

Comparison of plasmids from clinical *Lactobacillus* strains



**UNIVERSITY of the
WESTERN CAPE**

Lyle Keenan Harris

**A thesis submitted in partial fulfilment of the requirements for
the degree of MAGISTER SCIENTIAE (M.Sc.) Department of
Biotechnology, University of the Western Cape, Bellville**

Supervisor: Prof. Marla Trindade

Declaration

I, Lyle Keenan Harris, declare that “Comparison of plasmids from clinical *Lactobacillus* strains” is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Date: 2018-01-26

Signature: 



Acknowledgements

First and foremost, I would like to thank the Almighty God through which nothing is impossible.

I would like to thank my supervisor Prof. Marla Trindade for giving me the opportunity to pursue this dream.

I would especially like to thank Mr Lonnie van Zyl and Dr Bronwyn Kirby for the technical support and continuous assistance throughout the project.

I would like to thank Dr Heide Goodman for the guidance and support.

I would also like to thank Dr Leonard Damelin for his support and assistance.

I would also like to especially thank Dr Albert Abrie and Dr Kerry Gordon for their assistance and support.



Dedication

This is for my family, more especially my mother and father for the continuous support, Trevor, Benadette, Lucentio, Langston, Trevino, Troye and Annique.



ABSTRACT

The vaginal mucosa is dominated by Gram positive, rod shaped lactobacilli which serve as a natural barrier against infection. In both healthy and BV infected women *Lactobacillus crispatus* and *Lactobacillus jensenii* has been found to be the predominant *Lactobacillus* species. Many studies have been conducted to assess factors influencing lactobacilli dominance in the vaginal microbiome. However, no study has evaluated the impact of plasmids on the vaginal lactobacilli. In the present study two plasmids, pLc17 and pLc4, isolated from vaginal *Lactobacillus* species of both healthy and BV infected women were characterized. pLc4 was present in both *Lactobacillus crispatus* and *Lactobacillus jensenii* while pLc17 was only present in *Lactobacillus crispatus*. pLc17 (16663 bp in size) encoded a ribonucleotide diphosphate reductase (RNR), a filamentation induced by cAMP-like (FIC-like) protein and numerous mobile elements. The FIC-like protein may assist pLc17 to persist within the bacterial population, while RNR is commonly associated with phages and may indicate phage infection. pLc4 (4224 bp in size) encodes for a replication initiator protein and a plasmid partitioning protein. The replication protein on pLc4 shows 44% identity with the replication initiation protein of pSMQ173b_03. On further phylogenetic and sequence analysis with other Rolling Circle Replication (RCR) plasmids, pLc4 appears to be novel as the plasmid shows a low degree of similarity to these RCR plasmids. pLc17 appears to carry both a RCR replicon as well as a theta replicon, similar to pIP501, the broad-host-range plasmid from *Bacillus subtilis*. The relative Plasmid Copy Number (PCN) for pLc4 and pLc17 was analysed using quantitative polymerase chain reaction (qPCR) for the healthy state relative to the disease state from twenty-eight vaginal swab samples obtained from the National Institute for Communicable Diseases (NICD). The relative PCN for pLc4 and pLc17 had a fold increase of ~2.803 and ~1.693, respectively in the healthy patient samples relative to BV infected patient samples. However, there were not found to be significant differences when taking the standard error into account

Due to the novelty of these plasmids further analysis and characterisation is required for both plasmids, to establish what role they may play in the health of the vaginal milieu.

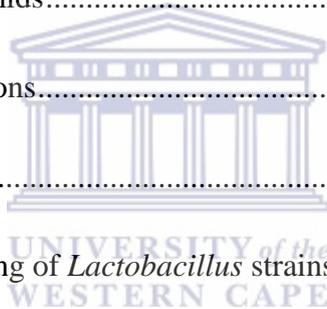
Keywords: *Lactobacillus crispatus*, *Lactobacillus jensenii*, pLc17, pLc4, Ribonucleoside-diphosphate reductase, Rolling Circle Replication (RCR), quantitative polymerase chain reaction (qPCR), plasmid copy number (PCN) and Theta replication

Contents

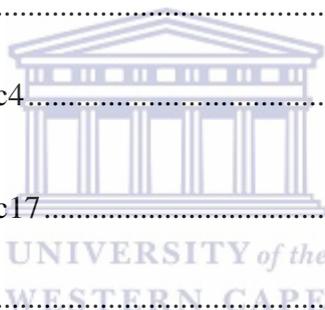
Declaration.....	ii
Acknowledgements.....	.iii
Dedication.....	iv
Abstract.....	v
Contents.....	vi
List of Tables.....	x
List of Figures.....	.xi
List of Abbreviations.....	xiii
CHAPTER 1 INTRODUCTION.....	1
1.1 Introduction.....	1
1.2 The Human Microbiome.....	1
1.3 The Vaginal Microbiome.....	4
1.4 Lactobacilli the resilient microbe.....	5
1.5 Lactobacilli defence mechanisms.....	6
1.5.1 Lactic acid production.....	6
1.5.2 Bacteriocin production.....	6
1.5.3 Hydrogen peroxide production.....	6
1.5.4 Lactobacilli adherence.....	7
1.6 Lactobacilli colonisation in the vaginal microbiome.....	8
1.7 Other associated vaginal microbes.....	14



1.8 Bacterial Vaginosis	14
1.9 Lactobacilli as a live microbiocide	16
1.10 Probiotics for BV	18
1.11 Present and past techniques used for vaginal microbiome studies	19
1.12.1 Plasmids	20
1.12.2 Lactobacilli plasmids	21
1.13 Research objectives.....	21
CHAPTER 2 MATERIALS AND METHODS.....	22
2.1 Bacterial strains and plasmids.....	22
2.2 Media and growth conditions.....	22
2.3 Plasmid extraction.....	24
2.4 Next generation sequencing of <i>Lactobacillus</i> strains.....	24
2.5 Genomic DNA extraction	25
2.6 PCR amplification.....	25
2.7 Analytical Techniques.....	27
2.7.1 DNA quantification.....	27
2.7.2 Agarose gel electrophoresis	27
2.8 Amplicon DNA purification	27
2.9 Relative quantitative PCR.....	27
2.9.1 Standard curve construction.....	27
2.9.2 Swab sample preparation and quantification	29



2.9.3 Bacterial genomic DNA test	29
2.9.4 Relative Expression Analysis	29
CHAPTER 3 Sequence Analysis using NGS	30
3.1 INTRODUCTION	30
3.2 RESULTS AND DISCUSSION	36
3.2.1 Plasmid Extraction	36
3.2.2 Isolation, sequencing and assembly of pLc4 and pLc17	37
3.2.3 Identification of plasmids using junction primers.....	38
3.2.4 pLc4 replication	39
3.2.5 Sequence analysis of pLc4.....	41
3.2.6 Sequence analysis of pLc17.....	44
3.3 CONCLUSION.....	52
CHAPTER 4 Plasmid Copy number determination	53
4.1 INTRODUCTION	53
4.2 RESULTS AND DISCUSSION	56
4.2.1 pLc4 and pLc17 plasmid copy number (PCN) determination	56
4.2.2. Swab sample DNA extraction.....	56
4.2.3 Quantitative real time PCR (qPCR).....	56
4.3 CONCLUSION.....	70
CHAPTER 5 GENERAL DISCUSSION AND CONCLUSION.....	71
REFERENCES	76



Appendices: Buffers and stock solutions	94
Supplementary figures and tables	95



List of Tables

Table 1.1: A summary of studies assessing the vaginal composition of women from various countries	9
Table 2.1: Bacterial strains used in this study.....	22
Table 2.2: Primers used in this study for PCR amplification of genes	26
Table 3.1: Analysis of sequence assembly.....	37
Table 3.2: Identification of pLc4 or pLc17 in individual samples from healthy or infected individuals.....	38
Table 3.3: Putative genes and their products, deduced from the plasmid nucleotide sequences for pLc4.....	40
Table 3.4: Inverted repeat sequences for pLc4	41
Table 3.5: Putative genes and their products, deduced from the plasmid nucleotide sequences for pLc 17.....	44
Table 3.6: Direct and inverted repeats for pLc17	50
Table 4.1 DNA concentrations used for spiking experiment.....	59
Table 4.2 Plasmid samples used to assess the presence of pLc4 and pLc17.....	61
Table 4.3: Relative PCN and mean DNA concentration of pLc4, pLc17 and reference gene (EGFR)	65
Table 4.4: Relative PCN ratio and SE of pLc4 and pLc17	68

List of Figures

Figure 1.1: Compositional differences in the microbiome by anatomic site	2
Figure 1.2: Acquisition of the microbiome in early life by vertical transmission and factors modifying mother-to-child microbial transmission.....	3
Figure 1.3: Overview of the level of LAB in the different body sites	5
Figure 3.1 Enzymatic tagmentation using Illumina Nextera XT library preparation	31
Figure 3.2: Summary of the sequencing-by-synthesis (SBS) process	32
Figure 3.3: Agarose gel electrophoresis on a 1% gel of plasmid DNA extracted from <i>L. crispatus</i> pLc4 using the Zyppy kit.....	36
Figure 3.4: Agarose gel electrophoresis on 1% gel of PCR DNA amplification using pL4F and pL4R junction primers.....	39
Figure 3.5: Annotated genetic map of pLc4. The sequenced genetic map of 4224 bp plasmid pLc4	40
Figure 3.6: The evolutionary history was inferred using the Neighbor-Joining method.	42
Figure 3.7: tBLASTx comparison of pLc4 with pSMQ173b and pSK11b	43
Figure 3.8: Annotated genetic map of pLc17	45
Figure 3.9: tBLASTx comparison of pLc17 with pHN1 and pR2.....	51
Figure 4.1: Agarose gel electrophoresis on 1% gel of PCR DNA amplification using qPL4F/R and qPL17F/R primers.....	57
Figure 4.2: qPCR standard curves for the pLc4, pLc17 (n=6) and EGFR (n=4) primer sets	58
Figure 4.3: Amplification curves generated to assess the primer specificity of pLc4 and pLc17.....	60
Figure 4.4: Amplification curve of clinical samples.....	62
Figure 4.5: Melt curve analysis of clinical samples	63
Figure 4.6: Agarose gel electrophoresis of representative qPCR amplicons from clinic samples for pLc4 and pLc17	64
Figure 4.7: Amplification curve of vaginal swab samples	67
Figure 4.8: Relative PCN ratio of pLc4 (n=26) and pLc17 (n=25) for healthy samples	

relative to diseased (BV infected) samples 69



List of Abbreviations

Aa Amino acid

ADP Adenosine 5'diphosphate

AMP Adenosine 5'monophosphate

ATP Adenosine 5'triphosphate

BHR broad host range

BLAST Basic local alignment sequencing tool

BLASTn Basic local alignment sequencing tool nucleotide

bp base pairs

BV bacterial vaginosis

BVAB BV Associated Bacteria

cDNA complementary DNA

Contig Contiguous

CsCl Caesium chloride

CV-N cyanovirin-N

Cq Quantification cycle

ddPCR droplet digital PCR

dH₂O distilled water

DNA deoxyribonucleic acid

dNTP deoxynucleotide triphosphate

doc death on curing

dsDNA double stranded DNA

dso double stranded origin

EDTA ethylene diamine tetra acetic acid

EF-TU translating elongation factor

EGFR epidermal growth factor receptor

et al. et alia (and others)

Fic Filamentation induced by cAMP

FISH Fluorescent In Situ Hybridization

gDNA genomic DNA

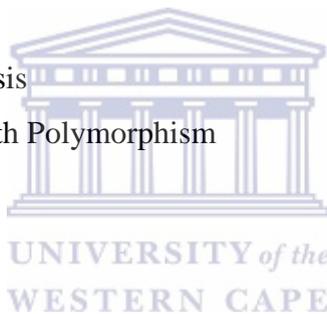
GC The proportion of guanine and cytosine bases in a DNA/RNA sequence



GIT gastro intestinal tract
g grams
g/L grams per liter
g/mol grams per mole
Gyr B gyrase B
HIV Human Immunodeficiency Virus
HPLC high performance liquid chromatography
HPV Human papillomavirus
H₂O₂ Hydrogen Peroxide
IS Insertion Sequences
kb kilobase
LAB Lactic Acid Bacteria
LB Luria-Bertani
LBA Luria-Bertani Agar
L Liter
µg/ml microgram per milliliter
µg/µL microgram per microliter
µL microliter
µM micromolar
mCV-N cyanovirin-N
MetaHIT Metagenomics of the Human Intestinal Tract
mg milligrams
mL milliliter
mM millimolar
Mob Mobilisation gene
MRS deMan, Rogosa and Sharpe
NA not applicable
NaCl sodium chloride
NCBI National Center for Biotechnology Information
NICD National Institute for Communicable Diseases
ng nanogram
NGS Next generation sequencing
°C degrees Celsius



OD optical density
OH hydroxyl
ORFs Open Reading Frames
PCR polymerase chain reaction
PCR-DGGE Polymerase Chain Reaction-Denaturing Gradient Gel Electrophoresis
PCN plasmid copy number
pg/ μ l picograms per microliter
pH percentage hydrogen
phd prevent host death
pRNA primer Ribonucleic acid
qPCR quantitative polymerase chain reaction
RAPD-PCR Randomly Amplified Polymorphic DNA Polymerase Chain Reaction
RC Rolling circle
RCR Rolling Circle Replication
REST Relative Expression Analysis
RFLP Restriction Fragment Length Polymorphism
rpm revolutions per minute
RNR ribonucleotide reductases
SBS sequencing-by-synthesis
scFV single chain variable fragments
SDS Sodium dodecyl sulphate
SNPs single nucleotide polymorphisms
sp. Species
ssDNA single stranded DNA
sso single stranded origin
STI Sexual Transmitted Infection
TA Toxin Antitoxin
TAE tris-acetic acid
tBLASTx Translated Basic local alignment sequencing tool
TE tris-EDTA
T-RFLP-terminal restriction fragment length polymorphism
Tris tris (hydroxymethyl)-aminomethane



Tris HCl tris hydrochloric acid

UV ultra violet

V volts

VDS vaginal discharge syndrome

VH variable heavy

VL variable light

v/v volume per volume

w/v weight per volume



CHAPTER 1 INTRODUCTION

1.1 Introduction

Studies have shown that the human microbiome plays an important role in development, physiology, immunity and nutrition (Ma *et al.* 2012). Different microbes colonize various parts of the body which include the vagina, oral cavity, skin, nasal cavity, gastrointestinal tract, and the urethral tract (Fettweis *et al.* 2012). The majority of these microbes form a mutualistic relationship with the human host, contributing to improved health (Ma *et al.* 2012). The vaginal mucosa is dominated by Gram positive, rod-shaped lactobacilli which serve as a natural barrier against infection (Damelin *et al.* 2010, 2011; Martin *et al.* 1999). *Lactobacillus* species inhibit pathogen colonisation and as such are recognised as a marker of good vaginal health (Damelin *et al.* 2011; Martin *et al.* 1999).

1.2 The Human Microbiome

The human body evolved to sustain a plethora of microorganisms, and there is a diverse array of bacteria associated with the human body (Pflughoeft & Versalovic 2012; Warinner *et al.* 2014). It is estimated that there are ten bacterial cells for every human cell (Pflughoeft & Versalovic 2012), estimated to account for as much as 2% of the total body mass of an adult. Thus the human-microbe can be viewed as a super organism (Warinner *et al.* 2014). In 2008 the National Institute of Health Human Microbiome project conducted in the United States and the Metagenomics of the Human Intestinal Tract (MetaHIT) conducted in Europe were created. These projects were established in an effort to measure and analyse the unique relationship between microbes and their human hosts by measuring the microbial diversity at various sites of the human body (Figure 1.1). Ultimately, the aim was to establish the role of the microbiome in vital human functions, the aetiology of disease, as well as the evolution of new species. Findings from the human genome project have revealed that the number of bacterial genes exceeds the number of human genes by a factor of 150 (Warinner *et al.* 2014). The host-microbe relationship in most cases can be advantageous to the health of the human host (Pflughoeft & Versalovic 2012). The benefits gained from this symbiosis include improved metabolism, digestion, vitamin production, and immune system, as well as the ability to outcompete potential pathogens (Warinner *et al.* 2014). However, the shift from symbiotic to dysbiotic state can result in various ailments and diseases. The microbiome of a healthy human can also be composed of a number of pathogenic bacteria such as

Streptococcus pneumoniae, *Haemophilus influenzae*, *Neisseria meningitidis*, *Clostridium difficile*, *Propionibacterium acnes* and *Staphylococcus aureus* (Warinner *et al.* 2014). In a study conducted on the human microbiome's metabolic and functional pathways, it was found that common core pathways existed in the microbiome of individuals and body habitats, which include ribosome and translational machinery, nucleotide charging, ATP synthesis and glycolysis (The Human Microbiome Project Consortium 2012). The microbial distribution varies at different points in or on the human body, thus the type of microbe is determined by the location on the human body and may also vary according to ethnic group, gender and age (Figure 1.1). The human microbiome is said to be resilient with the ability to recover after a disturbance has occurred. However, this phenomenon is not entirely true for all organisms, as studies on indicator organisms, like *Helicobacter pylori*, have shown that a population of these organisms can be completely removed from the host organism. Thus destabilization of the microbiome and major disturbance can result in loss of long term inherited organisms, which may impact the host's health (Cho & Blaser 2012).

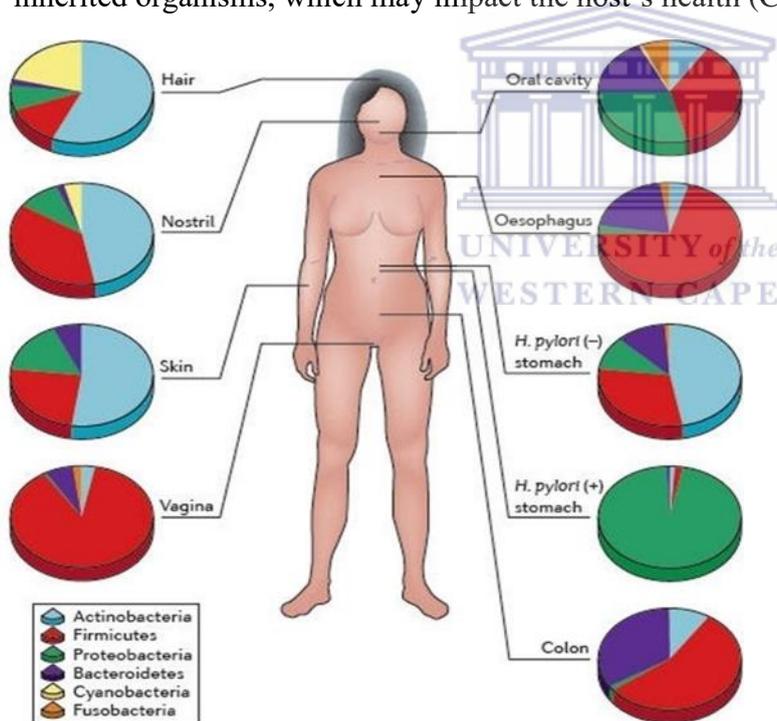


Figure 1.1: Compositional differences in the microbiome by anatomic site. High-throughput sequencing has revealed substantial intra-individual microbiome variation at different anatomical sites, and inter-individually for the same anatomical sites (Cho & Blaser 2012).

Microbial colonisation appears to occur at child birth. Studies have shown that the mother may be the primary source of initial microbial inheritance (Figure 1.2). During vaginal delivery the infant is exposed to lactobacilli from the vagina. In addition, Lactic Acid Bacteria (LAB) found in the breast milk, may introduce lactobacilli in the baby’s gastrointestinal tract (Cho & Blaser 2012).

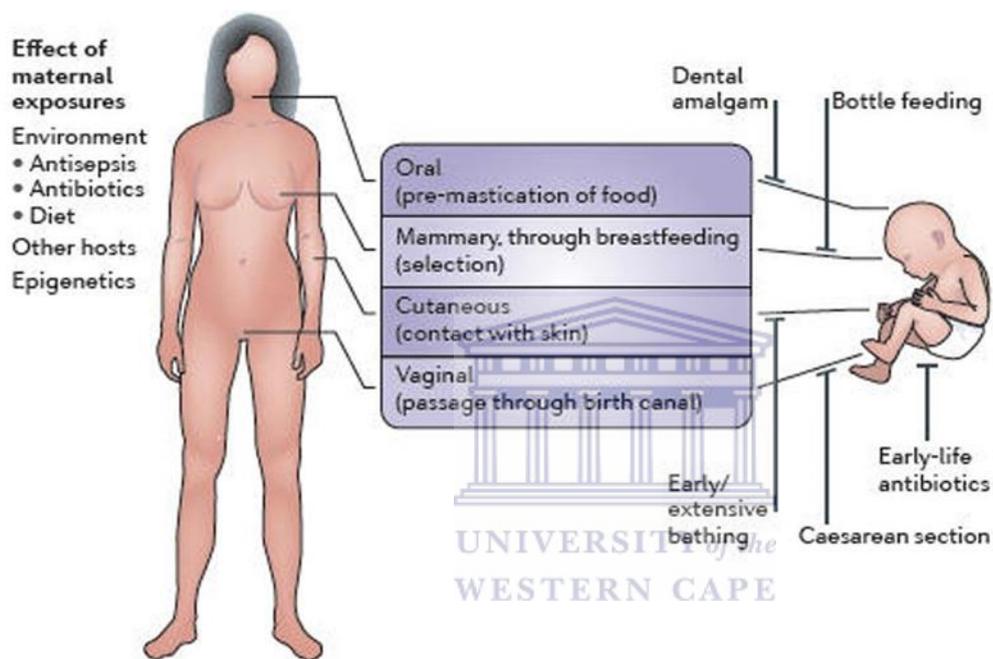


Figure 1.2: Acquisition of the microbiome in early life by vertical transmission and factors modifying mother-to-child microbial transmission. (Cho & Blaser 2012)

1.3 The Vaginal Microbiome

To fully understand the vaginal microbiome one has to consider the anatomical structure and physiology of the vagina. The human vagina is a stretchable lumen composed of stratified squamous epithelium, which undergoes various changes regulated by oestrogen levels. Vaginal secretions consist of 90-95% water, inorganic and organic salts, urea, carbohydrates, mucus, fatty acids, albumins, immunoglobulins, iron chelators, lysozyme and other macromolecules, leukocyte and epithelial debris (Boris & Barbés 2000). The complexity of bacteria colonizing the vagina is a result of anatomical and physiological changes that occur during the course of development (Turovskiy *et al.* 2011). Microbial colonisation, including vaginal colonisation, differs with birth mode. During vaginal delivery the infant is first exposed to the vaginal microbes, compared with Caesarian section delivery, where the infant is exposed to skin microbes (Cho & Blaser 2012). During the first few weeks of female infant development, the vaginal epithelium develops as a result of glycogen deposition. During the maturation process there is an increase in estrogen production resulting in thickening of vaginal epithelium and intracellular production of glycogen. These new environmental conditions allow for microorganisms capable of fermenting glycogen to lactic acid to result in the acidification of the vaginal environment, thereby shaping the initial vaginal microbiome (Hickey *et al.* 2012). The lactobacilli use the glycogen as an energy source and through anaerobic metabolism produce large amounts of lactic acid (Mirmonsef *et al.* 2014). A spike in oestrogen during puberty is linked with increased glycogen being deposited in the stratified squamous non-keratinized vaginal epithelium, resulting in further vaginal development. In most healthy women this change is associated with an increase in lactobacilli. *Lactobacillus* species are capable of producing lactic acid as an end product during fermentation of glycogen thereby lowering the pH of the vagina. It should be noted though, that *Gardenerlla vaginalis* and *Prevotella bivia* are also present during the pubescent phase but in low numbers (Hickey *et al.* 2012). Acidic conditions within the vaginal milieu is associated with vaginal health. However, the onset of menopause results in a reduction in glycogen levels leading to a reduction in lactobacilli. In addition, the onset of menopause also results in a reduction in vaginal secretions (Fettweis *et al.* 2012). Hormonal replacement therapy has been found to replenish lactobacilli numbers illustrating the importance of the physiological environment on the vaginal microflora (Boris & Barbés 2000; Turovskiy *et al.* 2011). Menstruation appears to be a vulnerable stage, as there is an increase in vaginal pH which results in an increase in the

number of potential pathogens that can lead to the development of bacterial vaginosis (BV) (Turovskiy *et al.* 2011). BV is an abnormal vaginal condition characterised by vaginal discharge that results from a shift in the normal vaginal microbiome and an overgrowth of atypical bacteria in the vagina (Pavlova *et al.* 2002; Teixeira *et al.* 2012). More detailed information pertaining to BV can be found in section (Section 1.8).

1.4 Lactobacilli the resilient microbe

LAB are ubiquitous in nature, colonising humans, plants and animals (Pavlova *et al.* 2002). They are found in many biomes of the human body such as the skin, mouth, gastro intestinal tract (GIT) and the vagina (Figure 1.3). However, the composition varies at different regions with the highest numbers being located in the mouth and the vagina (Douillard & de Vos 2014). Lactobacilli were first identified by Doderlein in 1892 as the dominant bacteria colonising the vagina (Boris & Barbés 2000; Verstraelen *et al.* 2009).

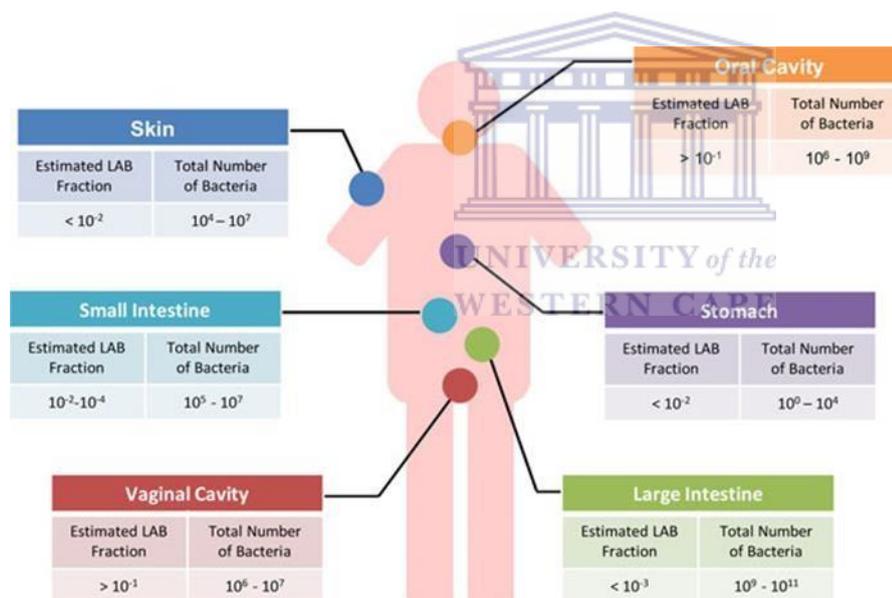


Figure 1.3: Overview of the level of LAB in the different body sites. The estimated LAB fraction is based on several complete, comprehensive phylogenetic and metagenomic datasets and is reported as the total number of bacteria per gram of homogenized tissue or fluid or square centimetre of skin. (Douillard & de Vos 2014)

1.5 Lactobacilli defence mechanisms

The lactobacilli dominating the genital tract have been found to suppress growth of other genital bacteria and this mechanism of suppression may also act on Human Immunodeficiency Virus (HIV) (Mirmonsef & Spear 2014; Vallor *et al.* 2001). There are several possible mechanisms which allow lactobacilli to dominate the vaginal mucosa (Mirmonsef & Spear 2014). These include the production of lactic acid, hydrogen peroxide, bacteriocins and lactobacilli adherence, which will be discussed separately (Boris & Barbés 2000).

1.5.1 Lactic acid production

The production of lactic acid during *Lactobacillus* metabolism is one of the major contributing factors for the low pH of the vagina, which is generally between pH 4-4.5. Many *in vitro* studies have shown the suppression of pathogenic bacteria as a result of this (Boris & Barbés 2000; McLean & Rosenstein 2000).

1.5.2 Bacteriocin production

Lactobacilli secrete bacteriocins; antimicrobial compounds that are either active against bacteria of the same species (narrow spectrum) or across genera (broad spectrum) (Mirmonsef & Spear 2014; Yang *et al.* 2012). Bacteriocins often target the cell membrane effecting its permeability (Ma *et al.* 2012). Bacteriocin production gives lactobacilli an added advantage in establishing themselves in particular environments (Boris & Barbés 2000). Although bacteriocins are not common antagonistic compounds produced by vaginal lactobacilli, they have been identified in various studies. In a study by Karaoğlu *et al.* (2003) where the antimicrobial activity of vaginal lactobacilli was analysed, it was found that six out of 100 lactobacilli strains tested were capable of producing bacteriocins. The bacteriocins had antagonistic activity against *G. vaginalis*, *Pseudomonas vulgaris*, *Pseudomonas aeruginosa*, *Escherichia coli* and *Escherichia cloacae*. These bacteriocins were, however, shown to also inhibit other vaginal lactobacilli which is interesting as this may play a role in depletion of vaginal lactobacilli observed during BV (Karaoğlu *et al.* 2003).

1.5.3 Hydrogen peroxide production

In vitro studies have shown that lactobacilli produce significant amounts of hydrogen peroxide (H_2O_2) which has been found to be the major defence mechanism against potential pathogens (Boris & Barbés 2000; Teixeira *et al.* 2012). H_2O_2 is either inhibitory or toxic to the surrounding bacteria and is also known to act on other exogenous lactobacilli (Boris & Barbés 2000; Turovskiy *et al.* 2011). BV-infected vaginal fluid has been shown to have reduced levels of H_2O_2 compared to normal vaginal fluid, suggesting that this may be a critical antagonistic compound (Turovskiy *et al.* 2011). Clinical studies performed by Wineslaus and Calver tested the efficacy of H_2O_2 by treating the vaginas of 30 BV-infected women with a 3% H_2O_2 solution. Eighteen women from the study were completely cured of the symptoms associated with BV according to the Amsel criteria (Wineslaus & Calver 1996). H_2O_2 Production by lactobacilli has also been shown to kill HIV *in vitro* (Mirmonsef & Spear 2014). The bacteriocidal activity demonstrated by H_2O_2 may be due to it acting on its own or in combination with a halide and a peroxidase (Boris & Barbés 2000; Vallor *et al.* 2001). It has been shown that 94-95% of *L. crispatus* and *L. jensenii* strains produce H_2O_2 (Antonio *et al.* 1999). *L. crispatus* and *L. jensenii* are known to inhibit growth of BV-related microorganisms such *G. vaginalis*, anaerobic Gram-negative rods and *Mycoplasma hominis* (Martin *et al.* 1999; Sha *et al.* 2005). This suggests that colonisation by H_2O_2 producing lactobacilli, particularly *L. crispatus* and *L. jensenii* may improve the health of the vaginal mucosa. Thus the presence of lactobacilli in the vagina may increase resistance to HIV directly through secretion of HIV inhibitors or indirectly through the inhibition of bacteria associated with the increased risk of HIV infection. However, some researchers have questioned the role of H_2O_2 *in vivo* as the vagina is devoid of oxygen, a requirement for H_2O_2 production by lactobacilli. Furthermore, a study has shown that H_2O_2 production was ineffective against 17 BV Associated Bacteria (BVAB) under anaerobic conditions. The vaginal fluid may also further reduce the effect of H_2O_2 (Ma *et al.* 2012). Thus the correlation between hydrogen peroxide producing lactobacilli and reduction in BV infection needs to be investigated further.

1.5.4 Lactobacilli adherence

Lactobacillus colonisation appears to be linked to the adherence of the bacteria to the vaginal epithelial cells (McLean & Rosenstein 2000). This may be dependent on the host factors such as changes in pH, glycogen content, and the sloughing and the restoration of vaginal

epithelial cells as a result of the female reproductive hormones (Boris & Barbés 2000; Mirmonsef & Spear 2014). This may affect the adhesive properties of the vaginal bacteria, thus impacting bacterial distribution. Improved colonisation due to glycogen in the vagina has led to the theory that glycogen may be a nutrient critical for *Lactobacillus* colonisation (Boris & Barbés 2000). This may be achieved by the breaking down of glycogen by the vaginal epithelium and *Lactobacillus* enzymes into glucose, which is then utilised by lactobacilli to produce lactic acid. It has been shown that lactobacilli are able to interfere with bacterial uropathogen colonisation of both uroepithelial and vaginal cells (Karaoğlu *et al.* 2003; Pyles *et al.* 2014). Lactobacilli cell walls prevent the attachment of uropathogens to uroepithelial cells thus allowing these bacteria to outcompete the pathogens via steric hindrance and not through receptor site blockage. However, it has been shown that vaginal lactobacilli can outcompete *G. vaginalis* and *Candida albicans* for vaginal receptor sites thus reducing their colonisation (Boris & Barbés 2000).

1.6 Lactobacilli colonisation in the vaginal microbiome

Numerous studies have been done to establish which of the *Lactobacillus* species are dominant in the human vagina irrespective of race, culture, demography or other differences. These studies are summarized in Table 1.1, with a specific focus on *L. crispatus*, *L. jensenii*, *L. iners* and *L. gasseri*, as these lactobacilli have been shown to play a role in vaginal health.

Table 1.1: A summary of studies assessing the vaginal composition of women from various countries

Aim of the study	Country	Number of participants	Findings	Reference
The study was performed to establish if there was a significant difference between the vaginal lactobacilli of African and Caucasian (American and European) women	Nigeria	241 healthy women	It was found that 85.8% of the women in the study had lactobacilli as part of their vaginal populations, and the majority of the women were colonised by non-H ₂ O ₂ producing <i>L. iners</i> then <i>L. gasseri</i> followed by <i>L. plantarum</i> , <i>L. stutoryeus</i> , <i>L. crispatus</i> and <i>L. rhamnosus</i> . These findings are in agreement with other vaginal microbiome studies conducted in Canada, the United States and Sweden	Anukam <i>et al.</i> (2006)
The study was performed to establish if there was a significant difference between the vaginal lactobacilli of African and Caucasian (American) women	North America	144 healthy black and white women	It was found that <i>L. iners</i> , <i>L. crispatus</i> , <i>L. gasseri</i> and <i>L. jensenni</i> were the dominant lactobacilli in both groups, lactobacilli were absent in more black women (33%) compared with white women (7%) and in these instances the vaginal microbiome had <i>Atopobuim</i> (LAB) and <i>Clostridia</i> . In addition, white women are more likely to have more than one type of <i>Lactobacillius</i> species.	Zhou <i>et al.</i> (2007)

The study was performed to establish if there was a significant difference between the vaginal lactobacilli of Asian women to that of African and Caucasian (American) women	Japan	73 healthy Japanese women	It was found that <i>L. iners</i> , <i>L. crispatus</i> , <i>L. gasseri</i> and <i>L. jensenni</i> were the dominant lactobacilli in Japanese women.	Zhou <i>et al.</i> (2010)
In this study four ethnic groups (black, white, Hispanic and Asian) were analysed	North America	396 asymptomatic women	It was found that of the majority of white and Asian women were dominated by <i>L. iners</i> , <i>L. crispatus</i> , <i>L. gasseri</i> and <i>L. jensenni</i> , which were less dominant in black and Hispanic women. Mixed microbial communities were found in the black and Hispanic women, which are composed of high numbers of LAB.	Ravel <i>et al.</i> (2011)
In this study 40 African women were analysed; 21 were healthy, 5 intermediate and 14 BV infected	South Africa	40 African women	The majority of the subjects were colonised by lactobacilli with <i>L. crispatus</i> being the dominant species, which is in accordance with the many studies conducted on European populations	Pendharkar <i>et al.</i> (2013)
This study was conducted to establish if there was a significant difference between the vaginal lactobacilli across women from 7 countries	Argentina, Brazil, China, India, South Korea, Turkey and the United States	400 women	It was found that the majority of isolates were <i>L. crispatus</i> , <i>L. jensenii</i> and <i>L. gasseri</i>	Pavalova <i>et al.</i> (2002)
This study was conducted to examine the vaginal composition of premenopausal South African women	South Africa	144 women	It was found that <i>L. crispatus</i> , <i>L. jensenii</i> , <i>L. iners</i> and <i>L. gasseri</i> were the dominant lactobacilli in the healthy vagina of the South African women examined. Furthermore, they found a reduction in lactobacilli associated with	Damelin <i>et al.</i> (2011)

			BV infection. There was a substantial reduction in <i>L. jensenii</i> and a partial reduction in <i>L. gasseri</i> . Interestingly there was not a major reduction in <i>L. crispatus</i> during BV infection and these <i>L. crispatus</i> isolates also displayed a great amount of lysogeny (77%).	
This study was conducted to examine the vaginal composition of premenopausal South American women	Mexico	108 Hispanic women	It was found that <i>L. crispatus</i> , <i>L. gasseri</i> and <i>L. jensenii</i> were the dominant <i>Lactobacillus</i> species	Martinez-Pena, <i>et al.</i> (2013).
This study was conducted to examine the vaginal composition of sexually active women	North America	302 sexually active women	It was found that <i>L. crispatus</i> , <i>L. jensenii</i> , L 1086V (later identified as <i>L. iners</i>) and <i>L. gasseri</i> were the dominant species. In this study they also examined the H ₂ O ₂ production of the isolated strains and it was found that 95% of <i>L. crispatus</i> isolates were capable of producing H ₂ O ₂ and 94% of <i>L. jensenii</i> were capable of producing H ₂ O ₂	Antonio <i>et al.</i> (1999)
This study was conducted to examine the vaginal isolates of premenopausal women using PCR-DGGE and real time PCR	Belgium	26 premenopausal women	In this study they found that there was a reduction in the number H ₂ O ₂ producing lactobacilli (<i>L. acidophilus</i> , <i>L. gasseri</i> and <i>L. vaginalis</i>) and an increase in the number of non-H ₂ O ₂ producing lactobacilli (<i>L iners</i>) with women suffering from <i>Candida</i> infection, which suggests that <i>Candida</i> infection may be linked to a lack of H ₂ O ₂ producing lactobacilli	Vitali <i>et al.</i> (2007)
This study was conducted to examine the vaginal isolates from non-pregnant women	North America	101 non pregnant women	They found that <i>L. jensenii</i> and <i>L. crispatus</i> were the dominant lactobacilli. They also found that H ₂ O ₂ producing lactobacilli tend to colonise sexually active women for longer	Vallor <i>et al.</i> (2001)

			periods as compared to non H ₂ O ₂ producing lactobacilli over the 8-month interval.	
The antagonism potential of lactobacilli was examined	Kenya	107 BV infected women	They found that the greatest antagonism was related to acidity, as a result of the lactic acid and not hydrogen peroxide production	Matu <i>et al.</i> (2010)
This study was conducted to evaluate whether there was a significant shift in the microbial population of the 4 dominant lactobacilli between healthy and BV infected patients	China	107 women	They found that <i>L. crispatus</i> and <i>L. gasseri</i> was reduced significantly in number in BV patients. Whereas <i>L. iners</i> was significantly high in BV infected patients. <i>L. jensenii</i> was found not to be affected by the shift.	Yan <i>et al.</i> (2009)
This study was conducted to examine the vaginal composition of Iranian women	Iran	178 women 	They found that <i>L. crispatus</i> , <i>L. gasseri</i> , <i>L. iners</i> , <i>L. jensenii</i> , <i>L. acidophilus</i> and <i>L. rhamnosus</i> were dominant in healthy women. <i>L. crispatus</i> , <i>L. gasseri</i> and <i>L. jensenii</i> were reduced in BV infected women. However, <i>L. iners</i> was found to be higher in BV infected women	Motevaselli <i>et al.</i> (2013)
This study was conducted to investigate the vaginal microflora associated with both BV infected and healthy women	China	108 BV infected and healthy women	They found an increase in bacterial diversity in BV women with increased numbers of lactobacilli present in healthy women. <i>L. iners</i> was found at high levels in healthy women followed by <i>L. crispatus</i> and <i>L. jensenii</i> . <i>Atopobium</i> was strongly associated with BV. Several new species of bacterium were also identified to be linked to BV infection.	Ling <i>et al.</i> (2010)

<p>This study was conducted to investigate the vaginal microflora of healthy pregnant women</p>	<p>Belgium</p>	<p>100 healthy pregnant white women</p>	<p>They found that there was a shift in lactobacilli composition over the three trimesters. Thirteen out of the 77 women that had normal vaginal microbiome (composed of lactobacilli) converted to a <i>lactobacillus</i> absent microbiome after the 3rd trimester. It was also found that long term lactobacilli dominance was dependent on what <i>lactobacillus</i> was dominant in the vaginal samples.</p>	<p>Verstralen <i>et al.</i> (2009)</p>
---	----------------	---	--	--



The production of H₂O₂ by the lactobacilli colonising the vagina seems to be a major contributor to their sustained colonisation of the vagina. However, there are several studies which suggest that the antagonism is linked to lactic acid production and not H₂O₂ production. One such study was done by O’Hanlon *et al.* (2011) where the antagonistic effects of H₂O₂ and lactic acid were compared under conditions which mimic the human vagina. It was found that lactic acid and not H₂O₂ had the greatest inhibitory potential on BVAB without effecting vaginal lactobacilli (O’Hanlon *et al.* 2011). The type of lactobacilli colonising the vagina seems to play a major role in the sustained colonisation and may reduce susceptibility to BV.

From the large number of studies conducted globally it is apparent that *L. crispatus*, *L. jensenii*, *L. iners* and *L. gasseri* are the dominant lactobacilli. The knowledge gained from these studies offer an insight into the effects of BV on the vaginal microbiome. Future focus needs to be on the use of lactobacilli as both a probiotic for BV and a live microbicide for HIV prevention.

1.7 Other associated vaginal microbes

As mentioned in the previous section there are women that lack substantial numbers of lactobacilli, yet do not exhibit any BV-related symptoms. These asymptomatic women are known to be colonised by diverse microbial communities, which are found to be facultative or strictly anaerobic. These microflorae include *Atopobium*, *Corynebacterium*, *Anaerococcus*, *Peptoniphilus*, *Prevotella*, *Gardnerella*, *Sneathia*, *Eggerthella*, *Mobiluncus* and *Fingoldia* species. The pH of vaginas colonised by these bacteria are marginally higher than those colonised with lactobacilli, with a pH of 5.3-5.5. It should be noted that many vaginal microorganisms, such as members from the genera *Atopobium*, *Streptococcus*, *Staphylococcus*, *Megasphaera* and *Leptotrichia*, also undergo lactic acid fermentation (Ma *et al.* 2012).

1.8 Bacterial Vaginosis

Bacterial Vaginosis (BV) is the most common lower genital tract infection found in women of childbearing age (Martin *et al.* 1999; Eade *et al.* 2012). BV is known to impact between

60% of women globally BV can be seen as a shift in the normal vaginal microbiome, with a decrease in lactobacilli and an increase in the numbers of aerobic, anaerobic and micro-aerophilic bacteria such as *Gardnerella vaginalis*, *Mycoplasma homines*, *Ureaplasma urealyticum*, *Peptostreptococcus* and *Mobilancoccus* species, as well as *Prevotella* and *Bacteroides* species (Antonio *et al.* 2009; McLean & Rosenstein 2000; Pavlova *et al.* 2002).

Through molecular analysis a further group of bacteria has been identified as being associated with this condition (Ngugi *et al.* 2011; Tamrakar *et al.* 2007). These are *Atopobium vaginae*, *Megasphaera*, *Leptotrichia* and *Eggerthella*-like species (Ngugi *et al.* 2011; Tamrakar *et al.* 2007). For many years, it has been believed that *G. vaginalis* was the only causative agent of BV (Turovskiy *et al.* 2011). However, it is now emerging that BV is not a result of the presence of the potential pathogens but rather an increase in the number of these pathogens above normal levels (Turovskiy *et al.* 2011; Teixeira *et al.* 2012). Over and above this, the condition is associated with a decrease in lactic acid producing bacteria (Ma *et al.* 2012), further suggesting the potential importance of lactic acid in protecting the vagina. It is not clear whether the microorganisms associated with BV are pathogenic in nature or merely opportunistic organisms that proliferate when the vaginal pH increases (Ma *et al.* 2012).

BV is thus associated with a complex mixture of pathogens, which may or may not interact with each other. In a study by Fredricks *et al.* (2005) where 73 subjects (27 with BV and 46 without BV) were analyzed, they found a high assortment of bacteria in BV infected subjects compared to healthy subjects, further suggesting that this condition is not dependent on the presence of a single organism (Fredricks, Fiedler & Marrazzo, 2005). This is substantiated by the finding that *G. vaginalis* is present in the genital tract of healthy females at low concentrations (Antonio *et al.* 2009; Vitali *et al.* 2007). This has fueled the debate among researchers regarding the contribution of *G. vaginalis* in the acquisition and progression of this condition. For a long time, no animal model existed for the *in vivo* study of BV to establish the impact of *G. vaginalis* on this condition. However, a study conducted by Gilbert *et al.* (2013) demonstrated the successful use of a murine model to study the clinical features of BV with *G. vaginalis* as the primary pathogen. This study successfully proved that *G. vaginalis* is involved in the acquisition and the phenotypic symptoms associated with BV. This model could also successfully mimic the BV condition found in the human vagina, which is the presence of sialidase activity and clue cells (Gilbert *et al.* 2013). Machado *et al.* (2013) looked at the ability of *G. vaginalis* to colonize the vagina when vaginal lactobacilli are present, as well as their ability to allow for further colonisation of other BVAB. In this

study they found increased *G. vaginalis* adherence when *L. crispatus* was present, and also found growth of *Prevotella bivia* and *Fusobacterium nucleatum* as a result of *G. vaginalis* colonisation suggesting that this species may be involved in the biofilm formation associated with BV (Machado *et al.* 2013).

Clinically, BV is diagnosed if a woman has three of the four symptoms described by Amsel *et al.* (1983). These symptoms are: i) vaginal discharge, ii) vaginal pH above 4.5, iii) the presence of 'clue' cells and iv) the release of amines (Amsel *et al.* 1983; Martin *et al.* 1999; Yan *et al.* 2009). However, as BV is also seen in asymptomatic women, the Nugent scoring system is used for further detection of the condition (McLean & Rosenstein 2000; Turovskiy *et al.* 2011). BV is known to cause Pelvic Inflammatory Disease, post caesarean delivery endometriosis, chorioamnionitis, premature rupture of membranes, late miscarriage and preterm labour (Antonio *et al.* 2009; Martin *et al.* 1999; Verstraelen *et al.* 2009). Women suffering from this condition are more susceptible to HIV, human papillomavirus (HPV) and sexually transmitted infections (STI) (Antonio *et al.* 2009; Anukam *et al.* 2006; Ngugi *et al.* 2011; Sha *et al.* 2005).

Risk factors associated with BV include cigarette smoking, douching, antibiotic treatment, unprotected sex and the use of intrauterine devices (Martin *et al.* 1999; Turovskiy *et al.* 2011). Condom usage may reduce this risk. In a study by Ma *et al.* (2013) it was found that condom usage increased the colonisation of microflora associated with a healthy vaginal mucosa (Ma *et al.* 2013). It has also been suggested that race may be a risk factor, as the incidence is higher in women of African ethnicity (Anukam *et al.* 2006; Turovskiy *et al.* 2011) with 10-20% white women and 30-50% African American women affected. It is estimated that BV may affect up to 85% of sex workers in Africa (Fettweis *et al.* 2012; Hickey *et al.* 2012). However, these risk factors are not the only determinants as clinical studies have found that many women without these risk factors acquire BV. The usual BV treatment involves the use of antibiotics such as metronidazole (oral or vaginal gel) or clindamycin vaginal cream.

1.9 Lactobacilli as a live microbicide

Many studies have investigated the potential benefits of externally applied probiotics, particularly regarding their use in treating viral infections. For example, lactobacilli have already been engineered to secrete HIV inhibitors and were also shown to inhibit bacteria associated with HIV susceptibility, therefore may serve as a possible preventative method

against HIV (Chancey *et al.* 2006; Damelin *et al.* 2010, 2011). In addition to this, the external application of lactobacilli to the vagina has been shown to prevent bacterial vaginosis (BV) associated with HIV infection (Damelin *et al.* 2010, 2011; Liu *et al.* 2006, 2008). Although the use of lactobacilli in HIV studies is not the focus of this study, numerous studies have been done where lactobacilli was genetically modified and used as a live microbiocide. These studies illustrate the potential to genetically modify lactobacilli for probiotic applications and for this reason I will highlight some examples which successfully demonstrates the potential of lactobacilli as a live microbiocide. The vaginal lactobacilli are good candidates for the live microbiocide approach as they show ideal growth characteristics (i.e. low pH and low oxygen requirements), have good adherence to mucosal cells and can be genetically manipulated (Liu *et al.* 2006). The use of endogenous lactobacilli that naturally colonise the vaginal mucosa is of great importance as it would limit the introduction of exogenous lactobacilli which may compete with resident microflora (Damelin *et al.* 2010, 2011). A good microbiocide must be cost effective and stable (Damelin *et al.* 2010; Liu *et al.* 2006). Several studies have engineered lactobacilli to secrete HIV inhibitors. Liu *et al.* (2006) were able to engineer a *L. jensenii* strain to secrete cyanovirin-N (CV-N) an inhibitor of CCR5 tropic HIV (Liu *et al.* 2006). This was achieved by creating a CV-N expression cassette that was able to produce CV-N *in vitro* and *in vivo* at levels capable of inhibiting CCR 5. Recently Yamamoto *et al.* (2013) were able to genetically modify *L. jensenii* strain 1153 to secrete a modified cyanovirin-N (mCV-N). This strain was also found not to have any detrimental effect on the human vaginal epithelial cells (did not induce an immune response). The ability of *L. jensenii* to form sustained colonies and its ability to secrete the HIV inhibitor mCV-N suggests that it also has the potential to be used as a live microbiocide (Yamamoto *et al.* 2013). Similarly, Chancey *et al.* (2006) were able to genetically engineer *L. casei* to secrete single chain variable fragments (scFV), which are proteins that contain variable regions of heavy (VH) and light (VL) chains. These are fused together by peptide linkers and can be expressed in *E. coli* (Chancey *et al.* 2006). These scFVs were able to disrupt cell associated HIV-1 transmission. This was achieved by transforming the *L. casei* strain 393 to secrete scFV. *L. casei* was shown to secrete scFV at levels required to inhibit HIV-1 p24 (Chancey *et al.* 2006). Liu *et al.* (2008) were able to genetically manipulate *L. jensenii* to express 2D CD4 molecules which are anti-viral proteins that act against HIV-1. This was achieved by transforming *L. jensenii* 1153 to express surface anchored 2D CD4 molecules (Liu *et al.* 2008). Similarly, Damelin *et al.* (2010) genetically manipulated *L. gasseri* via transduction to secrete cc chemokines and HIV

co-receptor antagonists CCL 5 and CCL 3 (Damelin *et al.* 2010). These studies illustrate that lactobacilli can be successfully transformed and could be used in future probiotic applications.

1.10 Probiotics for BV

Many vaginal probiotics have limited ability to provide sustained protection against BV and prevent recurrence. This failure suggests that either the incorrect *Lactobacillus* strains are being used as probiotics or that other factors are influencing the recurrence of BV. Another explanation could be that the antibiotics being used could be destroying the natural microbiome (Antonio *et al.* 2009).

Many clinical studies have been conducted to investigate the efficacy of BV-preventative therapeutics and pro-biotics. In a study conducted by Leite *et al.* (2011), the efficacy of a Brazilian pepper tree extract (*Schinus terebinthifolius* Raddi) was compared with metronidazole vaginal gel on BV. It was found that the vaginal metronidazole (56.4%) was more effective than the Brazilian pepper tree extract (21.2%). In addition, higher levels of lactobacilli were detected after metronidazole treatment (65%) compared to the Brazilian pepper tree extract (45%). This indicates that metronidazole treatment is not only able to reduce the condition but also ensures that the healthy microbiota is not affected (Leite *et al.* 2011). In a study conducted by Antonio *et al.* (2009) the potency and colonisation efficiency for *L. crispatus* CTV-05 was evaluated, when 87 sexually active women were treated with 2 gelatine capsules. The one group of women were treated with a gelatine capsule containing 10^6 CFU of *L. crispatus* CTV-05 and the other group was treated with a gelatine capsule containing a 100-fold more (10^8 CFU) *L. crispatus* CTV-05. In this study, they found that 36 of the 40 women that initially lacked *L. crispatus* (at enrolment) were colonised by *L. crispatus* CTV-05 after the initial treatment whereas only 24 of the 47 women who initially had *L. crispatus* (at enrolment) were colonised by *L. crispatus* CTV-05. In addition to this it was found that sexual activity during treatment reduced *L. crispatus* CTV-05 colonization. This was less prominent among women who engaged in protected sex. This may be due to the elevated pH associated with semen exposure during sexual intercourse or the semen may have inhibited lactobacilli adhesion. This also illustrates the potential of endogenous lactobacilli to outcompete newly introduced lactobacilli (Antonio *et al.* 2009).

1.11 Present and past techniques used for vaginal microbiome studies

For many years, microbial ecology has been studied using culture dependent methods. Lactobacilli species from *L. acidophilus* complex was identified through this method including *L. crispatus*, *L. jensenii*, *L. gasserrii*, *L. casei*, *L. plantarium*, *L. minutus* (subsequently renamed to *Atopobium minutus*), *L. plantarium*, *L. fermentum*, *L. cellobiosus*, *L. brevis* and *L. salivarius* (Hickey *et al.* 2012). In earlier studies *L. acidophilus* was identified as the dominant lactobacilli colonising the human vagina (Martínez-Peña *et al.* 2013; Burton *et al.* 2003; Pavlova *et al.* 2002). This has been shown not to be the case and in many studies it has been shown that in healthy women *L. crispatus*, *L. jensenii*, *L. gasserrii* and *L. iners* are in fact the dominant lactobacilli (Antonio *et al.* 1999; Martínez-Peña *et al.* 2013; Zhou *et al.* 2010).

Culture-based methods however, have limitations with regards to the variety of organisms that can be identified, as the majority are not able to be cultivated in the laboratory and go unidentified (Hickey *et al.* 2012; Ma *et al.* 2012). Previously undetected *L. iners*, which does not grow on conventional MRS or Rogosa-Sharp medium was overlooked in many vaginal microbiome studies as they have been grouped with one of the other lactobacilli (Anukam *et al.* 2006; Burton *et al.* 2003). These inaccuracies have led to the use of culture independent methods which are more sensitive, faster and have the potential for high throughput screening (Hickey *et al.* 2012). Through the analysis of genetic material extracted directly from the environment, including clinical samples, one can fully analyse the entire microbiome (Hickey *et al.* 2012; Ma *et al.* 2012). Molecular techniques such as Polymerase Chain Reaction-Denaturing Gradient Gel Electrophoresis (PCR-DGGE), Quantitative Polymerase Chain Reaction (qPCR), Randomly Amplified Polymorphic DNA (RAPD)-PCR, Fluorescent In Situ Hybridization (FISH) and Restriction Fragment Length Polymorphism (RFLP) have been used to accurately determine which lactobacilli are present in the female vagina (Anukam *et al.* 2006; Motevaseli *et al.* 2013; Pavlova *et al.* 2002). qPCR method detects DNA amplification in real time through the use of fluoresce dyes and probes. This method is employed for pathogen detection, gene expression analysis and single nucleotide polymorphism analysis (SNPs) (Deepak *et al.* 2007; Kubista *et al.* 2006). This method is considered to be relatively accurate for gene expression measurement studies and for this reason was used in this study. Further discussion on the method is covered in chapter 4.1.

1.12.1 Plasmids

Plasmids are additional chromosomal material capable of replicating independently from the host genome (Hughes *et al.* 2012; Svara & Rankin 2011; Yates *et al.* 2006). They play a vital role in the host survival mechanisms such as virulence, drug and heavy metal resistance, degradation of toxic compounds, and generally serve to improve host fitness (Kunnimalaiyaan & Vary 2005; Yates *et al.* 2006). However, the maintenance and persistence of plasmids has been shown to hamper the competitive fitness of the bacterial host when the selective pressure is no longer present in the environment, such as loss of antibiotics selection. The burden imposed by having the plasmid can be reduced or eliminated by changes to either host or plasmid genomes (Lenski *et al.* 1994; Yates *et al.* 2006). There are also plasmids with no apparent function yet they are continually maintained by the host, which are commonly referred to as 'Cryptic plasmids' (Heiss *et al.* 2015).

The two main modes of plasmid replication that have been identified are theta replication and rolling circle replication. Rolling circle (RC) replication is the common replication mode for small plasmids between 1.3-10 kb in size. This mode of replication requires a replication protein, double stranded origin and the single stranded origin (Heiss *et al.* 2015; Pan *et al.* 2011; Xi *et al.* 2013). RC plasmids are often unstable and promiscuous. RC replication was initially discovered in *Staphylococcus aureus*, but has since been found in many Gram-positive bacteria such as *Bacillus subtilis*, *Clostridium butyricum*, *Brevibacterium lactofermentum*, *Streptococcus agalactiae*, *Lactococcus lactis*, *Leuconostoc lactis*, and *Streptomyces* species, as well as Gram-negative bacteria such as *Actinobacillus actinomycetemcomitans*, *Bacteroides*, cyanobacteria, *Helicobacter pylori*, *Selmonas ruminantium*, *Shigella sonnei*, and *Zymomonas mobilis*. Theta replication is predominantly found in Gram-negative bacteria but also been identified in some Gram-positive species. Bacteria such as streptococci/enterococci, some *Lactococcus* species and at least one *Bacillus subtilis* plasmid have been shown to exhibit Theta replication. Theta replication requires the melting of the parental strands, production of a primer RNA (pRNA) and initiation of DNA production via covalent extension of the pRNA (del Solar *et al.* 1998).

The host in which the plasmid can replicate and be maintained is known as the host range. Plasmids can be grouped into narrow host range and broad host range (BHR) plasmids. BHR plasmids are defined as plasmids that can be transferred and maintained in phylogenetically diverse bacteria. These plasmids are of particular interest owing to their role in horizontal gene transfer and their replicons are a good source for vector construction (Brown *et al.* 2013;

Jain *et al.* 2013).

1.12.2 Lactobacilli plasmids

As mentioned in the previous sections lactobacilli are often used as probiotics to enhance human health. Plasmids found in certain probiotic lactobacilli have been found to play a vital role in their host survival. One such example is the *Lactobacilli Brevis* KB 290, which carries 9 plasmids that have been found to encode for proteins related to stress response and resistance (Fukao *et al.* 2013). Another is *Lactobacilli paracasei* NFBC 338 a gastrointestinal tract (GIT) probiotic, which contains numerous plasmids. These plasmids functions are possibly linked to biotin metabolism of the *L. paracasei* NFBC 338 as well as adherence (Desmond *et al.* 2005). Probiotic strains of vaginal lactobacilli were analysed by Martin *et al.* 2008. In this study, it was found that the majority of strains (90% of isolated probiotic strains) had plasmids (Martin *et al.* 2008). These plasmids may play a role in enhancing vaginal colonisation. However, very little to no work have been done with the focus solely on the plasmids isolated from vaginal lactobacilli.

1.13 Research Objectives

BV affects countless number of women worldwide. It is seen as a shift from a healthy vaginal microbiome composed of lactobacilli to an unhealthy vaginal microbiome composed of an array of microorganisms. This condition is of clinical significance as it is associated with an increased susceptibility to HIV and sexually transmitted diseases. Two plasmids (pLc4 and pLc17) were isolated and identified from healthy and BV infected samples from *Lactobacillus* strains used in the study conducted by Damelin *et al.* (2011). Plasmids are known to confer advantages to their host and may enhance their host survival in adverse conditions. The aim of this project is to investigate the potential role that pLc4 and pLc17 play in healthy and BV infected women. This will be done by firstly sequencing both plasmids, followed by the bioinformatic characterisation of both plasmids. A second objective is to measure the relative abundance of both plasmids in healthy vs BV infected women. The results from this study will give us a better indication of whether these plasmids play any role in the *L. crispatus* and *L. jensenii* fitness.

CHAPTER 2 MATERIALS AND METHODS

2.1 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 2.1. Stock cultures were maintained at 25% [v/v] glycerol cell suspensions at -80°C. *Lactobacillus* strains were isolated from vaginal swabs obtained in the Damelin *et al.* (2011) study.

Table 2.1: Bacterial strains used in this study

Organism	Sample Identification	Patient condition	Source/Supplier
<i>L. crispatus</i>	L1	Healthy patient	Swab samples from the Damelin <i>et al.</i> (2011) publication were used in the present study.
<i>L. crispatus</i>	L2	BV infected patient	
<i>L. jensenii</i>	L3	Healthy patient	
<i>L. jensenii</i>	L4	BV infected patient	
<i>L. crispatus</i>	L5	Healthy patient	
<i>L. crispatus</i>	L6	BV infected patient	
<i>L. crispatus</i>	L7	Healthy patient	
<i>L. crispatus</i>	L8	BV infected patient	
<i>E. coli</i>	E1	N/A	Affymetrix USB

2.2 Media and growth conditions

All media components were supplied by Sigma Aldrich Chemical Company (St. Louis, Missouri, USA) and Merck Chemical and Laboratory Supplies (Darmstadt, Germany). Unless otherwise stated the autoclave sterilization conditions were set at 121°C for 15 minutes.

The following media were used in this study:

deMan, Rogosa and Sharpe (MRS) Broth

MRS broth was used to grow lactobacilli for DNA extraction.

Constituent	L⁻¹
Special Peptone.....	10 g
Meat Extract.....	10 g
Yeast Extract	5 g
Dextrose	20 g
Sodium Acetate.....	5 g
Tween 80.....	1 g
di-Potassium Hydrogen Phosphate	2 g
tri-Ammonium Citrate	2 g
Magnesium Sulphate.....	0.1 g
Manganese Sulphate	0.05 g

This medium was aliquoted into 15 ml and 50 ml Falcon flasks before being sterilized by autoclaving, with 13 ml being aliquoted into the 15 ml falcon flask and 48 ml being aliquoted into the 50 ml flask. MRS agar medium contained 1.2% (w/v) agar.

Luria-Bertani (LB) broth

This medium was routinely used to grow bacterial strains

Constituent	L-1
Tryptone.....	10 g
Yeast Extract	5 g
NaCl	10 g
dH ₂ O up to.....	1000 ml

LB agar (LBA) medium contained 1.5 % (w/v) agar. When necessary the appropriate antibiotic was added after autoclaving.

2.3 Plasmid extraction

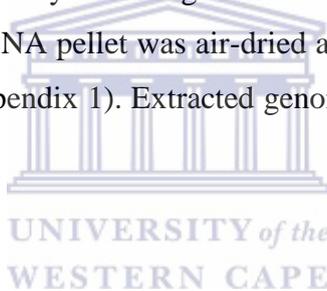
Plasmid DNA was extracted from *Lactobacilli* listed in Table 2.1 using the Zyppy™ Plasmid Miniprep Kit (Zymo Research, Irvine, CA, USA) with minor modifications to the lysis step. Cultures were grown in MRS broth at 37°C overnight until an OD_{600nm} >0.8. The cells were harvested by centrifugation at 7000 rpm for 5 minutes and the pellet was resuspended in a solution containing sucrose to a final concentration of 25% and lysozyme (final concentration of 30 mg/ml) in a final volume of 600 µl (recommended input volume for the Zyppy kit). The suspension was incubated at 37°C for 1 h (O’Sullivan & Klaenhammer 1993). The remaining steps were followed as per the manufacturer’s instructions. Extracted plasmid DNA was purified as outlined in section 2.8.

2.4 Next generation sequencing of *Lactobacillus* strains

The DNA was extracted and purified using the Qiagen gel extraction kit (Qiaex II; catalog no. 20021; Qiagen, Hilden, Germany), as per manufacturer’s instructions. Sequencing libraries were prepared using the Illumina Nextera XT library preparation kit (Illumina, San Diego, CA, USA) using 1 ng of input DNA. Libraries were individually indexed, pooled in equimolar concentrations and sequenced with MiSeq reagent cartridge V3 (2 x 300bp). A 10% PhiX V3 spike for low diversity libraries was included as per the manufacturer’s instructions (Illumina Nextera XT guide). The fastq files generated were analysed using CLC Genomic Workbench version 6.5. Paired end reads were merged, and sequential reference assemblies to Phi X174, *L. crispatus* (FN 692037.1), Human genome (GRCh37) and *L. jensenii* (FN557015.1) genomes were performed to remove contaminating DNA sequences. A *de novo* assembly was performed using all the remaining unmapped reads. The *de novo* assembly was performed using similarity and length fractions of 0.9 and 0.95, respectively, with scaffolding turned off. Annotated plasmid sequences are available on GenBank under the accession numbers KR052811 and KP984530 for pLc17 and pLc4, respectively. Softberry (www.softberry.com) and CLC Genomics Workbench were used for ORF prediction. BLAST analysis was performed to identify related ORFs using megaBLAST against the NCBI database, or in the case of replication proteins against the ACLAME database (<http://aclame.ulb.ac.be/>). Circoletto (<http://tinyurl.com/he9cqhm>) was used for plasmid comparisons. For phylogenetic tree construction, the full-length amino acid sequences of selected terminase proteins were aligned using MEGA6 (Tamura *et al.*, 2013).

2.5 Genomic DNA extraction

Genomic DNA extraction from lactobacilli isolates was done according to the method described by Wang *et al.* (1996) with minor modifications. Cells were harvested by centrifugation at 9449 rpm for 2 minutes. The pelleted cells were resuspended in 500 µl of lysozyme buffer (25 mM Tris-HCl pH 8, 50 mM glucose, 10 mM EDTA, 8 mg Proteinase K and 25 mM lysozyme) and incubated at 37°C for 2 hours/overnight. SDS was added to a final concentration of 1% and the tubes were mixed by inversion and incubated at 65°C for 30 minutes. An equal volume of phenol was added to the samples and the tubes were mixed by gentle inversion and centrifuged at 10774 rpm for 10 minutes. The upper aqueous phase was transferred to a new Eppendorf tube, an equal volume of chloroform:isoamyl alcohol (24:1; v/v) was added and the tubes were mixed gently. The mixture was centrifuged as before and the upper aqueous phase was transferred to a new tube. DNA was precipitated with 1 volume of ice-cold isopropanol, followed by centrifugation at 9449 rpm for 5 minutes and the supernatant was discarded. The DNA pellet was air-dried and resuspended in 100 µl of 1xTE buffer and stored at 4°C (See Appendix 1). Extracted genomic DNA was purified as outlined in section 2.8.



2.6 PCR amplification

PCR amplification was performed in 0.2 ml thin walled tubes using thermocyclers equipped with a heated lid. Primers (Inqaba Biotech, Pretoria, South Africa) employed in this study are listed in Table 2.2. PCR amplification was carried out containing the following (final concentrations) 1X Dream *Taq* Buffer, 2 mM dNTPs, 1µl of template DNA (>50 ng) and 1.25 U of Dream *Taq* DNA Polymerase (Thermo Fisher Scientific, Massachusetts, USA). Reactions were made up to a final volume of 50 µl with distilled water. PCR amplification products were purified as outlined in section 2.8.

Table 2.2: Primers used in this study for PCR amplification of genes

Primer Set	Sequence (5'-3')	Amplification cycle	Specificity	Reference
16S-F1	AGAGTTTGATCITGGCTCAG	98°C for 3 min, 45x (95°C for 10 s, 60°C for 20 s, 72°C and 72°C for 1 s). The ramp rate was set to 0.5°C/s.	Bacterial universal 16S rRNA	Babalola <i>et al.</i> (2009)
16S-R5	ACGGITACCTTGTTACGACTT			
pL17-F	TTAGAAATGCAGTTGCGTGG	95°C for 3 min, 34x (95°C for 30 s, 58°C for 30 s, 72°C for 1.35 min and 72°C for 7 min).	pLc17 junction designed to bind to the “start” and “end” of the assembled contig	This study
pL17-R	AATTGCAAAAAGTGCTGTCTG			
pL4-F	AGCACTTTAGGCGATCGATA	95°C for 3 min, 34x (95°C for 30 s, 58°C for 30 s, 72°C for 1.35 min and 72°C for 7 min).	pLc4 junction designed to bind to the “start” and “end” of the assembled contig	This study
pL4-R	ACTCCGCCTTGCTCCCGTGA			
EGFR E18 F	CTGGCACTGCTTTCCAGCAT	95°C for 3 min, 40x (95°C for 5 s, 60°C for 20 s and 72°C for 1 s). The ramp rate was set to 0.5°C/s.	Human epidermal growth factor receptor (EGFR) Gene	Lin <i>et al.</i> (2010)
EGFR E18 R	GCTTGCAAGGACTCTGGGCT			
qPL17F	AGGTACTCCTGCTACGCTCA	95°C for 3 min, 40x (95°C for 5 s, 60°C for 20 s and 72°C for 1 s). The ramp rate was set to 0.5°C/s.	pLc17, with forward and reverse primers target ORF 8	This study
qPL17R	ACAACCTGCGGAAAACCAACG			
qPL4F	GATTTCGCTCTTGCTGAAG	95°C for 3 min, 40x (95°C for 5 s, 60°C for 20 s and 72°C for 1 s). The ramp rate was set to 0.5°C/s.	pLc4, with forward and reverse primers target ORF 1	This study
qPL4R	ACACTCGCCATCTTCTTGGT			

2.7 Analytical Techniques

2.7.1 DNA quantification

For routine quantification DNA concentrations were determined using a Nanodrop ND-1000 spectrophotometer. DNA was resuspended in double distilled sterile water. For protocols requiring more accurate quantification the DNA concentration was determined using the Qubit™ DNA BR assay kit (Thermo Fisher Scientific, Massachusetts, USA) according to the manufacturer's instructions.

2.7.2 Agarose gel electrophoresis

Genomic DNA, plasmid DNA and PCR products were analysed using agarose electrophoresis. Ethidium bromide solution (0.5 µg/ml) was added to the molten agarose before the gels were cast. Samples were prepared by mixing 3-20 µL of DNA or PCR products with 3 µL of 6x loading buffer (Appendix 1) prior to loading the agarose gels. Electrophoresis was performed in 1xTAE buffer at 80-100V. DNA band sizes were determined according to their migration in the gel as compared to a molecular weight marker (phage lambda DNA digested with *Hind*III or *Pst*I endonuclease). Gels were visualised via ultraviolet (UV) light illumination at a peak wavelength of 302 nm and photographed with a digital imaging system (AlphaImager 2000, Alpha innotech, San Leandro, CA).

2.8 Amplicon DNA purification

DNA fragments were viewed under the UV illumination at a peak wavelength of 302 nm and excised from the agarose gel using a sterile scalpel blade. The NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, CA, USA) was used to purify the gel pieces according to the manufacturer's instructions.

2.9 Relative quantitative PCR

2.9.1 Standard curve construction

Plasmid DNA for pLc4 and pLc17 was extracted as described in section 2.3. The DNA copy number was calculated using the formula:

$$\text{number of copies} = \frac{(\text{amount} \times 6.022 \times 10^{23})}{(\text{length} \times 1 \times 10^9 \times 650)} = \frac{(\text{ng} \times \text{number/mole})}{(\text{bp} \times \text{ng/g} \times \text{g/mole of bp})}$$

The quantified plasmid DNA was serially diluted (10 fold dilutions) to obtain six standards ranging from 10-1000000 copies/µl of extracted sample for pLc4 and pLc17. For the human reference gene, the epidermal growth factor receptor (EGFR) was selected and an amplified product was obtained from Human Genomic DNA (Roche Diagnostics, Rotkreuz,

Switzerland) and diluted from 10-10000 copies/ μ l of extracted sample. The qPCR amplification was performed using a Rotor-gene® Q (QIAGEN) with the Kapa SYBR® Fast qPCR Master mix (2x) kit (Kapa Biosystems, Cape Town, South Africa). Human Genomic DNA (Roche Diagnostics, Rotkreuz, Switzerland) and plasmid DNA was added to each 10 μ l reaction containing 10 μ M of each of the forward and reverse primers and Master mix. The real time PCR primers were designed to meet the following criteria: primer length 18-24bp, T_m :59-68°C, Annealing Temperature: 59 or 60°C and product size: 80-150bp. Primers were designed using *Primer3* software. The thermal cycling parameters are as described in Table 2.2. The melting curve was measured by cycling from an annealing temperature of 60°C through to a melting temperature of 95°C at a ramp rate of 0.5 °C every 2 seconds. The plasmid extracted DNA was stored in 10 mM Tris/0.01% Tween-20 buffer (Appendix 1).

2.9.2 Swab sample preparation and quantification

Twenty-eight vaginal swabs were submerged in 1 ml of sterile normal saline (prepared with RNase free H₂O, pH 7.0) and vigorously agitated to dislodge adhering bacterial cells. Bacterial DNA was extracted from the vaginal swabs using Zyppy™ Genomic DNA clean-up Kit (Zymo Research, Irvine, CA, USA) with minor modifications to the lysis step (Section 2.3). The extracted swab samples were subjected to qPCR amplification as described in Section 2.9.1 in 10 μ l reaction volumes. The extracted DNA was stored in 10 mM Tris/0.01% Tween-20 buffer (Appendix 1).

2.9.3 Bacterial genomic DNA test

Escherichia coli (*E. coli*) genomic DNA from Affymetrix USB (Santa Clara, CA, USA) was used as a positive control for the bacterial gDNA assays and was tested at a concentration of 2.5 pg/ μ l. The qPCR amplification was performed using a Rotor-gene® Q (QIAGEN, Hilden, Germany) with the Kapa Probe Force qPCR Master mix (2x) kit (Kapa Biosystems, Cape Town, South Africa). DNA extracted from the swab samples were added to each 10 μ l reaction containing 10 μ M of the 16S rRNA forward and reverse primers and Master mix. The thermal cycling parameters are as described in Table 2.2.

2.9.4 Relative Expression Analysis

The relative PCN was determined using the Relative Expression Software Tool (REST: <http://rest.genequantification.info>). This is Excel®-based and programmed in Visual Basic and compares several gene expressions on Critical quantification (Cq) level. It compares two treatment groups, with multiple data points in the sample versus control groups, and

calculates the relative expression ratio between them. The mathematical model used is published and is based on the mean Cq deviation between sample and control group of target genes, normalized by the mean Cq deviation of one reference gene (Pfaffl *et al.*, 2002).



CHAPTER 3 Sequence Analysis using NGS

3.1 Introduction

Next generation sequencing (NGS) has revolutionized almost all fields of biological science through DNA sequencing (Ergan *et al.* 2012). As millions of DNA fragments can be sequenced in parallel, NGS technology allows enormous volumes of data to be generated at a relatively low cost compared to traditional Sanger sequencing (Bao *et al.* 2011; Buermans & den Dunnen, 2014; Luca *et al.* 2012; Metzker 2010; Sanger *et al.* 1977).

Despite the overwhelming benefits of NGS, researchers must be aware that the experimental procedures employed in both the wet lab (i.e. DNA extraction and library preparation method used) and dry lab (data analysis software, databases etc.) can all introduce bias and therefore ultimately influence the final results obtained. Optimization of both the wet and dry lab protocols are important to ensure that the results obtained are reproducible and accurate (Weber *et al.* 2017). DNA extraction methods play a major role in the recovery of DNA for sequencing. In bacterial metagenomic DNA extraction protocols the main challenges include DNA extraction methods which favour extraction from specific genera or DNA species (genomic over plasmid etc.) and DNA extraction methods which result in poor quality, sheared DNA. In addition, low DNA concentrations can influence results (Weber *et al.* 2017; Yuan *et al.* 2012). Low DNA yields can be overcome by using PCR amplification but this often results in biases being introduced in the sequencing (Oyola *et al.* 2012). Poor quality DNA can also result in the formation of chimeras during PCR amplification (Yuan *et al.* 2012). Thus a suitable extraction method or extraction kit should be used. Optimization of the software required to evaluate the DNA sequenced is equally important (Koonin & Galperin 2003).

Before sequencing is performed, sequencing libraries must be prepared. While there are several methods which can be used to prepare DNA libraries, Illumina's transposon-based technology is currently the preferred method for many applications. For library preparation using the Illumina Nextera XT kit, 1ng of input DNA from a genomic source is required. The DNA undergoes an enzymatic reaction, whereby an engineered transposome simultaneously fragments and tags the DNA, in a process called "tagmentation". Next, the fragmented DNA undergoes limited cycle amplification with primers specific to the transposon sequence.

Specific adapters and barcodes are added to the DNA fragments (Marine *et al.* 2011) (Figure 3.1).

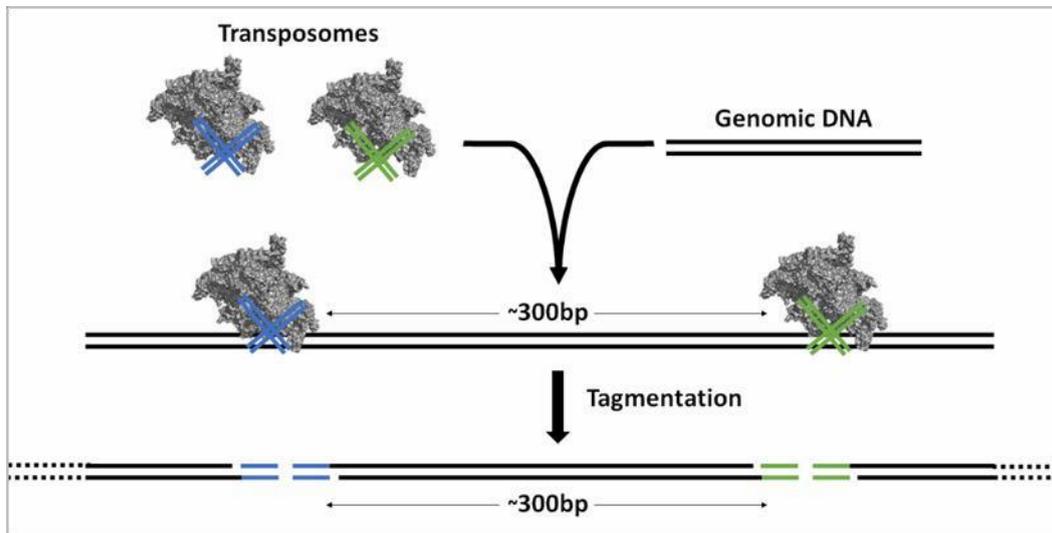


Figure 3.1 Enzymatic tagmentation using Illumina Nextera XT library preparation. Genomic DNA undergoes enzymatic reactions whereby an engineered transposome simultaneously fragments and tags the input DNA, in a process called “tagmentation (Adapted from Nextera DNA Library prep Illumina, San Diego, California, U.S.A.)

The DNA fragments are denatured and hybridised on a silica flow cell containing complementary oligos. This results in the flow cell being populated with a lawn of template molecules that will be used for the generation of clonal clusters. During cluster generation hybridization is enabled by the first of the two types of oligos. The first type of oligo is complimentary to the adapter region on one of the template strands and a polymerase makes a complimentary strand of DNA (Figure 3.2 E). This double strand is then denatured and the unbound strand is washed away (Figure 3.2 F). Template DNA is clonally amplified through bridge amplification. This occurs when the newly formed fragment folds over and the adapter region hybridizes to the second type of oligo by polymerisation and makes the complimentary strand forming a double stranded bridge (Figure 3.2 G-I). The bridge is then denatured resulting in two single copies of the molecules (i.e. forward and reverse strand) (Figure 3.2 J). This process is repeated resulting in millions of clonal copies per cluster (Figure 3.2 K). After bridge amplification the reverse strands are cleaved and washed off leaving the forward strand (Figure 3.2 L). To avoid unwanted priming the 3' ends are chemically blocked (Figure 3.2 M). The clonally copied DNA is sequenced by incorporating a single complementary base pair into the strand through DNA polymerization (Figure 3.2 N). The modified terminator

dNTP contains a fluorescent label, which allows it to be detected by a camera upon addition to the growing DNA strand (Figure 3.2 O, P and Q) (Janitz 2008). After detection, the 3'OH block containing the fluorescent label is cleaved and washed away. The process of a single nucleotide being incorporated, detection and cleavage is repeated for a predetermined number of cycles.

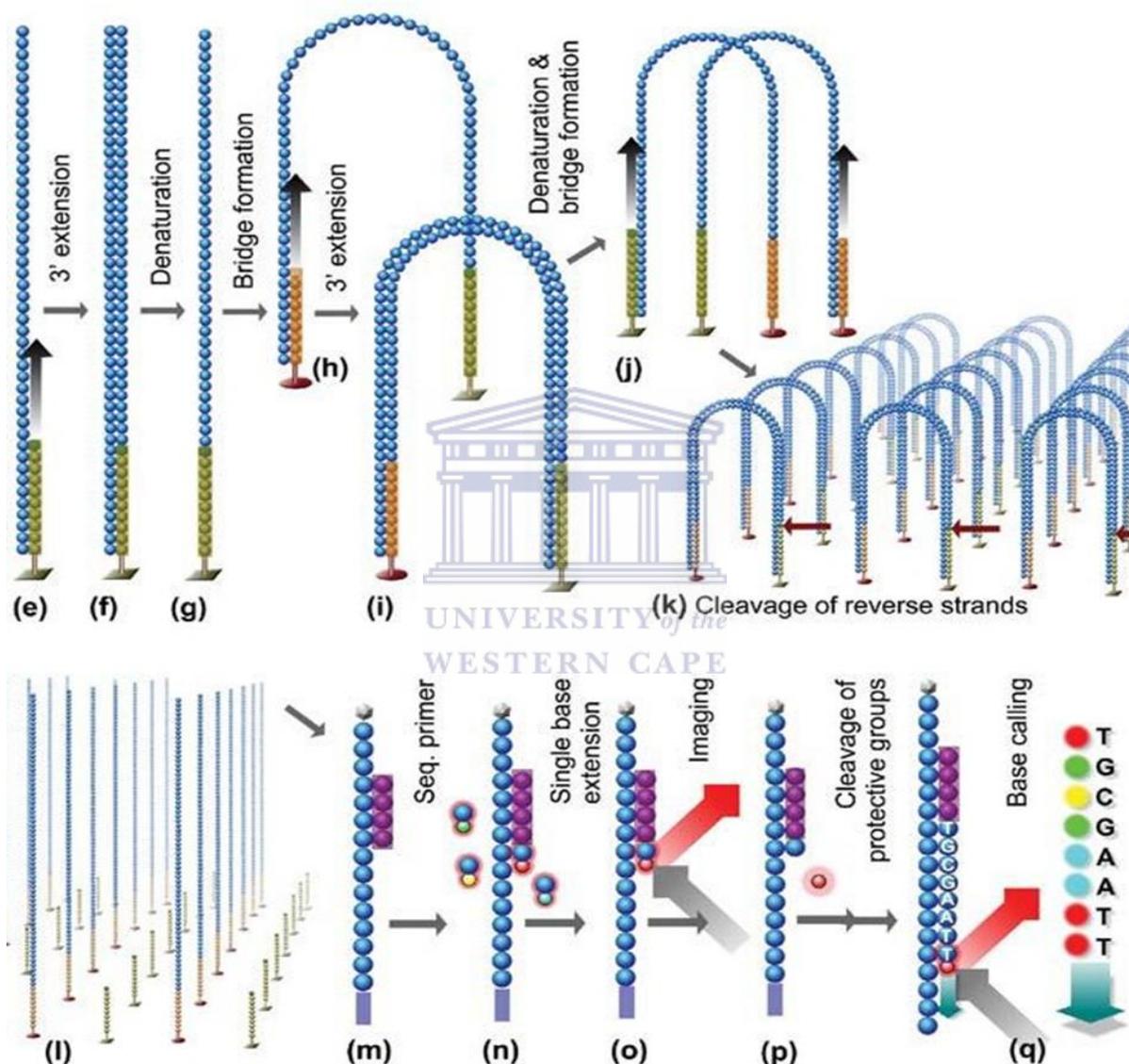


Figure 3.2: Summary of the sequencing-by-synthesis (SBS) process. E) Synthesis of complimentary DNA. F) Complimentary DNA is denatured. G-I) Template DNA is clonally amplified. The newly formed fragment folds over and the adapter region hybridizes with the second type of oligo to forming a double stranded bridge. J-K) The bridge is then denatured and process then repeated resulting in millions of clonal copies per cluster. L) Reverse strands are cleaved. M) 3' ends are chemically blocked ensuring no unwanted priming. N) clonally copied DNA is sequenced. O, P, Q) The modified terminator dNTP contains a fluorescent label which can be detected by a camera upon DNA extension. Library preparation:(Taken from Janitz 2008).

The assembly of the read sequences requires specialized software known as assemblers. Several assemblers are available for various NGS technologies, such as Euler assembler, Velvet, ALLPATHS, ABySS, SOAPdenovo, CABOG, MSR-CA, String Graph Assembler (SGA), Phusion2 assembler, CLC-Workbench, IDBA-UD, SPAdes and RayMeta (Henson *et al.* 2012; Hesse *et al.* 2017). The reads produced from this sequencing-by-synthesis (SBS) reaction are assembled *in silico* using any of a number of sequence assemblers to produce longer DNA strands (contigs), which can be aligned to a reference genome if one exists or through *de novo* assembly (Hodkinson & Grice 2015; Janitz 2008)

The scaffolding stage (through read pair data) determines the order and orientation of the contigs as well as producing gaps (Miller *et al.* 2010; Henson *et al.* 2012). The assembler's efficiency is dependent on the task. A recent study performed by Hesse *et al.* (2017) assessed the efficiency of several assemblers (CLC Genomics Workbench, IDBA-UD, ABySS, SPAdes and RayMeta) on notoriously difficult viral genomes. In this study it was found that the performance of CLC Genomics Workbench outperformed the other 4 assemblers (Hesse *et al.* 2017).

Although plasmids are often viewed as not essential for the survival of the host strain, they often carry special genes that can enhance the survival of the host in harsh environments (Cui *et al.* 2015; Zhong *et al.* 2011). Plasmids in LAB have been found to play a major role in the survival of their host and have also been used in a variety of industrial applications. The sequencing of plasmids is often required to obtain valuable information related to their role in virulence, gene transfer and many advantageous characteristics that they may impart onto the host (Cui *et al.* 2015).

The determination of which replication proteins the plasmid makes use of is also important, as it allows one to group the plasmids into their respective families. RCR plasmids can be grouped into several families namely; pT181, pE194/pMV158, pC194/pUB110 and pSN2. RCR plasmids are known to have a broad host range and are mostly found in Gram positive bacteria. Theta replicating plasmids are grouped into 6 classes based on their replicons. These are class A-F (Cui *et al.* 2015; Zhai *et al.* 2009).

Lactobacillus plantarum is known to carry many plasmids with the *L. plantarum* strain 16 carrying 11 plasmids (pLp16A-pLp16L). To date 56 *Lactobacillus* plasmids have been sequenced, many of which have been found to be cryptic, small RCR plasmids. These include pM4 (pC194), pLR1 (pC194), pXY3 (pMV158), pWCF101 and pWCF102 (Cui *et al.* 2015;

van Kranenburg *et al.* 2005). While others such as pMD5051, pWCFS103 and pST-III were found to replicate *via* theta replication (Cui *et al.* 2015). The sequence analysis of pCD033 a *L. plantarum* plasmid was found to contain toxin-anti-toxin (TA) systems typically associated with plasmid maintenance. This is a strategy often used by low copy number plasmids to remain in the bacterial population and is most times seen as a 'selfish' mechanism (Heiss *et al.* 2015). The same system was also identified on p256 (Sørvig *et al.* 2005). Both plasmids utilize unique replication mechanisms. In the case of pCD033 it was found that a possibly new RNA based theta replication mode is used, while p256 was found to use the host machinery to replicate in the absence of a self-encoded Rep protein (Heiss *et al.* 2015; Sørvig *et al.* 2005).

The sequence analysis of 3 *L. plantarum* plasmids pWCFS101, pWCFS102 and pWCFS103, revealed that 2 of these (pWCFS101 and pWCFS102) were cryptic RCR plasmids belonging to the pC194 and pMV158 family respectively. The third plasmid pWCFS103 a theta replicating plasmid which encodes a RepA and RepB protein. pWCFS103 carries a large range of mobilization genes (mob genes) involved in conjugation (van Kranenburg *et al.* 2005).

Lactobacillus delbrueckii rarely carries plasmids. Sequence analysis of pDOJ1 a plasmid isolated from *L. delbrueckii* revealed that this plasmid replicated *via* a theta mechanism. Lee *et al.*, 2007 found that the fused primase helicase replication gene was similar to *Streptococcus thermophilus* and that the plasmid replicon was functional in *S. thermophilus*. This suggests that the plasmid might be from *S. thermophilus* and moved to *L. delbrueckii* through horizontal gene transfer. This, as indicated by the sequence analysis, might have been due to the mob genes (Lee *et al.* 2007).

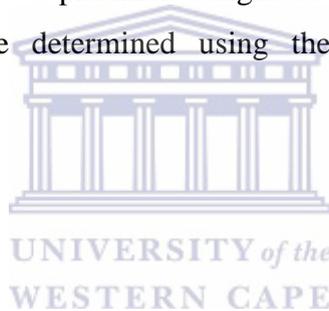
Lactobacillus sakei are known to carry one or more plasmids. A novel plasmid pYS18 was sequenced and found to be a RCR plasmid belonging to the pMV158 family. This was based on sequence similarity with pLH2 and pLC2 plasmids, which are from the pMV158 family (Zhai *et al.* 2009)

Two plasmids pCD01 and pCD02 isolated from the probiotic strain *Lactobacillus paracasei* showed that pCD01 RepA protein shared sequence homology with pLA103 and pNCD0151. pLA103, is a theta replicating plasmid from the pUCL287 family, while pNCD0151 is known to encode the Rep proteins for both RCR (belonging to pC194 family) and theta replication. The lack of a consensus nick site common for pC194-RCR families indicate that the plasmid

replicates *via* a theta mechanism. This plasmid also contained a RepB which in most theta replicating plasmids is involved in copy control. pCD02 had 2 Rep proteins, RepB found to have sequence homology with theta replicating plasmid pLA103, however the RepA protein had no sequence homology to any theta replicating plasmids. Desmond *et al.* (2005) hypothesized that this plasmid belongs to theta replicating family based on the fact that part of the replication machinery was from the pLA103 family. pCD01 encoded proteins involved in DNA replication and repair, while pCD02 encoded for large number of insertion sequences (IS), suggesting that the plasmid is highly plastic and polymorphic (Desmond *et al.* 2005).

There are challenges associated with sequencing plasmids such as issues with the assembly of the shafflon region of Inc plasmids due to the variability associated with this region most *de novo* assemblers would struggle with the assembly of the shafflon regions (Brouwer *et al.* 2015).

In this chapter the sequences of two plasmids designated pLc4 and pLc17 isolated from *L. crispatus* and *L. jensenii* were determined using the Illumina MiSeq and analysed bioinformatically.



3.2 RESULTS AND DISCUSSION

3.2.1 Plasmid Extraction

Plasmid extraction was performed on the samples using the Zyppy™ Plasmid Miniprep Kit (Zymo Research) in combination with the modification proposed by O' Sullivan (Section 2.3). The modified protocol successfully extracted plasmid DNA from *L. crispatus* and *L. jensenii* strains. Integrity of plasmid DNA was confirmed by electrophoresis (Fig. 3.3). Four bands were obtained for all samples, representing the different plasmid conformational forms that can be obtained during plasmid extraction, namely nicked, linear, supercoiled and circular, single stranded (Fig. 3.3). Supercoiled plasmid DNA is the optimal plasmid conformation as this is the native state and is most often required for cloning and transduction. However, this conformation has been shown to affect Cq scores, as well as reducing the quantification accuracy (Hou *et al.* 2010).

The lysozyme pre-treatment method described by O'Sullivan is an effective method for *Lactobacillus* species plasmid extraction, as more plasmid DNA was recovered when using the pre-treatment protocol than extraction without pre-treatment (Fig 3.3). It is probable that the effective lysis of the *Lactobacillus* cell wall is due to the high lysozyme concentration used in the protocol (O'Sullivan & Klaenhammer 1993).

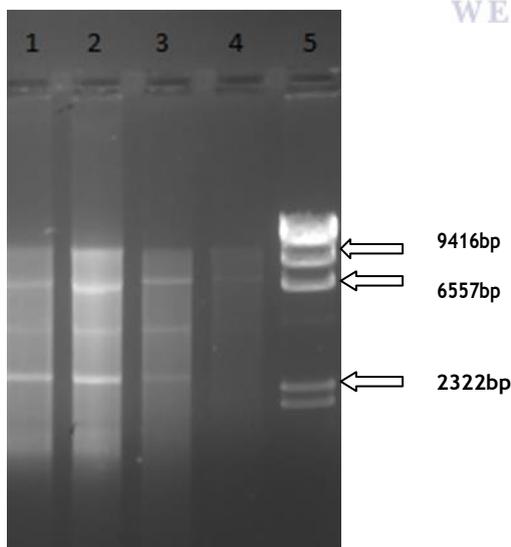


Figure 3.3: Agarose gel electrophoresis on a 1% gel of plasmid DNA extracted from *L. crispatus* pLc4 using the Zyppy kit. (1) 1ml of pellet harvested and lysozyme pre-treatment;(2) 1ml of pellet harvested and lysozyme pre-treatment;(3) 2ml of pellet harvested and lysozyme pre-treatment; (4) 2ml of pellet harvested and lysozyme pre-treatment; (5) Molecular marker: λ *Hind*III (IMBM Marker)

3.2.2 Isolation, sequencing and assembly of pLc4 and pLc17

The samples used for sequencing were plasmid extracts from *Lactobacillus crispatus* (L1, 2, 5, 6, 7, 8) and *Lactobacillus jensenii* (L3 and L4) vaginal isolates. Only plasmid DNA was expected to be extracted from the *Lactobacillus* cultures using the Zippy kit. BLASTn analysis of several of the contigs indicated that they were of *Lactobacillus* genome origin. Furthermore, high levels of human DNA contamination was also found. Despite every precautionary measure taken during library preparation, DNA contamination is a common occurrence possibly as a result of the kit or external factors (Lee *et al.* 2016). This necessitated a cleanup of the sequence data, by performing a reference assembly to the human genome and *Lactobacillus* genome data, prior to *de novo* assembly to identify plasmid sequences. The reads for the human genome were mapped (length and similarity fraction of 95%) against the human genome assembly GRCh37. Similarly, the reads for the *L. crispatus* and *L. jensenii* genomes were mapped (length fraction and similarity of 98% and 100% respectively) against the *L. crispatus* and *L. jensenii* genomes assembly FN692037.1 and FN557015.1 respectively. Following sequencing and *de novo* assembly of the remaining reads, two contigs among all the assemblies stood out as being of a suitable size and coverage to represent plasmid species and these were designated pLc17 and pLc4 (Table 3.1).

Table 3.1: Analysis of sequence assembly

Plasmid	Total Reads	Average Coverage	G+C %	Size (bp)	Average read length
pLc4	179549	10172.10	36%	4224 bp	242.60
pLc17	63640	835.59	36%	16663 bp	220.65

This identification was performed by visual inspection, and the approach may have missed mega plasmids, if present. Although unlikely, performing reference assembly may also have removed reads for related low copy number plasmids which may have been present in the reference genomes used. Interestingly, both pLc17 and pLc4 were detected in healthy and BV infected individuals (Table 3.2). pLc4 was found in both *L. crispatus* and *L. jensenii*, while pLc17 was only found in *L. crispatus*.

Table 3.2: Identification of pLc4 and pLc17 in individual samples from healthy and BV-infected individuals

Sample	Host	Patient condition	Plasmid
L1	<i>L. crispatus</i>	Healthy patient	pLc17
L2	<i>L. crispatus</i>	BV infected patient	pLc17
L3	<i>L. jensenii</i>	Healthy patient	pLc4
L4	<i>L. jensenii</i>	BV infected patient	pLc4
L5	<i>L. crispatus</i>	Healthy patient	pLc17
L6	<i>L. crispatus</i>	BV infected patient	pLc17
L7	<i>L. crispatus</i>	Healthy patient	pLc17
L8	<i>L. crispatus</i>	BV infected patient	pLc4

3.2.3 Identification of plasmids using junction primers

3.2.3.1 pLc4 and pLc17 PCR analysis

Sequence assembly resulted in a linear contig, therefore to confirm that the contig represented circular element, primer sets specific for pLc4 and pLc17 were designed to bind to the “start” and the “end” of the assembled contig to amplify the junction. Plasmid DNA was extracted from samples L2 and L8, purified by gel extraction and amplified using the junction primer sets pL17 and pL4, respectively (Table 2.2). The resulting amplicons were 1073bp for pLc4 and 1183bp for pLc17 (Figure 3.4). The amplified PCR products were gel purified, Sanger sequenced, and aligned with the sequence obtained *via* NGS. The resulting sequences matched the NGS reads of pLc4 and pLc17 100%, demonstrating that the plasmid sequences were complete.

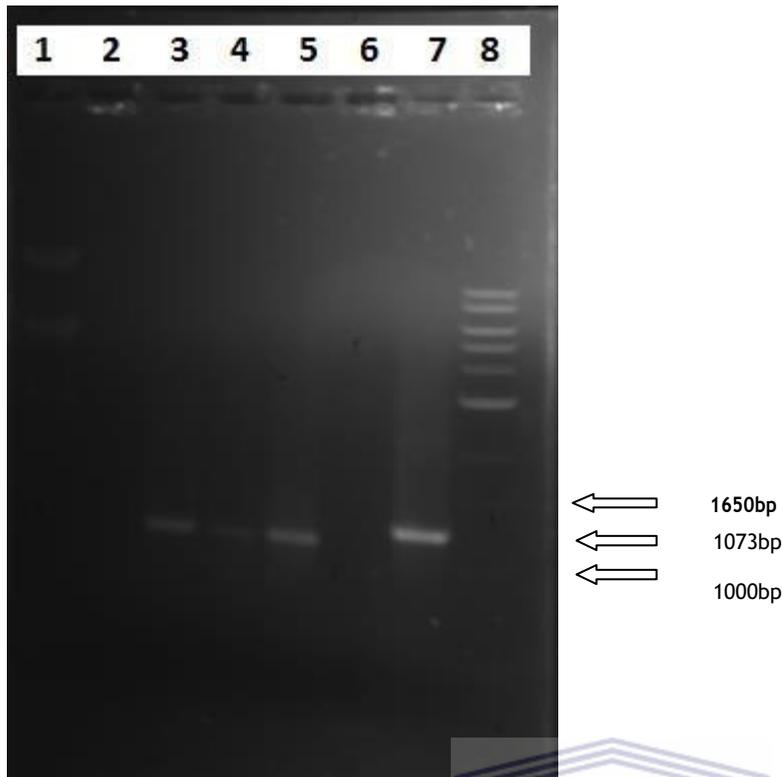


Figure 3.4: Agarose gel electrophoresis on 1% gel of PCR DNA amplification using pL4F and pL4R junction primers. (1) Molecular marker: λ *HindIII*. (2) Negative control. (3) Sample L3. (4) Sample L4. (5) Sample L8. (7) Positive control (plasmid DNA). (8) Molecular marker: 1Kb (Thermo Scientific).

3.2.4 pLc4 replication

The sequence coverage for pLc4 was very high (>10 000x), thus base calls could be made with a high level of confidence. Bioinformatic analysis revealed that pLc4 appears to be unique at the nucleotide level with no similarity to any nucleotide sequence on the NCBI database (discontiguous mega BLAST). Six ORFs were obtained after the sequences were subjected to NCBI BLASTx analysis (Table 3.3; Figure 3.5) which revealed only a few low similarity hits at an amino acid level, and this suggest that the *L. crispatus* and *L. jensenii* species from South African women may have more unique mobile elements.

Table 3.3: Putative genes and their products, deduced from the plasmid nucleotide sequences for pLc4.

ORF number	Size in amino acids	Start and end positions bp	Selected BLAST hits and comments; accession number; (length of protein on database in aa)	% Identity/Similarity (over number of aa)
1	385	+(81-1238)	Hypothetical protein <i>Streptococcus thermophilus</i> WP_011161226 (368) Replication initiation factor <i>Lactococcus lactis</i> WP_012881354(368)	44/64 (166/243) 43/66 (161/247)
2	364	+(1235-2329)	Hypothetical protein <i>Streptococcus suis</i> WP_032498644 (448) Plasmid partitioning protein <i>Lactococcus lactis</i> WP_014573491.1 (448)	32/52 (108/180) 28/49 (101/177)
3	92	+(2277-2553)	Hypothetical protein <i>Streptococcus suis</i> WP_032498132 (129)	39/60 (28/43)
4	120	+(2630-2995)	Hypothetical protein WP_051234001 (341)	43/61 (18/26)
5	137	+(2967-3383)	Hypothetical protein <i>Globicatella</i> sp. HMPREF 2811-00515 (577)	29/47 (34/54)
6	123	+(3829-4200)	Hypothetical protein <i>Lactococcus lactis</i> pSK11B YP_003329500.1 (121)	34/56 (40/65)

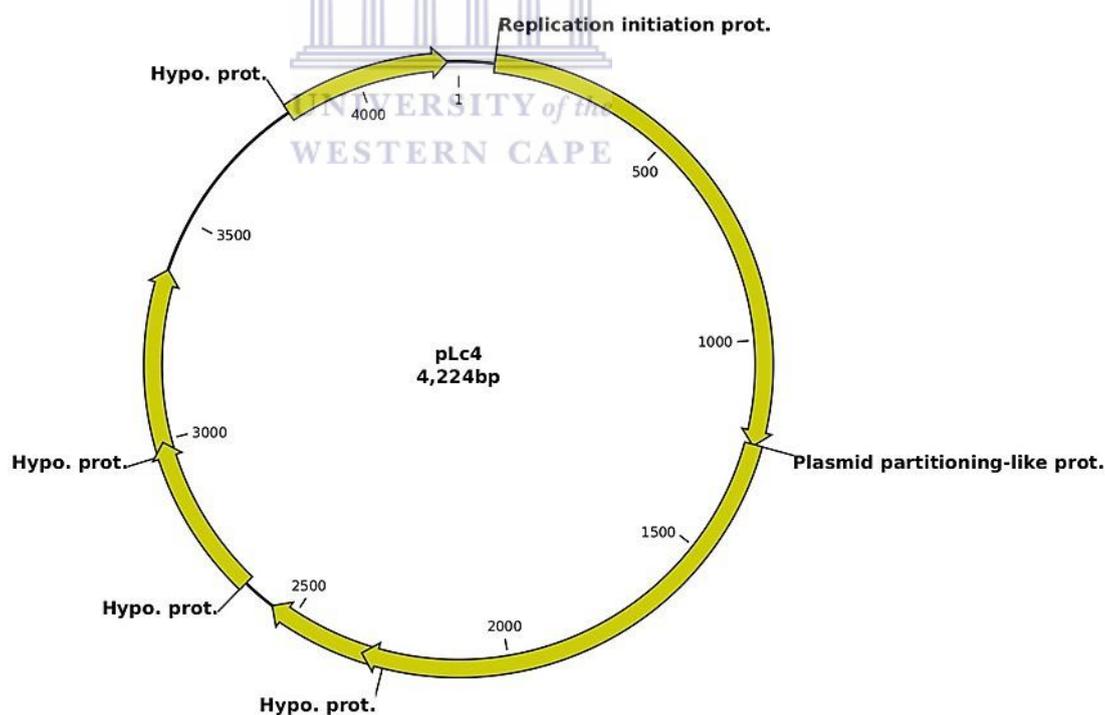


Figure 3.5: Annotated genetic map of pLc4. The assembly is based on 4224bp and was generated using CLC Genomics Workbench. ORFs and orientation of ORFs are indicated by the solid yellow arrows

3.2.5 Sequence analysis of pLc4

The top BLASTp hit to the replication protein identified on pLc4 (ORF 1) is similar to a protein encoded on pSK11B, a 13kb plasmid from *Lactococcus lactis*, with 43% amino acid identity. A comparison of 96 Rep proteins from *L. lactis* plasmid including those from pUB110, pA1 and pCW7 showed that the lactococcal plasmids belonged to 2 groups, with group 1 consisting of 88 out of the 99 Rep proteins. Group 1 is further divided into subgroups I-1 through to I-5. Plasmid members of subgroup I-1; pCI305, pWVO2, pW563, and pCD4 were experimentally shown to replicate *via* theta replication (Cui *et al.* 2015). pSK11B is thought to be theta replicating based on the similarity of one of the replication proteins encoded on it with the Rep proteins of pCI305 and pWV02, as well as the presence of three-and-one-half 22bp direct repeat (iterons) sequences directly upstream of the promoter for this replication protein (Papadimitriou *et al.* 2015). Both pCI305 and pWV02 fall under Class A of the theta replicon classes (Cui *et al.* 2015). No direct repeats of significance could be identified on pLc4. However, closer inspection of the pSK11B sequence, shows that it also encodes a Rep_Trans-like replication protein (YP_003329499.1) related to rolling circle replication (RCR) initiation proteins which is the ORF that the pLc4 protein shares homology with, and may indicate the presence of a second active replicon on pSK11B. The pLc4 Rep protein also shows 44% identity with the replication initiation protein of pSMQ173b_03, a 5kb RCR plasmid from *Streptococcus thermophilus*, classified as belonging to the pGI3 family of RCR plasmids (Turgeon & Moineau 2001; Turgeon *et al.* 2004). No HUH motif, characteristic of RCR plasmids from families pMV158, pC194 or phiX174 could be identified (del Solar *et al.* 1998). Two imperfect inverted repeats which could possibly form stem loop structures were identified on pLc4 (Table 3.4).

Table 3.4: Inverted repeat sequences for pLc4

Plasmid	Repeat location	Type of repeat	Repeat sequence
pLc4	260	Invert	ATAGTAGAAGCCTTTCTAGAGTATGCAATGA TATTATTTTCGGAAAGATTTTAGACGATGTT

Either of these could possibly serve as the double stranded origin (*dso*). Thus, pLc4 is likely to replicate *via* a rolling circle mechanism and based on the similarity with the pSMQ173b_03 RepB also belong to the pGI3 family of RCR plasmids. Efforts were also made to understand the phylogenetic relationship between the replication initiation protein of

pLc4 and the RepB proteins from pWV02, pCI305, pSMQ172, pSMQ173b 03, pG13, pER1-2, pMV158, pC194, pSTK1, pFX2, pLS1 (Figure 3.6).

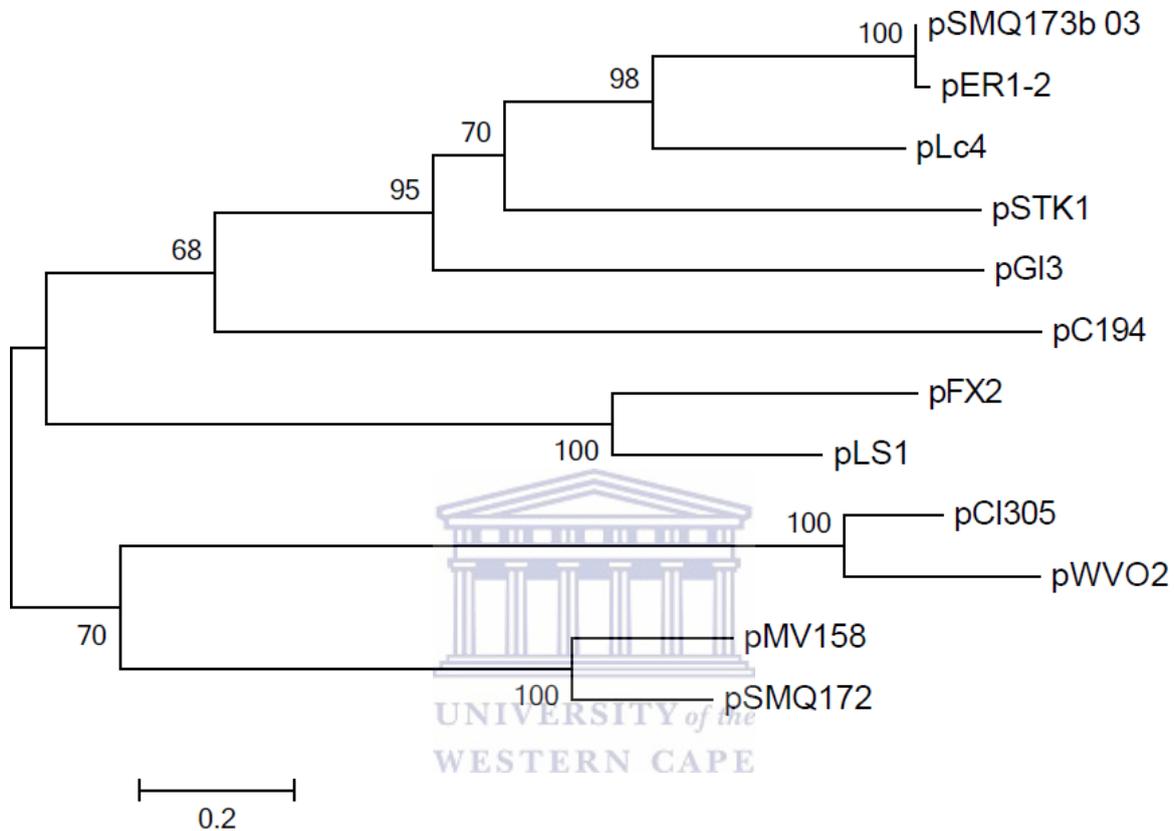


Figure 3.6: The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 7.73379931 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. There were a total of 146 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

The distances confirm the low identity, as well as distinct relationship of the replication initiation protein of pLc4 with the RepB proteins of pSMQ173b 03 and pSTK1. Based on the data, it can be concluded that the replication initiation protein is a novel Rep protein. The observation that pLc4 was present in both *L. crispatus* and *L. jensenii* could suggest that pLc4, while perhaps not having a particularly broad-host-range replicon (Boris & Barbés 2000) is not limited to replication in just these two hosts. Interestingly, many of the predicted proteins appear to be similar to a *Streptococcus* species and it is possible that pLc4 is originally derived from a *Streptococcus* plasmid. In addition to this many RCR plasmids from

Gram positive bacteria are known to have a broad host range (Cui *et al.* 2015). A comparison of closely related proteins between pLc4 and pSMQ173b as well as pSK11B shows high sequence homology between the 2 RCR plasmids (Figure 3.7).

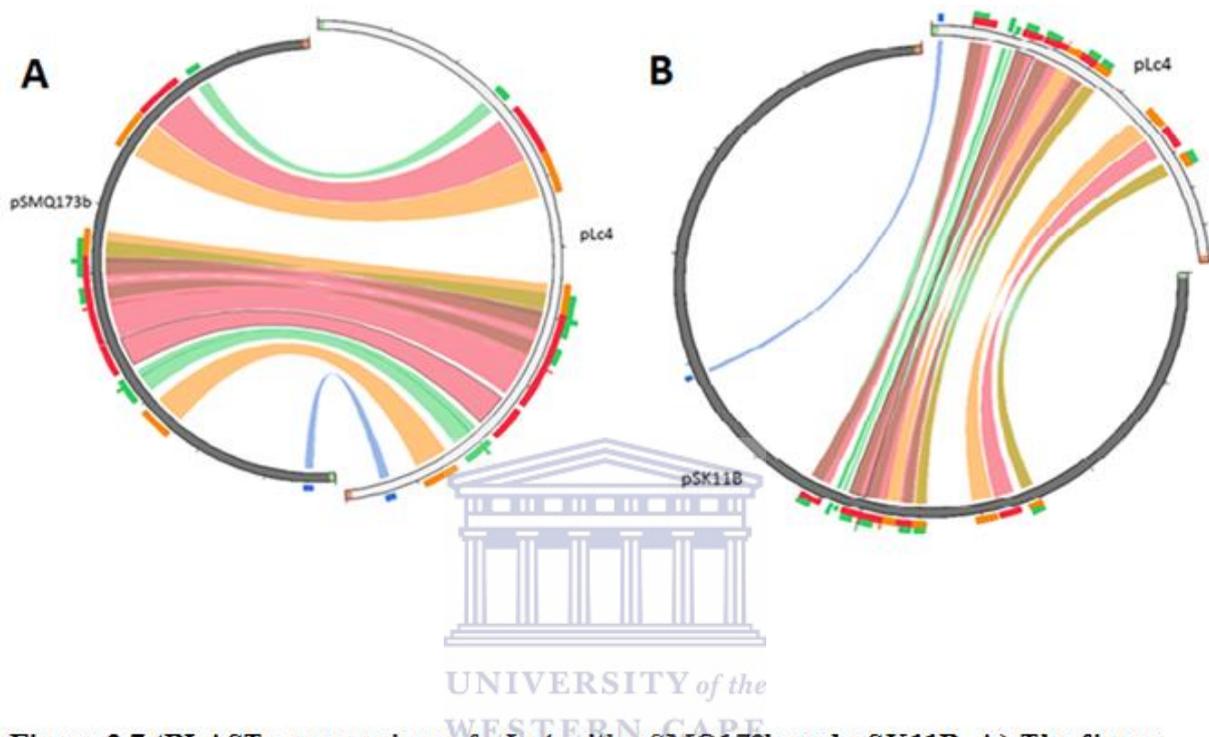


Figure 3.7 tBLASTx comparison of pLc4 with pSMQ173b and pSK11B. A) The figure shows a comparison of closely related proteins between pLc4 and pSMQ173b coloured by bitscore for BLAST hits. B) The figure shows a comparison of closely related proteins between pLc4 and pSK11B coloured by bitscore for BLAST hits. The coloured ribbons indicate the regions that are closely related on both plasmids.

RCR plasmids are known to be promiscuous plasmids and Del Solar *et al.* (1993) speculated that this could be due to the efficiency with which a given host recognises a plasmid's single stranded origin (*ssO*), the efficiency of the host's transcription/translation machinery to synthesize plasmid encoded gene products and host factors that can influence plasmid maintenance (del Solar *et al.* 1993). Further experimental work is required to better understand the possible broad-host-range nature of pLc4.

A plasmid partitioning protein (ORF 2) was identified on the pLc4, which may be involved in plasmid maintenance. These systems are often found in low copy number plasmids and suggest that this plasmid may be a low copy number plasmid (Baxter & Funnell 2014; Million-Weaver & Camps 2014)

3.2.6 Sequence analysis of pLc17

pLc17 was similar to various *Lactobacillus* species plasmids (high nucleotide identity) over portions of its genome (Table 3.5; Figure 3.8). 23 ORFs were identified based on NCBI BLASTx analysis.

Table 3.5: Putative genes and their products, deduced from the plasmid nucleotide sequences for pLc 17.

ORF Number	Size in amino acids	Start and end positions bp	Selected BLAST hits and comments; accession number; (length of protein on database in aa)	% Identity/Similarity (over number of aa)
1	107	+(10-334)	Hypothetical protein <i>Lactobacillus helveticus</i> WP_003625520 (80)	61/74 (33/40)
2	109	+(332-664)	Hypothetical protein <i>Lactobacillus crispatus</i> WP_005723210 (110)	100/100 (110/110)
3	207	+(678-1302)	Hypothetical protein HMPREF9250_02419, partial <i>Lactobacillus crispatus</i> FB049-03] EKB59528 (87)	100/100 (87/87)
4	343	+(1252-2286)	Transposase <i>Lactobacillus crispatus</i> WP_005728429 (344)	100/100 (344/344)
5	107	+(2204-2528)	Multidrug resistance protein MdtG, partial <i>Lactobacillus crispatus</i> SJ-3C-US] KFL92490 (165)	93/95 (38/39)
6	409	-(3795-2563)	Transposase <i>Lactobacillus crispatus</i> WP_035458603 (424)	100/100 (410/410)
7	243	-(4513-3779)	Resolvase Bacteria WP_005728427 (221)	100/100 (221/221)
8	237	-(5236-4520)	FIC-like protein <i>Lactobacillus crispatus</i> WP_005728426 (238)	100/100 (238/238)
9	123	-(5629-5255)	Hypothetical protein LBKG_02277 <i>Lactobacillus crispatus</i> CTV-05 WP_005728425 (58)	98/100 (57-58)
10	369	-(6747-5635)	Putative plasmid replication initiation protein <i>Lactobacillus crispatus</i> WP_005728424 (370)	100/100 (370/370)
11	193	+(7340-6757)	Replication initiator protein A, partial <i>Lactobacillus parafarraginis</i> WP_008213692 (203)	79/90 (26/33)
12	260	+(7326-8111)	ATPase <i>Lactobacillus</i> WP_005728422 (261)	100/100 (250/250)
13	173	+(8105-8629)	Hypothetical protein <i>Lactobacillus crispatus</i> WP_005728419 (71)	100/100 (71/71)
14	165	+(8622-9122)	Transposase <i>Lactobacillus helveticus</i> WP_005728418 (164) Hypothetical protein LBKG_02273 <i>Lactobacillus crispatus</i> CTV-05 EFQ43360.1 (166)	99/98 (143/145) 100/100(166/166)
15	138	+(9164-9583)	DDE endonuclease <i>Lactobacillus crispatus</i> WP_005729299 (139)	100/100 (139/139)
16	106	+(9497-9820)	Hypothetical protein HMPREF0548_0424 <i>Lactobacillus ultunensis</i> DSM 16047 EEJ72728 (145)	91/93 (30/31)
17	200	+(10061-10666)	Resolvase Bacteria WP_005728417 (201)	100/100 (201/201)

18	428	+(10668-11957)	Resolvase <i>Lactobacillus kefiranofaciens</i> WP_013855360.1 (197) Hypothetical protein <i>Lactobacillus crispatus</i> WP_005728416 (56)	54/70 (27/35) 98/100 (55/56)
19	342	+(11958-12986)	Ribonucleoside-diphosphatereductase, beta subunit <i>Lactobacillus crispatus</i> SJ-3C-US KFL92593 (342)	100/100 (342/342)
20	140	+(12967-13392)	Ribonucleotide reductase <i>Lactobacillus</i> WP_003648207 (141)	100/100 (141/141)
21	728	+(13364-15553)	Ribonucleoside-diphosphatereductase, alpha subunit <i>Lactobacillus crispatus</i> SJ-3C-US KFL92591 (729)	99/100 (728/729)
22	101	+(15554-15860)	Hypothetical protein <i>Lactobacillus crispatus</i> WP_005728412 (59)	98/100 (58/59)
23	273	+(15860-16684)	Replication initiation protein <i>Lactobacillus</i> <i>crispatus</i> WP_005728411 (262)	100/100 (262/262)

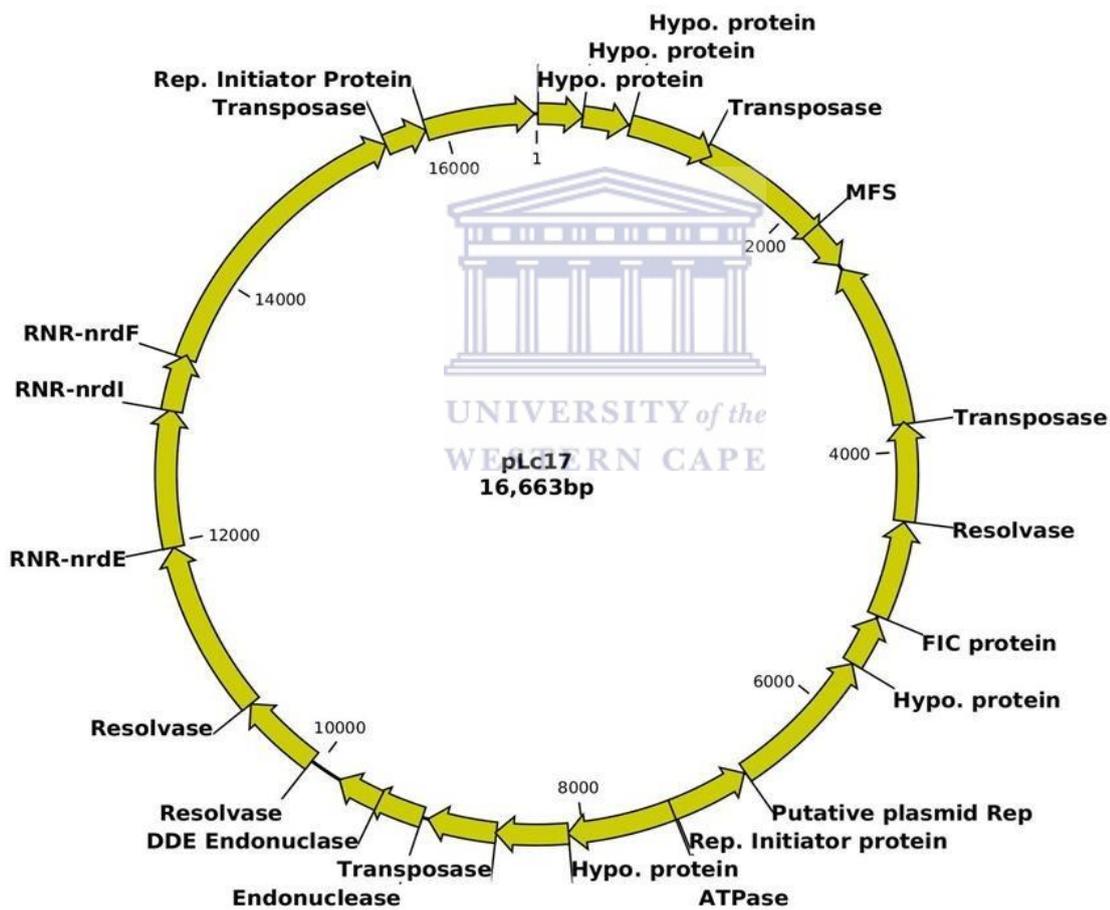


Figure 3.8: Annotated genetic map of pLc17. The assembly is based on 16663bp and was generated using CLC Genomics Workbench. ORFs and orientation of ORFs are indicated by the solid yellow arrows

Bioinformatic analysis for pLc17 resulted in the identification of the following accessory proteins; Transposase, Multi drug Resistance Protein, Resolvase, Filamentation induced by cAMP protein (FIC-like protein), endonuclease and three ribonucleotide reductases (RNR). pLc17 also has areas with high nucleotide identity to portions of *Lactobacillus* species genomes, with the largest being the region that encodes the RNR E, I and F encoded by ORF's 19, 20 and 21 respectively. RNRs are responsible for the conversion of ribonucleotides to deoxyribonucleotides, and play a pivotal role in cell survival by repairing damaged DNA through DNA synthesis (Stubbe 1990). This enzyme can be divided into several classes (Ia, Ib, Ic; II and III), of which Class I RNRs are oxygen dependent, Class IIs can function in the absence of oxygen, while Class III enzymes require anaerobic conditions to function (Dwivedi *et al.* 2013). Class Ib RNRs are encoded by the *nrdHIEF* suit of genes and rely on the glutaredoxin-like protein encoded by *nrdH* to generate the radical needed for catalysis (Torrents *et al.* 2014). All subunits, except *nrdH*, are encoded on pLc17, indicating that the plasmid encoded enzymes may use a host-encoded enzyme to fulfil this role. The presence of RNR on the plasmid is not unique and has been observed on other plasmids (Lundin *et al.* 2010; Siedow *et al.* 1999; Tannous *et al.* 2015). *Alcaligenes eutrophus* H16 harbours a mega plasmid, which is pivotal to its anaerobic growth in the presence of nitrate. The genes found to be responsible for this are *nrdD* and *nrdG*. Sequence alignment of the *A. eutrophus* H16 NrdD and NrdG proteins indicated that they were similar to anaerobic class III ribonucleotide reductases and their corresponding activases (Siedow *et al.*, 1999). The class I RNR was also found to enable the stable maintenance of mega plasmid pTT27 in *Thermus thermophilus* HB27, as a result of genome replication normalized by reversing the dNTP pool imbalance (Ohtani *et al.* 2016).

It is tempting to speculate that the RNRs on pLc17 provide the host with a metabolic advantage, allowing it to survive during BV infection. However, given the sequence identity to class Ib RNR's, it is speculated that the pLc17-encoded RNR's are likely only active under aerobic conditions and given that the vaginal milieu (whether in healthy or during BV infection) is likely microaerobic or anaerobic, it is not expected that their presence benefits the host under these conditions (Ghartey *et al.* 2014; Masalha *et al.* 2001). The plasmid is likely to represent more of a metabolic burden under these circumstances, rather than helping the host to survive and the plasmids' persistence is likely only due to the plasmid stability systems present.

Temperate phages are known to infect vaginal *Lactobacillus* sp, impacting these populations.

This has led some authors to speculate that the BV condition may be related to the phage infection. In a study conducted by Ojala *et al.*, 2014, 31 prophage like regions were identified on *L. crispatus* genomes. The majority of the *L. crispatus* genomes were from vaginal isolates, suggesting high phage exposure (Ojala *et al.* 2014). This was in agreement with Damelin *et al.* (2011) where they found 77% lysogeny in vaginal lactobacilli. RNR genes are abundant in many phages that infect both autotrophic and heterotrophic bacteria and are likely involved in phage replication (Dwivedi *et al.* 2013). A more possible explanation for the RNR on pLc17 may be a result of the *L. crispatus* host genome acquiring the RNR through phage infection and the plasmid acquiring this from the host as a result of a transposition event, as the RNRs are flanked by mobile elements such as transposons and endonucleases.

The FIC-like protein encoded on pLc17 shared 97% sequence homology with the protein on pHN1 from *L. salivarius*. This bacterium is found in human breast milk and has been identified as a viable probiotic (Jiménez *et al.* 2010; Malek *et al.* 2010). This strain carries 3 plasmids (pHN1, pHN2 and pHN3) (Jiménez *et al.* 2010). The FIC-like protein for pLc17 also shares 95% sequence homology with a *L. salivarius* UCC118 pSF118-44 protein, a plasmid that contains several toxin-antitoxin (TA) proteins associated with plasmid stability (Fang *et al.* 2008). The FIC-like protein (ORF 8) from pLc17 is suggested to be involved in death on curing proteins required for plasmid stability (Gerdes & Maisonneuve 2012; Stanger *et al.* 2016). These proteins are also known to retard/arrest the growth of bacteria via adenylation of the DNA gyrase B (GyrB), which plays a role in DNA replication (Lu *et al.* 2016; Stanger *et al.* 2016). Lu *et al.* (2016) found that FIC I solely targeted GyrB affecting bacterial or plasmid replication (Lu *et al.* 2016). This protein achieves this by adding the adenosine 5'-monophosphate (AMP) to a tyrosine residue (Tyr109), which is required for binding adenosine 5'-triphosphate (ATP). This blocking mechanism has been shown to inhibit GyrB activity (Lu *et al.* 2016). Thus, the FIC-like protein in pLc17 may be regulating the PCN in the host during times of environmental stress. This might ensure a low metabolic cost to the host and ensure plasmid maintenance in the host.

The TA systems commonly associated with these plasmids carry phd (prevent host death) antitoxin and doc (death on curing) systems (Castro-Roa *et al.* 2013; Cruz *et al.* 2014). This is a member of the FIC family and the TA toxins affect the cellular activity, switching the cell into a dormant state. In the dormant state the host cell is temporarily tolerant to antibiotics

(Castro-Roa *et al.* 2013). Thus the system encoded by pLc17 may allow the host cells an added advantage by increasing antibiotic resistance, while in the dormant state. The TA system phenomenon is common among drug resistant persisters and the mechanism of persistence has been hypothesised as being due to this change in metabolic activity (dormancy) (Cohen, Lobritz, & Collins, 2013). Furthermore, the role of the TA system in allowing the cell to enter a dormant resistant state has been suggested as a possible mechanism utilized by bacteria in unfavourable conditions. This persister cell formation has been seen in *E. coli* after ampicillin treatment. The toxin HipA of the *hipA/hipB* system was found to be essential in persister cell formation (Cohen *et al.* 2013; Unterholzner *et al.* 2013). Castro-Roa *et al.* (2013) and Cruz *et al.* (2014) suggested that this arrest in translation may be a result of a unique kinase activity whereby the threonine (Thr382) of the translating elongation factor (EF-TU) is phosphorylated reducing the binding capacity of EF-TU with the aminoacylated tRNA. This results in host cell death. The difference in enzymatic activity might be linked to the slight difference in the FIC motif, as the doc FIC motif is HXFX(D/N)(A/G) NKR compared to the conserved FIC protein motif HXFX(D/E)GNRXXXR. The TA toxin activity is then inhibited by the antitoxin reviving the cell. Castro-Roa *et al.* (2013) propose that this is due to the doc- catalysed dephosphorylation of EF-TU, resulting in an increase in nucleoside diphosphates (NDP's) during stasis, which in turn causes baseline levels of non-phosphorylated EF-TU that may allow initial translation.

Interestingly TA systems have been associated with phage exclusion. This has been seen in TA systems like *hok/sok* of plasmid R1 which excludes T4 phage in *E. coli* and the *mazE/mazF*, which induce abortive infection (Abi) during P1 bacteriophage infection. Another well-known TA system involved in Abi is the *toxI/toxN* from *Erwinia carotovora* subspecies *atroseptica*. The toxin ToxN is released and kills the phage-infected cells. As mentioned above *L. crispatus* genomes from vaginal isolates have a high phage exposure, which may be a possible reason for the TA system in pLc17 (Fineran *et al.* 2009; Unterholzner *et al.* 2013). Since the FIC-like protein encoded by pLc17 is flanked by mobile genetic elements it's not impossible that the region is host chromosome encoded. The FIC-like protein shows 100% sequence homology to that on the genome of *L. crispatus* (Table 3.5), which would suggest that the gene may have moved from the host to the plasmid in a transposition event. Another possibility is that the host obtained the FIC-like protein from another similar plasmid or episome. Episomes are known to replicate independently but can integrate into the host genome and are often beneficial as they improve host survival (Falkow

& Baron 1962).

The ATPase (ORF 12) from pLc17 is possibly involved in plasmid partitioning. These proteins play a critical role in proper cell division among bacterial genomes and ensure stable integration of low copy number plasmids through active partitioning systems. These systems also ensure that the plasmids are directed to their required intracellular sites (Fung *et al.* 2001). Par-A, a membrane associated ATPase, is involved in the correct positioning of the partitioning complex (Bignell & Thomas 2001; Erdmann *et al.* 1999). These proteins contain a specific Walker A ATP binding motif, as well as a Walker B motif. ParA in plasmid partitioning is involved in segregation reaction and acts as a transcriptional repressor. ATP and ADP binding has been shown *in vitro* to affect all ParA functions as well as the protein confirmation. The adenine nucleotide is strongly linked to the binding capacity of ParA to the inverted repeat sequence in the *par* promoter, ParOP. The ADP bound form (free form) of ParA is known to repress partitioning by binding to DNA rather than the partitioning complex but the ATP bound form is crucial to the partitioning (Bignell & Thomas 2001; Fung *et al.* 2001). ParB is capable of stimulating ParA ATPase activity and the binding to the ParOP. This however, is only when ATP is present. A study done by Fung *et al.* (2001) found that the conserved lysine and aspartate residues in Walker A, Walker B and motif 2 regions are required for ATPase activity (Fung *et al.* 2001). Furthermore, one of the replication proteins identified on pLc17 RepB-ORF 11 has an ATPase domain and shows similarity to ParA-like plasmid partitioning protein sequences, thus may play a role in the plasmid's stable inheritance. The ORFs are divergently transcribed and several repeats were identified in the intergenic space, as well as within the RepA ORF and downstream of RepB (Table 3.6), which is likely to form part of the replication origin. Two other replication proteins (RepA – ORF 10 and RepC –ORF 23) were identified on pLc17. pLc17 appears to have two potential replicon types, although we could identify many RepA-related protein sequences on sequenced *Lactobacillus* genomes, the plasmid replication protein that it is most similar to is from the recently sequenced pR2 from *Lactobacillus salivarius* (99.5% identity at the amino acid level) (Sun *et al.* 2015). The ORF on pR2 encoding the replication initiation protein in the GenBank entry contains several frame-shifts. The next closest homologue of RepA is the replication protein from *L. salivarius* pSF118-44 at 86% identity. pSF118-44 (44kb) is predicted to replicate via a theta replicating mechanism (Fang *et al.* 2008), and should pLc17 have a similar mechanism of replication as indicated by the similarity to this replication protein, it too is likely to be theta replicating. RepC appears to be related to the replication

initiation protein from pRS2 at 36% identity.

Table 3.6: Direct and inverted repeats for pLc17

Repeat location	Type of repeat	Repeat sequence
11751 / 11805 11703 / 11769 / 11823	Direct Direct	TCAACGAAGTCATGCGGACTTTCGATTTTCGAAATTCGGCT CGTAG CTTTCGATTTTCGAAATTCGGCTCGTAG
12486/ 12510	Invert	AATGGACTAAGTAG
12776 / 12813	Invert	ATGATTTATGATTAATC
12774 / 12789	Invert	ATATGATTTATGATT
12785 / 12813	Direct	GATTAATCATAAATCAT
13894 / 13921	Invert	TGATTGTGTGATTTATGATAAAAAGATT
4466-4497	Invert	AACCCCGAAATTCGGACAAGAATTTCCGGGGTT

pRS2 is a 2.5kb rolling circle replicating plasmid isolated from *Oenococcus oenii* (Mesas *et al.* 2004) pLc17 thus appears to have two potential replicon types, similar to what was found on *Bacillus subtilis* pIP501, a broad host range plasmid which carries both rolling circle and theta replicons (Pujol *et al.* 1998). In pIP501 the RepR protein plays a role in the inhibition of RCR allowing theta replication to take place (Pujol *et al.* 1998). The sequence alignment of the RepR sequence of the pIP501 replicon with the RepA and RepB sequence of pLc17 replicons revealed that they shared a 33.46% and 32.03% homology, respectively. This, together with the size of pLc17 of 16663bp, suggests that the plasmid may replicate via theta replication. There also appears to be an upper size limit for RCR plasmids (<12Kb) adding further evidence for it replicating via a theta mechanism (Khan 1997; Xi *et al.* 2013). A comparison of closely related proteins between pLc17 and pHN1 showed low homology with parts of pLc17 having similarity to pHN1. When pLc17 was compared with pR2 it also showed low levels of sequence homology (Figure 3.9).

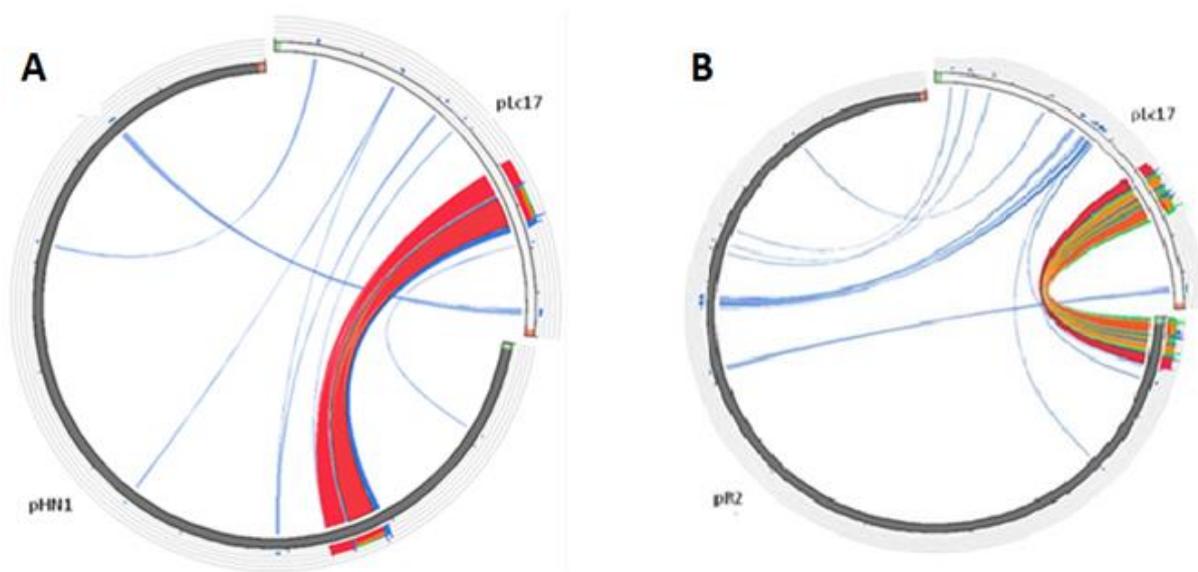


Figure 3.9 tBlastx comparison of pLc17 with pHN1 and pR2. A) The figure shows a comparison of closely related proteins between pLc17 and pHN1 coloured by bitscore for BLAST hits. B) The figure shows a comparison of closely related proteins between pLc17 and pR2 coloured by bitscore for BLAST hits. The coloured ribbons indicate the region that are closely related on both plasmids.

UNIVERSITY of the

There are a large number of mobile DNA elements on pLc17 (Table 3.3), which appears to divide the plasmid into the theta-like replication region and the rolling circle-like replication region together with the genes coding for the RNR and multidrug resistant protein. These elements could be responsible for the presence of the RNR on the plasmid through a transposition event. It is likely that they are responsible for the formation of the co-integrate plasmid either through recombination or a transposition event, and they may also play a role in the fitness of the host by providing a source of genome rearrangement and/or switching on/off of genes necessary for its adaptation to new environments (Bennett 2004).

The multidrug resistance protein (ORF 5) from pLc17 may be involved in antibiotic resistance and may allow the strains to survive antibiotic treatment, and could ensure that these beneficial probiotic strains dominates the vaginal milieu. The multidrug resistant gene is flanked by mobile elements suggesting that these resistant genes could potentially be transferred to surrounding pathogens (Gueimonde *et al.* 2013).

The resolvase proteins (ORF 7, 17 and 18), apart from potentially being part of a composite transposon, is likely to play a role in plasmid replication and allow segregational stability of the plasmid. These genes can be viewed as ‘selfish genes’ as their main purpose is to ensure that the plasmid is maintained in the population (Ito *et al.* 2009).

3.3 Conclusion

Sequence analysis revealed that pLc17 may encode for proteins that may improve the fitness and survival of *L. crispatus* in the vaginal environment, whereas pLc4 appears to be a selfish DNA element. The multi drug resistance protein may ensure resistance to antibiotics or other toxic compounds. The inclusion of the RNR into the plasmid may be a result of the numerous transposition events facilitated by the large number of mobile DNA elements on pLc17. This protein is likely inactive and therefore may not significantly contribute to the host’s ability to survive/thrive in its environment. The possibility does however, exist that despite the classification of these RNR’s into class Ib, these ORFs may have evolved to operate under low oxygen tension. This has yet to be investigated experimentally. These proteins may also indicate the high exposure of *L. crispatus* to phage in this environment. The FIC-like protein, interestingly, may confer antibiotic resistance by placing the host in a dormant state. In addition to this, in light of the high phage exposure experienced by the host, it may likely be involved in phage exclusion, an additional factor that could contribute to the BV state. pLc4 was shown to be a novel plasmid with a low number of proteins identified and is thus considered a cryptic plasmid. Cryptic plasmids tend to persist in bacterial populations despite imposing metabolic cost without any obvious benefit to the host. These plasmids are known to carry genes involved in core plasmid functions, required for plasmid survival known as ‘selfish DNA’ (Challacombe *et al.* 2017). The lack of proteins identified for pLc4 suggests that there are gaps in the databases that need to be filled by future experimental work, and points to novel mobile elements in the South African population.

CHAPTER 4 Plasmid Copy Number Determination

4.1 Introduction

PCR is a standard molecular technique that can amplify a small amount of template DNA or complementary DNA (cDNA), into large quantities in a few hours. This technique was invented by Kary Mullis and co-workers almost three decades ago (Mullis 1990), and has allowed scientists to manipulate DNA for cloning, genetic engineering and sequencing (Kubista *et al.* 2006). The problem associated with conventional PCR is the lack of reliable quantification and reproducibility (Valasek & Repa 2005). These problems were resolved two decades ago with the invention of real-time quantitative PCR (qPCR) by Higuchi *et al.* (1992) and Heid *et al.* (1996) (Navarro *et al.* 2015; Pabinger *et al.* 2014; Valasek & Repa 2005). Since its inception qPCR has become an invaluable tool in the life sciences, and is used in a variety of applications including basic research, pathogen detection and biomedical diagnostics (Navarro *et al.* 2015; Pabinger *et al.* 2014). An advantage of this technology is that it is high throughput, has increased sensitivity, accuracy and versatility (Pabinger *et al.* 2014)

qPCR is widely regarded as the gold standard for the analysis of DNA copy number, as it allows for the amplification and simultaneous quantification of target DNA. Unlike conventional PCR, qPCR allows for direct measurement of DNA amplification as it proceeds in real time and as the reaction is in a closed tube system post amplification manipulation is eliminated, thereby reducing possible contamination (Lin *et al.* 2011; Navarro *et al.* 2015; Pabinger *et al.* 2014). qPCR measures the amount of target DNA based on the emission of a fluorescent reporter dye. During the initial cycles the signal is weak and cannot be distinguished from the background. As the amount of PCR product increases, the fluorescent signal increases exponentially above the detectable threshold level. As the critical components (i.e. primers, polymerase and the dNTP's) become depleted the signal level reaches saturation resulting in the plateau phase. The slope of a typical qPCR reaction reflects the difference in the initial amounts of template molecules. The more target template present at the beginning of the reaction the fewer number of cycles it takes to reach the detection threshold. The threshold of a real time PCR reaction is the amount of signal that reflects a statistically significant increase over the calculated baseline signal. It is set to distinguish relevant fluorescent signal from the background (Kubista *et al.* 2006; Valasek & Repa, 2005). The quantification cycle (C_q) is the cycle number at which the fluorescence signal of the

reaction crosses the threshold (Laplante 2006).

The C_q can be determined either through absolute quantification or relative quantification. Absolute quantification determines the absolute/exact number of target DNA copies within an unknown sample by using a standard curve that is prepared from a dilution series of a control template of known concentration. Relative quantification determines the fold changes in expression between two samples and requires a reference gene. The target gene is normalized to the reference gene in the same sample, and the normalized values are compared between samples to obtain a fold change (Kubista *et al.* 2006; Navarro *et al.* 2015; Pabinger *et al.* 2014).

DNA standard curves can be determined using plasmid DNA, PCR amplicons, synthesized oligonucleotides, genomic DNA or cDNA, with plasmid DNA being utilized most frequently. However, despite this, studies have shown that changes in plasmid formation from supercoiled to nicked circular, closed-circular, or linear forms can influence qPCR amplification greatly. Conformational changes can be brought about by routine laboratory practices such as freeze-thawing and shearing by pipetting, therefore special care must be taken when working with plasmid DNA (Lin *et al.* 2011).

The primer efficiency is vital for qPCR as this value indicates the performance of the primer set in the DNA amplification. An efficiency of 2.00 indicates 100% amplification efficiency suggesting every template molecule results in two products at the end of a cycle (Anindyajati *et al.* 2016). Many factors may play a role in qPCR efficiency; these include primer/probe design, PCR inhibitors, inaccurate pipetting and plasmid conformation (Hou *et al.* 2010; Meijerink *et al.* 2001). There are two different quantification chemistries that can be employed for qPCR (Navarro *et al.* 2015; Pabinger *et al.* 2014) namely double stranded DNA intercalating dyes, such as SYBR Green 1, SYBR Gold and EvaGreen, or fluorophore-labeled oligonucleotides, which can be sub-divided into three categories: i) Primer-probes such as Scorpions, Amplifluor® and LUX™; ii) Hybridisation probes such as Hybprobe or FRET, Molecular Beacons, HyBeacon™, MGB-Pleiades; and hydrolysis probes such as TaqMan, MGB-TaqMan, Snake assay; and iii) Analogues of nucleic acids such as PNA, LNA®, ZNA™, non-natural bases: Plexor™ primer.

Double stranded DNA intercalating dyes are cost effective and easy to set up but due to the non-specific binding of the dye to any dsDNA target the accuracy of this quantification system is greatly reduced compared to other forms of qPCR as primer dimers and non-specific DNA products may also be detected. Thus melt curve analysis is critical in the

quantification of the desired DNA product. The melt curve is generated at the end of the amplification cycle by the melting of the dsDNA amplicon. The increased temperature causes the dsDNA to denature and the dye dissociates from the DNA (Kubista *et al.* 2006; Valasek & Repa 2005). Probes and primer probes tend to be more accurate as they only detect the target DNA molecules but are generally more expensive (Navarro *et al.* 2015; Pabinger *et al.* 2014).

Gerhardt *et al.* (1994) defines plasmid copy number (PCN) as the number of plasmid copies per chromosome in a bacterium (Lee *et al.* 2006; Skulj *et al.* 2008). PCN determination can be done either directly or indirectly. For indirect PCN assays, the enzyme activity of the reporter proteins of the plasmid is measured (Lee *et al.* 2006; Skulj *et al.* 2008). Direct PCN assays refer to the direct quantification of the target DNA through methods such as CsCl centrifugation, Southern blot hybridization, high performance liquid chromatography (HPLC), capillary electrophoresis, agarose gel electrophoresis and qPCR. Of the methods mentioned for PCN determination qPCR has been shown to be the most cost effective, fast, accurate and the safest method. In several studies where this method was employed for PCN determination it was found to be reliable and reproducible, particularly in PCN determination in fermentation processes. Lee *et al.* (2006) used the method and found a high sensitivity in the quantification of the plasmid NS3, transformed in *E. coli* DH5 α . The PCN of the plasmid pBR322 in DH5 α was found to be within the range of previously published data on PCN of pBR322 plasmid DNA. In addition, they found that a wide range of DNA concentrations (0.5 pg to 50 ng) could be used (Lee *et al.* 2006). Skulj and co-workers (2008) found the qPCR to be accurate, fast and reproducible upon optimization. In their study they avoided DNA isolation and instead used the whole cell as the template source (Skulj *et al.* 2008).

In the present study, the PCN fold change for pLc4 and pLc17 was analysed for the healthy state relative to the BV infected state. This was to assess the importance of these plasmids in host survival within the vaginal milieu.

4.2 RESULTS AND DISCUSSION

4.2.1 pLc4 and pLc17 plasmid copy number (PCN) determination

L. crispatus and *L. jensenii* have been found to be the dominant Lactobacilli in the healthy vaginal microbiome (Damelin *et al.* 2011). The exact reason for their dominance still remains unknown. Plasmids are known to enhance bacterial survival under adverse conditions. In this chapter the abundance of both plasmids in healthy and BV infected patients was investigated.

4.2.2 Swab sample DNA extraction

The Zyppy gDNA extraction kit with a modified lysis step was used (Section 2.9). However, low DNA yields were obtained. Due to the success obtained on plasmid extractions of *L. crispatus* and *L. jensenii* isolates (Section 2.3), the vaginal swab samples subjected to the modified lysis step were used for the qPCR experiments.

4.2.3 Quantitative real time PCR (qPCR)

4.2.3.1 Selection of qPCR primers

Due to the nature of vaginal swabs large volumes of eukaryotic DNA is often obtained (The Human Microbiome Project Consortium 2012). Although rRNA, β -actin and GAPDH have been historically used as reference genes for many qPCR studies on human tissue and samples, numerous studies where these genes were used as reference genes have found highly variable expression levels and therefore are often not reliable reference genes (Gorzelnik *et al.* 2001; Radonic *et al.* 2004). An initial qPCR run on the vaginal swab samples using 18S rRNA primers resulted in an over estimation of the reference gene (results not shown). Epidermal growth factor receptor (EGFR), which is constitutively expressed in normal cells (Voldborg *et al.* 1997) was therefore used instead. Furthermore, the primer set was chosen as it has been used in qPCR experiments where the source material was human DNA. Real time PCR conditions were optimised for the plasmid primers targeting the two plasmids pLc4 and pLc17, as well as the reference EGFR gene, to ensure that an optimal melt curve, relatively low Cq scores and single amplicon was obtained. The primers were checked on BLAST (Basic Local Alignment Search Tool) to confirm that only the target regions would be amplified. Due to the nature of vaginal swabs large volumes of eukaryotic DNA is often obtained (The Human Microbiome Project Consortium 2012). Furthermore, the qPCR primers were tested using conventional PCR (Figure 4.1).

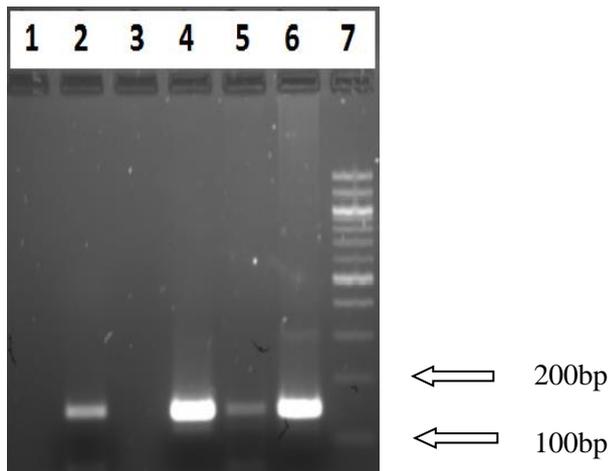


Figure 4.1: Agarose gel electrophoresis on 1% gel of PCR DNA amplification using qPL4F/R and qPL17F/R primers (1) Negative Control. (2) Sample L7. (4) Sample L5. (5) Sample L8. (7) Positive control (plasmid DNA). (8) Molecular marker: 100bp (Thermo Scientific).

4.2.3.1 qPCR standards on the Rotorgene

A dilution series of 10^{-n} was prepared for pLc4, pLc17 and EGFR gene amplicons (Chapter2) and the plasmid copy number (PCN) of each gene was analysed (Figure 4.2). Due to the low initial human DNA concentration obtained from conventional PCR, a smaller dilution range was generated for the EGFR amplicon. To ensure that the qPCR data are accurate and reliable it is recommended that the following parameters be followed: That the efficiency for each qPCR run is between 1.8-2.0; and the slope is between -3.6 and -3.3 (Pfaffi 2001). The efficiency of pLc4, pLc17 and EGFR primers was determined to be 1.92, 1.96 and 1.97, respectively. Additionally, the slope for pLc4, pLc17 and EGFR was -3.537, -3.4303 and -3.4073, respectively (Figure 4.2). In addition, the Y intercept also needs to be considered when evaluating the qPCR performance. A Y intercept above 37 illustrates that the DNA template has deteriorated (Laplante 2006). The Y intercept for pLc4, pLc17 and EGFR were 31.914, 31.385 and 31.108, respectively, indicating that there had been no template DNA deterioration during storage and processing. The state of the samples used in the qPCR experiment is another important factor. Storage buffer has been shown to influence qPCR. A 10 mM Tris/0.01% Tween-20 buffer is often recommended for the storage of Illumina sequencing libraries, as the buffer combination eliminates the sticking of DNA to the sides of the storage tubes. This buffer was utilized in this study to ensure accurate DNA concentrations, particularly at the lower concentrations, helping improve the efficiency further.

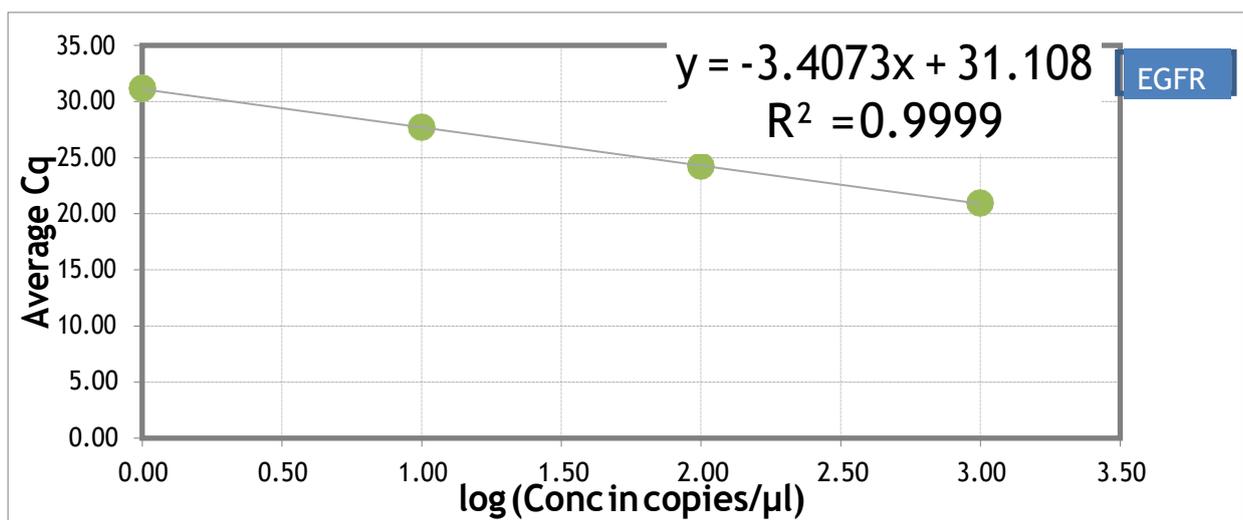
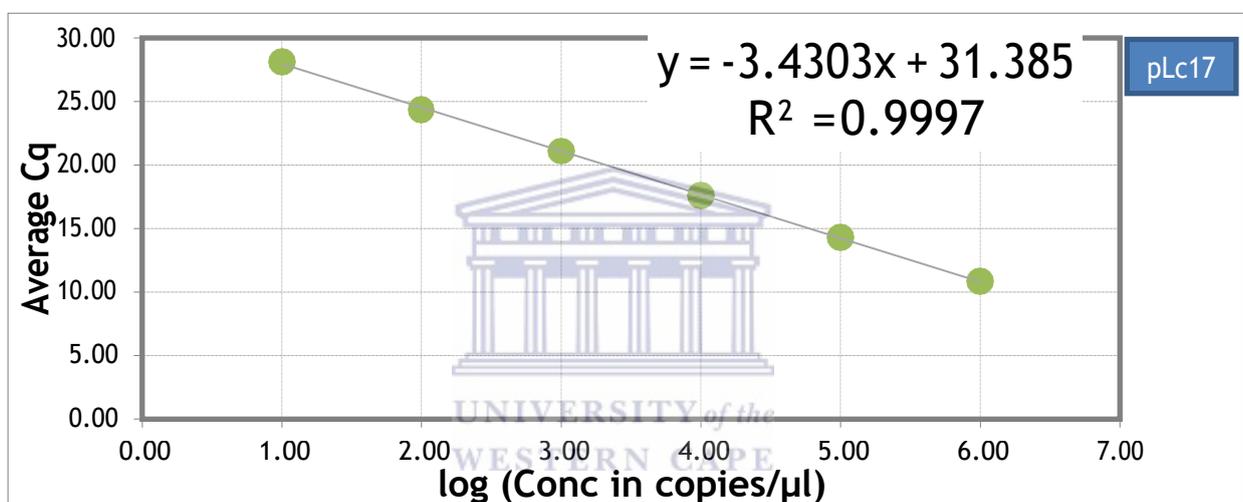
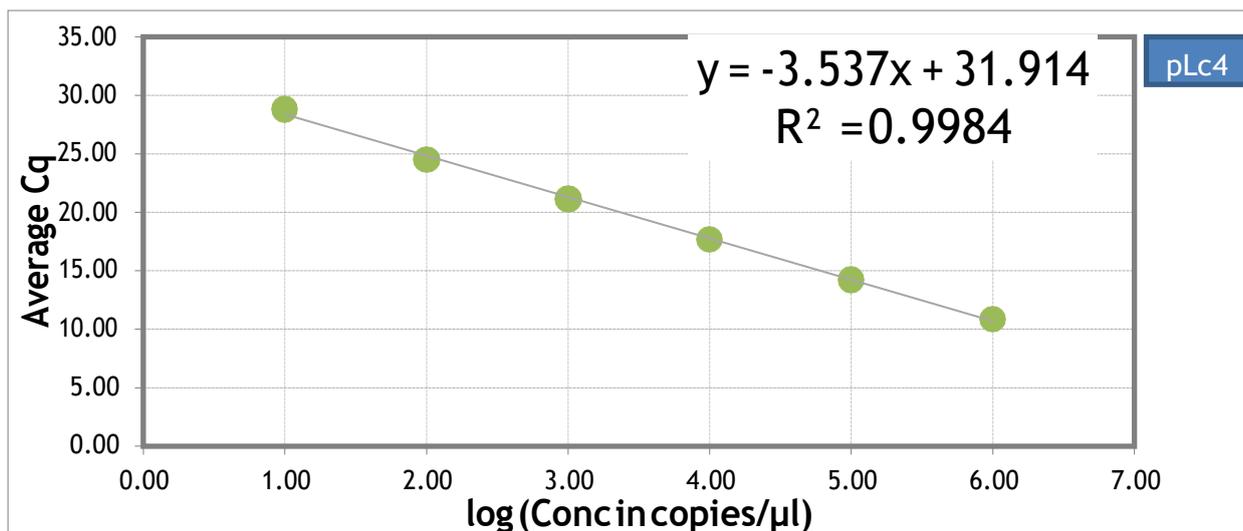


Figure 4.2: qPCR standard curves for the pLc4, pLc17 (n=6) and EGFR (n=4) primer sets. The correlation value (R^2) for the curves is indicated next to the trend line. The correlation value was determined using the Rotorgene software.

4.2.3.2 Primer specificity test

To test the specificity of the pLc4, pLc17 and EGFR primers, qPCR reactions were performed which were spiked with human DNA (placental DNA), plasmid DNA (pLc4 and pLc17) or bacterial genomic DNA (*E. coli*) at a range of concentrations (Table 4.1). This was done to test if the primers were binding to specific regions and not amplifying non-specifically.

Table 4.1 DNA concentrations used for spiking experiment

	Spiking combination				
DNA Type	Human	Human+ <i>E. coli</i>	Human+pLc4	Human+pLc17	All
Human DNA	0.05ng/ul	0.05ng/ul	0.05ng/ul	0.05ng/ul	0.05ng/ul
<i>E. coli</i> DNA		50 ng/ul			50 ng/ul
pLc4			500 copies/ul		500 copies/ul
pLc17				500 copies/ul	500 copies/ul

The primers designed for pLc4 and pLc17 were specific to the respective plasmids, as the presence and quantification of each of the plasmids was only detected with the respective primer set. This suggested that any co-extracted eukaryotic DNA would not be amplified during the qPCR run (Figure 4.3). The melt curve analysis (not presented) for pLc4 and pLc17, suggest that the desired product was obtained for samples subjected to qPCR using the qPL4, qPL17, 16S rRNA and EGFR primers.

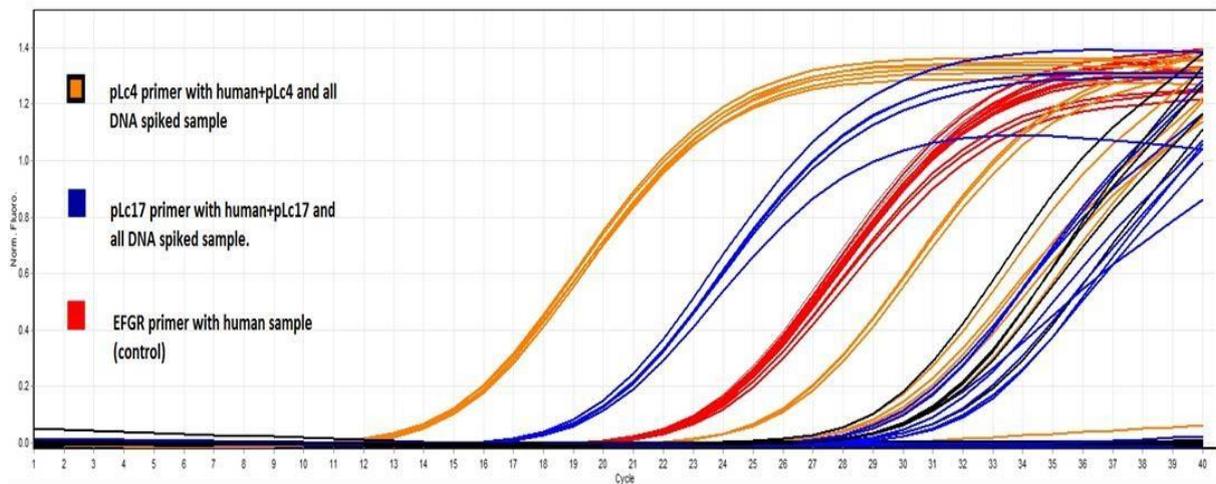


Figure 4.3: Amplification curves generated to assess the primer specificity of pLc4 and pLc17. ■ pLc4 primer with human+pLc4 and all DNA spiked sample. ■ pLc17 primer with human+pLc17 and all DNA spiked sample. ■ EGFR primer with human sample (control).

Furthermore, the specificity of the primers in detecting pLc4 and pLc17 and not any other bacterial DNA was assessed using human-DNA spiked with *E. coli* DNA. No non-specific amplification of the *E. coli* DNA was detected.

4.2.3.3 Assessment of pLc4 and pLc17 in Swab Samples using qPCR

The relative PCN for pLc4 and pLc17 extracted from swab samples (Section 2.9.2) was measured using real time PCR performed with a Rotorgene. In this study 28 swab samples obtained from NICD (previously used in the Damelin *et al.* 2011 publication) were assessed. These samples were collected from 16 healthy and 12 BV infected individuals (Table 4.2).

The DNA concentrations in the samples were standardized and equal volumes were subjected to qPCR amplification using the three primer sets (Table 2.2). The EGFR amplicon served to normalize concentration variations between the samples.

Table 4.2 Plasmid samples used in this study to assess the presence of pLc4 and pLc17.

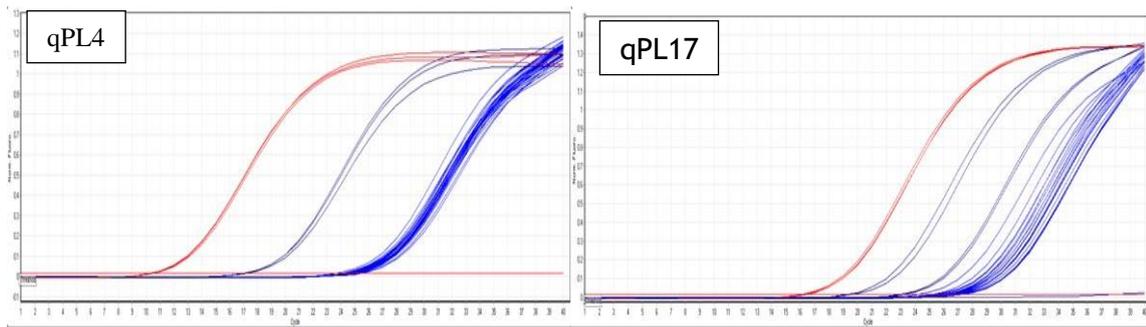
Healthy		BV infected	
Sample Name	Sample Label	Sample Name	Sample Label
CDC16-071-2015	SH2	GP-09-VDS-123	SD4
GP-09-VDS-106	SH3	GP-09-VDS-125	SD6
GP-09-VDS-126	SH5	GP-09-VDS-109	SD13
GP-09-VDS-131	SH10	GP-09-VDS-141	SD14
GP-09-VDS-099	SH11	CDC16-071-2021	SD15
GP-09-VDS-102	SH12	CDC16-071-2002	SD18
CDC16-071-2020	SH16	GP-09-VDS-132	SD20
CDC16-071-2018	SH17	CDC16-071-1487	SD23
GP-09-VDS-104	SH19	GP-09-VDS-138	SD25
CDC16-071-1500	SH21	GP-09-VDS-068	SD31
GP-09-VDS-136	SH22	GP-09-VDS-103	SD39
CDC16-071-1482	SH24	GP-09-VDS-078	SD41
GP-09-VDS-128	SH26		
CDC16-071-2014	SH30		
GP-09-VDS-097	SH40		
CDC16-071-1488	SH42		

 SH=Healthy patient sample

 SD= BV-infected patient sample

For all qPCR runs positive controls were included. For the pLc4 and pLc17 primer sets amplicons generated from plasmid DNA were diluted to 10000 copies/ μ l and were included in all runs. Similarly, amplicons for the reference gene underwent ten-fold serial dilutions and was included in all qPCR runs. This was done to ensure that during qPCR the primers amplified the target DNA. In addition, no template DNA controls were included for each qPCR reaction to monitor potential contamination and false positive results. Figure 4.4 illustrate the amplification curves for samples subjected to qPCR using the qPL4 and qPL17 primers.

A.



B.

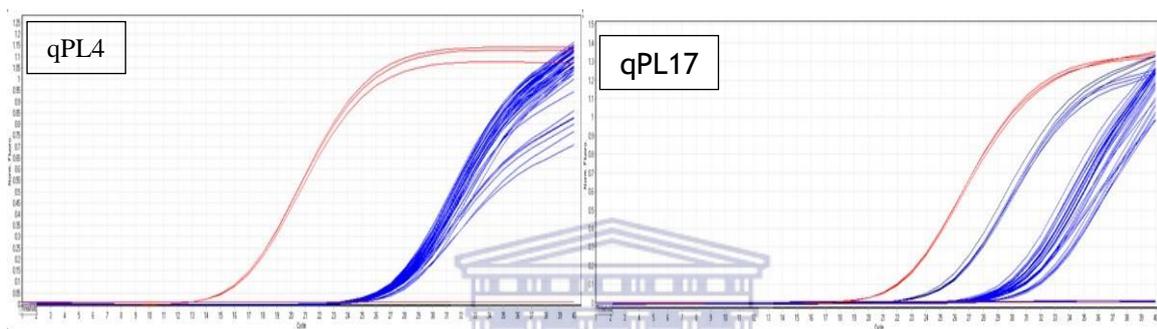
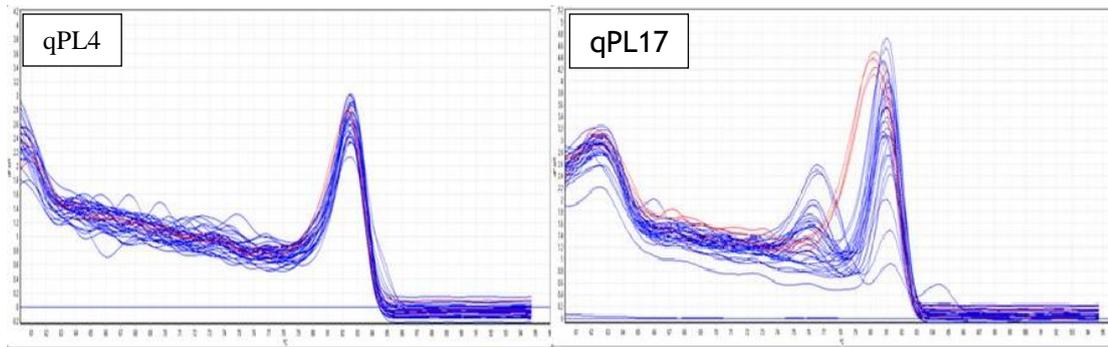


Figure 4.4: Amplification curve of clinical samples. A) Amplification of samples (SH2, SH3, SD4, SH5, SD6, SH10, SH11, SH12, SD13, SD14, SD15 and SH16) indicated by the blue curves were subjected to qPCR using qPL4 primers (left) and qPL17 primers (right). Both plasmids were serially diluted to 10000 copies/ μ l was used as the positive control, indicated by the red curve. The C_q values >31 cycles indicate levels outside the dynamic range of quantifiable PCN. B) Amplification curve of samples (SH17, SD18, SH19, SD20, SH21, SH22, SD23, SH24, SD25, SH26, SD31, SD39, SH40, SD41, SH42) indicated by the blue curves were subjected to qPCR using qPL4 primers (left) and qPL17 primers (right). Both plasmids were serially diluted down to 10000 copies/ μ l were used as the positive control, indicated by the red curve. The C_q values >31 cycles indicate levels outside the dynamic range of quantifiable PCN.

To assess whether the target DNA was amplified, melt curve analysis was generated for both amplification curves as seen in Figure 4.5.

A.



B.

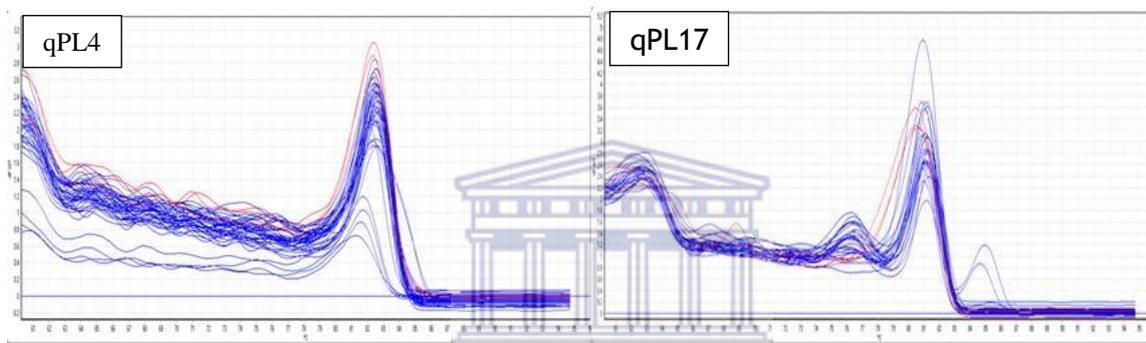


Figure 4.5 Melt curve analysis of clinical samples A) The melt curve of samples (SH2, SH3, SD4, SH5, SD6, SH10, SH11, SH12, SD13, SD14, SD15 and SH16) indicated by the blue curves were subjected to qPCR using qPL4 primers (left) and qPL17 primers (right). Both plasmids were serially diluted to 10000 copies/ μ l was used as the positive control, indicated by the red curve B) Melt curve of samples (SH17, SD18, SH19, SD20, SH21, SH22, SD23, SH24, SD25, SH26, SD31, SD39, SH40, SD41, SH42) indicated by the blue curves were subjected to qPCR using qPL4 primers (left) and qPL17 primers (right). Both plasmids were serially diluted to 10000 copies/ μ l was used as the positive control, indicated by the red curve.

The melt curves in Figure 4.5 suggest that the desired product was obtained for 26 out of the 28 samples subjected to qPCR using the qPL4 primers. For the qPL17 primers 25 of the 28 swab samples formed the desired product. This was illustrated by the overlapping peaks from the swab samples and the positive control (standards) indicating dissociation of the dsDNA products at a similar temperature. The swab samples SD4, SH5, SH10, SD13 and SD18 formed multiple peaks suggesting that either primer dimers had formed or, more likely, another product was amplified (nonspecific amplification) for these swab samples. Melt curve analysis assumes DNA melting is a 2-phase process (dsDNA and ssDNA) and does not

take into account the possibility of an intermediate phase brought about by secondary structures or stable GC rich regions. It is often advisable to perform agarose gel analysis on the qPCR products. The qPCR products from the swab samples were thus run on an agarose gel to confirm that the desired product was formed and that a single product was formed (Figure 4.6).

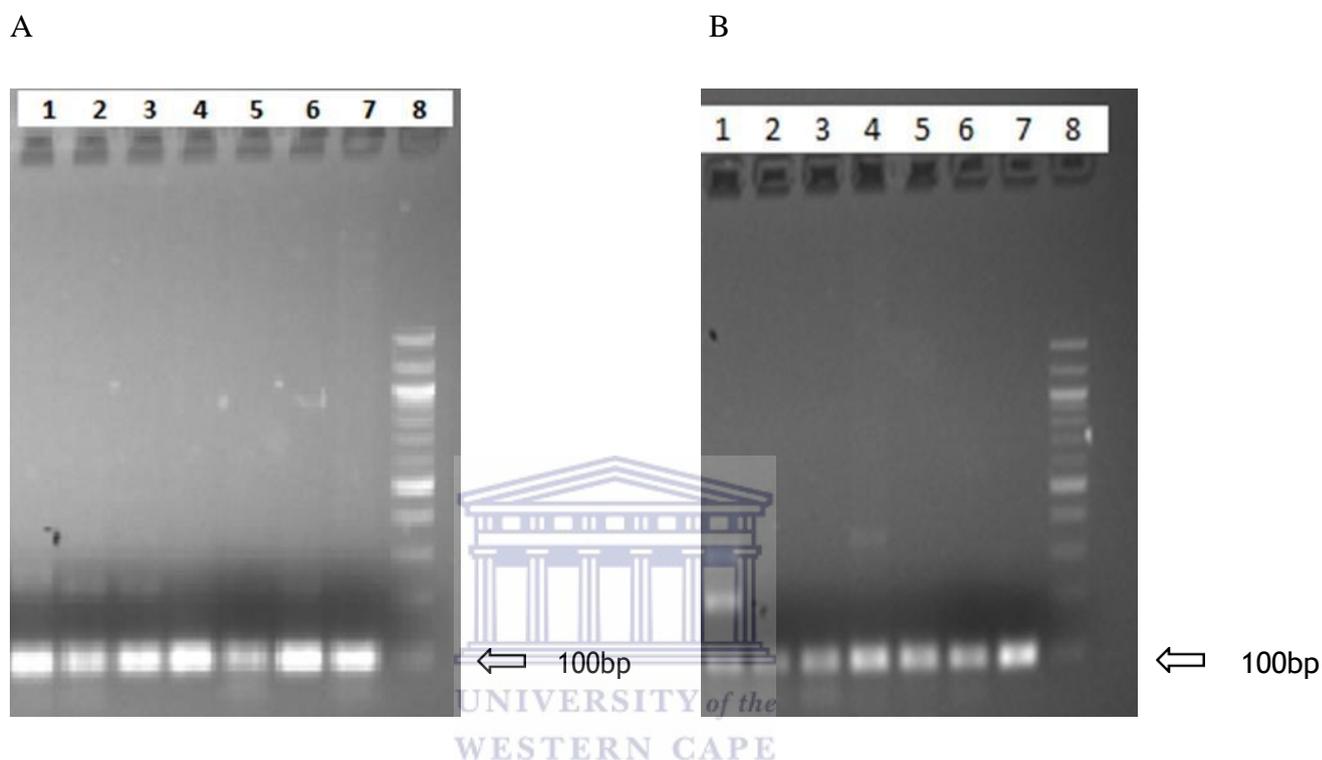


Figure 4.6 Agarose gel electrophoresis of representative qPCR amplicons from clinic samples for pLc4 and pLc17. A) pLc17 amplicons from samples (1) SD18. (2) SD4. (3) SH10. (4) SH16. (5) SD25. (6) SH40. (7) Positive Control (pLc17-diluted to 10000 copies/ μ l-expected size of 105bp) (8) Molecular marker: 1Kb (Thermo Scientific). B) pLc4 amplicons from samples (1) SD13. (2) SH3. (3) SH11. (4) SH5. (5) SD41. (6) SH2. (7) SH40. (7) Positive Control (pLc4-diluted to 10000 copies-expected size of 103bp) (8) Molecular marker: 1Kb (Thermo Scientific).

Figure 4.6 A) confirms that when screening with the pLc17 primers, the desired product (105bp) was formed for SD25, SH16 and SH40; while SD4, SH10 and SD18 appear to have additional bands suggesting that another PCR product was amplified, which corroborates the double peaks formed in the melt curve. In Figure 4.6 B) the desired product for pLc4 (103bp) was formed for SH3, SH11, SH2, SH40, SD41. SD13 and SH5 had multiple bands indicating that nonspecific amplification had occurred, confirming the melt curve analysis.

The 28 swab samples subjected to qPCR using both qPL4 and qPL17 had Cq values <31 cycles which was outside the dynamic range of quantifiable PCN (Figure 4.4). Thus the swab samples were within detectable/quantifiable range when normalized. In addition, positive ratios were obtained when swab samples were normalised with the reference gene (EGFR), illustrating that the DNA quantified was the plasmid DNA with low levels of human DNA (Table 4.3).

Table 4.3: Relative PCN and mean DNA concentration of pLc4, pLc17 and reference gene (EGFR)

Sample Name	Sample type	Sample label	Target/Ref ratio (Normalized ratios)		DNA concentration	
			pLc4	pLc17	pLc4	pLc17
CDC16-071-2015	Healthy	SH2	4,82E+02	1,19E+01	9,33E+02	2,30E+01
GP-09-VDS-106	Healthy	SH3	2,67E+02	2,37E+00	1,22E+03	1,09E+01
GP-09-VDS-123	BV	SD4	4,50E+00	4,52E+00	2,24E+01	2,25E+01
GP-09-VDS-126	Healthy	SH5	5,17E+00	4,64E+00	1,54E+01	1,38E+01
GP-09-VDS-125	BV	SD6	1,47E+01	2,31E+01	1,72E+01	2,71E+01
GP-09-VDS-131	Healthy	SH10	7,58E+00	2,99E+01	1,66E+01	1,32E+01
GP-09-VDS-099	Healthy	SH11	3,61E+01	1,71E+01	1,59E+01	1,82E+01
GP-09-VDS-102	Healthy	SH12	1,46E+01	1,76E+00	1,55E+01	9,86E+00
GP-09-VDS-109	BV	SD13	2,82E+00	2,08E+00	1,58E+01	1,11E+01
GP-09-VDS-141	BV	SD14	3,32E+00	1,93E+01	1,77E+01	2,90E+01
CDC16-071-2021	BV	SD15	9,98E+00	1,00E+01	1,49E+01	1,48E+01
CDC16-071-2020	Healthy	SH16	9,98E+00	4,82E+03	1,47E+01	5,69E+03

CDC16-071-2018	Healthy	SH17	1,30E+01	4,17E+00	1,53E+01	1,38E+01
CDC16-071-2002	BV	SD18	5,24E+00	1,00E+00	1,74E+01	7,93E+00
GP-09-VDS-104	Healthy	SH19	2,08E+00	5,71E+00	1,64E+01	1,27E+01
GP-09-VDS-132	BV	SD20	8,07E+00	2,28E+01	1,80E+01	5,75E+01
CDC16-071-1500	Healthy	SH21	7,56E+00	1,41E+00	1,91E+01	1,38E+01
GP-09-VDS-136	Healthy	SH22	1,79E+00	6,35E+00	1,75E+01	1,11E+01
CDC16-071-1487	BV	SD23	1,08E+01	2,65E+00	1,89E+01	1,95E+01
CDC16-071-1482	Healthy	SH24	1,83E+00	1,82E+01	1,35E+01	1,19E+01
GP-09-VDS-138	BV	SD25	2,28E+01	4,07E+00	1,49E+01	1,43E+01
GP-09-VDS-128	Healthy	SH26	4,65E+00	1,56E+01	1,63E+01	1,22E+02
CDC16-071-2014	Healthy	SH30	8,38E+00	4,79E+00	1,70E+01	3,04E+01
GP-09-VDS-068	BV	SD31	1,68E+00	2,04E+01	1,07E+01	2,92E+01
GP-09-VDS-103	BV	SD39	3,52E+00	3,22E+00	5,05E+00	4,63E+00
GP-09-VDS-097	Healthy	SH40	2,18E+03	2,22E+03	1,97E+03	2,67E+03
GP-09-VDS-078	BV	SD41	4,30E+00	3,06E+01	5,17E+00	2,36E+02
CDC16-071-1488	Healthy	SH42	1,57E+00	5,64E+00	2,25E+00	2,13E+01



SH=Healthy patient sample



SD=BV-infected patient sample

Zhong *et al.* (2011) describes PCN as either low (1-10 copies), medium (11-20 copies), or high (>20 copies). The low C_q scores observed for the majority of swab samples in Figures 4.5, suggests that if the plasmid is present, it is at low concentrations (1 to 10 copies/μl) or indicates the presence of contaminating genomic DNA. This may be due to external factors such as the collection method and extraction method used. The collection method used in the study (region on the vaginal tract swabbed, swab material used and the time the swab was collected i.e. pre-menstruation or post-menstruation) might have favoured the collection of certain microorganisms. The extraction method used might have favoured the extraction of more gDNA to plasmid DNA in samples that already had low amounts of the plasmid DNA. Furthermore, several authors have found that there is never a 100% recovery of the plasmid DNA after extraction, regardless of the method used (Carapuca *et al.* 2007; Friehs 2004; Skulj *et al.* 2008). To test whether the bacterial gDNA was more abundant in the samples, the samples were subjected to qPCR using 16S rRNA gene primers (Fig 4.7). The 9 swab samples selected, were selected based on their low plasmid DNA concentration (<10 copies/μl) (Table 4.3).

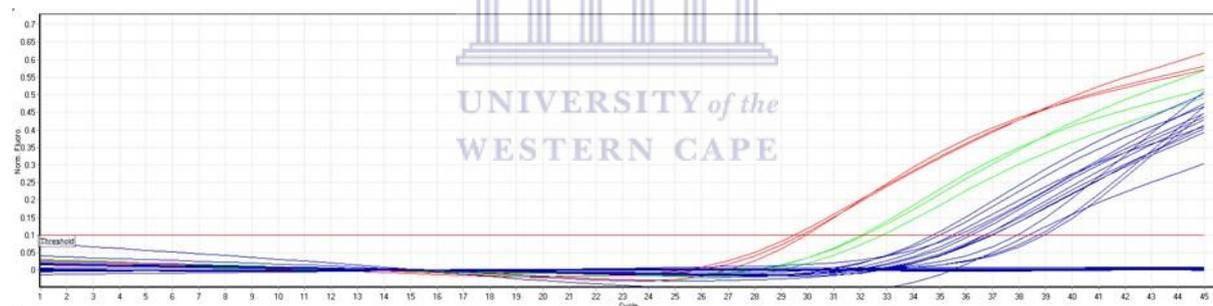


Figure 4.7: Amplification curve of vaginal swab samples. The swab samples (SD4, SH5, SH10, SD13, SH17, SD18, SH19, and SD20) indicated with blue curves and swab sample (SD25) indicated by the green curve, were subjected to qPCR using 16S rRNA primers. Positive control (*E. coli* diluted to 0.25 pg/μl) indicated by red curve.

Swab samples indicated in blue in Figure 4.7 were at DNA concentrations of 0.0025 pg/μl and SD25 was at 0.025 pg/μl. These low DNA concentrations suggest that little to no bacterial gDNA was extracted with the plasmid DNA. This partially excludes the possibility that bacterial gDNA was co-extracted with plasmid DNA. Agarose gel electrophoresis was not required as a probe based assay was used in this experiment (Chapter 2.9.3).

An alternate explanation could be the form of the plasmid DNA extracted- supercoiled versus relaxed (open circular or linear). Supercoiled DNA is known to suppress qPCR, where changes in conformation caused by DNA breaks increase detection of the relaxed DNA. This was hypothesized to be due to the DNA breaks allowing primers easy access to bind and elongation to proceed (Chen *et al.* 2007). Thus it is possible that during plasmid extraction the majority of plasmid DNA extracted might have been in the supercoiled form and not the relaxed form. Primer template mismatch has been shown to also impact amplification. In a study performed by Stadhouders *et al.* (2010) where the impact of primer template mismatch on qPCR reactions was investigated, it was found that primer template mismatch at the 3' end significantly reduced the qPCR amplification. The mismatches were thought to interfere with the polymerase affecting PCR amplification (Stadhouders *et al.* 2010).

Plasmids are seen as selfish elements which are a burden to the host, for this reason a host would benefit more by eliminating the plasmid (Zhong *et al.* 2011). Thus another possible explanation for the low copy number could be related to the plasmids regulating their rate of replication to ensure that they do not carry a great burden to their host.

The relative PCN according to the Relative Expression Software Tool (REST) for pLc4 and pLc17 increased by ~2.803 (P=0.133) and ~1.693 (P=0.494), respectively in healthy vs BV infected patient samples (Table 4.4). However, there were not found to be significant differences when taking the standard error into account (Figure 4.8).

Table 4.4: Relative PCN ratio and SE of pLc4 and pLc17

Plasmid	Ratio	Standard error (SE)
pLc4	2.803	±1.63201
pLc17	1.693	±1.0915

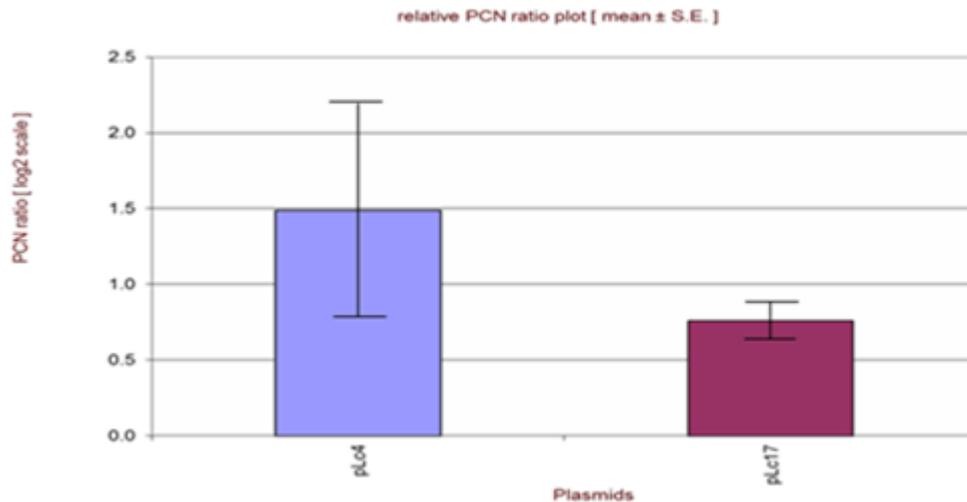


Figure 4.8: Relative PCN ratio of pLc4 (n=26) and pLc17 (n=25) for healthy samples relative to BV infected samples. The fold change was determined using the $2^{-\Delta\Delta Cq}$ method.

The sample size used in an experiment in part influences the precision and level of confidence. The larger the sample size, the more information one has and thus the uncertainty (caused by the variability) is reduced. In the current study only 28 samples were tested. Thus a larger sample size might have improved the statistical significance observed. In addition to this, the uneven distribution of healthy to BV infected samples (Table 4.2), may also have affected the precision (Faber & Fonseca 2014). Furthermore, pLc4 was found to be more abundant than pLc17 in the healthy patient samples. pLc4 was found in both *L. crispatus* and *L. jensenii*, which may account for the high abundance observed for pLc4. However, this finding may be due to experimental bias related to the relative size of the plasmids. Larger plasmids are known to be more difficult to extract compared to smaller plasmids. This has been found to be due to the inefficient cellular lysis and low copy numbers (Anderson & McKay 1983; Williams *et al.* 2006). This difference in extraction efficiency may have favored the selection of pLc4 (4224bp) compared to pLc17 (16663bp), resulting in the difference in abundance between the two plasmids. Methods used to improve extraction of larger plasmid DNA include the use of larger volumes of cell mass (Anderson & McKay 1983).

A large proportion of the samples subjected to qPCR in this study contained both plasmids, with 26/28 samples harbouring pLc4 and 25/28 samples harbouring pLc17. The following samples were excluded to avoid the over estimation of the quantification of both samples: SH5 and SD13 for pLc4 and SD4, SH10 and SD18 for pLc17 (Figure 4.6 and 4.7). This was

confirmed through melt curve analysis, and agarose gels. In addition, all samples were within the dynamic range of linearity of the qPCR assay (Cq value <31).

4.3 Conclusion

According to REST analysis there was an increase in relative PCN for both plasmids by ~2.803 fold for pLc4 and ~1.693 fold for pLc17 in healthy patient samples relative to BV infected samples. pLc4 was also found to be more abundant than pLc17 in healthy patients. This suggests that pLc4 and pLc17 might play a role in *L. crispatus* and *L. jensenii* abundance in healthy vaginal microbiomes. However, the standard deviation was too high for this to be considered a significant difference between the samples, which may be due to the small sample size used in this experiment (n=28). To enhance the statistical analysis, it is advisable to use a larger sample size which will allow one to extrapolate the findings to ensure that the total population is represented. In addition, to avoid skewing the results the proportion of healthy versus BV infected samples should be equal (Faber and Fonseca, 2014).

To properly correlate the impact of both plasmids on survival/fitness of *L. crispatus* and *L. jensenii* further analysis needs to be done. This could be achieved by measuring the relative PCN of both plasmids with the PCN per chromosome for both lactobacilli, especially considering that pLc4 was found to be harboured by both of the *Lactobacillus* species. The total DNA would need to be extracted, purified, and subjected to qPCR using the qPL4, qPL17 as well as appropriate *L. crispatus* and *L. jensenii* specific primers with appropriate reference genes included.

CHAPTER 5 GENERAL DISCUSSION AND CONCLUSION

Several studies have shown that *Lactobacillus crispatus* and *Lactobacillus jensenii* are essential for the maintenance of a healthy vaginal microbiome and possess several key mechanisms in order to gain dominance over other resident inhabitants of the vagina. These include the production of lactic acid, H₂O₂ and bacteriocins (Boris & Barbés 2000; Martínez-Peña *et al.* 2013; Mirmonsef & Spear 2014; Motevaseli *et al.* 2013). A previous study conducted on the distribution of lactobacilli in women with and without vaginal discharge syndrome (VDS) found that *L. crispatus*, *L. jensenii*, *L. iners* and *L. gasseri* were the dominant lactobacilli in the vagina (Damelin *et al.* 2011). Furthermore, *L. crispatus* was found to be equally distributed among women with and without VDS and were not significantly reduced in women with BV, while *L. jensenii* was reduced in women with VDS and BV. High levels of lysogeny was also observed in *L. crispatus* isolates (77% lysogeny), indicating high phage exposure (Damelin *et al.* 2011). In previous studies it has been established that prophages can contribute to host survival in adverse environments through the acquisition of advantageous prophage genes (Chibani-Chennoufi *et al.* 2004). Furthermore, plasmids are genetic mobile elements that have being shown to confer advantages to their host (Zhong *et al.* 2011; Cui *et al.* 2015). The aim of this study was to investigate if these plasmids may play a similar role to that of prophages. In order to investigate the potential role of the plasmids, the plasmids were first sequenced and bioinformatics analysis was performed (Chapter 3). Despite the potential benefits shown to be associated with the colonisation of *L. crispatus* and *L. jensenii* in the vagina, the present study is the first study to examine the plasmids found in these vaginal strains.

This study focused on plasmids extracted from clinical strains collected as part of the initial study conducted by Damelin *et al.* (2011). Two plasmids were identified and designated pLc4 and pLc17. The current study combined next generation sequencing and bioinformatics analysis, with a molecular biology approach. Key findings from the bioinformatics analysis indicated that, similar to *Bacillus subtilis* pIP501, pLc17 appears to be a broad host range plasmid which has two potential replicon types, rolling circle (RCR) and theta (Pujol *et al.* 1998). This has been found in many artificial constructs but is rare in natural plasmids, with pJD4, pSCP1 and pTB19 being the exceptions (Trembizki *et al.* 2014). The size limitation of RCR plasmids (>12kb) suggests that pLc17 is likely to employ theta replication as its mode

of replication (Xi *et al.* 2013; Khan 1997). Several accessory genes were identified on pLc17. The plasmid has numerous mobile DNA elements which might serve to allow integration of genomic elements in the host. RNR, a gene commonly associated with phages, was identified on pLc17. These genes on pLc17 may have been integrated onto the plasmid as a result of the numerous mobile elements flanking the genes. In addition, these genes also suggest phage infection has occurred which has been seen in numerous studies of vaginal isolates (Dwivedi *et al.* 2013; Ojala *et al.* 2014). The FIC-like protein, Rep B and resolvase identified may be involved in ensuring the stable maintenance of the plasmid in the host. The FIC-like protein may belong to phd (prevent host death) antitoxin and doc (death on curing) toxin (Castro-Roa *et al.* 2013; Cruz *et al.* 2014). This family of proteins is known to arrest the host activity placing the host in a dormant state, where the host becomes drug resistant (Cohen *et al.* 2013). The system may also play a role in assisting the host in phage exclusion (Fineran *et al.* 2009; Unterholzner *et al.* 2013). The ATPase may play a role in plasmid partitioning and is vital in proper cell division, particularly of low copy number plasmids (Fung *et al.* 2001). Interestingly, pLc4 was identified in both *L. crispatus* and *L. jensenii*. This suggests that pLc4 may replicate in more than one host and may have a broad host range (BHR), despite the fact that the plasmid does not possess a broad host range replicon. The mode of replication predicted for pLc4 is RCR based on the size of the plasmid, as many plasmids utilizing RCR are under 12kb (Cui *et al.* 2015). Bioinformatics analysis revealed that many of pLc4 genes were novel and could not be identified. These genes however, were similar to genes previously identified in *Streptococcus* species, which suggest that the plasmid may have been derived from a *Streptococcus* plasmid (del Solar *et al.* 1998). The Broad Host Range (BHR) nature exhibited by pLc4 may have led to the acquisition of genes from surrounding *Streptococcus* species. BHR plasmids as mentioned in Chapter 1 are of fundamental importance not only to the contribution to horizontal gene transfer but also the use of their replicons as cloning vectors for downstream biotechnology applications. Future studies should investigate the extent of pLc4 host range. This can be achieved by comparing the trinucleotide composition of the plasmid with sequenced bacterial chromosomes. If the plasmid signature is not similar to the chromosomal signature, then the plasmid is likely a BHR plasmid (Jain *et al.*, 2013). This method was validated and used in a study by Suzuki *et*

al., 2010 on known host range (Suzuki *et al.* 2010). Further characterisation of pLc4 is required to identify the genes associated with the plasmid and the possible role in enhancing the survival of *L. crispatus* and *L. jenssenii* in the adverse conditions of the vaginal microbiome.

The mode of replication employed by pLc4 also needs to be determined. This can be done through the use of Southern hybridization. This method is consistently used for the detection of ssDNA intermediates found on RCR plasmids (Li *et al.* 2015; Zhou *et al.* 2010). Further features such as the plasmid maintenance and stability for both pLc4 and pLc17 should also be investigated. This information would be of particular interest for pLc17 that encodes for the segregation stability elements FIC-like protein, ATPase and resolvase. This would be further confirmation of the putative proteins detected during sequencing.

In Chapter 4 the relative abundance of pLc4 and pLc17 in healthy patients and BV infected patients was determined. Due to its accuracy and sensitivity qPCR was used as the detection method to measure the relative abundance of the plasmids. Despite the benefits of real time PCR, its accurate quantification is influenced by DNA extraction method and sample type. Due to the nature of the *Lactobacillus* cell wall several DNA extraction methods were tested. The Zyppy gDNA extraction kit with a modified lysis step was found to be successful, however this kit tends to favour the extraction of gDNA which may have influenced the amount of plasmid DNA extracted. The gDNA contamination was evaluated using 16S rRNA as a proxy and all samples were found to contain very low to no bacterial gDNA. Further factors such as the selection of an appropriate reference gene and the quality of the standards play a huge role in the results obtained using this method. A recent technology droplet digital PCR (dd PCR) has been shown to accurately quantify nucleic acid avoiding the use of calibration curves and any external standards (Plotka *et al.* 2017), and should therefore be considered in future studies.

REST analysis revealed that the relative PCN for pLc4 and pLc17 had a fold increase of ~2.803 and ~1.693 fold, respectively in healthy patient samples relative to BV infected patient samples, suggesting that these plasmids might play a role in *L. crispatus* and

jensenii abundance in healthy vaginal microbiome. However, there was no significant difference between the samples. pLc4 was present in 26 out of the 28 samples and pLc17 was present in 25 out of the 28 samples analysed. pLc4 was found to be more abundant than pLc17, this may be attributed to the size of the plasmids. Due to inefficient cellular lysis and low plasmid copy number of the larger plasmid extraction methods tend to favour the extraction of smaller plasmids (Anderson & McKay 1983; Williams *et al.* 2006). The samples exhibited relatively low Cq values with plasmid concentration of 1-10 copies/ μ l of extracted sample. This may be due to the conformation of the plasmid extracted, as a possible mixture of supercoiled and relaxed (open circular or linear) may have been subjected to qPCR. Alternatively, this might be a result of the low amount of plasmid DNA present in the sample.

The sample size for any experiment can influence the research outcomes and for this reason careful consideration regarding the sample size should be taken into account (Faber & Fonseca 2014). The sample size used in this study was relatively low; 28 swab samples of which 16 belonged to healthy patients and only 12 from BV infected patients, which may have skewed the analysis towards the healthy patient samples. Furthermore, a sample size of 28 would not allow the findings to be extrapolated to the overall population. A larger sample size would allow for a more accurate account of the effects of these plasmids in healthy and BV infected individuals.

The selection of the reference gene should be carefully considered based on tissue type and experimental situation (Gorzelnik *et al.* 2001). Furthermore, as per the MIQE guideline the selection of an appropriate reference gene based on stable expression in various tissues and the selection based on the tissue and cell type used in the experiment, are vital to successful normalization. The MIQE guideline also encourages the use of more than one reference gene for normalization to accurate normalization (Bustin *et al.* 2013). Thus future qPCR studies on pLc4 and pLc17 should be conducted with more than one validated reference gene. Reference gene validation can be done using normalisation software such as NormFinder and GeNorm.

Although plasmids are often seen as a metabolic burden to the host (with plasmid bearing bacteria often outcompeted by plasmid free bacteria), several studies have shown that a host-plasmid adaptation can be obtained over evolutionary time through genetic changes in either the host, plasmid or both (Brown *et al.* 2013; Dahlberg & Chao 2003; Sota *et al.* 2010; Loftie-Eaton *et al.* 2017). In a study done by Loftie-Eaton *et al.* (2017), it was found that bacteria can adapt to conjugative multiple drug resistant plasmids when there is a shift from

plasmid cost to benefit and that the adaptation to one plasmid can lead to improve plasmid permissiveness (Loftie-Eaton *et al.* 2017). The same may be true for pLc4 and pLc17 persistence in *L. crispatus* and *L. jensenii*. This can be tested by plasmid evolution experiments and plasmid persistence can be measured.



Reference

- Amsel, R., Totten, PA, Spiegel, CA, Chen, KC, Eschenbach, D & Holmes, KK 1983, 'Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations', *Am. J. Med.*, vol. 74, no. 1, pp. 14–22.
- Anderson, DG & McKay, LL 1983. 'Simple and rapid method for isolating large plasmid DNA from lactic streptococci', *Appl. Environ. Microbiol.*, vol. 46, no. 3, pp. 549-552.
- Antonio, MAD, Hawes, SE & Hillier, SL 1999. 'The identification of vaginal *Lactobacillus* species and the demographic and microbiologic characteristics of women colonized by these species', *J. Infect. Dis.*, vol. 180, no. 6, pp. 1950-1956.
- Antonio, MAD, Meyn, LA, Murray, PJ, Busse, B & Miller, SL 2009, 'Vaginal colonization by probiotic *Lactobacillus crispatus* CTV-05 is decreased by sexual activity and endogenous lactobacilli', *J. Infect. Dis.*, vol. 199, no. 10, pp. 1506-1513.
- Anindyajati, A, Artarini, A, Riani, C & Retnoningrum, DS 2016, 'Plasmid copy determination by quantitative PCR', *Sci. Pharm.*, vol. 84, no. 1, pp. 89-101.
- Anukam, KC, Osazuwa, EO, Ahonkhani, I, & Reid, G 2006, '*Lactobacillus* vaginal microbiota of women attending in reproductive health care service in Benin city, Nigeria', *Sex. Transm. Dis.*, vol. 33, no. 1, pp. 59-62.
- Babalola, OO, Kirby, BM, Le Roes-Hill, M, Cook, AE, Cary, SC, Burton, SG, & Cowan, DA 2009, 'Phylogentic analysis of actinobacterial populations associated with Antarctic dry valley mineral soils', *Environ. Microbiol.*, vol. 11, no. 3, pp.566-576.
- Bao, S, Jiang, R, Kwan, RW, Wang, B, Ma, X & Song, YK 2011, 'Evaluation of next generation sequencing software in mapping and assembly', *J. Hum. Genet.*, vol. 56, no. 6, pp. 406-414.
- Baxter, JC & Funnell, BE 2014, 'Plasmid partitioning mechanisms', *Microbiol. Spectr.*, vol. 2, no. 6, pp. 1-20.

- Bennett, PM 2004, 'Transposable elements', Schaechter M (ed). The desk encyclopaedia of microbiology. Elsevier academic press: San Diego, CA.
- Bignell, C & Thomas, CM 2001, 'The bacterial ParA-ParB partitioning proteins', *J. Biotechnol.*, vol. 91, no. 1, pp. 1-34.
- Boris, S & Barbes, C 2000, 'Role played by lactobacilli in controlling the population of vaginal pathogens', *Microbes Infect.*, vol. 2, no. 5, pp. 543–546.
- Brouwer, MSM, Tagg, KA, Mevius, DJ, Iredell, JR, Bossers, A, Smith, HE & Partridge, SR 2015, 'IncI shufflons: assembly issues in the next-generation sequencing era', *Plasmid*, vol. 80, no. 8, pp. 111-117.
- Brown, CJ, Sen, D & Yano, H 2013, 'Diverse broad host range plasmids from freshwater carry few accessory genes', *Appl. Environ. Microbiol.*, vol. 79, no. 24, pp. 7684-7695.
- Buermans, HPJ & den Dunnen. JT 2014, 'Next generation sequencing technology: advances and applications', *Biochim. Biophys. Acta*, vol. 1842, pp.1932-1941.
- Burton, JP, Cadieux, PA & Reid, G 2003, 'Improved understanding of the bacterial vaginal microbiota of women before and after probiotic instillation', *Appl. Environ. Microbiol.*, vol. 69, no. 1, pp. 97-101.
- Bustin, SA, Benes, V, Garson, J, Hellems, J, Huggett, J, Kubista, M, Mueller, R, Nolan, T, Pfaffl, MW, Shipley, G, Wittwer, CT, Schjerling, P, Day, PJ, Abreu, M, Aguado, B, Beaulieu, JF, Beckers, A, Bogaert, S, Browne, JA, Ramiro, FC, Ceelen, L, Ciborowski, K, Cornillie, P, Coulon, S, Cuypers, A, De Brouwer, S, De Ceuninck, L, De Craene, J, De Naeyer, H, De Spiegelare, W, Deckers, K, Dheedene, A, Durinck, K, Teixeira, M, Fieuw, A, Gallup, JM, Flores, SG, Goossens, K, Heindryckx, F, Herring, E, Hoenicka, H, Icardi, L, Jaggi, R, Javad, F, Karampelias, M, Kibenge, F, Kibenge, M, Kumps, C, Lambertz, I, Lammens, T, Markey, A, Messiaen, P, Mets, E, Morais, S, Rubio, AM, Nakiwala, J, Nelis, H, Olsvik, PA, Novo, CA, Plusquin, M, Remans, T, Rihani, A, Santos, PR, Rondou, P, Sanders, R, Bleek, K, Skovgaard, K, Smeets, K, Tabera, L, Toegel, S, Acker, TV, Van den Broeck, W, Van der Meulen, J, Van Gele, M, Van Peer, G, Van Poucke, M, Van Roy, N, Vergult, S, Wauman, J, Wiklander, M, Willems, E, Zaccara, S, Zeka, F & Vandesompe, J 2013, 'The need for transparency and good practices in the qPCR literature', *Nature*, vol.

- 10, no. 11, pp.1063-1067.
- Carapuca, E, Azzoni, AR, Prazeres, DMF, Monterio, GA & Mergulhao, FJM 2007, 'Time course determination of plasmid content in eukaryotic and prokaryotic cells using real time PCR', *Mol. Biotechnol.*, vol. 37, no. 2, pp.120-126.
- Castro-Roa, D, Gieter, SD, Nuland, NAJV, Loris, R, Garcia-Pino, A & Zenkin, N 2013, 'The FIC protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu', *Nat. Chem. Biol.*, vol. 9, no. 12, pp. 811-817.
- Challacombe, JF, Pillai, S & Kuske, CR 2017, 'Shared features of cryptic plasmids from environmental and pathogenic *Francisella* species', *PLoS ONE*, vol. 12, no 8, pp.1-19.
- Chancey, CJ, Khanna, KV, Seegers, JFML, Zhang, GW, Hildreth, J, Langan, A & Markham, RB 2006, 'Lactobacilli-expressed single-chain variable fragment (scFv) specific for intercellular adhesion molecule 1 (ICAM-1) blocks cell-associated HIV-1 transmission across a cervical epithelial monolayer', *J. Immunol.*, vol. 176, no. 9, pp. 5627- 5636.
- Chen, J, Kadlubar, FF & Chen, JZ 2007, 'DNA supercoiling suppresses real-time PCR: a new approach to the quantification of mitochondrial DNA damaged and repair', *Nucleic Acids Res.*, vol.35, no. 4, pp. 1377-1388.
- Chibani-Chennoufi, S, Bruttin, A, Dillmann, ML & Brussow, H 2004, 'Phage-host interaction: an ecological perspective', *J. Bacteriol.*, vol.186, no. 12, pp. 3677-3686.
- Cho, I & Blaser, MJ 2012, 'The human microbiome: at the interface of health and disease', *Nat. Rev.*, vol. 13, no. 4, pp. 260-270.
- Cohen, NR, Lobritz, MA & Collins, JJ 2013, 'Microbial persistence and the road to drug resistance', *Cell Host Microbe*, vol.13, no. 6, pp. 632–642.
- Cruz, JW, Rothenbacher, FP, Maehigashi, T, Lane, WS, Dunham, CM & Woychik, NA 2014, 'Doc toxin is a kinase that inactivates elongation factor TU. *J. Biol. Chem.*, vol. 289, no. 11, pp. 7788-7798.
- Cui, Y, Hu, T, Qu, X, Zhang, L, Ding, Z & Dong, A 2015, 'Plasmids from food lactic acid bacteria: diversity, similarity, and new developments', *Int. J. Mol. Sci.*, vol. 16, no. 6, pp. 13172-13202.

- Dahlberg, C & Chao, L 2003, 'Amelioration of the cost of conjugative plasmid carriage in *Escherichia coli* K12', *Genet.*, vol. 165, no. 4, pp. 1641-1649.
- Damelin, LH, Mavri-Damelin, D, Klaenhammer, TR & Tiemessen, CT 2010, 'Plasmid transduction using bacteriophage λ adh for expression of CC chemokines by *Lactobacillus gasseri* ADH', *Appl. Environ. Microbiol.*, vol. 76, no. 6, pp. 3878-3885.
- Damelin, LH, Paximadis, M, Mavri-Damelin, D, Birkhead, M, Lewis, DA & Tiemessen, CT 2011, 'Identification of predominant culturable vaginal *Lactobacillus* species and associated bacteriophages from women with and without vaginal discharge in South Africa', *J. Med. Microbiol.*, vol. 60, no. 2, pp. 180-183.
- Deepak, SA, Kottapalli, KR, Rakwal, R, Oros, G, Rangappa, KS, Iwahashi, H, Masuo, Y & Agrawal, GK 2007, 'Real-time PCR: revolutionizing detection and expression analysis of genes', *Curr. Genomics*, vol. 8, no. 4, pp. 234-251.
- Desmond, C, Ross, RP, Fitzgerald, G & Stanton, C 2005, 'Sequence analysis of the plasmid genome *Lactobacillus Paracasei* NFBC338 which includes the plasmids pCD01 and pCD02', *Plasmid*, vol. 54, no. 2, pp. 160-175.
- del Solar, G, Moscoso, M, & Espinosa, M 1993, 'Rolling circle-replicating plasmids from gram-positive and gram-negative bacteria: a wall falls', *Mol. Microbiol.*, vol. 8, no. 5, pp. 789- 796.
- del Solar, G, Giraldo, R, Ruiz-Echevarría, MJ, Espinosa, M & Díaz-Orejas, R 1998, 'Replication and control of circular bacterial plasmids', *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 434-464.
- Douillard, FB & de Vos, WM 2014, 'Functional genomics of lactic acid bacteria: from food to health', *Microb. Cell Fact.*, vol. 13, no. 1, pp. 1-21.
- Dwivedi, B, Xue, B, Lundin, D, Edwards, RA & Breitbart, M 2013, 'A bioinformatics analysis of ribonucleotide reductase genes in phage genomes and metagenomes', *BMC Evol. Biol.*, vol. 13, no. 1, pp. 1-17.
- Eade, CR, Diaz, C, Wood, MP, Anastos, K, Patterson, BK, Gupta, P, Cole, AL & Cole, AM 2012, 'Identification and characterisation of bacterial vaginosis-associated pathogens using a comprehensive cervical-vaginal epithelial

- coculture assay', *PLoS ONE*, vol. 7, no. 11, pp.1-11.
- Erdmann, N, Petroff, T & Funnell, BE 1999, 'Intracellular localization of P1 ParB protein depends on ParA and ParS', *Proc. Natl. Acad. Sci.*, vol. 96, no. 26, pp. 14905–14910.
- Ergan, AN, Schlueter, J & Spooner, DM 2012, 'Applications of next generation sequencing in plant biology', *Am. J. Bot.*, vol. 99, no. 2, pp. 175-185.
- Faber, J & Fonesca, LM 2014, 'How sample size influences research outcomes', *Dental Press J. orthod.*, vol. 19, no. 4, pp. 27-29.
- Falkow, S & Baron, LS 1962, 'Episomic element in a strain of *Salmonella typhosa*', *J. Bacteriol.*, vol. 84, no. 3, pp. 581–589.
- Fang, F, Flynn, S, Li, Y, Claesson, MJ, van Pijkeren, J, Collins, JK, van Sinderen, D & O'Toole, PW 2008, 'Characterization of endogenous plasmids from *Lactobacillus salivarius* UCC118', *Appl. Environ. Microbiol.*, vol. 74, no. 10, pp. 3216-3228.
- Fettweis, JM, Serrano, MG, Girerd, PH, Jefferson, KK & Buck, GA 2012, 'A new era of the vaginal microbiome: advances using next-generation sequencing', *Chem. Biodivers.*, vol. 9, no. 5, pp. 965-976.
- Fineran, PC, Blower, TR, Foulds, IJ, Humphreys, DP, Lilley, KS & Salmond, GPC 2009, 'The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair', *Proc. Natl. Acad. Sci.*, vol. 106, no. 9, pp. 894–899.
- Fukao, M, Oshima, K, Morita, H, Toh, H, Suda, W, Kim, SW, Suzuki, S, Yakabe, T, Hattori, M & Yajima, N 2013, 'Genomic analysis by deep sequencing of the probiotic *Lactobacillus Brevis* KB290 harbouring nine plasmids reveal genomic stability', *PLoS ONE*, vol. 8, no. 3, pp. 1-12.
- Fredricks, DN, Fiedler, TL & Mrazek, JM 2005, 'Molecular identification of bacteria associated with bacterial vaginosis', *N. Engl. Med.*, vol. 353, no. 18, pp. 1899-1911.
- Friebs, K 2004, 'Plasmid copy number (PCN) and plasmid stability', *Adv. Biochem. Eng./Biotechnol.*, vol. 86, pp. 47-82.
- Fung, E, Bouet, JY & Funnell, BE 2001, 'Probing the ATP binding site of P1

- ParA: partitioning and repression have different requirements for ATP binding and hydrolysis', *EMBO J.*, vol. 20, no. 17, pp. 4901-4911.
- Gerdes, K & Maisonneuve, E 2012, 'Bacterial persistence and toxin antitoxin loci', *Annu. Rev. Microbiol.*, vol. 66, pp. 103-123.
- Gerhardt, P, Murray, RGE, Wood, WA & Krieg, NR 1994, 'Plasmids methods for general and molecular biotechnology', chapter 16. American Society for Microbiology, Washington D.C., USA.
- Ghartey, JP, Smith, BC, Chen, Z, Buckley, N, Lo, Y, Ratner, AJ, Herold, BC & Burk, RD 2014, '*Lactobacillus crispatus* dominant vaginal microbiome is associated with inhibitory activity of female genital tract secretions against *Escherchia coli*', *PLoS ONE*, vol. 9, no. 5, pp. 1-8.
- Gilbert, NM, Lewis, WG & Lewis, A 2013, 'Clinical features of bacterial vaginosis in a murine model of vaginal infection with *Gardnerella vaginalis*', *PLoS ONE*, vol. 8, no. 3, pp. 1-12.
- Gorzelnik, K, Janke, J, Engeli, S & Sharma, AM 2001, 'Validation of endogenous controls for gene expression studies in human adipocytes and preadipocytes', *Horm. Metab. Res.*, vol. 33, no. 10, pp. 625-627.
- Gueimonde, M, Sánchez, B, de los Reyes-Gavilán, CG & Margolles, A 2013, 'Antibiotic resistance in probiotic bacteria', *Front. Microbiol.*, vol. 4, pp. 1-6.
- Henson, J, Tischler, G & Ning, Z 2012, 'Next-generation sequencing and large genome assemblies', *Pharmacogenomics J.*, vol. 13, no. 8, pp. 901-915.
- Heid, CA, Stevens, JK, and Livak, KJ & Williams, PM 1996, 'Real time quantitative PCR', *Genome Res.*, vol. 6, no. 10, pp. 986-994.
- Heiss, S, Grabherr, R & Heintl, S 2015, 'Characterization of the *Lactobacillus plantarum* plasmid pCD033 and generation of the plasmid free strain *L. plantarum* 3NSH'. *Plasmid*, vol. 81, pp. 9-20.
- Hesse, U, van Heusden, P, Kirby, BM, Olonade, I, van Zyl, LJ & Trindade, M 2017, 'Virome assembly and annotation: A surprise in the Namib desert', *Front. Microbiol.*, vol. 8, no. 13, pp. 1-17.
- Hickey, RJ, Zhou, X, Pierson, JD, Ravel, J & Forney, LJ 2012, 'Understanding vaginal microbiome complexity from an ecological perspective', *Transl. Res.*,

- vol. 160, no. 4, pp. 267- 282.
- Higuchi, R, Dollinger, G, Walsh, PS & Griffith, R 1992, 'Simultaneous amplification and detection of specific DNA sequences', *Biotechnol.*, vol. 10, no. 4, pp. 413-417.
- Hughes, JM, Lohman, BK, Deckert, GE, Nichols, EP, Settles, M, Abdo, Z & Top, EM 2012, 'The role of clonal interference in the evolutionary dynamics of plasmid- host adaption', *M. Bio.*, vol. 3, no. 4, pp. 1-8.
- Hodkinson, BP & Grice, EA 2015, 'Next-generation sequencing: A Review of technologies and tools for wound microbiome research', *Adv. Wound Care*, vol. 4, no. 1, pp. 50–58.
- Hou, Y, Zhang, H, Miranda, L & Lin, S 2010, 'Serious overestimation in quantitative PCR by circular (supercoiled) plasmid standard: microalgal *pcna* as the model gene', *PLoS ONE*, vol. 5, no. 3, pp. 1-7.
- Ito, Y, Kawai, Y, Arakawa, K, Honme, Y, Sasaki, T & Saito, T 2009, 'Conjugative plasmid from *Lactobacillus gasserii* LA39 that carries genes for production of and immunity to the circular bacteriocin gassericin A', *Appl. Environ. Microbiol.*, vol. 75, no. 19, pp. 6340–6351.
- Jain, A. & Srivastava, P 2013, 'Broad host range plasmids', *FEMS Microbial Lett.*, vol. 348, no. 2, pp. 87-96.
- Janitz, M 2008, 'Next-generation genome sequencing: towards personalized medicine', Wiley-VCH Verlag GmbH & Co. KGaA, USA.
- Jime'nez, E, Mart'ın, R, Maldonado, A, Mart'ın, V, de Segura, AG, Fern'andez, L, & Rodr'ıguez, JM 2010, 'Complete genome sequence of *Lactobacillus salivarius* CECT 5713, a probiotic strain isolated from human milk and infant feces', *J. Bacteriol.*, vol. 192, pp. 5266–5267.
- Karaođlu, SA, Aydin, F, Kilic, SS, & Kilic, AO 2003, 'Antimicrobial activity and characteristics of bacteriocins produced by vaginal *Lactobacilli*', *Turk. J. Med. Sci.*, vol. 33, pp. 7- 13.
- Khan, SA 1997, 'Rolling-circle replication of bacterial plasmids', *Microbiol. Mol. Biol. Rev.*, vol. 61, no. 4, pp. 442-455.

- Koonin, EV & Galperin, MY 2003, 'Sequence - evolution - function: Computational approaches in comparative genomics', Boston, Kluwer academics, USA.
- Kubista, M, Andrade, JM, Bengtsson, M, Forootan, A, Jonak, J, Lind, K, Sindelka, R, Sjoback, R, Sjogreen, B, Strombom, L, Stahlberg, A & Zoric, N 2006, 'The real-time polymerase chain reaction', *Mol. Aspects Med.*, vol. 27, no. 2-3, pp. 95-125.
- Kunnimalaiyaan, M & Vary, PS 2005, 'Molecular characterization of plasmid pBM300 from *Bacillus megaterium* QM B1551', *Appl. Environ. Microbiol.*, vol. 71, no. 6, pp. 3068-3076.
- Laplante, PA 2006, 'Real time PCR', Taylor & Francis Group, 270 Madison Avenue New York, N Y 10016, USA.
- Lee, C, Kim, J, Shin, SG & Hwang, S 2006, 'Absolute and relative qPCR quantification of plasmid copy number in *Escherichia coli*', *J. Biotechnol.*, vol. 123, no. 3, pp. 273-280.
- Lee, HK, Lee, CK, Wei-TzeTang, J, Loh, TP & Siew-ChuanKoay, E 2016, 'Contamination-controlled high throughput whole genome sequencing for influenza A viruses using the MiSeq sequencer', *Sci. Rep.*, vol. 6, no. 33318, pp. 1-11
- Lee, JH, Halgerson, JS, Kim, JH & O'Sullivan, DJ 2007, 'Comparative sequence analysis of plasmids from *Lactobacillus delbrueckii* and construction of a shuttle cloning vector', *Appl. Environ. Microbiol.*, vol. 73, no. 14, pp. 4417-4424.
- Leite, SRRF, Amorim, MMR, Sereno, PFB, Leite, TNF, Ferreira, JAC & Ximenes, RAA 2011, 'Randomized clinical trial comparing the efficacy of the vaginal use of metronidazole with a Brazilian pepper tree (*Schinus*) extract for the treatment of bacterial vaginosis', *Braz. J. Med. Biol. Res.*, vol. 44, no. 3, pp. 245-252.
- Lenski, RE, Simpson, SC & Nguyen, TT 1994, 'Genetic analysis of a plasmid-encoded, host genotype-specific enhancement of bacterial fitness', *J. Bacteriol.*, vol. 176, no. 11, pp. 3140-3147.

- Li, Z, Liu, ZB, Xing, A, Moon, BP, Koellhoffer, JP, Huang, L, Ward, RT, Clifton, E, Falco, SC & Cigan, AM 2015, 'Cas9-Guide RNA directed genome editing in soybean', *Plant Physiol.*, vol. 169, no. 2, pp. 960–970.
- Lin, CH, Yeh, KT, Chang, YS, Hsu, NC, Chang, JG 2010, 'Rapid detection of epidermal growth factor mutations with multiplex PCR and primer extension in lung cancer', *J. Biomed. Sci.*, vol. 17, no. 1, pp. 1-6.
- Lin, CH, Chen, YC & Pan, TM 2011, 'Quantification bias caused by plasmid DNA conformation in quantitative real-time PCR assay', *PLoS ONE*, vol. 6, no. 12, pp. 1-10.
- Ling, Z, Kong, J, Liu, F, Zhu, H, Chen, X, Wang, Y, Li, L, Nelson, KE, Xia, Y, Xiang, C 2010, 'Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis', *BMC Genomics*, vol. 11, pp. 1-16.
- Liu, X, Lagenaur, LA, Simpson, DA, Essenmacher, KP, Frazier-Parker, CL, Liu, Y, Tsai, D, Rao, SS, Hamer, DH, Parks, TP, Lee, PP & Xu, Q 2006, 'Engineered vaginal *Lactobacillus* strain for mucosal delivery of the Human Immunodeficiency Virus inhibitor Cyanovirin-N', *Antimicrob. Agents Chemother.*, vol. 50, no. 10, pp. 3250-3259.
- Liu, X, Lagenaur, LA, Lee, PP & Xu, Q 2008, 'Engineering of a human vaginal *Lactobacillus* strain for surface expression of two-domain CD4 molecules', *Appl. Environ. Microbiol.*, vol. 74, no. 15, pp. 4626-4634.
- Loftie-Eaton, W, Bashford, K, Quinn, H, Dong, K, Millstein, J, Hunter, S, Thomason, MK, Merrikkh, H, Ponciano, JM & Top, EM 2017, 'Compensatory mutations improve general permissiveness to antibiotic resistance plamids', *Nat. Ecol. Evol.*, vol. 1, no. 9, pp. 1354-1363.
- Luca, F, Hudson, RR, Witonsky, DB & Rienzo, AD 2012, 'A reduced representation approach to population genetic analyses and application to human evolution', *Genome Res.*, vol. 21, no. 7, pp. 1087-1098.
- Lundin, D, Gribaldo, S, Torrents, E, Sjöberg, BM & Poole, AM 2010, 'Ribonucleotide reduction - horizontal transfer of a required function spans all three domains', *BMC Evol. Biol.*, vol. 10, no. 383, pp. 1-19.
- Lu, C, Nakayasu, ES & Zhang, LQ 2016, 'Identification of Fic-1 as an enzyme that

- inhibits bacterial DNA replication by AMPylating GyrB, promoting filament formation', *Sci. Signal*, vol. 9, no. 412, pp. 11-26.
- Malek, RA, Hamdan, SB, El Enshasy, HA, Othman, NZ, Zainol, NA, Sarmidi, MR & Aziz, RA 2010, 'Production of *Lactobacillus salivarius*, a new probiotic strain isolated from human breast milk, in semi-industrial scale and studies on its functional characterization', *Appl. Microbiol. Microb. Biotechnol.*, vol. 2, pp. 1196-1204.
- Ma, B, Forney, LJ & Ravel, J 2012, 'The vaginal microbiome: rethinking health and diseases', *Annu. Rev. Microbiol.*, vol. 66, pp. 371–389.
- Ma, L, Lv, Z, Su, J, Wang, J, Yan, D, Wei, J & Pei, S 2013, 'Consistent condom use increases the colonization of *Lactobacillus crispatus* in the vagina', *PLoS ONE*, vol. 8, no. 7, pp. 1-7.
- Machado, A, Jefferson, KK & Cerca, N 2013, 'Interactions between *Lactobacillus crispatus* and bacterial vaginosis (BV)-Associated bacterial species in initial attachment and biofilm formation', *Int. J. Mol. Sci.*, vol. 14, no. 6, pp. 12004-12012.
- Marine, R, Polson, SW, Ravel, J, Hatfull, G, Russell, D, Sullivan, M, Syed, F, Dumas, M & Wommack, KE 2011, 'Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA', *Appl. Environ. Microbiol.*, vol. 77, no. 22, pp. 8071– 8079.
- Martin Jr., HL, Richardson, BA, Nyange, PM, Lavreys, L, Hillier, SL, Chohan, B, Mandaliya, K, Ndinya-Achola, JO, Bwayo, J & Kreiss, J 1999, 'Vaginal lactobacilli, microbial flora, and risk of Human Immunodeficiency Virus Type 1 and sexually transmitted disease acquisition', *J. Infect. Dis.*, vol. 180, no. 6, pp. 1863-1868.
- Martin, R, Soberon, N, Vaneechoutte, M, Florez, AB, Vazquez, F & Suarez, JE 2008, 'Characterisation of indigenous vaginal lactobacilli from healthy women as probiotic candidates', *Int. Microbiol.*, vol. 11, no. 4, pp. 261-266.
- Martinez-Pena, MD, Castro-Escarpulli, G & Aguilera-Arreola, MG 2013, '*Lactobacillus* species isolated from vaginal secretions of healthy and bacterial

- vaginosis-intermediate Mexican women: a prospective study', *BMC Infect. Dis.*, vol. 13, pp. 1-9.
- Masalha, M, Borovok, I, Schreiber, R, Aharonowitz, Y, & Cohen, G 2001, 'Analysis of transcription of the *Staphylococcus aureus* aerobic class Ib and anaerobic class III ribonucleotide reductase genes in response to oxygen', *J. Bacteriol.*, vol. 183, no. 24, pp. 7260-7272.
- Matu, MN, Orinda, GO, Njagi, ENM, Cohen, CR & Bukusi, EA 2010, 'In vitro inhibitory activity of human vaginal *Lactobacilli* against pathogenic bacteria associated with bacterial vaginosis in Kenyan women', *Anaerobe*, vol. 16, no. 3, pp. 210-215.
- Mclean, NW & Rosenstein, IJ 2000, 'Characterisation and selection of a *Lactobacillus* species to re-colonise the vagina of women with recurrent bacterial vaginosis', *J. Med. Microbiol.*, vol. 49, no. 6, pp. 543-552.
- Meijerink J, Mandigers, C, van de Locht, L, Tonnissen, E, Goodsaid, F & Raemaekers, J 2001, 'A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR', *J. Mol. Diagn.*, vol. 3, no. 2, pp. 55-61.
- Mesas, JM, Rodriguez, MC & Alegrec, MT 2004, 'Plasmid curing of *Oenococcus oeni*', *Plasmid*, vol. 51, no. 1, pp. 37-40.
- Metzker. ML 2010, 'Sequencing technologies-the next generation', *Nat. Rev. Genet.*, vol. 11, no. 1, pp. 31-46.
- Miller, JR, Koren, S, & Sutton, G 2010, 'Assembly algorithms for next-generation sequencing data', *Genomics*, vol. 95, no. 6, pp. 1-31.
- Million-Weaver, S & Camps, M 2014, 'Mechanisms of plasmid segregation: have multicopy plasmids been overlooked?', *Plasmid*, vol. 0, pp. 27-36.
- Mirmonsef, P & Spear, GT 2014, 'The barrier to HIV transmission provided by genital tract *Lactobacillus* colonization', *Am. J. Reprod. Immunol.*, vol. 71, pp. 531-536.
- Mirmonsef, P, Hotton, AL, Gilbert, D, Burgad, D, Landan, A, Webber, KM, Cohen, M, Ravel, J & Spear, GT 2014, 'Free glycogen in vaginal fluid is associated with *Lactobacillus* colonisation and low vaginal pH', *PLoS ONE*,

- vol. 9, no. 7, pp. 1-11.
- Motevaseli, E, Shirzad, M, Raoofian, R, Hasheminasab, SM, Hatami, M, Dianatpour, M & Modarressi, MH 2013, 'Differences in vaginal lactobacilli composition of Iranian healthy and bacterial vaginosis infected women: A comparative analysis of their cytotoxic effects with commercial vaginal probiotics', *Iran Red Crescent Med. J.*, vol. 15, no. 3, pp. 199-206.
- Mullis, KB 1990, 'The unusual origin of the Polymerase Chain Reaction'. *Sci. Am.*, pp. 56-65
- Navarro, E, Serrano-Heras, G, Castano, MJ & Solera, J 2015, 'Real-time PCR detection chemistry', *Clin. Chim. Acta*, vol. 439, pp. 231-250.
- Ngugi, BM, Hemmerling, A, Bukusi, EA, Kikuvi, G, Gikunju, J, Shiboski, S, Fredricks, DN & Cohen, CR 2011, 'Effects of BV-associated bacteria and sexual intercourse on vaginal colonization with the probiotic *Lactobacillus crispatus* CTV-05', *Sex. Transm. Dis.*, vol. 38, no. 11, pp. 1020-1027.
- O'Hanlon, DE, Moench, TR & Cone, RA 2011, 'In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide', *BMC Infect. Dis.*, vol. 11, no. 200, pp. 1-8.
- O'Hanlon, DE, Moench, TR & Cone, RA 2013, 'Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota', *PLoS ONE*, vol. 8, no. 11, pp. 1-8.
- Ohtani, N, Tomita, M & Itaya, M 2016, 'Curing the mega plasmid pTT27 from *Thermus thermophilus* HB27 and maintaining exogenous plasmids in the plasmid-free strain', *Appl. Environ. Microbiol.*, vol. 82, no. 5, pp. 1537-1548.
- O'Sullivan, DJ & Klaenhammer, TR 1993, 'Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus* and *Lactobacillus* spp.', *Appl. Environ. Microbiol.*, vol. 59, no. 8, pp. 2730-2733.
- Ojala, T, Kankainen, M, Castro, J, Cerca, N, Edelman, S, Wikstrom, BW, Paulin, L., Holm, L & Auvinen, P 2014, 'Comparative genomics of *Lactobacillus crispatus* suggest novel mechanisms of exclusion of *Gardnerella vaginalis*', *BMC Genomics*, vol. 15, pp. 1-21.
- Oyola, SO, Otto, TD, Gu, Y, Maslen, G, Manske, M, Campino, S, Turner, DJ,

- MacInnis, B, Kwiatkowski, DP, Swerdlow, HP & Quail, MA 2012, 'Optimizing Illumina next-generation sequencing library preparation for extremely at-biased genomes', *BMC Genomics*, vol. 13, no. 1, pp. 1-12.
- Pabinger, S, Rodiger, S, Kriegner, A, Vierlinger, K & Weinhausel, A 2014, 'A survey of tools for the analysis quantitative PCR (qPCR) data', *Biomol. Detect. Quantif.*, vol. 1, no. 1, pp. 23-33.
- Pan, Q, Lin, Z, Jinchuan, L, Tian, C, Wei, C, Guangke, W & Jianhua, Y 2011, 'Plasmid characterization of pLP18, a novel cryptic plasmid of *Lactobacillus plantarum* PC518 isolated from Chinese pickle', *Plasmid*, vol. 65, pp. 204-209.
- Papadimitriou, K, Anastasiou, R, Maistrou, E, Plakas, T, Papandreou, NC, Hamodrakas, SJ, Ferreira, S, Supply, P, Renault, P, Pot, B & Tsakalidou, E 2015, 'Acquisition through horizontal gene transfer of plasmid pSMA198 by *Streptococcus macedonicus* ACA-DC198 points towards the dairy origin of the species', *PLoS ONE*, vol. 10, no. 1, pp. 1-17.
- Pavlova, SI, Kilic, AO, Kilic, SS, So, JS, Nader-Macias, ME, Simoes, JA & Tao, L 2002, 'Genetic diversity of vaginal lactobacilli from women in different countries based on 16S rRNA gene sequence', *J. Appl. Microbiol.*, vol. 92, no. 3, pp. 451-459.
- Pendharkar, S, Magopane, T, Larsson, PG, de Bruyn, G, Gray, GE, Hammarström, L & Marcotte, H 2013, 'Identification and characterisation of vaginal lactobacilli from South African women', *BMC Infect. Dis.*, vol. 13, no. 43, pp. 1-7.
- Petrova, MI, van den Broek, M, Balzarini, J, Vanderleyden, J & Lebeer, S 2013, 'Vaginal microbiota and its role in HIV transmission and infection', *FEMS Microbiol. Rev.*, vol. 37, no. 5, pp. 762-792.
- Pfaffi, MW 2001, 'A new mathematical model for relative quantification in real-time RT-PCR', *Nucleic Acids Res.*, vol. 29, no. 9, pp. 1-15.
- Pflughoef, KJ & Versalovic, J 2012, 'Human microbiome in health and disease', *Annu. Rev. Pathol. Mech.*, vol. 7, no. 99, pp. 99-122.
- Plotka, M, Wozniak, M & Kaczorowski, T 2017, 'Quantification of plasmid copy number with single colour droplet digital PCR', *PLoS ONE*, vol. 12, no. 1, pp.

1-17.

- Pujol, C, Chedin, F, Ehrlich, SD & Janniere, L 1998, 'Inhibition of naturally occurring rolling circle replicon in derivatives of the theta-replicating plasmid pIP501', *Mol. Microbiol.*, vol. 29, pp. 709-718.
- Pyles, RB, Vincent, KL, Baum, MM, Elsom, B, Miller, AL, Maxwell, C, Eaves-Pyles, TD, Li, G, Popov, VL, Nusbaum, RJ & Ferguson, MR 2014, 'Cultivated vaginal microbiomes alter HIV-1 infection and antiretroviral efficacy in colonized epithelial multilayer cultures', *PLoS ONE*, vol. 9, no. 3, pp. 1-12.
- Radonic', A, Thulke, S, Mackay, IM, Landt, O, Siegert, W & Nitsche, A 2004, 'Guidelines to reference gene selection for quantitative real time PCR' *Biochem. Biophys. Res. Commun.*, vol. 313, no. 4, pp. 856-862.
- Ravel, J, Gajer, P, Abdo, Z, Schneider, GM, Koenig, SSK, McCulle, SL, S. Karlebach, S, Gorle, R, Russell, J, Tacket, CO, Brotman, RM, Davis, CC, Ault, K, Peralta, L & Forney, LJ 2011, 'Vaginal microbiome of reproductive-age women', *Proc. Natl. Acad. Sci.*, vol. 108, no. 1, pp. 4680-4687.
- Sanger, F, Nicklen, S & Coulson, AR 1977, 'DNA sequencing with chain-terminating inhibitors', *Proc. Natl. Acad. Sci.*, vol. 74, no. 12, pp. 5463-5467.
- Siedow, A, Cramm, R, Siddiqui, RA & Friedrich, B 1999, 'A mega plasmid-borne anaerobic ribonucleotide reductase in *Alcaligenes eutrophus* H16', *J. Bacteriol.*, vol. 181, no. 16, pp. 4919-4928.
- Sha, BE, Zariffard, MR, Wang, QJ, Chen, HY, Bremer, J, Cohen, MH & Spear, GT 2005, 'Female genital-tract HIV load correlates inversely with *Lactobacillus* species but positively with bacterial vaginosis and mycoplasma hominis', *J. Infect. Dis.*, vol. 191, no. 1, pp. 25-32.
- Skulj, M, Okrslar, V, Jalen, S, Jevsevar, S, Slanc, P, Strukelj, B & Menart, V 2008, 'Microbial cell factories', *BMC*, vol. 7, no. 6, pp. 1-12.
- Sørvig, E, Mathiesen, G, Naterstad, K, Eijsink, VG & Axelsson, L 2005, 'High-level, inducible gene expression in *Lactobacillus sakei* and *Lactobacillus plantarum* using versatile expression vectors', *Microbiol.*, vol. 151, no. 7, pp. 2439-2449.
- Sota, M, Yano, H, Hughes, J, Daughdrill, GW, Abdo, Z, Forney, LJ & Top, EM

- 2010, 'Shifts in host range of a promiscuous plasmid through parallel evolution of its replication initiation protein', *Microb. Ecol.*, vol. 4, no. 12, pp. 1568-1580.
- Stadhouders, R, Pas, SD, Anber, J, Voermans, J, Mes, THM & Schutten, M 2010, 'The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay', *J. Mol. Diagn.*, vol. 12, no. 1, pp. 109-117.
- Stanger, FV, Harms, A, Dehio, C & Schirmer, T 2016, 'Crystal structure of the *Escherichia coli* Fic toxin like protein in complex with its cognate antitoxin', *PLoS ONE*, vol. 11, no. 9, pp. 1-21.
- Stubbe, J 1990, 'Ribonucleotide reductases: amazing and confusing', *J. Biol. Chem.*, vol. 265, no. 10, pp. 5329-5332.
- Sun, E, Ren, F, Liu, S, Ge, S, Zhang, M, Guo, H, Jiang, L, Zhang, H & Zhao, L 2015, 'Complete genome sequence of *Lactobacillus salivarius* Ren, a probiotic strain with anti-tumor activity', *J. Biotechnol.*, vol. 210, pp. 57-58.
- Suzuki, H, Yano, H, Brown, CJ & Top, EM 2010, 'Predicting plasmid promiscuity based on genomic signature', *J. Bacteriol.*, vol. 192, no. 22, pp. 6045-6055.
- Svara, F & Rankin, DJ 2011, 'The evolution of plasmid-carried antibiotic resistance', *BMC Evol. Biol.*, vol. 11, pp. 1-10.
- Tannous, A, Patel, N, Tamura, T & Herbert, DN 2015, 'Reglucosylation by UDP-glucose: glycoprotein glucosyltransferase 1 delays glycoprotein secretion but not degradation', *Mol. Biol. Cell*, vol. 26, no. 3, pp. 390-405.
- Tamura, K, Stecher, G, Peterson, D, Filipowski, A & Kumar, S 2013, 'MEGA6: Molecular Evolutionary Genetics Analysis version 6.0', *Mol. Biol. Evol.*, vol. 30, no. 12, pp. 2725-2729.
- Tamrakar, R, Yamada, T, Furuta, I, Cho, K, Morikawa, M, Yamada, H, Sakuragi, N & Minakami, H 2007, 'Association between *Lactobacillus* species and bacterial vaginosis scores in pregnant Japanese women', *BMC Infect. Dis.*, vol. 7, no. 128, pp. 1-8.
- Teixeira, GS, Carvalho, FP, Arantes, RME, Nunes, AC, Moreira, JLS, Mendonca, M, Almeida, RB, Farias, LM, Carvalho, MAR & Nicoli, JR 2012,

- ‘Characterisation of *Lactobacillus* and *Gardnerella vaginalis* from women with or without bacterial vaginosis and their relationship in gnotobiotic mice’, *J. Med. Microbiol.*, vol. 61, no. 8, pp. 1074-1081.
- The Human Microbiome Project Consortium 2012, ‘Structure, function and diversity of the healthy human microbiome’, *Nature*, vol. 486, no. 1, pp. 207-214.
- Torrents, E 2014, ‘Ribonucleotide reductase: essential enzymes for bacterial life’, *Front. Cell Infect. Microbiol.*, vol. 4, no. 52, pp. 4984-4985.
- Trembizki, E, Buckley, C, Lawrence, A, Lahra, M & Whiley, D 2014, ‘Characterization of a novel *Neisseria gonorrhoeae* penicillinase producing plasmid isolated in Australia in 2012’, *Antimicrob. Agents Chemother.*, vol. 58, no. 8, pp. 4984-4985.
- Turgeon, N & Moineau, S 2001, ‘Isolation and characterisation of *Streptococcus thermophilus* derived from a native plasmid closely related to the pMV158 family’, *Plasmid*, vol. 45, pp. 171-183.
- Turgeon, N, Frenetle, M & Moineau, S 2004, ‘Characterisation of a theta replicating plasmid from *Streptococcus thermophilus*’, *Plasmid*, vol. 51, pp. 24-36.
- Turovskiy, Y, Noll, KS & Chikindas, ML 2011, ‘The aetiology of bacterial vaginosis’, *J. Appl. Microbiol.*, vol. 110, no. 5, pp. 1105-1128.
- Unterholzner, SJ, Poppenberger, B & Rozhon, W 2013, ‘Toxin–antitoxin systems biology, identification, and application’, *Mob. Genet. elements*, vol. 3, no. 5, pp. 1-13.
- Valasek, MA & Repa, JJ 2005, ‘The power of real time PCR’, *Adv. Physiol. Educ.*, vol. 29, no. 3, pp. 151-159.
- Vallor, AC, Antonio, MAD, Hawes, SE & Hiller, SL 2001, ‘Factors associated with acquisition of, or persistent colonization by, vaginal lactobacilli: role of hydrogen peroxide production’, *J. Infect. Dis.*, vol. 184, no. 11, pp. 1431-1436.
- van Kranenburg, R, Golic, N, Bongers, R, Leer, RJ, de Vos, WM, Siezen, RJ & Kleerebezem, M 2005, ‘Functional analysis of three plasmids from *Lactobacillus plantarum*’, *Appl. Environ. Microbiol.*, vol. 71, no. 3, pp. 1223-

- Verstraelen, H, Verhelst, R, Claeys, G, De Backer, E, Temmerman, M & Vaneechoutte, M 2009, 'Longitudinal analysis of the vaginal microflora in pregnancy suggests that *L. crispatus* promotes the stability of the normal vaginal microflora and that *L. gasseri* and/or *L. iners* are more conducive to the occurrence of abnormal vaginal microflora', *BMC Microbiol.*, vol. 9, no. 116, pp. 1-10.
- Vitali, B, Pugliese, C, Biagi, E, Candela, M, Turrone, S, Bellen, G, Donders, GGG & Brigidi, P 2007, 'Dynamics of vaginal bacterial communities in women developing bacterial vaginosis, candidiasis, or no infection, analysed by PCR-Denaturing Gradient Gel Electrophoresis and Real-Time PCR', *Appl. Environ. Microbiol.*, vol. 73, no. 18, pp. 5731- 5741.
- Voldborg, BR, Damstrup, L, Thomsen, MS & Poulsen, HS 1997, 'Epidermal growth factor receptor (EGFR) and EGFR mutations, functions and possible role in clinical trials', *Annals of oncol.*, vol. 8, no. 12, pp. 1197-1206.
- Warinner, C, Speller, C, Collins, MJ & Lewis Jr., CM 2014, 'Ancient human microbiomes', *J. Hum. Evol.*, vol. 79, pp. 1-12.
- Weber, L, DeForce, E & Apprill, A 2017, 'Optimization of DNA extraction for advancing coral microbiota investigations', *Microbiome*, vol. 5, no. 18, pp. 1-14.
- Williams, LE, Detter, C, Barry, K, Lapidus, A, & Summers, AO 2006, 'Facile recovery of individual high-molecular-weight, low-copy-number natural plasmids for genomic sequencing', *Appl. Environ. Microbiol.*, vol. 72, no. 7, pp. 4899-4906.
- Winceslaus, SJ & Calver, G 1996, 'Recurrent bacterial vaginosis-an old approach to a new problem', *Int. J. STD AIDS*, vol. 7, no. 4, pp. 284-287.
- Xi, X, Fan, J, Hou, Y, Gu, J, Shen, W & Li, Z 2013, 'Characterisation of three cryptic plasmids from *Lactobacillus plantarum* G63 that was isolated from Chinese pickle', *Plasmid*, vol. 70, pp. 321-328.
- Yamamoto, HS, Xu, Q & Fichoroval, RN 2013, 'Homeostatic properties of *Lactobacillus jensenii* engineered as a live vaginal anti-HIV microbicide', *BMC*

- Microbiol.*, vol. 13, no. 4, pp. 1-14.
- Yang, E, Fan, L, Jiang, Y, Doucette, C & Fillmore, S 2012, 'Antimicrobial activity of bacteriocin-producing lactic acid bacteria isolated from cheeses and yogurts', *Appl. Microbiol. Biotechnol. Express*, vol. 2, pp. 1-12.
- Yates, CM, Shaw, DJ, Roe1, AJ, Woolhouse, MEJ & Amyes, SGB 2006, 'Enhancement of bacterial competitive fitness by apramycin resistance plasmids from non-pathogenic *Escherichia coli*', *Biol. Letters*, vol. 2, no. 3, pp. 463-465.
- Yuan, Y, Norris, C, Xu, Y, Tsui, KW, Ji, Y & Liang, H 2012, 'BM-Map: an efficient software package for accurately allocating multireads of RNA-sequencing data', *BMC Genomics*, vol. 13, no. 8, pp. 1-5.
- Zhai, Z, Hao, Y, Yin, S, Luan, S, Zhang, L, Zhao, L, Chen, D, Wang, O & Luo, Y 2009, 'Characterisation of a novel rolling circle replication pYS18 from *Lactobacillus sakei* YS18', *Plasmid*, vol. 62, pp. 30-34.
- Zhong, C, Peng, D, Ye, W, Chai, L, Qi, J, Yu, Z, Ruan, L & Sun, M 2011, 'Determination of plasmid copy number reveals the total plasmid DNA amount is greater than the chromosomal DNA amount in *Bacillus thuringiensis* YBT-1520', *PLoS ONE*, vol. 6, no.1, pp. 1-8.
- Zhou, X, Brown, CJ, Abdo, Z, Davis, CC, Hansmann, MA, Joyce, P, Foster, JA & Forney, LJ 2007, 'Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women', *Microb. Ecol.*, vol. 1, no. 2, pp. 121-133.
- Zhou, X, Hansmann, MA, Davis, CC, Suzuki, H, Brown, CJ, Schutte, U, Pierson, JD & Forney, LJ 2010, 'The vaginal bacterial communities of Japanese women resemble those of women in other racial groups', *FEMS Immunol. Med. Microbiol.*, vol. 58, no. 2, pp. 169-181.

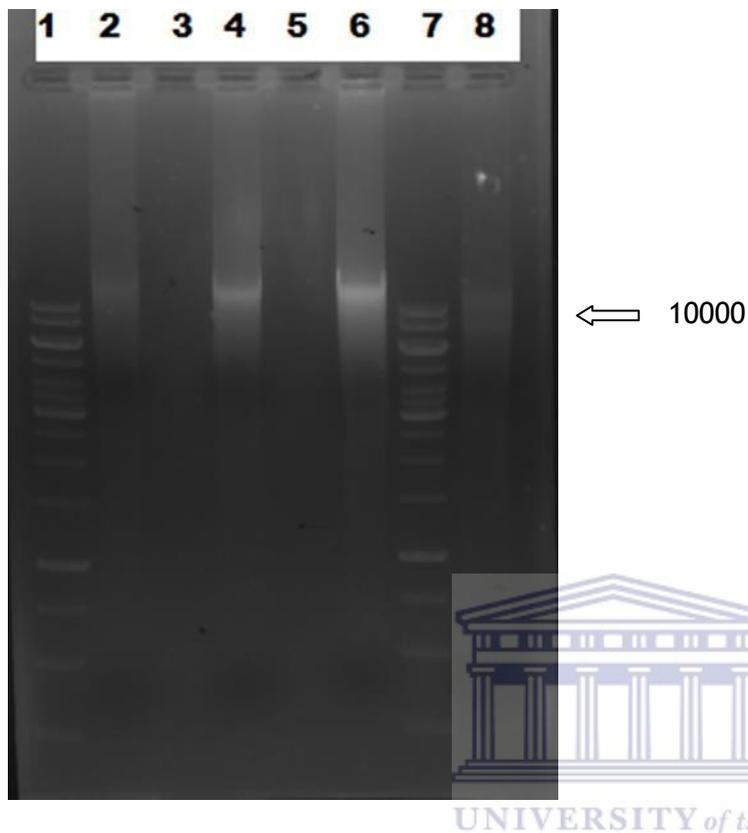
Appendices: Buffers and stock solutions

Appendix 1: Buffers, Stock solutions

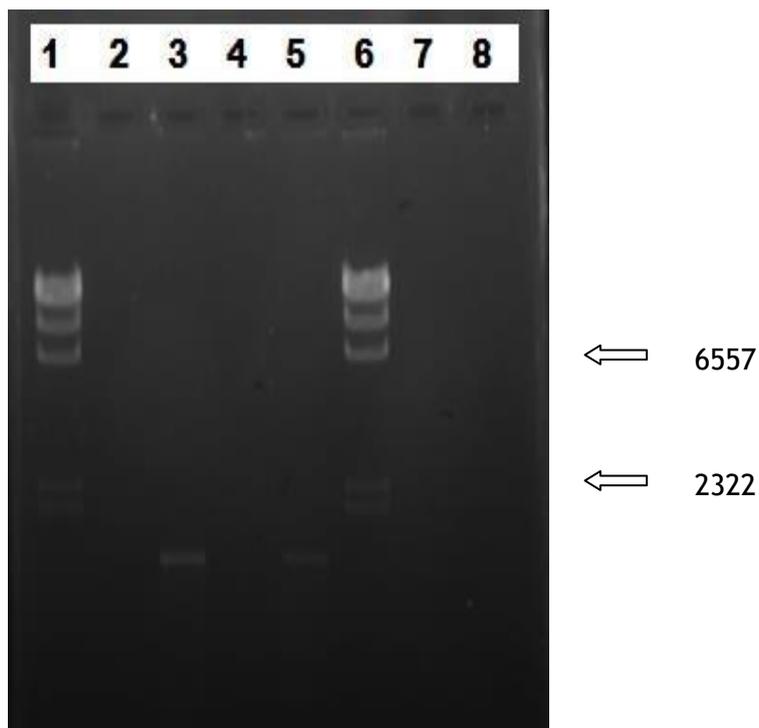
Buffer	Components	pH
10X Orange G loading buffer	60% Glycerol 0.25% (w/v) Orange G	
50X TAE	2 M Tris base 10 mM Glacial acetic acid 0.5 M EDTA	8
1X TAE	40 mM Tris acetate 1 mM EDTA 0.2 mM glacial acetic acid	8
1X TE	10 mM Tris-HCl (pH 8.0) 1 mM EDTA (pH 8.0)	8
Lysis buffer	50 mM Tris-HCl (pH 7.6) 50 mM NaCl 5% SDS (pH 8.0)	
DNA loading buffer	20 % [v/v] Glycerol 1 % [w/v] SDS 0.1 M EDTA 0.25 % Bromophenol blue	
Storage Buffer	10 mM Tris pH 8 0.01% Tween-20	8

UNIVERSITY of the
WESTERN CAPE

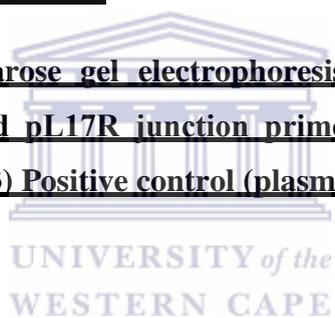
Supplementary Figures and Tables

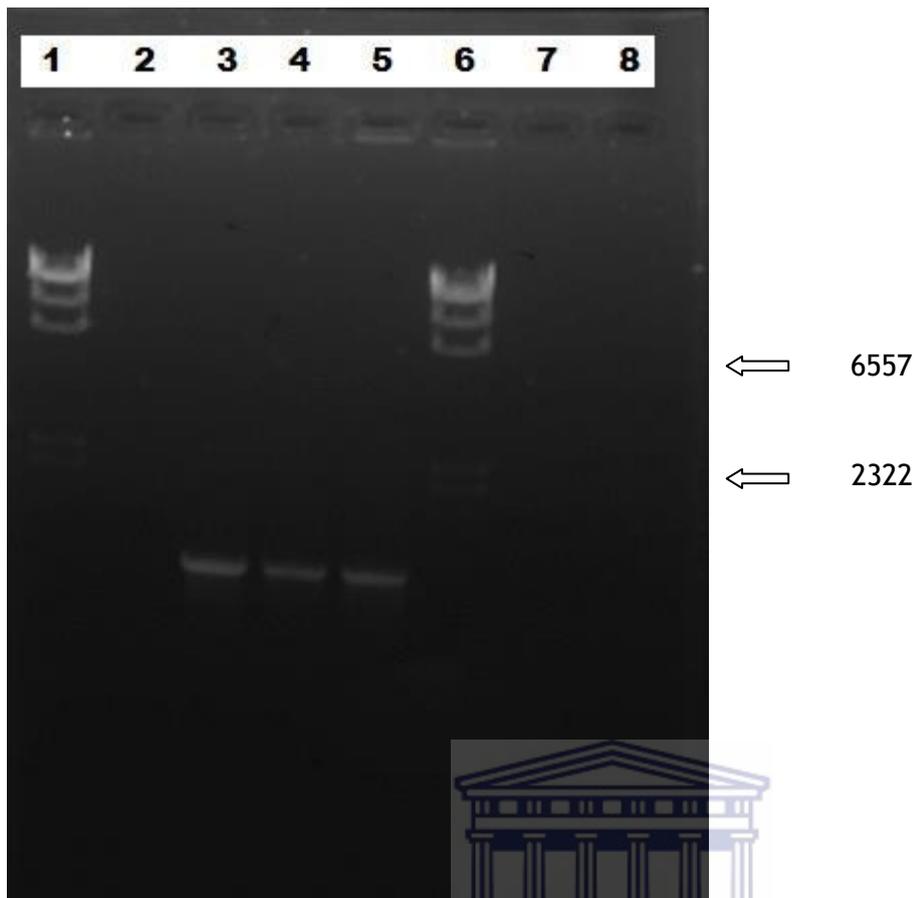


Supplementary Figure 1: Agarose gel electrophoresis on a 1% gel of plasmid DNA extracted from *L. crispatus* pLc17 using the Zyppy kit. (2) 0.5ml of pellet harvested and lysozyme pre- treatment;(4) 1ml of pellet harvested and lysozyme pre-treatment;(6) 2ml of pellet harvested and lysozyme pre-treatment; (8) 2.5ml of pellet harvested and lysozyme pre- treatment; (1,7) 100bp Marker (NEB)



Supplementary Figure 2: Agarose gel electrophoresis on a 1% gel of PCR DNA amplification using pL17F and pL17R junction primers. (1.6) Molecular marker: λ HindIII. (2) Negative control. (3) Positive control (plasmid DNA). (5) Sample L1.





Supplementary Figure 3: Agarose gel electrophoresis on a 1% gel of PCR DNA amplification using pL17F and pL17R junction primers. (1,6) Molecular marker: λ HindIII. (2) Negative control. (3) Positive control (plasmid DNA). (4) Sample L6. (5) Sample L7

Supplementary Table 1: Relative PCN and mean DNA concentration of pLc4, pLc17 and reference gene (EGFR)

Sample Name	Sample type	Sample label	Target/Ref ratio (Normalized ratios)		DNA concentration		DNA concentration (Ref gene)
			pLc4	pLc17	pLc4	pLc17	
CDC16-071-2015	Healthy	SH2	4,82E+02	1,19E+01	9,33E+02	2,30E+01	1,94E+00
GP-09-VDS-106	Healthy	SH3	2,67E+02	2,37E+00	1,22E+03	1,09E+01	4,59E+00
GP-09-VDS-123	BV	SD4	4,50E+00	4,52E+00	2,24E+01	2,25E+01	4,98E+00
GP-09-VDS-126	Healthy	SH5	5,17E+00	4,64E+00	1,54E+01	1,38E+01	2,98E+00
GP-09-VDS-125	BV	SD6	1,47E+01	2,31E+01	1,72E+01	2,71E+01	1,17E+00
GP-09-VDS-131	Healthy	SH10	7,58E+00	2,99E+01	1,66E+01	1,32E+01	2,19E+00
GP-09-VDS-099	Healthy	SH11	3,61E+01	1,71E+01	1,59E+01	1,82E+01	4,41E-01
GP-09-VDS-102	Healthy	SH12	1,46E+01	1,76E+00	1,55E+01	9,86E+00	1,06E+00
GP-09-VDS-109	BV	SD13	2,82E+00	2,08E+00	1,58E+01	1,11E+01	5,62E+00
GP-09-VDS-141	BV	SD14	3,32E+00	1,93E+01	1,77E+01	2,90E+01	5,34E+00
CDC16-071-2021	BV	SD15	9,98E+00	1,00E+01	1,49E+01	1,48E+01	1,50E+00
CDC16-071-2020	Healthy	SH16	9,98E+00	4,82E+03	1,47E+01	5,69E+03	1,47E+00
CDC16-071-2018	Healthy	SH17	1,30E+01	4,17E+00	1,53E+01	1,38E+01	1,18E+00
CDC16-071-2002	BV	SD18	5,24E+00	1,00E+00	1,74E+01	7,93E+00	3,31E+00
GP-09-VDS-104	Healthy	SH19	2,08E+00	5,71E+00	1,64E+01	1,27E+01	7,90E+00
GP-09-VDS-132	BV	SD20	8,07E+00	2,28E+01	1,80E+01	5,75E+01	2,23E+00
CDC16-071-1500	Healthy	SH21	7,56E+00	1,41E+00	1,91E+01	1,38E+01	2,52E+00
GP-09-VDS-136	Healthy	SH22	1,79E+00	6,35E+00	1,75E+01	1,11E+01	9,76E+00
CDC16-071-1487	BV	SD23	1,08E+01	2,65E+00	1,89E+01	1,95E+01	1,75E+00
CDC16-071-1482	Healthy	SH24	1,83E+00	1,82E+01	1,35E+01	1,19E+01	7,36E+00
GP-09-	BV	SD25	2,28E+01	4,07E+00	1,49E+01	1,43E+01	6,55E-01

VDS-138							
GP-09-VDS-128	Healthy	SH26	4,65E+00	1,56E+01	1,63E+01	1,22E+02	3,51E+00
CDC16-071-2014	Healthy	SH30	8,38E+00	4,79E+00	1,70E+01	3,04E+01	2,03E+00
GP-09-VDS-068	BV	SD31	1,68E+00	2,04E+01	1,07E+01	2,92E+01	6,33E+00
GP-09-VDS-103	BV	SD39	3,52E+00	3,22E+00	5,05E+00	4,63E+00	1,44E+00
GP-09-VDS-097	Healthy	SH40	2,18E+03	2,22E+03	1,97E+03	2,67E+03	1,44E+00
GP-09-VDS-078	BV	SD41	4,30E+00	3,06E+01	5,17E+00	2,36E+02	9,02E-01
CDC16-071-1488	Healthy	SH42	1,57E+00	5,64E+00	2,25E+00	2,13E+01	1,20E+00

- SH=Healthy patient sample
- SD=BV-infected patient sample

Reference gene (EGFR)

