n Hoë inter-monster variasie in mikrobiota is gevind. Daar was geen betekenisvolle verskille in bakteriële diversiteit of gemeenskapsamestelling tussen die gesonde en siek ram groepe nie. Die mikrobiota bevolking was soortgelyk tussen siek en gesonde ramme, met 'n paar OTE'e van 'n hoë biologiese relevansie (d.i. genera *Fusobacterium* en *Porphyromonas*, asook nie-gekarakteriseerde genera binne *Aerococaceae* en *Bacteroidales*) wat meer volop in die siek was. Die bakterieë MMMLC en *Trueperella pyogenes* kon nie in dié studie in verband gebring word met die voorkoms van UB nie, wat teenstrydig is met die vorige verslae. 'n Nuwe mikoplasma spesie, *Mycoplasma hyopharyngis*, wat nog nie voorheen in skape en gevalle van UB in Dorper ramme geïsoleer is nie, is geïdentifiseer en alhoewel nie betekenisvol nie, het in hoër getalle in die siek ramme voorgekom. Die voorkoms en oorheersing van *Corynebacterium* in beide die siek en gesonde ramme dui aan dat dié genus deel vorm van die kern mikrobioom van die penis omgewing. Hoë inter-monster variasie in mikrobiota kan normaal wees, maar daar word aanbeveel dat toekomstige studies waar moontlik homogene ram bevolkings (d.w.s. dieselfde ras, ouderdom, ens.) onder dieselfde omgewingsfaktore (bv. bestuur, voeding, ens.) gedoen moet word om hierdie bevindinge te bevestig. Ulseratiewe balanoposthitis word dus nie veroorsaak deur 'n verandering in bakteriële diversiteit of gemeenskap samestelling nie, maar die OTE's wat meer volop in die siek ramme voorgekom het mag siekte-spesifiek wees of verband hou met UB. Derhalwe moet hierdie organismes se rol in die voorkoms van UB verder ondersoek word. Geen definitiewe etiologiese agent is dus

geïdentifiseer nie, maar die OTE's geïdentifiseer in hierdie studie kan leiding verskaf oor die ontwerp van toekomstige studies vir die identifisering van 'n etiologiese agent van UB.

DEDICATION

To my parents who have given me endless support throughout my infinite journey as a student. For giving me all the opportunities in the world and always making sure I had everything I have ever needed. This masters is for you.

ACKNOWLEDGMENTS

To Dr Helet Lambrechts, my supervisor, thank you for giving me the opportunity to take on this challenging and interesting project. I would have never believed anyone that said I could accomplish a project of this depth on my own. The newly found patience and skills I have acquired will surely take me places I never imagined.

Special thanks to my co-supervisor Prof Johan Burger for understanding my frustrations and always being honest with me. Thank you for helping me deal with problems in the best possible way and for always having an open door and advice.

To the Vitis lab for allocating me space in your laboratory so that I could undertake my project, for your constant support, listening to my rants and for making me feel part of the Genetics department.

I am deeply thankful to Tracy Meiring and Harris Onywera, whom without this project would have literally been impossible, for giving me the necessary tools to start my bioinformatics journey and consistent guidance and advice throughout that led me to completion. For giving me your time although you had no responsibility to, I owe this project to you both.

Thank you to Angela Verde and Prof Cowan for being accommodating and allowing me to visit your department at the University of Pretoria and pointing me in the right direction.

Thank you to my boyfriend and my friends for your support in all ways and for getting me through this master's thesis. I am sure you will be happy never to hear the word thesis again.

Most importantly thank you to the Northern Cape Department of Land Reform and Rural Development and the National Red Meat Producers Organization for the financial support of the study.

Acknowledgement also goes out to the respective farms involved for making their animals available for collection of samples.

LIST OF FIGURES

Figure 4.18. Unifrac diversity measures. Principle coordinate analysis, using weighted Unifrac distances to explore and visualize diversity between samples. Each point represents a sample plotted by a principal component on the X- axis and another principal component on the Y- axis. Square (red) data points represent samples from healthy rams and triangular (orange) from diseased rams. PC1 is principal component 1, PC2 is principle component 2 and PC3 is principle component 3. **A)** PC1 versus PC2, **B)** PC1 versus PC3, **C)** PC3 versus PC2.

Figure 4.19. Differentially abundant taxa (classified from the phylum to genus) as revealed by LEfSe. **A)** Histogram showing taxa most abundant in the healthy group. Taxa are ranked according to linear discriminant analysis and only taxa with p-value <0.05 and LDA scores ≥2 were included in the results. **B)** Cladogram displaying the phylotypes that were most abundant in the healthy ram population with the size of the circles increasing with increasing relative abundance.

Figure 4.20. Histogram depicting the OTUs identified as significantly differential in healthy and diseased groups. Only taxa with p-value <0.05 and LDA scores ≥2 were included in the results. Taxa more abundant in the healthy group are indicated with positive LDA scores (green), and taxa more abundant in the diseased group are indicated with negative LDA scores (red).

Figure 4.21. Distribution of the relative abundance of phylotypes flagged by LEfSe as significantly more abundant. Histograms showing the relative abundance of **A)** the genus *Arcanobacterium*. **B)** the genus *Pasteurella* **C)** the family Streptococacceae and **D)** the family Leptotrichiaceae The mean and median relative abundance of each phylotype is indicated with a solid and dashed line, respectively. Each bar on the x-axis represents a sample.

Figure 4.22. Heatmap of the square root-transformed relative abundances of OTUs flagged by LEfse as significantly abundant in healthy and diseased groups. The yellow colour bars represent the diseased samples and the green bars the healthy samples. The colour scale above the heatmap depicts the range in relative abundance, the light colour indicates a low proportion and red shows a high proportion. OTU identified as significantly more abundant in either the healthy or diseased groups are marked by a (H) or (D), respectively.

Figure 4.23. Bar graph representing the range in percentage relative abundance of the OTU identified as *Trueperella pyogenes* by BLASTn analysis, across healthy and diseased samples. Each label on the horizontal access is an individual sample.

Figure 4.24. Broken axis bar graph displaying the distribution of percentage relative abundance of three OTUs identified as *Mycoplasma hyopharyngis, Mycoplasma arginini and Mycoplasma bovigenitalium* across healthy and diseased samples. The relative abundance on the y-axis stops at 1.2% and restarts at 9% due to the large inter-sample variation. Each label on the horizontal access is an individual sample.

Figure 4.25. The percentage contribution of each sample to the total relative abundance of *Truperella pyogenes* observed in the healthy and diseased groups, separately. Each label on the horizontal access is an individual sample.

Figure 4.26. The percentage contribution of each sample to the total relative abundance of *Mycoplasma hyopharyngis, Mycoplasma arginini and Mycoplasma bovigenitalium* observed in the healthy and diseased groups, separately. Each label on the horizontal access is an individual sample.

SUPPLEMENTARY FIGURES

Figure S4.1. Unifrac diversity measures. Principle coordinate analysis, using unweighted Unifrac distances to explore and visualize group similarities between healthy and diseased samples. Each point represents a sample, plotted by a principal component on the X- axis and another principal component on the Y- axis. Square (red) data points represent samples from healthy rams and triangular (orange) from diseased rams. PC1 is principal component 1, PC2 is principle component 2 and PC3 is principle component 3. **A)** PC1 versus PC2, **B)** PC1 versus PC3, **C)** PC3 versus PC2.

LIST OF TABLES

SUPPLEMENTARY TABLES

Table S3.1. The total healthy and diseased ram population sampled in the study grouped according to age.

LIST OF ABBREVIATIONS

UB	ulcerative balanoposthitis
UV	ulcerative vulvovaginitis
MMMLC	*Mycoplasma mycoides mycoides* large colony
DNA	deoxyribonucleic acid and ribonucleic acid
RNA	ribonucleic acid
rRNA	ribosomal RNA
PCR	polymerase chain reaction
et al	and others
SDS	sodium lauryl sulphate
dNTP	deoxynucleotide triphosphate
GB	gigabases
bp	base pairs
OTU	operational taxonomic units
SD	standard deviation
HT-NGS	high throughput next generation sequencing
CO_2	carbon dioxide
WC	Western Cape
NC	Northern Cape
USA	United State of America
UK	United Kingdom
^oC	degrees Celsius
&	and
%	percentage
<	less than
>	greater than
~	approximately

≥	more than or equal to
≤	less than or equal to
e.g.	for example
mg	milligram
ng	nanogram
mm	millimetre
mM	millimolar
µM	micromolar
mL	millilitre
µL	microlitre

.

TABLE OF CONTENTS

Chapter 1

General introduction

Large areas of South Africa are characterized as semi-arid, with little potential to be used for the production of commercial crops (Schoeman *et al.,* 2010). Consequentially, livestock farming is considered the most important agricultural practice in the country, using as much as 70% of the agricultural land available (Goldblatt, 2010; Schoeman *et al.,* 2010; Spies, 2011). South Africa's population is growing at almost 2% per year (Goldblatt, 2010), and this places increasing pressure on livestock farmers to optimise production in order to supply in the national mutton demand. Cost-efficient and improved livestock production is however hampered by factors such as water scarcity within the country and a decline in farming profitability due to impacts such as higher input costs, with this effect evident in a decrease in the number of farms by two thirds since the early 1990s (Goldblatt, 2010; DAFF, 2014).

Of the different livestock species, sheep is a preferred species in the more arid parts of South Africa due to the fact that they can utilize a wide variety of vegetation types, and are hardy to arid conditions (Conradie & Landman, 2015). As a result, sheep comprises the largest livestock population in South Africa, with 20 breeds that currently contribute to the national economy through the production of animal fibre and mutton/lamb (DAFF, 2014). Fluctuations in wool and mutton relative prices (driven by supply and demand) over the last 50 years have resulted in a continuous fluctuation in the composition of sheep flocks in South Africa, driving a shift in farmer preferences towards mutton producing sheep (Conradie & Landman, 2015). Cost-efficient mutton production highlighted the need for a sheep breed with a high reproductive rate, and that could produce a high quality carcass with an even fat distribution on sparse vegetation that occurs in the more arid regions of South Africa. These requirements led to the development of the Dorper breed in 1942 by crossbreeding Dorset Horn rams and Black-head Persian ewes (Dorper Sheep Breeders' Society of South Africa, 2016). The Dorper breed exceeded these expectations and is now considered the top mutton breed in the country, representing about 24% of the national sheep population.

Despite South Africa's top quality mutton population, it is still considered a net importer of mutton with declining sheep numbers due to diseases, one of the many factors contributing to this decline. According to Modisane (2009), an increase in the number and intensity of infectious diseases in South Africa driven by poor knowledge of the aetiology and epidemiology of such diseases and/or poor management plans to prevent, control or treat known diseases, has negatively affected livestock production. Sheep are susceptible to many infectious diseases, and reproductive diseases in particular have a direct effect on losses pertaining to sheep numbers by negatively affecting fertility in sheep flocks. Amongst the important reproductive diseases prevalent in sheep, ulcerative balanoposthitis (UB) is a venereal disease affecting up to 50% of affected flocks causing increasing concern to the South African mutton industry (Gummow & Staley, 2000). Gummow and Staley (2000) suggest the disease

as primarily affecting Dorper populations in South Africa with fewer cases observed in other sheep and goat breeds.

Ulcerative balanoposthitis was first reported in Dorper flocks in 1976 in the Calvinia district of the Northern Cape, and has since spread to the Free State, Kwazulu Natal, and Eastern and Western Cape provinces following the distribution of Dorper flocks, with the highest incidence of the disease observed in the Free State Dorper populations (Trichard *et al.,* 1993; Gummow & Staley, 2000). The disease affects both rams and ewes (ulcerative vulvovaginitis), with UB more readily observed in males. In rams UB is characterized by erosions and ulceration of the glans penis, which has been observed to vary in appearance due to severity, age of lesions and secondary bacterial infection complicating diagnosis of the disease.

Dorper flocks affected by UB have experienced depressed lambing percentages as much as a 50% decline due to the refusal of rams and/or ewes to mate because of the inflammation and resulting discomfort and/or pain associated with UB infection (Bath & De Wet, 2000). The incidence of UB also prevent Dorper producers to participate in breeder and stud auctions, which in turn restrict the national trade of sheep stock. These consequences of UB along with forced culling of diseased sheep, as a means to control the disease, makes UB a disease of economic importance in SA.

Almost 40 years after the discovery of the disease, the causative agent of UB has not yet been conclusively identified and with few studies related to the disease in SA, researchers have attributed the aetiology of the disease to different infectious organisms. Initially, studies of UB in South Africa were through conventional microbiological and molecular methods using morphological appearance; biochemical, serological (Trichard *et al.,* 1993; Kidanemariam, 2003; Ali, 2012) and DNA based-tests (Ali, 2012) to identify and classify cultured bacterial species recovered from affected and unaffected individuals. Their combined findings revealed a number of mollicutes, mostly *Mycoplasma* and unidentified Ureaplasma, isolated from the mucosal lining of the prepuce and penis of sheep. The more commonly isolated *Mycoplasma* species amongst studies included *Mycoplasma mycoides subspecies mycoides* large colony *(MMMLC), Mycoplasma arginini, Mycoplasma bovigenitalium, Acholeaplasma laidlawii, Mycoplasma agalactiae, Mycoplasma mycoides capri, Mycoplasma capricolum*, and *Mycoplasma* species Group 7. Trichard *et al.* (1993) and Kidanemariam (2003) also observed a number of other bacteria concurrently isolated, some with established pathogenic properties (e.g. *Trueperella pyogenes, Corynebacterium renale, Eschericia coli, Staphyloccocus aurens*, *Staphylococcus epidermidis, Enterococcus faecalis* and *Histophilus ovis*). In summary, the respective studies on the incidence and cause of UB in SA have illustrated inconsistent isolation of different bacterial species from clinically infected specimens and have suggested that the aetiology of the disease may be polymicrobial. No viruses or fungi have yet been isolated from cases of UB in South Africa.

While studies of UB using traditional microbiological methods have provided insight into some of the bacterial species surrounding the penis of healthy and diseased Dorper rams, the results from Ali (2012)

that suggests molecular techniques as more specific in identifying *Mycoplasma*, and the fact that not all bacteria are cultivable (Wade, 2002), emphasizes the need for newer molecular methods to characterize and compare these bacterial communities. In addition, easily cultivable bacterial species have the ability to overcrowd slower growing bacteria that may be of clinical importance under *in-vitro* conditions, leading to misinterpretations of the true prevalence of bacterial species present in clinical samples (Kong, 2011). Many studies exist in the literature that compare culture-dependent and -independent methods, and illustrate the superiority of molecular phylogenetic techniques in identifying bacteria not previously observed (Sakamoto *et al.,* 2002; Dekio *et al.,* 2005) or cultivated with conventional-culture dependent techniques (Petti *et al.,* 2005; Siqueira *et al.,* 2005; Rogers *et al.,* 2009; Pallen *et al.,* 2010).

Rapid advances in sequencing technologies have led to the development of 16S amplicon sequencing a method of high throughput next generation sequencing (HT-NGS) (Behjati & Tarpey, 2013), a molecular technique like Sanger sequencing that uses conserved regions of the 16S rRNA bacterial gene found across most groups of bacterial species, to amplify and sequence bacterial DNA directly from samples, producing a large magnitude of small DNA fragments from many samples in a short period of time (Kircher & Kelso, 2010; Buermans & Den Dunnen, 2014). 16S amplicon sequencing has permitted researchers to identify almost complete microbiomes. This was previously thought impossible through the use of conventional methods due to the skill, time and various media, reagents and environmental conditions required to grow and identify bacteria; making it impossible to culture all species present in clinical samples, especially one of high diversity (Petti *et al.,* 2005). Molecular profiling of bacterial communities within samples using the universal 16S rRNA gene, not only is able to recover almost all species present but can also provide reliable estimates of relative abundance of different bacterial species present, moving past simple presence and absence criteria (Pallen *et al.,* 2010). The application of 16S amplicon sequencing has led to an improved understanding of diseases and the importance of profiling bacterial microbiomes in healthy and diseased states in order to understand the underlying biological and clinical role of bacteria in diseases (Lamont *et al.,* 2011; Dickson *et al.,* 2013; Lee *et al.,* 2013; Wade, 2013).

The existing studies on UB in sheep have highlighted the need to identify the causative agent(-s) of UB, and to formulate guidelines that can limit the transmission of the infection, or prevent the occurrence of the infection in commercial Dorper sheep flocks. Studies have been conducted on the microbial population present in the reproductive tract of the ewe (El-Arabi *et al*., 2014; Swartz *et al.,* 2014), but information for rams is scant. In the proposed study, a molecular approach will be followed to conduct a preliminary investigation of the microbial population from the mucosal membrane of the prepuce and the glans penis of healthy and diseased Dorper rams and the comparison thereof. This approach will include targeted amplification of DNA extracted from samples collected from the genital mucosa, and diseased tissue using the 16S rRNA gene, universal across all bacterial organisms, which will then be sequenced using 16S amplicon sequencing technology.

The aims of the study are therefore to:

1. Develop a protocol for the culture-independent study of ulcerative balanoposthitis in Dorper rams to characterize the bacterial population present in the penile and preputial mucosa of Dorper rams.
2. Compare the microbial bacterial diversity and composition between healthy and diseased Dorper rams.
3. Provide baseline information for future studies on the causative agent (-s) of UB in Dorper rams, which will assist with the formulation of management programs to minimise or prevent the incidence and transmission of the disease. This will assist Dorper producers to potentially optimise the production and reproduction efficiency of their production systems.

In this study, we hypothesize that the bacterial diversity and composition of healthy Dorper rams will differ from that of diseased Dorper rams. The alternative hyposthesis suggests that the bacterial diversity and composition of healthy Dorper rams does not differ from that of diseased Dorper rams.

References

Ali, H. 2012. Molecular characterization of *Mycoplasma*s species isolated from the genital tract of Dorper sheep in South Africa. Doctoral dissertation. Pretoria: University of Pretoria.

Bath, G. & De Wet, J. 2000. Pizzle Disease, in Sheep and Goat Diseases. Cape Town: Tafelberg Publishers Ltd. 74-76.

Behjati, S. & Tarpey, P.S. 2013. What is next generation sequencing? *Archives of disease in childhood-Education & Practice Edition,* 98(6):236-238.

Buermans, H.P.J. & Den Dunnen, J.T. 2014. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842(10):1932-1941.

Conradie, B. & Landman, A. 2015. Wool versus mutton in extensive grazing areas. *South African Journal of Agricultural Extension*, 43(1):22-31.

DAFF 2014. *A profile of the South African mutton market value chain.* Department of Agriculture Forestry and Fisheries [Online]. Available:http://www.nda.agric.za/doaDev/sideMenu/Marketing/Annual%20Publications/Commodity%20Profiles/Livestock/Mutton%20Market%20Value%20Chain%20Profile%202014.pdf [2016, August 30].

Dekio, I., Hayashi, H., Sakamoto, M., Kitahara, M., Nishikawa, T., Suematsu, M. & Benno, Y. 2005. Detection of potentially novel bacterial components of the human skin microbiota using culture-independent molecular profiling. *Journal of Medical Microbiology*, 54: 1231–8.

Dickson, R.P., Erb-Downward, J.R. & Huffnagle, G.B. 2013. The role of the bacterial microbiome in lung disease. *Expert Review of Respiratory Medicine*, 7(3):245-257.

Dorper Sheep Breeders' Society of South Africa 2016. *Breed history* [Online]. Available: http://www.dorpersa.co.za/dorper-breed-info/breed-history/. [2016, June 29].

Goldblatt, A. 2010. Agriculture: Facts & Trends, South Africa, in T. von Bormann (ed.), *World Wildlife Forum*.

Gummow, B. & Staley, G.P. 2000. A retrospective survey of ulcerative balanoposthitis and vulvovaginitis in South African Dorper sheep. *Livestock Health and Production,* 12:31-36.

Kidanemariam, A., Gouws, J., Van Vuuren, M. & Gummow, B. 2005. Ulcerative balanitis and vulvitis of Dorper sheep in South Africa: a study on its aetiology and clinical features. *Journal of the South African Veterinary Association,* 76(4):197-203.

Kong, H.H. 2011. Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends in Molecular Medicine*, *17*(6):320-328.

Modisane, B.M. 2009. Field Services: Eradication and Control of animal diseases. *Onderstepoort Journal of Veterinary Research,* 76:115-121

Pallen, M.J., Loman, N.J. & Penn, C.W. 2010. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. *Current Opinion in Microbiology,* 13(5):25-631.

Rogers, G.B., Carroll, M.P. & Bruce, K.D. 2009. Studying bacterial infections through culture-independent approaches. *Journal of Medical Microbiology*, *58*(11):1401-1418.

Sakamoto, M., Huang, Y., Umeda, M., Ishikawa, I. & Benno, Y. 2002. Detection of novel oral phylotypes associated with periodontitis. *FEMS Microbiology Letters,* 217(1):65-69.

Schoeman, S.J., Cloete, S.W.P. & Olivier, J.J. 2010. Returns on investment in sheep and goat breeding in South Africa. *Livestock Science*, 130(1):70-82.

Siqueira, J.F. & Rôças, I.N. 2005. Uncultivated phylotypes and newly named species associated with primary and persistent endodontic infections. *Journal of Clinical Microbiology, 43*(7):3314-3319.

Spies, D.C. 2011. Analysis and quantification of the South African red meat value chain. Doctoral dissertation. Bloemfontein: University of the Free State.

Trichard, C.J.V., Prozeskey, l. Jacobsz, E.P. & Henton, M.M. 1993. The identification of *Mycoplasma* mycoides subspecies mycoides large colony as the aetiological agent of balanoposthitis and vulvovaginitis in sheep in South Africa. *Onderstepoort Journal of Veterinary Research,* 60:29-37.

Wade, W. 2002. Unculturable bacteria—the uncharacterized organisms that cause oral infections. *Journal of the Royal Society of Medicine,* 95(2):81-83.

Wade, W.G. 2013. The oral microbiome in health and disease. *Pharmacological Research*, 69(1):137-143.

Chapter 2

Literature review

2.1 The South African sheep industry

Almost 72 million hectares of land in South Africa is located in semi-arid and arid climatic zones. The arid regions are most suitable for livestock production, and currently 70% of these areas are used exclusively for this purpose, representing it as the largest and most important agricultural practice within the country (Schoeman *et al.,* 2010; Spies, 2011). The ability of sheep to thrive under adverse conditions and the fast return on capital investment have resulted in sheep being the most abundant livestock species in the country, with an estimated population of 24.3 million (DAFF, 2014). The distribution of sheep flocks throughout South Africa and the percentage distribution within the different provinces can be seen in Figures 2.2 and 2.3, respectively.

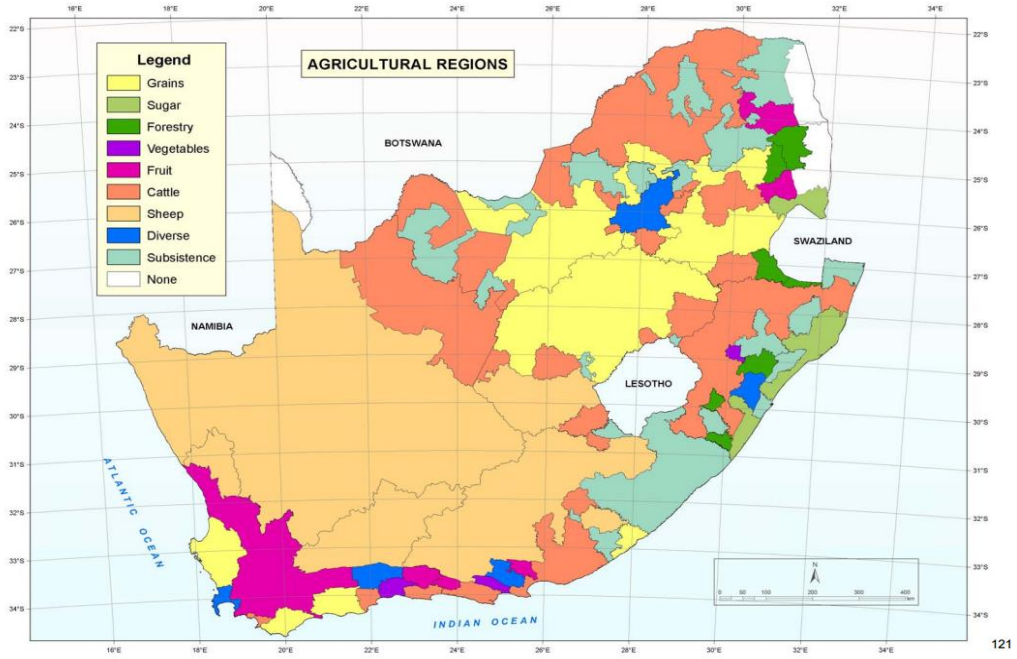

Figure 2.1. Regional distribution of sheep within South Africa (source: Agricultural Geo-Referenced Information System, 2007).

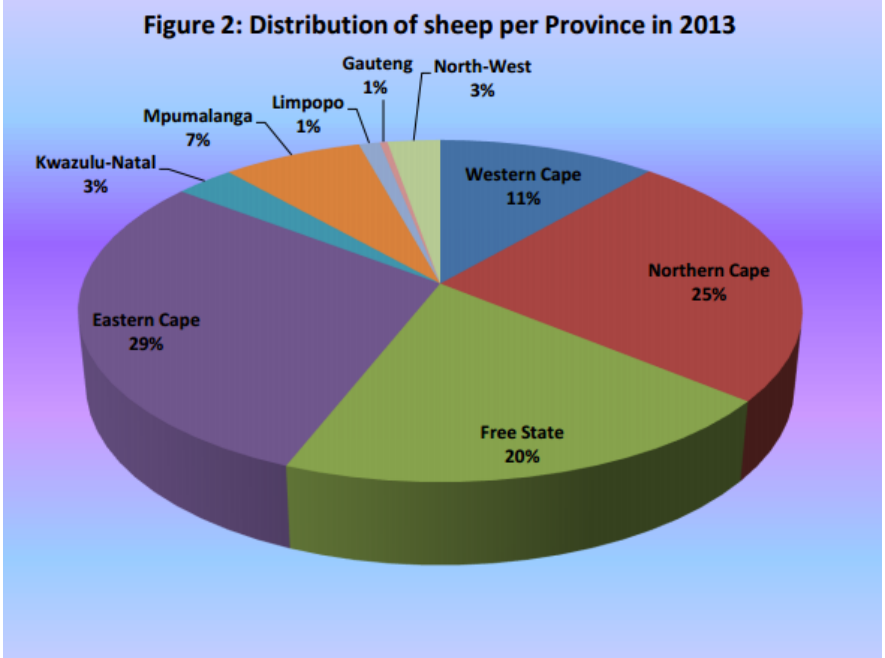

Figure 2.2. The percentage distribution of sheep per province according to 2013 statistics (source: DAFF, 2014).

In South Africa, there are currently twenty breeds that are commercially produced for wool and/or mutton. The mutton industry in South Africa plays an important part in food security, assisting in supplying the growing demand for animal protein. The demand is however usually higher than the supply, and currently South Africa is considered a net importer of mutton (DAFF, 2014). When breeds are considered, the Dorper breed represents the largest non-wool mutton sheep population and is recognized as the top mutton breed in South Africa. This breed however is currently affected with a disease of unknown aetiology termed ulcerative balanoposthitis that affects both ewes and rams, and is predominant in this breed. The incidence of this disease in flocks can have a significant effect on the reproduction efficiency of Dorpers reducing lambing percentages and increasing the culling of diseased animals, ultimately reducing their population numbers. Determining the cause of this disease in order to prevent infection and transmission through proper management programs has become an essential goal for mutton sheep production in order to increase the national flock reproduction efficiency of this breed as a means to supply the mutton demand in the country.

The following sections will give a brief outline on the reproductive system highlighting the components known to be affected by reproductive diseases followed by a more detailed history of the ulcerative balanoposthitis in South Africa.

2.2 The anatomy of the reproductive tract of the ram

The male reproductive system consists of the bilateral testes, the duct system which are found within the penis and the accessory sex organs. The testes are housed in a two-lobed sac, and internally consist of seminiferous tubules that join to form the rete testis, and again join to form the *ductus efferens* (Ross & Pawlina, 2011). Continuing from the *ductus efference, the* duct system consists of a three-part

epididymis (head, cauda and caput sections) and *vas deference*, with the latter that joins to the urethra which runs through the penis and out of the glans penis (Hafez, 1987). The penis is the male copulatory organ and consists of the shaft and the glans penis. When the penis is not erect, it is retracted into an invagination of skin termed the sheath that protects the penis and is made up of an externall and internal portion. The internal portion is termed the prepuce, and is made up of a mucous membrane that extends from the external portion of the sheath to the glans penis. The glans penis is the enlarged free end of the penis (Ross & Pawlina, 2011). The prostate gland, and the bilateral Cowper's glands, the seminal glands and the ampullae make up the accessory sex organs of the male reproductive tract (Hafez, 1987).

Successful reproduction relies on this system to work together to produce, nourish, transport and expel viable sperm as well as produce reproductive hormones which are responsible for the maintenance and development of the reproductive system.

2.3 Factors influencing the reproduction efficiency of rams

The success of any sheep enterprise, given the breed and the resources they are farmed under, is measured by the productive rate of the flock, which in turn is directly related to the reproductive performance. The reproduction performance of a flock is determined by the number of lambs born, weaned and marketed per ewe per year in a given production system. These measured outputs can be used to calculate a number of reproductive traits in sheep that are in turn used to detect reproduction problems in flocks that can be traced back to either the ewes or the rams (Sidwell & Miller, 1971; Salhab *et al.,* 2003).

An understanding of reproduction and environmental effects that affect reproduction is necessary to make sound management decisions that will result in maximum lamb production and ultimately ensure the profitability of a sheep enterprise. A number of factors have been reported to affect the reproductive performance of rams, e.g. nutrition, seasonality, age, mating strategy, frequency of ejaculation, competitive behaviour and diseases (Foote, 1978; Bester, 2006; Marai *et al.,* 2008). The abovementioned factors have been considered to exert an influence on the reproductive efficiency of rams, and this can range from reduced libido and ability to mate to structural and functional defects of the reproductive system which may be temporary or permanent (Foote, 1978). Functional defects of the reproductive system are most often than not related to interference of the natural processes of spermatogenesis, which is responsible for the development and growth of spermatozoa into mature sperm cells (Bearden & Fuquay, 1997). These effects on the reproductive efficiency of sheep can be minimized mainly through efficient reproductive management measures and indirectly more slowly (due to low heritability) through genetic selection of sheep with higher reproductive traits and greater resilience to environmental effects.

In the next section we discuss the effects of diseases, specifically that of reproductive diseases, on the reproductive efficiency of rams.

2.3.1 Reproductive diseases

Reproductive diseases affecting the production efficiency of livestock production systems have been identified across the world in a number of livestock species including bulls (Hancock *et al.,* 2015), pigs (Teankum, 2006), goats (Gouletsou & Fthenakis, 2015) and sheep (Kidanemariam, 2003). Both non-infectious and/or infectious agents can contribute to the incidence of diseases, with a myriad of microorganisms involved in infections, and in the case of the reproductive system, functioning can be impaired or inhibited. In some cases, microbial pathogens contributing to the incidence of a disease may not yet have been established or can be complicated and hard to determine due to multiple organisms contributing to the disease (Yoo, 2010).

Infectious reproductive diseases adversely affect the performance and the welfare of animals, resulting in most cases in significant economic losses in flocks affected. Economic losses occur as a consequence of the lower reproductive output in flocks, high culling rates of infected individuals, bans on international trade and restricted participation of rams in national sales and auctions (Picard-Hagen *et al.,* 2015).The degree to which the production efficiency of flocks are affected will depend on the aetiological agent and the manner in which it originally gained access to the animal population, and it is transmitted between animals in the same or a different population.

The effects of infectious reproductive diseases range from structural or functional defects of the reproductive system that can cause partial or complete reproductive failure, to tampering with the libido and mating ability of males. A decline in semen quality and fertility has also been associated with reproductive diseases (Toe *et al.,* 1994).

2.3.1.1. Common reproductive diseases of livestock

The most common reproductive diseases affecting livestock include epididymitis, orchitis, contagious ecthyma, ulcerative dermatosis and ovine posthitis. These diseases are known to affect different parts of the reproductive system and the aetiological agent has been identified in most cases.

Epididymitis is a disease affecting the epididymides that forms part of duct system and is characterized by lesions which cause pathological changes that negatively affects semen quality and fertility. Epididymitis occurs as a direct result of invasion into the epididymis from specific and non-specific bacterial agents. In sheep, the specific bacterial aetiology is thought to differ between mature and pre-pubertal rams, suggesting two forms of the disease, with *Brucella ovis* mainly associated with mature rams, and *Actinobacillus seminis* and other Gram-positive pleomorphic bacteria with pre-pubertal animals (Bagley *et al.,* 1985).

Orchitis is a reproductive disease of livestock characterized by inflammation of the testes and can be divided into an acute and chronic phase that depends on the aetiological agent and severity of infection.

Most causes of orchitis have been associated to preceding infections by epididymitis, with the disease less frequently occurring as a separate entity (Gouletsou & Fthenakis, 2006). The influence of orchitis is evident with reduced fertility with ensued economic losses due to the reduction in the number of offspring lambed and early culling of breeding rams. Bacterial organisms are the most commonly recognised cause of the disease (Gouletsou & Fthenakis, 2006). Viral aetiological agents have also been isolated from cases of orchitis in rams (Gouletsou & Fthenakis, 2015). Specific pyogenic organisms, *Brucella* species and numerous non-specific bacteria are involved in the clinical manifestation of orchitis (Ribeiro *at el.,* 2015) such as *Arcanobacterium pyogenes* (Gouletsou *et al.,* 2004; Gouletsou *et al.,* 2006), *Brucella ovis* and *Brucella melitensis,* which are both important pathogens of small ruminants (Burgess, 1982; Bulgin, 1990), *Actinobacillus seminis* (Heath *et al.,* 1991; Al-Katib & Dennis, 2007)*, Corynebacterium pseudotuberculosis* (Van Vuuren & Trichard, 2004c)*, Mannheimia haemolytica*, *Bibersteinia trehalosi* or *Pasteurella multocida* (Garcia-Pastor *et al.*, 2009)

Contagious ecthyma is a contagious disease charcaterized by lesions on the preputial orifice, scrotum and penis which come about in a progressive pattern (Billinis *et al.,* 2012). Initially local erythema occurs at the site of infection, followed by the formation of papules, vesicles, pustules and proliferative ulcers that gradually develop into scabs. The cause of contagious echtyma has been described as a *Parapox virus* that affects both genders of domesticated ruminants, primarily sheep and goats (Hosamani *et al.,* 2009). Secondary infections of contagious ecthyma by bacterial, fungal and maggot origins have been suggested to be common adding to complications of the disease (Hosamani *et al.,* 2009). Contagious ecthyma is transmitted venereally as descrined as a well as by direct or indirect contact (Nandi & Chowdhury*,* 2011).

Ulcerative dermatosis has been characterized as a disease of sheep. The appearance of the disease is seen in both younger and mature animals, most commonly affecting the latter (Van Vuuren & Trichard, 2004d). The disease manifests in two ways, i.e. crusted ulcerations can occur on either the skin of the lips, legs and feet or it can take on a venereal form, which is characterized with pustulous lesions that develop into ulcers enclosed by scabs on the vulva, the preputial orifice and less commonly the glans penis (Aiello & Moses, 2003). Diagnosis of the disease is often confused with other diseases such as blue tongue, sheep pox, ulcerative balanoposthitis, foot rot and contagious ecthyma. The source of the disease has yet to be classified but it has been suggested to take on a viral form closely related to that of the contagious ecthyma virus.

Ovine posthitis is also known as sheath rot/pizzle rot, *Pisgoed*, urine scald, balanoposthitis, balanitis or non-contagious posthitis. The disease is clearly defined has occurring in two progressive stages (Southcott, 1970). The initial infection occurs on the external prepuce, with necrosis of the external tissue occurring near or around the preputial orifice. Necrosis of the tissue develops into ulcers that become covered by dark scabs. This initial stage is not usually of economic importance, but scabs may join to cover and occlude the orifice with subsequent accumulation of urine in the prepuce. Secondary infection can develop and spread to the internal mucosa of the prepuce forming internal ulcerations. Accumulation of pus, necrotic material and urine around the sheath (Southcott, 1965) as well as swelling

of the prepuce (Pemberton, 1959) are common features of the disease. Sometimes severe cases can erupt spreading ulcerations to the glans penis (Van Vuuren & Trichard, 2004d). Rams affected by the disease are less likely to mate with ewes due to the painful nature of the symptoms and also due to the inability of males to extrude the penis due to swelling and scabbing of the prepuce). The aetiology of the disease has thus been characterized into three important factors: a legume (protein) rich diet that increases the presence of urea in urine and therefore ammonia production; the increased proportion of *Corynebacterium renale* as a consequence of elevated urea levels (Greig, 2007); and thirdly, low testosterone levels which applies only to whethers.

2.3.1.2. Ulcerative balanoposthitis

Ulcerative balanoposthitis (UB) and the analogous female condition, ulcerative vulvovaginitis (UV) is a reproductive disease that has been associated with numerous aetiological agents in a number of livestock species including horses (Allen & Umphenour, 2004), cattle (Van Vuuren & Trichard, 2004d; Pritchard *et al.,* 1997), goats (Tarigan *et al.,* 1990; Grewal & Wells, 1986) and sheep (Kidanemariam, 2003; El-Arabi *et al.,* 2014). In the literature the name of the disease may vary and the syndrome has been described in relation to the location of symptoms on the genitalia, with further classification of the female condition according to the clinical appearance of the symptoms (e.g. ulcerative, granulous).

In sheep, UB has been investigated less frequently than UV, with a small number of reports in the UK (Deas, 1983; Rutten, 2012; El-Arabi *et al.,* 2014), Australia (Webb & Chick, 1976), Argentina (Robles *et al.,* 2016) and South Africa (Kidanemariam, 2003; Trichard *et al.,* 1993; Ali, 2012). Furthermore, when the diagnostic method is based solely on clinical evaluation, it is possible that UB may be confused with other reproductive disorders of the penis and prepuce (e.g. contagious echtyma, ulcerative dermatosis and ovine posthitis) due to the similarity of the lesions on the external genitalia; possibly influencing the number of reports on the disease in the literature. For example, Bath and de Wet (2000) described the aetiology and clinical signs of a disease that are characteristic to UB, but named it pizzle rot, which refers to the reproductive disease caused by a nitrogen-rich diet and the microbial organism *Corynebacterium renale*. El-Arabi *et al.* (2014) also classified UB and pizzle rot as the same disease, incriminating both *Mycoplasma mycoides* species and a *Corynebacterium* species as causative agents involved in the disease.

Other studies such as Bush *et al.* (2006), reported on the occurrence of lesions on the prepuce and penis, but the lack of a comprehensive description of the clinical manifestation and microbial organisms involved cannot differentiate the lesions from that of other infectious diseases of the reproductive system, and the disease involved in producing the lesions remains unclear. Although other reproductive disorders of the external genitalia are not generally associated with mating (Van Vuuren & Trichard, 2004d), thorough investigation of the infectious nature of the lesions and/or the description of the pathogen involved will be required to differentiate them from UB.

2.3.1.2.1 Epidemiology

Natural transmission of the infectious agent between ram and ewe during sexual activity seems to be the most important way in which the disease is spread throughout and between flocks (Deas, 1983; Trichard *et al.,* 1993), with non-venereal transfer also suggested (Van Vuuren & Trichard, 2004d; Robles *et al.,* 2016). The clinical manifestation has occasionally been reported singly in ewes, with no simultaneous incidence in rams (Cottew *et al.,* 1974; Doig & Ruhnke, 1977; Ball & McCaughey, 1987).

In South Africa, UB is thought to be more prominent in rams, especially young rams (Gummow & Staley, 2000; Kidanemariam, 2003). Higher prevalence in rams may be explained by the higher sexual activity during the breeding season leading to a higher chance of ram-to-ram infection via an infected ewe or an increased likelihood of trauma to the genitals as a result of multiple servicing (Gummow & Staley, 2000). Gummow and colleagues (2010) further implied that these findings may be biased in that rams are examined for breeding soundness prior to the mating season thus are more readily identified with UB than ewes who may be affected but are not readily examined for the disease.

The incidence of UB has been observed before and during the mating season where the outbreak of the acute stage of UB is generally observed a few days after introduction of infected individuals into mating flocks (Trichard *et al.,* 1993; Gummmow & Staley, 2000; Robles *et al.,* 2016). An incubation period of 4-6 days and 18-20 days has been described by Van Vuuren and Trichard (2004) and Deas (1983), respectively. Infected rams are able to copulate and ewes experience normal pregnancy upon fertilization. Depending on the severity of the disease rams may then become reluctant or completely refuse to undergo sexual intercourse due to the painful nature of the lesions involved. Furthermore, in more advanced stages the animal may become depressed and sometimes stand aside and assume an arched back stance (Trichard *et al.,* 1993; Van Vuuren & Trichard, 2004d).

In South Africa, the outbreak of the disease is predominant in Dorper sheep, and is thought to spread to 100% of sheep on primary outbreak with lower incidence of about 50% thereafter, with 2-4% of national flocks affected at any one time (Trichard *et al.,* 1993; Van Vuuren & Trichard, 2004d). These results were concurrent with that of Webb and Chick (1976) that claimed that out of 240 Australian Merino ewes approximately 50% of them had developed a degree of vulvovaginitis. In the UK, Deas (1983) described a lower rate on infection between 20 and 30% of flocks but explained that this may vary according to the severity of the disease.

Due to the implications of the disease on the sexual activity of rams and its high infection rates within and across flocks, the disease has been identified as economically important to the South African mutton industry (Trichard *et al.,* 1993). Bath and de Wet (2000) estimated that as a consequence of the repressed reproductive performance of rams lambing percentages may be reduced to 50% comparing to usual standards of 100% or more. Results from a survey conducted by Gummow and Staley (2000) on ulcerative balanoposthitis in South Africa were in agreement with these results stating that the 56.4%

of ewes infected with the disease observed a lambing percentage less than a 100%. Furthermore, economic losses have also been described as a result of increased culling of diseases animals and inability of stock and stud owners to participate in national sales and auctions (Kidanemariam, 2003).

2.3.1.2.2 Clinical manifestations

The clinical symptoms of the male condition are more consistent in the literature than in ewes, and differ only in their location on the male genitalia. In rams the disease can affect the glans penis or the glans penis and the prepuce with ulcerative balanitis (Webb & Chick, 1976; Ball *et al.,* 1991; Kidanemariam, 2003) and ulcerative balanoposthitis (El- Trichard *et al.,* 1993; Gummow & Staley, 2000; Arabi *et al.,* 2014; Robles *et al.,* 2016) describing the disease, respectively.

The onset of UB can generally be defined by hyperaemia and inflammation of the mucosal membrane of the penis (Van Vuuren & Trichard, 2004d). As the infection progresses small scattered papulo-vesicular lesions which develop into erosions or ulcers are observed on the glans penis (Webb & Chick, 1976; Deas, 1983; Trichard *et al.,* 1993; Kidanemariam, 2003; El-Arabi *et al.,* 2014; Robles *et al.,* 2016) and the penis may be covered with fibrinous or mucopurulent exudate (Ball *et al.,* 1991; Van Vuuren & Trichard, 2004d; El-Arabi *et al.,* 2014; Robles *et al.,* 2016). The acute ulcers are filled with blood clots that are sensitive and easily tear, and during servicing attempts blood may ooze from the lesions staining the wool surrounding the vulva and prepuce in ewes and rams, respectively (Deas, 1983; Bath & de Wet, 2000; Kidanemariam, 2003; Robles *et al.,* 2016). Blood markings surrounding the genitals are usually regarded as the first sign that the infection has been established in a flock (Pritchard *et al.,* 2008).

In more progressive stages erosion can be extensive and ulcers may cover most of the glans penis and advance to the preputial orifice (Deas, 1983; Trichard *et al.,* 1993; El-Arabi, 2014; Robles *et al.,* 2016). Furthermore, scabs can form over the ulcers leaving exposed raw, bleeding surfaces if removed. In less frequent but more complicated cases, phimosis and para-phimosis of the prepuce can occur and the penis cannot be retracted into the sheath (Trichard *et al.,* 1993; Kidanemariam, 2003). Sometimes necrotic material can occlude the preputial opening and the prepuce may be swollen and oedematous making it difficult and in some cases impossible to extrude the penis, even manually (Deas, 1983; Webb & Chick, 1976). The variations in the gross clinical appearance described by several authors have been suggested to depend on the invasion from secondary bacteria and the stage and severity of infection (Vuuren & Trichard, 2004d).

2.3.1.2.3 Aetiology

Ulcerative lesions of the penis and prepuce have been described in the literature since the early 1900's. It is only now towards the later part of the 20th century that ovine ulcerative balanoposthitis has become more prominent in the literature, especially in South Africa and the UK. Greig (2007) suggests that, ulcerative balanitis (balanoposthitis) that cannot be explained by a *Parapox virus* (ORF) or a urease-

producing diphteroid such as *Corynebacterium renale*, is one that is classified as having an unknown aetiology. In several countries, many assumptions of infectious agents associated with the disease have been deduced, however this has been achieved with little confidence and inconsistent experimental isolation of specific organisms to back it up (Webb & Chick, 1976; Deas, 1983; Linklater & Smith, 1993; Kidanemariam, 2003; Van Vuuren & Trichard, 2004d; Pritchard *et al.,* 2008; Ali, 2012; Robles *et al.,* 2016).

Although it was previously thought that the disease was not associated with Chlamydophila or a virus (Webb & Chick, 1976; Jones *et al.,* 1983; Deas, 1983; Kidanemariam, 2003; Trichard *et al.,* 1993); more recent studies conducted on flocks in Switzerland and the UK have isolated ovine *Herpes virus* type 2 from vulvar and penile specimens, suggesting its possible role in UB/UV (Pritchard *et al.,* 2008; Rutten, 2012). Pritchard *et al.* (2008) observed the possible association between shedding ovine *Herpers virus* type 2 and the early stages of the condition.

Growing evidence suggests that the microbial cause of UB/UV could be a mollicute belonging to the genera *Mycoplasma*, *Ureasplasma, Achoeloplasma* or a combination thereof (Van Vuuren & Trichard, 2004d). Although *Ureaplasma* species have been isolated in the normal genital tract of ewes, a number of serotypes and unidentified species have been associated with UB/UV (Cottew *et al.,* 1974; Livingstone & Gauer, 1983; Trichard *et al.,* 1993; Kidanemariam, 2003; Ali, 2012). *Acholeplasma axanthum* and *Acholeplasma laidlawi* have been isolated from cases of UV and UV/UB, respectively; their role in the disease is still unknown (Jones *et al.,* 1983; Kidanemariam, 2003; Ali, 2012). *Mycoplasma* species from vulvar lesions (Jones *et al.,* 1983) and cases of vulvovaginitis (Cottew *et al.,* 1974), vulvitis (Ball & McCaughey, 1987), ulcerative vulvovaginitis (Trichard *et al.,* 1993) ulcerative vulvitis (Kidanemariam, 2003; Ali, 2012) and ulcerative balanoposthitis (Trichard *et al.,* 1993) and ulcerative balanitis and rams have been reported (Kidanemariam, 2003; Ali, 2012). Furthermore, unpublished data that granular-vesicular vaginitis and balanitis could be reproduced in sheep following the application of a combination of mollicutes to the mucous membrane of the genitals is proof that either one or more species could thus be involved in the disease (Van Vuuren & Trichard, 2004d).

The first *Mycoplasma* isolate to be associated with the disease in sheep was *Mycoplasma* species 2D (Cottew *et al.,* 1974; Linvingstone & Gauer, 1983), however its presence as a natural inhabitant of the reproductive tract (Linvingstone & Gauer, 1983) with no subsequent research to verify its role as an aetiological agent warranted further investigation of the disease by other authors.

In South Africa, the disease was first identified in 1976 in Dorper flocks in the Calvinia district of the Northern Cape. It has since been presumed to have spread throughout the country where sheep are reared and Dorper flocks prevail (Gummow & Staley, 2000). The first bacteriological study on the incidence of UB in Dorper flocks was conducted by Trichard *et al.* (1993) who proposed another *Mycoplasma* species, *Mycoplasma mycoides mycoides* large colony (MMMLC), as the primary causative agent of UB. Kidanemariam (2003) supported these findings and reported consistent isolation

of MMMLC in diseased Dorper sheep around the country compared to healthy ones, with an isolation rate of 61.5% and 6%, respectively. The MMMLC has not been clearly defined as being associated with reproductive diseases in sheep, and is generally considered a *Mycoplasma* occurring in goats, isolated in cases of polyarthritis, conjunctivitis, keratitis, pneumonia and cervical abscesses (Rosendal, 1994). Kidanemariam (2003) also went on to suggest a synergistic relationship with *Arcanobacterium pyogenes* in that it had concurrently high isolation rates (31.7%) with MMMLC and its pyogenic nature could be involved in the more progressive clinical stages of the disease, with 74% of strains isolated from severe clinical cases of UB/UV.

Ali (2012) went on to molecularly characterize the *Mycoplasma* species isolated from cases of UB, from both new samples and samples previously collected by Kidanemariam (2003). Ali (2012) found that although the number of isolates used in this study was small, comparing to Kidanemariam (2003) and Trichard *et al.* (1993) MMMLC had a poor isolation rate (~6%), with *Mycoplasma arginini* being the most frequently isolated *Mycoplasma* species. Although *Mycoplasma arginini* has been isolated from animals suffering from mastitis, pneumonia, arthritis and reproductive diseases (Ali, 2012) its role in UB is undetermined and its natural occurrence in the genital tract (Kidanemariam, 2003) as well as presumed low pathogenicity in animals would propose otherwise (Ali, 2012)-rephrase sentence-too long. The findings from these studies suggest that the aetiological agent involved in ulcerative lesions of the external genitalia in South Africa is still unresolved.

Apart from MMMLC, a mixed microflora of other mollicutes and pathogenic and non-pathogenic bacteria have also been isolated less frequently from the mucosal membranes of the penis and prepuce in affected sheep in South Africa. Mollicutes isolated included *Mycoplasma agalactiae* (Kidanemariam, 2003), *Mycoplasma bovigenatalium*, *Mycoplasma mycoides capri*, *Mycoplasma* species group 7, unidentified *Mycoplasma* and Ureaplasma species (Trichard *et al.,* 1993; Kidanemariam, 2003), *Mycoplasma arginini* (Trichard *et al.,* 1993; Kidanemariam, 2003; Ali, 2012), *Acholeplasma Laidlawii* (Kidanemariam, 2003; Ali, 2012), *Mycoplasma ovine/caprine serogroup II*, and *Mycoplasma canadense* (Ali, 2012). A number of these mollicutes are known to be pathogenic and have been involved in other livestock diseases (Ali, 2012).

The opportunistic bacteria *Arcanobacterium pyogenes* has been most consistently isolated from field cases of UB/UV in South Africa (Van Vuuren & Trichard, 2004d) and is presumed to be a secondary bacterial infection to *Mycoplasma* causing the progressive purulent clinical manifestations observed (Kidanemariam, 2003). *Arcanobacterium pyogenes* has also been significantly isolated in a number of other cases of UV/UB around the world (Pritchard *et al.,* 2008; Robles *et al.,* 2016). Some other known pathogenic bacteria included *Erysipelothrix rhusiopathiae, Pasteurella multocida, Enterococcus faecalis, Haemophilus somnus* also known as *Histophilus somni, Eschericia coli* and Staphylococci and Streptococc*i* species. These were isolated less frequently and most had concurrently similar isolation rates in healthy animals (Kidanemariam, 2003).

Other studies on UB/UV where *Mycoplasma* species were not identified in diseased specimens were also reported in the literature. In Argentina, *Arcanobacterium pyogenes* and *Pasteurella Multocida* were the most prevalent bacterial species isolated in Patagonian Merinos that experienced UB/UV (Robles *et al.,* 2016). Bacteriological findings in Scottish sheep, showed that the main predominant bacterial species isolated in rams and ewes infected with the disease differed. *Mannheimia haemolytica* and Streptococcus species formed the largest percentage of isolates in rams with concurrently higher isolation rates of Corynebacteria and Staphylococci species in ewes (El-Arabi *et al.,* 2014). Intravaginal inoculation of a mixture of bacterial strains of the *Haemophilus/Histophilus* group isolated from field cases of UV in Northern Ireland was able to reproduce a disease identical to that of the field condition in both ewes and rams, following mating (Ball *et al.,* 1991). *Mycoplasma*, *Ureaplasma* and viruses were not isolated prior to or during the experiment. *Haemophilus/Histophilus* strains were the only bacteria consistently isolated from the mucosa of the vulva and glans penis of experimental cases.

2.4 History of analytical methods used in the identification of bacteria in South African cases of ulcerative balanoposthitis

In order to make a valid clinical diagnosis and develop an effective treatment that can aid in the management and control of an infectious disease, identification and classification of microbes involved in the incidence of the diseases, is essential. Identification of bacterial species can be described as assigning an unknown microbial organism to a particular class in an existing classification (Priest, 2003). In clinical microbiology the aim is usually directed at characterizing bacteria to the most appropriate classification level that is informative enough to associate an organism(-s) to a disease. In some cases a bacterial taxon (not necessarily at the species level) may stand out from the expected microbiome, and therefore species level identification is not required. In others, species level identification is essential to differentiate organisms from that of the normal microflora where no differences can be seen at the higher classification levels. Some diseases may be linked with a change in the microbial diversity, which then result in a shift in community composition between the healthy and the diseased state. In other scenarios, an increase in the relative abundance of bacteria which normally occur in the animal, to higher than normal levels, may also result in the incidence of a disease.

Previous research on ulcerative balanoposthitis in South Africa has been aimed at identifying a causative agent from swabs and scrapings taken from the mucosal lining of the penis and prepuce of rams. These studies investigated suspected pathogens for their isolation rate in healthy and diseased sheep, with significant differences further suggesting the association of a pathogen(-s) to the disease. The methods used to identify bacteria in clinical samples included broad categorization of bacteria by means of culturing, followed by more specific definitive tests that allowed researchers to characterize the bacteria present.

The above-mentioned UB-related studies have focussed primarily on identifying bacteria from the class mollicutes, specifically *Mycoplasma*, which have previously been associated with other reproductive

diseases in livestock species (Basemen & Tuly, 1997; Nicholas *et al.,* 2008) and are currently considered potential aetiological agents of UB, although the cause is still unresolved. Attempts to identify other pathogenic bacteria have also also been condcted. No attempts at characterizing and comparing the entire bacterial microbiome surrounding the penis in healthy and diseased Dorper rams were carried out. A description of the diagnostic methods previously used as well as their advantages and disadvantages are described in the sections below.

2.4.1 Culture identification

Culturing of bacteria is a technique that has been practiced for over 50 years and is described as the golden standard in diagnostic microbiology (Padmanabhan *et al.,* 2013). This technique involves the isolating and proliferation of bacteria from clinical specimens on various growth media under different environmental conditions, allowing researchers to make initial presumptions of bacterial identity according to their growth requirements and characteristics. Studies on cases of UB in South Africa have been carried out with the aim to culture and investigate the bacterial composition of clinical samples; to determine bacterial colonies that could be involved in the infection; and obtain sufficient growth of pure organisms for further identification and classification using various types of tests (Didelot *et al*., 2012).

The initial step involved in the clinical diagnosis of an aetiological agent(-s) in UB in South Africa has been based on the cultivation of bacteria (Trichard *et al.,* 1993; Kidanemariam, 2003) and *Mycoplasma* (Trichard *et al.,* 1993; Kidanemariam, 2003; Ali, 2012) on a battery of standard media known to support the growth and routine isolation of most pathogenic bacteria (Quinn *et al.,* 2011) and *Mycoplasma* species (Nicholas *et al.,* 2008Along with specialised media, the environmental conditions under which these cultures were grown differed across studies but remained within the scope of what is suggested optimal growth conditions for culturing most pathogens (Razin, 1994; Quinn *et al.,* 2011).

In order for bacteria to be cultivated, similar growth conditions found *in-vivo* need to be established *in-vitro*, and the use of specific media and conditions relating but not limited to the nutrient requirements, temperature, pH, moisture, and oxygen and carbon dioxide requirements should allow for such organisms to proliferate under artificial conditions (Quinn *et al.,* 2011). For example, Trichard *et al.* (1993) incubated some cultures at 10% CO_2 in order to cultivate bacteria from the *Haemophilus* and *Actinobacillus* groups that have previously been identified as causing reproductive diseases in sheep (Bagley *et al.,* 1985; Van Vuuren & Trichard, 2004a & b), which can only optimally grow in higher than normal CO_2 concentrations.

The need to create a perfect *in-vitro* environment plays an important role in diagnostic microbiology in that it is not possible to produce a culture environment specific to all bacteria and growth conditions will be chosen according to pathogens suspected of being present in clinical samples, thus limiting the number of pathogens tested for. It is well accepted in the literature that not all bacterial species can be cultured and it is believed that about 10% of infectious bacterial pathogens are unculturable or difficult

to grow *in-vitro* (Didelot *et al.,* 2012). For example, the diagnosis of *Mycoplasma* from infectious diseases using conventional culture techniques can be complicated in that some *Mycoplasma* species are unculturable and the minority of *Mycoplasma* existing in different habitats that have been cultivated grow slowly (several days to weeks) and poorly even on the best *Mycoplasma* medium available and are often overgrown by other bacteria in samples (Razin, 1994; Cai *et al*., 2014). In addition, bacteria that grow rapidly in a culture often outgrow species that grow slowly in the same culture, thus removing what could be an "important" organism from further analysis. Bacterial cultivation can be very laborious, time-consuming and in many cases requires skill and equipment (Bowler *et al.,* 2001; de Boer *et al.,* 2010) thus making it near impossible to test for an entire bacterial population within a clinical sample.

2.4.2 Definitive identification

Colonial morphology, physical traits such as colour and motility, and other key growth characteristics observed on primary bacterial isolation media are used to make a presumptive bacterial identification to higher levels of classification, with the choice of more specific identification procedures thereafter based on the need for a definitive identification of an organism down to species level (Snyder & Atlas, 2006). Different types of tests have been described in the literature that can be used as additional or alternative definitive bacterial identification methods to bacterial culture (Quinn *et al*., 2011). Assays previously used in South Africa for definitive bacterial identification were based on previously cultured bacteria and included biochemical, serological and molecular assays, depending on the nature of the study.

2.4.2.1 Biochemical assays

Conventional culture methods are the most widely used methods in diagnostic microbiology and base definitive identification of bacteria on phenotypic analysis that look most commonly at biochemical characteristics with other physiological, chemotypic (particularly fatty acid components) and metabolic characteristics also studied (Quinn *et al.,* 2011).

Bacteria are divided into different groups that are known to have specific biochemical characteristics (Healing, 1993) and the type and number of biochemical tests required to identify a bacterium will vary from one group to another. Some bacterial taxa require fewer tests and are easier to differentiate than others. A number of biochemical assays can be routinely used for many groups of bacteria (e.g. fermentation or utilization of carbohydrates, oxidase, amino acid degrading enzymes, nitrate reduction) whilst others are confined to a single family, genus or species (e.g. coagulase test for Staphylococci, pyrrolidonyl arylamidase test for Gram-positive cocci) (Baron, 1996). Specific growth media, nutrients, chemicals or growth conditions are applied by biochemical tests to elicit observable and measurable responses from microorganisms that can in turn be used to identify and classify them (Gupta, 2011). These tests are abundant throughout the literature, ranging in their specificity, sensitivity and

reproducibility, and the type used largely depends on the pathogen suspected as well as the knowledge, expertise and preferences of the researcher in terms of the protocols followed (Healing, 1993).

Using broader as well as more specific biochemical assays specifically designed to test the catabolic properties of different types of bacteria, Trichard *et al.* (1993) and Kidanemariam (2003) followed standard criteria and interpretation of results from laboratory protocols, literature (Stephens *et al.,* 1983; Razin & Freundt, 1984; Quinn *et al.,* 2011) and purchased commercial analytical systems (API 10S, API Coryn, Microbat 12A and 12B) and kits (staph and strep kits). Using these specific biochemical tests, they were able to isolate and characterize a number of pathogenic bacteria from healthy and diseased rams that fall under characteristics of Gram-negative and –positive bacteria along with other mollicutes (Trichard *et al.,* 1993; Kidanemariam, 2003).

Conventional methods have been perfected over many years and are still the preferred method in diagnostic microbiology. This approach does however, like any other method, have several drawbacks which limit accurate diagnosis. Phenotypic tests can be subjective and interpretation of results can be highly biased according to the technologist and his or her experience (Stager & Davis, 1992). When phenotypic profiles do not follow standard criteria as a result of for example evolution, environmental stress factors, unusual microorganisms which are rare or may result from new strains or poorly updated commercial bacterial identification databases, the accuracy of biochemical tests may be questioned and it is possible that tests are interpreted to fit expectations (Kolbert & Persing, 1999; Petti *et al.,* 2005). Furthermore, some biochemical tests cannot sufficiently differentiate bacteria to species level, for several reasons that may include similarity in phenotypic profiles with other bacteria or phenotypic paucity rendering them indistinguishable (Vandamme *et al.,* 1996). Bosshar *et al.* (2003) demonstrated that out of a 136 known aerobic gram-positive rods isolated from clinical specimens only a minority of isolates were identified using phenotypic methods. Classification was only achieved at the genus level, with 71 of 136 (52.2%) isolates phenotypically characterized at the genus level and the remaining 65 of 136 (47.8%) isolates could not be discriminated at any taxonomic level. In this case additional non-phenotypic tests will be required to further characterize these isolates (Gupta, 2011).

Studies on *Mycoplasma* suggest that characterization of species from cultures can be achieved through the aid of biochemical tests (Gois *et al*., 1974). Although biochemical tests are still widely used, the diagnosis of *Mycoplasma* through these assays are limited, with consequential increase in the dependence of other tests such as serological and molecular assays (Gois *et al.,* 1974). This is especially true for the *Mycoplasma mycoides* cluster, under which the suspected aetiological agent of UB, MMMLC, is grouped, where only a few biochemical and physiological properties can be used to differentiate them (Cottew *et al.,* 1987).

Many authors have described biochemical tests used in the classification of *Mycoplasma* which include but is not limited to glucose fermentation, arginine utilization, urea hydrolosis, sensitivity to digitonin, serum digestion, tetrazolium HCL reduction, phosphatase activity, catalase and oxidase tests and

metabolisms of a number of carbohydrates (Kidanemariam, 2003). Only one study in South Africa carried out by Ali (2012), used biochemical assays to characterize *Mycoplasma* species in rams affected with UB. Ali (2012) biochemically characterized *Mycoplasma* present in clinical samples by following standard protocols developed by Erno and Stipkovits (1973) that involved the use of several of the biochemical tests previously mentioned, preceded by a standard staining method (Gram's stain), additional growth on a selective medium and further differentiation according to lactose fermentation. The biochemical tests used by Ali (2012) proved to be limited in their ability to differentiate *Mycoplasma* to species level and additional molecular tests were used to further differentiate these *Mycoplasma*s.

2.4.2.2 Serological identification

Serological tests have been extensively used in differentiating *Mycoplasma* species, and can include amongst others immunoperoxidase, immunofluorescent antibody assays (IFA), Western immunoblotting, complement fixation, and enzyme linked immunoabsorbent assays (ELISA) (Clyde, 1964; Rosendal & Black, 1972; Krogsgaard-Jensen, 1972; Goll, 1994; Thacker & Talkington, 1995; Nicholas *et al.,* 2008), with some of these techniques being more specific and/or sensitive than others. Trichard *et al.* (1993) and Kidanemariam (2003) made use of the direct-IFA (uses only one antibody) and indirect-IFA (uses two antibodies) techniques, respectively, which have proven to be highly species-specific tests for the identification of various *Mycoplasma* species observed in the penile samples of South African Dorper rams, with higher sensitivity suggested with the indirect-IFA technique (Rosendal & Black, 1972). As a result, the indirect-IFA technique is now the most commonly used method of the two (Turgeon, 2015).

The serological techniques rely on fluorophore-conjugated antibodies that fluoresce on attachment to its complementary *Mycoplasma* antigen. These tests although specific, may be subjective in that fluorescent signal intensity used to determine the presence of a specific *Mycoplasma* cannot be quantified and grading of the intensity signal depends on the opinion of the researcher, which could lead to false positives Furthermore, this test is generally suitable to test for a small number of bacteria (Gillespie, 1994). In *Mycoplasma* identification, the antibodies filtered from antisera prepared from experimental animals as a response to a number of reference *Mycoplasma* strains are required which makes this test time-consuming and limits the identification of *Mycoplasma*s to only those reference strains used. Trichard *et al.* (1993) and Kidanemariam (2003) limited *Mycoplasma* identification to 9 *Mycoplasma* strains, although suggesting that several mollicutes could be involved in the disease.

The diagnosis of *Mycoplasma*s is complicated by the fact that members of the genus seem to be closely related phenotypically and genotypically. Serologically, the results achieved for members of the *Mycoplasma mycoides* cluster are often difficult to interpret with serological cross-reactions not uncommon. Although not believed to be observed by Trichard *et al.* (1993) and Kidanemariam (2003), cross-reactions have been observed amongst *Mycoplasma* species such as between *Mycoplasma capricolum* subspecies *capripneumoniae* and *Mycoplasma* species strain PG50 as well as some other

strains of *Mycoplasma capricolum* subspecies *capricolum* (Bolske *et al.,* 1988). Confirmation of these *Mycoplasma* species with serological assays often require further tests to characterize them, which include protein profiling (Thiaucourt *et al.,* 2000) and DNA hybridization (Bonnet *et al.,* 1993).

2.4.2.3 16S rRNA gene sequencing

Although conventional phenotypic methods and serological tests are readily used to identify bacteria these tests require live cultures and experience in diagnostic microbiology. The methods are also considered to be subjective resulting in misclassifications; and are sometimes limited in their ability to identify bacteria to species level or differentiate bacteria within the same bacterial taxa, which will then require additional tests to further characterize them.

Some of the limitations of phenotypic and serological bacterial characterization techniques can be overcome by using molecular genotypic assays as a complementing or alternative mode of identification (Fihman *et al.,* 2007; Petti, 2007; Monteserin *et al.,* 2016). Ali (2012) was the first researcher in South Africa to make use of molecular technology as a complementary test to characterize previously cultured *Mycoplasma* isolated from Dorper rams affected with UB. Genotypic identification methods in general are considered to be quicker, objective, highly sensitive and specific, and reproducible compared to phenotypic methods (Fihman *et al.,* 2007; Srinivasan, 2015).

Genotypic tests are not limited to providing a complementary test to bacterial cultures, and can be conducted from bacterial DNA extracted directly from clinical samples. It can be used to overcome the limitations of culture-based identification of bacteria mainly related to the identification of bacterial species that are unculturable, rare or difficult to grow *in-vitro*, with the added benefit of novel species discovery of previously unculturable organisms (Woo *et al.,* 2003). The most obvious advantages of 16S gene sequencing of bacterial DNA directly from samples is that researchers are not required to be experienced in diagnostic microbiology and that there is no need to suspect infectious agents prior to experimentation, omitting other potential pathogens through limited culture media, growth conditions and phenotypic tests aimed at specific pathogens, as well as poor expertise. All bacteria present have the ability to be identified.

The analytical properties of genetic material, directed at deoxyribonucleic acid and ribonucleic acid (DNA and RNA), have allowed a number of genotypic methods to be developed under headings such as molecular hybridization, molecular fingerprinting, microarrays, PCR and gene sequencing (Quinn *et al.,* 2011).

The molecular technique gene sequencing is currently the gold standard for defining bacteria and can be described as the sequencing of a specific section of a bacterial genome previously amplified using polymerase chain reaction (PCR). The 16S ribosomal RNA (rRNA) gene is approximately 1550 nucleotide base pairs (bp) long and in certain bacterial taxa multiple copies can exist throughout the

genome (Větrovský & Baldrian, 2013). The 16S gene has widely been accepted in the literature as a tool for identifying bacterial isolates (Patel, 2001; Mignard & Flandrois, 2006; Revetta *et al.,* 2010; de Melo Oliveira *et al.,* 2013; Monteserin *et al.,* 2016) and diagnosing microbial diseases (Trotha *et al.,* 2001; Lecouvet, 2004; Lau *et al.,* 2006; Woo *et al.,* 2007; Kuhn *et al.,* 2011; Srinivasan *et al.,* 2015). It is currently the most extensively used bacterial gene in clinical microbiology and can be used in its entirety or in smaller sections to characterize bacteria, depending on the sequencing technology used (which will be described in the next section). Its popularity lies in its degree of conservation/universality across the domain bacteria with few other genes as equally conserved (Clarridge 3rd, 2004).

The 16S gene consists of eight highly conserved regions that flank nine hypervariable regions across the bacterial domain (Armougom & Raoult, 2009). The regions that exhibit variable nucleotide sequences can be used to compare DNA homology between bacteria in turn differentiating them according to species. Thus, the 16S rRNA sequence of a bacterium is a genotypic feature which allows the identification of organisms at the species level. Furthermore, the conserved regions of the 16S gene can be used to identify bacteria covering most taxa or differentiate organisms belonging to a specific group (Baker *et al.,* 2003). In the most recent study on UB in South Africa, Ali (2012) characterized the *Mycoplasma* diversity in sheep affected with UB by means of 16S rRNA sequencing. He characterized *Mycoplasma* from sequenced PCR and cloned PCR products previously amplified from a forward primer derived from a section of the 16S gene universal to all bacteria and a reverse primer from a section of the 16S gene conserved across all *Mycoplasma*s; generating a DNA fragment that covered approximately 1078bp of the 16S gene (~1550bp) and that was specific to *Mycoplasma*.

During bacterial identification, 16S rRNA sequences are assigned according to the similarity of the query (amplified) 16S rRNA sequence to 16S rRNA reference sequence located in a database. Thus 16S sequence classification and taxonomic resolution is constrained to reference databases and to organisms that have previously been identified and sequenced. Like phenotypic techniques, 16S gene sequencing also uses a cut-off value for acceptable levels of similarity to identify bacterial species. Although the value is not as well-defined as should be and no consensus algorithm has been used to calculate this value, it is suggested by most taxonomists that a percent identity score of ≥97% and ≥99% between a query sequence and a reference sequence is usually sufficient to define the 16S rRNA sequence to a genus and a species, respectively (Petti, 2007). In the literature, disparities have been observed in the total percentage of bacteria identified to species level across all cut-off definitions (even with no definition) (Clarridge 3rd, 2004), suggesting that this parameter plays an important role in definitive pathogen diagnosis. Higher percentage of species are usually classified at no or lower identity cut-off values, this however increases the chance of misclassification.

Several studies compared the ability of 16S gene sequencing with conventional methods in identifying clinically important bacterial groups from cultures (Bosshard *et al.,* 2003; Heikens *et al.,* 2005; Song *et al.,* 2005; Adderson, 2008; Rhoads *et al.,* 2012). From these studies it is deduced that when 16S rRNA

sequencing is used to identify rare bacteria and bacteria with unusual biochemical profiles, a more effective degree of identification can be achieved.

Two separate studies investigated the potential of the MicroSeq 500 16S gene-based identification system to identify bacteria and found that
this system was able to characterize 81% of clinically significant bacterial isolates with ambiguous phenotypic profiles (Woo *et al.,* 2003) and 89.2% of unusual aerobic Gram-negative bacilli to species level (Tang *et al.,* 1998). When 16S rRNA sequencing was used to resolve the identification of phenotypically unidentified bacterial isolates the method proved to be effective, with more than almost 90% of isolates defined (Drancourt *et al.,* 2000).

When testing the ability of routine identification by the 16S rRNA in the clinical laboratory, 16S rRNA was able to characterize 243 out of 382 clinical isolates of mycobacteria which are known to be clinically important pathogens and are in general slow-growing and/or difficult to identify phenotypically (Clarridge 3rd, 2004). Species-specific identification of *Campylobacter* has been reported as problematic due to the absence of suitable biochemical assays. 16S rRNA sequencing was demonstrated as a reliable tool in differentiating most *Campylobacter* species with exceptions observed with *Campylobacter jejuni, Campylobacter coli* and *Campylobacter lari* that could not be differentiated due to identical or almost identical 16S rRNA gene profiles (Gorkiewicz *et al.,* 2003). This is one of the limitations of 16S gene sequencing, and usually alternative genes may be used to further differentiate these organisms. One such gene, ropB, has been described as a complementary gene in differentiating between *Bacillus cereus* and *Bacillus anthracis* that have identical 16S gene sequences (Blackwood *et al.,* 2004).

In South Africa, Ali (2012) made use of biochemical assays to detect isolates from the genus *Mycoplasma*, and characterized *Mycoplasma*s to species level using 16S rRNA sequencing (Sanger sequencing technology), in an attempt to verify previous findings on UB suspecting MMMLC as the primary aetiological agent. Ali (2012) isolated species that had previously been identified in other studies (Trichard *et al.,* 1993; Kidanemariam, 2003) as well three new species of *Mycoplasma*, two that had never been identified in cases of UB (*Mycoplasma* Mmm. jvc1 and *Mycoplasma* species USP 120) and one (*Mycoplasma canadense*) never observed in sheep in South Africa. Ali (2012) succeeded in identifying MMMLC, however noted that the rate at which it was isolated (2 out of 34 isolates) was much smaller than expected, when compared to much higher isolation rates suggested in previous studies on the disease (Trichard *et al.,* 1993; Kidanemariam, 2003). Although the number of isolates tested for was small, the low isolation rate of MMMLC may suggest an overestimation in the role of MMMLC as an aetiological agent brought about as a result of serological cross-reactions with other *Mycoplasma* species, which is a common flaw of serological identification methods (Bolske *et al.,* 1988).

Janda and Abbott (2007) reviewed the use of 16S gene sequencing in clinical diagnosis and found that in most cases (>90%) only genus identification is possible with species identification that ranges from 65-83% according to studies. Bacterial groups including but not limited to Enterobacteriaceae,

Mycobacteria, Achromobacter, Stenotrophomonas and Actinomyces have demonstrated resolution problems at these levels. As is suggested from these findings, no technique can be a 100% effective in identifying bacteria and although 16S rRNA sequencing can be useful in classifying bacteria especially those difficult to classify using other techniques, it is not without limitations and species level identification is not always guaranteed.

The lack of resolution to the species level can be attributed to a number of reasons mainly related to the size of the 16S gene DNA fragment sequenced; the chosen hypervariable region if only a section of the 16S gene is used; the sequence reference database used; and the quality of the 16S sequences produced. Although Clarridge, 3rd (2004) has suggested that the initial 500bp of the gene is sufficient to classify most clinical bacterial isolates to species level, sequencing of the whole 16S gene is often required to distinguish bacteria between taxa, where sometimes even the whole 16S gene is not enough (Gorkiewicz *et al.,* 2003). It is not to say that species characterization cannot be achieved using DNA fragments that are less than 500bp long, but that the smaller a DNA fragment is the smaller the resolution and the less confidently the sequence can be attributed to a bacterial taxon at the species level. Tewari *et al.* (2011) compared the use of a DNA fragment of ~500bp and another of ~40bp, and found that the longer 16S gene fragment could identify 87% of isolates to species level compared to 43% using the shorter section. Furthermore, not all short fragments could be identified to the genus, family or order taxonomic levels as compared to 100% identification by the 500bp.

Some short 16S rRNA sequenced gene fragments have proved to have taxonomic resolution closer to that of full length 16S sequences as a result of the variable region they cover. Some hypervariable regions are more variable across bacterial taxa than others, thus have higher potential of discriminating between bacterial taxa. Recovery and coverage levels of bacterial taxa using different 16S regions measured relative to the results achieved using full length sequences, observed that the V6 hypervariable region did poorer than other regions in identifying organisms from the gut microbiota at the genus level (Lui *et al.,* 2008).

The bacterial identification and discriminative ability of the 16S rRNA gene using a reference sequence database relies on four key components, i.e. the deposition of complete and good quality nucleotide sequences into databases; correct taxonomic classification of each sequence deposited; the number of 16S rRNA sequences deposited; and the taxonomic-resolution of sequences available, especially as longer sequences are added to databases. It is no secret that the sequences found in reference databases contain errors especially those that were deposited years back prior to the development of high-fidelity, automated systems (Ashelford *et al.,* 2005). Although these errors are believed to be corrected, they and errors due to newer sequencing technologies are still observed at varying degree throughout the available databases such as Genbank, RDP and Greengenes (DeSantis *et al.,* 2006; Cole *et al.,* 2009; Benson *et al.,* 2013). For example, Genbank accepts any taxonomically named sequence deposited and is not peer reviewed (Clarridge 3rd, 2004). A study by Clayton *et al.* (1995)

revealed that for every two 16S rRNA gene sequence coding for the same bacteria, a minimum of 26% had more than 1% sequencing errors, and 50% had more than 2% sequencing errors.

Poor quality reference sequences diverge the identity score between a query sequence and the reference sequence, where ambiguous bases reduce the score preventing an identity match to the desired level of taxonomy (e.g. family instead of species). In addition, these errors may lead to wrong best identity match between a query and a reference sequence. In other words, it is possible that a query sequence may match, by random chance, to another reference sequence rather than the erroneous reference sequence it is actually meant to match to, thus assigning a wrong taxonomic classification to the query sequence.

Partial 16S rRNA reference sequences in databases also reduce the taxonomic resolution potential of longer query sequences, for the latter will be classified according to the classification of the shorter reference sequence if longer reference sequences are not available. Furthermore, it is also possible that a number of sequences are misclassified in databases. Bacterial 16S sequences are usually assigned to a taxon based on phenotypic tests; and in some cases when phenotypic traits are poorly described or tests are faulty, the wrong taxonomic label may be assigned to a 16S rRNA reference sequence and subsequently to query sequences (Clarridge 3rd, 2004). When many sequences are analysed at once it is hard to determine whether a sequence is difficult to identify or that it is incorrectly identified.

As mentioned by Drancourt *et al.* (2000) errors are also observed in query sequences affecting taxonomic resolution and correct species identification. These errors may arise as a result of inefficient PCR amplification and/or poor sequencing. It is possible that during PCR amplification, suboptimal reaction conditions, such as improper temperature, may cause amplification to be less sensitive or non-specific or may stop earlier than expected (producing 16S rRNA sequences that are shorter than expected). These types of errors cannot always be avoided all together but can be minimized by testing various reaction conditions that provide the most sensitive and specific PCR reaction (Wilson, 1997). PCR amplification with reduced sensitivity (i.e. increased base mismatch) and specificity (i.e. binding primers to the wrong region of the 16S gene) can produce sequences that are classified at lower resolutions, that are wrongly classified, and that are not amplified from the desired variable region. This can cause biases in the bacterial diversity and representation of organisms. In addition, incomplete amplified sequences can act as a primer template which can bind to related DNA fragments, generating a product of two sequence combinations which is otherwise known as a chimera, and as a result the new chimeric sequence sequenced can be mistaken for new bacterial taxa, amplifying the bacterial diversity of a sample (Edgar *et al.,* 2011). The rate of chimerism has been reported by Haas *et al.* (2011) to range between 5% and 45%.

The number of sequence errors produced during sequencing vary according to the sequencing platform used. For example an error rate (i.e. number of errors per total nucleotide bases sequenced) of 1-2%

has been reported using the Roche 454 sequencing platform, mainly due to errors related to homopolymers (Margulies *et al.,* 2005). Sequence errors in query sequences have the same effect on bacterial identification than sequencing errors observed in 16S reference sequences, previously mentioned.

The reference databases available to researchers vary in diversity and overall taxonomic structure and do not always agree on species names or taxonomy (Werner *et al*., 2011; Santamaria *et al*., 2012).This will influence the taxonomic classification of 16S sequences, ultimately affecting the perceived diversity. It is possible that a bacterial taxon (phylum to genus level) is under-represented in one database as compared to another, and the representation of this bacterial taxon will thus also be under-represented in the 16S data when assigning an identity to 16S query sequences, as compared to using a database where the taxon is highly represented (Huse *et al.,* 2008). In addition, for some genera too few species have been sequenced and deposited in the databases and the identity score for a particular query sequence never exceeds the genus level (i.e. 97% identity score) (Drancourt *et al.,* 2000). For example, Genbank comprises the largest databank of nucleotide sequences with over 100 000 sequences of the 16SrRNA gene deposited (http://www.ncbi.nlm.nih.gov/genbank/) to match against, as compared to a commercially available MicroSeq database with ~2000 16S rRNA sequences (Petti, 2007). Furthermore, not all databases are updated regularly, thus new sequences and sequences with higher taxonomic resolution available cannot be accessed. The Greengenes reference database (DeSantis *et al*., 2006), for example is updated only periodically.

Apart from limitations of 16S gene sequencing to correctly classify sequences and provide sequences of high taxonomic resolution, as determined according to the sequences available in the reference databases, the observed diversity can also be influenced by DNA extraction and primer selection. DNA extraction protocols have been recognised in the literature with special attention to their ability and efficiency in extracting DNA from all bacterial taxa and the effects thereof on bacterial diversity analysis (Feinstein *et al.,* 2009; Wesolowska-Andersen *et al.,* 2014; Hart *et al.,* 2015). The disruption and lysis of bacterial membranes during the first step of DNA extraction can be biased towards specific bacterial taxa due to differences in cell wall integrity and structure. Gram-positive bacteria generally require bead-beating to break down the cell wall and expose the DNA molecule. Drancourt *et al.* (2000) observed that 2% of clinical bacterial isolates could not be identified due to inappropriate DNA extraction methods.

The choice of a primer or primer set used to amplify regions of the 16S rRNA gene and the hypervariable regions sequenced vary across the literature *(*Baker *et al.,* 2003; Edwards *et al.,* 2006; *Sogin et al.,* 2006; Roesch *et al.,* 2007; Andersson *et al., 2008;* Frank *et al.,* 2008*;* Huse *et al.,* 2008; Liu *et al.,* 2008; Wang & Qian, 2009). The selection of a primer or primer set has been described as the most important step in accurate 16S gene sequencing analysis, in that suboptimal primer pairs can lead to under- or over-representation of bacterial taxa, or select against a specific species or group if they match the consensus sequence poorly (Hamady & Knight, 2009; Klindworth *et al.,* 2012). For example, where a primer is not conserved across all bacterial taxa present in a clinical sample, some mismatches

between a primer and the complementary 16S DNA fragment may occur, decreasing the efficiency by which it amplifies the DNA fragment, causing that taxa to be under-represented in the sample. This effect can usually be decreased by the use of degenerative primers.

No true consensus on what the best universal (i.e. conserved across all bacterial taxa) primer pair has yet been agreed upon (Schloss *et al.,* 2011) and the coverage of bacterial taxa differ according to the taxa present in microbial communities analysed (Huse *et al.,* 2008; Soergel *et al*., 2012). Klindworth *et al.* (2012) experimented with a number of universal primer pairs and found variations in the coverage percentages, which ranged from 76.5% - 96.7%, with differences in the number of phyla that could be recovered. For example, they reported that through modelling experiments the primer pair S-D-Bact-0785-b-A-18 and S-D-Bact-0785-a-A-21 is believed to cover 49 out of 59 phyla from the bacterial domain as compared to the primer pair S-D-Bact-0564-a-S-15 and S-D-Bact-0785-b-A-18 that had a slightly lower overall coverage for bacteria but only failed to detect 4 bacterial phyla.

The limitations of 16S rRNA sequencing to some extent affects final bacteria identifications and the observed bacterial diversity. A number of methods have however been developed to overcome or reduce these constrains, such as choosing a curated reference database that is frequently updated and has a low rate of sequence errors; maximizing the length of the DNA fragment sequenced; choosing a hypervariable region that is highly variable and can maximize bacterial differentiation; and making sure the best primer pairs, PCR reaction conditions and DNA extraction methods are chosen. The ability of 16S gene sequencing as a viable method in characterizing bacteria cannot be rejected.

2.5 High throughput next generation sequencing

With the advent of sequencing technology bacterial identification is no longer limited to cultures and identifying a small number of bacteria per sample. Microbial communities can now be sequenced and identified with limited bias by using universal 16S rRNA primers that during a PCR reaction amplify all microbial nucleic acids present in a sample. The ability to identify all bacteria has several advantages such as the identification of unculturable bacteria and the identification of polymicrobial infections (Cai *et al*., 2014), which is presumed a challenge for culture-based methods (Fenollar *et al.,* 2006; Roger *et al.,* 2009). In addition, characterizing all bacteria removes the need for researchers to make presumptions about the pathogen(-s) that may be involved in a disease of unknown aetiology, allowing all pathogens to be identified rather than testing for a select few. UB is both a disease of unknown aetiology and one suspected of polymicrobial infection.

High throughput next generation sequencing (HT-NGS) is a new era of sequencing technologies, that has only been available for the last decade, with an ever growing number of technologies surpassing the abilities of older systems (Van Dijk *et al.,* 2014), varying in their sequencing chemistries, read lengths and throughput capacities (Loman *et al.,* 2012). High-throughput next generation sequencing technologies can generate thousands to millions of sequences in a single run, generating sequence

information data of larger magnitudes in a small period of time (Kircher & Kelso, 2010; Buermans & Den Dunnen, 2014). A number of methods of HT-NGS exists and currently 16S amplicon sequencing is one that is widely used to characterize microbial communities. Its ability to generate thousands to millions of sequences in a single run has described 16S amplicon sequencing as an efficient tool in characterizing bacteria from clinical samples with high microbial diversity, where both dominant and lowly abundant or rare bacterial taxa (of which can be pathogens) can be characterized (Siqueira *et al.,* 2012). This means that 16S amplicon sequencing is able to recover greater species richness and characterize almost complete microbial communities from clinical samples with the theoretical ability to identify all bacterial species present, including those not previously cultured. Along with qualitative characterization, molecular profiling of bacterial communities using 16S amplicon sequencing also provides quantitative characterization through more reliable estimates of the relative abundance of bacterial taxa present in clinical samples (Pallen *et al*., 2010). This permits researchers to compare bacterial taxa within and between communities beyond simple presence or absence diagnostic criteria. Further advantages of using16S amplicon sequencing include lower cost, the ability to sequence many samples at once and avoiding cloning biases.

The use of 16S amplicon sequencing in veterinary clinical laboratories is still limited (Tewari *et al.,* 2011) but its use in the veterinary research sector is fairly well documented in the literature (Gill *et al.,* 2006; Costa *et al.,* 2012; Oikonomou *et al.,* 2012; Steelman *et al.,* 2012). These studies and other studies related to human health (Lamont *et al.,* 2011; Dickson *et al.,* 2013; Wade, 2013) have illustrated the role of 16S amplicon sequencing in understanding diseases and the importance of profiling bacterial microbiomes in healthy and diseased states in order to understand the underlying biological and clinical role of bacteria, especially if the disease state is related to changes in community composition from that of the healthy state (Roger *et al.,* 2009). For example, a study that compared the microbial community within the gut of healthy horses and horses that had coilitis, found that the microbiome differed significantly between the two groups and that a shift in microbiome rather than an enrichment in an individual bacterial pathogen may be associated with the disease (Costa *et al.,* 2012).

The ability to characterize a healthy microbiome in order to understand specific diseases has been an innovation of HT-NGS technology such as 16S amplicon sequencing. It is thought that a core microbiome exists within body sites and maintain a healthy state. Altered microbiomes from that of the healthy states can help researchers determine if the shift in community composition is merely associated as markers for diseases or can also actively contribute to the pathology of the disease (Goodrich *et al.,* 2014). If a shift in the community composition is observed from one state to the other, it is conceivable to narrow down to individual bacterial taxa that are observed to be involved in driving this shift and determine their biological and clinical roles in disease. Furthermore, defining healthy microbiomes allows researchers to identify biological systems that demonstrate natural inter-individuals variations in diversity, limiting later complications in the identification of microbial components and imbalances that may cause or reflect a disease state in such a system (Lloyd-Price *et al*., 2016). An

understanding of the structure of a healthy microbiome in the absence of disease is therefore an essential initial step to identifying bacteriological patterns implicated in disease.

Although 16S amplicon sequencing technology surpass Sanger technology on many levels, it is at a cost of 16S rRNA fragment length and sequence quality. High throughput next generation sequencing technologies are only capable of sequencing smaller 16S rRNA fragments, thus affecting the taxonomic resolution of bacteria identified. Nonetheless, newer 16S amplicon sequencing platforms are being developed at a rapid rate to overcome this with sequencing platforms like Illumina Miseq (http://www.illumina.com) that make use of paired-end sequencing exploiting larger read lengths of up to ~550bp that were not previously possible (Schimer *et al.,* 2015). Inevitably, all sequencing technologies produce sequencing errors, some more than others, which if sufficient can mistake a sequence for a new species, inflating species richness of a sample. Higher sequencing depth can however make these errors less relevant, where many copies of the same sequence (with fewer or no errors) can be retrieved. Several strategies have been proposed to deal with these errors which include removing parts of sequences that are of poor quality or ambiguous, or removing sequences all together so that only good quality sequences remain for downstream analysis (Schloss *et al.,* 2011).

With the development of 16S amplicon sequencing, the number of sequences produced per sequencing run has expanded into the millions producing data in the gigabytes range, which poses a challenge for both data storage and analysis (Belak *et al*., 2013). Normal computers do not often have sufficient memory, storage and processing power to analyse these types of datasets and alternative resources such as online servers are usually required. In addition, 16S amplicon sequencing data analysis requires knowledge of bioinformatics, which is the research field focussing on the study of methods for retrieving, analysing and storing biological data (Belak *et al.*, 2013). Researchers also require to select amongst a plethora of tools/software available to analyse and make sense of the data, each with their own algorithms and variable parameters, which is too much to describe for the scope of this study. No tool/software is a “one-size-fits-all” and users are required to have knowledge and understanding of the various analyses steps in a given application (such as characterizing a microbiome) and how different software operates at each step, affecting the end results. The type of data analysis steps and the order in which they occur depends on the researcher and no set protocol exists for how data should be analysed. Often the way data is processed depends on the data, the research question and whether sequence quality or quantity or a balance between the two is most important.

Despite all the limitations of 16S rRNA sequencing along with the complications of data analysis, the potential of this method is growing with the development of newer technologies and it remains evident that this is a valuable tool in investigating microbial communities directed at understanding diseases. Although it cannot be expected that all 16S rRNA sequences will be characterized to species level, the microbial diversity no matter the taxonomic resolution, can contribute insight into “what is there “and “how does it differ within and between groups”, forming the basis from which future studies can be directed in an attempt to identify a causative agents of UB.

References

Adderson, E.E. 2008. Identification of clinical coryneform bacterial isolates: comparison of biochemical methods and sequence analysis of 16S rRNA and rpob genes. *Journal of Clinical Microbiology*, 46(3):921-927.

Agricultural Geo-Referenced Information System. 2007. Agricultural regions [Online]. Available: http://www.agis.agric.za/agisweb/$WEB_HOME?MIval=content2_h.html&id=nr_atlas_agric_regions [2016, August 30].

Aiello, S.E. & Moses, M.A. (eds.). 2015. Overview of Ulcerative Dermatosis of Sheep, in *The Merck Veterinary Manual*. New Jersey: Merck and Co. [online]. Available: http://www.merckvetmanual.com /mvm/index.html [2016, September 1].

Al-Katib, W. & Dennis, S. 2007. Epididymal and testicular lesions in rams following experimental infection with *Actinobacillus Seminis*. *New Zealand Veterinary Journal,* 55(3):125-129.

Allen, G.P. & Umphenour, N.W. 2004. Equine coital exanthema. *Infection*, *41*(43):44.

Ali, H. 2012. Molecular characterization of *Mycoplasma* species isolated from the genital tract of Dorper sheep in South Africa. Master's thesis. Pretoria: University of Pretoria.

Armougom, F. & Raoult, D. 2009. Exploring microbial diversity using 16s rRNA high-throughput methods. *Journal of Computer Science & Systems Biology*, 2:074-092, doi:10.4172/jcsb.1000019.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J. & Weightman, A.J. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology,* 71(12):7724-7736.

Bagley, C.V., Paskett, M.E., Matthews, N.J. & Stenquist, N.J. 1985. Prevalence and causes of ram epididymitis in Utah. *Journal of the American Veterinary Medical Association,* 186(8):798-801.

Baker, G., Smith, J.J. & Cowan, D.A. 2003. Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods*, 55(3):541-555.

Ball, H., Kennedy, S. & Ellis, W. 1991. Experimental reproduction of ovine vulvitis with bacteria of the Haemophilus/Histophilus Group. *Research in Veterinary Science,* 50(1):81-85.

Ball, H.J. & McCaughey, W.J. 1987. Experimental intramuscular inoculation of Tylosin in the elimination of Ureaplasmas from ewes. *The Veterinary Record,* 120(23):557-558.

Baron, E.J. 1996. Chapter 3. Classification, in S. Baron (ed.). *Medical microbiology*. Galveston: The University of Texas Medical Branch at Galveston [Online]. Available: http://www.ncbi.nlm.nih.gov/ books/NBK8406/ [2016, August 04].

Bath, G. & de Wet, J. 2000. Pizzle disease, in *Sheep and Goat Diseases*. Cape Town: Tafelberg Publishers Limited. 74-76.

Bearden, H.J. & Fuquay, J.W. 1997. *Applied Animal Reproduction* (4th ed.). New Jersey: Prentice-Hall Inc.

Belák, S., Karlsson, O., Leijon, M. & Granberg, F. 2013. High-throughput sequencing in veterinary infection biology and diagnostics. *Revue Scientifique et Technique,* 32:893-915.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. 2013. GenBank. *Nucleic Acids Research*, *41*(D1):D36-D42, doi:10.1093/nar/gks1195.

Bester, N. 2006. Effect of different dietary energy levels on productive and reproductive traits in Dorper rams. Doctoral dissertation. Bloemfontein: University of the Free State.

Billinis, C., Mavrogianni, V.S., Spyrou, V. & Fthenakis, G.C. 2012. Phylogenetic analysis of strains of orf virus isolated from two outbreaks of the disease in sheep in Greece. *Virology Journal,* 9(1):1.

Blackwood, K.S., Turenne, C.Y., Harmsen, D. & Kabani, A.M. 2004. Reassessment of sequence-based targets for identification of Bacillus species. *Journal of Clinical Microbiology,* 42(4):1626-1630.

Bolske, G., Msami, H., Humlesjo, N.E., Erno, H. & Jonsson, L. 1988. *Mycoplasma capricolum* in an outbreak of polyarthritis and pneumonia in goats. *Acta Veterinaria Scandinavica*, 29(3-4):331-338.

Bosshard, P.P., Abels, S., Zbinden, R., Bottger, E.C. & Altwegg, M. 2003. Ribosomal DNA sequencing for identification of aerobic gram-positive rods in the clinical laboratory (an 18-month evaluation). *Journal of Clinical Microbiology*, 41(9):4134-4140.

Bowler, P.G., Duerden, B.I. & Armstrong, D.G. 2001. Wound microbiology and associated approaches to wound management. *Clinical Microbiology Reviews*, 14(2):244-269.

Buermans, H.P.J. & Den Dunnen, J.T. 2014. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1842*(10):1932-1941.

Bulgin, M.S. 1990. Epididymitis in rams and lambs. *Veterinary Clinics of North America: Food Animal Practice*, 6(3):683-690.

Burgess, G. 1982. Ovine contagious epididymitis: A review. *Veterinary Microbiology*, 7(6):551-575.

Cai, H.Y., Caswell, J.L. & Prescott, J.F. 2014. Nonculture molecular techniques for diagnosis of bacterial disease in animals a diagnostic laboratory perspective. *Veterinary Pathology. Online*, 51(2):341-350.

Clarridge, J.E., 3rd. 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews,* 17(4):840-62.

Clayton, R.A., Sutton, G., Hinkle Jr, P.S., Bult, C. & Fields, C. 1995. Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *International Journal of Systematic and Evolutionary Microbiology*, *45*(3):595-599.

Clyde, W.A., Jr. 1964. *Mycoplasma* species identification based upon growth inhibition by specific antisera. *Journal of Immunology*, 92:958-965.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M. & Tiedje, J.M., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, *37*(suppl 1):D141-D145, doi:10.1093/nar/gkn879.

Costa, M.C., Arroyo, L.G., Allen-Vercoe, E., Stämpfli, H.R., Kim, P.T., Sturgeon, A. & Weese, J.S. 2012. Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput

sequencing of the V3-V5 region of the 16S rRNA gene. *PloS One, 7*(7):p.e41484, doi.org/10.1371/journal.pone.0041484.

Cottew, G.S., Breard, A., DaMassa, A.J., Ernø, H., Leach, R.H., Lefevre, P.C., Rodwell, A.W. & Smith, G.R. 1987. Taxonomy of the *Mycoplasma* mycoides cluster. *Israel journal of medical sciences*, *23*(6):632-635.

Cottew, G., Lloyd, L., Parsonson, I. & Hore, D. 1974. Isolation of a *Mycoplasma* from vulvovaginitis in sheep. *Australian Veterinary Journal*, 50(12):576-577.

DAFF. 2014. A profile of the South African mutton market value chain. Department of Agriculture, Forestry and Fisheries [Online].Available:http://www.nda.agric.za/doaDev/sideMenu/Marketing/ Annual%20Publications/Commodity%20Profiles/Livestock/Mutton%20Market%20Value%20Chain%20Profile%202014.pdf [2016, August 30].

de Boer, R.F., Ott, A., Kesztyus, B. & Kooistra-Smid, A.M. 2010. Improved detection of five major gastrointestinal pathogens by use of a molecular screening approach. *Journal of Clinical Microbiology*, 48(11):4140-4146.

de Melo Oliveira, M.G., Abels, S., Zbinden, R., Bloemberg, G.V. & Zbinden, A. 2013. Accurate identification of fastidious gram-negative rods: Integration of both conventional phenotypic methods and 16S rRNA gene analysis. *BMC Microbiology*, 13(1):1.

Deas, D.W. 1983. Chapter 31. Ulcerative balanitis and ulcerative vulvitis, in W.B. Martin (ed.). *Diseases of Sheep*. London: Blackwell Scientific Publications. 144-146.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7):5069-5072.

Dickson, R.P., Erb-Downward, J.R. & Huffnagle, G.B. 2013. The role of the bacterial microbiome in lung disease. *Expert Review of Respiratory Medicine*, *7*(3):245-257.

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E. & Crook, D.W. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics,* 13(9):601-612.

Doig, P.A. & Ruhnke, H.L. 1977. Isolation of Ureaplasma from sheep with granular vulvitis. *The Veterinary Record,* 100(9):179-180.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J.P. & Raoult, D. 2000. 16S Ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *Journal of Clinical Microbiology*, 38(10):3623-3630.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16):2194-2200.

Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D.M., Saar, M.O., Alexander, S., Alexander, E.C. & Rohwer, F. 2006. Using Pyrosequencing to Shed Light on Deep Mine Microbial Ecology. *BMC Genomics,* 7(1):1.

El-Arabi, A.A., Taylor, D.J., Logue, D.N., Benothman, M. & Abuargob, O. 2014. Clinical and bacteriological findings of balanoposthitis and vulvitis in Scottish sheep. *Journal of Veterinary Advances*, 1(4):481-487.

Erno, H. & Stipkovits, L. 1973. Bovine *Mycoplasmas*. *Acta Veterinaria Scandinavica*, 14(3):450-463.

Feinstein, L.M., Sul, W.J. & Blackwood, C.B. 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Applied and Environmental Microbiology*, 75(16):5428-5433.

Fenollar, F., Roux, V., Stein, A., Drancourt, M. & Raoult, D. 2006. Analysis of 525 samples to determine the usefulness of PCR amplification and sequencing of the 16S rRNA gene for diagnosis of bone and joint infections. *Journal of Clinical Microbiology*, 44(3):1018-1028.

Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A. & Olsen, G.J. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA Genes. *Applied and Environmental Microbiology,* 74(8):2461-2470.

García-Pastor, L., Blasco, J. & Barberán, M. 2009. Pasteurellosis as a cause of genital lesions in rams. A descriptive study. *Small Ruminant Research,* 87(1):111-115.

Gill, J.J., Sabour, P.M., Gong, J., Yu, H., Leslie, K.E. & Griffiths, M.W. 2006. Characterization of bacterial populations recovered from the teat canals of lactating dairy and beef cattle by 16S rRNA gene sequence analysis. *FEMS Microbiology Ecology*, 56(3):471-481.

Gillespie, S.H. 1994. Investigation of specimens from the genital tract and diagnosis of sexually transmitted diseases (STDs), in S.H. Gillespie (ed.). *Medical Microbiology Illustrated.* Oxford: Butterworth-Heinemann Ltd. 222-233.

Goiŝ, M., Kuksa, F., Franz, J. & Taylor-Robinson, D. 1974. The antigenic differentiation of seven strains of *Mycoplasma* hyorhinis by growth-inhibition, metabolism-inhibition, latex-agglutination, and polyacrylamide-gel-electrophoresis tests. *Journal of Medical Microbiology*, 7(1):105-116.

Goll, R. 1994. Identification of *Mycoplasma*s isolated from domestic animals, in H.W. Whitford, R.F. Rosenbusch & L.H. Lauerman (eds.). *Mycoplasmas in Animals: Laboratory Diagnosis*. Iowa: Iowa State University Press. 15-30.

Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., Knight, R. & Ley, R.E. 2014. Conducting a microbiome study. *Cell*, *158*(2):250-262.

Gouletsou, P. 2004. Experimentally induced orchitis associated with arcanobacterium pyogenes: clinical, ultrasonographic, seminological and pathological features. *Theriogenology,* 62(7):1307-1328.

Gouletsou, P.G. & Fthenakis, G.C. 2006. Ovine orchitis, with special reference to orchitis associated with Arcanobacterium pyogenes. *Small Ruminant Research*, *62*(1):71-74.

Gouletsou, P.G. & Fthenakis, G.C. 2015. Microbial diseases of the genital system of rams or bucks. *Veterinary Microbiology,* 181(1):130-135.

Gouletsou, P., Fthenakis, G., Tzora, A., Cripps, P. & Saratsis, P. 2006. Isolation of *Arcanobacterium pyogenes* from the scrotal skin and the prepuce of healthy rams or from rams with testicular abnormalities. *Small Ruminant Research*, 63(1):177-182.

Greig, A. 2007. Ulcerative balanitis and vulvitis, in I.D. Aitkin (ed.). *Diseases of sheep* (4th ed.). Oxford: Blackwell Science. 143-145.

Grewal, A. & Wells, R. 1986. Vulvovaginitis of goats due to a *Herpesvirus*. *Australian Veterinary Journal,* 63(3):79-82.

Gummow, B. & Staley, G. 2000. A retrospective survey of ulcerative balanoposthitis and vulvovaginitis in South African Dorper sheep. *Livestock Health and Production*, 12:31-36.

Gupta, A.K. 2011. Molecular studies on gut associated microbiota in housefly and fleshfly. Doctoral thesis. Pune: National Centre for Cell Science.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E. & Methé, B. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research*, 21(3):494-504.

Hafez, E.S.E. 1987. *Reproduction in farm animal* (5th ed.). Philadelphia: Lea and Febiger printing.

Hamady, M. & Knight, R. 2009. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome research,* 19(7):1141-1152.

Hancock, A., Younis, P., Beggs, D., Mansell, P. & Pyman, M. 2015. Infectious reproductive disease pathogens in dairy herd bulls. *Australian Veterinary Journal,* 93(10):349-353.

Hart, M.L., Meyer, A., Johnson, P.J. & Ericsson, A.C. 2015. Comparative evaluation of DNA extraction methods from faeces of multiple host species for downstream next-generation sequencing. *PloS One*, 10(11):e0143334, doi:10.1371/journal.pone.0143334.

Healing, D. 1993. Cowan and Steel's manual for the identification of medical bacteria. *Journal of Medical Microbiology*, 39(4):319-319.

Heath, P.J., Davies, I.H., Morgan, J.H. & Aitken, I.A. 1991. Isolation of *Actinobacillus Seminis* from rams in the United Kingdom. *The Veterinary Record,* 129(14):304-307.

Heikens, E., Fleer, A., Paauw, A., Florijn, A. & Fluit, A.C. 2005. Comparison of genotypic and phenotypic methods for species-level identification of clinical isolates of coagulase-negative staphylococci. *Journal of Clinical Microbiology*, 43(5):2286-2290.

Hosamani, M., Scagliarini, A., Bhanuprakash, V., McInnes, C.J. & Singh, R.K. 2009. Orf: An update on current research and future perspectives. *Expert Review of Anti-Infective Therapy*, 7(7):879-893.

Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A. & Sogin, M.L. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics*, 4(11):e1000255, doi:10.1371/journal.pgen.1000255.

Janda, J.M. & Abbott, S.L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology,* 45(9):2761-2764.

Jones, G.E., Rae, A.G., Holmes, R.G., Lister, S.A., Jones, J.M., Grater, G.S. & Richards, N. 1983. Isolation of exotic *Mycoplasma*s from sheep in England. *The Veterinary Record,* 113(23):540.

Kidanemariam, A. 2003. Identification and characterization of the primary infectious agents associated with ovine ulcerative balanoposthitis and vulvovaginitis in South Africa. Master's thesis. Pretoria: University of Pretoria.

Kircher, M. & Kelso, J. 2010. High-throughput DNA sequencing–concepts and limitations. *Bioessays*, *32*(6):524-536.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glöckner, F.O. 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation

sequencing-based diversity studies. *Nucleic Acids Research*, 41(1):e1, doi: 10.1093/nar/gks808.

Kolbert, C.P. & Persing, D.H. 1999. Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. *Current Opinion in Microbiology,* 2(3):299-305.

Krogsgaard-Jensen, A. 1972. *Mycoplasma*: Growth precipitation as a serodiagnostic method. *Applied Microbiology*, 23(3):553-558.

Kuhn, C., Disque, C., Muhl, H., Orszag, P., Stiesch, M. & Haverich, A. 2011. Evaluation of commercial universal rRNA gene PCR plus sequencing tests for identification of bacteria and fungi associated with infectious endocarditis. *Journal of Clinical Microbiology,* 49(8):2919-2923.

Lamont, R.F., Sobel, J.D., Akins, R.A., Hassan, S.S., Chaiworapongsa, T., Kusanovic, J.P. & Romero, R., 2011. The vaginal microbiome: new information about genital tract flora using molecular based techniques. *BJOG: An International Journal of Obstetrics & Gynaecology*, *118*(5):533-549.

Lau, S.K., Teng, J.L., Leung, K.W., Li, N.K., Ng, K.H., Chau, K.Y., Que, T.L., Woo, P.C. & Yuen, K.Y. 2006. Bacteremia caused by *Solobacterium moorei* in a patient with acute proctitis and carcinoma of the cervix. *Journal of Clinical Microbiology*, 44(8):3031-3034.

Linklater, K. A. & Smith, M. C. 1993. Diseases of the reproductive system and udder, in *Color Atlas of diseases of the sheep and goat*. London:Wolfe. 123-135.

Liu, Z., DeSantis, T.Z., Andersen, G.L. & Knight, R. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, 36(18):e120, doi: 10.1093/nar/gkn491.

Livingston, C.W., Jr & Gauer, B.B. 1983. Occurrence of *Mycoplasma* Sp (2D) in Texas sheep flocks. *American Journal of Veterinary Research*, 44(5):868-869.

Loman, N.J., Constantinidou, C., Chan, J.Z., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R. & Pallen, M.J. 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*, 10(9):599-606.

Low, J.C. & Graham, M.M. 1985. *Histophilus ovis* epididymitis in a ram in the UK. *The Veterinary record*, 117(3):64-65.

Marai, I., El-Darawany, A., Fadiel, A. & Abdel-Hafez, M. 2008. Reproductive performance traits as affected by heat stress and its alleviation in sheep. *Tropical and Subtropical Agroecosystems*, 8(3):209-234.

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. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057):376-380.

Mignard, S. & Flandrois, J. 2006. 16S rRNA Sequencing in Routine Bacterial Identification: A 30-Month Experiment. *Journal of Microbiological Methods,* 67(3):574-581.

Monteserin, J., Paul, R., Lopez, B., Cnockaert, M., Tortoli, E., Menéndez, C., García, M.J., Palomino, J.C., Vandamme, P., Ritacco, V. & Martin, A. 2016. Combined approach to the identification of clinically infrequent non-tuberculous mycobacteria in Argentina. *The International Journal of Tuberculosis and Lung Disease*, *20*(9):1257-1262.

Nandi, S., De, U.K. & Chowdhury, S. 2011. Current status of contagious ecthyma or orf disease in goat and sheep—a global perspective. *Small Ruminant Research*, 96(2):73-82.

Oikonomou, G., Machado, V.S., Santisteban, C., Schukken, Y.H. & Bicalho, R.C. 2012. Microbial diversity of bovine mastitic milk as described by pyrosequencing of metagenomic 16S rDNA.

Padmanabhan, R., Mishra, A.K., Raoult, D. & Fournier, P. 2013. Genomics and metagenomics in medical microbiology. *Journal of Microbiological Methods*, 95(3):415-424.

Pallen, M.J., Loman, N.J. & Penn, C.W. 2010. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. *Current Opinion in Microbiology*, *13*(5):625-631.

Patel, J.B. 2001. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis,* 6(4):313-321.

Pemberton, P. 1959. Non-contagious ovine posthitis ("sheath rot"): its treatment with cetrimide. *Australian Veterinary Journal,* 35(8):359-365.

Petti, C.A. 2007. Detection and identification of microorganisms by gene amplification and sequencing. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America,* 44(8):1108-1114.

Petti, C.A., Polage, C.R. & Schreckenberger, P. 2005. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. *Journal of Clinical Microbiology*, *43*(12), pp.6123-6125.

Picard-Hagen, N., Berthelot, X., Champion, J.L., Eon, L., Lyazrhi, F., Marois, M., Peglion, M., Schuster, A., Trouche, C. & Garin-Bastuji, B. N. 2015. Contagious epididymitis due to *Brucella ovis*: relationship between sexual function, serology and bacterial shedding in semen. *BMC Veterinary Research,* 11(1):1.

Price, E., Estep, D., Wallach, S. & Dally, M. 1991. Sexual performance of rams as determined by maturation and sexual experience. *Journal of Animal Science*, 69(3):1047-1052.

Priest, F.G. 2003. Rapid identification of microorganisms, in F.G. Priest & I. Campbell (eds.). *Brewing Microbiology*. New York: Springer. 305-328.

Pritchard, G., Cook, N. & Banks, M. 1997. Infectious pustular vulvovaginitis/infectious pustular balanoposthitis in cattle. *The Veterinary Record*, 140(22):587.

Pritchard, G.C., Scholes, S.F.E., Foster, A.P., Mitchell, E.S.E., Lawes, J., Ibata, G. & Banks, M. 2008. Ulcerative Vulvitis and Balanitis in Sheep Flocks. *Veterinary Record: Journal of the British Veterinary Association*, 163(3):86.

Quinn, P.J., Markey, B.K., Leonard, F.C., FitzPatrick, E.S., Fanning, S. & Hartigan, P. 2011. Section II Introductroy Bacteriology, in P.J. Quinn, B.K. Markey, F.C. Leonard, E.S. FitzPatrick, S. Fanning & P. Hartigan (eds.). *Veterinary Microbiology and Microbial Disease*. UK: Wiley-Blackwell. 115-178.

Razin, S. 1994. DNA probes and PCR in diagnosis of *Mycoplasma* infections. *Molecular and Cellular Probes*, 8(6):497-511.

Razin, S. & Freundt, E. A. 1984. The *Mycoplasma*s, in N. R. Krieg & R. G. Holt (eds.). *Bergey 's manual of Systematic Bacteriology.* London: Williams & Wilkins. 740-793.

Revetta, R.P., Pemberton, A., Lamendella, R., Iker, B. & Santo Domingo, J.W. 2010. Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. *Water research,* 44(5):1353-1360.

Rhoads, D.D., Cox, S.B., Rees, E.J., Sun, Y. & Wolcott, R.D. 2012. Clinical identification of bacteria in human chronic wound infections: culturing vs. 16S Ribosomal DNA sequencing. *BMC Infectious Diseases*, 12(1):1.

Ribeiro, M.G., Risseti, R.M., Bolaños, C.A.D., Caffaro, K.A., de Morais, A.C.B., Lara, G.H.B., Zamprogna, T.O., Paes, A.C., Listoni, F.J.P. & Franco, M.M.J. 2015. *Trueperella pyogenes* multispecies infections in domestic animals: a retrospective study of 144 cases (2002 to 2012).*Veterinary Quarterly*, *35*(2):82-87.

Robles, C., Paramidani, M. & Armando, S. 2016. Balanopostitis y vulvitis ulcerativa infecciosa en ovinos Merino de la Patagonia Argentina. *Veterinaria*, 33(338):1852-317X.

Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K., Kent, A.D., Daroub, S.H., Camargo, F.A., Farmerie, W.G. & Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal*, *1*(4):283-290.

Roger, G.B., Carroll, M.P. & Bruce, K.D. 2009. Studying bacterial infections through culture-independent approaches. *Journal of Medical Microbiology*, *58*(11):1401-1418.

Rosendal, S. 1994. Ovine and Caprine *Mycoplasma*s, in H.W., Whiteford, R.F., Rosenbusch & L.H. Lauerman (eds.). *Mycoplasmosis in Animals: Laboratory Diagnosis*. 84-95.

Rosendal, S. & Black, F.T. 1972. Direct and indirect immunofluorescence of unfixed and fixed *Mycoplasma* colonies. *Acta Pathologica Microbiologica Scandinavica Section B Microbiology and Immunology*, *80*(4):615-622.

Ross, M.H. & Pawlina, W. 2011. Male Reproductive System, in M.H. Ross & W. Pawlina (eds.). *Histology: A text and atlas* (6th ed.). Philadelphia: Lippincott Williams and Wilkins. 604-657.

Rust, J. & Rust, T. 2013. Climate change and livestock production: A review with emphasis on Africa. South *African Journal of Animal Science*, 43(3):256-267.

Rütten, M. 2012. Outbreak of ulcerative vulvitis in a sheep flock caused by ovine *Herpesvirus* Type 2. *Journal of Comparative Pathology,* 146(1):71.

Salhab, S., Zarkawi, M., Wardeh, M., Al-Masri, M. & Kassem, R. 2003. Characterization and evaluation of semen in growing Awassi ram lambs. *Tropical Animal Health and Production*, 35(5):455-463.

Santamaria, M., Fosso, B., Consiglio, A., De Caro, G., Grillo, G., Licciulli, F., Liuni, S., Marzano, M., Alonso-Alemany, D., Valiente, G. & Pesole, G. 2012. Reference databases for taxonomic assignment in metagenomics. *Briefings in bioinformatics*, 13(6):682-695.

Schloss, P.D., Gevers, D. & Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS One*, 6(12):e27310, doi:10.1371/journal.pone.0027310.

Schoeman, S., Cloete, S. & Olivier, J. 2010. Returns on investment in sheep and goat breeding in South Africa. *Livestock Science*, 130(1):70-82.

Sidwell, G.M. & Miller, L.R. 1971. Production in some pure breeds of sheep and their crosses. I. reproductive efficiency in ewes. *Journal of Animal Science*, 32(6):1084-1089.

Siqueira, J.F., Fouad, A.F. & Rocas, I.N. 2012. Pyrosequencing as a tool for better understanding of human microbiomes. *Journal of Oral Microbiology*, 4:10743, doi:10.3402/jom.v4i0.10743.

Snyder, J.W. & Atlas, R.M. 2006. *Handbook of Media for Clinical Microbiology* (2nd ed.). Florida: CRC Press.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. & Herndl, G.J. 2006. Microbial Diversity in the Deep Sea and the Underexplored "Rare Biosphere". *Proceedings of the National Academy of Sciences of the United States of America,* 103(32):12115-12120.

Song, Y., Liu, C., Bolanos, M., Lee, J., McTeague, M. & Finegold, S.M. 2005. Evaluation of 16S rRNA sequencing and re-evaluation of a short biochemical scheme for identification of clinically significant Bacteroides species. *Journal of Clinical Microbiology,* 43(4):1531-1537.

Southcott, W.H. 1965. Epidemiology and control of ovine posthitis and vulvitis. *Australian Veterinary Journal*, 41(8):225-234.

Southcott, W.H. 1970. Studies on posthitis and vulvitis of sheep. Doctoral dissertation. Sydney: University of Sydney.

Spies, D.C. 2011. Analysis and quantification of the South African red meat value chain, Doctoral dissertation. Bloemfontein: University of the Free State Bloemfontein.

Srinivasan, R., Karaoz, U., Volegova, M., MacKichan, J., Kato-Maeda, M., Miller, S., Nadarajan, R., Brodie, E.L. & Lynch, S.V. 2015. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PloS One,* 10(2):e0117617, doi:10.1371/journal. pone.0117617.

Stager, C.E. & Davis, J.R. 1992. Automated systems for identification of microorganisms. *Clinical Microbiology Reviews*, *5*(3):302-327.

Steelman, S.M., Chowdhary, B.P., Dowd, S., Suchodolski, J. & Janečka, J.E. 2012. Pyrosequencing of 16s rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Veterinary Research*, 8(1):1.

Stephens, L. R., Humphrey, J. D., Utile, P. B. & Barnum, D. A. 1983. Morphological, biochemical, antigenic and cytochemical relationships among *Haemophilus somnus*, *Haemophilus agni* and *Haemophilus haemoglobinophilus, Histophilus avis* and *Actinobacillus seminis*. *Journal of Clinical Microbiology,* 17(5):728-737.

Tang, Y.W., Ellis, N.M., Hopkins, M.K., Smith, D.H., Dodge, D.E. & Persing, D.H. 1998. Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative Bacilli. *Journal of Clinical Microbiology,* 36(12):3674-3679.

Tarigan, S., Ladds, P. & Foster, R. 1990. Genital pathology of feral male goats. *Australian Veterinary Journal*, 67(8):286-290.

Teankum, K. 2006. Detection of chlamydiae in boar semen and genital tracts. *Veterinary Microbiology*, 116(1):149-157.

Tewari, D., Cieply, S. & Livengood, J. 2011. Identification of bacteria recovered from animals using the 16s ribosomal RNA gene with pyrosequencing and Sanger sequencing. *Journal of Veterinary*

Diagnostic Investigation: official publication of the American Association of Veterinary Laboratory Diagnosticians Inc, 23(6):1104-1108.

Thacker, W.L. & Talkington, D.F. 1995. Comparison of two rapid commercial tests with complement fixation for serologic diagnosis of *Mycoplasma* pneumoniae infections. *Journal of Clinical Microbiology*, 33(5):1212-1214.

Thiaucourt, F., Lorenzon, S., David, A. & Breard, A. 2000. Phylogeny of the *Mycoplasma* mycoides cluster as shown by sequencing of a putative membrane protein gene. *Veterinary Microbiology*, 72(3):251-268.

Toe, F., Lahlou-Kassi, A. & Mukasa-Mugerwa, E. 1994. Semen characteristics of Ile-De-France rams of different age and physical condition. *Theriogenology*, 42(2):321-326.

Trichard, C.J.V., Jordaan, P., Prozesky, L., Jacobsz, E.P. & Henton, M.M. 1993. The identification of *Mycoplasma* mycoides mycoides LC as the aetiological agent of balanoposthitis and vulvovaginitis in sheep in South Africa. *The Onderstepoort Journal of Veterinary Research,* 60:29–37.

Trichard C. J .V. & van Tonder, E. M. 1994. Ulcerative balanoposthitis and vulvovaginitis in sheep and goats, in J. A. W. Coetzer, G. R. Thomson, R. C. Tustin & N. P.J. Kriek (eds.). *Infectious Diseases of Livestock with special reference to South Africa,* Vol 3. Cape Town: Oxford University Press. 2:1599-1602.

Trotha, R., Hanck, T., König, W. & König, B. 2001. Rapid ribosequencing–an effective diagnostic tool for detecting microbial infection. *Infection*, 29(1):12-16.

Van Dijk, E.L., Auger, H., Jaszczyszyn, Y. & Thermes, C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics,* 30(9):418-426.

Van Tonder, E.M. 1973. Infection of rams with *Actinobacillus seminis*. *Journal of the South African Veterinary Association*, 44(3):235-240.

Van Vuuren, M. & Trichard C.J.V. 2004a. Actinobacillus spp. infections, in J.A.W. Coetzer & R.C. Tustin (eds.). *Infectious Diseases of Livestock with special reference to Southern Africa* (2nd ed.), Vol 3. Cape Town: Oxford University Press.1646-1670.

Van Vuuren, M. & Trichard C.J.V. 2004b. Haemophilus spp. infections, in J.A.W. Coetzer & R.C. Tustin (eds.). *Infectious Diseases of Livestock with special reference to Southern Africa* (2nd ed.), Vol 3. Cape Town: Oxford University Press.1628-11634.

Van Vuuren & Trichard. 2004c. Irregular, Non-Sporing, Gram-Positive Rods, in J.A.W. Coetzer & R.C. Tustin (eds.). *Infectious Diseases of Livestock with special reference to Southern Africa* (2nd ed.), Vol 3. Cape Town: Oxford University Press.1915-1962.

Van Vuuren & Trichard. 2004d. Ulcerative balanoposthitis and vulvovaginitis in sheep and goats, in J.A.W. Coetzer & R.C. Tustin (eds.). *Infectious Diseases of Livestock with special reference to Southern Africa* (2nd ed.), Vol 3. Cape Town: Oxford University Press. 2151-2156.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews*, 60(2):407-438.

Wade, W.G. 2013. The oral microbiome in health and disease. *Pharmacological Research*, *69*(1):37-143.

Wang, Y. & Qian, P. 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PloS One*, 4(10):e7401, doi:10.1371/journal.pone.0007401.

Webb, R. & Chick, B. 1976. Balanitis and vulvo-vaginitis in sheep. *Australian Veterinary Journal*, 52(5):241-242.

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R. & Ley, R.E. 2012. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *The ISME journal*, *6*(1):94-103.

Wesolowska-Andersen, A., Bahl, M.I., Carvalho, V., Kristiansen, K., Sicheritz-Pontén, T., Gupta, R. & Licht, T.R. 2014. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome*, 2(1):1.

Wilson, I.G. 1997. Inhibition and facilitation of nucleic acid amplification. *Applied and environmental microbiology*, *63*(10):3741.

Woo, P.C., Ng, K.H., Lau, S.K., Yip, K.T., Fung, A.M., Leung, K.W., Tam, D.M., Que, T.L. & Yuen, K.Y. 2003. Usefulness of the MicroSeq 500 16S Ribosomal DNA-based bacterial identification system for identification of clinically significant bacterial isolates with ambiguous biochemical profiles. *Journal of Clinical Microbiology*, 41(5):1996-2001.

Woo, P.C., Lau, S.K., Lin, A.W., Curreem, S.O., Fung, A.M. & Yuen, K. 2007. Surgical site abscess caused by *Lactobacillus fermentum* identified by 16S Ribosomal RNA gene sequencing. *Diagnostic Microbiology and Infectious Disease*, 58(2):251-254.

Yoo, H.S. 2010. Infectious causes of reproductive disorders in cattle. *Journal of Reproduction and Development*, 56(S):S53-S60.

Chapter 3

Materials and Methods

Ethical approval for the study was obtained from the Stellenbosch University Animal Research Ethics Committee (SU-ACUM14-00029).

3.1 Study site and sampling

The ZF Mgcawu district in the Northern Cape Province and the West Coast region of the Western Cape Province of South Africa were selected for sampling, as these areas are characterized as prominent Dorper sheep production regions. The two regions are currently experiencing a high incidence of UB infections. The participating farms were selected based on the producer's willingness to make their animals as well as their records available for research purposes. Farm owner and relevant demographic information of the study participants are omitted to ensure anonymity and to keep details of participating farmers confidential.

Samples were collected on farms where UB outbreaks were confirmed prior to collection of field samples. The location and the number of animals sampled that were healthy, diseased or suspected of being diseased on each respective farm is indicated in Table 3.1.

Table 3.1. Location and number of farms that participated in the sampling collection during the overall study.

Province	Farm number	Rams sampled (n)	Healthy	Diseased	Suspected
Western Cape	Farm 1	20	4	16	0
	Farm 2	13	5	7	1
	Farm 3	10	3	4	3
Northern Cape	Farm 4	36	30	6	0
	Farm 5	34	14	14	6

Farms sampled in the Northern Cape region experienced mean annual rainfall of ≤ 200mm, while those from the Western Cape were from areas that had mean annual rainfall of ≤ 200mm and between 201–400mm. The different maximum temperatures experienced by the two farms sampled in the Northern Cape were >31^{o}C and 29.3-31^{o}C, whilst the maximum temperature experienced by farms in the Western Cape was 27-29.2^{o}C.

3.2 Collection of swab samples and ticks

The samples were collected from black head and white head Dorper sheep (Figure 3.3) due to the predominant incidence of UB in this breed (Kidanemariam, 2003; Ali, 2012;). Although UB is reported to affect both rams and ewes, the higher incidence of the disease in rams (Gummow & Staley, 2000)

motivated the use of only rams in this study. The total number of diseased and healthy rams sampled was not predetermined and was dependent on ram availability and the status of the outbreak of the disease in ram flocks at the time of sampling.

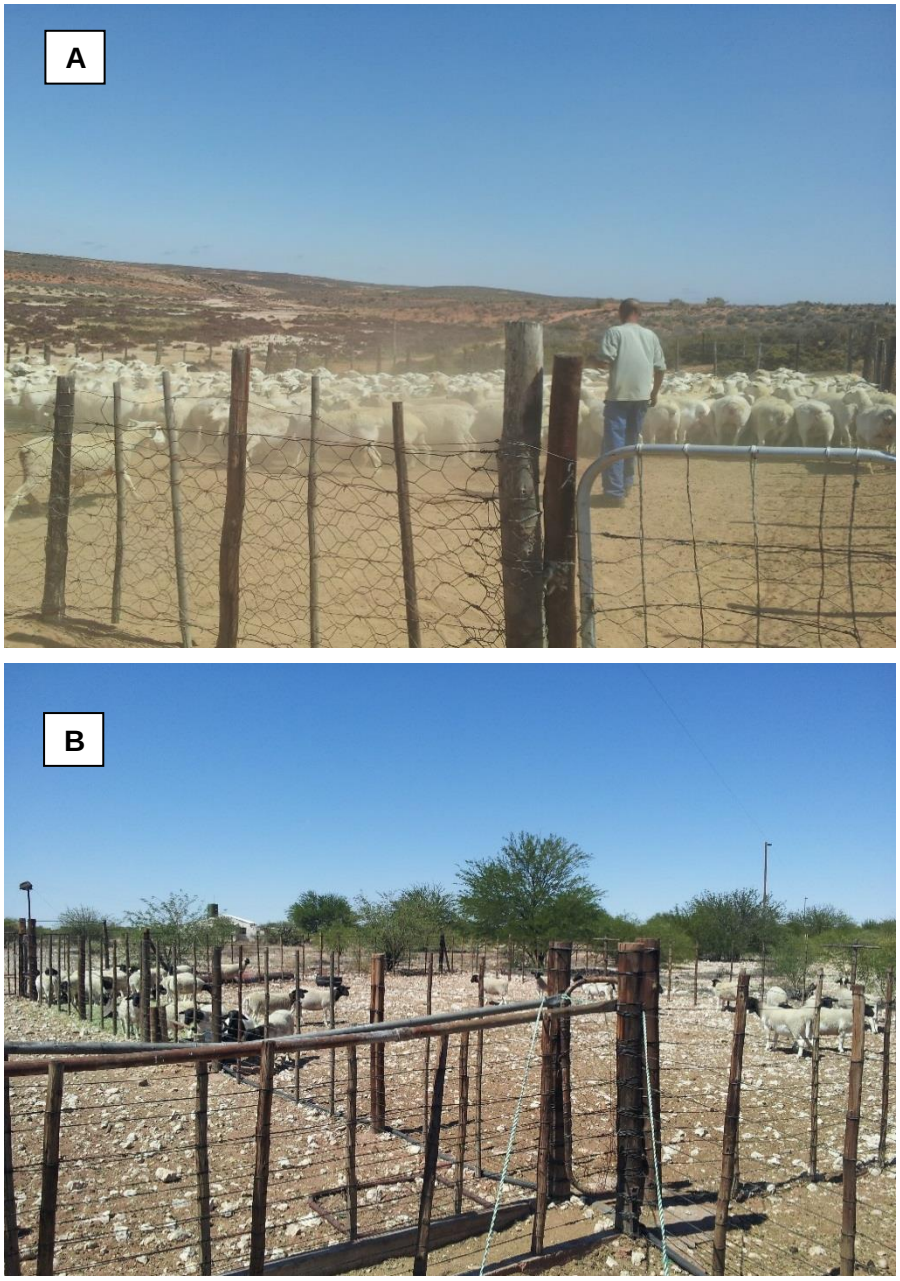

Figure 3.1. A flock of **A)** white head Dorper rams and **B)** black head Dorper rams sampled from farms in the Northern and Western Cape Province, respectively.

Swab samples were collected from a total of 113 rams; the origin of the samples was previously indicated in Table 3.1. Prior to sampling, the sheep were put in a pen, where all handling and sample collection procedures were performed. Rams were restrained individually, and each ram was classified as being healthy or diseased (based on visual assessment of typical symptoms) prior to the sample being collected.

Each swab sample was collected by using HydraFlock swabs (Puritan Diagnostic, Guilford, Maine, USA). Sampling was performed by gently extruding the penis from its sheath. The penis was then prevented from moving back into the sheath by holding it with a piece of sterile gauze without obstructing entrance into the sheath. The swab was then rolled over the entire surface of the penis and the inner membrane of the sheath for a minimum of 20 seconds. Two swab samples were collected from each ram. Immediately after contact with the penis and sheath surfaces, each swab was transferred to empty sterile Eppendorf tubes (Lasec, Ndabeni, Cape Town, South Africa) and placed on dry ice for the duration of sampling. All samples were stored at -20^{o}C under sterile conditions until further processing. To prevent cross-contamination of samples, all persons involved in sampling disinfected their hands with disinfectant between each sampling occasion.

Ticks surrounding the genital area of rams were collected using a sterilised pinsette. No limits were placed on the number and type of ticks collected. Collected ticks were placed in labelled, sterile tubes containing 70% isopropyl alcohol and stored at room temperature until later processing.

3.3 Data recorded

3.3.1 Clinical manifestations of ulcerative balanoposthitis

The classification system of the disease was adapted solely on the absence or presence of previously described clinical signs, which can range from acute to severe, in order to distinguish the health status of the rams sampled. Symptoms were previously described as occurring either on the mucosal membrane of the glans penis or the mucosal membrane of the glans penis and prepuce.

The health status of each animal was discerned after a thorough examination of the total surface area of the penis on extrusion from the sheath. Sampling criteria that included the respective clinical signs associated with the incidence of UB were recorded and are presented in Table 3.2. The presence of clinical symptoms was recorded using categorical variables “Yes”, “No” and “Suspected”. Animals exhibiting one or more clinical signs of UB (marked as “Yes” in Table 3.2) were classified as diseased and individuals displaying no visual symptoms (marked as “No” in Table 3.2) were categorised as healthy. In cases where disease was uncertain, rams were recorded as “Suspected” of being diseased.

Table 3.2. Descriptive characteristics associated with the clinical manifestations of ulcerative balanoposthitis in Dorper sheep.

Clinical sign	Presence of clinical sign
Ulcerations on the glans penis; shaft of the penis; prepuce	Yes/No/Suspected
Hyperaemia and inflammation of the penile mucosa	Yes/No/Suspected
Inability to extrude the penis from the sheath	Yes/No/Suspected
Inability to retract the penis into the sheath	Yes/No/Suspected
Bleeding	Yes/No/Suspected

3.3.2 Study population demographics

A total of 113 rams were examined for known symptoms of ulcerative balanoposthitis. A total of 55 rams were classified as healthy and 48 exhibited symptoms of UB and were thus classified as diseased. Of the 113 rams sampled, 10 rams were classified as being suspected due to the lesions observed that were not typical of UB. The prevalence of UB in the total sheep flock sampled was 46.6%, which was consistent with that previously reported by Kidanemariam (2003) who indicated a prevalence of about 47% in sheep flocks also sampled around the Northern and Western Cape Provinces. The demographic information of the sample site, and number of animals sampled per site of the entire 113 ram population are given in Table S3.1.

Rams of all ages were sampled in this study and age was determined by using the number and appearance of the teeth, according to the system indicated by the South African National Department of Agriculture (Table 3.3; Food and Agriculture Organisation, 2000). Ulcerative balanoposthitis was observed in rams of all age groups. The highest percentage of all rams sampled (~43%) was from the “full mouth” age group. The largest percentage of diseased rams (~44%) also fell within the “full mouth” age group followed by the “2 teeth” age group which made up 27% of the total diseased population.

Table 3.3. Age determination of sheep according to teeth number and appearance (source: Food and Agriculture Organisation, 2000).

Teeth number	Age
0 tooth	< 1 year old
2 teeth	1-1.5 years old
4 teeth	2 years old
6 teeth	2.5 years old
Full mouth	> 3 years old

Twenty healthy and twenty diseased rams from a total of fifty-five healthy and forty-eight diseased rams sampled were chosen for further analysis in this study. The demographics of the study subpopulation used is given in Table 3.4. Due to a larger number of black head Dorper sheep made available to this study, a higher percentage (65%) were sampled than white head Dorper (35%). The percentage of each type of sheep was however the same across healthy and diseased groups (Table 3.4). In each province (Northern and Western Cape) the percentage of healthy and diseased rams was the same (Table 3.4) but a higher total number of rams (26 out of 40) were from the Northern Cape Province as compared to the Western Cape Province (14 out of 40), as a result of higher availability of rams in this region.

Table 3.4. The demographic information relating to the sampling sites and animals sampled in this study.

		Healthy		Diseased			
		Number	%	Number	%	Total	%
Head colour	black	13	65	13	65	26	65
	white	7	35	7	35	14	35
Age (teeth number)	0 tooth	5	25	1	5	6	15
	2 teeth	2	10	5	25	7	17.5
	4 teeth	3	15	2	10	5	12.5
	6 teeth	5	25	2	10	7	17.5
	Full mouth	5	25	10	50	15	37.5
Province	Northern Province	13	65	13	65	26	65
	Western Province	7	35	7	35	14	35

Comparable with the whole population, grouping rams within the study population according to age based on their teeth number revealed that the highest percentage of rams sampled (37.5%) in this study population were from the age group “full mouth”. All age groups were affected by UB and the highest percentage of diseased rams (50%) was from the “full mouth” age group followed by the “2 teeth” age group which made up 25% of the total diseased rams. Albeit the small percentage differences in the number of diseased rams assigned to each age group in the study population compared to the entire

sampled population (Table S3.1), the study population of 20 healthy and 20 diseased rams was a good representation of the age demographic of the bigger population.

Statistical analysis to test the association between all ages (0 teeth – full mouth) and health status (diseased or healthy) was carried on the entire ram population (i.e. 113 rams) by means of a contingency table and chi-square statistics within the statistical program XLSTAT version 2016.05.33324. This study also grouped the ages from 0 teeth to 6 teeth into one variable, young, and the full mouth age group as a second variable, adult, in order to compare our results to that of a study by Kidanemariam (2003) that investigated the prevalence of UB in sheep flocks, and also grouped sheep into young or adult. The association of young and adult rams against health status (diseased or healthy) was also tested using a contingency table and chi-square statistics in XLSTAT.

3.4 Laboratory analyses

3.4.1 Microbial mock community as a control for analysis

Acquisition of a permit to import pathology specimens and raw materials for laboratory or pharmaceutical use into the Republic of South Africa was authorised by the Department of Agriculture, Forestry and Fisheries (permit No:13/1/1/30/2/0-2014/11/002080). The following reagent was obtained through the Biodefense and Emerging Infections (BEI) Research Resources Repository, NIAID, NIH, as part of the Human Microbiome Project: Genomic DNA from Microbial Mock Community B (Even, Low Concentration), v5.1L, for 16S RNA Gene Sequencing, HM-782D. The mixture consisted of genomic DNA from 20 known bacterial strains commonly found on and within the human body, containing equimolar ribosomal RNA operon counts of 100 000 copies per organism per µL.

The bacterial composition of microbial mock community B can be viewed online on the BEI website (www.beiresources.org). The use of a mock community is widely applied in 16S amplicon sequencing as a means to evaluate common errors such as chimera formation and substitution errors that may occur during polymerase chain reaction (PCR) and sequencing runs, respectively (Acinas *et al.,* 2005; Schloss *et al.,* 2011; Nelson *et al.,* 2014; Salipante *et al.,* 2014).

3.4.2 Nucleic acid-based tests

3.4.2.1 DNA extraction

The Eppendorf tubes containing the swab samples were thawed on ice and 400 µL of TE [1M Tris (pH 8.0), 0.5M Na_2EDTA (pH 8.0)] buffer was added to each sample, after which each sample was vortexed (Vortex Genie2; Lasec) at maximum speed for 5 min to suspend cells. Each suspended sample was then transferred to a new sterile Eppendorf tube, and 30 µL of lysozyme (10mg/ml; Sigma-Aldrich, St. Louis, MO, USA) was added to each tube followed by incubation in a water bath (Lasec) for 1 hour at 37°C. Each sample was then transferred to a bead beating tube (ZR Bashing bead Lysis Tubes (0.5

mm); Inqaba Biotechnical Industries (Pty) Ltd., Hatfield, Pretoria, South Africa) which was subsequently vortexed on maximum speed for 30 sec in order to mechanically disrupt hard to lyse bacterial cell walls. The mixture was transferred to a new sterile Eppendorf tube, and homogenised after the addition of: 70 µL of a 10% SDS (sodium lauryl sulfate) solution; 10 µL of proteinase K (20mg/ml; Biolabs, Ipswich, MA, USA) and 0.2 µL of RNAse A (10mg/mL, ThermoFisher Scientific, Germiston, Gaunteng, South Africa). Samples were subsequently incubated at 37°C for 30 min, and then for another hour at 55°C. Subsequent steps included protein precipitation with 100 µL of 5m NaCl; further cell lysis using a lysis buffer of 100 µL pre-warmed (65°C) CTAB solution (cetyltrimethylammonium bromide, 3% beta-mercaptoethanol); and incubation of the final mixture at 65°C for 30 min.

The DNA purification step was repeated twice and carried out by adding 700 µL of a chloroform/isoamyl mix (24:1) to the lysate and centrifuging at maximum speed at 4°C for 10 min. After each centrifugation step the top aqueous layer was removed and transferred to a new tube. DNA precipitation was then achieved by centrifuging the samples for 10 min at maximum speed at 4°C after the addition of an equal volume of isopropanol. The DNA was further purified by removing the supernatant and adding 300 µL of 70% ethanol to the remaining pellet and centrifuging at 4°C at maximum speed for 5min. The DNA pellet was air dried for 15 min and resuspended at room temperature in 50 µL of TE buffer. Each DNA sample was then tested for purity and concentration by means of gel electrophoresis on a 0.8% agarose gel stained with ethidium bromide and spectrophotometry using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific). The remaining resuspended DNA was then stored at – 20°C until further processed.

From the 113 samples, DNA from 20 healthy and 20 diseased samples, originating from 4 of the farms were chosen for further analysis. A balance between the number of samples and the sequencing depths required per sample was the deciding factor for the number of samples used to describe the penile microbiome of rams. The samples were chosen according to DNA quality and quantity to provide the best possible DNA for amplifying and sequencing.

3.4.2.2 The polymerase chain reaction (PCR) and pooling of amplicons

Two sets of 50 µL PCR mixtures were prepared for each of the 40 genomic DNA samples, using two separate bacterial primer pairs (Whitehead Scientific Integrated DNA Technologies, Stikland, Cape Town, South Africa) targeting different regions of the 16S rRNA gene. Each primer pair was synthesized with an adapter sequence specific to the MiSeq Illumina sequencer. Each forward primer was tagged with adapter sequence A and reverse primers with adapter sequence B as follows: Adapter A - 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG- 3' and adapter B - 5' -GTCTCGTGGGCTC GGAGATGTGTATAAGAGACA- 3'. An optimized forward primer, 27f-YM- 5' AGAGTTTGATYMT GGCTCAG- 3' (Y is C or T and M is A or C) and standard universal reverse primer, 534R- 5' –ATT ACCGCGGCTGCTGGCA- 3' previously described in Romero *et al.* (2014) were used to amplify the V1-V3 hypervariable region of the 16S gene. The forward and reverse primer U347F- 5' –GGA

GGCAGCAGTRRGGAAT- 3' (R is A or G) and U803R- 5' –CTACCRGGGTATCTAATCC- 3', respectively, were used as described from Nossa *et al.* (2010), amplifying the V3-V4 hypervariable region of the 16S gene. Primers with observed degeneracies have been optimized to reduce loss of amplification efficiency and specificity by allowing all variant forms of a nucleotide to be amplified.

Each 50 µL PCR reaction was prepared in a sterile PCR hood and run on a 2720 thermal cycler (ThermoFisher Scientific). The PCR mix consisted of 10 µL5X KAPA HiFi buffer, 1 µL of 1U/µL KAPA HiFi polymerase (KAPA Biosystems (Pty) Ltd., Cape Town, South Africa), 25.5 µL of PCR grade water (Celtic Molecular Diagnostic (Pty) Ltd., Wynberg, Cape Town, South Africa), 1.5 µL of 10 mM dNTP (ThermoFischer Scientific), 1 µL each of 20 µM forward and reverse primer and 10 µL of genomic DNA at a concentration of 10ng/µL. PCR amplification of the V1-V3 primer pair took place under the following cycling conditions: initial denaturation at 95°C for 3 min, 35 cycles at 98°C for 20 s (denaturation), 62°C for 15 s (annealing), 72°C for 30 s (extension) and a final extension cycle at 72°C for 1min. The V3-V4 PCR reactions were amplified using the same cycling conditions but with an annealing temperature of 57°C. Due to the high concentration of bacterial DNA per microliter of sample, DNA from the mock community was diluted at a ratio of 1:5 in PCR grade water. Ten microliter of the diluted DNA was used in PCR amplification following the same protocol as for genomic DNA samples. Negative controls without a template were included in each PCR run. If amplicons were observed in the negative controls, DNA samples and negative controls were rerun with a fresh batch of primer sets and PCR reagents.

Gel electrophoresis was carried out for each PCR reaction on a 1% agarose gel stained with ethidium bromide. Where low yields or no amplicons were generated, DNA samples were re-amplified. Amplicons of the correct size range for the V1V3 and V3V4 regions were excised from the gel and purified with the Zymoclean gel DNA recovery kit (Inqaba Biotechnical Industries (Pty) Ltd.) using the protocols recommended by the manufacturers. Purified DNA was eluted in 20 µL of PCR grade water (70°C) and quantified using Qubit 2.0 Fluorometer (Life Technologies, Johannesburg, Gauteng, South Africa) at the Central Analytical Facility (CAF) (Stellenbosch University, Stellenbosch). Equimolar concentrations of the two amplicons were pooled in a single tube for each sample.

3.5 Sequencing analysis

3.5.1 Sequencing procedure

All 41 samples containing equimolar concentrations of pooled V1V3 and V3V4 amplicons were sent to the Agricultural Research Council Biotechnology Platform (ARC BTP; Pretoria, South Africa; http://www.arc.agric.za/Pages/BTP.aspx) for library preparation and sequencing, which was carried out according to the standard Illumina 16S rRNA amplicon sequencing protocol (http://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html). In brief, the library preparation included the addition of dual indices and Illumina sequencing adapters to both ends of the amplicon sequences by means of PCR (as proposed by the Illumina amplicon

sequencing protocol) using the Nextera XT index kit. Dual indices (index 1 and 2) were used as a means to differentiate samples according to different combinations of 8 nucleotide bases attached to both ends of the amplicons. Adapter sequences (P5 and P7) were used as a means for the amplicons to adhere to the flow cell of the Illumina sequencer during sequencing. A unique combination of index 1 and 2 are given to each sample and each index is linked to an adapter sequence, with index 1 adjacent to the P7 adapter sequence and index 2 adjacent to the P5 adapter sequence. Equimolar concentrations of amplified amplicons containing the indices and adapter sequences were pooled together from each sample. The library prepared amplicons are then denatured and loaded onto the sequencing wells along with a PhiX spike-in of ~5% and paired-end sequenced using the 2x 300bp on the Illumina MiSeq sequencer. DNA sequences were retrieved from the ARC BTP as demultiplexed (i.e. separated according to sample) unidirectional raw forward and reverse reads ready for processing.

3.5.2 Data analyses

A summary of the steps used for data analysis are displayed in Figure 3.2.

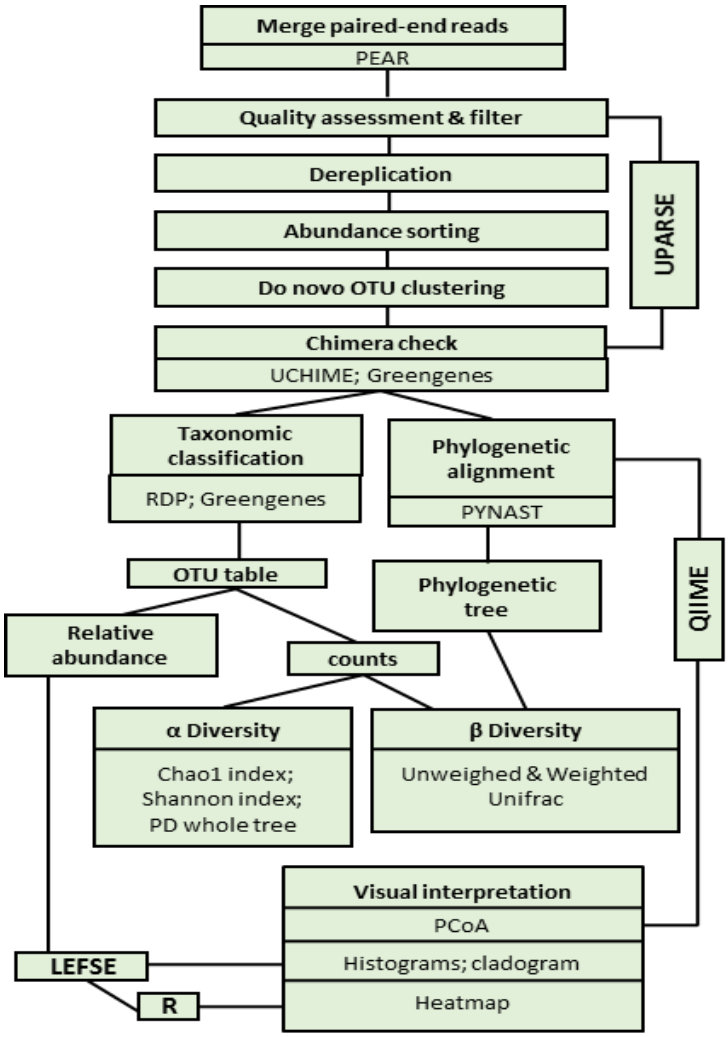

Figure 3.2. Bioinformatic workflow for data analysis of the penile microbiome.

2.5.2.1 Pre-processing and merging of forward and reverse reads

A total of 67.53 Gigabases (GB) of data was returned from the ARC BTP. Raw forward and reverse reads were analysed for quality using fastQC (Andrews, 2010) which calculates per base sequence quality in terms of phred scores, a common metric used to assess the quality of sequencing runs. The forward primer was used as a "barcode" to split the reads into V1V3 and V3V4 using FASTX_toolkit (http://hannonlab.cshl.edu/ fastx_toolkit/). Schloss *et al.* (2011) investigated the error rates of sequences (i.e. wrong nucleotide base calls) testing for different numbers of primer mismatches and found that 2 or lower mismatches drastically reduced the error rate of sequences. It was decided that in order to retain as many good quality reads as possible, a maximum primer mismatch of 1 would be allowed. Sequences with more than one mismatch were discarded. Primer sequences were removed from the 5' end of all forward and reverse reads, using the software Trimmomatic, version 0.33 (Bolger *et al.,* 2014).

As its ability to merge a higher number of read pairs comparing to other programs tested (including UPARSE) in this study was recognised the remaining paired forward and reverse reads were merged in Paired-End reAd merger (PEAR) version 0.9.6 32 bit (Zhang *et al.,* 2014). The minimum overlap allowed between paired reads for V1V3 and V3V4 were 30 and 50 nucleotides, respectively. Custom to this dataset, preliminary quality filtering steps in PEAR included trimming of bases with a quality Phred score of less than 25; discarding reads smaller than 200 base pairs (bp) after trimming; and discarding amplicons smaller than 350bp after merging.

3.5.2.2 Data analysis using UPARSE

V1V3 and V3V4 sequences were analysed separately throughout the remainder of the bioinformatics analyses. Further processing of 16S rRNA sequences was done following the pipeline set out by UPARSE (http://drive5.com/usearch/manual/uparse_pipeline.html), which is implemented in USEARCH version 8.1.1861 (Edgar, 2013). Parameters can be viewed on the UPARSE website on selection of scripts relating to the analysis at hand. Programs and algorithms used for the different steps were preloaded in UPARSE.

The initial step carried out in UPARSE involved additional quality filtering of the merged sequences and sequences that had a total expected error of more than 1 were discarded, as recommended by Edgar and Flyvberg (2015). In order for sequences to globally align more efficiently using the UPARSE alignment algorithms, sequences were truncated to a chosen length of 440bp and 394bp for V1V3 and V3V4, respectively, discarding sequences that did not meet this threshold. Sequences truncated at this length provided a fair representation of the bacterial mock community (the positive sample) whilst at the same time maintained an adequate balance between keeping as many reads as possible; as long as possible; and with the least number of total errors per base as possible.

To reduce the computational need, sequences were de-replicated so that only unique sequences remained. The unique sequences were then sorted by decreasing abundance. Singletons were removed, in line with literature which suggest that sequences arising from sequencing errors and PCR artefacts, which are observed occurring at low abundances, are reduced by this method (Huse *et al.,* 2010; Sibley *et al.,* 2011; Poretsky *et al.,* 2014).

To identify sequences that may be of biological relevance, similar sequences were grouped into clusters using the de novo clustering method employed in UPARSE, which is suggested to produce highly accurate operational taxonomic units (OTU) from 16S sequences compared to reference-based methods (Westcott & Schloss, 2015). The UPARSE approach consists of a greedy algorithm that follows a maximum parsimony model comparing each input sequence to the remaining sequence dataset, grouping sequences that are ≥ 97% similar to each other; starting with the most abundant sequences and working down to the least abundant (Edgar, 2013). The centroid sequence of the grouped clusters become the group representative and is termed the OTU; each OTU is assigned a number. A *de novo* chimera checking step is implemented during OTU clustering, which removes any sequences flagged as chimeric. Chimeric sequences can greatly increase the number of false positive sequences and inflate diversity estimates, it was thus decided to use a second chimera checking step that identifies chimeric sequences against a reference database (Haas *et al.,* 2011). The UCHIME-based algorithm (Edgar *et al.,* 2011) was used to align OTU reads to the Greenegenes reference database version gg_13_8 (http://qiime.org/ home_static/dataFiles.html; DeSantis *et al.,* 2006) and OTUs identified as chimeric were excluded from subsequent analysis. Each sequence that initially passed the quality filtering stages was mapped back to each OTU representative and assigned the OTU number of a representative if observed to be 97% similar to each other. Further processing included only non-chimeric OTUs.

3.5.2.3 Data analysis using QIIME

All taxonomic, phylogenetic and statistical analyses that follow in section 3.6.2.3.1 were conducted by following the pipeline set out by Quantitative Insights Into Microbial Ecology (QIIME) version 1.9.1 (Caporaso *et al.,* 2010), which is available for download at http://qiime.sourceforge.net. Standard python scripts and suggested default parameters (Kuczynski *et al.,* 2012) for all taxonomic and phylogenetic analyses were used unless otherwise stated. Parameters can be viewed on the QIIME website on selection of scripts relating to the analysis at hand (http://qiime.org/scripts/). Programs and algorithms used for the different steps were preloaded in QIIME.

3.5.2.3.1. Taxonomical and phylogenetic classification of operational taxonomic units

Non-chimeric OTUs were assigned taxonomic classification based on the 97% sequence identity, pre-clustered Greengenes reference database using the Ribosomal Database Project (RDP) classifier version 2.2 (Wang *et al.,* 2007; McDonald *et al.,* 2012). The RDP classifier uses a Bayesian approach and also runs a bootstrapping algorithm (Wang *et al.,* 2007). It was decided to use a confidence threshold value of 80% for assigning OTUs to a specific taxonomy. If a sequence could not be classified down to species level, the nearest taxonomical class was reported. Based on taxonomic identities the classified sequences were also grouped into phylotypes ranging from the phylum to genus level.

In order to compare these results from that of previous work on UB in South Africa, OTUs that were classified up to the genera *Mycoplasma*, *Arcanobacterium* and *Trueperella* level were further classified to the species level using the tool BLASTn against the GenBank reference database. Taxonomic assignment using BLASTn analysis was only accepted if an OTU sequence and the reference sequence shared ≥99% identity and did not observe the same identity percentage with another taxonomic assignment.

Phylogenetic trees are necessary inputs for various phylogenetic diversity metrics. In order to build a phylogenetic tree our OTU sequences were aligned to a reference Greengenes alignment template (≥85% sequence similarity) using PyNAST (Caporaso *et al.,* 2010), which was subsequently optimized by filtering all gaps and highly variable regions using the lanemask file available at http://greengenes.lbl.gov/Download /Sequence_Data/ (Kuczynski *et al.,* 2012). A Newick formatted phylogenetic tree was then built from the filtered sequence alignment using FastTree (Price *et al.,* 2010) and visualized in FigTree version 1.4.2 (http://tree.bio.ed.ac. uk/software/figtree/).

3.5.2.3.2 Filtering contaminants and spurious operational taxonomic units using the mock community

Mapped reads with an OTU number generated from UPARSE were used in conjunction with taxonomically classified OTUs to compile an OTU table containing information on per sample OTU presence, absolute abundance and taxonomic classification. On analysis of the V1V3 OTU table, the remaining number of OTUs and sequences per sample, after being processed through the pipelines, were significantly too few to infer diversity and were therefore omitted from further analysis.

A mock bacterial community was included in the sequencing as a final quality control measure. In order to filter what we expected to be spurious and contaminant OTUs, we decided to use an OTU filtering method that combines the ideas from Narrowe's *et al.* (2015) and Bokulich *et al.* (2013). OTUs identified in the V3V4 mock community sample were examined and the relative abundance of each species

present was calculated. Mock OTUs that could not be identified down to species level were further classified using the command-line BLASTn tool (Altschul *et al*., 1990; Camacho *et al*., 2009), which classified our mock OTU sequences using the manually curated GenBank reference database (Benson, 2013) compiled from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov).

A total of 99.98% of the reads in the mock community sample belonged to expected mock sequences, the remaining 0.02% constituted of 16S sequences from non-mock bacteria. The relative abundance of the most abundant non-mock species (n= 0.004%) found in the mock sample was used as a threshold by which to filter out spurious and contaminant OTUs from our OTU table. OTUs observed to have a sequence relative abundance lower than this threshold over all samples were removed and the resulting OTU table was used in subsequent diversity analyses.

3.5.2.3.3 Diversity measurements and statistical analysis

Before conducting diversity analysis, due to certain alpha and beta diversity measures being sensitive to differences in sampling efforts, the OTU table was rarefied (normalised) to 56347 sequences per sample (Soetaert & Heip, 1990; Magurran, 2013). This minimum sequence number per sample was chosen according to the sample containing the least number of sequences, and as a result no sample was omitted from the analyses. Estimations of community richness (Chao1 index), diversity (Shannon index) and phylogenetic diversity (PD-whole tree) were used for alpha diversity analysis by examining the bacterial communities of both healthy and diseased samples, separately (Faith, 1992; Chao & Bunge, 2002; Chao & Shen, 2003). The Chao1 index was further used to generate rarefaction curves describing the OTU richness of both bacterial communities. Non-parametric, two sample t-tests using 999 Monte Carlo permutations were performed for each alpha diversity metric testing for significant differences between the means of healthy and diseased groups.

Beta-diversity analyses to test community differences between healthy and diseased groups incorporated two phylogenetic beta-diversity metrics, weighted and unweighted UniFrac, which are well cited in literature as effective tools in community difference analysis (Lozupone, 2011). Similarity matrices of the communities, based on weighted and unweighted UniFrac distances were calculated and subsequently used by principle coordinates analysis (PCoA) to cluster samples on a 2-dimensional scale, in an effort to detect inter-group differences. The statistical test, permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001) was performed using the UniFrac matrices and testing the assumption of no difference amongst groups. Due to the non-parametric nature of PERMANOVA, statistical significance was determined through permutations (N=999). All tests for significance were two-sided and statistical significance was only considered when $p \leq 0.05$.

3.5.3 LEfSe and visual representation of penile microbiota

Multivariate diversity analysis allowed us to test for differences in microbiota composition when an animal is healthy and diseased but it cannot determine whether differences occurred between healthy and diseased rams at the OTU and phylotype levels. To determine OTUs and phylotypes that may distinguish penile microbiota specific to healthy and diseased groups, the linear discriminant analysis (LDA) effect size (LEfSe) method was used (Segata *et al.,* 2011) which can be accessed online at http://huttenhower.sph.harvard.edu/galaxy/ in the galaxy workflow framework. LEfSe makes use of a normalised relative abundance matrix to identify bacterial taxa at the OTU or phylotype level with significant differences in relative abundance between groups, using the non-parametric Kruskal-Wallis rank-sum test and performs LDA to estimate the effect size of each of the biological features flagged as significant. Healthy and diseased were designated as two classes (groups) with no subclass indicated. The LDA scores were computed using the logarithm (base 10) and default parameters including a 0.05 significance level and an effect size threshold of 2 were used for biomarker discovery.

To visualize relative abundances of OTUs flagged by LEfSe as statistically and biologically relevant, a heatmap was generated via the gplots (Warnes *et al.,* 2009), vegan (Dixon, 2003) and heatplus (Ploner, 2011) packages of the R statistical software (RDC Team, 2006).

References

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. & Polz, M.F. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and environmental microbiology*, *71*(12): 8966-8969.

Ali, H. 2012. Molecular characterization of *MycoplasmaMycoplasma* species isolated from the genital tract of Dorper sheep in South Africa. Master's thesis. Pretoria: University of Pretoria.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of molecular biology*, *215*(3):403-410.

Agricultural Geo-Referenced Information System. 2007a. Mean annual rainfall [Online]. Available: http://www.agis.agric.za/ [2016, September 14].

Agricultural Geo-Referenced Information System. 2007b. Maximum summer temperature [Online]. Available: http://www.agis.agric.za/ [2016, September 14].

Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32-46.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc [2016, June 10].

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. 2013. GenBank. N*ucleic Acids Research*, 41(D1):D36-D42, doi:10.1093/nar/gks1195.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A. & Caporaso, J.G. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10(1): 57-59.

Bolger, A.M., Lohse, M. & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114-2120, doi:10.1093/bioinformatics/btu170.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1):1.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I. & Huttley, G.A. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5):335-336.

Chao, A. & Bunge, J. 2002. Estimating the number of species in a stochastic abundance model. *Biometrics*, 58(3):531-539.

Chao, A. & Shen, T.J. 2003. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics*, 10(4):429-443.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7):5069-5072.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14(6):927-930.Edgar, R.C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10):996-998.

Edgar, R.C. & Flyvbjerg, H. 2015. Error filtering, pair assembly, and error correction for next generation sequencing reads. *Bioinformatics,* 31(21), 3476-3482, doi:10.1093/ bioinformatics/btv401.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16):2194-2200.

Faith, D.P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1):1-10.Food and Agriculture Organisation. 2000. Red Meat Marketing. Paper No.7 on Livestock Marketing [Online]. Available: http://www.nda.agric.za/docs/GenPub/7livestock.pdf [2016, September 14].

Gummow, B. & Staley, G. P. 2000. A retrospective survey of ulcerative balanoposthitis and vulvovaginitis in South African Dorper sheep. *Livestock Health and Production*, 12:31–36.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E. & Methé, B. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, 21(3):494-504.

Huse, S.M., Welch, D.M., Morrison, H.G. & Sogin, M.L. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, *12*(7):1889-1898.

Kidanemariam, A. 2003. Identification and characterization of the primary infectious agents associated with ovine ulcerative balanoposthitis and vulvovaginitis in South Africa. Master's thesis. Pretoria: University of Pretoria.

Kuczynski, J., Stombaugh, J., Walters, W.A., González, A., Caporaso, J.G. & Knight, R. 2012. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current Protocols in Microbiology*.1E.5, doi:10.1002/9780471729259.mc01e05s27.

Lozupone, C. & Knight, R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12):8228-8235.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J. & Knight, R. 2011. UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal,* 5(2):169.

Magurran, A.E. 2004. *Measuring Biological Diversity*. Oxford: Blackwell.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. & Hugenholtz, P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, 6(3):610-618.

Narrowe, A.B., Albuthi-Lantz, M., Smith, E.P., Bower, K.J., Roane, T.M., Vajda, A.M. & Miller, C.S. 2015. Perturbation and restoration of the fathead minnow gut microbiome after low-level triclosan exposure. *Microbiome*, 3(6):1.

Nelson, M.C., Morrison, H.G., Benjamino, J., Grim, S.L. & Graf, J. 2014. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PloS One*, *9*(4):e94249*,* doi:10.1371/journal.pone.0094249.

Nossa, C.W., Oberdorf, W.E., Yang, L., Aas, J.A., Paster, B.J., DeSantis, T.Z., Brodie, E.L., Malamud, D., Poles, M.A. & Pei, Z.. 2010. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World Journal Gastroenterology,* 16(33):4135-4144.

Ploner, A. 2011. Heatplus: Heatmaps with row and/or column covariates and colored clusters. R package version 2.1. 0. R Foundation for Statistical Computing, Vienna, Austria [Online]. Available: http://www.bioconductor.org/packages/2.10/bioc/html/Heatplus.html [2016, May 10].

Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D. & Konstantinidis, K.T. 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One*, *9*(4):e93827, doi:10.1371/journal.pone.0093827.

Price, M.N., Dehal, P.S. & Arkin, A.P. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS One*, 5(3):e9490, doi:10.1371/ journal.pone.0009490RDC Team. 2006 R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing [Online]. Available: http://www.R-project.org [2016, May 2].

Romero, R., Hassan, S.S., Gajer, P., Tarca, A.L., Fadrosh, D.W., Bieda, J., Chaemsaithong, P., Miranda, J., Chaiworapongsa, T. & Ravel, J. 2014. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome*, 2(18):1.

Salipante, S.J., Kawashima, T., Rosenthal, C., Hoogestraat, D.R., Cummings, L.A., Sengupta, D.J., Harkins, T.T., Cookson, B.T. & Hoffman, N.G. 2014. Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology*, *80*(24):7583-7591.

Schloss, P.D., Gevers, D. & Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS One*, 6(12):e27310, doi:10.1371/journal.pone.0027310.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. & Huttenhower, C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6):1.

Sibley, C.D., Grinwis, M.E., Field, T.R., Eshaghurshan, C.S., Faria, M.M., Dowd, S.E., Parkins, M.D., Rabin, H.R. & Surette, M.G. 2011. Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PloS One*, *6*(7):e22702, doi:10.1371/journal.pone.0022702.

Soetaert, K. & Heip, C.H.R. 1990. Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Marine Ecology Progress Series*, 59:305-307.

Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16):5261-5267.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S. & Schwartz, M. 2009. gplots: Various R programming tools for plotting data. R package version, 2(4):1.

Westcott, S.L. & Schloss, P.D. 2015. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ,* 3:e1487, doi 10.7717/peerj.1487.

Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30(5):614-620.

Chapter 4

Results & Discussion

4.1 Clinical observations of ulcerative balanoposthitis in Dorper rams

In this study the symptoms of ulcerative balanoposthitis (UB) observed for each ram sampled were used to classify the rams into three categories, i.e. healthy, diseased, or suspected. The severity of the symptoms observed varied between the diseased rams, which can potentially be ascribed to a difference in the onset and progression of the disease in the respective rams.

Figure 4.1 represents two examples of a healthy Dorper ram penis, free from any of the typical lesions that are observed when a ram is affected by UB. In contrast, Figure 4.2 shows typical inflammation and hyperaemia of the mucosal membrane of the penis that is observed when a ram is affected by UB. In this study, lesions associated with UB were confined to the mucosal membrane of the glans penis. The remaining penile and preputial mucosal membranes appeared to be normal. Symptoms ranged from ulcers that were confined to a small portion (Figures 4.3) of the glans penis to completely covering the glans penis (Figure 4.4). In few cases, bleeding and pus material was also observed in the affected area.

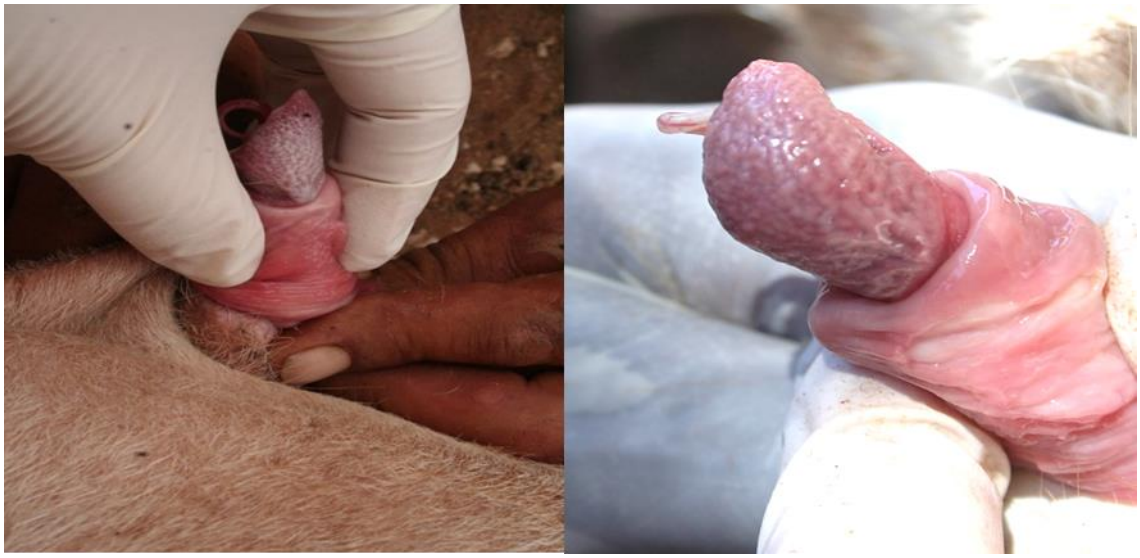

Figure 4.1. Normal penises of Dorper rams where no lesions associated with ulcerative balanoposthitis were observed.

Figure 4.2. Hyperaemia and inflammation of the preputial and penile mucosa as observed in Dorper rams infected with ulcerative balanoposthitis.

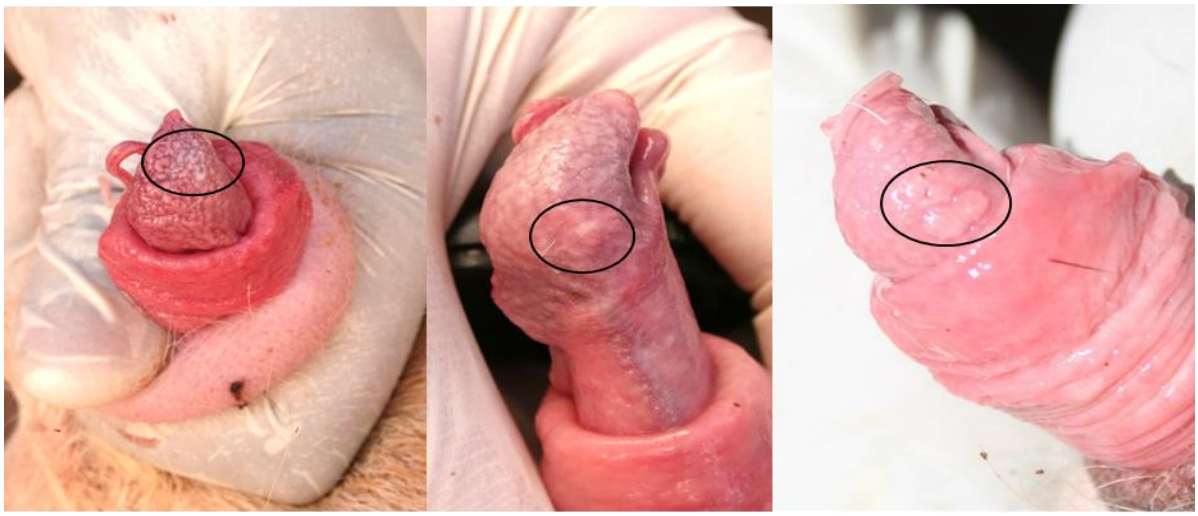

Figure 4.3. Ulcers confined to a small portion of the glans penis (as indicated by the circles) of dorper rams infected with ulcerative balanoposthitis.

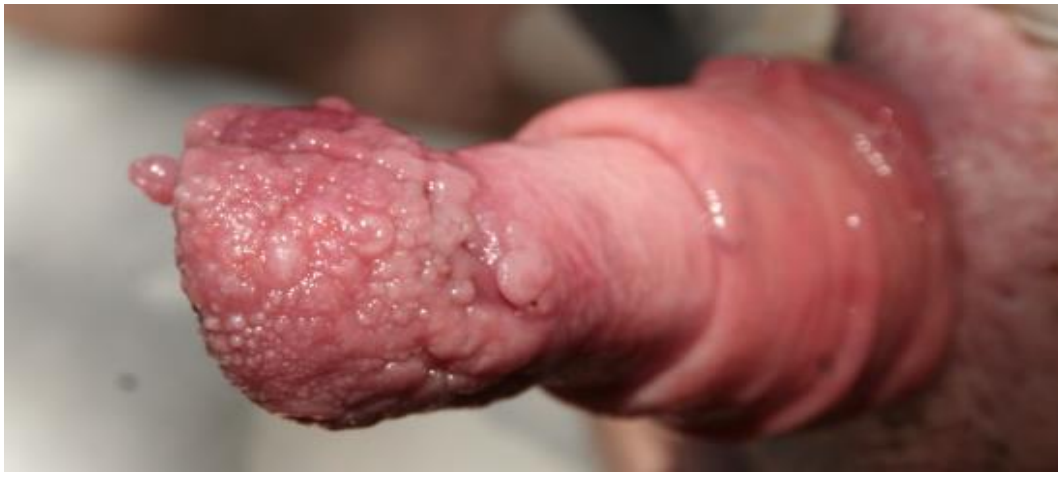

Figure 4.4. The distribution of ulcers on the entire surface of the glans penis in a Dorper ram infected with ulcerative balanoposthitis.

In some diseased rams where the disease was characterized as being in the early stages, blot clots were observed in the craters of the ulcers (Figure 4.5). The most severe cases of UB were characterized by the presence of ulcers that eventually were converted into scabs (Figure 4.6) or severe inflammation of the genital mucosal tissue preventing the penis from either retracting (Figure 4.7) or protruding from the sheath (Figure 4.8). On inspection of a number of rams, healthy and diseased, perforations on the glans penis were observed, which were indicative of a previous UB infection, where the ulcerations became so severe that the membrane became permanently damaged (Figure 4.9).

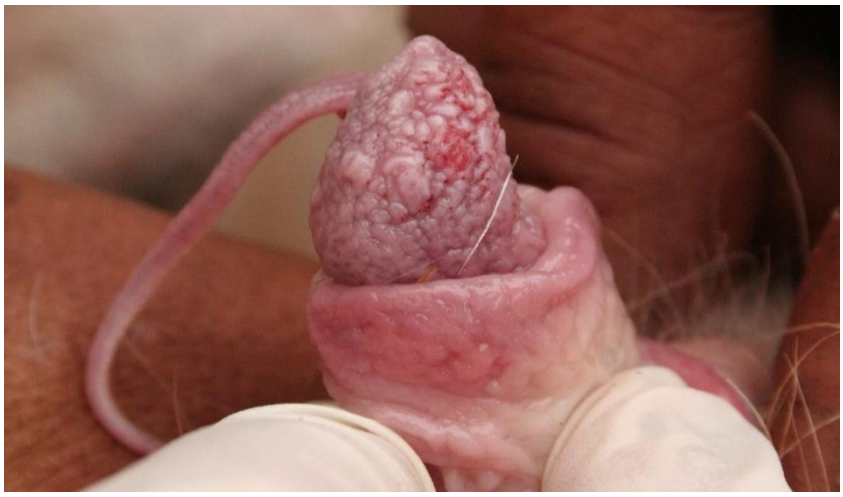

Figure 4.5. Ulcers filled with coagulated blood in what is presumed an early stage of ulcerative balanoposthitis, as observed in Dorper rams.

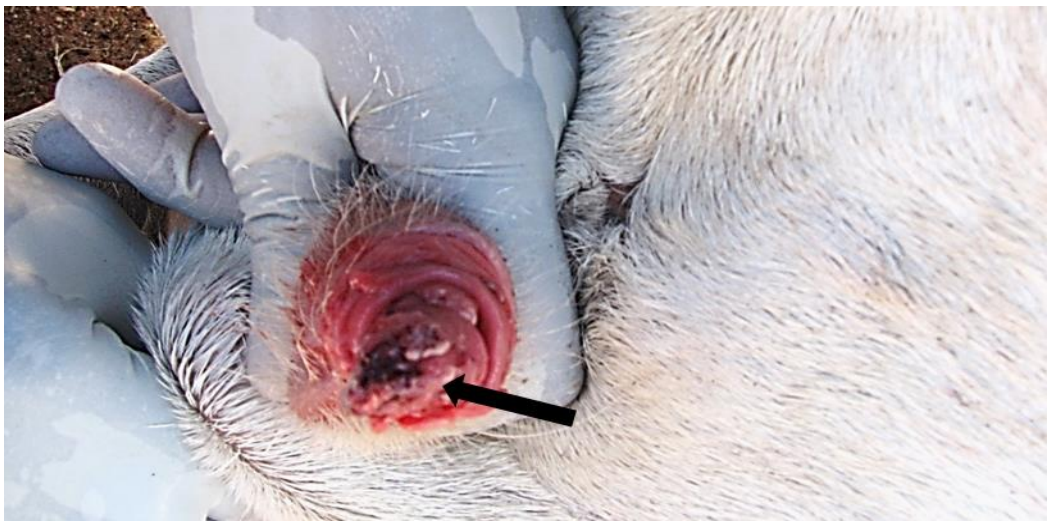

Figure 4.6. A case of permanent damage to the glans penis as a result of an ulcerative balanoposthitis infection in Dorper rams, as evident in the scabs present on the surface of the glans penis.

Figure 4.7. A progressive case of UB infection in Dorper rams, with an extensive swelling of the preputial mucosal membrane to the point that the penis could not be extruded from the sheath.

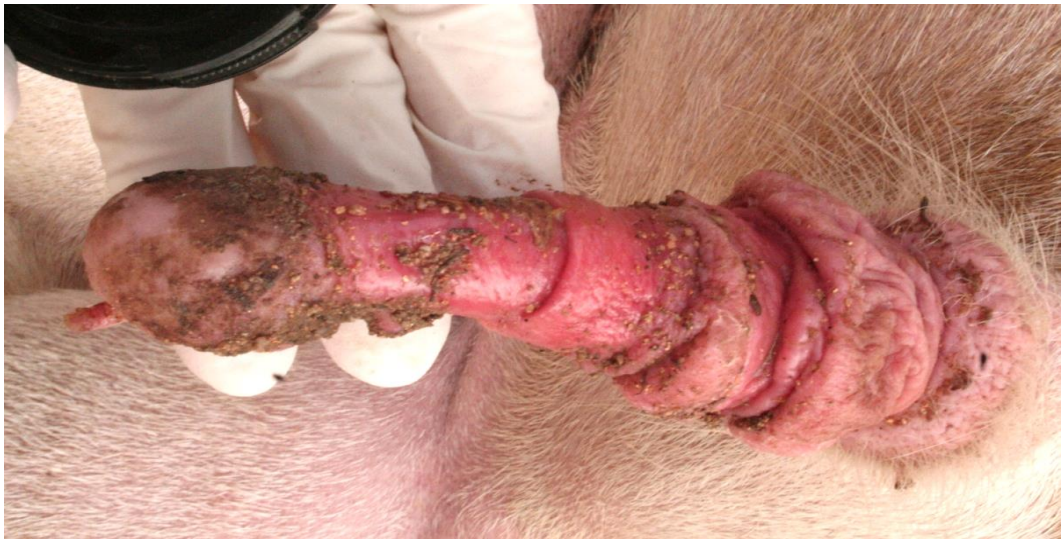

Figure 4.8. Severe inflammation of the penis of a Dorper ram infected with ulcerative balanoposthitis, preventing the penis from being retracted into the sheath.

Figure 4.9. Permanent tissue damage of the glans penis in Dorper rams as a result of progressive ulceration from previous UB infections.

Although it was suggested by other authors (Trichard *et al*., 1993; Gummow & Staley, 2000) that the clinical manifestations of UB could extend from the glans penis to the preputial mucosal membrane of the penis (Trichard *et al*., 1993; Gummow & Staley, 2000; El-Arabi *et al*., 2014; Robles *et al*., 2016), this study and an earlier study by Kidanemariam (2003) observed that the lesions were confined to the mucosal membrane of the glans penis. This study is thus in agreement with the suggestions made by Kidanemariam (2003) that posthitis is not observed as part of the disease syndrome and the disease in South Africa would most appropriately be described as ulcerative balanitis.

Although recording the severity of the disease did not form part of the metadata, on examining the genitals of rams, symptoms of differing severity were observed, which can be seen in Figure 4.2-4.8. It is believed that secondary bacterial infection may be responsible for the symptoms observed in the more progressive stages of the disease (Vuuren & Trichard, 2004) with the severity of the lesions influenced by the particular pathogenicity of the aetiological agent (Vuuren & Trichard, 2004). The symptoms observed in this study were in accordance with previous work on UB in South Africa (Trichard *et al*., 1993; Gummow & Staley, 2000; Kidanemariam, 2003) as well as findings from elsewhere (Webb & Chick, 1976; El-Arabi *et al*., 2014; Robles *et al*., 2016). Inflammation of the penis and deep ulcers on the glans penis were observed by all studies (Webb & Chick, 1976; Trichard *et al*., 1993; Gummow & Staley, 2000; Kidanemariam, 2003; El-Arabi *et al*., 2014; Robles *et al*., 2016). The presence of coagulated blood in the cavity of ulcers and pus material was supported by Bath and de Wet (2000), Kidanemariam (2003) and Robles *et al*. (2016). A few studies also reported the inability of rams to retract (Trichard *et al.,* 1993; Robles *et al.,* 2016) or extrude (Webb & Chick, 1976) their penises in more progressive stages. Ulcers that developed into scabbing in this study were only reported in South African cases of UB (Trichard *et al*., 1993; Kidanemariam, 2003). It seems that the development of UB was more severe in the latest cases of UB observed in other countries, where ulcers were also found to affect the preputial mucosa, and penises were covered with extensive blackened bloody ulcers (El-Arabi *et al*., 2014; Robles *et al*., 2016) of which 2 animals suffered and died from (El-Arabi *et al*., 2014). No animals died as a result of disease in this study or in any study previously reported in South Africa, and the ulcers did not appear to form a dark covering over the penis (glans penis and prepuce). It is likely that the disease observed in other countries is not as a result of the same aetiological agent than in South Africa.

Findings of this study indicated that a clear definition of all possible disease manifestations as well as the stage at which they occur is still lacking for the South African disease. Future work on the disease should aim to examine as many rams affected with UB as possible in an attempt to record and clinically describe all possible manifestations of the disease as it progresses. It is reported that histopathological analysis can be used to determine the progression of diseases (Nielsen *et al*., 2016). Histopathological analysis of the affected penile tissue may thus provide additional information on the progression of the disease. Together these methods can provide a more comprehensive picture of the symptoms observed.

4.2 Association of ram age and disease status

Rams were grouped according to age based on their teeth number (as described in Chapter 3 section 2.4.2), where less teeth was related to being younger. In order to test whether there was a significant association between age groups and health status (healthy or diseased), the entire population of rams sampled, 55 healthy and 48 diseased, was used for statistical analyses. Although some age groups observed higher numbers of diseased rams, testing the association between age groups (0 teeth – full mouth) and health status (healthy or diseased) by means of a contingency table and chi-square statistics revealed that no significant association ($p>0.05$) between the different age groups and health status was observed.

Grouping the rams in this study according to young (0 teeth – 6 teeth) and adult (full mouth) and testing whether younger rams were more likely to be diseased compared to adult rams revealed that a larger number of young rams were diseased (27 vs 21) but this was not significantly different ($p>0.05$). A study by Kidanemariam (2003) also compared the number of young (0 teeth – 6 teeth) and adult sheep (full mouth) that were affected with UB, and contrary to this study, found that the number of young sheep (72/104) affected with UB was significantly higher compared to adult sheep (32/104). He went further to suggest that young sheep were 2.5 times more likely to be affected with UB. It is possible that our smaller sample size of 48 versus 104, influenced the non-significant results observed in this study.

The reason why younger rams seem to be more susceptible to UB, as observed in this study, although not statistically significant, and by Kidanemariam (2003), is still unclear. One explanation offered by Gummow and Staley (2000) referred to a higher rate of sexual activity of younger rams, where vigorous mating activities can result in abrasion of the mucosal surface of the glans penis, thereby increasing the susceptibility of rams to bacterial infection. Natural mating practices, whereby a few rams run in the field with a larger group of ewes and with the rams that are responsible for servicing all ewes multiple times throughout the mating season, is common practice amongst the farms where we sampled. It is assumed that younger experienced rams have a higher mating fitness and dexterity than older rams (Toe *et al*., 1994) which under natural mating conditions can mean that they cover more services than older rams and are thus more likely susceptible to UB.

Another potential explanation suggests that older rams previously exposed to the disease may develop a certain level of immunity to re-infection (Kidanemariam, 2003). Van Vuuren and Trichard (2004) observed that on primary outbreak UB can spread to 100% of sheep and a lower incidence of about 50% occurs in outbreaks thereafter. This immune resistance however, does not provide complete immunity. According to the sampling site records and observations of Bath & de Wet (2000), rams that were previously infected with UB and recovered, were prone to become infected again if conditions were conducive. This suggests that future work on UB should aim to study the immunopathology of sheep (young and adult) directed at understanding the immune responses that occur as a result of

disease, adding to the collective information on the disease that could lead to the discovery of an aetiological agent.

4.3 DNA extraction from penile swab samples

DNA was extracted from the penile swab samples from the 20 healthy and 20 diseased rams using a modified DNA extraction protocol that included enzymatic and bead-beating lysis steps (as described in Chapter 3 section 3.5.2.1).

It is known that different cell lysis methods differ in their abilities to lyse bacterial taxa. Bacteria differ in their cell wall structure and integrity, and thus extraction protocols can be biased for specific bacterial taxa (Wesolowska-Andersen *et al*., 2014), resulting in different representations of bacterial composition (Yuan *et al*., 2012; Henderson *et al*., 2013; Wesolowska-Andersen *et al*., 2014; Wen *et al*., 2016). In an attempt to characterize the most representative penile bacterial microbiome the lysis steps were included to maximize DNA extraction from both Gram-negative bacteria and the harder to lyse Gram-positive bacteria (Mahalanabis *et al*., 2009). The combination of lysis steps included in this study has been described to produce significantly better DNA yields and representation of the bacterial communities than those that do not include both of these steps (Yuan *et al*., 2012). A DNA extraction protocol that combines the use of enzymatic and mechanical cell lysis methods has been described in a 16S amplicon study that characterized the penile microbiome of pre-circumcised and post-circumcised men (Price *et al*., 2010).

The purity and concentration of the extracted DNA was tested using agarose gel electrophoresis and spectrophotometry. DNA was successfully extracted from all penile swab samples with the DNA concentrations per sample ranging from 82.6 ng/µL to 1389 ng/µL with a mean and standard deviation (SD) of 449 ±.390 ng/µL. Only samples that extracted DNA with a minimum 260:280 purity ratio of greater than 1.5 were selected for further analysis.

4.4 Amplification of bacterial 16S V1V3 and V3V4 hypervariable regions

In this study, the V1-V3 and V3-V4 hypervariable regions of the bacterial 16S rRNA genes were chosen to characterize the penile bacterial diversity of Dorper rams. It is thought that different hypervariable regions may be biased towards certain bacterial taxa, producing different bacterial community profiles when singly used (Chakravorty *et al*., 2007; Kumar *et al*., 2011; Guo *et al*., 2013). Many microbiome studies therefore sequence multiple hypervariable regions. Nelson *et al*. (2010) for example characterised the penile microbiome of men by sequencing V1–V3, V3–V5 and V6–V9 variable regions of the 16S rRNA gene and was able to recover a wide array of bacterial diversity, with 306 genera. The use of two hypervariable regions was thus preferred in this study in order to characterize the bacterial community composition more accurately.

The V1-V3 and V3-V4 regions were amplified from the extracted DNA as described in Chapter 3, section 3.5.2.2. Amplicons were analysed by agarose gel electrophoresis. The expected amplicon size for the V1V3 and V3V4 PCRs were approximately 550bp and 460bp, respectively. Amplicons of the expected sizes were observed following agarose gel electrophoresis of the PCRs for all DNA samples. The gel electrophoresis results observed for a representative subset of samples are illustrated in Figure 4.10. Lanes 7-11 and 17-22 in Figure 4.10 show a single band of the expected size representing the bacterial 16S V1V3 regions. The remaining lanes 2-6 and 14-16 displayed single bands of the expected size for the bacterial 16S V3V4 regions. A negative control can be seen in lane 12 and was used as a means to identify any potential contaminants present in the PCR reagents. The smaller band (<100bp) seen in the negative control in Figure 4.10 was as a result of primers binding to each other through small strings of complementary bases and forming primer dimers and not as a result of DNA contamination. Although the DNA extraction protocol was optimized numerous times to produce the most sensitive and specific PCR reactions, we found that the chosen V1V3 and V3V4 primers pairs were not very specific to the genomic region they flanked and non-specific amplification was observed, as seen by the two faint single bands of unexpected band sizes (higher than 1000bp) for all V1V3 and V3V4 PCRs in Figure 4.10. To overcome this drawback, DNA from bands of the correct sizes were extracted from the gel and used in further analysis.

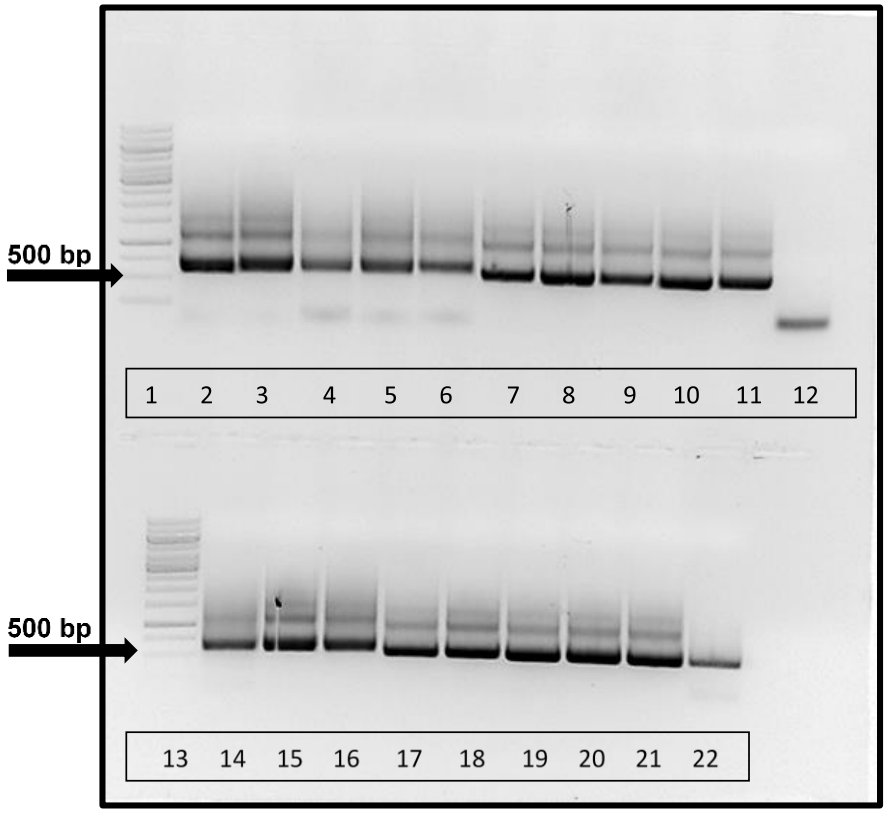

Figure 4.10. PCR amplification products viewed on a 1% agarose gel. Lane 1 and 13 are the DNA ladders. Lanes 2-6 and 14-16 represent amplicons amplified using the V1V3 primer set. Lanes 7-11 and 17-22 show amplicons of a different size, amplified using the V3V4 primer set. Lane 12 is the negative control. Arrows point to DNA fragments in the DNA ladder that are 500bp long.

4.5 Sequence analysis

4.5.1 Per sample sequence results

To characterise the bacterial communities in the penile samples of healthy and diseased rams, paired-end Illumina sequencing of the 16S V1V3 and V3V4 amplicons was performed. A total of ~113 million combined raw forward and reverse reads were sequenced with a per sample mean of 2 745 003 sequences, with a SD of 1 345 810 (Figure 4.11).

The number of reads per sample ranged from 120 084 to 5 392 476. The uneven number of reads observed across all samples was not expected, as equal concentrations of amplicons belonging to each sample were pooled together during library preparation before being sequenced. It is likely that a number of technical reasons rather than biological reasons are responsible for these differences such as inaccurate measurements of DNA concentration prior to pooling samples together.

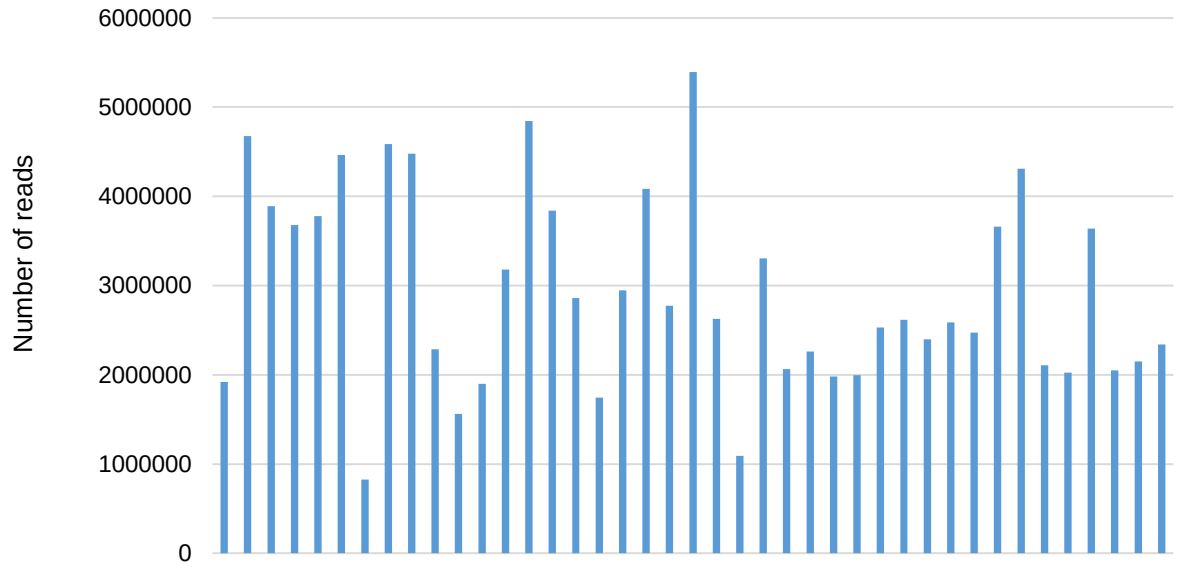

Figure 4.11. Bar chart of the combined number of forward and reverse reads per sample. Each bar on the x-axis represents an individual sample.

Additionally, samples prepared for sequencing consisted of pooled equimolar concentrations of V1V3 and V3V4 amplicons. On sequencing the amplicons an approximately equal number of V1V3 and V3V4 reads (forward and reverse reads) was also expected. Separating the samples according to V1V3 and V3V4 reads, discarding reads with more than 1 primer mismatch, revealed that the number of V3V4 read pairs sequenced were approximately three-fold more than the number of V1V3 read pairs, with an average number of read pairs of 415 071 (SD of 203 524) and 1 253 812 (SD of 505 033) per sample for V1V3 and V3V4, respectively. We can only assume that impurities or unknown compounds (although purity ratios were > 1.5) were derived from DNA gel extraction and purification may have biased the ratio by disturbing the accurate measurement of the concentration of DNA, along with minor pipetting

errors (Farkas, 1993). In future studies, more accurate DNA concentration estimation methods, such as qPCR or ddPCR (Robin *et al*., 2016) could be used to prevent such large disparities in numbers.

4.5.2 Sequence quality control

In order to see how much of the sequencing data could be used to accurately characterize the penile microbiome, the quality of the V1V3 and V3V4 forward and reverse reads was assessed with fastQC (Andrews, 2010). The per nucleotide base quality phred score which represents the probability that a base was incorrectly called, was calculated for all reads. The higher the phred score the higher the chance that a base was correctly called. The per base inter-quartile range in phred score for the forward and reverse V1V3 and V3V4 reads calculated from all samples is displayed in Figure 4.12. The different colours observed in the background of Figure 4.12 divide the phred quality scores on the y-axis into ranges which depict base calls that are of very good quality (green), reasonable quality (orange), and base calls of poor quality (red).

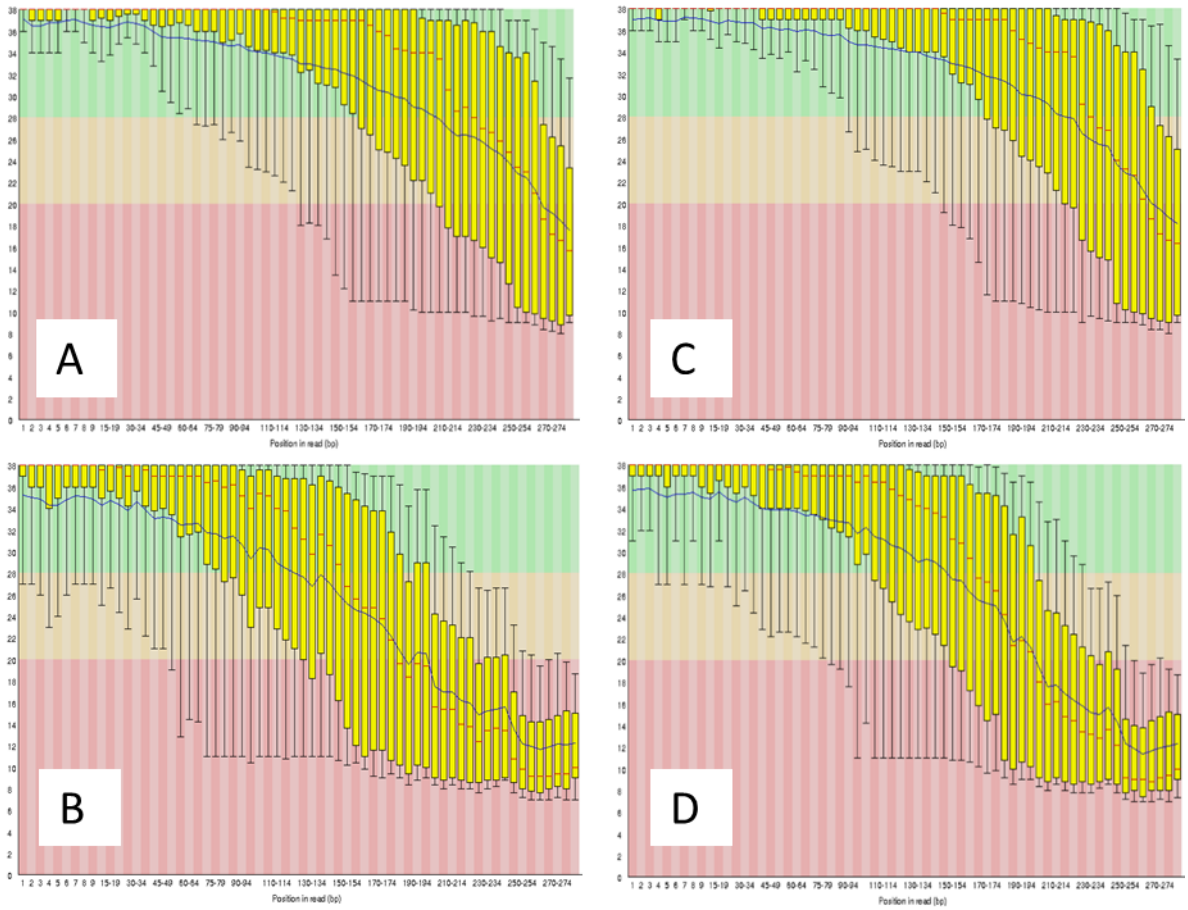

Figure 4.12. FastQC report comparing the phred quality scores across all bases for the V1V3 A) forward and B) reverse reads and V3V4 C) forward and D) reverse reads. The vertical axis represents the Phred quality scores. The red line is the median value at each base and the blue line represents the mean quality. The box and whiskers plot at each nucleotide position represents the inter-quartile range, with the upper and lower whiskers representing the first and last inter-quartile range and the yellow box representing the inter-quartile range from 25% to 75%.

The results in Figure 4.12 revealed that the per base quality ranged across all V1V3 and V3V4 reads, with a prominent loss in quality towards the 3' end of all reads and poorer quality observed in the reverse reads, as typically observed in Illumina amplicon sequencing data (Edgar & Flyvberg, 2015; Schirmer,*et al.*, 2015). The V1V3 reads had poorer per base quality than V3V4 reads, especially in the reverse read where the median base quality score dropped below phred score 25 at about 158bp, as compared to 178bp in the V3V4 read.

Although the Miseq system is a powerful sequencing platform for high-throughput 16S amplicon sequencing, it has known limitations related to the technology which affects the quality of the sequencing data generated. Low sequence diversity in the first several bases sequenced are known to compromise base calling and sequencing on the Illumina platform (Tremblay *et al*., 2015). This problem is encountered in samples that are dominated by one or very few types of bacteria (Fadrosh *et al*., 2014). Amplicon sequences that contain regions of extreme G+C content and long stretches of G's or C's are also thought to produce sequence regions of lower quality (Tremblay *et al*., 2015). It is possible that the low quality observed for the V1V3 and V3V4 reads are as a result of these limitations, with sequences produced from the V1V3 amplicon worst affected. Although current sequencing methods have managed to reduce this effect through the addition of PhiX DNA spike-in to increase diversity and improve sequence quality (Caporaso *et al*., 2011), some 16S rRNA amplicon libraries require the addition of different percentages of PhiX of up to 50% (Fadrosh *et al*., 2014). In this study the Illumina recommended minimum 5% PhiX spike-in addition was used. Increasing this may have facilitated improved data quality in the sequencing run.

The reduced quality of sequences towards the 3' end of reads as well as the lower quality observed for the reverse reads have further been assumed to be due to accumulation of phasing and pre-phasing events throughout the sequencing process (Edgar & Flyvberg, 2015; Schirmer *et al*., 2015). Some sequencing protocols have further been optimized to reduce this effect (Fadrosh *et al*., 2014; Tremblay *et al*., 2015), by adding random nucleotide bases to primers (heterogeneity spacer) used during PCR amplification in the library preparation steps, thus producing reads with nucleotide diversity.

Sequencing data with bases of suboptimal quality can result in false interpretation of bacterial communities (Bokulich *et al.,* 2013) and upon looking at the V1V3 and V3V4 read quality it was decided that stringent quality control steps were required for accurate characterisation of the penile microbiome. Quality filtering of the reads was initially carried out during read merging. Bases with a phred score less than 25 were trimmed and, reads and merged sequences that fell below a size threshold were (i.e. uninformative) discarded. Merging of the reads obtained sequences with an average length of ~500bp and ~430bp for the V1V3 and V3V4 hypervariable regions, respectively. Additional discarding of merged sequences that did not pass a quality threshold was carried out as described in Chapter 3, section 3.6.2.2, followed by truncating V1V3 and V3V4 sequences to 440bp and 394bp, respectively. Reads below these sizes were discarded.

The per base quality profile of the V1V3 and V3V4 merged sequences can be seen in Figure 4.13A and 4.13B, respectively. The red line depicts per base median quality score and the blue line the per base mean quality score. The box and whiskers plot demonstrates the inter-quartile phred score range at each nucleotide position, with the upper and lower whiskers representing the first and last inter-quartile ranges and the yellow box representing the inter-quartile range from 25% to 75%. Quality filtering retained merged sequences of good quality, with the median quality above phred score 28 at each nucleotide position for both hypervariable regions (Figure 4.13). The V3V4 merged sequences performed better in terms of per base inter-quartile range in phred score (i.e. smaller range) compared to the V1V3 merged sequences due to the higher quality of V3V4 forward and reverse reads. In addition, sequence overlaps produced during read merging increase the quality and confidence for a given base call in the overlapping region (Zhou *et al*., 2011). Merging V3V4 forward and reverse reads produced a sequence with a larger overlap than the V1V3 paired-reads (i.e. ~140bp compared to ~50bp), resulting in a higher base call confidence and quality for a higher number of bases for V3V4 sequences.

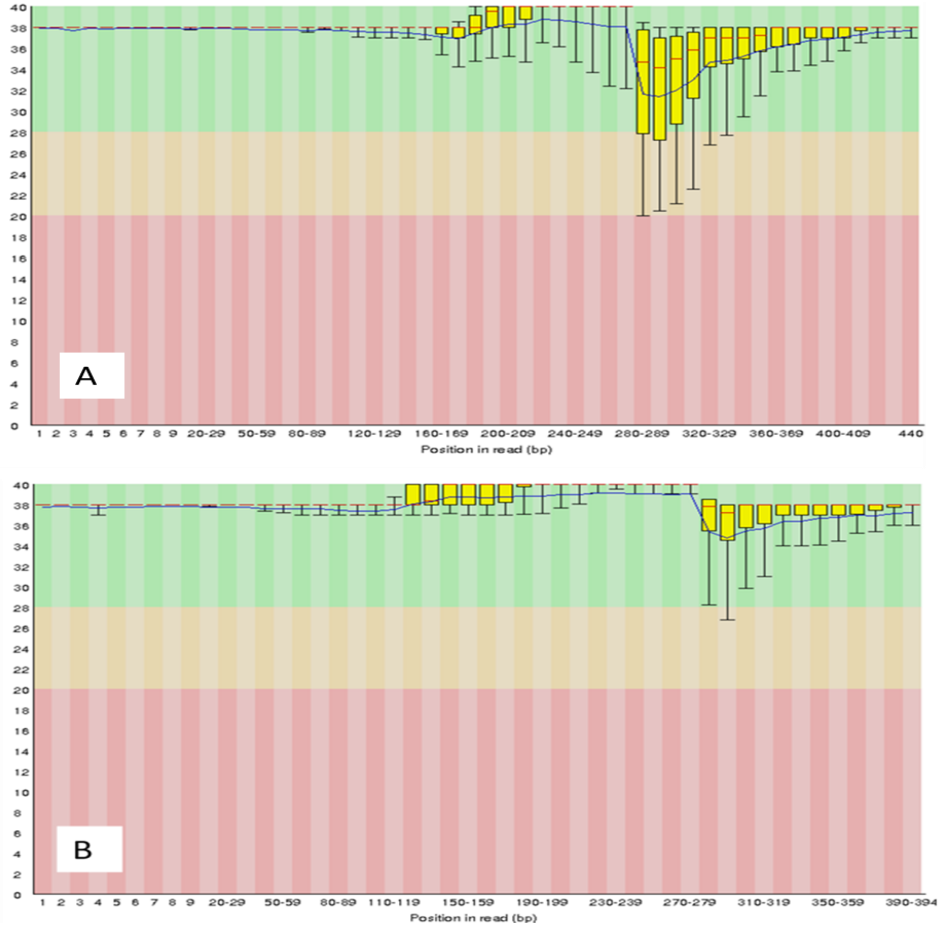

Figure 4.13. Per base sequence quality after quality filtering and truncating the V1V3 and V3V4 merged sequences to 440bp and 394bp, respectively. **A)** Sequences from the V1V3 hypervariable regions **B)** sequences from the V3V4 hypervariable regions**.** The different colours observed in the background divide the phred quality scores on the y-axis into ranges, which depict base calls that are of very good quality (green), reasonable quality (orange), and of poor quality (red).

Comparing the number of paired reads with the number of sequences that remained approximately 13% and 34% of the V1V3 and V3V4 sequence data, respectively, was determined to be of good quality. A study by Sinclair *et al*. (2015) that also sequenced the V3V4 regions of the 16S rRNA gene using Illumina technology encountered similar results. They found that from the initial number of read-pairs produced by sequencing only about 50% were retained after a number of pre-processing steps and quality filtering. Additional studies that conducting sequencing on the Illumina platform showed that typically between 40-85% of raw reads are discarded through quality filtering (Gloor *et al.*, 2010; Caporaso *et al.*, 2011; Degnan & Ochman, 2012). Although these studies made use of older Illumina sequencing chemistries, the problem of amplicon sequencing quality still persists in the newer chemistries. The V3V4 sequence data performed within the previously reported ranges, however the V1V3 performed worse than previously reported.

Lowly abundant sequences are likely as a result of error, such as sequencing errors and PCR artefacts (Huse *et al*., 2010; Sibley *et al*., 2011; Poretsky *et al*., 2014) and an additional quality control step included the removal of sequences that occurred only once (Singletons). We found that of the unique sequences in the dataset more than 80% of the V1V3 and V3V4 sequences were singletons. Schloss *et al*. (2011) described that the number of erroneous sequences, such as singletons, increases with sequencing depth. This study included deep-sequencing with millions of reads produced per sample and could be used as one of the explanations for the high proportion of singletons observed in this study. Subsequently, PCR chimera sequences that were not filtered out as singletons were discarded from the dataset by a *de novo* method during OTU clustering and thereafter using UCHIME-based algorithm (Edgar *et al.,* 2013). Singletons and PCR chimeras can greatly inflate the diversity of a community during OTU clustering by acting as separate OTUs which are displayed as new “species”. By removing singletons and chimeric sequences in this study we were able to reduce the number of erroneous sequences, providing higher confidence of the described penile community diversity.

4.5.3 Mock community analysis

A bacterial mock community (HM-278D) composed of 20 bacterial strains, with equimolar rRNA copies per strain (100 000 copies of each bacteria per µL), representing 17 genera was included in this study as a positive control for the 16S rRNA PCR and sequencing. The use of bacterial mock communities to examine and evaluate things like performance and error characteristics of DNA processing protocols, sequencing technology and bioinformatics pipelines on bacterial community composition was first suggested by the human microbiome project (Jumpstart Consortium Human Microbiome Project Data Generation Working Group, 2012) and has become an essential component of microbiome research with many recent studies including them in their research (Diaz *et al*., 2012; May *et al*., 2014; Lluch *et al*., 2015; Zheng *et al*., 2015).

Sequence processing through the pipeline included the clustering of sequence reads with >97% identity into groups termed operational taxonomic units (OTUs). The UPARSE pipeline employed in this study is believed to generate OTUs that are highly accurate compared to other pipelines and was thus chosen as the method for OTU clustering (Edgar, 2013). The OTU based approach is standardly used in microbiome analysis where sequences that group into an OTU are phylogenetically close and therefore presumed to represent an ecological bacterial subpopulation, in most cases strain or species. A total of 79 OTUs were identified in the bacterial mock community. Assuming each of the 20 bacterial strains in the mock community is represented by a single OTU roughly 20 OTUs were then expected. The deviation from the expected OTU number has also been observed in a study by Bokulich *et al*. (2013). Of these 79 OTUs, 47 OTUs were classified as mock species representing 98% of the total abundance in the sample. The remaining 32 OTUs were observed at very low abundances and were not classified as known mock species. These contaminant OTUs were also observed in the ram penile microbiota. It was assumed that these OTUs may have derived as a result of cross-contamination between samples during the 16S PCR's and the library preparation step prior to sequencing. Cross contamination of OTUs between samples has been described as a challenge of multiplexed (multiple samples sequenced in parallel) Illumina sequencing (Sloan *et al*., 2013).

Only 16 mock genera were identified in the mock sample, *Staphylococcus* could not be identified further than the class *Bacilli*. A total of 7 OTUs were correctly identified to the species level using the chosen bioinformatics pipeline. Thirteen percent of the OTUs were incorrectly classified to unexpected species belonging to the mock genera *Listeria, Acinobacter* and *Bacillus.* Correct classification of these OTUs, along with OTUs that could not be classified up to species level, was performed using the BLASTn tool (Altschul *et al*., 1990; Camacho *et al*., 2009) and the GenBank nucleotide reference database (Benson *et al*., 2013). As indicated by the use of a second reference database (GenBank), the reference sequences available within databases play an important role in the classification of 16S rRNA sequences (Werner *et al*., 2011). On analysis of the mock data the Greengenes reference database (DeSantis *et al.,* 2006) used in this study proved to have some limitations in classifying the short (394bp) 16S rRNA sequences correctly and down to the species level at the chosen confidence threshold (≥80% bootstrap confidence). It is also important to mention that classifiers (i.e. Greenegenes and NCBI) use different taxonomies and also vary in the methods they use for reporting confidence in the prediction, which cannot be directly compared but can be used to explain the different results observed when using the chosen pipeline in this study and the BLASTn classifier tool. In conclusion the V3V4 primer pair selected for this study had a good coverage of the mock bacteria species, but had several drawbacks in terms of correct classification and species level classification which were hindered by the tools and reference database used to classify sequences.

Sequencing the mock community sample is expected to ideally yield an approximately equal relative abundance of ~5% (100%/20) for each of the 20 mock bacterial species. A comparison of the expected versus the observed proportion of mock bacterial genera in the sequence data is shown in Figure 4.14. A number of mock species namely, *Escherichia coli*, *Bacillus cereus*, *Streptococcus agalactiae*,

Lactobacillus gasseri, *Listeria monocytogenes* and *Rhodobacter sphaeroides* were observed at the expected relative abundances of ~5%. Lower and higher than expected relative abundances were however observed for the remaining bacteria, with *Bacteroides vulgatus* having the highest relative abundance (~14%) and *Propionibacterium acnes* the lowest (~0.5%). These findings suggest that some OTUs may be under- or overrepresented across the non-mock samples.

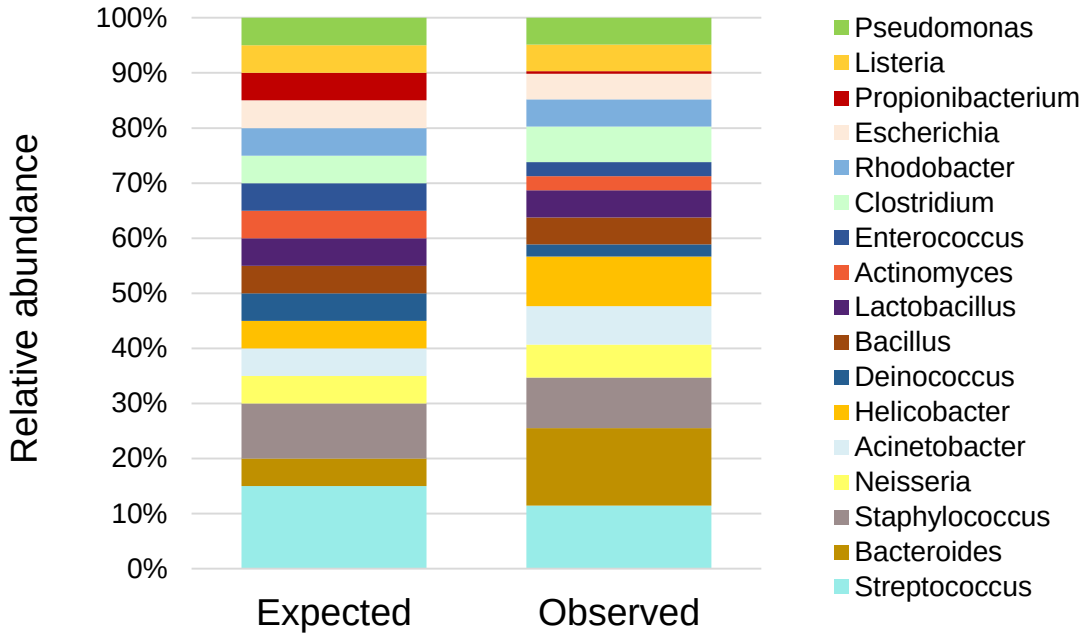

Figure 4. 14. Genus level comparison of the expected and observed percentage relative abundance of mock bacteria.

These results are in agreement with other studies (Diaz *et al*., 2010; Jumpstart Consortium Human Microbiome Project Data Generation Working Group, 2012; Hang *et al*., 2014; Fouhy *et al.,* 2016; Krohn *et al*., 2016) which have observed that the proportion of mock species vary from that of expected, indicating this as a common challenge of 16S rRNA gene sequencing. This bias has been described as arising from many sources including systematic errors experienced during sample preparation such as PCR and sequencing errors (Krohn *et al*., 2016; Singer *et al*., 2016), PCR amplification (Diaz *et al*., 2012; Ekblom *et al*., 2014), primer set design (Bergmann *et al*., 2011), sampling depths (Dubourg *et al*., 2013) and the reference database used (Brooks *et al*., 2015). Biological bias can also be observed in that some bacterial taxa have multiple copies of the 16S rRNA gene and can thus inflate their relative abundance within a sample (Vetrovsky & Baldrian, 2013).

From the examples suggested above we suspect that two of the most prominent sources of bias in relative abundance in the mock sample and dataset as a whole could be explained by the choice of primers and the quality distribution of reads. The efficiency by which primers anneal to target bacterial sequences varies, which will affect the apparent abundance of some bacterial taxa within the community causing some to be under- or over-estimated (Hamady & Knight, 2009; Klindworth *et al.,* 2012; Nelson

et al., 2012). Findings from the mock analysis may indicate a limitation in the chosen primer set, in that the V3V4 primers in this study may be responsible for the over-representation of *Bacteroides vulgatus* and under-representation of *Propionibacterium acnes*.

Methods used for filtering and analysing 16S sequence data assumes that the probability of having a number of errors is the same for all reads, regardless of where they originated from (Puente-Sánchez *et al*., 2016). This however does not take into account that sequences from different taxa may have different quality distributions and when quality filtering is applied removing sequences that do not pass quality thresholds, this can artificially enrich some taxa versus others, compromising confident quantitative taxa allocation and interpretation of results obtained by 16S rRNA high-throughput sequencing. The large proportion of sequences that were quality filtered may thus have influenced the quantitative results of each taxon identified. This reinforces the need for future studies to produce sequences with higher quality profiles minimizing the number of sequences lost during quality filtering.

4.5.4 Operational taxonomic unit clustering of sequences from penile microbiomes

From the 20 healthy and 20 diseased samples a total of 6320 OTUs were obtained based on clustering the V3V4 quality filtered sequences (97% similarity). The total number of V1V3 sequences left after OTU picking was low (22663 reads) with an average of 54 sequences per sample and was considered too low to infer diversity. No further analysis of the V1V3 hypervariable regions was therefore carried out, again suggesting the importance of producing sequences of high quality. The V3V4 hypervariable regions was thus singly used for the purpose of characterizing the penile bacterial taxonomic composition and diversity.

On review of results from analysis of the mock community sample in the previous section the presence of non-mock species and their relative abundance revealed that despite the stringent quality filtering and the removal of singletons and chimeras, erroneous OTUs were present within mock and non-mock samples. An additional step to remove spurious and contaminant OTUs was thus carried out and the final number of unique OTUs was further reduced to 789 (as described in Chapter 3 section 3.6.2.3.2). This study made use of a combination between the methods developed by Bokulich *et al*. (2013) and Narrowe *et al*. (2015) and used the relative abundance (0.004%) of the highest non-mock species within the mock community as a threshold by which to filter out OTUs that were likely present in this dataset as a result of error.

The 789 OTUs represented 9 964 842 V3V4 sequences of high quality which were used for further downstream analysis. The number of OTUs per sample in the healthy and diseased communities ranged from 292 to 690 and 147 to 738, respectively. The mean number of sequences per sample was 186 016 (SD of 81 893) for healthy rams, and 312 226 (SD of 134 322) for diseased rams. A total of

776 OTUs were shared amongst healthy and diseased groups, and 4 and 9 OTUs were observed to be specific to the healthy and diseased groups, respectively, (Figure 4.15). The classification and relative abundance of the OTUs observed in only the healthy or diseased groups are shown in Table 4.1.

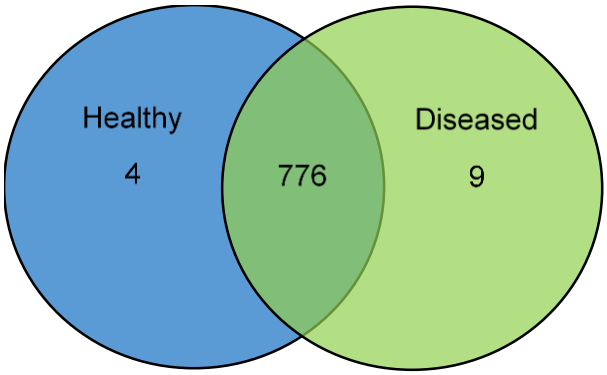

Figure 4.15. Venn diagram depicting the number of OTUs shared between or distinctive to the healthy and diseased groups. Only a small number of OTUs calculated at the 3% dissimilarity level were observed to be specific to either the healthy (4 OTUs) or diseased (9 OTUs) group. 776 OTUs are shared across healthy and diseased individuals.

The majority (12/13) of the unique OTUs were observed in less than half of rams in their respective group (20 rams per group; Table 4.2). In the diseased group *Porphyromonas*_Unclassified_192 and *Porphyromona*s_Unclassified_197 observed the highest (11) and lowest (1) prevalence in diseased rams, respectively. Leptotrichiaceae_Unclassified_416 had the highest prevalence in the healthy group. Large inter-sample variation in the percentage relative abundance of each OTU was observed with the mean and SD given for each OTU in Table 4.2. The association of these OTUs with either the healthy or diseased group is presented in Section 4.8.

Table 4.1. Prevalence (mean ± SD) of the operational taxonomic units (OTU) observed in the healthy or diseased Dorper rams in the study.

OTU	Health Status	Number of rams	Prevalence of OTU
Bacteroidetes_Unclasified_147	Disease	7	0.0093± 0.0328
Bacteroidales_Unclasified_170	Disease	7	0.0024 ± 0.0107
Bacteroidales_Unclasified_171	Disease	3	0.0002 ± 0.0009
Porphyromonas_Unclasified_192	Disease	11	0.0046 ± 0.0207
Porphyromonas_Unclasified_197	Disease	1	0.0002 ± 0.0007
Porphyromonas_Unclasified_198	Disease	2	0.0001 ± 0.0003
Porphyromonas_Unclasified_201	Disease	4	0.0009 ± 0.0004
Pedobacter_Unclasified_687	Disease	3	0.0002 ± 0.0007
Clostridiales_Unclasified_374	Disease	4	0.0018 ± 0.0079
Arcanobacterium_Unclasified_646	Healthy	4	0.0004 ± 0.0013
Streptococcus_Unclasified_371	Healthy	4	0.0007 ± 0.0033
Leptotrichiaceae_Unclasified_416	Healthy	6	0.0002 ± 0.0006
Leptotrichiaceae_Unclasified_420	Healthy	4	0.0002 ± 0.0005

4.6 Taxonomic description of the bacterial community observed in the penile microbiota

The 16S high throughput next generation sequencing (HT-NGS) technology was used to characterize the underlying microbiome of the penis of Dorper rams and to determine observable differences between healthy rams and rams affected by UB in an attempt to understand the aetiology of the disease. In order to identify the bacterial taxa that exist in the penile microbiome of rams, taxonomic assignment at the phylum, class, order, family and genus level was conducted on OTUs from each sample using the Ribosomal Database Project (RDP) Naïve Bayesian classifier (Wang *et al*., 2007) against the Greengenes reference database.

In this section we provide a description of the bacterial taxa that were present in the healthy and diseased microbiome of the penis of Dorper rams. No definition of what constitutes highly abundant taxa has been described and most studies have looked at taxa that have a relative abundance of 1% and higher. Only the most abundant taxa that had a total relative abundance ≥1% across all samples (healthy and diseased) at each taxonomic level are thus described here and are presented in Table 4.2

Classifying the short DNA sequences (394bp) using the RDP Naïve Bayesian Classifier at ≥80% bootstrap confidence allowed 99%, 96%, 95.8%, 79%, 52% and 13% of sequences to successfully be assigned to the phylum, class, order, genus and species taxonomic levels, respectively. The lower frequency of sequences assigned a taxonomy at lower taxonomic levels, especially at the species level,

is expected when using short DNA sequences from hypervariable regions of the 16S gene generated from 16S amplicon sequencing (Janda *et al*., 2007). For example, Price *et al*. (2010) characterised the penile microbiome of men by sequencing the same 16S hypervariable regions (V3V4) and also reported a decrease in the proportion of sequences that were classified at lower taxonomic levels (≥ 95% bootstrap confidence level), with 97.4%, 96.44%, 93.3%, 90.1% and 65.9% of sequences assigned a phylum, class, order, family and genus, respectively.

Table 4.2. Bacterial taxa with total relative abundance ≥1% identified at each taxonomic level. Relative abundance of the taxa across all samples (healthy and diseased) was used to identify bacterial taxa of relative importance.

Phylum	Class	Order	Family	Genus
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Actinomycetaceae*	*Trueperella*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Intrasporangiaceae*	unclassified *Intrasporangiaceae*
Firmicutes	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Salinicoccus*
Firmicutes	*Bacilli*	*Bacillales*	*Planococcaceae*	*Planomicrobium*
Firmicutes	*Bacilli*	*Bacillales*	unclassified *Bacilli*	unclassified *Bacilli*
Firmicutes	*Bacilli*	*Lactobacillales*	*Aerococcaceae*	unclassified *Aerococcaceae*
Firmicutes	*Bacilli*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus*
Proteobacteria	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Moraxella*
Proteobacteria	*Gammaproteobacteria*	*Cardiobacteriales*	*Cardiobacteriaceae*	*Suttonella*
Proteobacteria	*Gammaproteobacteria*	*Pasteurellales*	*Pasterurellaceae*	*Actinobacillus*
Proteobacteria	*Gammaproteobacteria*	*Pasteurellales*	*Pasterurellaceae*	unclassified *Pasteurellaceae*
Proteobacteria	*Epsilonproteobacteria*	*Campylobacterales*	*Campylobacteraceae*	*Campylobacter*
Proteobacteria	*Betaproteobacteria*	*Neisseriales*	*Neisseriaceae*	*Neisseria*
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*	*
Fusobacteria	*Fusobacteriia*	*Fusobacteriales*	*Fusobacteriaceae*	*Fusobacterium*
Fusobacteria	*Fusobacteriia*	*Fusobacteriales*	*Fusobacteriaceae*	unclassified Fusobacteriaceae
Fusobacteria	*Fusobacteriia*	*Fusobacteriales*	*Leptotrichiaceae*	*Leptotrichia*
Fusobacteria	*Fusobacteriia*	*Fusobacteriales*	*Leptotrichiaceae*	unclassified *Leptotrichiaceae*
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	*Porphyromonadaceae*	*Porphyromonas*
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	unclassified Bacteroidales	unclassified *Bacteroidales*
Bacteroidetes	*Flavobacteria*	*Flavobacteriales*	*Flavobacteriaceae*	*

*No bacterial taxa identified as relatively abundant at this taxonomic level

The low sequence assignment at the species level (13%) and the analysis of the mock community that revealed some inaccurate sequence assignment at the species level, excluded species level taxonomic description from this study. Although not directly comparable due to the differences in the study such as sequencing platform used, the higher taxonomic assignment observed at the family and genus level by Price *et al*. (2010) mentioned in the afore paragraph indicates that further optimization of the protocols and methods used in this study could result in better taxonomic assignment of the penile microbiome using the V3V4 hypervariable regions.

Nonetheless, the V3V4 hypervariable regions was able to identify an array of bacterial taxa at the ≥80% bootstrap confidence level, with sequences that could be classified to 14 phyla, 39 classes, 69 orders, 134 families and 248 genera. Bacterial taxa at the phylum to family level were shared amongst the healthy and diseased populations. Two genera, *Arcanobacterium* and *Pedobacter*, were present in the healthy group only, and were only identified in 2 and 3 samples, respectively. One sample comprised ~99% of the sequences in *Pedobacter*. At this prevalence and high variation in numbers, these genera are not likely to represent that of the stable microbiome of the penis or the microflora involved in disease.

4.6.1 Bacterial taxa identified with total relative abundance ≥1% in healthy rams

Bacterial taxa observed at a total relative abundance ≥1% at each taxonomic level are presented in Table 4.3 and their relative abundance is shown in Table 4.4. The bacterial taxa listed in Table 4.4 are in the order of most abundant to least abundant according to total relative abundance. Five out of fourteen phyla had higher relative abundances of more than 1% (Table 4.2). Actinobacteria, which had a relative abundance of 30.6% was identified as the most relatively abundant in healthy rams and together with the phylum Firmicutes (25.1%) and Proteobacteria (22.8%) made up approximately 77% of the sequences at the phylum level, indicating these 3 phyla as important component of the mucosal microbiome of the penis in rams. The phyla Fusobacteria and Bacteroidetes were the next most abundant phyla and were observed at lower relative abundances of 12.6% and 8.1%, respectively. The remaining 9 phyla made up only 1.2% of the sequences.

On analysis of the total relative abundance at the class level, we found 9 classes of bacteria that formed part of the healthy penile microbiota at a relative abundance ≥1% (Table 4.2). The class Actinobacteria was the most relatively abundant class observed and made up ~95% of the organisms detected in the most abundant phylum Actinobacteria. The order of the subsequent classes from most to least relatively abundant included Bacilli (25.4%), Fusobacteriia (11.1%), Gammaproteobacteria (9.3%) and Epsilonproteobacteria (6.5%). The remaining 4 phyla hadrelative abundances between 2.1% and 4.2%.

The most dominant orders within the penile microbiome, representing 58.5% of sequences, were Actinomycetales, Lactobacillales and Fusobacteriales (Table 4.3), which form part of the 3 phyla that make up majority of residence in the penile microbiome. Actinomycetales observed the highest relative abundance (29.2%), but the most dominant order varied between Actinomycetales, Lactobacillales and Fusobacteriales according to samples. Other orders that were dominant included Bacteroidales, Flavobacteriales and Bacillales and their relative abundance can be seen in Table 4.3.

The family Corynebacteriaceae was the most relatively abundant (20.9%) and represented 77% of the phylum Actinobacteria (71%). The family Aerococcaceae was revealed as the second most relatively abundant (14.6%) family making up the microbiota of the ram penis. The family Leptotrichiaceae was

the next most relatively abundant, followed by Campylobacteraceae. One family observed as relatively abundant in healthy rams could not be classified and was revealed having higher taxonomic classification in the order Bacilli. This unclassified family observed a low relative abundance of 1.1%.

On analysis of the genera present it was very apparent that the penile microbiome was dominated by one genus. The genus *Corynebacterium* was the most prevailing genera in healthy samples at an observed relative abundance of 20.9%. *Corynebacterium* species are small pleomorphic, Gram-positive, non-motile, facultative anaerobes (Quinn *et al*., 2011c). Brightling (1988) suggested species from this genus as part of the natural microbiota of the genital tract in sheep. This genus was isolated from two separate studies (Kidanemariam, 2003; Gouletson & Fthianiks, 2006) on the penile mucosa of healthy sheep. This genus is suggested as a commensal group of bacteria, and also resides in other mucosal membranes such as in the respiratory tract and the intestine of animals (Markey *et al*., 2004). Some members are also found in the urethra of healthy animals (Hiramune *et al*., 1970).

All other genera made up smaller proportions of the penile microbiome, with a high proportion of sequences unclassified at this taxonomic level, and were identified at higher taxonomic levels belonging to the families Aerococcaceae, Leptotrichiaceae, Pasteurellaceae, Fusobacteriaceae, Intrasporangiacea, the class Bacilli, and the order Bacteroidales. Unclassified sequences (OTUs) are not unusual in 16S amplicon studies and studies conducted on the microbiome of the penis (Price *et al*., 2010; Lui *et al*., 2013) and another study on the vaginal microbiome of cows (Laguardia-Nascimento *et al*., 2015) indicated the same findings. Combined with poor taxonomic assignment of short DNA sequences, poor classification of sequences can also reflect the yet unexplored bacteria (i.e. not culturable). These unclassified “genera” ranged from abundant to least abundant with unclassified Aerococcaceae identified as the second most relatively abundant genus followed by unclassified Leptotrichiaceae. *Campylobacter* and *Facklamia* were the next most relative abundant genera and observed equal relative abundance in the penile microbiome of healthy sheep.

The family Aerococcaceae consists of Gram-positive, non-spore forming, non-motile, and facultative anaerobic, coccibacilli cocci or catalase-negative ovoid (Lawson, 2014). This family does not seem to be thoroughly documented in the literature and its natural occurrence in the reproductive tract of male animals is not evidently attained. The genus *Facklamia* also identified as abundant in the penile microbiome, belongs to the family Aerococcaceae. This genus consists of Gram-negative, catalase-negative cocci that are facultative anaerobes (Hoyles *et al*., 2014). This genus has mostly been isolated from the teats and milk of cows (Takamatsu *et al*., 2006; Verdier-Metz *et al*., 2012) and seems to form part of the reproductive tract of female animals (Collins *et al*., 1999; Takamatsu *et al*., 2006). Literature on *Facklamia* has not suggested members of this genus as part of the natural microflora of the reproductive tract of male animals and thus the prevalence and high relative abundance of *Facklamia* in the penile environemt of rams will require validation in future studies.

Table 4.3. The percentage of sequences assigned to the most abundant bacterial taxa at each taxonomic level.

	All samples	Healthy samples		Diseased samples	
	relative abundance	Relative abundance (%)		Relative abundance (%)	
Phylum	**(%)**	**total**	**mean (SD)**	**total**	**mean (SD)**
Actinobacteria	29.2	30.6	29.67 (14.36)	27.9	28.69 (13.38)
Firmicutes	25.1	26.0	24.24 (8.90)	24.3	24.10 (11.72)
Proteobacteria	22.8	25.7	27.16 (13.13)	20.0	20.02 (8.79)
Fusobacteria	12.6	11.1	13.04 (13.99)	14.1	20.02 (8.79)
Bacteroidetes	8.1	4.7	4.91 (3.15)	11.5	10.57 (10.71)
Class					
Actinobacteria	28.3	29.5	27.56 (14.35)	27.1	27.81 (13.27)
Bacilli	24.3	25.4	23.54 (8.97)	23.2	22.85 (12.91)
Fusobacteriia	12.6	11.1	13.04 (13.99)	14.1	13.87 (14.59)
Gammaproteobacteria	9.2	11.1	12.12 (13.16)	7.4	7.89 (6.67)
Epsilonproteobacteria	6.7	6.5	6.08 (5.54)	7.0	5.76 (6.79)
Bacteroidia	5.5	2.1	2.11 (2.51)	8.9	8.13 (11.04)
Betaproteobacteria	3.4	4.2	4.99 (7.39)	2.6	3.11 (4.99)
Alphaproteobacteria	3.4	3.8	3.83 (3.79)	2.9	3.17 (2.81)
Flavobacteria	2.0	1.8	2.11 (2.42)	2.1	1.82 (3.55)
Order					
Actinomycetales	28.1	29.2	27.31 (14.37)	27.0	27.72 (13.25)
Lactobacillales	18.8	18.2	16.32 (11.04)	19.3	19.96 (12.49)
Fusobacteriales	12.6	11.1	13.04 (13.99)	14.1	13.87 (14.59)
Campylobacterales	6.7	6.5	6.08 (5.54)	7.0	5.76 (6.79)
Bacteroidales	5.5	2.1	2.11 (2.42)	8.9	1.82 (3.55)
Bacillales	3.8	4.8	4.38 (4.75)	2.7	2.80 (4.18)
Pasteurellales	3.5	4.6	5.10 (10.39)	2.4	2.44 (3.98)
Neisseriales	3.1	4.0	4.75 (7.38)	2.3	2.73 (4.94)
Cardiobacteriales	2.7	2.6	2.78 (3.22)	2.8	3.11 (3.90)
Flavobacteriales	2.2	1.8	2.11 (2.51)	2.1	8.13 (11.04)
Pseudomonadales	2.0	2.6	2.82 (5.98)	1.3	1.40 (2.13)
Rhizobiales	1.5	1.6	1.67 (1.68)	1.3	1.37 (1.18)
unclassified Bacilli	1.1	1.2	1.14 (2.56)	0.9	0.89 (1.24)
Family					
Corynebacteriaceae	20.9	20.9	19.36 (14.42)	20.9	21.72 (13.19)
Aerococcaceae	16.2	14.6	13.10 (11.46)	17.8	17.50 (12.69)
Campylobacteraceae	6.7	6.5	6.08 (5.54)	7.0	5.76 (6.79)
Fusobacteriaceae	6.7	4.0	5.37 (9.61)	9.5	10.10 (14.66)
Leptotrichiaceae	5.9	7.1	7.66 (7.02)	4.7	3.78 (5.30)
Porphyromonadaceae	3.6	1.6	1.54 (2.06)	5.6	5.36 (8.12)
Pasterurellaceae	3.5	4.6	5.10 (10.39)	2.4	2.44 (3.98)
Neisseriaceae	3.1	4.0	4.75 (7.38)	2.3	2.73 (4.94)
Cardiobacteriaceae	2.7	2.6	2.78 (3.22)	2.8	3.11 (3.90)
Streptococcaceae	2.5	3.5	3.05 (3.28)	1.5	1.37 (1.86)
Actinomycetaceae	1.7	1.3	1.23 (2.38)	2.2	1.64 (2.37)
Staphylococcaceae	1.7	2.2	1.90 (2.48)	1.1	1.26 (2.88)
Planococcaceae	1.6	2.1	1.98 (2.16)	1.1	1.18 (1.16)
Moraxellaceae	1.6	2.2	2.13 (5.86)	1.1	1.12 (2.13)
Intrasporangiaceae	1.3	1.6	1.33 (1.94)	0.9	1.06 (1.19)
unclassified *Bacteroidales*	1.2	0.0	0.13 (0.40)	2.3	1.65 (3.11)
Flavobacteriaceae	1.2	1.0	1.32 (2.22)	1.4	1.20 (3.22)
unclassified *Bacilli*	1.1	1.2	1.14 (2.56)	0.9	0.89 (1.24)
Genus					
Corynebacterium	20.9	20.9	19.36 (14.42)	20.9	21.72 (13.19)
unclassified *Aerococcaceae*	10.0	8.1	5.71 (5.92)	11.9	9.10 (7.08)
Campylobacter	6.7	6.5	6.08 (5.54)	7.0	5.76 (6.79)
Facklamia	6.2	6.5	5.45 (6.32)	5.9	5.94 (5.40)
Fusobacterium	5.0	3.0	4.61 (9.66)	6.9	8.28 (14.78)
unclassified *Leptotrichiaceae*	4.8	5.9	3.52 (4.18)	3.9	0.63 (2.25)
Porphyromonas	3.6	1.6	1.52 (2.05)	5.6	5.35 (8.11)
Neisseria	2.8	3.8	4.41 (7.29)	1.9	2.15 (4.49)
Suttonella	2.7	2.6	2.74 (3.20)	2.8	3.08 (3.86)
Streptococcus	2.5	3.5	3.05 (3.28)	1.5	1.37 (1.86)
unclassified *Pasteurellaceae*	1.8	3.1	2.42 (7.83)	0.5	0.56 (1.00)
unclassified *Fusobacteriaceae*	1.8	0.9	0.77 (1.80)	2.6	1.82 (3.06)
Trueperella	1.7	1.3	1.13 (2.24)	2.1	1.59 (2.37)
Moraxella	1.6	2.1	2.00 (5.86)	1.0	1.08 (2.14)
Salinicoccus	1.3	1.7	1.43 (1.91)	0.9	0.97 (2.34)
Planomicrobium	1.3	1.8	1.58 (1.75)	0.9	0.93 (0.95)
unclassified *Bacteroidales*	1.2	0.0	0.13 (0.40)	2.3	1.65 (3.11)
unclassified *Intrasporangiaceae*	1.1	1.4	1.15 (1.69)	0.8	0.92 (1.06)
Actinobacillus	1.1	1.2	2.42 (7.83)	1.0	0.56 (1.00)
unclassified *Bacilli*	1.1	1.2	1.14 (2.56)	0.9	0.89 (1.24)
Leptotrichia	1.0	1.2	1.71 (3.51)	0.8	0.63 (2.25)

The family Leptotrichiaceae within the phylum Fusobacteria consists of fewer explored and infrequently isolated microorganisms and members have obligate anaerobic or capnophilic growth. They are Gram-negative fastidious bacteria and are mostly anaerobic, with a few species that are aerobic (Eisenberg *et al*., 2016). Bacteria from this family are known as natural colonizers of mucous membranes that form the inner lining of the oral cavities, gastrointestinal or urogenital tracts of animals (Dewhirst *et al*., 2012; Sturgeon *et al*., 2014; Bik *et al*., 2016; Knudsen *et al*., 2016).

The genus *Campylobacter* consists of bacteria that are non-spore forming, Gram-negative, spiral shaped rods. They demonstrate microaerophilic metabolic characteristics, and some species are also able to grow in aerobic and anaerobic conditions (Fitzgerald *et al.,* 2006). This genus has been highlighted as part of the normal intestinal flora of many animals which include sheep, cattle and chickens (Workman *et al*., 2005). They are also sometimes commensal in the reproductive tract of animals (Quinn *et al*., 2011a) and the oral cavity (Rowe & Madden, 2014). For example, one species, *Campylobacter fetus*, appears to be adapted to the environment of the preputial mucosa of bulls (Quinn *et al*., 2011a).

The glans penis is a lipid rich environment high in fat and protein which is gained from the nutrient smegma. The urethral passage is also closely linked to the environment of the glans penis and provides further nutrition that is high in nitrogen sources (Wilson, 2005). It is thus not surprising that a wide array of bacteria occur in this nutrient-rich environment. The core microbiome of the ram species present in the penile environment has not been objectively investigated. Studies previously cited on the penile microbiome of rams have used culture-based methods (Kidanemariam, 2003; Gouletsou *et al*., 2006). Findings that *Corynebacterium* was the most relatively abundant genus (making up 20.9% of the microflora present), as well as its prevalence across all rams sampled may suggest this group of organisms as part of the core microbiome of the ram penile environment (glans penis and prepuce). This warrants further investigation.

All bacterial taxa were represented across the 20 healthy samples, besides at the genus level. Some of the more lowly abundant genera observed a decrease in the representation across all samples. For example, the genus *Leptotrichia* that had a total relative abundance of exactly 1%, was not represented in 2 samples. A vast degree of per sample variation in relative abundance was observed and was a shared characteristic at all taxonomic levels (Table 4.3). For example, the mean relative abundance and standard deviation across all healthy samples for the phylum *Firmicutes* was given 24.24% ± 8.90%. At the genus level the most abundant genus *Corynebacterium* revealed the highest per sample variation with mean relative abundance of 19.36% ± 14.42%.

The lack of previous work on the penile microbiome of rams and the lack of a larger sample size, limits the ability to identify the inter-sample variations as a natural characteristic of the penile microbiome. However, studies on the penile microbiome of men suggest that high inter-sample variation of the bacterial taxa observed in the penile microbiome appears to be a natural phenomenon (Price *et al*.,

2010; Lui *et al*., 2013). We can speculate that this inter-sample variation observed may be explained by environmental (e.g. climate, diet, etc.) and biological heterogeneity (e.g. age, head colour, sexual maturity), and differences in animal history, combined with inherent bias introduced during PCR amplification, sequencing and data processing. The number of biological and environmental variables that differ between individuals sampled must be minimized to reduce this quantitative variation, and factors like general management of livestock must be considered when designing and interpreting microbiome studies, which can influence an individual ram's penile microbiome and thus generate variations.

4.6.2 Comparison of the healthy penile microbiome of rams to culture-based studies

Due to the lack of 16S amplicon studies to date that have been carried out on the ram penile microbiome or that of other livestock species, the healthy microbiome of the penis characertized in this study is compared to culture-based studies that investigated the bacteria present in the penile environment of rams. We compare only the bacterial genera, as the ram penile microbiomes have been most extensively studied at this taxonomic level. Table 4.4 shows the genera that were previously identified in healthy rams using culture-based identification methods and that were in common amongst studies. We report only on the most prevalent genera observed in these studies. Within the genus *Arcanobacterium* only the species *Arcanobacterium pyogenes* was identified in both Kidanemariam (2003) and Gouletsou *et al*. (2006), which has since been renamed as *Trueperella pyogenes* (Yassin *et al*., 2011) and as a result *Arcanobacterium pyogenes* identified in their study will be changed to belonging to the genus *Trueperella* for the purpose of comparison.

A study by Kidanemariam, (2003) that identified bacteria using culture-based methods observed that the species from the genus *Streptococcus* (30%) followed by *Corynebacterium* (26.7%) were the most prevalent in healthy rams. Species identified as belonging to the genera *Staphylococcus, Enterococcus, Trueperella* and *Pasteurella* were subsequently isolated, in order of highest to lowest prevalence. A later study by Gouletsou *et al*. (2006) that used microbiological methods to culture and identify the bacteria present in the preputial cavity of 48 healthy rams, also isolated *Streptococcus*, *Corynebacteriurm*, *Staphylococcus* and *Trueperella* species as prevalent species. The prevalence of *Staphyloccocus* was higher in Gouletsou *et al*. (2006) with 28.6%, compared to 22.4% in Kidanemariam (2003) and indicated the highest prevalence. *Corynebacterium* and *Trueperella* had the same prevalence of 6.8%, which was lower than that reported by Kidanemariam (2003), i.e. 26.7% and 17.2%, respectively. *Streptococcus* and *Pasteurella* were also observed however, they formed part of the lowly prevalent species (3.4%). *Enterococcus* was not identified by Gouletsou *et al*. (2006) but another genus *Eschericia* made up the most prevalent (25.4%) genera in the ram penis.

Examination of this study revealed that all genera identified as the most prevalent in Kidanemariam (2003) and Gouletsou *et al*. (2006) were also identified by this study. Interestingly, only two of the most

prevalent genera, *Corynebacterium* and *Streptococcus*, identified by culture-based methods were genera identified having a relative abundance ≥1% in the ram population in this study. The genus *Staphylococcus* was not identified by this study. Results from the mock community analysis in section 4.5 revealed that the pipeline used in this study was inadequate in assigning sequences to the genus *Staphylococcus*, which might explain the absence of this genus in this study. Sequences that were analysed as *Staphylococcus* species in the mock sample using BLASTn analysis were derived from unclassified Bacilli. Unclassified *Bacilli* was one of the most relatively abundant bacterial "genus" identified, so it is possible that with higher taxonomic classification of these sequences identification of the genus *Staphylococcus* could be observed amongst the sequences classified as unclassified *Bacilli.* Further research has not yet been carried out to investigate this theory.

The genera *Streptococcus* and *Corynebacterium* were a 100% prevalent in theram population used in this study. The genera *Trueperella* and *Escherichia* were both observed in 95% (19/20) of samples in this study. The genera *Enterococcus* and *Pasteurella* were isolated in 90% and 80% of rams. Although it is not possible to directly compare results from Kidanemariam (2003) and Gouletsou *et al*. (2006) to this study due to different methodologies used, the higher prevalence of bacterial genera observed across all sheep sampled in this study, especially at a smaller sample size of 20 versus 48 and 116 in Kidanemariam (2003) and Gouletsou *et al*. (2006), respectively, may indicate a higher specificity of molecular methods (i.e. 16S rRNA gene) in identifying bacteria, and also limitations in successfully culturing bacteria with some species that are unculturable. Differences in the bacteria observed as well as differences in the prevalence of shared bacteria observed by Kidanemariam (2003) and Gouletsou *et al*. (2006) may also be accounted for by inefficient culturing of bacteria and different microbiological tests used to identify them.

Table 4.4. List of bacterial genera that have previously been identified in the penile environment of rams.

Bacterial genera	Studies
Acinetobacter	Kidanemariam, 2003; Gouletsou *et al*., 2006
Actinobacillus	Kidanemariam, 2003; Gouletsou *et al*., 2006
Trueperella	Kidanemariam, 2003; Gouletsou *et al*., 2006
Corynebacterium	Kidanemariam, 2003; Gouletsou *et al*., 2006
Eschericia	Kidanemariam, 2003; Gouletsou *et al*., 2006
Lactobacillus	Kidanemariam, 2003; Gouletsou *et al*., 2006
Pasteurella	Kidanemariam, 2003; Gouletsou *et al*., 2006
Staphylococcus	Kidanemariam, 2003; Gouletsou *et al*., 2006
Streptococcus	Kidanemariam, 2003; Gouletsou *et al*., 2006
Alcaligenes	Kidanemariam, 2003
Flavobacterium	Kidanemariam, 2003
Rhodococcus	Kidanemariam, 2003
Moraxella	Kidanemariam, 2003
Enterococcus	Kidanemariam, 2003
Erysipelothrix	Kidanemariam, 2003
Bacillus	*Gouletsou et al. (2006)*
Clostridium	*Gouletsou et al. (2006)*
Micrococcus	*Gouletsou et al. (2006)*
Mycoplasma	Kidanemariam, 2003

For a group of organisms (such as genus) to be described as part of the core microbiome of a body site, it is said that a large proportion of this group in that particular body site must be observed in every participant in a large group of individuals that have the same age and gender (limiting further variation in factors such as diet and location), and similar findings must be observed in separate studies (Wilson, 2005). These strict criteria could not be met on all levels in this study, however the consistent isolation of *Streptococcus, Corynebacterium, Trueperella* and *Pasteurella* across all three studies of the ram penile environment may indicate these as part of the core microbiota. Especially, *Streptococcus* and *Corynebacteriurm* that had 100% prevalence in this study and formed part of the largest proportion of bacteria isolated in Kidanemariam (2003) and Gouletsou *et al*. (2006), along with the high relative abundances of these genera identified in the previous section in this study. In addition to only the most prevalent taxa which were reported here, 16S amplicon sequencing was able to characterize a higher bacterial richness and diversity than previously identified in the ram penile environment using these culture-based methods. Since this study was the first metagneomic study to characterize the healthy microbiome of the penile environment further studies should aim to replicate this work reducing possible variability brought about by environment and biological differences in an attempt to verify findings from this study and identify the core microbiome of healthy rams.

4.6.3 Comparison of the penile microbiome in healthy and diseased rams

Comparison of the microbiota present in the penile environment of healthy and diseased sheep was carried out, and the distribution of bacterial taxa at the different taxonomic levels are presented in Figure 4.16. The total, mean and standard deviation of the relative abundances of each bacterial taxa in healthy and diseased groups can also be viewed in Table 4.3.

At the phylum level, all phyla perceived high relative abundance in both healthy and diseased samples. The most observable differences between the healthy and diseased groups were from the phyla *Bacteroidetes, Fusobacteria* and *Proteobacteria*. Fusobacteria and *Bacteroidetes* had higher relative abundance in the diseased ram population, whilst *Proteobacteria* was more relatively abundant in the healthy group (Figure 4.16A). The phyla *Bacteroidetes* demonstrated the biggest difference in relative abundance to that of the healthy group (11.5% Diseased; 4.7% healthy). The most abundant to least abundant phyla were consistent across healthy and diseased groups.

The classes *Actinobacteria* and *Bacilli* were the most abundant classes in both the healthy and diseased groups, thereafter the order of the most to least abundant class varied between the two groups (Figure 4.18B). Higher relative abundance for the diseased group was indicated in the classes *Flavobacteria*, Fusobacteria, *Epsilonproteobacteria* and *Bacteroidiia*. The biggest difference in relative abundance between the healthy and diseased groups at the class level was from the class *Bacteroida* belonging to the *Bacteroidetes* phylum and the class *Gammaproteobacteria*, which was more abundant in healthy rams.

The top 3 orders, which included Actinomycetales, Lactobacillales and Fussobacteriales, consistent with the most abundant classes, were the most abundant in both the healthy and diseased groups, with the remaining orders that differed in relative abundance changing the order of most to least abundant (Figure 4.16C). For example, the order Campylobacterales was the fourth most abundant in the healthy group whereas *Bacteroidales* consisted the fourth most abundant order in diseased rams. Seven orders were more relatively abundant in the diseased group which included orders such as Lactobacillales and Cardiobacteriales, with Bacteoridales exhibiting the most evident difference between the groups (Figure 4.16C). The order *Pasteurellales* and *Bacilli* were observably more relatively abundant than in the healthy group.

At the family level Corynbacteriaceae and Aerococcacea were the most relatively abundant families in healthy and diseased rams. Corynebacteriaceae revealed the same relative abundance across both groups whilst a higher relative abundance of Aerococcaceae was observed in the diseased rams. The families Pasteurellaceae, Leptotrichiaceae and Neisseriacecae observed relatively higher abundances in the healthy group, but by far the largest differences were observed in the relative abundance of Porphyromonadaceae and Fusobacteriaceae, which were higher in the diseased group (Figure 4.16D).

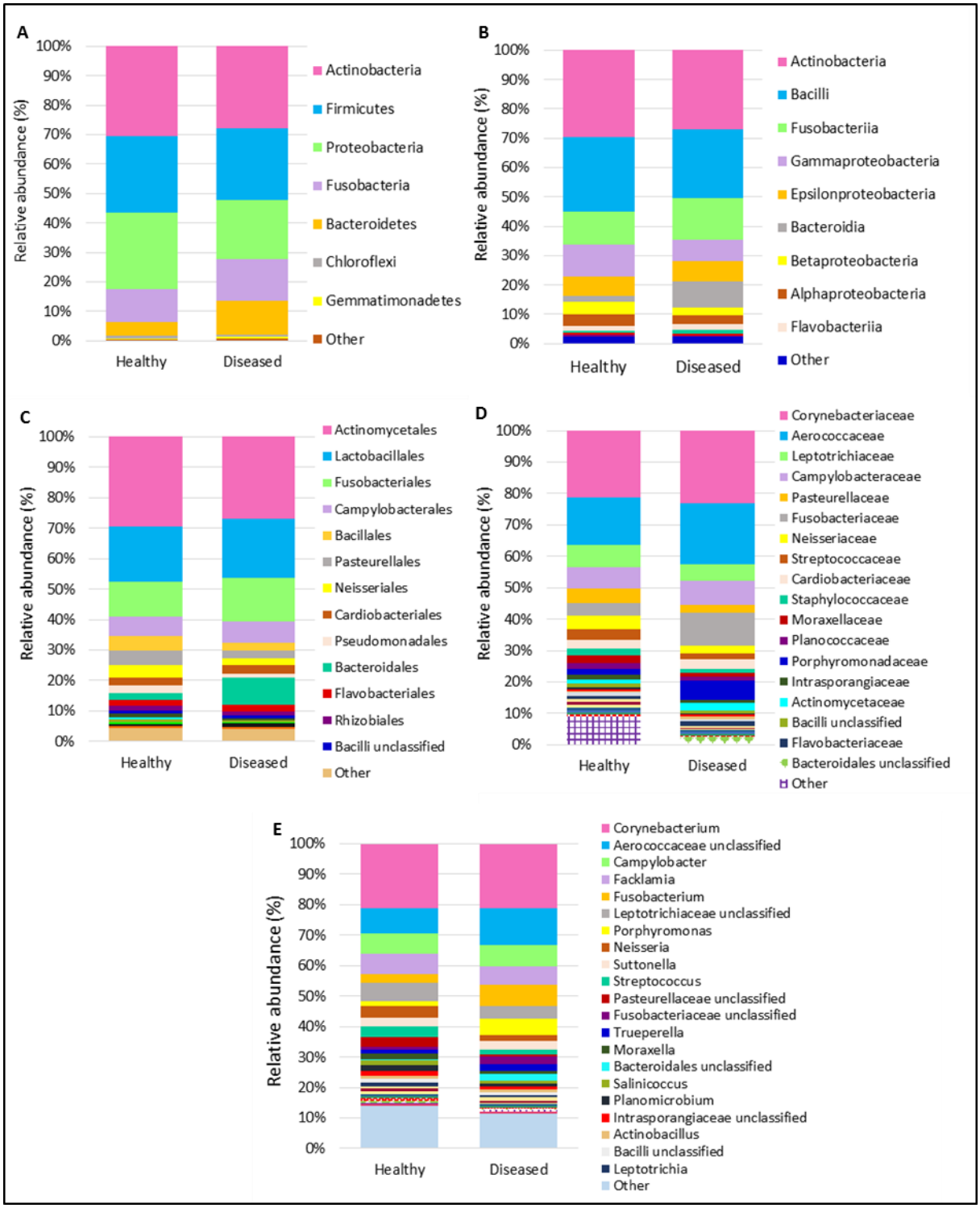

Figure 4.16. Relative abundance of sequences assigned to respective taxonomic classification for healthy and diseased groups. **A)** Phylum, **B)** Class, **C)** Order, **D)** Family, **E)** Genus. Only the taxonomic groups that had total relative abundance ≥1% across all samples were indicated in the legend.

At the family level Corynbacteriaceae and Aerococcacea were the most relatively abundant families in healthy and diseased rams. Corynebacteriaceae revealed the same relative abundance across both groups whilst a higher relative abundance of Aerococcaceae was observed in the diseased rams. The families Pasteurellaceae, Leptotrichiaceae and Neisseriacecae observed relatively higher abundances

in the healthy group, but by far the largest differences were observed in the relative abundance of Porphyromonadaceae and Fusobacteriaceae, which were higher in the diseased group (Figure 4.16D).

The genus *Corynebacterium*, consistent with the family Corynebacteriaceae from which it belongs, was the most relatively abundant genus in both groups, observing the same relative abundance. The most apparent differences between the diseased and healthy groups were from the genera *Fusobacterium* and *Porphyromonas* as well as from sequences that could not be classified to the genus level and could only be classified to the order Bacteroidales, the family Intraspongiaceae and Pasteurellaceae (Figure 4.16E). *Fusobacterium* and *Porphyromonas* were more relatively abundant in the diseased group. A genus classified as *Streptococcus* was noticeably more relatively abundant in healthy rams.

Inter-sample variation in the relative abundances of all bacterial taxa that were indicated as highly abundant in the diseased ram population was also a characteristic in diseased animals (Table 4.4).

This is the first study that used a non-culture method to compare the healthy microbiome in the penile environment of rams to that of rams that had UB. Examination of the bacterial taxa identified and their relative abundances using the non-culture based method indicated that the microbiota that dominated the healthy penile mucosa was also abundant in the diseased penile mucosa, however, differences in the relative abundances at all taxonomic levels were indicated between the groups. The most abundant taxon (e.g. *Corynebacterium* at the genus level) was always consistent between healthy and diseased rams but the order of the most dominant to least dominant was not the same throughout each taxonomic level, suggesting some relationship to UB whereby they could be involved/associated with the disease or their relative abundance may be affected simply by the increase or decrease of other bacterial taxa as a response to disease and changes in environmental conditions. This also can imply that if changes in relative abundances are significantly different, that the disease agent may be naturally occurring in the penile environment and becomes an opportunistic pathogen.

4.7. Alpha and beta diversity of the penile microbiota in healthy and diseased rams

The diversity of the penile microbiota in healthy and diseased rams was compared by calculating alpha and beta ecological metrics. Prior to alpha and beta diversity analysis the number of sequences per sample was rarefied to 56 347 sequences per sample. This was done to normalise the number of reads per sample and avoid sampling error (Soetaert & Heip, 1990; Magurran, 2004). Certain alpha and beta diversity measures are sensitive to differences in sampling efforts (Navas-Molina *et al*., 2013). The number of reads (56 347) was chosen according to the sample with the least amount of sequences.

In order to determine if the selected number of reads per sample was sufficient to accurately represent the bacterial community, rarefaction curves were generated of the healthy and diseased groups. The chao1 index of richness was used as a measure of alpha diversity (Figure 4.17), which estimates the number of OTUs present in a sample. The chao1 curves presented in Figure 4.17 indicated that at the

chosen sequence depth (56 347), most of the bacterial diversity was covered for both communities (healthy and diseased). This is indicated by the levelling off of the curves. The sampling procedure was thus considered adequate for diversity analysis and the use of additional sequences would not considerably influence the diversity estimates. Hence, the analysis of the samples should provide an indication of the whole microbiome and not only a portion of the residing microbiota, although there will always be some under-estimation of overall diversity and species number, with random detection of uncommon species.

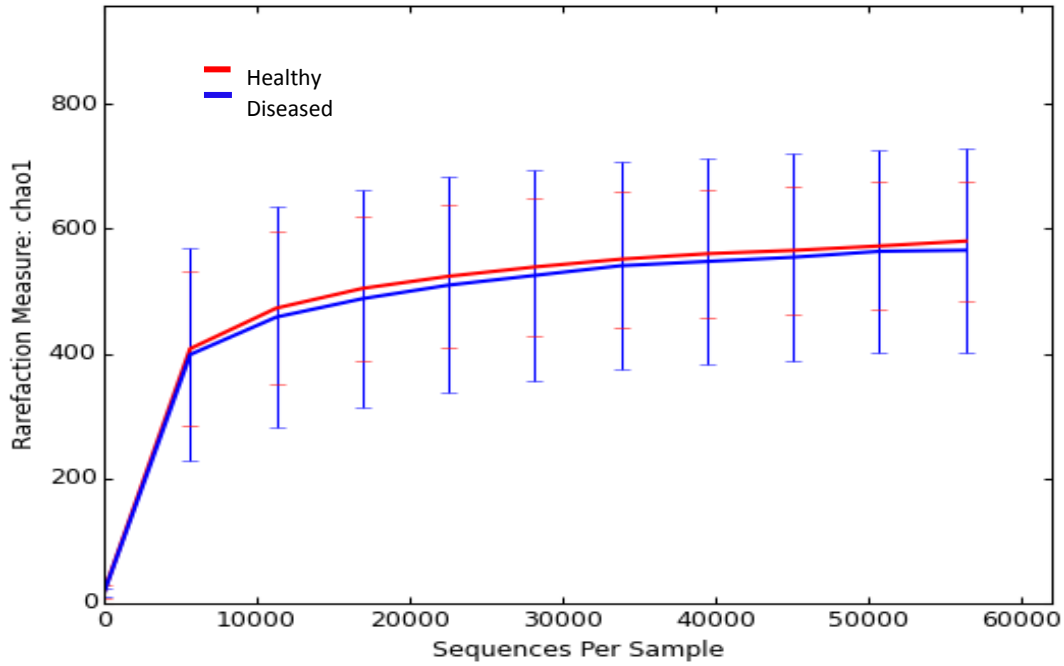

Figure 4.17. Rarefaction curves of the chao1 estimated OTU richness (97% similarity threshold) amongst rams grouped as healthy and diseased. The vertical axis shows the estimated number of OTUs at a given sequencing depth selected on the x-axis. The maximum sequencing depths is the rarefied value of 56 347.

Rarefaction analyses are further used as a tool for evaluating bacterial richness. The rarefaction curves in Figure 4.17 indicated that the OTU richness of the penile microbiota followed the same trend across healthy and diseased ram groups, with the estimated OTU number slightly higher in the healthy group than in the diseased group. Along with chao1 index of richness, two additional alpha diversity measures were calculated, Shannon's diversity index and phylogenetic diversity (PD whole tree) (Table 4.5). These metrics estimate the diversity within samples based on different measures such as phylogenetic relatedness, richness and evenness. Diversity estimates using Shannon's diversity index and phylogenetic diversity (PD whole tree) also revealed higher bacterial diversity in the healthy rams. Comparison of the mean diversity indices using non-parametric, two sample t-tests showed that no significant difference in alpha diversity was observed between the healthy and diseased microbial communities (Table 4.5).

Table 4.5. Comparison of alpha diversity indices of healthy and diseased penile microbiota.

	Healthy		Diseased		
	Mean	SD*	Mean	SD	P-value #
Chao1	580.19	96.08	565.55	162.98	0.33
Phylogenetic Distance	37.58	6.97	36.32	10.71	0.42
Shannon index	4.92	1.22	4.66	0.99	0.70

*Standard deviation (SD)

Bacterial species richness and diversity are thought to be important components of a healthy microbiome. Various conditions have been associated with changes in richness and diversity such as chronic diarrhea in humans (Khoruts *et al*., 2010), but others observed no associations. For example, coilitis in horses was found not to be associated with a change in diversity between that found in healthy horses (Costa *et al*., 2012). Although UB had lower richness and diversity, it was not associated with loss of diversity and richness, but further studies using less variable groups of rams are required. The diversity metrics used to decipher between differences in bacterial diversity take into account richness, evenness and phylogenetic relatedness. In the sequence data 776 OTUs were shared in common between the healthy and diseased groups, which left only 9 and 4 OTUs that were specific to the diseased and healthy communities (previously presented in section 4.5.3; Table 4.1), respectively. So when looking at bacterial richness, the number of bacterial species (i.e. OTUs) were practically the same in each group (780 OTUs in healthy and 789 OTUs in diseased) and even with specific OTUs in each group this was not enough to suggest a difference. This would show the same results when comparing phylogenetic relatedness, 776 OTUs are shared between the two groups and therefore OTUs of a particular phylogeny in one group will also be present in the other group. This indicates that the lack of significant findings is thus unsurprising.

When considering the evenness of the population, two populations may have the same OTU distribution of abundance, however at the individual OTU level the relative abundance may have changed from one group to another and another OTU with more or less the same abundance may have done the same in the opposite direction. So although one OTU loses abundance in one group the abundance lost in that group may have been regained from another OTU that increased abundance in the same group at relatively the same abundance. This would then still consider the two groups as having even abundance distributions, although differences in relative abundance occurred at the OTU level. Communities may be identical in terms of richness, evenness and phylogeny but differ in their taxonomic diversity of their species, as well as their relative abundances (Magurran, 2004). So although the diversity metrics suggested no significant difference in bacterial diversity according to these measures, some differences in relative abundance in the healthy and diseased population may have occurred at the individual OTU level and these are therefore tested for significance in the following section of this chapter. We also

grouped these OTUs into phylotypes from phylum to genus and examined for any significant differences in relative abundance between the healthy and diseased groups.

Beta diversity analysis tests the extent of similarity or dissimilarity between microbial communities by measuring the degree to which structure is shared or different between communities. The microbial communities were compared using weighted and unweighted Unifrac distance beta diversity metrics. UniFrac measures the phylogenetic distance between samples, with the weighted metric taking into account the relative abundance of each phylogenetic lineage within samples (Lozupone & Knight, 2005). All-by-all distance matrices of the weighted and unweighted UniFrac distances between all samples were visualized on 2-dimensional Principle Coordinates Analysis (PCoA) plots of the three principle components (PC1 to PC3) that explained the greatest portion of variation (Figure 4.18; Figure S4.1). The penile microbiota of the healthy and diseased groups did not show clear separation into different clusters according to community composition using UniFrac and were not separated clearly on the PCoA plot. Consistent with these findings, PERMANOVA analysis on the Unifrac distance matrices indicated that no significant difference in community composition exists between the bacterial communities from healthy and diseased rams (weighted Unifrac, p= 0.10; unweighted Unifrac, p= 0.35). Figure 4.18 shows high inter-sample variation across all samples, including within healthy and diseased groups.

These findings revealed that the community composition of healthy samples was similar to samples from the diseased group and overall bacterial diversity between healthy and diseased groups was similar and no patterns in community composition could be used as a basis to discriminate between healthy and diseased rams. This suggests that diseased subjects did not have altered or unusual penile bacterial communities. It is possible that the limitations of 16S amplicon sequencing that affect taxonomic resolution to lower levels (i.e. species) may influence the non-significant results observed. This study was only able to assign taxonomy to 13% of sequences at the species level thus we cannot exclude the possibility that some of the bacterial shifts that affect disease states such as UB can occur at these lower taxonomic levels. However, it may also be that the high degree of inter-individual variation masks differences in composition related to disease. This study consisted of taking swab samples of bacteria present in the lesions observed on the glans penis, but also extended to the portion of the penis that is free of knonw UB symptoms. It is possible that having swabbed beyond the lesions on the glans penis, in affect increasing the bacterial diversity present in the swab, that any minor change in community from the lesions may have been masked by the larger overall diversity of the entire surface of the penis. The study may therefore have benefited from including a swab of "healthy" mucosa from the diseased rams (e.g. a site next to a lesion), this would allow one to compare within each individual if there was a shift in composition in the healthy and affected region, or simply sampling from a smaller region concentrating only on the areas that are affected.

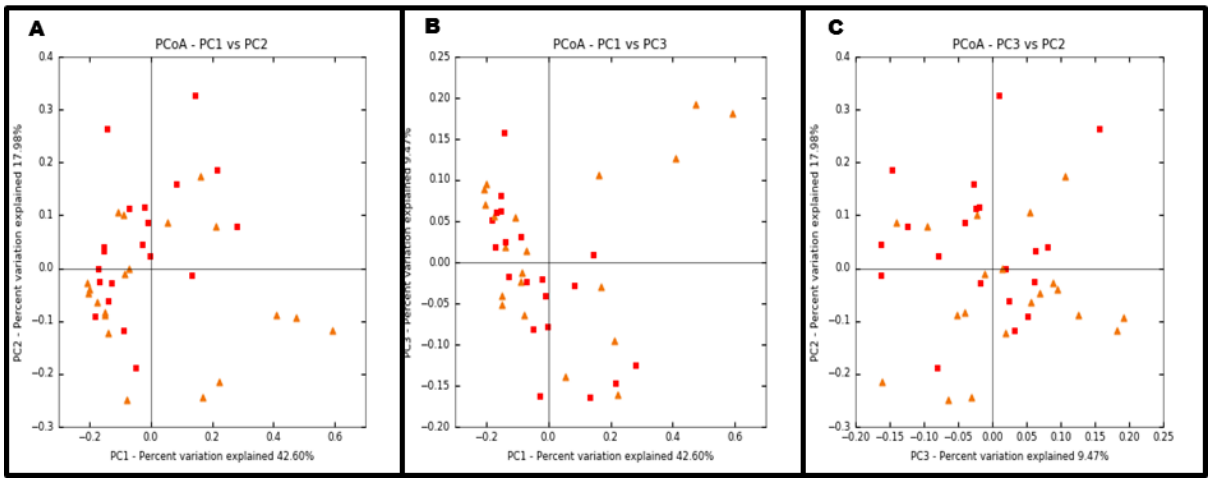

Figure 4.18. Unifrac diversity measures. Principle coordinate analysis, using weighted Unifrac distances to explore and visualize diversity between samples. Each point represents a sample plotted by a principal component on the X- axis and another principal component on the Y- axis. Square (red) data points represent samples from healthy rams and triangular (orange) from diseased rams. PC1 is principal component 1, PC2 is principle component 2 and PC3 is principle component 3. **A)** PC1 versus PC2, **B)** PC1 versus PC3, **C)** PC3 versus PC2.

4.8 Identification of phylotypes and operational taxonomic units with differential abundance in the penile microbiota

In order to determine which bacterial taxa occurred at significantly different abundances in the penile microbiota of healthy versus diseased rams the LEfSe analysis tool was used (Segata *et al.,* 2011). LEfSe analysed OTUs grouped into phylotypes according to taxonomical classification, from the phylum to genus level. Figure 4.19A shows the taxa identified as differentially represented and ranked according to effect size, which is described as a Linear Discriminant Analysis (LDA) score. According to LEfSe no taxa, from the phylum to genus level, were identified to be significantly differential in the diseased group. In the healthy community 3 taxa were identified as having significantly higher abundance at the family and genus level when compared to the diseased community. No taxon was identified as significantly more abundant at the phylum, class and order level in the healthy community.

Differences in relative abundance between the healthy and diseased groups were only identified at the family and genus level. The two most biologically relevant taxa in the healthy community were the family Leptotrichiaceae and Streptococcaceae, with Leptotrichiaceae demonstrating the highest LDA score, and therefore displayed the highest difference in relative abundance between the groups. Within the family Streptococcaceae, the genus *Streptococcus* was also identified by LEfSe as being significantly more abundant in the healthy community and was exclusively responsible for the significantly greater abundance of Streptococcacceae at the family level as revealed by the cladogram in Figure 4.19B. At the lower biological relevance with LDA scores between 2 and 4, two additional taxa identified to the genus level, *Pasteurella* and *Arcanobacterium*, were significantly more abundant in the healthy community. The family Myxococcaceae proved significantly abundant but had the lowest biological relevance as compared to the rest of the taxa flagged by LEfSe as significantly differential. The taxa

Myxococcaceae, *Pasteurella* and *Arcanobacterium* identified as significantly differentially abundant in the healthy group had low relative abundances and were not from dominant taxa that had total percentage relative abundances ≥1%. Findings indicate that these taxa during the incidence of disease may decrease significantly in relative abundance from that naturally occurring in the healthy state.

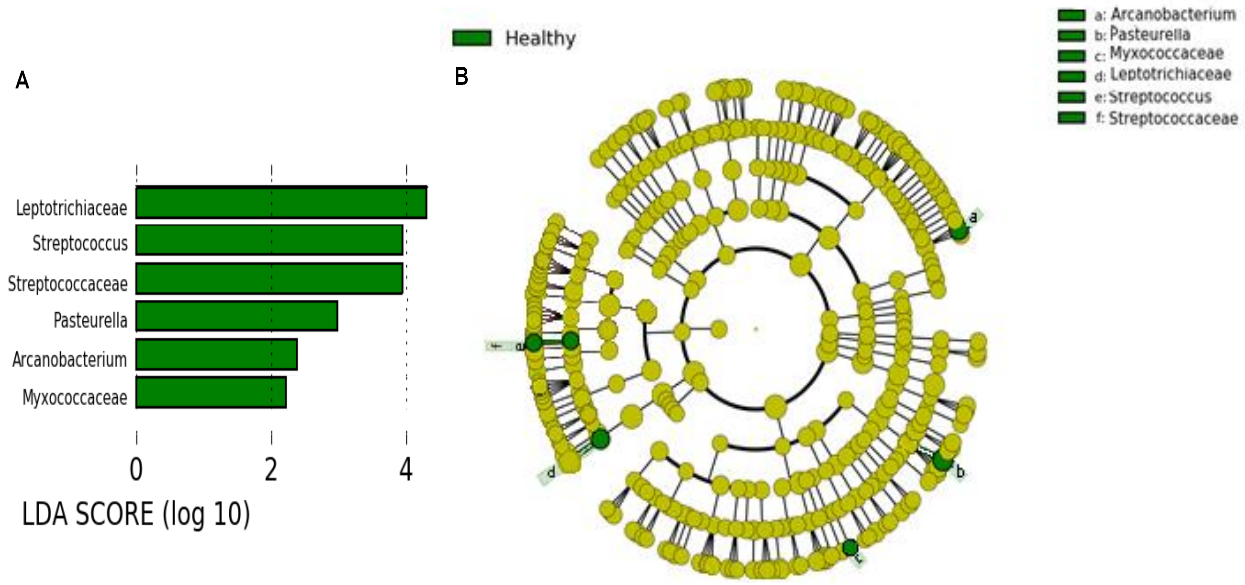

Figure 4.19. Differentially abundant taxa (classified from the phylum to genus) as revealed by LEfSe. **A)** Histogram showing taxa most abundant in the healthy group. Taxa are ranked according to linear discriminant analysis and only taxa with p-value <0.05 and LDA scores ≥2 were included in the results. **B)** Cladogram displaying the phylotypes that were most abundant in the healthy ram population with the size of the circles increasing with increasing relative abundance.

Differentially represented OTUs were also analysed with LEfSe, and results are shown in Figure 4.20. LEfSe calculates the effect size of each OTU as determined by the LDA score and no OTUs with LDA score lower than 2 was flagged by LEfSe as having biological relevance in either healthy or diseased communities. Due to various factors such as the short reads (394bp) used and the chosen database many OTUs were not able to be classified down to the genus or species level and as a result many OTUs remained unclassified at various taxonomical levels (from phylum to genus). Nonetheless, these OTUs were included in the LEfSe analysis. A total of 40 OTUs were observed as significant, with 21 and 19 OTUs more abundant in the healthy and diseased groups, respectively.

The top 5 OTUs represented has significantly more abundant having the highest biological relevance (LDA score >3) in the healthy community (in order of effect size) were associated to Leptotrichiaceae, *Leptotrichia, Corynebacterium,* Pasteurellaceae and *Trueperella* (Figure 4.20). The remaining OTUs that had significantly higher abundances in the healthy community observed lower LDA scores that ranged between 2 and 3. Apart from Leptotrichiaceae_Unclassified_504 that observed the highest LDA score of all 21 OTUs, other OTUs belonging to the Leptotrichiacea family, with three having the lowest LDA scores (Leptotrichiaceae_Unclassified_498, Leptotrichiaceae_Unclassified_497, Leptotrichiaceae _Unclassified_ 506) were also identified as significantly more abundant in the healthy community.

These results suggest that at the Leptotrichiaceae family level, these 6 individual OTUs may have largely been responsible for Leptotrichiaceae's significant abundance in healthy individuals seen in previous LEfSe analysis of phylotypes (Figure 4.19A).

Four of the remaining OTUs that had significantly higher abundances in the healthy community were associated to the families, Micrococcaceae and Bradyrhizobiaceae, and the orders Gemm-1 and iii1-15 from the class Acidobacteria. At the genus level, a further 5 OTUs from 5 different genera, ranging in their LDA scores had higher abundances in the healthy community, with *Corynebacterium*_Unlclassified_OTU77 displaying the highest LDA score of the group (Figure 4.20). One of these OTUs, *Arcanobacterium*_Unclassified_OTU63, was the only OTU that contributed to the relative abundance of the genus *Arcanobacterium* and was therefore exclusively responsible for the significant abundance of this phylotype in the healthy community seen in Figure 4.19A. The species belonging to *Arcanobacterium* are non-motile, non-spore forming, Gram-positive bacteria. They are thought to colonize the mucous membranes of animals such as the nasopharyngeal mucosa of cattle, sheep and pigs (Quinn *et al*., 2011b). Their occurrence in the healthy mucosa of the penis is thus expected.

While the study was not designed to determine the aetiology of UB, some notable species-level identification was investigated and the fact that biological variance exists within OTUs suggests that OTUs identified at the species level should be accepted with caution. In this study two OTUs (*Pasteurella_multocida*_OTU743 and *Escherichia_coli*_OTU726) were identified as the species *Pasteurella multocida* and *Escherichia coli.* *Pasteurella_multocida*_OTU743 was amongst those with high LDA scores, (LDA score of 3) and observed high biological relevance in the healthy community. *Escherichia_coli*_OTU726 had a lower LDA score of approximately 2.5 and was of less biological relevance than *Pasteurella_multocida*_OTU743, but of higher relevance than more than half of the OTUs identified by LEfSe as more abundant in the healthy community. *Pasteurella multocida* is a Gram-negative, non-motile, facultative anaerobic coccobacillus that is a commensal bacteria of the nasopharyngeal and gastrointestinal tract of animals such as cats and dogs, as well as poultry and other domestic and wild animals (Lalloo, 2003, Holman *et al*., 2015). *Eschericia coli* is a Gram-negative bacteria and is typically found in the mucosal surfaces of the gastrointestinal tract of cattle, sheep and goats (Weese & Fulford 2011). Both these species have previously been identified as residents of the urogenital microflora in healthy rams (Kidanemariam, 2003; Gouletsou *et al.,* 2006).

When looking at the diseased community, 19 OTUs were observed to be significantly more abundant and are shown in Figure 4.20. The OTUs with the most significant biological relevance (LDA score >4) were from the genus *Fusobacterium* and the family Aerococcaceae, with *Fusobacterium*_Unclassified _OTU494 representing the OTU with the largest LDA score amongst the 19 OTUs, thus indicating the most biological relevance of the OTUs. Both of these OTUs correlate with our results that found that the majority of difference in the relative abundance between healthy and diseased were across these genera, and these OTUs contributed the most to relative abundance in these taxa. The genus

Fusobacterium includes several species of obligate anaerobic, non-spore forming, motile or non-motile, Gram-negative rods which are found to occur naturally in the mucosal environment of animals such as the alimentary canal of pigs, the respiratory tract of a number of animals such as ruminants, horses and pigs (Jang & Hirsh, 1994) and also in the urogenital tract of animals (Nagaraja *et al*., 2005). A number of species are believed to be pathogenic in animals and are related to well characterize diseases (Hofstad, 2006). The available literature on species from the genus *Fusobacterium* that are pathogenic to animals, is based mostly on the species *Fusobacterium Necrophorum*, although other members are also known to cause disease in animals (Nagaraja *et al*., 2005). In livestock, *Fusobacterium necrophorum* cause necrotic lesions and has been found as the aetiology of a number of diseases such as footrot in sheep, cattle and goats (Zhou *et al*., 2009); liver abscesses in cattle (Tadepalli *et al.,* 2009). This bacterial species has previously been isolated in genital diseases of animals such as contagious ecthyma where they are suggested to be involved in the progressive forms of the necrotic lesions (Radostits *et al.,* 2006). Ulcerative lesions of the prepuce and penis of bison have also been associated with this species (Jakob *et al*., 2000). *Fusobacterium necrophorum* has indicated synergistic relationships with other bacteria as it is often encountered in mixed infections (Tan *et al*., 1996) and Sheldon *et al*. (2008) suggested its synergistic relationship with *Trueperella pyogenes* and *Prevotella* species in enhancing uterine disease, and increase the risk of clinical endometritis and its severity. Although it was not possible in this study to identify the species within the genus *Fusobacterium*, the high relative abundance of this genus in diseased animals may indicate some role in the disease, with *Fusobacterium necrophorum* that has previously been implicated in genital disorders.

As previously mentioned, little information exists on the family Aerococcaceae and the natural occurrence of this family within animals. They have particularly been isolated from human and veterinary clinical sources (Facklam & Elliott, 1995). Most of the information relates to the genus *Aaerococcus* which for example has been isolated from the respiratory tract of healthy rodents and rabbits (Hansen, 2000). Members within the family are identified as aetiological agent of animal diseases, such as mastitis in cows (Devriese *et al*., 1999), arthritis in pigs (Martin *et al*., 2007), septicaemia in mice (Dagnaes-Hansen *et al*., 2004) and causing gaffkemia in marine lobsters (Battison *et al*., 2004). To our knowledge, there is no information on their involvement in diseases of the urogenital tract of sheep, but one study by El-Arabi (2014) isolated an *Aeroccocus* species from the vagina of an ewe that was infected with vulvitis (the female counterpart of UB). Also members from the *Aeroccocus* genus are aetiological components of urinary tract infections in humans (Zhang *et al*., 2000).

The succeeding 6 OTUs with high LDA scores ranging between 4 and 3, were from the phylum *Bacteroidetes* with 5 OTUs associated to *Bacteroidales* and 2 OTUs associated to *Porphyromona*s which belong to the order *Bacteroidales*, with Bacteroidale_Unclassified_OTU203 displaying the highest biological relevance of the group. Two other OTUs from the phylum *Bacteroidetes,* Porphyromomonas_ unclassified_OTU 229 and Bacteroidetes_Unclassified_OTU179 were also observed to be more abundant in the diseased community, although displayed lower LDA scores of approximately 2 compared to the other OTUs from the *Bacteroidetes* phylum and thus had less

biological relevance. These results correlated with previous findings in this study that saw this phylum, order and genus having higher abundance in diseased animals, which were chiefly as a result of these OTUs. The phylum Bacteroidetes represents 7000 species which consist of Gram-negative, non-spore forming, anaerobic or aerobic, and rod-shaped bacteria which have colonized all types of habitats on earth, displaying various biological functions. They are major members of the microbiota of animals and have been thoroughly isolated from the mucosal membranes of the gastrointestinal tract in sheep (Lopes *et al*., 2015) and a large number of other animals (Thomas *et al*., 2011). They have also be identified as forming part of the urogenital tract (Jarvinen & Kinyon, 2010; Swartz *et al*., 2014) and the nasopharyngeal tract of animals (Holman *et al*., 2015). Members from this phylum of the genera *Riemerella*, *Ornithobacterium*, and *Coenonia* have also been identified causing septicemia and respiratory tract infections in birds (Segers *et al*., 1993; Vandamme *et al*., 1999).

The bacterial species from the phylum Bacteroidetes that are the most important in livestock species and are of veterinary concern are from the genera *Bacteroides, Porphyromonas* and *Prevotella* which form part of the order Bacteroidales (also identified as 4 unclassified OTUs significantly higher in diseased rams). All genera consist of Gram-negative, anaerobic bacilli. *Bacteroides*, *Prevotella* and *Porphyromonas* species have all been implicated in foot-rot in sheep (Frisken *et al*., 1987; Hurtado *et al*., 1999). A number of other diseases caused by these three genera have involved broken-mouth periodontitis in sheep (Frisken *et al*., 1987), ovine periodontitis (Dreyer *et al*., 1992) and uterine infection (Sheldon *et al*., 2008). Three OTUs that were significantly more abundant in diseased animals were from the genus *Porphyromonas.*

The pathogenicity of members of the Bacteroidetes phylum in the urogenital tract of rams has not been suggested, but a similar condition in men (balanitis), that also causes lesions on the glans penis, has been ascribed to species belonging to the genus *Bacteroides* which were indicated as responsible for the anaerobic infection (Edwards, 1996). A member from the *Porphyromonas* genus has been isolated in cases of vulvovaginitis (Elad *et al*., 2014), which causes lesions of the vulva, similar to that of ulcerative balanoposthitis.

The remaining OTUs had LDA scores ranging from 2 to 3 and were of less biological importance in the diseased community such as Neisseriaceae_Unclassified_OTU672 and Propionibacteriaceae_ Unclassified_OTU154. As in the healthy community, an OTU assigned as an unclassified *Leptotrichia* and another as an unclassified *Corynebacterium* were observed as significantly more abundant in the diseased group. Relative to the healthy group, the diseased group had a higher abundance of OTUs from the phylum Firmicutes which were associated to (in order of the OTU with the highest to the lowest LDA scores) *Facklamia,* Clostridales, Lachnospiraceae, Bacilli and Lactobacillales (Figure 4.20).

Interestingly, as indicated in Figure 4.19 an OTU from the *Corynebacterium* genus (*Corynebacterium*_unclassified _OTU77) was also identified as significantly abundant in diseased rams. Members from this genus are commensal bacteria of the mucous membranes, such as the

reproductive tract (Markey *et al*., 2004), however when the environment changes some species may become opportunistic and there is a long list of members that cause disease in animals (Moore *et al*., 2010) typically causing pyogenic infections. In particular, the species *Corynebacterium renale*, which has been described as highly adapted to the bovine and ovine urinary tract, has previously been identified as the aetiological agent of ovine posthitis, a disease of the genitals which forms lesions on the glans penis and prepuce of rams. This disease however has been associated with a high protein diet, which increases the number of urea-fixing *Corynebacterium renale* in the penile environment due to a protein-rich diet (Loste *et al*., 2005) and as a result of a high urea content in the urine, allows for the proliferation of these urea-fixing individuals. We were not able to identify the OTU to species level, but this genus should thus be kept in mind when conducting future studies to rule out this genus as a potential pathogen. *Corynebacterium* have also been isolated from cases of orchitis, a disease that causes inflammation of the scrotum in rams (Malone *et al*., 2010) and the species *Corynebacterium mastitidis* has been found in preputial gland abscesses in mice (Radaelli *et al*., 2010).

The prevalence and the distribution of the relative abundances of the LEfSe identified differentially abundant phylotypes and OTUs in samples belonging to the healthy and diseased groups were examined. Examples of the distribution in relative abundance of phylotypes are given in Figure 4.21. The relative abundance of all OTUs identified by LEfSe are visualized by means of a heatmap in Figure 4.22 and a few examples are discussed.

The distribution of the relative abundance of the genus *Arcanobacterium* (significantly abundant in healthy individuals) across healthy and diseased samples is shown in Figure 4.21A. Only one OTU was responsible for the relative abundance of *Arcanobacterium*. For the healthy group the bar chart in Figure 4.21A revealed that only two samples out of the 4 that observed the genus *Arcanobacterium* contributed to the total abundance in the healthy community. No samples were identified having organisms classified as *Arcanobacterium* in the diseased group.

The remaining phylotypes identified by LEfSe were observed in samples from both the healthy and diseased communities, ranging in the total number of samples they were prevalent in. For example, the genus *Pasteurella* was observed in 16 samples in the healthy community as compared to 9 in the diseased community (Figure 4.21B) compared to the rest of the phylotypes (Figure 4.21C and Figure 4.21D) that were observed in all 40 samples.

High inter-sample variation in the relative abundance was observed for all phylotypes, as seen in the examples in Figure 4.21. In *Pasteurella* the majority of the relative abundance was observed in only 4 out of 16 (25%) samples and 1 out of 9 (11%) samples in the healthy and diseased communities, respectively. The remaining phylotypes also had varying relative abundances across healthy and diseased samples. In Figure 4.21C and Figure 4.21D, the distribution of the relative abundance of the family Streptococcaceae and Leptotrichiacea revealed that high inter-sample variation occurred and that less than half of the samples in both the healthy and diseased communities were responsible for a

large percentage of the total relative abundance observed in both groups. For example, in Figure 4.21D the relative abundance of Leptotrichiaceae ranged from 0.0001% to 0.2029% in the diseased community and from 0.0035% to 0.2401% in the healthy community. Sixty-eight percent of the relative abundance in diseased rams were observed in 4 samples and sixty seven percent of the relative abundance in healthy rams were observed in 5 samples.

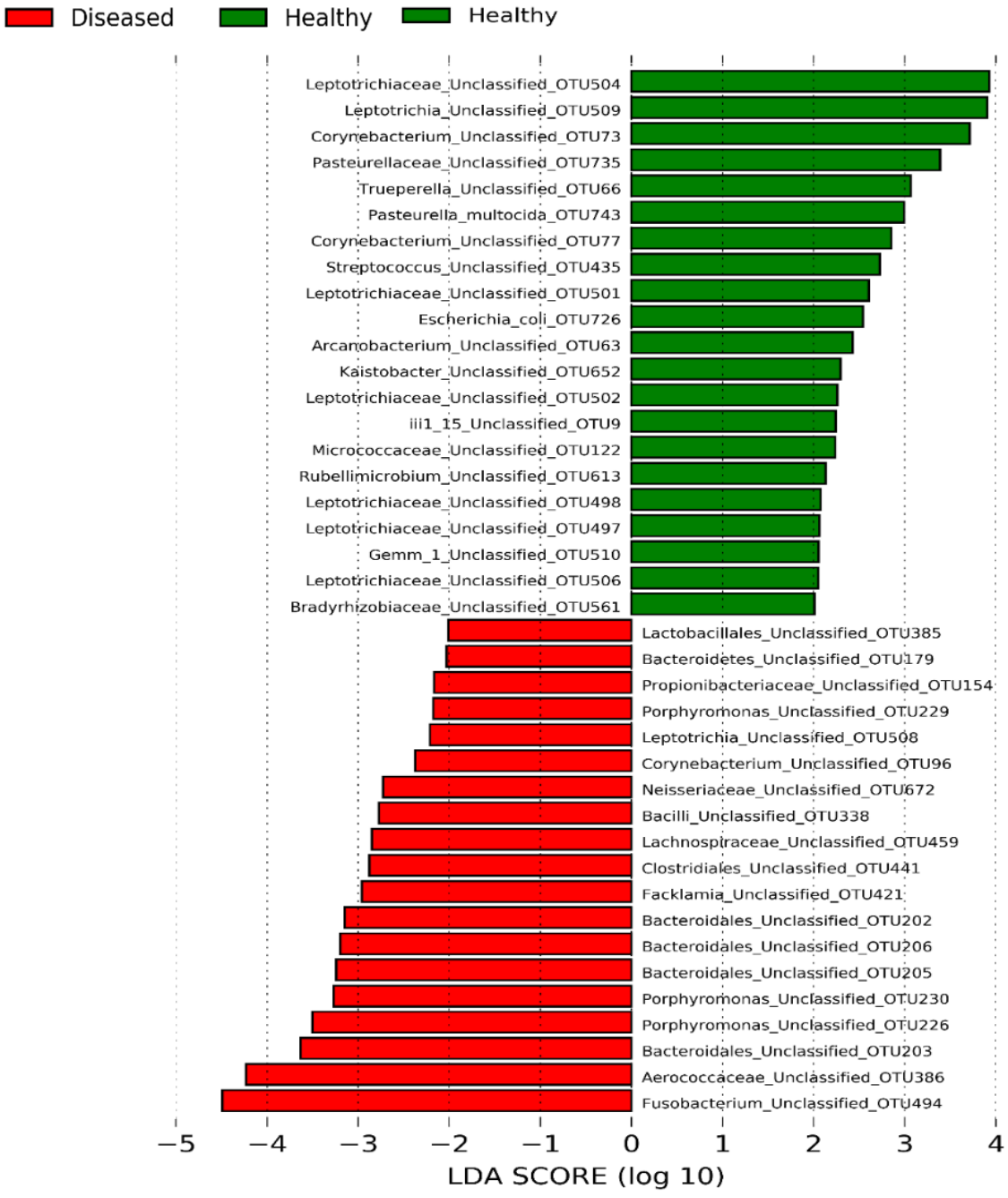

Figure 4.20. Histogram depicting the OTUs identified as significantly differential in healthy and diseased groups. Only taxa with p-value <0.05 and LDA scores ≥2 were included in the results. Taxa more abundant in the healthy group are indicated with positive LDA scores (green), and taxa more abundant in the diseased group are indicated with negative LDA scores (red).

When examining the per sample distribution in relative abundance at the OTU level, we observed that a number of OTUs were not prevalent across all samples, including the community from which they were identified as significantly more abundant. In addition, when comparing the distribution of relative abundances of OTUs across communities only a few samples, most often less than half and with some OTUs performing better than others, contributed more than 90% of the total relative abundance in each community. These results were observed in OTUs that had high and low LDA scores. Examples of the distribution of the relative abundance of 4 OTUs identified as significantly abundant in the healthy and diseased communities with high LDA scores are described in the next paragraph in no particular order. Visual representation of the distribution of the relative abundance of each OTU across healthy and diseased samples can be seen in the heatmap in Figure 4.22.

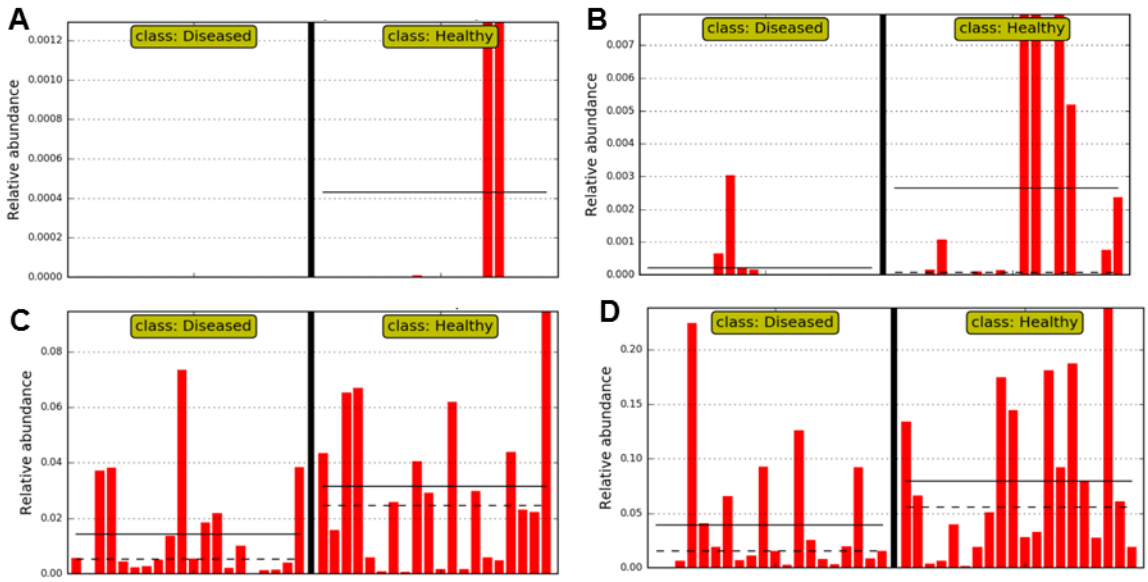

Figure 4.21. Distribution of the relative abundance of phylotypes flagged by LEfSe as significantly more abundant. Histograms showing the relative abundance of **A)** the genus *Arcanobacterium*. **B)** the genus *Pasteurella* **C)** the family *Streptococacceae* and **D)** the family *Leptotrichiaceae* The mean and median relative abundance of each phylotype is indicated with a solid and dashed line, respectively. Each bar on the x-axis represents a sample.

For example, Fusobacterium_Unclassified_OTU494 was identified by LEfSe has having significantly greater abundance in the penile microbiota of diseased rams. This OTU was identified in 19 out of 20 (95%) of the diseased rams and 17 out of 20 (85%) of the healthy rams. The percentage relative abundance in diseased rams ranged from 0.0026% to 50.12% and 0.00084% to 2.77% in the healthy rams (Figure 4.22), with 90% of the relative abundance concentrated in 4 samples in the diseased community and 5 samples making up 95% of the relative abundance in the healthy community. Another example of the distribution in the relative abundance of OTUs was Leptotrichiaceae_ Unclassified_OTU504 which obtained a high LDA score and was identified by LEfSe as significantly more abundant in the penile microbiota of healthy rams. This OTU was identified in 14 out of 20 (70%) of the diseased rams and 19 out of 20 (%) of the healthy rams ranging in their percentage relative abundance from 0.0002% to 0.29% and 0.0004% to 15.40%, respectively (Figure 4.22). In the diseased

community 6 samples made up 95% of the relative abundance and in the healthy community 6 samples made up about 98% of the relative abundance.

The relative abundance across all samples for another OTU, Aerococcacceae_Unclassified_OTU386, with a high LDA score and significantly abundant in diseased individuals was also examined and observed better results than the previously discussed OTUs. We found that this OTU was identified in all 40 samples and had a range in percentage relative abundance across the diseased and healthy samples of 0.0036% to 19.77% and 0.0163% to 17.44%, respectively (Figure 4.22). In the diseased community the range in percentage relative abundance would have been narrower, with only one sample (outlier) responsible for the large difference seen in the range (0.0036%). This OTU observed 91% of the total relative abundance in 11 samples in the diseased community and 97% in 8 samples in the healthy community.

*Trueperella*_Unclassified_OTU66 is an example of an OTU that also observed higher abundance in the healthy community with a high biological relevance (LDA score >3) but that observed poorer prevalence across all samples. This OTU was identified in 10 out of 20 (50%) diseased rams and 15 out of 20 (75%) healthy rams. *Trueperella*_Unclassified_OTU66 observed low relative abundance across all samples ranging from 0.0005% to 0.51% in diseased rams and 0.0006% to 6.21% in healthy rams (Figure 4.22). More than 90% of the relative abundance was obtained in 2 samples in both the healthy and diseased communities.

To analyse the penile microbiota of rams LEfSe was used to identify significant differences in relative abundance of phylotypes and OTUs between healthy and diseased ram populations. This study did not find any phylotypes significantly more abundant in diseased rams, however a number of OTUs were enriched in the diseased population. This preliminary microbiome study of the ram preputial and penile mucosa in the healthy and diseased population, revealed that OTUs classified as belonging to the family Aerococcaceae, the genus *Fusobacterium* and majorly from the phylum Bacteroidetes (i.e. *Porphyromonas* and Bacteroidales) may be disease-specific/ disease-associated. A number of known pathogens in animal diseases have previously been identified within these bacterial taxa but differentiating cause and effect is not possible without species-level identification and a greater understanding of pathophysiology. However, the identification of these above mentioned OTUs disproportionately present in rams with UB warrants further investigation.

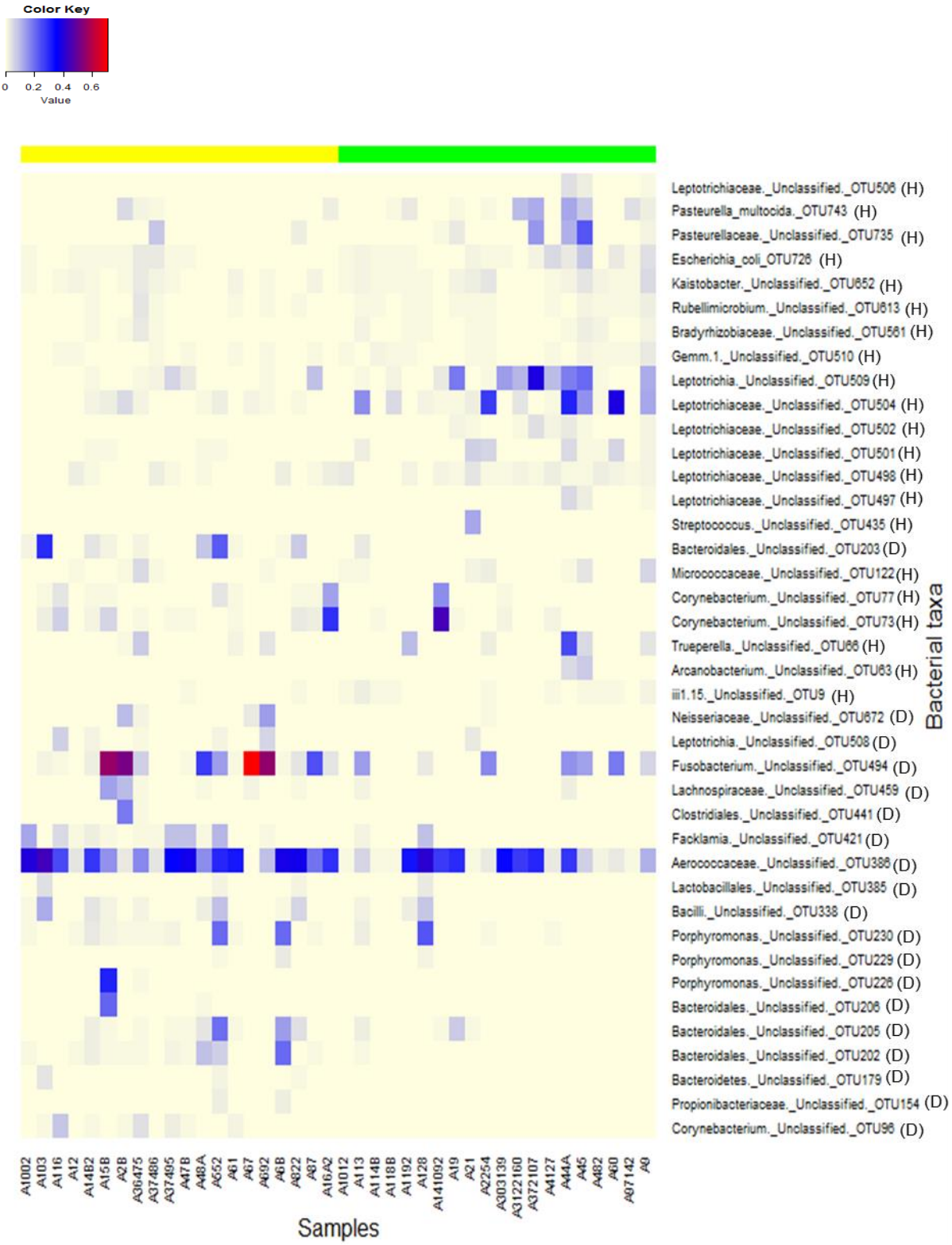

Figure 4.22. Heatmap of the square root-transformed relative abundances of OTUs flagged by LEfse as significantly abundant in healthy and diseased groups. The yellow colour bars represent the diseased samples and the green bars the healthy samples. The colour scale above the heatmap depicts the range in relative abundance, the light colour indicates a low proportion and red shows a high proportion. OTU identified as significantly more abundant in either the healthy or diseased groups are marked by a (H) or (D), respectively.

If we are to associate a significantly higher relative abundance of these OTUs with the disease phenotype, possibly suggesting their involvement in UB, one would expect that the OTUs are prevalent in all diseased samples and that the relative abundances across samples would be somewhat similar, thus the inter-sample variations observed reduces the confidence in these OTUs being involved in aetiology. However, 16S amplicon sequencing data is only semi-quantitative as a result of inherent bias brought about during PCR amplification, sequencing and data processing, thus affecting the abundance observed across samples and could reflect bigger differences than naturally occurring. It is also possible that the variations may be as a result of different aetiology amongst rams, especially if secondary bacterial infection is present and thus these variations in prevalence and relative abundance will warrant further investigation.

4.9 Classification and comparison of *Mycoplasma* and *Trueperella* species in healthy and diseased communities

Previous research on UB (Kidanemariam, 2003) in South Africa has suggested that a combination of *Trueperella pyogenes,* previously known as *Arcanobacterium pyogenes* (Yassin *et al*., 2011) and *Mycoplasma mycoides mycoides* large colony (MMMLC) or a combination of *Mycoplasma* species may be involved in the aetiology of UB (Trichard *et al*., 1993; Kidanemariam, 2003; Ali, 2012). The presence and abundance of these taxa in the penile microbiotas was therefore examined here. One of the complications with bacterial taxa being renamed accordingly over the years is that, its taxonomic classification, newer and older versions, may exist in reference databases and therefore sequences may be named after either or both separating them into *Trueperella* and *Arcanobacterium*. Thus for this analysis both the *Trueperella* and *Arcanobacterium* genera were analysed to make sure we describe the true representation of the *Trueprella pyogenes* species in the penile microbiome.

On analysis of the data we were able to identify 3, 1 and 4 OTUs belonging to the *Mycoplasma*, *Arcanobacterium* and *Trueperella* genera, respectively. Further characterization of these OTUs to species level using our pipeline was not possible, and further classification was achieved using the BLASTn tool against the GenBank reference database. The ability of the BLASTn tool and GenBank to classify the OTUs compared to the Naïve Bayesian classifier and the Greengenes reference database used in this pipeline is a direct result of differences observed in the bacterial sequences available, differences in the taxonomic naming system of bacteria and the methods used to assign taxonomy and report confidence in the prediction.

Results from the BLASTn analysis revealed 1 OTU as belonging to the species *Trueperella pyogenes. Arcanobacterium pyogenes* was not identified. Three *Mycoplasma* species, *Mycoplasma arginini*, *Mycoplasma bovigenitalium* and *Mycoplasma hyopharyngis* were identified. *Mycoplasma mycoides mycoides* large colony (MMMLC) was not identified in this study. This was surprising in that it has been associated as an aetiological agent of UB and has consistently been identified in cases of the disease in South Africa (Trichard *et al*., 1993; Kidanemariam, 2003; Ali, 2012). The taxonomic description,

percentage identity, E-value and accession number for OTUs assigned a taxonomic classification using BLASTn is shown in Table 4.6.

This study identified 3 *Mycoplasma* species as compared to 6 (Trichard *et al*., 1993; Ali, 2012) or 8 (Kidanemariam, 2003) species detected in previous studies that examined the *Mycoplasma* community present in cultures derived from the genital mucosa of healthy and diseased Dorper sheep. The lower *Mycoplasma* richness of 3 may come as a true biological representation. However, we speculate that the lower *Mycoplasma* richness in this study is due to the chosen universal V3V4 primer set which may lack conservancy across all members of the genus *Mycoplasma*. This speculation was justified by the fact that although Ali (2012) first cultured *Mycoplasma* he also used the 16S rRNA gene to identify *Mycoplasma* but amplified the 16S rRNA gene using *Mycoplasma* specific primers, and was therefore able to identify 6 species of *Mycoplasma* compared to the 3 species identified in this study. With this in mind future studies should aim to sequence the 16S rRNA gene of bacteria using primers that are universal to all bacteria and some that are specific to *Mycoplasma* in order to validate findings of this study.

Table 4.6. BLASTn taxonomic assignment of operational taxonomic units.

OTU	Taxonomic Assignment	E-value	Identity %	Accession number
OTU834287193	*Trueperella pyogenes* strain LYH 16S ribosomal RNA gene, partial sequence	0	100%	KU738726.1
OTU112723901	*Mycoplasma bovigenitalium* gene for 16S ribosomal RNA, complete sequence, strain: PG11	0	100%	LC158833.1
OTU395215088	*Mycoplasma arginini* strain Dak-2/M.arg/EG014 16S ribosomal RNA gene, partial sequence	0	100%	KP972459.1
OTU892673693	*Mycoplasma hyopharyngis* 538N 16S rRNA gene, partial sequence	0	99%	U04652.1

Prevalence of each bacterial species in healthy and diseased groups identified by BLASTn analysis is shown in Table 4.7. *Trueperella pyogenes* typically resides in the membrane of the gastrointestinal, genital and upper respiratory tract mucosa (Moore *et al*., 2010) and their high prevalence in the healthy genital tract is thus proof of this in this study. Kidanemariam (2003) also indicated a high prevalence in

healthy sheep and found that a significantly higher isolation rate occurred in diseased sheep, suggesting an association of *Trueperella pyogenes* with UB. This would not be surprising in that *Trueperella pyogenes* has been implicated in various diseases in a wide variety of animal hosts (Carlton *et al*., 2010). In agreement with Kidanemariam (2003) this study observed a higher prevalence of *Trueperella pyogenes* in the diseased group (100%) compared to the healthy one (65%).

Kidanemariam (2003) looked at the prevalence of *Trueperella pyogenes* in ewes and rams individually as well as combined. Due to this study being exclusively on rams we decided to compared the prevalence of *Trueperella pyogenes* in only rams. We found that this study isolated *Trueperella pyogenes* from a higher percentage of healthy (65% vs 13.85%) and diseased (100% vs 38.03%) rams compared to the prevalence reported by Kidanemariam (2003). The differences observed may be as a result of the larger sample size used by Kidanemariam (2003). It is also probable that the 16S rRNA gene was more specific in identifying *Trueperella p*yogenes than culture-based methods. A study by Oikonomou *et al*. (2012) that used both culture and 16S rRNA sequencing to identify bacteria associated with mastitis in cows, found that samples that were culture negative for *Trueperella pyogenes* were positively identified with *Trueperella pyogenes* using 16S rRNA sequencing.

Mycoplasma bovigenitalium (Trichard et la., 1993; Kidanemariam, 2003; Ali, 2012) and *Mycoplasma arginini* (Kidanemariam, 2003; Ali, 2012) have both previously been identified in healthy and diseased sheep in studies carried out to identify the aetiology of UB in South Africa, whilst the identification of *Mycoplasma hyopharyngis* is the first to be reported. *Mycoplasma bovigenitalium* has been implicated in genital disorders of cattle and sheep. In Nigeria, *Mycoplasma bovigenitalium* was isolated from the genital tract of ewes that experienced a similar genital disease to UB/UV (Chima *et al*., 1995). Vaginal swabs taken from granulo-pustular lesions on the vulva of Dairy cows from a herd in Israel revealed the presence of *Mycoplasma bovignetialium* in about 11/20 clinical cases (Lysnyansky *et al*., 2009). Twelve heifers inoculated with *Mycoplasma bovignetialium* demonstrated granular vulvovaginitis days after inoculation, suggesting the pathogenicity of this species in the genital tract of cows (Saed & Al-Aubaidi, 1983).

Mycoplasma arginini is a naturally occurring bacteria in the genital tract of small ruminants (Rosendal *et al*., 1994). Its capabilities as a pathogen is still not well-documented and in fact it is believed to have low pathogenicity (Leach, 1970). It has however been isolated from various diseases. Chima *et al*. (1995) found that along with *Mycoplasma bovigenitalium*, *Mycoplasma arginini* was also isolated from vaginal symptoms similar to UV in ewes from Nigeria. A study by Leach (1970) described the isolation of *Mycoplasma arginini* in a case of ovine keratoconjunctivitis and from an arthritic joint in goats. Its isolation rate in healthy versus diseased rams has not proved to be significantly different in studies of UB (Kidanemariam, 2003). However the study by Ali (2012) suggested that the isolation rate of *Mycoplasma arginini* using the 16S rRNA gene to identify *Mycoplasma* from cultures was much higher than previously reported using serological identification methods in the study by Kidanemariam (2003) and that the pathogenicity of this species should thus be reviewed.

Mycoplasma hyopharyngis was first identified and classified in 1986 by Erickson *et al*. (1986) as originating from the upper respiratory tract of pigs. Since then its occurrence in the upper respiratory tract of pigs has been validated in other studies but its occurrence is thought to be rare (Friis *et al*., 2003). Pathogenic properties of this species has been suggested by Hartmann *et al*. (2010) and Friis *et al*. (2003) who described its isolation from cases of conjunctivitis in cats and inflamed joints and adjacent subcutaneous abscesses in pigs, respectively. Little is known about its pathogenic capabilities (Kobisch & Friis, 1996) and its involvement in genital diseases has not been described. Prevalence of this *Mycoplasma* in healthy and diseased sheep and its involvement in UB will have to be validated further in future studies.

Table 4.7 shows that 20% and 50% of identified *Mycoplasma hyopharyngis* were observed in healthy and diseased rams, respectively. When comparing the prevalence of *Mycoplasma bovigenitalium* and *Mycoplasma arginini* in healthy and diseased rams from this study to that of a study by Kidanemariam (2003) that looked at their prevalence across ewes and rams combined, we found that this study had a higher prevalence across both groups. *Mycoplasma bovigenitalium* was observed in 15% versus 11.2% of healthy sheep and 30% versus 19.2% of diseased sheep in this study and Kidanemariam (2003), respectively. *Mycoplasma arginini* was observed in 25% versus 0.86% of healthy sheep and 50% versus 3.8% of diseased sheep in this study and Kidanemariam (2003), respectively. The difference in the percentage prevalence of *Mycoplasma bovigenitalium* and *Mycoplasma arginini* observed between our studies may be explained by the larger sample size (40 rams versus 120 sheep) of the study conducted by Kidanemariam (2003). However the substantial 24.14% (25%-0.86%) and 46.2% (50%-3.8%) difference in *Mycoplasma arginini* for the healthy and diseased groups, respectively, may suggest that the16S rRNA gene is more specific in identifying *Mycoplasma arginini* than culture-based serological methods. *Mycoplasma* that are cultivated are known to grow slowly and poorly even on the best *Mycoplasma* medium available and are often overgrown by other bacteria in samples (Razin, 1994; Cai *et al*., 2014)

Table 4.7. A comparison of the prevalence of *Trueperella* and *Mycoplasma* species in healthy and diseased Dorper rams.

OTU	Bacterial species	Healthy		Diseased	
		Number	%	Number	%
OTU834287193	*Trueperella pyogenes*	13	65	20	100
OTU112723901	*Mycoplasma bovigenitalium*	3	15	6	30
OTU395215088	*Mycoplasma arginini*	5	25	10	50
OTU892673693	*Mycoplasma hyopharyngis*	4	20	10	50

The range in the percentage relative abundance observed for the OTU classified as *Trueperella pyogenes* is shown in Figure 4.23. The percentage relative abundance *Trueperella pyogenes* was higher in the diseased community than in the healthy one, with a mean of 0.06% with SD of 0.001%

and 0.25% with SD of 0.005% across healthy and diseased samples, respectively. The significant difference in the percentage relative abundance of the healthy and diseased group for the OTU assigned as *Trueperella pyogenes* (Table 4.6) was tested by LEfSe analysis in section 4.9. We found that the OTU identified as *Trueperella pyogenes* was not flagged by LEfSe as significantly different between the two groups.

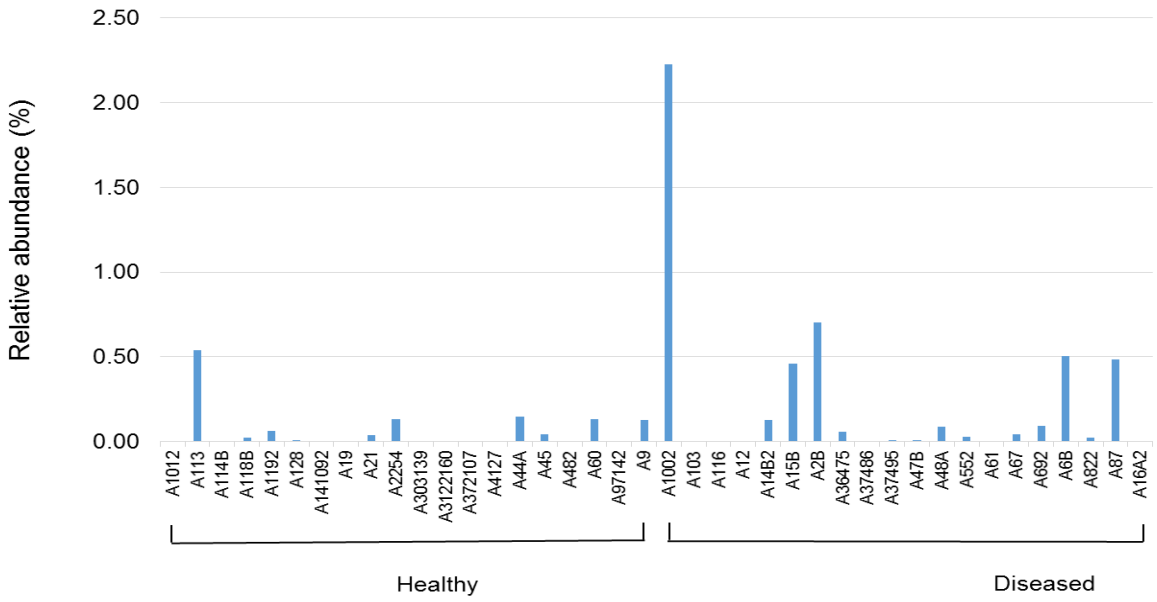

Figure 4.23. Bar graph representing the range in percentage relative abundance of the OTU identified as *Trueperella pyogenes* by BLASTn analysis, across healthy and diseased samples. Each label on the horizontal access is an individual sample.

The percentage relative abundances of the OTUs classified as *Mycoplasma hyopharyngis, Mycoplasma arginini* and *Mycoplasma bovigenatalium*, respectively for all samples is shown in Figure 4.24. The OTUs belonging to *Mycoplasma hyopharyngis, Mycoplasma arginini* and *Mycoplasma bovigenatalium* demonstrated a mean percentage relative abundance with SD of 0.04% ± 0.001%, 0.004% ± 0.0001% and 0.003% ± 0.0001%, respectively for the healthy group and 0.46% ± 0.02%, 0.06% ± 0.003% and 0.02% ± 0.0005%, respectively, for the diseased group. A higher total percentage relative abundance was experienced in the diseased community compared to the healthy one for all *Mycoplasma* species, however according to LEfSe analysis no statistically significant difference was observed for the OTUs belonging to the *Mycoplasma* genus.

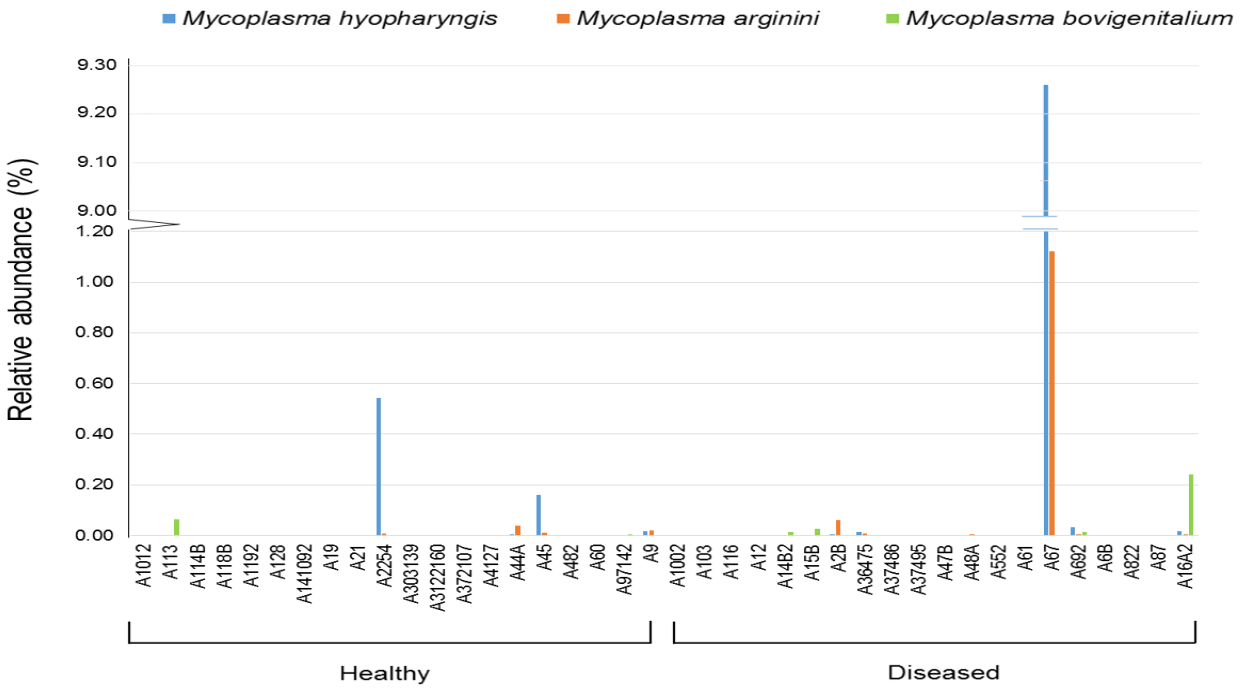

Figure 4.24. Broken axis bar graph displaying the distribution of percentage relative abundance of three OTUs identified as *Mycoplasma hyopharyngis, Mycoplasma arginini and Mycoplasma bovigenitalium* across healthy and diseased samples. The relative abundance on the y-axis stops at 1.2% and restarts at 9% due to the large inter-sample variation. Each label on the horizontal access is an individual sample.

From Figure 4.23 and 4.24 it is evident that a small number of samples are responsible for most of the OTU relative abundances observed in healthy and diseased groups. For example, for the OTU classified as *Trueperella pyogenes* samples A1002 and A113 contributed more than 40% of the relative abundance within the healthy and diseased groups, respectively (Figure 4.25). In the healthy group, the remaining approximately 60% of the relative abundance was distributed amongst 12 samples, with ~40% distributed across 4 samples. In the diseased community, the same trend was observed with 4 out of the remaining 19 samples providing more than 40% of the relative abundance.

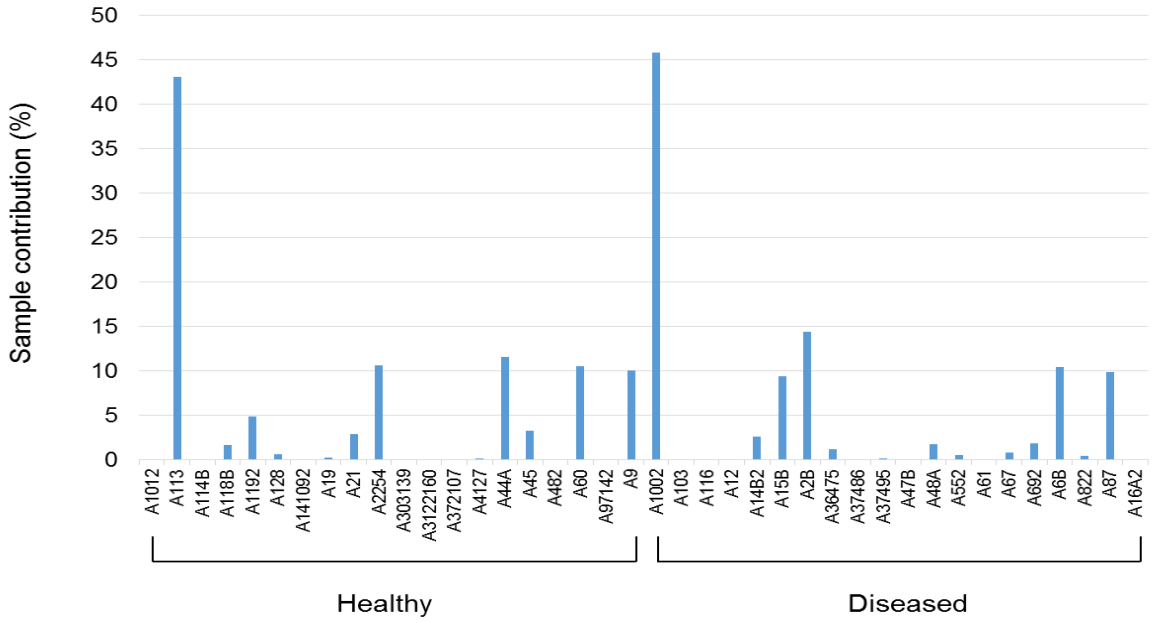

Figure 4.25. The percentage contribution of each sample to the total relative abundance of *Truperella pyogenes* observed in the healthy and diseased groups, separately. Each label on the horizontal access is an individual sample.

For the *Mycoplasma* species (Figure 4.26), samples A113 and A16A2 accounted for more than 80% of the relative abundance of *Mycoplasma bovigenitalium* in the healthy and diseased groups, respectively. Even though *Mycoplasma hyopharyngis* and *Mycoplasma arginini* were observed in more samples in the healthy and diseased groups (Table 4.5) than *Mycoplasma bovigentialium*, a similar trend was observed with only one sample, A67, accounting for ~ 99% and ~93% of the relative abundance of *Mycoplasma hyopharyngis* and *Mycoplasma arginini* in the diseased community, respectively. In the healthy group, more than 90% of the relative abundance of *Mycoplasma hyopharyngis* was distributed between 2 (A2254 and A45) of the 4 samples that observed this OTU (species). The relative abundance of *Mycoplasma arginini* in the healthy community was distributed amongst 5 samples with 76% of the relative abundance concentrated in 2 samples (A44A and A9).

This study revealed that *Trueperella pyogenes, Mycoplasma arginini, Mycoplasma bovigenatalium* and *Mycoplasma hyopharyngis* were not associated with UB. Kidanemariam (2003) based significant findings on prevalence in healthy and diseased rams. In contrast this study compared the relative abundance of organisms across healthy and diseased groups as a measure of association, which was the first use of this method in the study of the aetiology of UB which can account for the differences in significant results observed. It is also possible that the lack of significant differences observed between healthy and diseased samples for the *Trueperella pyogenes* OTU and the 3 OTUs identified as *Mycoplasma* species is likely due to the high variation in the percentage relative abundance of these organisms across samples, as well as the small sample size which affects the statistical power.

The relative abundance of these organisms in healthy and UB affected Dorper rams has not previously been described, and again it is not possible to determine whether the inter-sample variation in relative abundance is a true biological representation. As 16S amplicon sequencing sequence data can only be used as a semi-quantitative measure of relative abundance, the inter-sample variations of these bacterial taxa and the non-association of these organisms to UB should be validated through additional 16S amplicon studies of the penile microbiota on a greater number of samples from healthy and diseased Dorper rams using 16S high-throughput next generation sequencing. Additionally, more accurate quantification methods to determine bacterial abundance per sample such as qPCR and ddPRC (Robin *et al*., 2016) should also be carried out.

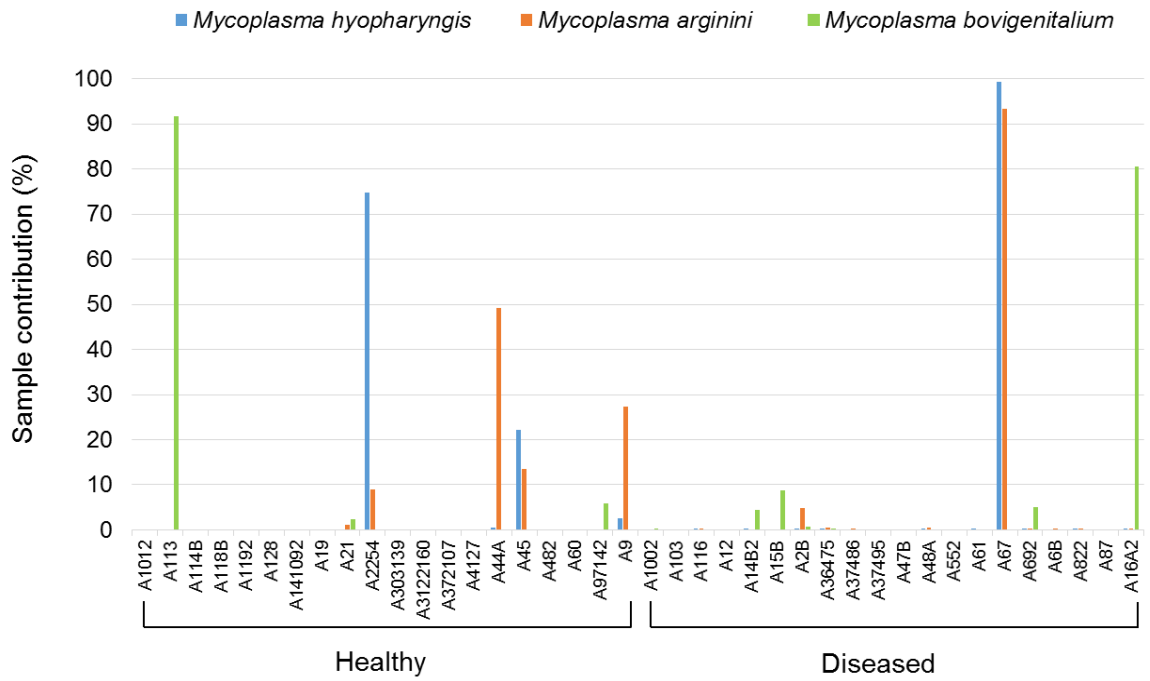

Figure 4.26. The percentage contribution of each sample to the total relative abundance of *Mycoplasma hyopharyngis, Mycoplasma arginini and Mycoplasma bovigenitalium* observed in the healthy and diseased groups, separately. Each label on the horizontal access is an individual sample.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403-410.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc [2016, 30 October].

Bath, G. & de Wet, J. 2000. Pizzle disease, in *Sheep and Goat Diseases*. Cape Town: Tafelberg Publishers Limited. 74-76.

Battison, A. L., Cawthorn, R. J. & Horney, B. 2004. Response of American lobsters Homarus americanus to infection with a field isolate of Aerococcus viridans var. homari (Gaffkemia): survival and hematology. *Diseases of Aquatic Organisms*, 61:263-268.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. 2013. GenBank. *Nucleic Acids Research*, 41(D1):D36-D42, doi:10.1093/nar/gks1195.

Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A., Knight, R. & Fierer, N. 2011. The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biology and Biochemistry*, 43(7):1450-1455.

Bik, E.M., Costello, E.K., Switzer, A.D., Callahan, B.J., Holmes, S.P., Wells, R.S., Carlin, K.P., Jensen, E.D., Venn-Watson S. & Relman, D.A. 2016. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. *Nature Communications*, 7:e10516, doi: 10.1038/ncomms10516.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A. & Caporaso, J.G. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10(1): 57-59.

Brooks, J.P., Edwards, D.J., Harwich, M.D., Rivera, M.C., Fettweis, J.M., Serrano, M.G., Reris, R.A., Sheth, N.U., Huang, B., Girerd, P. & Strauss, J.F. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*, 15(1):1.

Cai, H.Y., Caswell, J.L. & Prescott, J.F. 2014. Nonculture molecular techniques for diagnosis of bacterial disease in animals a diagnostic laboratory perspective. *Veterinary Pathology Online*, 51(2):341-350.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1):1.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N. & Knight, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Science,* 108:4516–4522.

Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods,* 69(2):330-339.

Chima, J.C., Ojo, M.O. & Adetosoye, A. I. 1992. Association of *Mycoplasma*s with genital tract infection of goats in Nigeria. Proceeding of the West African Commonwealth Veterinary Association/ Ghana Veterinary Medical Association. Accra International Conference Centre, Ghana. 87-100.

Collins, M.D., Jovita, M.R., Lawson, P.A., Falsen, E. & Foster, G. 1999. Characterization of a novel Gram-positive, catalase-negative coccus from horses: description of Eremococcus coleocola gen. nov. sp. nov. International. *Journal of Systematic and Evolutionary Microbiology,* 49(4):1381-1385.

Costa, M.C., Arroyo, L.G., Allen-Vercoe, E., Stämpfli, H.R., Kim, P.T., Sturgeon, A. & Weese, J.S. 2012. Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3-V5 region of the 16S rRNA gene. *PloS One*, 7(7):p.e41484, doi.org/10.1371/journal.pone.0041484.

Dagnaes-Hansen, F., M. Kilian & K. Fuursted. 2004. Septicaemia associated with an Aerococcus viridans infection in immunodeficient mice. *Laboratory Animals*, 38:321–325.

Degnan, P.H. & Ochman, H. 2012. Illumina-based analysis of microbial community diversity. *The ISME Journal*, 6(1):183-194.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology,* 72(7):5069-5072.

Devriese, L. A., Hommez, J., Laevens, H., Pot, B., Vandamme, P. & Haesebrouck, F. 1999. Identification of aesculin-hydrolyzing streptococci, lactococci, aerococci, and enterococci from subclinical intramammary infections in dairy cows. *Veterinary Microbiology*, 70:87–94.

Dewhirst, F.E., Klein, E.A., Thompson, E.C., Blanton, J.M., Chen, T., Milella, L., Buckley, C.M., Davis, I.J., Bennett, M.L. & Marshall-Jones, Z.V. 2012. The Canine Oral Microbiome. *PLoS One*, 7(4):e36067, doi: 10.1371/journal.pone.0036067.

Diaz, P.I., Dupuy, A.K., Abusleme, L., Reese, B., Obergfell, C., Choquette, L., Dongari-Bagtzoglou, A., Peterson, D.E., Terzi, E. & Strausbaugh, L.D. 2012. Using high throughput sequencing to explore the biodiversity in oral bacterial communities. *Molecular Oral Microbiology,* 27(3):182-201.

Dreyer, W.P. & Basson, N.J. 1992. Porphyromonas gingivalis as putative pathogen in ovine periodontitis. The Journal of the Dental Association of South Africa= Die Tydskrif van die Tandheelkundige Vereniging van Suid-Afrika, 47(12):513-516.

Dubourg, G., Lagier, J.C., Armougom, F., Robert, C., Hamad, I., Brouqui, P. & Raoult, D. 2013. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *European Journal of Clinical Microbiology & Infectious Diseases*, 32(5):637-645.

Edgar, R.C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10):996-998.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics,* 27(16):2194-2200.

Edwards, S. 1996. Balanitis and balanoposthitis: a review. *Genitourinary Medicine,* 72(3):155-159.

Eisenberg, T., Fawzy, A., Nicklas, W., Semmler, T. & Ewers, C. 2016. Phylogenetic and comparative genomics of the family Leptotrichiaceae and introduction of a novel fingerprinting MLVA for Streptobacillus moniliformis. *BMC Genomics*, 17(1):864.

Elad, D., Friedgut, O., Alpert, N., Stram, Y., Lahav, D., Tiomkin, D., Avramson, M., Grinberg, K. & Bernstein, M. 2004. Bovine necrotic vulvovaginitis associated with Porphyromonas levii. *Emerging Infectious Disease*, 10(3):505-507.

El-Arabi, A.A., Taylor, D.J., Logue, D.N., Benothman, M. & Abuargob, O. 2014. Clinical and bacteriological findings of balanoposthitis and vulvitis in Scottish sheep. *Journal of Veterinary Advances,* 1(4):481-487.

Erickson, B.Z., Ross, R.F., Rose, D.L., Tully, J.G. & Bove, J.M. 1986. *Mycoplasma* hyopharyngis, a new species from swine. International. *Journal of Systematic and Evolutionary Microbiology,* 36(1):55-59.

Facklam, R.R. & Elliott, J.A. 1995. Identification, classification, and clinical relevance of catalase-negative, gram-positive cocci, excluding the streptococci and enterococci. Clinical Microbiology Reviews 8, 479–495.

Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M. & Ravel, J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome,* 2(1):1.

Farkas, D.H. 1993. Quality control, in D.H. Farkas (ed.) *Molecular Biology and Pathology: A Guidebook for Quality Control*. USA: Academic Press Inc. 51-64.

Fitzgerald, C., Nachamkin, I., Murray, P., Baron, E., Jorgensen, J., Landry, M. & Pfaller, M. 2006. Campylobacter and Arcobacter. *Manual of Clinical Microbiology,* Vol 1. Washington: American Society for Microbiology: 933-46.

Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M.J. & Cotter, P.D., 2016. 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiology,* 16(1):123.

Friis, N.F., Ahrens, P., Hagedorn-Olsen, T., Nielsen, E.O. & Kokotovic, B. 2003. *Mycoplasma* hyopharyngis isolation from swine. *Acta Veterinaria Scandinavica,* 44(2):1.

Frisken, K.W., Tagg, J.R., Laws, A.J. & Orr, M.B. 1987. Black-pigmented Bacteroides associated with broken-mouth periodontitis in sheep. *Journal of Periodontal Research*, 22(2):156-159.

Gloor, G.B., Hummelen, R., Macklaim, J.M., Dickson, R.J., Fernandes, A.D., MacPhee, R. & Reid, G. 2010. Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PloS one*, 5(10), p.e15406, doi.org/10.1371/journal.pone.0015406.

Gouletsou, P.G., Fthenakis, G.C., Tzora, A., Cripps, P.J. & Saratsis, P. 2006. Isolation of Arcanobacterium pyogenes from the scrotal skin and the prepuce of healthy rams or from rams with testicular abnormalities. *Small Ruminant Research,* 63(1):177-182.

Gummow, B. & Staley, G. 2000. A retrospective survey of ulcerative balanoposthitis and vulvovaginitis in South African Dorper sheep. *Livestock Health and Production,* 12:31-36.

Hamady, M. & Knight, R. 2009. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research,* 19(7):1141-1152.

Hang, J., Desai, V., Zavaljevski, N., Yang, Y., Lin, X., Satya, R.V., Martinez, L.J., Blaylock, J.M., Jarman, R.G., Thomas, S.J. & Kuschner, R.A. 2014. 16S rRNA gene pyrosequencing of reference and clinical samples and investigation of the temperature stability of microbiome profiles. *Microbiome,* 2(1):1.

Hartmann, A.D., Hawley, J., Werckenthin, C., Lappin, M.R. & Hartmann, K. 2010. Detection of bacterial and viral organisms from the conjunctiva of cats with conjunctivitis and upper respiratory tract disease. *Journal of Feline Medicine and Surgery,* 12(10):775-782.

Henderson, G., Cox, F., Kittelmann, S., Miri, V.H., Zethof, M., Noel, S.J., Waghorn, G.C. & Janssen, P.H. 2013. Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One,* 8(9): p.e74787, doi.org/10.1371/journal.pone.0074787.

Hofstad, T. 2006. The genus fusobacterium, in M. Dworkin, s. Falkow, e. Rosenberg, K.H. Schleifer & E. Stackerbrandt (eds.).*The prokaryotes* (4th ed.), Vol 7. New York: Springer. 1016-1027.

Holman, D.B., Timsit, E. & Alexander, T.W. 2015. The nasopharyngeal microbiota of feedlot cattle. *Scientific Reports*, 5.

Hoyles, L. 2014. The genus Facklamia, in W. H. Holzapfel and B. J.B. Wood (eds.), *Lactic Acid Bacteria: Biodiversity and Taxonomy*. Chichester:John Wiley & Sons, Ltd. 91-98.

Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A. & Sogin, M.L. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics*, 4(11):e1000255, doi:10.1371/journal.pgen.1000255

Jakob, W., Schröder, H.D., Rudolph, M., Krasinski, Z.A., Krasinska, M., Wolf, O., Lange, A., Cooper, J.E. & Frölich, K. 2000. Necrobacillosis in free-living male European bison in Poland. *Journal of Wildlife Diseases*, 36(2), pp.248-256.

Janda, J.M. & Abbott, S.L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology,* 45(9):2761-2764.

Jarvinen, J.A. & Kinyon, J.M. 2010. Preputial microflora of llamas (Lama glama) and alpacas (Vicugna pacos). *Small Ruminant Research*, 90(1):156-160.

Jumpstart Consortium Human Microbiome Project Data Generation Working Group. 2012. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PloS One*, 7(6), p.e39315.

Khoruts, A., Dicksved, J., Jansson, J.K. & Sadowsky, M.J. 2010. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *Journal of Clinical Gastroenterology,* 44(5):354-360.

Kidanemariam, A. 2003. Identification and characterization of the primary infectious agents associated with ovine ulcerative balanoposthitis and vulvovaginitis in South Africa. Master's thesis. Pretoria: University of Pretoria.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glöckner, F.O. 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research,* 1–11.

Knudsen, L.R.V., Karstrup, C.C., Pedersen, H.G., Angen, Ø., Agerholm, J.S., Rasmussen, E.L., Jensen, T.K. & Klitgaard, K., 2016. An investigation of the microbiota in uterine flush samples and endometrial biopsies from dairy cows during the first 7 weeks postpartum. *Theriogenology,* 86(2):642-650.

Kobisch, M. & Friis, N.F. 1996. Swine mycoplasmoses. *Revue Scientifique et Technique,* 15(4):1569-1605.

Krohn, A., Stevens, B., Robbins-Pianka, A., Belus, M., Allan, G.J. & Gehring, C. 2016. Optimization of 16S amplicon analysis using mock communities: implications for estimating community diversity. *PeerJ Preprints*, No. e2196v3.

Kumar, P.S., Brooker, M.R., Dowd, S.E. & Camerlengo, T. 2011. Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. *PloS One*, 6(6):e20956, doi:10.1371/journal.pone.0020956.

Laguardia-Nascimento, M., Branco, K.M.G.R., Gasparini, M.R., Giannattasio-Ferraz, S., Leite, L.R., Araujo, F.M.G., de Matos Salim, A.C., Nicoli, J.R., de Oliveira, G.C. & Barbosa-Stancioli, E.F. 2015. Vaginal microbiome characterization of Nellore cattle using metagenomic analysis. *PloS One,* 10(11):p.e0143294, doi.org/10.1371/journal.pone.0143294.

Lalloo, D. 2003. D.A. Yersinia, Pasteurella, and Francisella, in J.D. Warrell, T.M. Cox & J.D. Firth (eds.), *Oxford Textbook of Medicine* (4th ed.), Vol 1. Oxford: Oxford University Press. 533-537.

Lawson, P.A. 2014. The family Aerococcaceae, in W.H. Holzapfel & B.J. Wood (eds.), *Lactic Acid Bacteria: Biodiversity and Taxonomy*. Chichester:John Wiley & Sons. 71-74.

Leach, R. H., 1970. The occurrence of *Mycoplasma* arginini in several animal hosts. *The Veterinary Record,* 87:319-320.

Liu, C.M., Hungate, B.A., Tobian, A.A., Serwadda, D., Ravel, J., Lester, R., Kigozi, G., Aziz, M., Galiwango, R.M., Nalugoda, F. and Contente-Cuomo, T.L., 2013. Male circumcision significantly reduces prevalence and load of genital anaerobic bacteria. *MBio*, 4(2):p.e00076-13, doi: 10.1128/mBio.00076-13.

Lluch, J., Servant, F., Païssé, S., Valle, C., Valière, S., Kuchly, C., Vilchez, G., Donnadieu, C., Courtney, M., Burcelin, R. & Amar, J. 2015. The characterization of novel tissue microbiota using an optimized 16S metagenomic sequencing pipeline. *PloS One*, 10(11), p.e0142334, doi.org/10.1371/journal.pone.0142334.

Lozupone, C. & Knight, R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12):8228-8235.

Mahalanabis, M., Al-Muayad, H., Kulinski, M.D., Altman, D. & Klapperich, C.M. 2009. Cell lysis and DNA extraction of gram-positive and gram-negative bacteria from whole blood in a disposable microfluidic chip. *Lab on a Chip*, 9(19):2811-2817.

Mahalanabis F.E., Hartley, H.M. & Skuce, R.A. 2010. Bacteriological examinations in sheep health management. *Small Ruminant Research,* 92(1):78-83.

Markey, B., Leonard, F., Archambault, M., Cullinane, A. & Maguire, D. 2004. Bacteriology. Clinical veterinary microbiology (2nd ed). *Elsevier Health Sciences*.118-345.

Martin, V., Vela, A.I., Gilbert, M., Cebolla, J., Goyache, J., Domínguez, L. & Fernández-Garayzábal, J.F. 2007. Characterization of Aerococcus viridans isolates from swine clinical specimens. *Journal of Clinical Microbiology,* 45(9):3053-3057.

May, A., Abeln, S., Crielaard, W., Heringa, J. & Brandt, B.W. 2014. Unravelling the outcome of 16S rDNA-based taxonomy analysis through mock data and simulations. *Bioinformatics,* 30(11):1530-1538.

Moore, R., Miyoshi, A., Pacheco, L.G.C., Seyffert, N. & Azevedo, V. 2010. Corynebacterium and Arcanobacterium, in C.L. Carlton, J.F. Prescott, G. Songer & C.O. Thoen, *Pathogenesis of Bacterial Infections in Animals.* Iowa: Blackwell Publishing. 133-141.

Nagaraja, T.G., Narayanan, S.K., Stewart, G.C. & Chengappa, M.M. 2005. Fusobacterium necrophorum infections in animals: pathogenesis and pathogenic mechanisms. *Anaerobe,* 11(4):239-246.

Narrowe, A.B., Albuthi-Lantz, M., Smith, E.P., Bower, K.J., Roane, T.M., Vajda, A.M. & Miller, C.S. 2015. Perturbation and restoration of the fathead minnow gut microbiome after low-level triclosan exposure. *Microbiome,* 3(6):1.

Navas-Molina, J.A., Peralta-Sánchez, J.M., González, A., McMurdie, P.J., Vázquez-Baeza, Y., Xu, Z., Ursell, L.K., Lauber, C., Zhou, H., Song, S.J. & Huntley, J. 2013. Advancing our understanding of the human microbiome using QIIME. *Methods in Enzymology,* 531:.371.

Nelson, D.E., Dong, Q., Van Der Pol, B., Toh, E., Fan, B., Katz, B.P., Mi, D., Rong, R., Weinstock, G.M., Sodergren, E. & Fortenberry, J.D. 2012. Bacterial communities of the coronal sulcus and distal urethra of adolescent males. *PloS One,* 7(5):p.e36298, doi.org/10.1371/journal.pone.0036298.

Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D. & Konstantinidis, K.T. 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One*, 9(4):e93827, doi:10.1371/journal.pone.0093827.

Price, L.B., Liu, C.M., Johnson, K.E., Aziz, M., Lau, M.K., Bowers, J., Ravel, J., Keim, P.S., Serwadda, D., Wawer, M.J. & Gray, R.H. 2010. The effects of circumcision on the penis microbiome. *PloS One*, 5(1):p.e8422, doi.org/10.1371/journal.pone.0008422.

Quinn, P.J., Markey, B.K., Leonard, F.C., FitzPatrick, E.S., Fanning, S. & Hartigan, P. 2011a. Section III: Campylobacter and Helicobacter species. *Veterinary Microbiology and Microbial Disease*. London: Wiley-Blackwell. 342-350.

Quinn, P.J., Markey, B.K., Leonard, F.C., FitzPatrick, E.S., Fanning, S. & Hartigan, P. 2011b. Section III: The Actinobacteria. *Veterinary Microbiology and Microbial Disease*. London: Wiley-Blackwell. 196-206.

Quinn, P.J., Markey, B.K., Leonard, F.C., FitzPatrick, E.S., Fanning, S. & Hartigan, P. 2011c. Section III: Corynebacterium species. *Veterinary Microbiology and Microbial Disease.* London: Wiley-Blackwell. 207-213.

Radaelli, E., Manarolla, G., Pisoni, G., Balloi, A., Aresu, L., Sparaciari, P., Maggi, A., Caniatti, M. & Scanziani, E. 2010. Suppurative adenitis of preputial glands associated with Corynebacterium mastitidis infection in mice. *Journal of the American Association for Laboratory Animal Science,* 49(1):69-74.

Radostits, O.M., Gay, C.C., Hinchcliff, K.W. & Constable, P.D. (eds.). 2006. *Veterinary medicine: a textbook of the diseases of cattle, horses, sheep, pigs and goats*. New York: Saunders Elsevier.

Razin, S. 1994. DNA probes and PCR in diagnosis of *Mycoplasma* infections. *Molecular and Cellular Probes*, 8(6):497-511.

Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E. and Shay, J.W. 2016. Comparison of DNA Quantification Methods for Next Generation Sequencing. *Scientific reports*, 6: e24067, doi:10.1038/srep24067.

Robles, C., Paramidani, M. & Armando, S. 2016. Balanopostitis y vulvitis ulcerativa infecciosa en ovinos Merino de la Patagonia Argentina. *Veterinaria,* 33(338):1852-317X.

Rosendal, S. 1994. Ovine and caprine *Mycoplasma*s, in H.W. Whitford, R.F. Rosenbusch & L.H. Lauerman (Eds.). *Mycoplasmas in Animals: Laboratory Diagnosis.* Ames:Iowa State University Press. 84-92.

Saed, O.M. & Al-Aubaidi, J.M. 1983. Infertility in heifers caused by pathogenic strain of *Mycoplasma* bovigenitalium. *Cornell Veterinarian,* 73:125-130.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T. & Quince, C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, 43(6):e37, doi:10.1093/nar/gku1341.

Schloss, P.D., Gevers, D. & Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS One*, 6(12):e27310, doi:10.1371/journal.pone.0027310.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. & Huttenhower, C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6):1.

Segers, P., Mannheim, W., Vancanneyt, M., De Brandt, K., Hinz, K.H., Kersters, K. & Vandamme, P. 1993. Riemerella anatipestifer gen. nov., comb. nov., the causative agent of septicemia anserum exsudativa, and its phylogenetic affiliation within the Flavobacterium-Cytophaga rRNA homology group. *International Journal of Systematic and Evolutionary Microbiology*, 43(4):768-776.

Sheldon, I.M., Williams, E.J., Miller, A.N., Nash, D.M. & Herath, S. 2008. Uterine diseases in cattle after parturition. *The Veterinary Journal,* 176(1):115-121.

Sibley, C.D., Grinwis, M.E., Field, T.R., Eshaghurshan, C.S., Faria, M.M., Dowd, S.E., Parkins, M.D., Rabin, H.R. & Surette, M.G. 2011. Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PloS One,* 6(7):e22702, doi:10.1371/journal.pone.0022702.

Sinclair, L., Osman, O.A., Bertilsson, S. and Eiler, A., 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. *PloS One*, 10(2):p.e0116955, doi.org/10.1371/journal.pone.0116955.

Singer, E., Andreopoulos, B., Bowers, R.M., Lee, J., Deshpande, S., Chiniquy, J., Ciobanu, D., Klenk, H.P., Zane, M., Daum, C. & Clum, A. 2016. Next generation sequencing data of a defined microbial mock community. *Scientific Data*, 3:e160081, doi:10.1038/sdata.2016.81.

Sloan, D.B., Bennet, G.M., Engel, P. & Williams, D. 2013. Disentangling Associated Genomes in E.F. DeLong (ed.). *Methods in Enzymology.* Vol 531. USA: Academic press. 445-465.

Soetaert, K. & Heip, C.H.R. 1990. Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Marine Ecology Progress Series*, 59:305-307.

Sturgeon, A., Pinder, S.L., Costa, M.C. & Weese, J.S. 2014. Characterization of the oral microbiota of healthy cats using next-generation sequencing. *The Veterinary Journal,* 201(2):223-229.

Swartz, J.D., Lachman, M., Westveer, K., O'Neill, T., Geary, T., Kott, R.W., Berardinelli, J.G., Hatfield, P.G., Thomson, J.M., Roberts, A. & Yeoman, C.J. 2014. Characterization of the vaginal microbiota of ewes and cows reveals a unique microbiota with low levels of lactobacilli and near-neutral pH. *Frontiers in Veterinary Science,* 1.

Tadepalli, S., Narayanan, S.K., Stewart, G.C., Chengappa, M.M. & Nagaraja, T.G. 2009. Fusobacterium necrophorum: a ruminal bacterium that invades liver to cause abscesses in cattle. *Anaerobe,* 15(1):36-43.

Takamatsu, D., Ide, H., Osaki, M. & Sekizaki, T. 2006. Identification of Facklamia sourekii from a lactating cow. *Journal of Veterinary Medical Science*, 68(11):1225-1227.

Tan, Z.L., Nagaraja, T.G., Chengappa, M.M., 1996. Fusobacterium necrophorum infections: Virulence factors, pathogenic mechanism and control measures. *Veterinary Research Communications* 20, 113-140.

Toe, F., Lahlou-Kassi, A. & Mukasa-Mugerwa, E. 1994. Semen characteristics of Ile-De-France rams of different age and physical condition. *Theriogenology*, 42(2):321-326.

Tremblay, J., Singh, K., Fern, A., Kirton, E.S., He, S., Woyke, T., Lee, J., Chen, F., Dangl, J.L. & Tringe, S.G. 2015. Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology,* 6:e771, doi:10.3389/fmicb.2015.00771.

Trichard, C.J.V., Jordaan, P., Prozesky, L., Jacobsz, E.P. & Henton, M.M. 1993. The identification of *Mycoplasma* mycoides mycoides LC as the aetiological agent of balanoposthitis and vulvovaginitis in sheep in South Africa. *The Onderstepoort Journal of Veterinary Research*, 60:29–37.

Van Vuuren, M. & Trichard C.J.V. 2004. Ulcerative balanoposthitis and vulvovaginitis in sheep and goats, in J.A.W. Coetzer & R. C. Tustin (eds.). *Infectious Diseases of Livestock with Special Reference to Southern Africa* (2nd ed.), Vol 3. Cape Town: Oxford University Press. 2151-2156.

Verdier-Metz, I., Gagne, G., Bornes, S., Monsallier, F., Veisseire, P., Delbès-Paus, C. & Montel, M.C., 2012. Cow teat skin, a potential source of diverse microbial populations for cheese production. *Applied and Environmental Microbiology,* 78(2):326-333.

Vetrovsky, T. & Baldrian, P. 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PloS One*, 8(2):p.e57923, doi.org/10.1371/journal.pone.0057923.

Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology,* 73(16):5261-5267.

Weese, J.S. & Fulford, M. 2011. Bacterial Diseases, in J.S. Weese & M., Fulford, *Companion Animal Zoonoses*. Oxford: Wiley-Blackwell. 109-240.

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R. & Ley, R.E. 2012. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *The ISME Journal*, 6(1):94-103.

Wesolowska-Andersen, A., Bahl, M.I., Carvalho, V., Kristiansen, K., Sicheritz-Pontén, T., Gupta, R. & Licht, T.R. 2014. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome,* 2(1):1.

Wilson, M. 2005. An introduction to the human-microbe symbiosis. *Microbial inhabitants of humans: Their Ecology and Role in Health and Disease.* Cambridge: Cambridge University Press.1-47.

Workman, S.N., Mathison, G.E. & Lavoie, M.C. 2005. Pet dogs and chicken meat as reservoirs of Campylobacter spp. in Barbados. *Journal of Clinical Microbiology,* 43(6):2642-2650.

Yassin AF, Hupfer H, Siering C & Schumann P. 2011. Comparative chemotaxonomic and phylogenetic studies on the genus Arcanobacterium Collins *et al*. 1982 emend. Lehnen *et al*. 2006: proposal

for Trueperella gen. nov. and emended description of the genus Arcanobacterium. *International Journal Systematic Evolution Microbiology,* 61:1265-1274.

Yuan S, Cohen DB, Ravel, Abdo, Z. & Fornery, L.J. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One,* 7:e33865, doi.org/10.1371/journal.pone.0033865.

Zhang, Q., Kwoh, C., Attorri, S. & Clarridge, J.E. 2000. Aerococcus urinae in urinary tract infections. *Journal of Clinical Microbiology,* 38(4):1703-1705.

Zheng, W., Tsompana, M., Ruscitto, A., Sharma, A., Genco, R., Sun, Y. & Buck, M.J. 2015. An accurate and efficient experimental approach for characterization of the complex oral microbiota. *Microbiome,* 3(1):1.

Zhou, H.W., Li, D.F., Tam, N.F.Y., Jiang, X.T., Zhang, H., Sheng, H.F., Qin, J., Liu, X. & Zou, F. 2011. BIPES, a cost-effective high-throughput method for assessing microbial diversity. *The ISME Journal,* 5(4):741-749.

Chapter 5

General conclusions and recommendations

This study is the first culture-independent study to characterize the penile microbiome of rams and specifically Dorper rams, providing the most complete description of such a bacterial community to date. The use of 16S amplicon sequencing of the V3V4 hypervariable region of the 16S rRNA gene proved to be a powerful method for evaluating the microbiota present, revealing a high bacterial diversity in the penile environment, higher than previously reported and isolated using culture-based bacterial identification methods. The most relatively abundant taxa (≥1%) identified were shared between the healthy and diseased groups with the genus *Corynebacterium* indicating the most dominant in both the healthy and diseased rams. The prevalence of this genus in all rams sampled as well as its high relative abundance (20.9%) in healthy rams compared to all other genera may suggest this genus as forming part of the core microbiome of the penile environment of rams. This finding should be validated in future studies and its role in the establishment and maintenance of a healthy penile environment determined.

No differences in the bacterial diversity were observed between healthy and diseased rams. No single or a few OTUs (i.e. specific to the diseased group) could therefore be definitively identified as a potential aetiological agent(-s). The disease was also not associated with significant changes in microbiota composition, and the overall composition remained similar in healthy and diseased rams. Fifty percent of our data set was classified to the genus level. The use of a longer DNA fragment or a different hypervariable region may have increased the number of sequences classified at this taxonomic level and potentially may have provided different results. It is also possible that significant changes in community composition occur at the species level rather than at higher taxonomic levels. However, the use of 16S amplicon sequencing is limited in its ability to classify sequences to the species level and also with high confidence and it is not possible to test this theory until newer sequencing technology is developed that can sequence longer DNA fragments at the same sequencing depth.

In addition, high inter-sample variation in the prevalence and relative abundance of OTUs was observed and this could be masking differences in community composition between healthy and diseased ram populations. It is unclear whether the high inter-sample variation observed is a true biological representation of the bacterial community in the penile environment of rams as this is the first 16S amplicon study conducted, and the first time that the relative abundance of bacteria present in this environment has been measured. A number of biological and environmental factors may be implicated in the high inter-sample variation observed. For example things such as the differences in the age of the rams sampled, possible differences brought about by sampling from white head versus black head Dorper rams, the different time period in which rams were sampled, the different stages of the disease progression at which rams were sampled and the different management system used at each farm sampled. The latter include factors such as mating practices and feeding regime. Although these variables (i.e. the type of rams, age, disease onset etc.) were considered during the design of this study,

the limited number of farmers willing to participate in this study and the sporadic nature of the disease meant that we could not plan prior to an outbreak of UB and were forced to sample the sheep made available to us. This study therefore formed a basis from which more controlled future studies on the aetiology of UB will be carried out (e.g. more homogenous ram population).

In addition, inter-sample variation in relative abundance is suspected as arising from inherent bias produced during PCR amplification, DNA preparation, sequencing and data analysis. This effect should be minimized as much as possible by refinement of the relevant protocols, with higher emphasis being placed on per sample DNA quantification methods such as qPCR prior to sequencing and maximizing the quality profiles of sequences during sequencing.

We can also not rule out that an incubation period exists for the disease which could mean that a number of “healthy” rams may have already been diseased although no symptoms were present, thus providing a bacterial community that would closely resemble that of diseased rams. A deeper understanding of the disease and its affects is thus required and studies on factors such as immune and histopathological repsonses which could help narrow down uncertainties about UB should be carried out.

Future studies should sample the directly affected areas in an effort to narrow down the potential organisms that might be considered as playing a causative role in the incidence of UB. In the present study the entire penile mucosa comprising of affected and unaffected areas were sampled, which may have produced the high bacterial diversity observed and which potentially masked any changes in community composition at the affected area.

Although no significant differences in bacterial diversity and community composition was observed, a few phylotypes (phylum to genus) and OTUs were found to occur at significantly different abundances in the diseased versus healthy populations. No phylotypes were enriched in the diseased state. OTUs that were differentially abundant in the diseased population and that indicated the highest biological relevance were characterized as the genera *Fusobacterium* and *Porphyromonas*, as well as uncharacterized genera within Aerococcaceae and Bacteroidales. These few OTUs that had a significantly higher abundance in the diseased state may be disease-specific/disease-associated. Future studies need to investigate the involvement of these genera in the incidence of UB.

The previously suggested primary aetiological agent *Mycoplasma mycoides mycoides* large colony was not identified in this study. *Trueperella pyogenes (*previously known as *Arcanobacterium pyogenes*) thought to be associated with the progressive stage of the disease was identified in this study with higher prevalence in diseased rams than previously reported, but was not exclusively associated with UB. Three *Mycoplasma* species, two of which have consistently been identified in diseased rams (*Mycoplasma arginini* and *Mycoplasma bovigenitalium*), and one not previously identified (*Mycoplasma hyopharyngis)* were observed in this study, with higher prevalence in diseased rams, higher than

previously reported. These were however, in the two previous South African studies, not associated with UB. *Mycoplasma hyopharyngis*, which was characterized for the first time in the penile environment of rams, warrants further investigation into its role in the penile environment. High inter-sample variation in the relative abundance and prevalence of these species was observed. The relative abundance of these organisms in healthy and UB affected Dorper rams has not previously been described, and it is thus not possible to determine whether the inter-sample variation in relative abundance observed for these species is a true biological representation. The large inter-sample variation observed and the small sample size may however have affected the statistical power which resulted in no significant differences in the relative abundances of these species in healthy versus diseased sheep. These findings require validation through additional 16S amplicon sequencing studies of the penile microbiota, concomitantly with more accurate quantification methods using methods such as qPCR on a greater number of samples from healthy and diseased Dorper rams. From these findings we thus conclude that the aetiology of UB is still unresolved. Results from this study will however be used to direct future studies on the disease.

Although this study provides some evidence of a bacterial aetiology which must be further validated, the viral and fungal communities of the penile environment of rams have not previously been thoroughly investigated and the isolation of ovine *Herpes virus* type 2 in vaginal and penile specimens of sheep affected with UB in a recent study conducted in Europe may suggest the need for future studies to characterize the fungal and viral bacterial communities in healthy and diseased rams.

During analysis of the mock community, we observed that the Greengenes reference database and the Ribosomal Database Project Naïve Bayesian classifier chosen for this study had limitations in taxonomically classifying our sequences, with some species that were wrongly classified. The use of a mock community as a positive sequencing control is thus imperative in interpreting sequencing data and can reveal limitations of 16S amplicon sequencing.

Finally, results from all sequence analysis steps revealed the complexity of 16S amplicon sequencing data analysis, and the implications of the protocols and tools used for DNA preparation, sequencing and sequence analysis in characterizing the bacterial population within the penile mucosa of Dorper rams. Interpretation of the bacterial community in future studies will be based on the findings from this study, and therefore refinement of the protocols used in this study and incorporating potential new techniques may indicate signficiant differences between the bacterial microbiome of healthy and diseased Dorper rams, thus providing crucial information required for the formulation of management programs to minimize of prevent the spread of UB in South Africa.

APPENDIX A

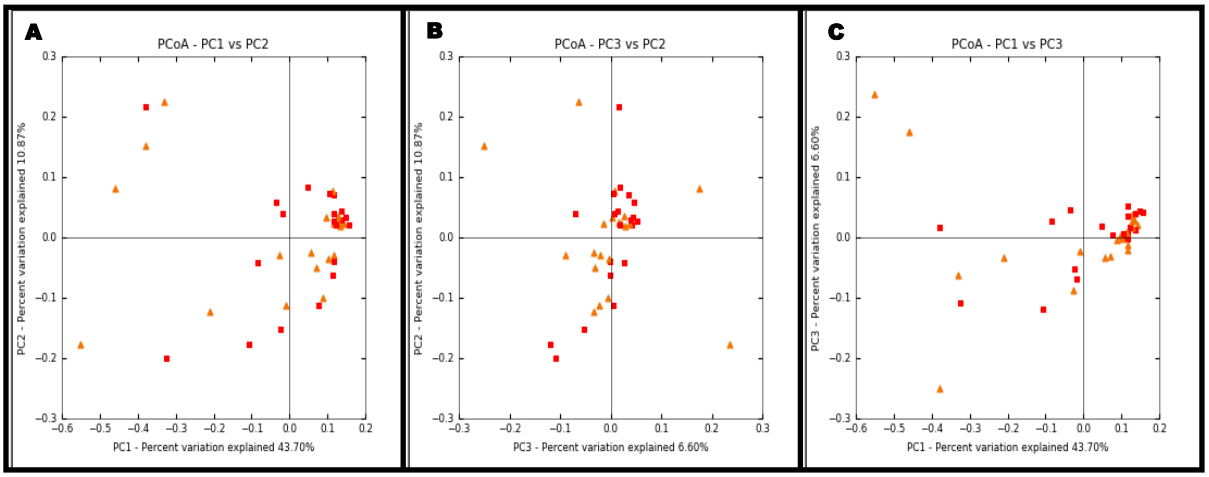

Figure S4.1. Unifrac diversity measures. Principle coordinate analysis, using unweighted Unifrac distances to explore and visualize group similarities between healthy and diseased samples. Each point represents a sample, plotted by a principal component on the X- axis and another principal component on the Y- axis. Square (red) data points represent samples from healthy rams and triangular (orange) from diseased rams. PC1 is principal component 1, PC2 is principle component 2 and PC3 is principle component 3. **A)** PC1 versus PC2, **B)** PC1 versus PC3, **C)** PC3 versus PC2.

Table S3.1. The total healthy and diseased ram population sampled in this study grouped according to age.

		Healthy		Diseased			
		Number	**%**	**Number**	**%**	**Total**	**%**
Age (teeth number)	0 tooth	6	10.9	1	2.1	7	6.8
	2 teeth	7	12.7	13	27.1	20	19.4
	4 teeth	5	9.1	6	12.5	11	10.7
	6 teeth	14	25.5	7	14.6	21	20.4
	Full mouth	23	41.8	21	43.7	44	42.8
Total		55	100	48	100	103	100