Comparison of plasmids from clinical
*Lactobacillus* strains

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: [Signature]

UNIVERSITY of the WESTERN CAPE
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.
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 jensennii* 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 jensennii* 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 jensennii*, 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

http://etd.uwc.ac.za
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 Lactobacillus 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

http://etd.uwc.ac.za
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 L. crispatus 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 access 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 69
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

http://etd.uwc.ac.za
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/µl 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

http://etd.uwc.ac.za
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 pneumonia, 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. Lactobacillius 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 Gardnerella 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) (Mirmonef & Spear 2014; Vallor et al. 2001). There are several possible mechanisms which allow lactobacilli to dominate the vaginal mucosa (Mirmonef & 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) (Mirmonef & 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 aerugonosa, 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 Winceslaus 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 (Winceslaus & Calver 1996). $H_2O_2$ Production by lactobacilli has also been shown to kill HIV in vitro (Mironsef & 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

<table>
<thead>
<tr>
<th>Aim of the study</th>
<th>Country</th>
<th>Number of participants</th>
<th>Findings</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>The study was performed to establish if there was a significant difference between the vaginal lactobacilli of African and Caucasian (American and European) women</td>
<td>Nigeria</td>
<td>241 healthy women</td>
<td>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-H₂O₂ producing <em>L. iners</em> then <em>L. gasseri</em> followed by <em>L. plantarum</em>, <em>L. stutoryeus</em>, <em>L. crispatus</em> and <em>L. rhamnosus</em>. These findings are in agreement with other vaginal microbiome studies conducted in Canada, the United States and Sweden.</td>
<td>Anukam et al. (2006)</td>
</tr>
<tr>
<td>The study was performed to establish if there was a significant difference between the vaginal lactobacilli of African and Caucasian (American) women</td>
<td>North America</td>
<td>144 healthy black and white women</td>
<td>It was found that <em>L. iners</em>, <em>L. crispatus</em>, <em>L. gasseri</em> and <em>L. jensenni</em> 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 <em>Atopobuim</em> (LAB) and <em>Clostridia</em>. In addition, white women are more likely to have more than one type of <em>Lactobacillus</em> species.</td>
<td>Zhou et al. (2007)</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Country</th>
<th>Sample Description</th>
<th>Findings</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>73 healthy Japanese women</td>
<td>It was found that <em>L. iners</em>, <em>L. crispatus</em>, <em>L. gasseri</em> and <em>L. jensenni</em> were the dominant lactobacilli in Japanese women.</td>
<td>Zhou <em>et al.</em> (2010)</td>
</tr>
<tr>
<td>North America</td>
<td>396 asymptomatic women</td>
<td>It was found that of the majority of white and Asian women were dominated by <em>L. iners</em>, <em>L. crispatus</em>, <em>L. gasseri</em> and <em>L. jensenni</em>, 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.</td>
<td>Ravel <em>et al.</em> (2011)</td>
</tr>
<tr>
<td>South Africa</td>
<td>40 African women</td>
<td>The majority of the subjects were colonised by lactobacilli with <em>L. crispatus</em> being the dominant species, which is in accordance with the many studies conducted on European populations</td>
<td>Pendharkar <em>et al.</em> (2013)</td>
</tr>
<tr>
<td>Argentina, Brazil, China, India, South Korea, Turkey and the United States</td>
<td>400 women</td>
<td>It was found that the majority of isolates were <em>L. crispatus</em>, <em>L. jensenni</em> and <em>L. gasseri</em></td>
<td>Pavalova <em>et al.</em> (2002)</td>
</tr>
<tr>
<td>South Africa</td>
<td>144 women</td>
<td>It was found that <em>L. crispatus</em>, <em>L. jensenni</em>, <em>L. iners</em> and <em>L. gasseri</em> were the dominant lactobacilli in the healthy vagina of the South African women examined. Furthermore, they found a reduction in lactobacilli associated with</td>
<td>Damelin <em>et al.</em> (2011)</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Participants</td>
<td>Findings</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>This study was conducted to examine the vaginal composition of premenopausal South American women</td>
<td>Mexico</td>
<td>108 Hispanic women</td>
<td>It was found that L. crispatus, L. gasseri and L. jensenii were the dominant Lactobacillus species</td>
</tr>
<tr>
<td>This study was conducted to examine the vaginal composition of sexually active women</td>
<td>North America</td>
<td>302 sexually active women</td>
<td>It was found that L. crispatus, L. jensenii, L. 1086V (later identified as L. iners) and L. gasseri were the dominant species. In this study they also examined the H₂O₂ production of the isolated strains and it was found that 95% of L. crispatus isolates were capable of producing H₂O₂ and 94% of L. jensenii were capable of producing H₂O₂</td>
</tr>
<tr>
<td>This study was conducted to examine the vaginal isolates of premenopausal women using PCR-DGGE and real time PCR</td>
<td>Belgium</td>
<td>26 premenopausal women</td>
<td>In this study they found that there was a reduction in the number H₂O₂ producing lactobacilli (L. acidophilus, L. gasseri and L. vaginalis) and an increase in the number of non-H₂O₂ producing lactobacilli (L iners) with women suffering from Candida infection, which suggests that Candida infection may be linked to a lack of H₂O₂ producing lactobacilli</td>
</tr>
<tr>
<td>This study was conducted to examine the vaginal isolates from non-pregnant women</td>
<td>North America</td>
<td>101 non pregnant women</td>
<td>They found that L. jensenii and L. crispatus were the dominant lactobacilli. They also found that H₂O₂ producing lactobacilli tend to colonise sexually active women for longer</td>
</tr>
<tr>
<td>Study Description</td>
<td>Location</td>
<td>Participants</td>
<td>Findings</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>----------</td>
<td>--------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>The antagonism potential of lactobacilli was examined</td>
<td>Kenya</td>
<td>107 BV infected women</td>
<td>They found that the greatest antagonism was related to acidity, as a result of the lactic acid and not hydrogen peroxide production.</td>
</tr>
<tr>
<td>This study was conducted to evaluate whether there was a significant shift in the</td>
<td>China</td>
<td>107 women</td>
<td>They found that L. crispatus and L. gasseri was reduced significantly in number in BV patients. Whereas L. iners was significantly high in BV infected patients. L. jensenii was found not to be affected by the shift.</td>
</tr>
<tr>
<td>microbial population of the 4 dominant lactobacilli between healthy and BV infected patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This study was conducted to examine the vaginal composition of Iranian women</td>
<td>Iran</td>
<td>178 women</td>
<td>They found that L. crispatus, L. gasseri, L. iners, L. jensenii, L. acidophilus and L. rhamnosus were dominant in healthy women. L. crispatus, L. gasseri and L. jensenii were reduced in BV infected women. However, L. iners was found to be higher in BV infected women.</td>
</tr>
<tr>
<td>This study was conducted to investigate the vaginal microflora associated with both BV infected and healthy women</td>
<td>China</td>
<td>108 BV infected and healthy women</td>
<td>They found an increase in bacterial diversity in BV women with increased numbers of lactobacilli present in healthy women. L. iners was found at high levels in healthy women followed by L. crispatus and L. jensenii. Atopobium was strongly associated with BV. Several new species of bacterium were also identified to be linked to BV infection.</td>
</tr>
</tbody>
</table>

http://etd.uwc.ac.za
This study was conducted to investigate the vaginal microflora of healthy pregnant women. 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 *lactobacillus* absent microbiome after the 3rd trimester. It was also found that long term lactobacilli dominance was dependent on what *lactobacillus* was dominant in the vaginal samples.

<table>
<thead>
<tr>
<th>Location</th>
<th>Sample Size</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>100 women</td>
<td>Healthy pregnant white women</td>
<td>Verstralen <em>et al.</em> (2009)</td>
</tr>
</tbody>
</table>
The production of \( \text{H}_2\text{O}_2 \) 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 \( \text{H}_2\text{O}_2 \) production. One such study was done by O’Hanlon et al. (2011) where the antagonistic effects of \( \text{H}_2\text{O}_2 \) and lactic acid were compared under conditions which mimic the human vagina. It was found that lactic acid and not \( \text{H}_2\text{O}_2 \) 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 \textit{L. crispatus}, \textit{L. jensenii}, \textit{L. iners} and \textit{L. gasserii} 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 microbiocide 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 \textit{Atopobium}, \textit{Corynebacterium}, \textit{Anaerococcus}, \textit{Peptoniphilus}, \textit{Prevotella}, \textit{Gardnerella}, \textit{Sneathia}, \textit{Eggerthella}, \textit{Mobiluncus} and \textit{Finegoldia} 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 \textit{Atopobium}, \textit{Streptococcus}, \textit{Staphylococcus}, \textit{Megasphaera} and \textit{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-aerophillic bacteria such as *Gardenerella vaginalis*, *Mycoplasma homines*, *Ureaplasma urealyticin*, *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 nucleotum* 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 microbiocide

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 antiviral 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. gasseri*, *L. casei*, *L. plantarium*, *L. minutus* (subsequently renamed to *Atopobium minutus*), *L. plantarium*, *L. fermentum*, *L. celpobiosus*, *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. gasseri* 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, Fluorecent 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, Selenomonas 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 has 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

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

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

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Peptone</td>
<td>10 g</td>
</tr>
<tr>
<td>Meat Extract</td>
<td>10 g</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>5 g</td>
</tr>
<tr>
<td>Dextrose</td>
<td>20 g</td>
</tr>
<tr>
<td>Sodium Acetate</td>
<td>5 g</td>
</tr>
<tr>
<td>Tween 80</td>
<td>1 g</td>
</tr>
<tr>
<td>di-Potassium Hydrogen Phosphate</td>
<td>2 g</td>
</tr>
<tr>
<td>tri-Ammonium Citrate</td>
<td>2 g</td>
</tr>
<tr>
<td>Magnesium Sulphate</td>
<td>0.1 g</td>
</tr>
<tr>
<td>Manganese Sulphate</td>
<td>0.05 g</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>10 g</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>5 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>10 g</td>
</tr>
<tr>
<td>dH₂O up to</td>
<td>1000 ml</td>
</tr>
</tbody>
</table>

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 OD600nm >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 NCBInr 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

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Sequence (5’-3’)</th>
<th>Amplification cycle</th>
<th>Specificity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S-F1</td>
<td>AGAGTTTGATCITGGCTCAG</td>
<td>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.</td>
<td>Bacterial universal 16S rRNA</td>
<td>Babalola et al. (2009)</td>
</tr>
<tr>
<td>16S-R5</td>
<td>ACGGITACCTTTGTTGCACTTT</td>
<td>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).</td>
<td>pLc17 junction designed to bind to the “start” and “end” of the assembled contig</td>
<td>This study</td>
</tr>
<tr>
<td>pL17-F</td>
<td>TTAGAAATGCAGTTGCGTGG</td>
<td>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).</td>
<td>pLc4 junction designed to bind to the “start” and “end” of the assembled contig</td>
<td>This study</td>
</tr>
<tr>
<td>pL17-R</td>
<td>AATTGCAAAAGTGCTGCGG</td>
<td>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).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pL4-F</td>
<td>AGCAGTTTAGGCCATCGATA</td>
<td>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.</td>
<td>Human epidermal growth factor receptor (EGFR) Gene</td>
<td>Lin et al. (2010)</td>
</tr>
<tr>
<td>pL4-R</td>
<td>ACTCCGCTTGTCCGCTGGA</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR E18 F</td>
<td>CTGGCACTGCTTTCCAGCAT</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR E18 R</td>
<td>GCTTGCAAGGACTCTGGGCT</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qPL17F</td>
<td>AGGTACTCCTGCTACGCTCA</td>
<td>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.</td>
<td>pLc17, with forward and reverse primers target ORF 8</td>
<td>This study</td>
</tr>
<tr>
<td>qPL17R</td>
<td>ACAACTGCCAGAAACAACAG</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qPL4F</td>
<td>GATTTCCGCTCTTGCTGAAG</td>
<td>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.</td>
<td>pLc4, with forward and reverse primers target ORF 1</td>
<td>This study</td>
</tr>
<tr>
<td>qPL4R</td>
<td>ACACCTGCCATCTTTTTGT</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://etd.uwc.ac.za
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 HindIII or PstI 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:

\[ \text{number of copies} = \frac{(\text{amount} \times 6.022 \times 10^{23})}{(\text{length} \times 1 \times 10^9 \times 6.02) \times \frac{\text{number (bp) \times \text{ng/g} \times \text{g/mole of bp)}}{\text{ng x number/mole}}} \]

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).
Switzerland) and diluted from 10-10000 copies/µl 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 µl reaction containing 10 µM 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 H\(_2\)O, 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 µl 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/µl. 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 µl reaction containing 10 µM 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) accessed 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 seeing 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. delbruckii 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 *HindIII* (IMBM Marker)](http://etd.uwc.ac.za)
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>
<thead>
<tr>
<th>Plasmid</th>
<th>Total Reads</th>
<th>Average Coverage</th>
<th>G+C %</th>
<th>Size (bp)</th>
<th>Average read length</th>
</tr>
</thead>
<tbody>
<tr>
<td>pLc4</td>
<td>179549</td>
<td>10172.10</td>
<td>36%</td>
<td>4224 bp</td>
<td>242.60</td>
</tr>
<tr>
<td>pLc17</td>
<td>63640</td>
<td>835.59</td>
<td>36%</td>
<td>16663 bp</td>
<td>220.65</td>
</tr>
</tbody>
</table>

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*. 

http://etd.uwc.ac.za
Table 3.2: Identification of pLc4 and pLc17 in individual samples from healthy and BV-infected individuals

<table>
<thead>
<tr>
<th>Sample</th>
<th>Host</th>
<th>Patient condition</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>L. crispatus</td>
<td>Healthy patient</td>
<td>pLc17</td>
</tr>
<tr>
<td>L2</td>
<td>L. crispatus</td>
<td>BV infected patient</td>
<td>pLc17</td>
</tr>
<tr>
<td>L3</td>
<td>L. jensenii</td>
<td>Healthy patient</td>
<td>pLc4</td>
</tr>
<tr>
<td>L4</td>
<td>L. jensenii</td>
<td>BV infected patient</td>
<td>pLc4</td>
</tr>
<tr>
<td>L5</td>
<td>L. crispatus</td>
<td>Healthy patient</td>
<td>pLc17</td>
</tr>
<tr>
<td>L6</td>
<td>L. crispatus</td>
<td>BV infected patient</td>
<td>pLc17</td>
</tr>
<tr>
<td>L7</td>
<td>L. crispatus</td>
<td>Healthy patient</td>
<td>pLc17</td>
</tr>
<tr>
<td>L8</td>
<td>L. crispatus</td>
<td>BV infected patient</td>
<td>pLc4</td>
</tr>
</tbody>
</table>

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 a 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.
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.

<table>
<thead>
<tr>
<th>ORF number</th>
<th>Size in amino acids</th>
<th>Start and end positions bp</th>
<th>Selected BLAST hits and comments; accession number; (length of protein on database in aa)</th>
<th>% Identity/Similarity (over number of aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>385</td>
<td>+(81-1238)</td>
<td>Hypothetical protein <em>Streptococcus thermophilus</em> WP_011161226 (368) Replication initiation factor <em>Lactococcus lactis</em> WP_012881354 (368)</td>
<td>44/64 (166/243)</td>
</tr>
<tr>
<td>2</td>
<td>364</td>
<td>+(1235-2329)</td>
<td>Hypothetical protein <em>Streptococcus suis</em> WP_032498644 (448) Plasmid partitioning protein <em>Lactococcus lactis</em> WP_014573491.1 (448)</td>
<td>32/52 (108/180)</td>
</tr>
<tr>
<td>3</td>
<td>92</td>
<td>+(2277-2553)</td>
<td>Hypothetical protein <em>Streptococcus suis</em> WP_032498132 (129)</td>
<td>39/60 (28/43)</td>
</tr>
<tr>
<td>4</td>
<td>120</td>
<td>+(2630-2995)</td>
<td>Hypothetical protein WP_051234001 (341)</td>
<td>43/61 (18/26)</td>
</tr>
<tr>
<td>5</td>
<td>137</td>
<td>+(2967-3383)</td>
<td>Hypothetical protein <em>Globicattella</em> sp. HMPREF 2811-00515 (577)</td>
<td>29/47 (34/54)</td>
</tr>
<tr>
<td>6</td>
<td>123</td>
<td>+(3829-4200)</td>
<td>Hypothetical protein <em>Lactococcus lactis</em> spSK11B YP_003329500.1 (121)</td>
<td>34/56 (40/65)</td>
</tr>
</tbody>
</table>

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 trough 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 pWVO2, 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 pWVO2 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>
<thead>
<tr>
<th>Table 3.4: Inverted repeat sequences for pLc4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>pLc4</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

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).

![Evolutionary tree of plasmid proteins](http://etd.uwc.ac.za)  

**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 pLc17.

<table>
<thead>
<tr>
<th>ORF Number</th>
<th>Size in amino acids</th>
<th>Start and end positions bp</th>
<th>Selected BLAST hits and comments; accession number; (length of protein on database in aa)</th>
<th>% Identity/Similarity (over number of aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>107</td>
<td>+(10-334)</td>
<td>Hypothetical protein Lactobacillus helveticus WP_003625520 (80)</td>
<td>61/74 (33/40)</td>
</tr>
<tr>
<td>2</td>
<td>109</td>
<td>+(332-664)</td>
<td>Hypothetical protein Lactobacillus crispatus WP_005723210 (110)</td>
<td>100/100 (110/110)</td>
</tr>
<tr>
<td>3</td>
<td>207</td>
<td>+(678-1302)</td>
<td>Hypothetical protein HMPREF9250_02419, partial Lactobacillus crispatus FB049-03 EKB59528 (87)</td>
<td>100/100 (87/87)</td>
</tr>
<tr>
<td>4</td>
<td>343</td>
<td>+(1252-2286)</td>
<td>Transposase Lactobacillus crispatus WP_005728429 (344)</td>
<td>100/100 (344/344)</td>
</tr>
<tr>
<td>5</td>
<td>107</td>
<td>+(2204-2528)</td>
<td>Multidrug resistance protein MdtG, partial Lactobacillus crispatus SJ-3C-US] KFL92490 (165)</td>
<td>93/95 (38/39)</td>
</tr>
<tr>
<td>6</td>
<td>409</td>
<td>-(3795-2563)</td>
<td>Transposase Lactobacillus crispatus WP_035458603 (424)</td>
<td>100/100 (410/410)</td>
</tr>
<tr>
<td>7</td>
<td>243</td>
<td>-(4513-3779)</td>
<td>Resolvase Bacteria WP_005728427 (221)</td>
<td>100/100 (221/221)</td>
</tr>
<tr>
<td>8</td>
<td>237</td>
<td>-(5236-4520)</td>
<td>FIC-like protein Lactobacillus crispatus WP_05728426 (238)</td>
<td>100/100 (238/238)</td>
</tr>
<tr>
<td>9</td>
<td>123</td>
<td>-(5629-5255)</td>
<td>Hypothetical protein LBKG_02277 Lactobacillus crispatus CTV-05 WP_005728425 (58)</td>
<td>98/100 (57-58)</td>
</tr>
<tr>
<td>10</td>
<td>369</td>
<td>-(6747-5635)</td>
<td>Putative plasmid replication initiation protein Lactobacillus crispatus WP_005728424 (370)</td>
<td>100/100 (370/370)</td>
</tr>
<tr>
<td>11</td>
<td>193</td>
<td>+(7340-6757)</td>
<td>Replication initiator protein A, partial Lactobacillus parafarraginis WP_08213692 (203)</td>
<td>79/90 (26/33)</td>
</tr>
<tr>
<td>12</td>
<td>260</td>
<td>+(7326-8111)</td>
<td>ATPase Lactobacillus WP_005728422 (261)</td>
<td>100/100 (250/250)</td>
</tr>
<tr>
<td>13</td>
<td>173</td>
<td>+(8105-8629)</td>
<td>Hypothetical protein Lactobacillus crispatus WP_005728419 (71)</td>
<td>100/100 (71/71)</td>
</tr>
<tr>
<td>14</td>
<td>165</td>
<td>+(8622-9122)</td>
<td>Transposase Lactobacillus helveticus WP_05728418 (164) Hypothetical protein LBKG_02273 Lactobacillus crispatus CTV-05 WP_05729299 (139)</td>
<td>99/98 (143/145)</td>
</tr>
<tr>
<td>15</td>
<td>138</td>
<td>+(9164-9583)</td>
<td>DDE endonuclease Lactobacillus crispatus WP_05729299 (139)</td>
<td>100/100 (139/139)</td>
</tr>
<tr>
<td>16</td>
<td>106</td>
<td>+(9497-9820)</td>
<td>Hypothetical protein HMPREF0548_0424 Lactobacillus ultunensis DSM 16047 EEJ72728 (145)</td>
<td>91/93 (30/31)</td>
</tr>
<tr>
<td>17</td>
<td>200</td>
<td>+(10061-10666)</td>
<td>Resolvase Bacteria WP_005728417 (201)</td>
<td>100/100 (201/201)</td>
</tr>
<tr>
<td>18</td>
<td>428</td>
<td>+(10668-11957)</td>
<td>Resolvase <em>Lactobacillus kefiranofaciens</em> WP_013855360.1 (197) Hypothetical protein <em>Lactobacillus crispatus</em> WP_005728416 (56)</td>
<td>54/70 (27/35) 98/100 (55/56)</td>
</tr>
<tr>
<td>19</td>
<td>342</td>
<td>+(11958-12986)</td>
<td>Ribonucleoside-diphosphatereductase, beta subunit <em>Lactobacillus crispatus</em> SJ-3C-US KFL92593 (342)</td>
<td>100/100 (342/342)</td>
</tr>
<tr>
<td>20</td>
<td>140</td>
<td>+(12967-13392)</td>
<td>Ribonuclease reductase <em>Lactobacillus</em> WP_003648207 (141)</td>
<td>100/100 (141/141)</td>
</tr>
<tr>
<td>21</td>
<td>728</td>
<td>+(13364-15553)</td>
<td>Ribonucleoside-diphosphatereductase, alpha subunit <em>Lactobacillus crispatus</em> SJ-3C-US KFL92591 (729)</td>
<td>99/100 (728/729)</td>
</tr>
<tr>
<td>22</td>
<td>101</td>
<td>+(15554-15860)</td>
<td>Hypothetical protein <em>Lactobacillus crispatus</em> WP_005728412 (59)</td>
<td>98/100 (58/59)</td>
</tr>
<tr>
<td>23</td>
<td>273</td>
<td>+(15860-16684)</td>
<td>Replication initiation protein <em>Lactobacillus crispatus</em> WP_005728411 (262)</td>
<td>100/100 (262/262)</td>
</tr>
</tbody>
</table>

**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 1b RNRS’s, it is speculated that the pLc17-encoded RNRS’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)GNGRXXR. 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 Erwina 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
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 portioning 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
Table 3.6: Direct and inverted repeats for pLc17

<table>
<thead>
<tr>
<th>Repeat location</th>
<th>Type of repeat</th>
<th>Repeat sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>11751 / 11805</td>
<td>Direct</td>
<td>TCAACGAAGTCATGCAGACTTTCTCGATTTTCGAAATTCCGCTCGTAG</td>
</tr>
<tr>
<td>11703 / 11769 / 11823</td>
<td>Direct</td>
<td>TCAACGAAGTCATGCAGACTTTCTCGAAATTCCGCTCGTAG</td>
</tr>
<tr>
<td>12486 / 12510</td>
<td>Invert</td>
<td>AATGGACTAAGTAG</td>
</tr>
<tr>
<td>12776 / 12813</td>
<td>Invert</td>
<td>ATGATTTATGATTAATC</td>
</tr>
<tr>
<td>12774 / 12789</td>
<td>Invert</td>
<td>ATATGATTTATGATT</td>
</tr>
<tr>
<td>12785 / 12813</td>
<td>Direct</td>
<td>GATTAATCATAAATCAT</td>
</tr>
<tr>
<td>13894 / 13921</td>
<td>Invert</td>
<td>TGATTGTGTGATTATGATAAAAGATT</td>
</tr>
<tr>
<td>4466-4497</td>
<td>Invert</td>
<td>AACCCCGAAATTCGGACAAAGAAATTTCCGGGT</td>
</tr>
</tbody>
</table>

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).
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).

Figure 3.9 tBlastx comparison of pLc17 with pH1N and pR2. A) The figure shows a comparison of closely related proteins between pLc17 and pH1N 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.
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 (Cq) is the cycle number at which the fluorescence signal of the
reaction crosses the threshold (Laplante 2006).

The Cq 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 DH5α. The PCN of the plasmid pBR322 in DH5α 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 (Gorzelniak *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. A10 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 (R²) 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

<table>
<thead>
<tr>
<th>DNA Type</th>
<th>Spiking combination</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human DNA</td>
<td>Human</td>
<td>Human+E. coli</td>
<td>Human+pLc4</td>
<td>Human+pLc17</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>0.05ng/ul</td>
<td>0.05ng/ul</td>
<td>0.05ng/ul</td>
<td>0.05ng/ul</td>
<td>0.05ng/ul</td>
</tr>
<tr>
<td>E. coli DNA</td>
<td>50 ng/ul</td>
<td></td>
<td></td>
<td>50 ng/ul</td>
<td></td>
</tr>
<tr>
<td>pLc4</td>
<td></td>
<td>500 copies/ul</td>
<td></td>
<td>500 copies/ul</td>
<td></td>
</tr>
<tr>
<td>pLc17</td>
<td></td>
<td></td>
<td>500 copies/ul</td>
<td></td>
<td>500 copies/ul</td>
</tr>
</tbody>
</table>

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 access 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.**

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

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/µl 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.

61
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/µl was used as the positive control, indicated by the red curve. The Cq 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/µl were used as the positive control, indicated by the red curve. The Cq 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.
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/µl 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/µl 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/µl-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)**

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample type</th>
<th>Sample label</th>
<th>Target/Ref ratio (Normalized ratios)</th>
<th>DNA concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>pLc4</td>
<td>pLc17</td>
</tr>
<tr>
<td>CDC16-071-2015</td>
<td>Healthy</td>
<td>SH2</td>
<td>4.82E+02</td>
<td>1.19E+01</td>
</tr>
<tr>
<td>GP-09-VDS-106</td>
<td>Healthy</td>
<td>SH3</td>
<td>2.67E+02</td>
<td>2.37E+01</td>
</tr>
<tr>
<td>GP-09-VDS-123</td>
<td>BV</td>
<td>SD4</td>
<td>4.50E+00</td>
<td>4.52E+00</td>
</tr>
<tr>
<td>GP-09-VDS-126</td>
<td>Healthy</td>
<td>SH5</td>
<td>1.51E+00</td>
<td>1.64E+00</td>
</tr>
<tr>
<td>GP-09-VDS-125</td>
<td>BV</td>
<td>SD6</td>
<td>1.47E+01</td>
<td>2.31E+01</td>
</tr>
<tr>
<td>GP-09-VDS-131</td>
<td>Healthy</td>
<td>SH10</td>
<td>7.58E+00</td>
<td>2.99E+01</td>
</tr>
<tr>
<td>GP-09-VDS-099</td>
<td>Healthy</td>
<td>SH11</td>
<td>3.61E+01</td>
<td>1.71E+01</td>
</tr>
<tr>
<td>GP-09-VDS-102</td>
<td>Healthy</td>
<td>SH12</td>
<td>1.46E+01</td>
<td>1.76E+00</td>
</tr>
<tr>
<td>GP-09-VDS-109</td>
<td>BV</td>
<td>SD13</td>
<td>2.82E+00</td>
<td>2.08E+00</td>
</tr>
<tr>
<td>GP-09-VDS-141</td>
<td>BV</td>
<td>SD14</td>
<td>3.32E+00</td>
<td>1.93E+01</td>
</tr>
<tr>
<td>CDC16-071-2021</td>
<td>BV</td>
<td>SD15</td>
<td>9.98E+00</td>
<td>1.00E+01</td>
</tr>
<tr>
<td>CDC16-071-2020</td>
<td>Healthy</td>
<td>SH16</td>
<td>9.98E+00</td>
<td>4.82E+03</td>
</tr>
<tr>
<td>Sample ID</td>
<td>Group</td>
<td>Patient ID</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>CDC16-071-2018</td>
<td>Healthy</td>
<td>SH17</td>
<td>1.30E+01</td>
<td>4.17E+00</td>
</tr>
<tr>
<td>CDC16-071-2002</td>
<td>BV</td>
<td>SD18</td>
<td>5.24E+00</td>
<td>1.00E+00</td>
</tr>
<tr>
<td>GP-09-VDS-104</td>
<td>Healthy</td>
<td>SH19</td>
<td>2.08E+00</td>
<td>5.71E+00</td>
</tr>
<tr>
<td>GP-09-VDS-132</td>
<td>BV</td>
<td>SD20</td>
<td>8.07E+00</td>
<td>2.28E+01</td>
</tr>
<tr>
<td>CDC16-071-1500</td>
<td>Healthy</td>
<td>SH21</td>
<td>7.56E+00</td>
<td>1.41E+00</td>
</tr>
<tr>
<td>GP-09-VDS-136</td>
<td>Healthy</td>
<td>SH22</td>
<td>1.79E+00</td>
<td>6.35E+00</td>
</tr>
<tr>
<td>CDC16-071-1487</td>
<td>BV</td>
<td>SD23</td>
<td>1.08E+01</td>
<td>2.65E+00</td>
</tr>
<tr>
<td>CDC16-071-1482</td>
<td>Healthy</td>
<td>SH24</td>
<td>1.83E+00</td>
<td>1.82E+01</td>
</tr>
<tr>
<td>GP-09-VDS-138</td>
<td>BV</td>
<td>SD25</td>
<td>2.28E+01</td>
<td>4.07E+00</td>
</tr>
<tr>
<td>GP-09-VDS-128</td>
<td>Healthy</td>
<td>SH26</td>
<td>1.46E+00</td>
<td>1.56E+01</td>
</tr>
<tr>
<td>CDC16-071-2014</td>
<td>Healthy</td>
<td>SH30</td>
<td>1.88E+00</td>
<td>4.79E+00</td>
</tr>
<tr>
<td>GP-09-VDS-068</td>
<td>BV</td>
<td>SD31</td>
<td>1.68E+00</td>
<td>2.04E+01</td>
</tr>
<tr>
<td>GP-09-VDS-103</td>
<td>BV</td>
<td>SD39</td>
<td>3.52E+00</td>
<td>3.22E+00</td>
</tr>
<tr>
<td>GP-09-VDS-097</td>
<td>Healthy</td>
<td>SH40</td>
<td>2.18E+03</td>
<td>2.22E+03</td>
</tr>
<tr>
<td>GP-09-VDS-078</td>
<td>BV</td>
<td>SD41</td>
<td>4.30E+00</td>
<td>3.06E+01</td>
</tr>
<tr>
<td>CDC16-071-1488</td>
<td>Healthy</td>
<td>SH42</td>
<td>1.57E+00</td>
<td>5.64E+00</td>
</tr>
</tbody>
</table>

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

http://etd.uwc.ac.za
Zhong et al. (2011) describes PCN as either low (1-10 copies), medium (11-20 copies), or high (>20 copies). The low Cq 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**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Ratio</th>
<th>Standard error (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pLc4</td>
<td>2.803</td>
<td>±1.63201</td>
</tr>
<tr>
<td>pLc17</td>
<td>1.693</td>
<td>±1.0915</td>
</tr>
</tbody>
</table>
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, \( \text{H}_2\text{O}_2 \) 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. gasserii* 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

http://etd.uwc.ac.za
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 being 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 portioning 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. jensennii. 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
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. jensenii* 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/µl 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 being 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 (Gorzelniak et al. 2001). Furthermore, as per the MIQE guideline the selection of an appropriate reference gene based on stable expression in various tissue 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

http://etd.uwc.ac.za
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


http://etd.uwc.ac.za


10, no. 11, pp.1063-1067.


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.


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.


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.


Henson, J, Tischler, G & Ning, Z 2012, ‘Next-generation sequencing and large genome assemblies’, *Pharmacogenomics J.*, vol. 13, no. 8, pp. 901-915.


vol. 160, no. 4, pp. 267-282.


Laplante, PA 2006, ‘Real time PCR’, Taylor & Francis Group, 270 Madison Avenue New York, N Y 10016, USA.


Lu, C, Nakayasu, ES & Zhang, LQ 2016, ‘Identification of Fic-1 as an enzyme that


Martinez-Pena, MD, Castro-Escarpulli, G & Aguilera-Areveola, MG 2013, ‘*Lactobacillus* species isolated from vaginal secretions of healthy and bacterial


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.


Million-Weaver, S & Camps, M 2014, ‘Mechanisms of plasmid segregation: have multicopy plasmids been overlooked?’, *Plasmid*, vol. 0, pp. 27-36.


Oyola, SO, Otto, TD, Gu, Y, Maslen, G, Manske, M, Campino, S, Turner, DJ,


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.


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.


Yamamoto, HS, Xu, Q & Fichoroval, RN 2013, ‘Homeostatic properties of Lactobacillus jensenii engineered as a live vaginal anti-HIV microbicide’, BMC


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.


## Appendices: Buffers and stock solutions

### Appendix 1: Buffers, Stock solutions

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Components</th>
<th>pH</th>
</tr>
</thead>
</table>
| 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  |
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)**

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample type</th>
<th>Sample label</th>
<th>Target/Ref ratio (Normalized ratios)</th>
<th>DNA concentration (Ref gene)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC16-071-2015</td>
<td>Healthy</td>
<td>SH2</td>
<td>pLc4: 4.82E+02, pLc17: 1.19E+01</td>
<td>pLc4: 9.33E+02, pLc17: 2.30E+01</td>
</tr>
<tr>
<td>GP-09-VDS-106</td>
<td>Healthy</td>
<td>SH3</td>
<td>pLc4: 2.67E+02, pLc17: 2.37E+00</td>
<td>pLc4: 1.22E+03, pLc17: 1.09E+01</td>
</tr>
<tr>
<td>GP-09-VDS-123</td>
<td>BV</td>
<td>SD4</td>
<td>pLc4: 4.50E+00, pLc17: 4.52E+00</td>
<td>pLc4: 2.24E+01, pLc17: 2.25E+01</td>
</tr>
<tr>
<td>GP-09-VDS-126</td>
<td>Healthy</td>
<td>SH5</td>
<td>pLc4: 5.17E+00, pLc17: 4.64E+00</td>
<td>pLc4: 1.54E+01, pLc17: 1.38E+01</td>
</tr>
<tr>
<td>GP-09-VDS-125</td>
<td>BV</td>
<td>SD6</td>
<td>pLc4: 1.47E+01, pLc17: 2.31E+01</td>
<td>pLc4: 1.72E+01, pLc17: 2.71E+01</td>
</tr>
<tr>
<td>GP-09-VDS-131</td>
<td>Healthy</td>
<td>SH10</td>
<td>pLc4: 7.58E+00, pLc17: 2.99E+01</td>
<td>pLc4: 1.66E+01, pLc17: 1.32E+01</td>
</tr>
<tr>
<td>GP-09-VDS-099</td>
<td>Healthy</td>
<td>SH11</td>
<td>pLc4: 3.61E+01, pLc17: 1.71E+01</td>
<td>pLc4: 1.59E+01, pLc17: 1.82E+01</td>
</tr>
<tr>
<td>GP-09-VDS-102</td>
<td>Healthy</td>
<td>SH12</td>
<td>pLc4: 1.36E+01, pLc17: 1.76E+00</td>
<td>pLc4: 1.55E+01, pLc17: 9.86E+00</td>
</tr>
<tr>
<td>GP-09-VDS-109</td>
<td>BV</td>
<td>SD13</td>
<td>pLc4: 2.82E+00, pLc17: 2.08E+00</td>
<td>pLc4: 1.58E+01, pLc17: 1.11E+01</td>
</tr>
<tr>
<td>GP-09-VDS-141</td>
<td>BV</td>
<td>SD14</td>
<td>pLc4: 9.32E+00, pLc17: 1.93E+01</td>
<td>pLc4: 1.77E+01, pLc17: 2.90E+01</td>
</tr>
<tr>
<td>CDC16-071-2021</td>
<td>BV</td>
<td>SD15</td>
<td>pLc4: 9.98E+00, pLc17: 1.00E+01</td>
<td>pLc4: 1.49E+01, pLc17: 1.48E+01</td>
</tr>
<tr>
<td>CDC16-071-2020</td>
<td>Healthy</td>
<td>SH16</td>
<td>pLc4: 9.98E+00, pLc17: 4.82E+03</td>
<td>pLc4: 1.47E+01, pLc17: 5.69E+03</td>
</tr>
<tr>
<td>CDC16-071-2018</td>
<td>Healthy</td>
<td>SH17</td>
<td>pLc4: 1.30E+01, pLc17: 4.17E+00</td>
<td>pLc4: 1.53E+01, pLc17: 1.38E+01</td>
</tr>
<tr>
<td>CDC16-071-2002</td>
<td>BV</td>
<td>SD18</td>
<td>pLc4: 5.24E+00, pLc17: 1.00E+00</td>
<td>pLc4: 1.74E+01, pLc17: 7.93E+00</td>
</tr>
<tr>
<td>GP-09-VDS-104</td>
<td>Healthy</td>
<td>SH19</td>
<td>pLc4: 2.08E+00, pLc17: 5.71E+00</td>
<td>pLc4: 1.64E+01, pLc17: 1.27E+01</td>
</tr>
<tr>
<td>GP-09-VDS-132</td>
<td>BV</td>
<td>SD20</td>
<td>pLc4: 8.07E+00, pLc17: 2.28E+01</td>
<td>pLc4: 1.80E+01, pLc17: 5.75E+01</td>
</tr>
<tr>
<td>CDC16-071-1500</td>
<td>Healthy</td>
<td>SH21</td>
<td>pLc4: 7.56E+00, pLc17: 1.41E+00</td>
<td>pLc4: 1.91E+01, pLc17: 1.38E+01</td>
</tr>
<tr>
<td>GP-09-VDS-136</td>
<td>Healthy</td>
<td>SH22</td>
<td>pLc4: 1.79E+00, pLc17: 6.35E+00</td>
<td>pLc4: 1.75E+01, pLc17: 1.11E+01</td>
</tr>
<tr>
<td>CDC16-071-1487</td>
<td>BV</td>
<td>SD23</td>
<td>pLc4: 1.08E+01, pLc17: 2.65E+00</td>
<td>pLc4: 1.89E+01, pLc17: 1.95E+01</td>
</tr>
<tr>
<td>CDC16-071-1482</td>
<td>Healthy</td>
<td>SH24</td>
<td>pLc4: 1.83E+00, pLc17: 1.82E+01</td>
<td>pLc4: 1.35E+01, pLc17: 1.19E+01</td>
</tr>
<tr>
<td>GP-09-</td>
<td>BV</td>
<td>SD25</td>
<td>pLc4: 2.28E+01, pLc17: 4.07E+00</td>
<td>pLc4: 1.49E+01, pLc17: 1.43E+01</td>
</tr>
</tbody>
</table>

http://etd.uwc.ac.za
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Health</th>
<th>SH26</th>
<th>4.65E+00</th>
<th>1.56E+01</th>
<th>1.63E+01</th>
<th>1.22E+02</th>
<th>3.51E+00</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP-09-VDS-128</td>
<td>Healthy</td>
<td>SH26</td>
<td>4.65E+00</td>
<td>1.56E+01</td>
<td>1.63E+01</td>
<td>1.22E+02</td>
<td>3.51E+00</td>
<td></td>
</tr>
<tr>
<td>CDC16-071-2014</td>
<td>Healthy</td>
<td>SH30</td>
<td>8.38E+00</td>
<td>4.79E+00</td>
<td>1.70E+01</td>
<td>3.04E+01</td>
<td>2.03E+00</td>
<td></td>
</tr>
<tr>
<td>GP-09-VDS-068</td>
<td>BV</td>
<td>SD31</td>
<td>1.68E+00</td>
<td>2.04E+01</td>
<td>1.07E+01</td>
<td>2.92E+01</td>
<td>6.33E+00</td>
<td></td>
</tr>
<tr>
<td>GP-09-VDS-103</td>
<td>BV</td>
<td>SD39</td>
<td>3.52E+00</td>
<td>3.22E+00</td>
<td>5.05E+00</td>
<td>4.63E+00</td>
<td>1.44E+00</td>
<td></td>
</tr>
<tr>
<td>GP-09-VDS-097</td>
<td>Healthy</td>
<td>SH40</td>
<td>2.18E+03</td>
<td>2.22E+03</td>
<td>1.97E+03</td>
<td>2.67E+03</td>
<td>1.44E+00</td>
<td></td>
</tr>
<tr>
<td>GP-09-VDS-078</td>
<td>BV</td>
<td>SD41</td>
<td>4.30E+00</td>
<td>3.06E+01</td>
<td>5.17E+00</td>
<td>2.36E+02</td>
<td>9.02E-01</td>
<td></td>
</tr>
<tr>
<td>CDC16-071-1488</td>
<td>Healthy</td>
<td>SH42</td>
<td>1.57E+00</td>
<td>5.64E+00</td>
<td>2.25E+00</td>
<td>2.13E+01</td>
<td>1.20E+00</td>
<td></td>
</tr>
</tbody>
</table>

SH=Healthy patient sample
SD=BV-infected patient sample
Reference gene (EGFR)

http://etd.uwc.ac.za