The microbiological context of HIV resistance

by

John J. Schellenberg

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Medical Microbiology and Infectious Diseases

University of Manitoba

Winnipeg

ABSTRACT

Immune activation is increasingly recognized as a critical element of HIV infection and pathogenesis, causing expansion of virus founder populations at the mucosal port of entry and eventual exhaustion of cellular immune effectors. A cohort of HIV-resistant (HIV-R) commercial sex workers (CSW) in Nairobi, Kenya, have increased levels of anti-inflammatory factors in vaginal secretions and reduced peripheral immune activation ("immune quiescence"). The mucosal immune micro-environment underlying HIV susceptibility is well-known to be influenced by concurrent sexually transmitted infections, however the role of commensal microbiota is poorly characterized. Bacterial vaginosis (BV), characterized by a shift from *Lactobacillus* to *Gardnerella* and *Prevotella* as dominant members of vaginal microbiota, is a risk factor for HIV acquisition in studies worldwide. However, the etiology and ecological dynamics of BV remain enigmatic, and the mechanisms by which BV increases HIV susceptibility are not fully defined. Protective functional characteristics of *Lactobacillus* microbiota, including acid and hydrogen peroxide (H_2O_2) production, may reinforce physicochemical defences of vaginal mucus, stimulate innate epithelial defences and/or modulate activation status of HIV target cells. Therefore, the goal of this study was to determine if reduced BV and increased *Lactobacillus* colonization are the basis for resistance to HIV in this cohort. Vaginal specimens from a group of 242 CSW were examined, including microscopic diagnosis of BV, culture-based functional analyses and phylogenetic profiling by ultra-deep sequencing. HIV-R individuals were just as likely to have BV compared to other HIV-negative (HIV-N) individuals, and no more likely to be colonized with acid- or H_2O_2-

producing bacteria, however two BV-related phylotypes identified by deep sequencing were significantly more likely to be observed in HIV-N individuals ($p=0.0002$ and $p=0.006$). HIV+ individuals were significantly more likely than HIV– individuals to have *E. coli* detected by deep sequencing ($p<0.0001$) and less likely to have *Lactobacillus crispatus* ($p=0.0006$). A coherent set of differences in culture-based and culture-independent characteristics were observed in individuals with BV diagnoses compared to BV– individuals. This study has generated an unprecedented amount of information regarding the composition, structure and function of the vaginal microbiota in African CSW, fundamentally defining many aspects of BV microbiology. Elucidation of the relationship between complex microbial communities and protective mucosal responses against HIV infection should be a priority for future research.

ACKNOWLEDGEMENTS

Most importantly, I would like to express my sincerest thanks to the research participants for providing the biological samples that are the foundation of this work, especially those members of the Majengo cohort in Nairobi, Kenya, and also to the young women in Winnipeg who participated in the Good Bacteria Study.

This project has benefitted from the kind collaboration of so many: Jane Njeri Mungai, Dr. Joshua Kimani, Helen Wandaka, Dr. Rich Lester, Dr. Charles Wachihi, Jane Njoki, Jane Kamene, Edith Amahwa, Dr. Margo Lane, Heather McIntosh, Charisse Maliwanag, Leslie Slaney, Dr. Ian McLean, Sue Ramdahin, John Rutherford, Steve Wayne, Joanna Brunka, Angela Nelson, Carol Sigurdson, Lynn Gauthier, Jude Zieske, Dr. Daryl Hoban, Dr. George Zhanel, Donna Collins, Nancy Laing, Dr. Tim Dumonceaux, Dr. Alberto Severini, Vanessa Goleski, Dr. Janet Hill, Matthew Links, Dr. Morag Graham, Dr. Shaun Tyler, Geoff Peters, Brynn Kaplan, Erika Bellamy, Claude Ouellette, Shari Tyson, Dr. Gregor Reid, Dominique Lam, Dr. Ma Luo, Dr. Keith Fowke, Dr. Ruey Su, Dr. Maritim Songok, Dr. Adam Burgener, Dr. Adrienne Meyers, Dr. James Sainsbury, Dr. Shehzad Iqbal, Dr. Julius Oyugi, Dr. Sandra Kiazyk, Dr. Paul McLaren, Dr. Lyle McKinnon, Ben Liang, Jill Waruk, Meika Richmond and Catherine Card.

Members of my PhD Advisory Committee, Dr. Michelle Alfa, Dr. Denis Krause and Dr. John Wylie provided excellent guidance over the course of this project. Dr. Blake Ball and Dr. Frank Plummer have inspired this work and kept the project on track. I gratefully acknowledge a 3-year scholarship from the CIHR Doctoral Research Award program and a concurrent 4-year scholarship from the CIHR/International Centre for Infectious Diseases Training Program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS USED

AIDS	acquired immune deficiency syndrome
APC	antigen-presenting cell
ARV	antiretroviral
ATCC	American Type Culture Collection
BA	blood agar
BCP	bromocresol purple
BCU	bacterial cell-units
BHI	brain-heart infusion
bp	base pair
BV	bacterial vaginosis
CA	chocolate agar
CBA	Columbia blood agar
CCR5	chemokine (C-C motif) receptor 5
CD4	cluster of differentiation 4
CD8	cluster of differentiation 8
CFU	colony-forming units
CI	confidence interval
*cpn*60	chaperonin-60
cpnDB	chaperonin-60 database
CSW	commercial sex worker
CTL	cytotoxic T lymphocyte
CVL	cervicovaginal lavage
CXCR4	chemokine (CXC motif) receptor 4
DC	dendritic cell
DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DGGE	denaturing gradient gel electrophoresis
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxy N (A-C-G-T) triphosphate
DSMZ	Deutsche Sammlung von Mikroorganismen and Zellkulturen
EDTA	ethylenediamenetetriacetic
ESN	exposed seronegative
FGT	female genital tract
FL1	fluorescence channel 1
FL3	fluorescence channel 3
FSC	forward scatter
GalC	galactosyl cerebroside
GALT	gut-associated lymphoid tissue
gp120	glycoprotein 120
GPC	Gram-positive cocci
GPR	Gram-positive rods
GUD	genital ulcer disease
GVR	Gram-variable rods
H_2O_2	hydrogen peroxide
HEX	hexachlorofluorescein
HIV	human immunodeficiency virus
HIV-N	HIV-negative
HIV-R	HIV-resistant
HIV+	HIV-positive
HLA	human leukocyte antigen
HLM	hierarchical linear modelling
HPF	high-powered field
HPLC	high pressure liquid chromotography
HPV	human papilloma virus
HRP	horseradish peroxidase
HSV	herpes simplex virus
IFNγ	interferon gamma

IgA	immunoglobulin A
IgG	immunoglobulin G
IL	interleukin
IPTG	isopropylthiogalactoside
IRF-1	interferon regulatory factor-1
ISR	interspacer region
LAB	lactic acid bacteria
LB	Luria broth
LC	Langerhans cell
LFA-1	leukocyte function-associated antigen-1
LP	lamina propria
LPS	lipopolysaccharide
LTR	long terminal repeat
mBru	modified Brucella
MCU	mucosal cell-units
MHC	major histocompatibility complex
MIP	macrophage inflammatory protein
Mo	macrophage
MRS	Mann-Rogosa-Sharpe media for Lactobacillus
MSM	men who have sex with men
NHP	non-human primate
NK	natural killer
NNN	no near neighbour
NVB	no visible bacteria
OCT	oral contraceptive therapy
OR	odds ratio
PBMC cells	peripheral blood mononuclear
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PI	propidium iodide
RANTES	regulated upon activation, normal T-cell expressed, and secreted
rDNA	ribosomal deoxyribonucleic acid
RISA	ribosomal interspacer analysis
ROX	X-rhodamine
rRNA	ribosomal ribonucleic acid
SD	standard deviation
SIV	simian immunodeficiency virus
SLPI	secretory leukocyte protease inhibitor
SNP polymorphisms	single nucleotide
SSC	side scatter
STI	sexually transmitted infection
TGFβ	transforming growth factor beta
Th17	T helper-17
TLR	Toll-like receptor
TMB	tetramethylbenzidine
TNFα	tumor necrosis factor alpha
Treg	T-regulatory cell
TZ	transformation zone
UT	universal target
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

LIST OF PUBLICATIONS

SCHELLENBERG, J., Ball. T.B., Lane, M., Cheang, M., Plummer, F.A. 2008. Flow cytometric quantification of bacteria in vaginal swab samples self-collected by adolescents attending a gynecology clinic. J Microbiol Meth. 73(3): 216-26.

SCHELLENBERG, J.. Links, M.G., Hill, J.E., Dumonceaux, T.J., Peters, G. A., Tyler, S., Ball, T.B., Severini, A., Plummer, F.A. 2009. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. Appl Env Microbiol. 75(9):2889-2898.

Dumonceaux, T.J., SCHELLENBERG, J., Goleski, V., Hill, J.E., Jaoko, W., Kimani, J., Money, D., Ball, T.B., Plummer, F.A., Severini, A. 2009. Multiplex detection of bacteria associated with normal microbiota and with bacterial vaginosis in vaginal swabs by use of oliognucleotide-coupled fluorescent microspheres. J Clin Micro. 47(12): 4067-77.

SCHELLENBERG, J.. Links, M.G., Hill, J.E., Dumonceaux, T.J., Peters, G. A., Tyler, S., Ball, T.B., Severini, A., Plummer, F.A. 2010. Pyrosequencing of the chaperonin-60 (cpn60) amplicons as a means of determining microbial community composition. In: High-throughput sequencing: Applications to Microbiology, Kwon, Y.M. and Ricke, S.C., (volume editors). *Methods in Molecular Biology*, Walker, J. (series editor). Humana Press, NY, USA. Book chapter accepted.

SEQUENCE ACCESSIONS

A total of 248 full-length *cpn*60 universal target sequences generated during this project were deposited in cpnDB, the chaperonin database (www.cpndb.cbr.nrc.ca) and Genbank (www.ncbi.nlm.nih.gov/Genbank):

	cpnDB ID #:	Genbank accession #:
For 61 vaginal isolates:	15945-16005	FJ577558-FJ577638
For 187 library clones:	16204-16390	FJ588254-FJ588440
Search keywords:	Schellenberg, Nairobi	

A. INTRODUCTION

1. Review

AIDS and the immune system – Fatal opportunistic infections associated with what is now known as the acquired immunodeficiency syndrome (AIDS) were first described in 1981 among young, previously healthy homosexual men in the United States (74, 75). The causative agent of AIDS is the human immunodeficiency virus (HIV), discovered in 1983 (19). The global pandemic spread of a deadly, largely sexually transmitted viral infection is unprecedented in human history, with an estimated 33 million people living with HIV worldwide by the end of 2008 (214). Over 2 million AIDS-related deaths and 2.7 million new HIV infections are estimated to have occurred in that year alone (214). Despite large numbers of individuals infected through homosexual sex, blood-to-blood contact, and transmitted from mother to child, the vast majority of new infections occur through heterosexual sex, with sub-Saharan African countries continuing to bear a disproportionate share of the global disease burden.

It was recognized very early in the epidemic that individuals with AIDS had abnormally low counts of CD4+ or "helper" T lymphocytes (Th), a type of white blood cell recognized as the principal co-ordinator of the communicative network of cells and tissues that work together during normal immune function. Clearance of infectious agents depends on the interaction of innate (early, non-specific) and adaptive (late, specific) responses (22). The innate response depends on invariant pattern recognition receptors (PRR) able to detect conserved microbe-associated molecular patterns (MAMP) that are unique to micro-organisms (eg. lipopolysaccharide or LPS, flagellin, unmethylated

bacterial DNA, single-stranded viral RNA), while adaptive responses depend on the limitless variability of lymphocyte receptors able to recognize specific protein or carbohydrate structures of infectious agents (antigens), due to random rearrangement of gene segments encoding these receptors (22).

Rather than detecting infectious agents directly, Th depend on professional antigen-presenting cells (APC) called macrophages (Mo) and dendritic cells (DC), that specialize in phagocytosis and presentation of protein fragments derived from microbes in surface-bound receptors (major histocompatibility complex class II or MHC-II molecules). APC are ubiquitous throughout peripheral tissues such as skin and mucosal surfaces where most infectious agents enter the body, while Th are concentrated in lymph nodes and in lymphoid structures underlying mucosal surfaces, especially in the intestine (gut-associated lymphoid tissue or GALT). Upon detection of infectious agents by cell-surface PRR such as Toll-like receptors (TLR), APC communicate with other immune cells by secreting proteins called cytokines and chemokines. For example, stimulation of TLR results in upregulation of the cytokines interleukin (IL)-1, IL-6 and IL-8, which activate the immune response, initiate inflammation and facilitate increased migration of macrophages, other phagocytes and leukocytes to the site of infection. APC migrate via the lymphatic circulation to lymphoid tissue, where loaded MHC-II receptors interact with Th cell receptors (TCR). T cell activation results from this interaction only in the presence of costimulatory molecules. Increased IL-2 production by activated Th cells induces proliferation of antigen-specific Th cells, initiating the adaptive immune response (22).

The profile of Th effector functions depends on cytokines released by APC subsequent to TLR stimulation, which varies in type and intensity according to the

infectious agent detected (22). Naïve Th cells (Th0) may become functionally skewed to a Th1 (cell-mediated) or Th2 (antibody-mediated) phenotype. For example, IL-12 secreted by APC favours the development of Th1 cells secreting IL-2, TNFα and IFNγ, which in turn promote activation of macrophages and CD8+ cytotoxic T lymphocytes (CTL) that act against intracellular pathogens by direct killing of infected cells. APC secreting IL-4 are instrumental in development of Th2 cells secreting IL-4, activating production of antigen-specific proteins (antibodies) by specialized lymphocytes called B cells (213).

Although Th1/Th2 profiles are often described as dichotomous, designed to eliminate intra- and extra-cellular pathogens respectively, other Th subsets are increasingly recognized (48). Inducible T regulatory cells (Th3 or Tr1) develop from naïve Th under the influence of IL-10, producing immunosuppressive cytokines TGFβ and IL-10, while "natural" Tregs develop in the thymus and also exert immunosuppressive effects in order to check the inflammatory response (48). The recently described Th17 subset develops under the influence of proinflammatory cytokines IL-1 and IL-6, and secretes the proinflammatory cytokine IL-17 (155). This subset has important functions in stimulation of innate epithelial defences against extracellular pathogens and stimulates chemokines that attract phagocytes to the site of infection (155).

Although most Th in a clonally expanded pool are deleted through programmed cell-death (apoptosis) subsequent to clearance of the infectious agent, some survive and become long-lived memory cells enabling efficient elimination of the same infectious agent upon re-exposure (22). Given the critical functions of Th subsets, AIDS pathogenesis is characterized by the progressive loss of immune response coordination and memory due to HIV infection and destruction of Th and other immune cell populations.

Origins and evolution of HIV – The human immunodeficiency virus (HIV) is a T-cell lymphotropic, single-stranded RNA virus belonging to the lentivirus subfamily of retroviruses (79). The term retrovirus refers to the integration of these viruses into the host genome during productive infection. Lentiviruses, named for the typically slow course of disease that they cause, infect a wide variety of mammals including felines, ovines, bovines and primates. The lentiviruses most closely related to HIV are the simian immunodeficiency viruses (SIV) of Old World monkeys in sub-Saharan Africa.

Two major types of HIV have been described, HIV-1 and -2, with distinct origins, geographic range and pathogenesis. HIV-1 is most closely related to SIV in chimpanzees (SIVcpz) from western equatorial Africa (186). SIVcpz is a hybrid of two viruses from different monkey species and was most likely acquired by chimps via predation (15). Three main groups of HIV-1 (M, N and O) likely emerged from three separate zoonotic chimp-to-human jumps in Cameroon (186). A newly described lineage of HIV-1 (group P), recently isolated from an individual in Cameroon, is most closely related to the gorilla SIV (SIVgor) (160). HIV-2, less pathogenic than HIV-1 and largely confined to central African populations, is most closely related to SIV from the sootey mangabey (SIVsm). There are 8 types of HIV-2, again thought to reflect unique transmission events (174).

All pandemic HIV-1 is caused by clades (subgroups) belonging to Group M, which emerged in central Africa early in the twentieth century. Although other theories have been expounded at length, most researchers favour "natural" transfer of simian viruses to humans via predation, most likely during butchering of chimps and monkeys for food (186). The earliest confirmed HIV-1 positive sample was identified retrospectively using sera collected in 1959 in Leopoldville, Zaire (now Kinshasa, Democratic Republic of

Congo) (229). Recently, a lymph node sample collected in 1960 in the same city was found to contain a divergent strain of HIV-1, indicating that extensive diversification of the virus had already occurred by this time (225). When considering these two early isolates, the most recent common ancestor for HIV-1 Group M is predicted to have already existed in humans by approximately 1908 (225). Some of the earliest reported cases of fatal immunodeficiency were observed in European countries with links to central Africa, including Belgium, Norway and Denmark. Military and economic links between the newly independent Republic of Congo and Haiti likely brought HIV-1 to the Caribbean by the early 1960s, where it spread via tourism and emigration to North American capitals and beyond around 1969 (72, 166) (Figure 1.1).

While global media and health authorities were focussed on stigmatized "risk groups" in North America, including homosexuals and injection drug users, largely heterosexually-transmitted epidemics were already well established throughout sub-Saharan Africa by the time HIV was discovered. Two early reports, published back to back, described characteristics of AIDS patients in Zaire and Rwanda (158, 216), while another described alarming increases in HIV incidence among commercial sex workers (CSW) in Nairobi, Kenya (115).

In the early 1980s, a large cohort-based study of Nairobi CSW was established through collaboration between the University of Nairobi and the University of Manitoba, in order to understand and prevent sexually transmitted infections (STI) in this high-risk group. They found that the rate of HIV in the cohort had increased from 4% to over 60% between 1981 and 1985 (115). Interestingly, HIV was not observed in any samples prior to 1980, indicating the rapidity of HIV spread in this context (157).

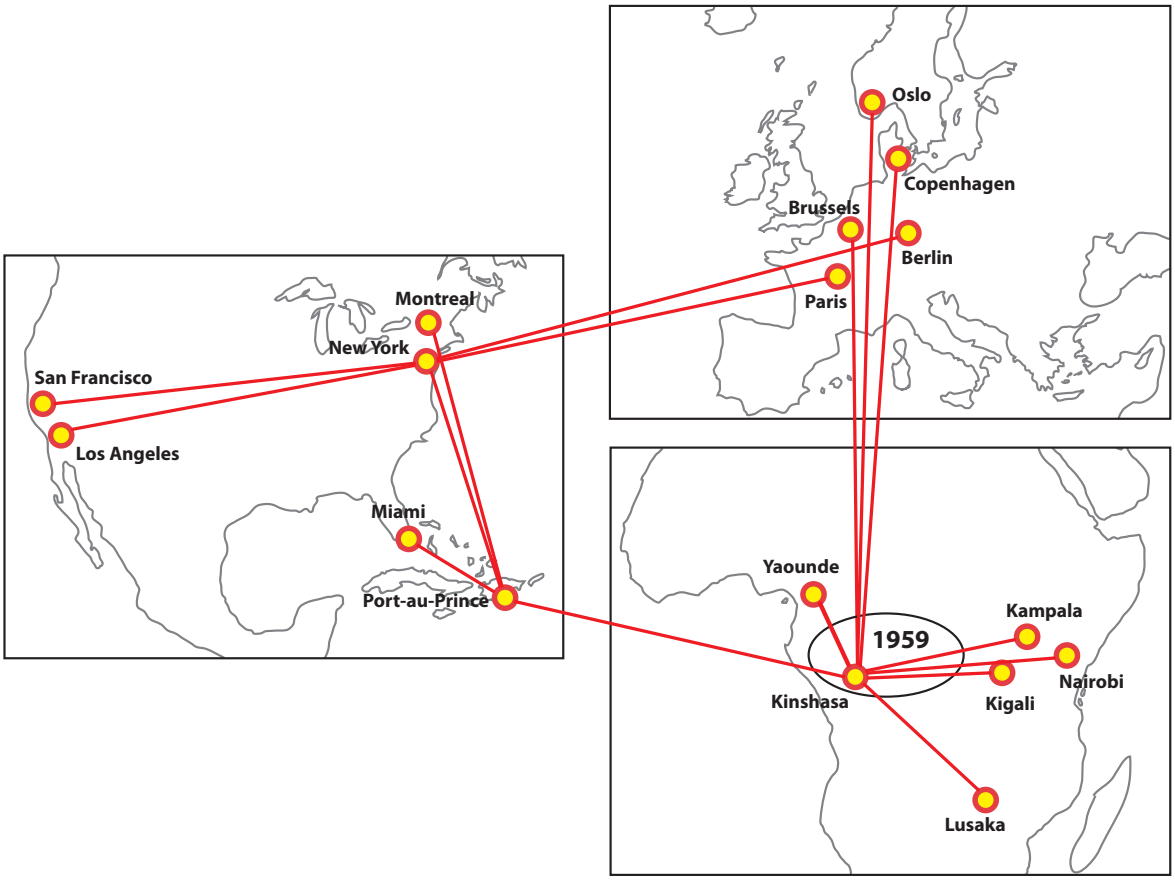

Figure 1.1 – Emergence of HIV/AIDS, 1959-1984

Maps show hypothetical global spread of HIV-1 Group M from the central African "epicentre", with earliest confirmed HIV-1 positive sera in 1959 from Kinshasa, Democratic Republic of Congo. Chimpanzees infected with SIVcpz believed to be the closest relative of HIV-1 are found in a region of southwest Cameroon 700km to the northwest of Kinshasa.

Since chimps have always been hunted, the factors that caused the AIDS pandemic in the twentieth century require elucidation. Serial transfer and amplification of virus through injection with contaminated needles during mass vaccination campaigns in west Africa during the 1950s and 1960s has been postulated as a "second step" transforming largely non-pathogenic SIV into infections causing AIDS in humans (131). This hypothesis is consistent with the origin of the well-described model SIV known to be highly pathogenic in macaque monkeys (SIVmac). This strain arose during accidental transfer of SIV from sootey mangabey monkeys (SIVsm) to macaques during transfer of brain tissue from one species to another during laboratory studies of the brain-wasting disease kuru in the 1970s. Interestingly, primary SIVsm is much less pathogenic in macaques compared to SIVmac, indicating that the adaptive process resulting in SIVmac is responsible for amplification of pathogenesis, possibly naturally constrained by host defences in skin and mucosae (10). Bypassing these actively protective barriers through transplantation (in the case of SIVmac) or mass vaccination campaigns transformed newly emerged HIV-1 into a virulent strain causing the global pandemic (131).

This hypothesis should not be confused with the largely discredited theory that HIV originated from SIV-contaminated primate kidney cells used to produce polio vaccine in Central Africa in the 1950s (92, 113). Rather, natural transfer of SIV to humans via predation was followed by alteration and amplification of HIV due to the widespread contemporary practice of needle re-use for multiple vaccinations. Since a single amino acid change in the HIV accessory protein Vif has been shown to dramatically increase viral pathogenesis (178), a change of this type during the expansion of the epidemic mid-century may explain the enhanced virulence of HIV-1 Group M strains.

Mechanisms of vaginal HIV transmission – The efficiency of male-to-female sexual transmission is very low, in the range of one productive infection for every 200-2000 exposures (91), demonstrating the formidable physicochemical and immunological barrier of the female genital tract (FGT) (Figure 1.2). Semen from infected individuals, especially those with bacterial co-infections, is known to contain very high concentrations of free virus as well as large numbers of infected lymphocytes and monocytes that may directly transfer infectious virus to vaginal epithelium, APC or Th lymphocytes at the mucosal "port of entry" (119).

The first barrier that the virus must cross is the mucus layer, a dense multi-layered network of glycoprotein molecules (mucins) colonized by high concentrations of diverse micro-organisms ("vaginal microbiota"). As well as creating a dense fluid gel overlying the epithelium, mucins bind covalently to epithelial cells to create a tightly adherent layer called the glycocalyx (224). Both cell-free and cell-associated virus may become trapped in these networks and subsequently inactivated by low pH and hydrogen peroxide (H_2O_2) of bacterial origin, or by epithelial-derived proteins with anti-viral properties, such as defensins or secretory leukocyte protease inhibitor (SLPI) and related effectors (80, 149).

Even under artificial infection conditions with a very high inoculum of pure virus, only a tiny proportion of virions gain access to susceptible cells in the epithelium and lamina propria (LP) (80), with productive infection frequently initiated by a single viral particle (119). Changes in the amount and consistency of excreted mucus, the profile and activity of colonizing microbiota, the amount and nature of epithelial defence molecules due to phase of menstrual cycle, sexual intercourse, co-infections, stress and nutrition are all likely to influence the likelihood that HIV can cross this initial physical barrier.

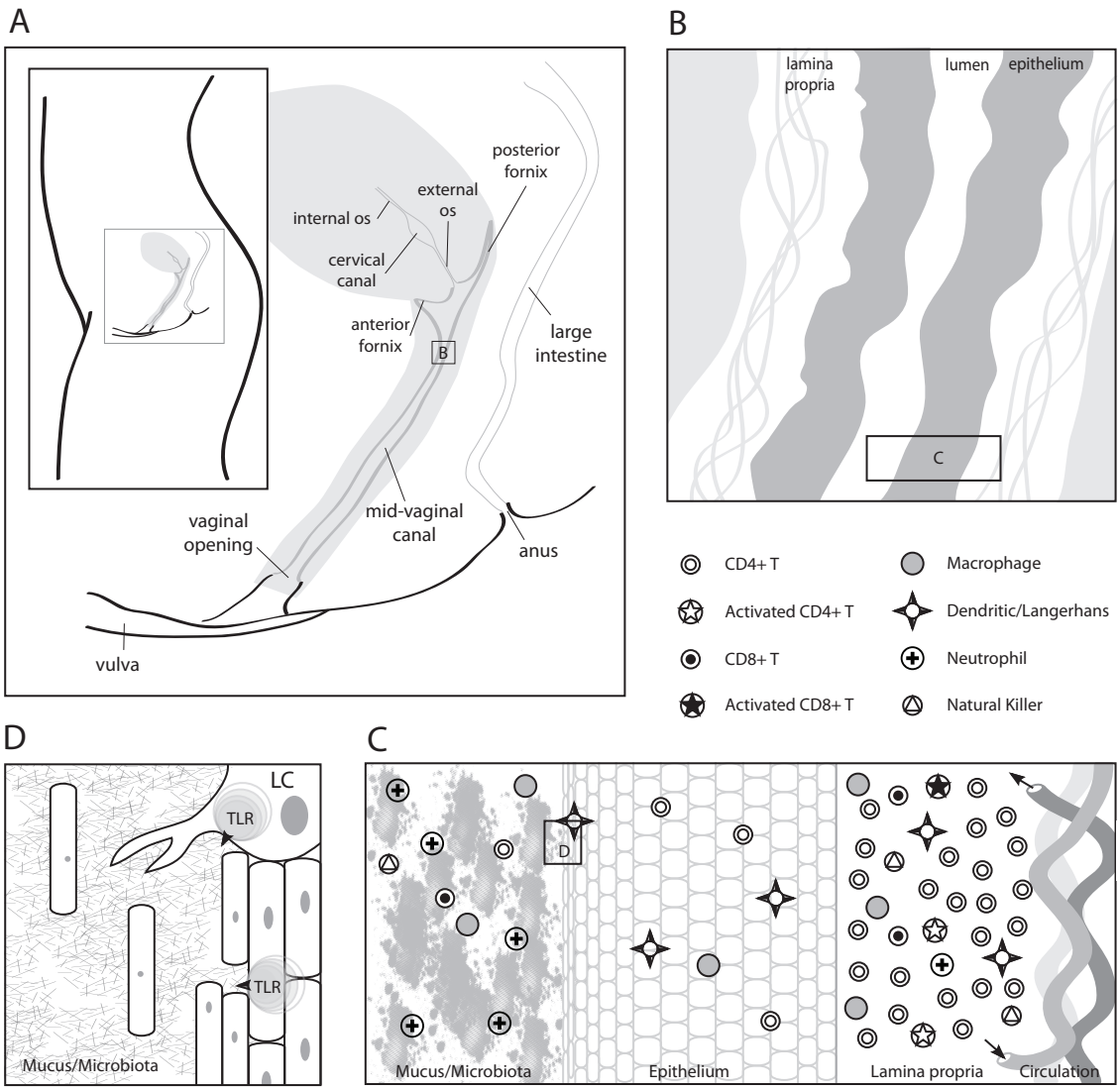

Figure 1.2 – The female genital tract

Overview of the major anatomical and histological features of mucosal surfaces in the female genital tract (FGT). A. Side view of FGT highlighting epithelial surfaces of vagina and cervix. B. Enlargement showing vaginal epithelium, underlying lamina propria and circulatory system (veins/arteries/lymphatics). C. Section of vaginal epithelium from lumen to lamina propria, highlighting common immune cell types in presence. D. Close-up of mucosal surface highlighting Langerhans cell (LC) in top right-hand corner and Toll-like receptors (TLR) on cell surfaces, as well as dense microbiota present at this site.

Several mechanisms for crossing the stratified squamous epithelium of the vagina and ectocervix or the simple columnar epithelium of the endocervix and upper reproductive tract to access target cells have been proposed (Figure 1.3), however the relative importance of each during infection *in vivo* has not been defined. Both HIV and SIV enter target cells in the FGT primarily after binding the CD4 cell-surface glycoprotein (receptor) and the chemokine ligands CCR5 or CXCR4 (co-receptors), found primarily on Th lymphocytes and APC, such as Mo, DC and Langerhans cells (LC) (23). Target cells are found throughout FGT epithelial surfaces, as well as in the LP, however distribution of various immune cell types has been shown to vary greatly between individuals and anatomical sites, and is strongly influenced by mucosal inflammation (163).

Normal vaginal epithelium is characterized by intraepithelial CTL, relatively few Th restricted to the LP and low numbers of LC in basal epithelial layers. However, greatly increased intraepithelial Th and LC are observed in inflamed vaginal tissues (163). Normal ectocervical epithelium contain higher numbers of LC, Mo, Th and CTL compared to normal vaginal epithelium, and these numbers are even higher in inflamed tissues.

CTL are often found in clusters at the base of the epithelium, or in association with dermal papillae, regularly-spaced “dimples” where vascularised LP extends close to the epithelial surface, possibly enhancing access for migration of leukocytes and monocytes into the epithelial layer. The highest concentrations of Mo, Th and CTL are found associated with the “transformation zone” (TZ) of the exocervical os, where stratified epithelium abruptly ends and the endocervical monolayer begins, indicating a critical position in immune control for ascending pathogens and also increased vulnerability to HIV infection (163).

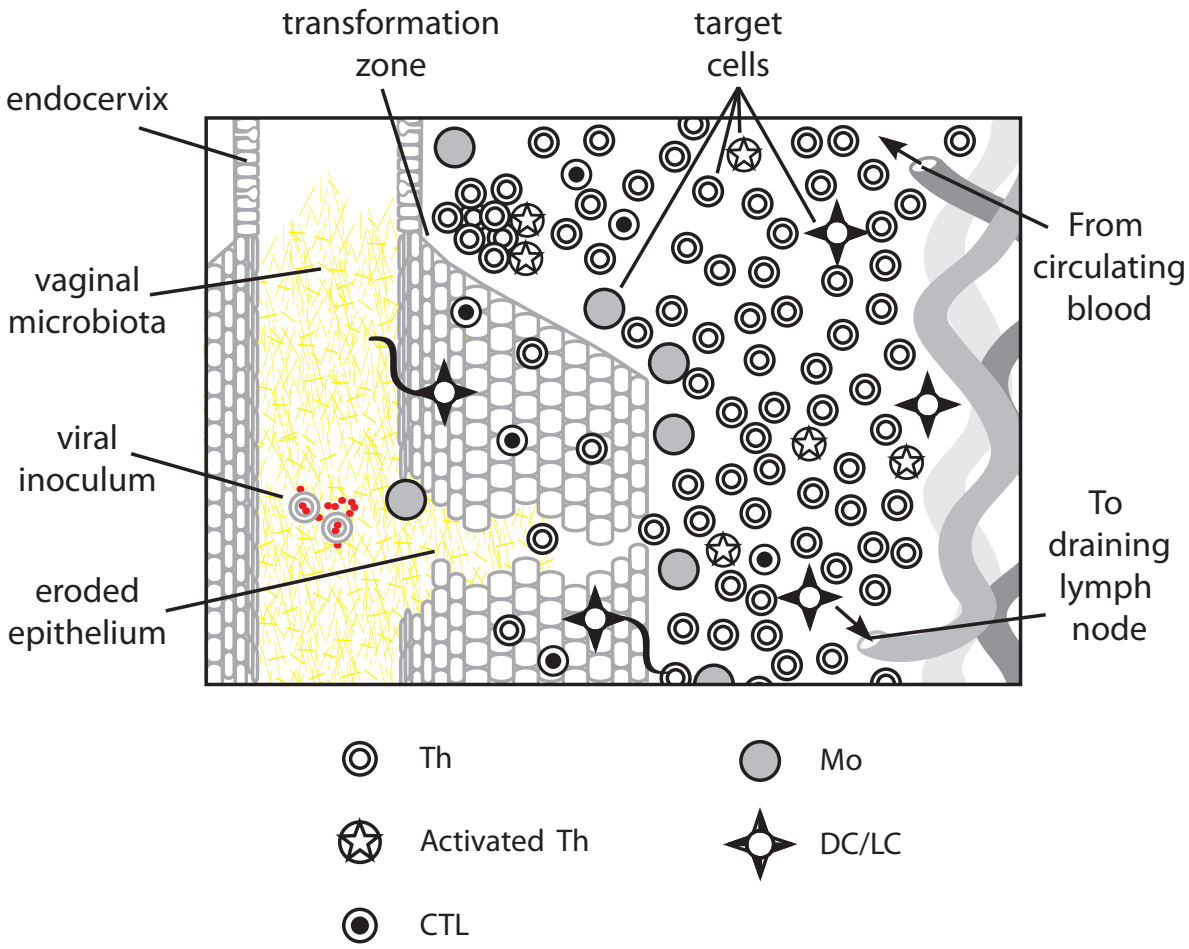

Figure 1.3 – Mechanisms of vaginal HIV transmission

Overview of hypothetical mechanisms of HIV transmission in the female genital tract. HIV enters the female genital tract as either free virions or as infected cells from an infected partner. Initiation of a productive infection depends on the concentration and activation status of target cells, including Th, DC/LC and Mo in the epithelium and lamina propria. The virus may be more likely to pass through the simple epithelium of the endocervix than the multi-layered squamous epithelium of the vagina and ectocervix, however the surface area of the latter is much greater. Under normal conditions, Th are concentrated in the lamina propria and transformation zone, while cervicovaginal inflammation greatly increases the distribution of targets in the epithelium. Inflammation also causes an increase in Th activation, increasing vulnerability to infection and expansion of viral populations. LC/DC sampling luminal contents may also transfer virus to Th without being productively infected themselves. Epithelial erosion due to physicochemical factors may also increase infection via the vaginal or ectocervical epithelium. Vaginal microbiota may influence HIV transmission by altering the concentration or vulnerability of target cells to HIV infection.

In the normal endocervix, relatively few Th are found in epithelium or sub-mucosally, although numerous Th and CTL can be observed just below the epithelial layer in inflamed tissue. Although LC are not found in the endocervix, intraepithelial macrophages are a consistent feature. Abundant B lymphocytes expressing antibodies (immunoglobulin A or IgA, IgG) are also a feature of the endocervical LP but only rarely present in ectocervical or vaginal LP (163).

These findings indicate the existence of several "immune micro-environments" associated with different anatomical sites within the FGT. Although the cervix is likely to be a preferred site of entry for HIV given higher concentrations of target cells in these regions, especially in the TZ, large increases in numbers of target cells in the vaginal epithelium observed during inflammation increase vulnerability to HIV infection in these tissues as well. Studies showing a lack of protective effect in those using diaphragms to protect the cervix from exposure corroborate the vulnerability of the vaginal epithelium to HIV infection (91).

Penetration of the epithelial layer has been shown to happen within minutes of exposure in experiments with non-human primates (91). Although CD4 and CCR5 receptors have been found on the surface of uterine epithelial cells (226), cervicovaginal epithelial cells have been shown to be free of these receptors (163). Alternatively, HIV can bind lipid moieties (for example, galactosyl cerebroside or GalC) or other receptors on vaginal epithelial cells (68), leading to transcytosis across or productive infection of the epithelial layer (42, 227). *In vitro*, transcytosis has been demonstrated across epithelial monolayers typical of the upper female genital tract (endocervix, uterus, fallopian tubes) (28), and basal layers of the pluristratified vaginal epithelium (91), but not across

superficial epithelial layers (31). However, virus may also penetrate the superficial layers by passing through inter-cellular space (91).

Intraepithelial and submucosal APC such as LC and DC are thought to play an important role in viral transport to large populations of susceptible cells in the lymphatic tissue. Terminal mannose residues on HIV surface glycoproteins bind to related receptors or lectins, called langerin in intraepithelial LC, or DC-specific-intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) in submucosal DC (47). Capture of HIV on LC or DC lectins can facilitate productive infection of the cells themselves (called *cis*-infection), or present infectious virus to Th without being productively infected (called *trans*-infection) (47).

Since LC form a tight network at the mucosal surface, and are known to sample the luminal space with specialized cell extensions (91), they are likely to be among the first cells to come into contact with HIV. In contrast to DC-SIGN, viral binding to langerin has been found to trigger degradation of viral particles and block *cis*-infection via an endosomal pathway that is unique to LC. In the absence of activating stimuli, LC resist infection with HIV and transmit virus much less efficiently to Th when compared to submucosal DC. However, activation of LC by TLR stimulation and epithelial production of the cytokine TNFα in the context of vaginal co-infections abolishes the protective role of langerin and facilitates *cis*-infection and migration of LC to lymphatic tissue (47).

Erosion of the cervicovaginal epithelium due to mechanical stress or ulceration is also hypothesized to facilitate transfer of infectious viral particles or infected cells to the sub-mucosa. An inflammatory micro-environment, possibly resulting from frequent sexual intercourse or vaginal co-infections, may exacerbate mucosal permeability and result in

recruitment of large numbers of target cells to the site of viral entry (47, 163).

Although the precise pathway travelled by HIV during the majority of sexually transmitted infections is unknown, it is clear that genital co-infections or other factors that: 1) increase shedding of free virus and infected cells by the infecting partner, 2) increase permeability of the protective mucus layer, 3) cause changes in the composition and function of vaginal microbiota, 4) disrupt the physical integrity of the epithelial barrier, or 5) increase the abundance of target cells at the mucosal port-of-entry, will increase the likelihood that HIV will be transmitted during heterosexual intercourse.

Immune activation and HIV pathogenesis – HIV infection results in profound, multi-factoral immune dysregulation eventually leading to AIDS usually after a period of 7-10 years. This lengthy asymptomatic period is characteristic of the lentiviridae family of viruses to which SIV and HIV belong. Progressive destruction of Th populations via direct viral killing is generally believed to explain slow disease progression, however recent studies in non-human primates have revealed that critical events in the earliest phases of infection determine SIV/HIV immunopathogenesis (33, 35, 80, 81).

Once founder populations of virus have adsorbed and penetrated target cells at the port of entry, there follows a period of several days in which there is little or no productive expansion of viral populations. This “eclipse phase” represents a critical opportunity to mitigate expansion of very small populations of virus in spatially dispersed foci amid a relatively small concentration of target cells in LP relative to lymphoid tissue (80). Stimulation of innate immune molecules such as toll-like receptors (TLR) on epithelial cells and APC by a variety of bacterial and viral components is likely to trigger increased production of pro-inflammatory cytokines such as IFNγ, IL-12, IL-6, IL-8 and others.

Inflammation and immune activation near the port of entry are likely to determine expansion or decay of HIV founder populations (105). If the virus reproductive rate exceeds unity (ie. each infected cell results in the infection of at least one other cell), the threshold to productive infection (the "fast phase") is crossed on or about day seven. The virus-as-signal is amplified and broadcast distally from the port of entry in LP to draining lymph nodes, the systemic circulation and all lymphatic tissue, rising to peak levels by day ten to fourteen (80).

As a result of direct viral killing and Fas/FasL-mediated apoptosis in the presence of high concentrations of viral gp120, mucosal Th populations, concentrated in gut-associated lymphoid tissue (GALT), vanish almost entirely within four days (35, 80). This massive immunological insult is not reflected in the systemic circulation, where only a minority of Th reside, explaining earlier models of gradual Th depletion in disease progression based on measuring Th in peripheral blood only (80).

A novel aspect of this model of HIV infection is that "resting" Th cells, which are the most abundant cells in the genital tract LP, are the primary target and amplifier of viral signal in acute infection. The activation state of these cells is ambiguous, however even low levels of expression of the HIV co-receptor CCR5 are sufficient for viral entry and replication and an increased abundance relative to activated cells compensates for reduced production of virions in "resting" cells (80).

Increased viral production in "resting" cells may also be explained by the ability of HIV gp120 to bind the integrin $\alpha 4\beta 7$, a gut mucosal homing receptor on Th. Binding of the integrin leads to activation of lymphocyte function-associated antigen (LFA)-1, which in turn is known to amplify HIV replication in Th (11).

A robust CTL response in the genital tract does not occur until a week after peak viral replication, and is too late to control viral populations, particularly considering extensive diversification of viral epitopes by random mutation over several generations (80). A general decline in viremia after the peak may not be determined by CTL control but rather to substrate exhaustion, with up to 80% of mucosal Th cells eradicated (35).

The "slow phase" of infection following the initial insult is marked by chronic, generalized immune activation, including increased activation of polyclonal B cells, increased activation and turnover of Th, and increased systemic proinflammatory cytokines (12, 50). Immune activation has been found to be a better predictor of disease progression than viral load, but whether the virus drives immune activation or vice versa is unresolved. Immune activation increases target cells for HIV infection and the self-perpetuating relationship between activation and viral replication leads to eventual clonal exhaustion of Th memory pools, as reflected in rising viral loads and declining Th counts as disease progresses (50).

Although antiretroviral therapy can mitigate immune defects by reducing viral replication, continued immune activation remains a barrier to immune reconstitution following treatment (50). In "elite controllers", defined as HIV-infected individuals who maintain very low viral loads in the absence of treatment, progressive loss of Th correlates with immune activation (50). Therefore, regardless of virus control in later stages of infection, chronic immune activation continues to strongly influence disease progression.

Although the idea that immune activation underlies HIV pathogenesis is not new (12), possible mechanisms of the association have only recently been elucidated. A now widely accepted hypothesis is that translocation of immune-stimulating microbial products

such as lipopolysaccharide (LPS) across damaged mucosal barriers in the gut may explain chronic immune activation in HIV pathogenesis (36, 50). Consistent with this hypothesis, a recent study has found higher LPS levels in the blood of chronically HIV-infected people and SIV-infected macaques (36).

In contrast to SIVmac infection in rhesus macaques, many SIV, including SIVagm from African green monkeys, cause no immunopathology in their natural hosts, despite intense viral replication and massive depletion of Th in systemic and mucosal compartments during acute infection (121, 153). Surprisingly, this initial insult to Th populations does not result in enteropathy or chronic immune activation in non-pathogenic infections, and the function of various immune cell subsets is preserved.

Although Th numbers in peripheral blood and mucosal tissues rise in the weeks following acute infection by both pathogenic and non-pathogenic SIV, steep declines in the chronic phase are only observed in pathogenic SIV infection. No increases in Th activation or apoptosis is observed in chronic non-pathogenic SIV infection, while activation and proliferation of CTL occurs early and then subsides, with very little detectable LPS in serum indicating preservation of the gut barrier and absence of microbial translocation. This contrasts with increased Th activation/apoptosis and CTL activation/proliferation in pathogenic infections, concomitant with increasing LPS levels in serum (153).

In contrast to pathogenic SIVmac or HIV-1 infection, upregulation of inflammatory cytokines such as IFNγ or TNFα and beta chemokines such as regulated upon activation, normal T-cell expressed, and secreted (RANTES) (chemotactic for monocytes and memory Th cells) and macrophage inflammatory protein (MIP)-1α/β (chemotactic for CTL) are not observed in serum of SIVagm-infected green monkeys (114,

153). Instead, upregulation of TGF-β stimulates expression of *FoxP3* in Th and CTL within one day of infection, leading to high levels of immunosuppressive IL-10 at early timepoints (114). This early induction of the Treg subset and IL-10 may counteract immune activation signals due to chronic infection in SIVagm-infected green monkeys, although Treg expansion and increased IL-10 in later stages of pathogenic SIVmac and HIV-1 infections have been associated with immunosuppression and AIDS (114).

An altered balance between Treg populations and the related Th17 subset has also been identified in pathogenic SIV infection of the pigtailed macaque, but not in non-pathogenic SIV infection of the green monkey (58). Th17 cells express IL-17 and IL-22, cytokines that have been shown to be important for immune responses to extracellular bacteria in the gut, where Th17 cells are concentrated (34). IL-17 induces IL-8 to promote recruitment of neutrophils to sites of bacterial infection, while IL-22 induces proliferation of intestinal epithelial cells and repair of the mucosal barrier, as well as stimulating epithelial production of anti-bacterial defensins.

Th17 cells and Tregs are derived from a common progenitor cell and their reciprocal differentiation depends on the cytokine milieu generated by DC with TLR stimulated by bacterial products in peripheral tissues (58). In humans, the Th17 phenotype is induced in naïve T cells in the presence of IL1β and IL-6 (1). Interestingly, most α4β7+ cells in GALT of rhesus macaques have been shown to be Th17 cells (103). Therefore, preferential depletion of Th17 cells in pathogenic SIV, as well as in HIV infection (34), may be due to the affinity of SIV/HIV gp120 for α4β7 integrin on gut mucosal Th, leading to permeability of mucosal barriers in the gut, microbial translocation and chronic immune activation. The early upregulation of Tregs in non-pathogenic SIV infection may

contribute to balanced ratios of Th17/Treg cell populations in GALT and mitigate immune dysfunction leading to AIDS (34).

An important lesson from these studies is that the timing and pattern of host immune responses to infection play a critical role in determining disease outcome. Anti-HIV microbicides (ie. vaginally-applied agents designed to block HIV transmission) or vaccines eliciting early robust immune responses may inadvertently cause amplification of the viral signal from the port of entry. Alternatively, induction of early tolerogenic responses to founder populations may block virus propagation and broadcasting into systemic compartments, or mitigate damage to mucosal Th populations thereby reducing immune activation resulting from microbial translocation.

HIV resistance in Kenyan CSW – As in "non-pathogenic" SIV infections, some untreated HIV-infected people remain healthy for very long periods ("long-term non-progressors" or LNTP). In contrast, "rapid progressors" develop AIDS relatively soon after HIV infection, analogous to what is observed during pathogenic SIV infection in the rhesus macaque model. Altered susceptibility to HIV is also observed among individuals able to resist infection entirely. Although rarely reported in non-human primates challenged with SIV (156), many individuals in diverse groups worldwide, including homosexual men, seronegative partners of seropositive individuals, hemophiliacs and children born to infected mothers, appear to resist HIV infection despite repeated exposure.

This variability in susceptibility to HIV infection and/or its pathogenic consequences has been intensively studied and found to correlate with a variety of viral, genetic, immunological and sociobehavioural variables (128). The immune activation hypothesis of HIV pathogenesis suggests that multiple factors may contribute to reduced

viral broadcast at the mucosal port of entry, either blocking productive infection entirely or limiting damage to lymphoid tissue subsequent to peak viremia. Therefore, these phenomena may be seen as being part of a continuum of HIV resistance (128).

Following up on the identification of rapid expansion of HIV in a cohort of CSW in a market area of Nairobi, Kenya (referred to from now on as the "Majengo cohort") (115, 157), further cross-sectional studies revealed that use of oral contraceptives and STI such as chlamydia and genital ulcer disease (GUD) were independently associated with HIV+ serostatus (162, 188). These factors are hypothesized to alter the integrity of the mucosal barrier or result in recruitment of susceptible cells to the mucosal port of entry, facilitating HIV infection.

Although rates of other STI including gonococcal cervicitis, syphilis and genital warts were also very high in the cohort, only chlamydia and GUD were independently associated with positive serostatus (162, 188). In one study, condom use was shown to be significantly protective against HIV infection, especially in those with GUD (162). Paradoxically, younger age and shorter duration of prostitution were also associated with being HIV-positive, although this was only statistically significant in one study (162, 188). Relative resistance to HIV infection was suggested as a possible explanation for this association, since a rising proportion of the CSW population should consist of persistently HIV seronegative women as HIV seropositive individuals drop out of CSW (188).

A prospective study of 434 individuals from the cohort followed between 1985 and 1994 was the first to formally propose that some individuals in the cohort were resistant to HIV infection (66). Despite prevention efforts, 239 individuals (56%) seroconverted during the study period. As in the earlier cross-sectional studies, shorter

duration of CSW and GUD were independently associated with seroconversion. In contrast to earlier studies, gonococcal cervicitis was independently associated with seroconversion while oral contraceptive use was not. Condom use was found to be associated with a reduced risk of seroconversion in univariate but not multivariate analysis (66).

Over the course of the study, infection "pressure" (defined as the estimated number of CSW clients who are seropositive in relation to the number of clients per day and frequency of condom use) was estimated to have increased from 24 to 64 exposures per year per individual in the cohort, yet each additional year of CSW while remaining seronegative was associated with a drop in the odds of an individual seroconverting. A group of 43 seronegative individuals with more than 3 years of CSW were confirmed to be HIV– using an ultra-sensitive polymerase chain reaction (PCR) assay, and were subsequently defined as "HIV-resistant" (HIV-R) (66). Possible confounders such as condom use and concomitant STI were assessed and did not alter the study's conclusions, indicating the possibility of a biological mechanism of resistance in a proportion of HIV– CSW in this cohort (66).

Although the precise mechanisms of protection in HIV-R individuals have not been elucidated, the phenomenon has been described in many highly exposed groups worldwide. Study of this phenomenon is hampered by lack of a clear terminology or definition of who should or should not be classified as HIV-resistant, difficulties in comparing groups of individuals who have been classified as HIV-resistant, and the small numbers of individuals included in these studies (140).

In the Majengo cohort, HIV resistance was originally defined epidemiologically as remaining HIV-1 seronegative despite longer than three years of CSW activity (66).

Since the length of time an individual has been involved in CSW cannot be independently verified, the three-year definition of resistance in practice has been based on length of follow-up in the cohort, rather than on self-reported length of CSW.

Several challenges are inherent in a time-based definition of resistance. First, conclusions made at one timepoint may no longer be valid at a later timepoint, since individuals may have changed categories in the interval. Second, infection pressure has likely fallen since the HIV resistance phenomenon was first identified, due to increased condom use and access to treatment for STI. Certainly, the incidence of HIV in the Majengo cohort has fallen dramatically in recent years (110). As a result, all new studies now define relative resistance to HIV as remaining HIV seronegative for longer than seven years of follow-up (39). However, this strategy only exacerbates the considerable and highly statistically significant age difference that is already observed between HIV-R and other HIV-negative (HIV-N) individuals using the three-year definition. Many women originally identified as HIV-R and whose immunogenetics are best described are now in their late 50s and 60s, with little or no ongoing engagement in sex work.

These issues reinforce assertions that the nature of resistance to HIV infection in exposed seronegative (ESN) groups around the world is likely to be complex and multifactoral (128, 140). Various genetic and immunological correlates of protection that have been identified in studies to date have not been proven to be absolute; rather, they are associated with "reduced susceptibility" to HIV infection (140).

Some of the most well-defined cases of HIV resistance have been demonstrated in individuals with mutations in the gene encoding the cellular HIV co-receptor CCR5, thereby inhibiting HIV entry in to target cells (128). This mutation is largely restricted to

Caucasian populations, and has not been observed in the Majengo cohort. Further, Th from HIV-R individuals in this cohort are readily infected with HIV *ex vivo* (64), indicating that cellular resistance to HIV is not a factor in this cohort as has been reported in other groups of ESN (171). In contrast to these early studies, however, a polymorphism in the coding region for the CD4 molecule, common in African populations, has recently been shown to be associated with increased HIV incidence in this cohort (152).

HIV-specific Th and CTL responses in persistently seronegative sex partners of seropositive individuals were observed as early as 1989, indicating that exclusively cell-mediated responses to HIV, possibly subsequent to an aborted infection, may result in protection against the establishment of HIV systemically (140, 161). Th and CTL with specificity for HIV antigens have also been demonstrated in many, but not all, HIV-R individuals in the Majengo cohort (65), and these responses have been shown to be qualitatively different from those in HIV+ cohort members (3, 4). CTL from HIV-R women have been found to respond to multiple conserved HIV epitopes *ex vivo*, important since these CSW are exposed to a wide variety of HIV clades and subtypes (124). Late seroconversion of HIV-R individuals with pre-existing HIV-specific CTL responses, possibly following a break from CSW, indicates that ongoing exposure to CSW and/or HIV may be required to maintain the resistant phenotype (106).

Interestingly, HIV resistance has also been observed in mothers and sisters of HIV-R cohort members (161), suggesting a genetic basis for HIV resistance in this cohort. Follow-up studies have led to the identification of several polymorphisms in immune-related genes. For example, several of the epitopes recognized by HIV-specific CTL in HIV-R women were found in association with specific human leukocyte antigen (HLA)

Class I molecules encoded by major histocompatibility complex (MHC) alleles that were also associated with HIV resistance in this group (124, 170). Since different MHC alleles produce HLA with different epitope specificities, it has been hypothesized that HLA from HIV-R individuals will select for conserved HIV epitopes, altering the likelihood of seroconversion when exposed to that virus (124).

Although several recent studies have identified MHC alleles that are strongly associated with HIV resistance and susceptibility in the Majengo cohort (83, 116, 211), these alleles have been found to be neither necessary nor sufficient for HIV resistance, indicating that favourable MHC alleles may contribute to HIV resistance in combination with other factors (124).

The observation that Th from HIV-R individuals in the Majengo cohort are more likely to produce the Th1-related cytokine IFNγ and not Th2-related IL-4 compared to HIV+ CSW (210) may highlight the effects of HIV infection rather than the dynamics of HIV resistance. Nonetheless, follow-up studies have led to the identification of single nucleotide polymorphisms (SNP) in non-coding regions of the interferon regulatory factor (IRF)-1 gene in the IL-4 gene cluster of HIV-R women, possibly explaining observed Th1-skewed responses to HIV in this group (16).

IRF-1 is critical in regulating immune responses to IFNγ through binding of promoter regions in IFNγ-stimulated genes, and has also been implicated in increased HIV replication through binding of HIV long terminal repeats (LTR) in infected cells. Since observed SNPs are associated with reduced expression of IRF-1, stimulation of HIV replication may be reduced in HIV-R individuals, thereby decreasing the reproductive rate of HIV and blocking productive infection (16).

Reduced IRF-1 production may also lead to generalized hyporesponsiveness to pro-inflammatory stimuli mediated by IFNγ. Generally lower levels of basal gene transcription and down-regulation of pro-inflammatory cytokines in Th from HIV-R CSW suggest that resistant individuals may be characterized by a "quiescent" immune profile (39). Multiparametric flow cytometry of peripheral blood mononuclear cells (PBMC) has demonstrated reduced activation of Th/CTL and increased frequency of CD4+CD25+FoxP3+ Tregs in HIV-R compared to HIV-N individuals, indicating that increased Tregs may potently downregulate Th activation levels, thereby reducing the likelihood of viral expansion beyond the mucosal port of entry (39).

Reduced immune activation has been proposed as a mechanism of protection in several other studies of ESN, including among men who have sex with men (MSM) in the Netherlands (112), CSW in Cote d'Ivoire (99), serodiscordant couples in Central African Republic (24), and neonates born to infected mothers in Brazil (118). However, increased immune activation has been observed in several other groups of ESN (27, 100, 198, 209), indicating that mechanisms of protection may vary among different groups of ESN.

Most studies of ESN discussed above have focussed on immune cells isolated from peripheral blood, however increasing attention is directed at the immunological characteristics of mucosal sites of the FGT (the mucosal "immune milieu") (105), where small HIV founder populations are most vulnerable to clearance.

HIV-specific neutralizing IgA has been identified in mucosal secretions of ESN from around the world, including in the Majengo cohort (71, 90, 107, 134). Mucosal IgA is regulated and expressed in the FGT independently of systemic IgA or IgG and can block HIV infection of target cells or transcytosis across epithelial barriers (140). However, HIV-

specific IgA has not been observed in all groups of ESN (140), and the association between mucosal IgA and HIV resistance (but not HIV exposure) in the Majengo cohort has recently been questioned (93).

Increased total T cells, as well as increased Th, CCR5+ Th and CTL have been reported in cervicovaginal samples from HIV-R compared to HIV-N individuals in the Majengo cohort (96). This finding is somewhat counterintuitive since it shows increased target cells available to be infected in the FGT. Interestingly, elevated mucosal (but not systemic) levels of the β-chemokine RANTES (chemoattractive for T cells at sites of inflammation) in HIV-R CSW may explain why elevated T cells are observed in the FGT, as well as protect against infection of susceptible cells by binding CCR5 (96).

A follow-up study focussed on screening proteins present in cervicovaginal lavage (CVL) samples from individuals in the Majengo cohort revealed increased concentrations of the anti-inflammatory serine protease inhibitor trappin-2 (also known as elafin) in CVL from HIV-R compared to both HIV-N and HIV+ individuals. This result was confirmed in an independent cohort of CSW in a different part of Nairobi (97).

Recently, the epithelial origin and suspected anti-viral properties of this molecule were confirmed, with studies showing direct inactivation of HIV particles during co-incubation with trappin-2 (70). Trappin-2 is related to the known anti-HIV molecule secretory leukocyte protease inhibitor (SLPI), with antibacterial, antifungal, antiviral and anti-inflammatory properties in addition to well-described antiprotease functions (17, 141). Further proteomic studies have identified upregulation of several other antiproteases in vaginal fluid of HIV-resistant individuals, including several serine protease inhibitors (serpins) and cystatin, a known anti-HIV factor (38).

HIV-R CSW in the Majengo cohort are among the most striking and well-studied examples of ESN individuals worldwide. Ongoing characterization of this group has revealed many immunogenetic factors that may contribute to relative resistance in a subgroup of CSW, including recent findings indicating a role for reduced T cell activation or "immune quiescence". Mucosal studies have revealed the possibility that HIV-R CSW may have increased anti-inflammatory factors in mucosal fluids, mirroring reduced immune activation in peripheral blood. Together, these studies suggest a unified set of factors in HIV-R women that reduce susceptibility of target cells *in vivo* or hamper viral expansion from the mucosal port of entry.

Since the HIV-R phenotype is relative and not absolute, environmental factors may plausibly contribute to HIV resistance in this group. Little is known about vaginal microbiota in African individuals or in CSW, therefore an evaluation of the dynamics of microbiota and their interactions with HIV at the earliest stages of infection may contribute to an overall understanding of the HIV resistance phenomena in the Majengo cohort.

Vaginal microbiology and HIV susceptibility – Genital tract infections such as chlamydia, gonorrhea, syphilis, herpes simplex virus (HSV), human papilloma virus (HPV) and trichomoniasis are known to increase the risk of HIV acquisition (26, 60, 104, 184). The "enigmatic" clinical entity known as bacterial vaginosis (BV) (117, 222) is associated with increased susceptibility to many STI and has been shown to increase risk for HIV infection in numerous studies from around the world (13). The mechanism of this association has not been fully elucidated, however a testable hypothesis is that microbiological changes associated with BV create a permissive physiological environment for expansion of HIV founder populations at the mucosal port of entry (80).

BV is a frequently asymptomatic clinical condition defined by a reduction in vaginal *Lactobacillus* populations and overgrowth of anaerobic and Gram-negative organisms (222). Increasingly linked with negative reproductive health outcomes, such as miscarriage, premature birth, post-operative infections, pelvic inflammatory disease and HIV infection, its etiology and clinical course remain poorly defined with a consequent lack of effective prevention and treatment strategies (62, 63, 117).

The current "gold standard" for diagnosis of BV is based on observation of clinical markers, known as Amsel's criteria, including thin, milky discharge, high vaginal pH, presence of "clue cells" (shed epithelial cells covered in Gram-negative bacteria), and amine ("fishy") odour, (7). In the laboratory setting, most studies rely on profile of microbial types in Gram-stained vaginal swab smears (146). These scoring systems are based on the relative abundance of *Lactobacillus* morphotypes (short to long, straight-sided rods) compared to BV bacteria morphotypes (Gram-negative/variable coccobacilli and curved rods).

Quantitative bacteriology and microbial ecology are of particular concern in understanding BV, which is defined by changes in levels of total bacteria and alterations in abundance of bacterial sub-populations, rather than the presence of a specific pathogen (62). For example, BV-associated bacteria such as *Gardnerella*, *Mobiluncus*, *Atopobium* and *Mycoplasma* species are routinely isolated from women without BV (63). Simultaneous isolation of *Mycoplasma*, *Gardnerella* and anaerobes, described as a "pathologic core", combined with failure to isolate *Lactobacillus* in a vaginal sample, has been found to be strongly associated with BV (206). However, current anaerobic culture techniques required to detect these organisms are not feasible for routine diagnosis (63).

Despite a general lack of external signs of inflammation in BV (as indicated by the *–osis* suffix signifying overgrowth, rather than the *–itis* suffix signifying inflammation) (94), many studies associate BV with increased proinflammatory signalling via TLR stimulation (123, 126, 139). BV is characterized by an increase in the overall bacterial load in the vagina, with possible implications for TLR stimulation and immune activation. The transition from *Lactobacillus*-dominated to BV microbiota implies an increased number of different bacterial species or phylotypes (density) and an increased relative abundance of a range of phylotypes (diversity) (77). These ecological parameters may also have important implications for epithelial defenses and immune modulation related to HIV susceptibility.

The presence of various proteolytic enzymes during BV demonstrate the importance of mucin lysis as a source of energy for the bacterial consortium (167), and may also explain nutritional synergies observed in culture-based experiments with BV bacteria (164). Enzymes such as sialidase and prolidase have been shown to impair innate and adaptive immune effectors in the vagina, including cytokines, immunoglobulins and cellular receptors, and are likely to explain lack of overt signs of inflammation despite local production of pro-inflammatory signals during BV (41).

Given a pre-eminent ecological position in the vaginal environment (84), colonization of the female genital tract by *Lactobacillus* species is increasingly recognized as critical for overall vaginal health and resistance to infection by bacterial and viral pathogens, including HIV (9, 25, 167). Under the influence of estrogen, vaginal epithelial cells store glycogen, which is hydrolysed to glucose and metabolized by vaginal bacteria (89). Excreted lactate, especially by *Lactobacillus* species, reduces the overall pH of the vaginal lumen to “normal” levels, in the range of 3.9 to 4.5 (29, 30). Along with effectors

such as H_2O_2 and bacteriocins, acid production by *Lactobacillus* is believed to discourage overgrowth of other bacterial genera including *Streptococcus*, *Gardnerella*, *Bacteroides* and *Mycoplasma* (168). H_2O_2 produced by a *Lactobacillus* strain has been shown to inactivate HIV *in vitro* (111). Production of acid and H_2O_2 may be synergistic, since bacterial H_2O_2 is more likely to remain stable at lower pH *in vitro* (61).

The ability of *Lactobacillus* to produce acid from mannose scavenged as a carbon source has also been hypothesized to block HIV infection through digestion of mannosylated residues on viral glycoproteins (204). Interestingly, two promising molecules currently being studied as potential topical agents for blocking HIV infection (microbicides) also bind mannose residues on the viral surface (37, 147).

Absence of culturable H_2O_2-producing *Lactobacillus* has been shown to be an important contributing factor in elevated risk for cervicitis observed in those with BV (129). This finding confirms that BV-related organisms are associated with an elevated risk of genital tract inflammation, the converse of which may be a non-/anti-inflammatory role for *Lactobacillus* organisms. The possibility of anti-inflammatory effects by *Lactobacillus* organisms in the vagina is all but unexplored, although several studies concerning the gut epithelium indicate induction of epithelial defence molecules, anti-inflammatory signalling and increased Tregs by specific strains of *Lactobacillus* (32, 41, 120, 207).

Since HIV resistance has been shown to be associated with increased anti-inflammatory effectors of epithelial origin, reduced activation of Th and an increased frequency of Treg cells, a testable hypothesis is that induction of this "quiescent" phenotype is due at least in part to functional characteristics of "anti-inflammatory" strains of vaginal *Lactobacillus*.

Therefore, vaginal microbiology is predicted to influence susceptibility of mucosal surfaces to HIV infection via a number of direct and indirect mechanisms (Figure 1.4). The quantity, viscosity, and/or permeability of the mucin network in mucus layer may be altered via proteases or other effectors of bacterial origin. Viral particles or infected allogenic cells may be directly inhibited via binding, mannose scavenging or bacterial metabolites with antiviral properties (acid, H_2O_2). Innate mucosal defence molecules and cells may be stimulated via bacterial interactions with TLR. Th activation and susceptibility may be modulated by bacterial conditioning of TLR-stimulated DC, Mo or epithelium. Whether any of these factors are involved in observed resistance to HIV in the Majengo cohort has not been addressed by previous studies.

Further work will be required to understand how microbiological and host factors interact with each other as part of the normal functioning of mucosal surfaces exposed to a wide variety of endogenous and exogenous stimuli. How these interactions influence the likelihood that HIV introduced at the interface of host and microbiota will cross formidable physicochemical barriers and irreversibly alter host immune function should be a central question of HIV prevention research in coming years.

Current and novel methods to characterize vaginal microbiota – Scientific interest in human microbial communities is growing, and basic concepts about the "human microbiome" are evolving rapidly (49, 212, 224). From the invention of the microscope, to the advent of anaerobic culture in the 1970s, to the molecular revolution of the 1990s, the range of observable micro-organisms in natural samples has continually expanded.

Current methods of studying vaginal microbiota include some of the oldest techniques (microscopy) to the most modern (deep sequencing).

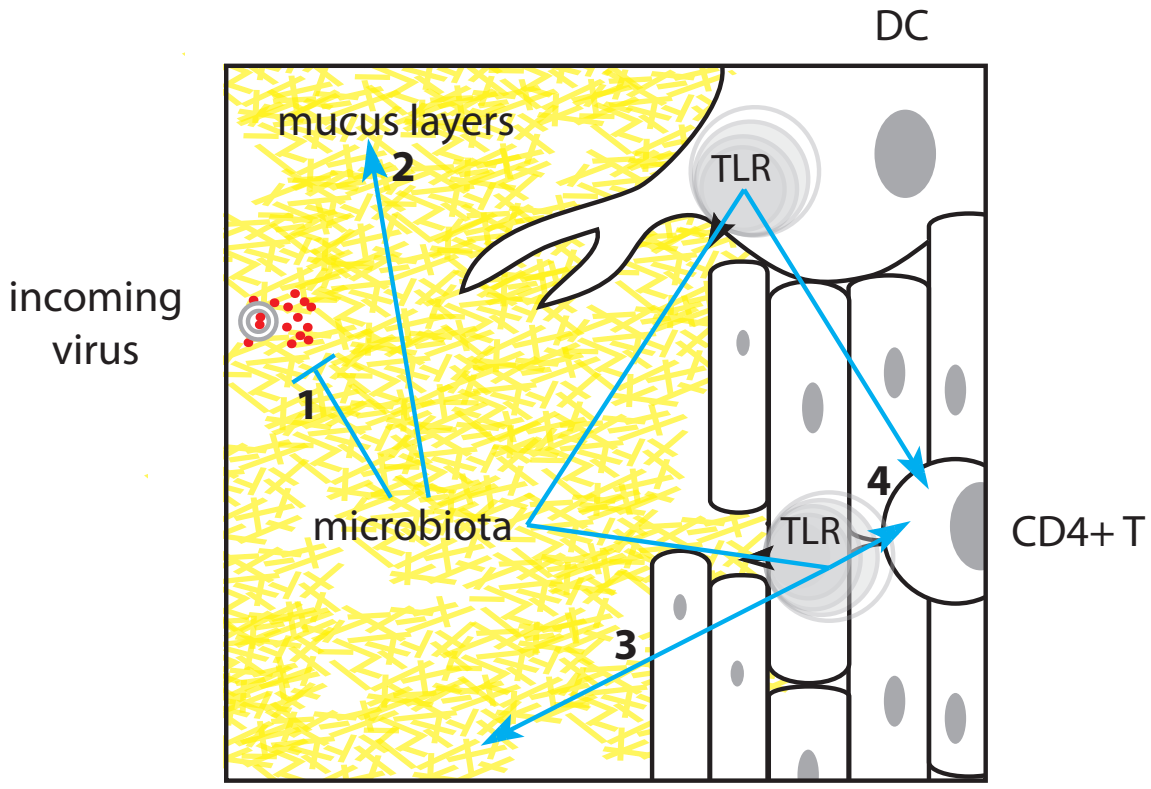

Figure 1.4 – The microbiological context of HIV resistance

Vaginal microbiology influences susceptibility of the mucosal layer to HIV infection in several ways, including: 1. direct effects on viral particles or infected allogenic cells via binding, mannose scavenging or bacterial metabolites (acid, H_2O_2), 2. alteration of quantity/viscosity/permeability of mucin network via proteases, 3. stimulation of innate mucosal defense molecules and cells via TLR, 4. modulation of T cell activation and susceptibility via conditioning by TLR-stimulated dendritic cells, macrophages or epithelial cells.

Molecular phylogenetic analysis of phylogenetically informative DNA sequences has revealed a vast diversity of uncultured microbial symbionts that influence animal physiology in ways only beginning to be understood. In particular, microbial species inhabiting the human vagina are thought to play an important role in host health (84).

Current in-depth culture-based and molecular methods for analysis of vaginal microbial communities provide very rich information about presence and abundance of different bacterial types but are labour-intensive such that only cross-sectional studies in small- to moderately-sized groups and longitudinal studies in small groups have been reported so far (9, 85, 95, 151, 228). A recent review describes many reports of in-depth culture-based and molecular analyses, designed to more fully characterize the range of organisms present in women with and without BV (122).

In order to address study hypotheses and goals, several accepted and innovative approaches for characterizing vaginal bacteria were assessed, developed and applied. Many studies of vaginal bacteria use the pre-weighed, dual-swab technique described in this study, allowing for expression of culture-based measurements, such as colony-forming units or CFU, as concentrations per gram of sample (150, 151). Measuring sample weight allows for control of variability based on the amount of secretions absorbed by the collection device and has been shown to greatly enhance the reproducibility of quantitative results (132).

The self-collect approach has been used since the 1970s (20), and has been validated for BV diagnosis by comparing results from self-collected swabs vs. physician-collected swabs (142, 180). Self-collect is also increasingly used for diagnosis of sexually transmitted infections (76, 183). The principal benefit of the approach is lack of

requirement for pelvic examination, reducing clinician time and improving privacy and comfort of the participant (182). However, vaginal inflammation or other symptoms cannot be assessed.

Since Amsel's criteria are not routinely assessed in Nairobi, BV is defined in this study based only on microscopy of Gram-stained slides and not on "gold standard" clinical criteria. Despite some concerns about its accuracy as a predictor of BV (205), many studies rely exclusively on Gram stain data to draw conclusions about hormonal, sociodemographic or behavioural factors related to BV, particularly for longitudinal studies and in the developing world (108, 142, 181, 190).

Flow cytometry is a novel approach to rapidly quantify all bacterial and mucosal cells in a vaginal sample (175). The power of flow cytometry to rapidly quantify total bacterial cells in natural samples has been clearly demonstrated in numerous microbiological fields and documented in several comprehensive reviews (6, 45, 189, 195). However, most acknowledge that this method has not been utilized to its full potential in microbiology. Flow cytometry can precisely enumerate bacteria in different kinds of samples using a simple latex micro-particle protocol (219). Further, all bacterial cells are counted by flow cytometry regardless of their ability to form colonies on culture media (177).

Combining flow cytometry with species-specific probes (fluorescent in situ hybridization or FISH) or immunofluorescence (tagged antibodies) has great potential to provide quantitative information about bacterial subpopulations in vaginal samples. Originally, both of these approaches were pursued in order to address Hypothesis 4, however several attempts to stain pure cultures of bacteria with *Lactobacillus*-specific

fluorescent ribosomal probes and an attempt to make *Lactobacillus*-specific monoclonal antibodies in collaboration with J. Berry (NML) were unsuccessful. Therefore, flow cytometry experiments focussed only on quantitation of total bacteria and mucosal cells present in vaginal swabs.

Although some exhaustive culture-based analyses of vaginal bacterial populations using multiple culture media have been reported (150, 151), many studies semi-quantitatively assess *Lactobacillus* colonization and isolate strains for further analysis via culture of vaginal specimens on selective media, most often on Rogosa (8, 101, 145, 208). Production of H_2O_2 by *Lactobacillus* isolates is frequently determined using Rogosa with added tetramethylbenzidine (TMB) and horseradish peroxidase (HRP), which turns blue in the presence of H_2O_2 upon exposure to oxygen after anaerobic culture (2, 165, 193).

Many studies use molecular "signatures" to rapidly group together and identify vaginal isolates (14, 56, 218). Ribosomal interspacer analysis (RISA), which measures inter-strain variability in the number and size of interspacer regions (ISR) located between 16S and 23S rDNA, is frequently described to characterize complex microbial samples as well as individual isolates (40, 46), but has not previously been applied to vaginal samples or isolates.

Several recent studies demonstrate increasing interest in determining the composition of the vaginal microbiota by culture-independent methods (9, 67, 85, 95, 148, 194, 220, 221). The construction of clone libraries is considered the gold standard for culture-independent analysis of complex microbial communities. The application of this technique using the *cpn*60 UT is well-established for analysis of microbial communities and has recently been applied to the vaginal microbiota (85). One drawback to this

approach is time- and labor-intensiveness, limiting the reach of phylogenetic or functional surveys of microbial communities across body sites, individuals, geographic areas and scales of time.

The advent of next-generation ultra-high throughput sequencing technologies, in particular the GS FLX (454/Roche, Laval, PQ), has removed an important quantitative barrier in molecular analysis by increasing the number of reads from a gene or genome by orders of magnitude in a single run (127). Unfortunately, the short average length of pyrosequencing reads (~200bp, compared to ~700 bp using dideoxy sequencing) presents a new set of problems. The results of recent application of this technology to analysis of 16S rRNA gene sequences from microbes in vaginal samples have demonstrated that short reads are more likely to generate matches to multiple sequences in the ribosomal RNA sequence database. Taxonomic and phylogenetic resolution was limited due to strong similarities between 16S rRNA sequences from closely related species (197). Since the *cpn*60 UT is shorter, has improved species resolution, and even distribution of variability across the target, it is an ideal candidate target for deep pyrosequencing studies.

2. Hypotheses, Goals & Objectives

The primary goal of this study is to determine a microbiological basis for resistance to HIV infection in the Majengo CSW through in-depth characterization of bacteria in vaginal swab specimens.

The main hypothesis to be addressed in this study is that differential colonization of the female genital tract by protective *Lactobacillus* organisms explains why some women remain HIV-negative despite prolonged occupational exposure to HIV.

By comparing individuals defined in epidemiological terms as HIV-R (ie. remaining HIV– after 3 years of follow-up) to HIV-N members of the Majengo cohort, the following objectives and hypotheses will be addressed:

1. Determination of BV prevalence and vaginal pH in cohort. *Methods:* Microscopy of Gram-stained slides prepared from vaginal swabs; Measurement of vaginal pH. HIV-R individuals will:
   a) be less likely to be diagnosed with BV by microscopy,
   b) have lower vaginal pH.
2. Quantification of total bacterial concentrations. *Methods:* Flow cytometry. HIV-R individuals will:
   c) have lower overall bacterial concentrations in vaginal specimens.
3. Quantification of *Lactobacillus* concentrations. *Methods:* Enumeration of colonies on selective media; Quantitative PCR. HIV-R individuals will:
   d) have an increased absolute abundance of *Lactobacillus*.

4. Determination of acid production from mannose and H_2O_2 production by vaginal isolates. *Methods:* Culture-based isolation on selective media; Phenotyping of isolates on chromogenic media detecting acid and H_2O_2. HIV-R individuals will:
    e) be more likely to be colonized with bacteria that produce H_2O_2,
    f) be more likely to be colonized with bacteria that produce acid from mannose.
5. Determination of overall phylotype density and diversity as well as presence and relative abundance of *Lactobacillus* vs. BV-related bacteria. *Methods*: Ribosomal interspacer analysis and sequencing of cultured isolates; Construction of clone libraries; Deep sequencing using GS-FLX. HIV-R individuals will:
    g) be more likely to be colonized with *Lactobacillus* phylotypes,
    h) be less likely to be colonized with BV-related phylotypes,
    i) have an increased relative abundance of *Lactobacillus*,
    j) have a reduced relative abundance of BV-related bacteria,
    k) have reduced overall density and/or diversity of bacterial phylotypes.

A number of secondary study goals addressing gaps and ambiguities in the scientific literature were also pursued, further expanding the utility of in-depth culture-based and molecular information generated during this project, including:

Effects of HIV on vaginal microbiology – HIV infection is well known to alter the body's relationship to infectious agents in the environment, increasing susceptibility to a range of opportunistic infections. However, little is known about how HIV might alter commensal microbial communities such as those in the vagina. Therefore, microbiological parameters were compared between HIV+ and HIV– individuals, taking into account the

potential effects of antiretroviral therapy and antibiotic prophylaxis in HIV+ women.

Defining BV – BV is an enigmatic, ecological condition that is poorly understood within current biomedical paradigms and is currently characterized primarily using clinical and microscopic criteria. Therefore, studies were undertaken to expand on traditional definitions of BV by comparing phenotypic and molecular profiles of those with and without BV. The role of other biological and behavioural confounders, such as age, douching practices and concomitant sexually transmitted infections among others, was also evaluated.

Vaginal health and HIV prevention – The precise mechanisms underlying the well-known association between BV and vulnerability to HIV infection have not been elucidated. Therefore, the results of this study were considered in relation to factors that might improve vaginal health and reduce susceptibility to HIV infection for CSW in the Majengo cohort.

B. MATERIALS & METHODS

Study groups – Initially, a pilot study was conducted based on self-collected vaginal samples from young women attending a gynecology clinic in inner-city Winnipeg. The "Good Bacteria Study" was carried out with a convenience sample of 32 patients attending Children's Hospital, Health Sciences Centre, during June-August, 2005. As well as providing a total of 103 self-collected vaginal swabs, participants completed a short, self-completed survey (age, type of menstrual protection, time since last menstruation, douching, last sexual intercourse and hormonal birth control use).

These samples were used to optimize most experimental procedures in preparation for the main study carried out with samples provided by members of the Majengo cohort. All study procedures were approved by the Ethics Review Board, University of Manitoba.

The main study group was selected from the Majengo cohort as part of a well-established collaborative study for STI/HIV prevention in the Departments of Medical Microbiology at the University of Manitoba and University of Nairobi (64, 66). A cross-section of 242 individuals provided vaginal specimens in June-July, 2007. This group is not a random sample of CSW in Nairobi, but rather consists of all CSW attending the clinic during the study period.

The main test group consists of HIV-resistant CSW (HIV-R; remaining HIV-negative after 3 years or more of follow-up). The main comparison group is made up of other HIV-negative CSW (HIV-N; HIV-negative with less than 3 years of follow-up). Remaining individuals were HIV-positive (HIV+), and were included in order to assess the effects of HIV infection on vaginal microbiology in relation to HIV– groups.

A subset of 32 individuals in each of HIV-R, HIV-N and HIV+ groups was selected for in-depth culture-based analysis (Culture subset, N=96), and further subsets were selected for construction of clone libraries (Clone subset, N=10), pyrosequencing studies (Pyro subset, N=48) and quantitative PCR (qPCR subset, N=40).

Since BV was shown in preliminary studies to be a critical parameter defining the vaginal microbiota, subsets were selected primarily in order to ensure adequate representation of individuals with different BV diagnoses. Since the HIV-R group was significantly older on average than the HIV-N group (described below), an attempt was also made to select younger HIV-R individuals for the subsets. Although this slightly minimized the difference in age, it did not eliminate the significance of the difference (described below).

All Nairobi participant data, including age, length of time in cohort, current CSW activity, date of last menstrual period, douching practices, type of birth control used, number of CSW clients per day, concurrent STI, vaginal symptoms and other clinical parameters, were derived from an extensive database of survey and laboratory data collected on the same day as the vaginal samples.

A retrospective, longitudinal analysis using information extracted from the project database was carried out in order to characterize BV diagnoses in 1,005 individuals providing a total of 3,539 vaginal specimens over the period of November 2004-July 2008.

All study procedures were approved by the Ethics Review Boards of the University of Manitoba and the University of Nairobi.

Sample collection and processing – In the pilot study, samples were self-collected using double-headed Starplex devices (Starplex Scientific, Inc., Etobicoke), pre-weighed to

the nearest milligram. All swab sets were marked under aseptic conditions with felt pen at a distance of 5 cm from the swab tip. Participants were instructed to insert the swab to that point, rotate the device to sample the entire mid-vaginal lumen, while avoiding contact with outside skin and surfaces. Samples were processed within three hours of sample collection. Swab devices were weighed to the nearest milligram upon receipt and the initial weight subtracted from this value to estimate sample weight.

One swab head was rolled onto a glass slide, air-dried, fixed with heat and Gram-stained for microscopy using standard protocols. For assays using fresh samples, swab heads were snipped into 1ml phosphate-buffered saline (PBS: 0.8% w/v sodium chloride, 0.02% w/v potassium chloride, 0.14% w/v sodium phosphate dibasic and 0.024% w/v potassium phosphate monobasic) (all chemicals from Sigma-Aldrich Canada, Oakville, ON) in 1.5 ml Eppendorf tubes and vortexed vigorously to dislodge specimen material from the swab. The swab head was removed with forceps, squeezed against the side of the tube to remove excess liquid, and discarded. Sample was centrifuged, supernatant discarded, and pellet resuspended in 1ml PBS. For frozen storage, swab heads were snipped into 1ml RTF-glycerol buffer (0.045% w/v potassium phosphate monobasic, 0.045% w/v potassium phosphate dibasic, 0.09% w/v sodium chloride, 0.018% w/v magnesium sulphate, 0.038% w/v EDTA, 0.04% w/v sodium bicarbonate, 0.02% dithiothreitol and 10% v/v glycerol) (all from Sigma) (109) in 1.5ml Eppendorf tubes.

For the Nairobi study, specimens were collected using pre-weighed, double-headed Starplex swabs of the mid-vaginal canal by a single physician, stored without buffer in coolers and delivered to the lab from the clinic after each day's sample collection (max. 3 hours). One swab head was immediately frozen at –80°C in RTF-glycerol buffer as

above for shipment to Winnipeg in liquid nitrogen tanks. The other swab was dabbed onto pH-sensitive paper with a range of 4-7 (Whatman, New Jersey, NJ) the value recorded according to manufacturer's directions, then snipped into 1ml PBS and sample eluted and washed as described above prior to subsequent assays.

Diagnosis of BV by microscopy – In the pilot study, diagnosis of BV was made using the criteria of Hay and Ison (98). Slides were scored on a three-point ordinal scale, where 1 denotes dominance of *Lactobacillus* morphotypes (BV–), 2 denotes reduced *Lactobacillus* in a mixture of bacterial morphotypes (BV-intermediate or BVI), and 3 denotes absence of *Lactobacillus* and predominance of BV morphotypes (BV+). Slides with no visible bacteria were scored as BV0. Slides were blinded and the entire set examined and scored three times, without reference to patient identifiers or previous scores. Slides with discrepant scores were evaluated in more detail and re-scored.

In Nairobi, slides are prepared in the clinic from separate vaginal swabs and Gram-stained by laboratory staff using standard procedures. BV diagnosis is based on the Nugent score according to the following criteria. The number of *Lactobacillus* and BV morphotypes in several high-powered fields (HPF) is summarized in a semi-quantitative abundance score with 5 levels (0=None, 1=<1 per field, 2=1-5 per field, 3=6-30 per field, 4=>30 per field). The number of curved morphotypes is summarized in a 3-level score (0=None, 1=1-30 per field, 2=>30 per field).

To generate the Nugent score, abundance scores for BV and curved morphotypes receive the same number of points (ie. 0=0, 1=1 etc.), while for *Lactobacillus* morphotypes, abundance scores receive the inverse number of points (ie. 0=4, 1=3, 2=2, 3=1, 4=0). This results in a Nugent score based on 10 points total, with 4 points for the

highest abundance of BV morphotypes, 4 points for the complete absence of *Lactobacillus* morphotypes and 2 points for highest abundance of curved morphotypes (Figure 2.1).

BV diagnosis is made by collapsing the 10-point scale into a 3-level scale: 0-3 = 1 (BV–), 4-6 = 2 (BVI), 7-10 = 3 (BV+). Slides with very different characteristics can have the same Nugent score. For example, complete absence of bacterial morphotypes or the highest abundance of *Lactobacillus* and BV morphotypes in the same smear both lead to a score of 4 or BVI. As in the pilot study, complete absence of visible bacteria was defined as BV0 and considered separately from other slides with BVI scores. Semi-quantitation of yeast cells and leukocytes was also carried out during microscopic analysis.

For comparison with other study variables, BV diagnosis was considered either as a 4-level categorical variable (BV–, BVI, BV+, BV0) or sometimes as a 2-level variable comparing BV– to BVI/BV+/BV0 diagnoses combined (BV– vs. any BV), or sometimes excluding BVI and BV0 diagnoses and comparing BV– to BV+ diagnoses (BV–/BV+).

For the main study, slides were scored independently by two evaluators and slides with discrepant diagnoses (32/242 or 13%) were re-analyzed until consensus was reached. In all but one case, the diagnosis made by the more experienced evaluator was retained as correct. The consensus diagnosis was used in further data analysis. For the retrospective study, all BV diagnosis data was extracted from the project database, therefore all slides were read by a single, highly experienced evaluator.

Quantitative flow cytometry – Quantification of microbial and mucosal cell populations in the sample was carried out in reference to a solution of fluorescent micro-particles at a known concentration. Flow cytometry was carried out on undiluted sample and on one tenfold dilution of the sample.

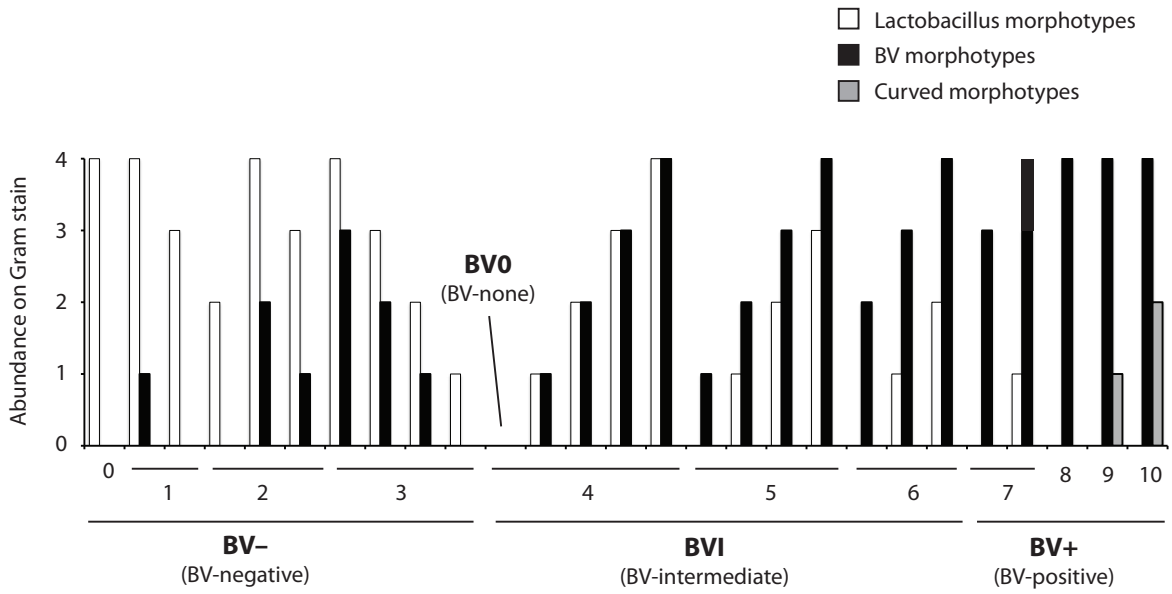

Figure 2.1 – Laboratory diagnosis of bacterial vaginosis (the "Nugent score")

The Nugent score is derived by recording semi-quantitative abundance of *Lactobacillus*, BV and curved morphotypes in several high-powered fields of a Gram-stained vaginal swab smear. Highest abundance of *Lactobacillus* morphotypes receives 0 points while highest abundance of BV morphotypes receives 4 points and highest abundance of curved morphotypes receives 2 points. Combining points leads to the Nugent score on a 10-point scale, which is collapsed into a three-level BV diagnosis: 0-3 = BV– (BV-negative), 4-6 = BVI (BV-intermediate), 7-10 = BV+ (BV-positive). In this study, complete absence of bacteria is termed BV0 (BV-none).

Sample fixation was carried out with 1 volume (75µl) of fresh, cold 6% w/v paraformaldehyde (Sigma) to 2 volumes (150µl) of fresh, washed sample, eluted from swab head (undiluted or diluted tenfold), and incubated overnight at 4°C. Fixed sample (150µl) was stained with 50µl 0.002% propidium iodide (PI; Sigma) at 37°C for 20min. Prior to flow cytometry, each sample was split into 2 aliquots and 50µl of a solution of yellow-green fluorescent micro-particles (Molecular Probes, Eugene, OR) at a known concentration (~10^6 particles/ml) was added to each aliquot.

Flow cytometry was carried out on a FacsCalibur (Becton-Dickinson, Mississauga, ON) using standardized settings (FSC=E00, SSC=350, FL1=443, FL3=381). Threshold was set on FL3 to exclude acellular debris (particles with no PI staining indicates lack of DNA). Regions were defined for fluorescent micro-particles, bacterial and mucosal cell populations (see Figure 3.8) and total events for each region calculated using FlowJo 8.0 software (TreeStar Inc., Ashland, OR).

Total bacterial cell-units (BCU) and mucosal cell-units (MCU) were calculated using a standard formula in which the known number of micro-particles is multiplied by the observed number of cell-units and then divided by the observed number of micro-particles (185). The "cell-unit" is similar to the "colony-forming unit" concept in which a single event in flow cytometry may actually consist of several cells (for example, chains or clusters) (177). Values are adjusted according to dilution in fixation and staining buffers (two-fold), and expressed as total BCU or MCU per sample. These values were then adjusted to sample weight and expressed as total BCU or MCU per gram sample.

Fresh and previously frozen pilot study samples were used for optimization of flow cytometry procedures (175). For the main study, fresh samples were analyzed in

Nairobi using the above protocol. Analysis was also repeated on a subset of previously frozen samples in order to assess replicability of quantitative values after freeze-thaw.

Semi-quantitative culture on Lactobacillus-selective media – In this study, initial culture of fresh samples in Nairobi was carried out on Rogosa agar (EMD Chemicals, Gibbstown, NJ), prepared according to manufacturer's directions with addition of 0.025% w/v TMB (Sigma) and 0.001% w/v HRP (Sigma), in order to simultaneously assess *Lactobacillus* colonization and H_2O_2 production (Rogosa-TMB). This media is selective for *Lactobacillus* organisms due to final pH of 5.4 after the addition of 0.132% v/v glacial acetic acid (Sigma). Fresh samples (N=242) were washed twice, diluted tenfold in PBS and 100μl was spread-plated onto a single plate of Rogosa-TMB. Plates were incubated anaerobically using the Gaspak anaerobic system (Becton-Dickinson) for 48h at 37°C. Plates were removed from jars and colour allowed to develop for 30min. prior to scoring and photographing of plates. Ordinal scores were assigned based on level of growth (none/low/medium/high) and production of H_2O_2 (negative/weak/strong).

Quantitative culture and strain isolation – In-depth culture-based assays were optimized using a subset of 48 pilot study samples from 16 individuals (1-4 samples per individual), and then applied to a subset of 96 samples ("culture subset") from the Nairobi study. This subset included 32 individuals in each of HIV-R, HIV-N and HIV+ groups, selected for balanced representation of BV diagnoses by Nugent score.

In order to keep culture procedures manageable, it was decided to focus on *Lactobacillus*-selective media only for strain isolation. Since not all samples in initial studies produced growth on Rogosa media, an alternative media was required in order to ensure isolation of strains in all samples. Recently, a new culture media based on Brucella

with TMB/HRP (mBru-TMB) has been optimized to support growth and H_2O_2 production in a wider range of *Lactobacillus* strains (165). Therefore, both Rogosa-TMB and mBru-TMB were used for quantification of colony-forming units and strain isolation.

Instead of random sampling of colonies on plates, the strategy for isolate selection was to pick every colony with a distinct appearance, sampling the greatest diversity of colonies possible based on their appearance. Most studies isolate *Lactobacillus* colonies on Rogosa media without chromogen and then test resulting isolates for H_2O_2 production on Rogosa with chromogen (2, 165, 193). In this study, primary isolation was carried out on media with H_2O_2-sensitive chromogens, allowing for colonies to be selected based on colour as well as shape, increasing types of colonies with a distinct appearance.

Frozen swabs from Winnipeg (N=48) or shipped from Nairobi (N=96) were thawed, vortexed vigorously and squeezed against the side of the tube prior to being discarded. Eluted sample was centrifuged and resuspended in 1ml PBS. Decimal dilutions were prepared by transferring 100µl of eluted sample into 900µl PBS (10X dilution), 100µl of 10X dilution into 900µl PBS (100X dilution), and so on to 100,000X dilution (Figure 2.2). The 10X dilution was used for detection of acid production from mannose by whole sample (see next section). A 50µl aliquot of each dilution was spread-plated on each of 3 types of media: 1) Rogosa (without chromogen), 2) Rogosa-TMB, and 3) mBru-TMB (per litre: 41g Brucella agar (EMD), 20g soluble starch (Becton-Dickinson), 0.86g magnesium sulfate anhydrous (Sigma), 0.18g manganese sulfate monohydrate (Sigma), 1ml 5% w/v hemin (Sigma) dissolved in 1N NaOH (Sigma), 2.5ml 0.5% v/v Vitamin K (Sigma) dissolved in 95% EtOH, 50ml horse serum (Sigma), 250mg TMB and 1ml 1% w/v HRP dissolved in filter-sterilized water).

Plates were incubated for 48h in an anaerobic chamber. Rogosa and Rogosa-TMB plates were used to count colony-forming units (CFU), with the average count of colonies on the two plates, adjusted to dilution level, used as an estimate of total culturable *Lactobacillus* in the sample. Rogosa-TMB and mBru-TMB were used for strain isolation by selecting all colonies with a unique appearance as described above. Up to 10 strains were picked per sample, streaked to Mann-Rogosa-Sharpe agar (MRS; EMD) or mBru (without chromogen) and incubated in an anaerobic chamber at 37°C for 48h.

Determination of acid production from mannose using whole sample – A second mBru media was created by using mannose as the only source of added carbohydrate and bromocresol purple (BCP) to detect acid production by whole samples (mBru-mannose). A simple assay, inspired by the "Snyder test" for estimating acid-producing organisms in saliva (192), was designed in order to estimate levels of organisms able to produce acid by fermenting mannose in vaginal specimens. Aliquots of diluted whole sample (10X) were transferred to a replicator and stamped onto mBru-mannose.

Since commercial Brucella agar contains glucose in its formulation, mBru-mannose was prepared using its constituent ingredients except for glucose; per litre: 10g meat peptone (Becton-Dickinson), 10g tryptone from casein (Becton-Dickinson), 2g yeast extract (Becton-Dickinson), 5g sodium chloride (Sigma), 15g agar (Becton-Dickinson), 25g D+mannose (Sigma), 0.15g BCP (Sigma) and soluble starch, magnesium sulfate, manganese sulfate, hemin, Vitamin K and horse serum as described above.

Each sample was inoculated on 3 plates using a replicator, incubated in an anaerobic chamber at 37°C, and plates removed at 24h intervals to assess colour development at each timepoint (BCP is purple at neutral pH and yellow at pH 4).

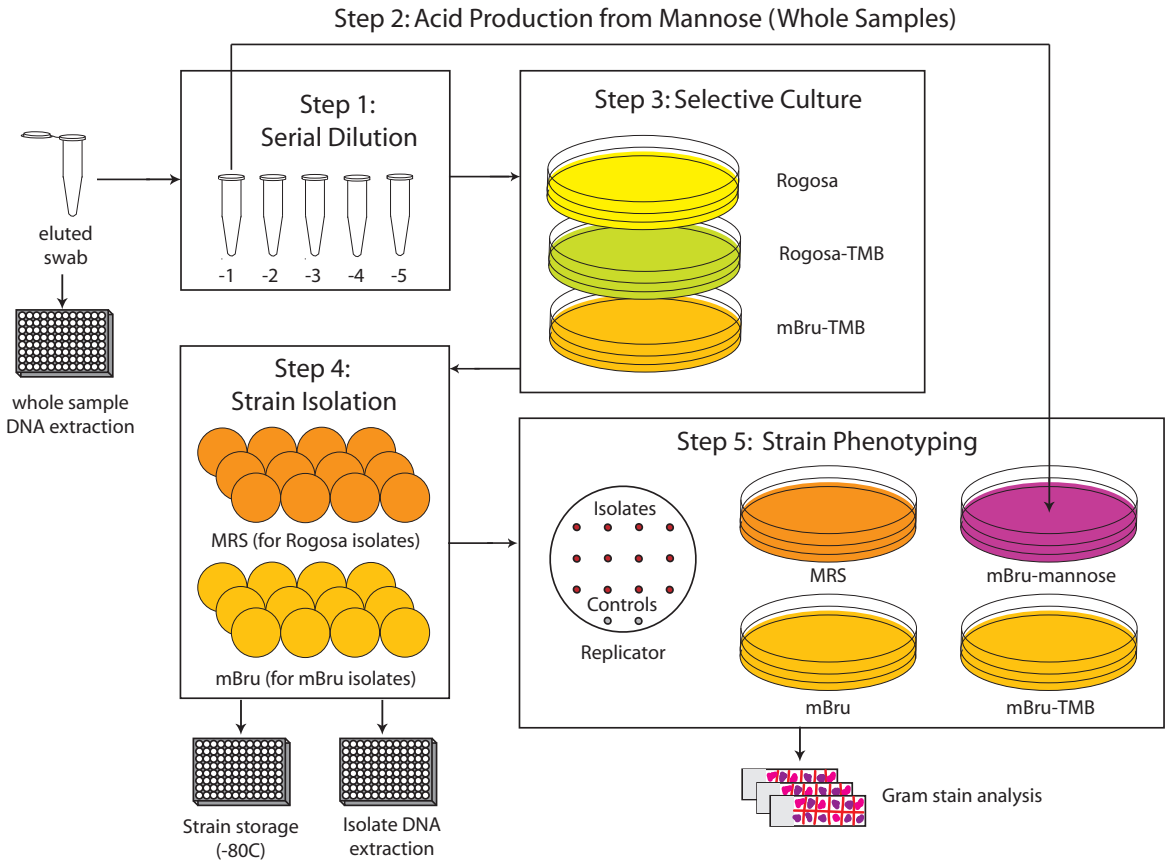

Figure 2.2 – Overview of culture procedure

Vaginal specimens (swabs) were eluted and tenfold serial dilutions prepared in PBS (Step 1). An aliquot of the tenfold dilution was used to assess the capacity of whole sample to produce acid from mannose (Step 2). Appropriate dilutions were plated onto three different media (Step 3). Rogosa plates (with and without chromogen) were used to calculate CFU, while isolated colonies on Rogosa-TMB and mBru-TMB plates with distinct morphology and colour were selected and isolated on MRS (for Rogosa-TMB isolates) and/or mBru (for mBru-TMB isolates) (Step 4). Phenotype of selected isolates was assessed by spotting on MRS, mBru, mBru-TMB and mBru-mannose (Step 5). DNA was extracted from whole samples and isolates for downstream analyses, isolated strains were stored in skim milk/ glycerol, and Gram stains of isolates were prepared and photographed.

Resulting spots were scored on a 4 point scale in which 0 is no yellow colour and 3 is a very bright yellow halo surrounding the spot. Controls were not included on each plate, but were included on identical plates incubated simultaneously (see next two sections).

Panel of reference organisms – Phenotypic assays were optimized and control strains selected using three sets of relevant organisms selected from culture collections (Table 2.1, A), probiotic strains (Table 2.1, B), and strains isolated from Nairobi samples in preliminary studies (Table 2.1, C).

Phenotypic characterization of isolates – For isolates, growth from isolation plates (Figure 2.2, Step 4) was washed off with 1ml PBS and adjusted to 0.5 McFarland in 3ml PBS. Aliquots were transferred to replicator plates and stamped onto 4 different media: MRS, mBru, mBru-mannose and mBru-TMB (as for mBru-mannose, except with 12.5g D+mannose (Sigma), 12.5g D+glucose (Sigma), and HRP/TMB at concentrations described above). Plates were incubated 37°C for 48h in an anaerobic chamber.

After removal, colour was allowed to develop for 30min., after which plates were scored and photographed. Growth on MRS and mBru was scored as positive or negative. H_2O_2 production on mBru-TMB was scored as either positive (blue) or negative (no colour). Acid production on mBru-mannose was scored on a 4-point scale based on intensity of yellow colour as for whole samples, as described above. A Gram stain of each isolate was prepared and photographed using standard procedures, and an aliquot stored at –80°C in an equal volume of 2X freezing media (4% w/v skim milk, 1% D+ glucose and 20% v/v glycerol) (all from Sigma).

Table 2.1 – Panel of reference organisms

A. Culture collection strains

Organism	Strain[1]	Media[2]	Source[3]
Lactobacillus crispatus	DSMZ 20584	MRS	DSMZ
Lactobacillus gasseri	DSMZ 20243	MRS	DSMZ
Lactobacillus iners	DSMZ 13335	CBA	DSMZ
Lactobacillus jensenii	DSMZ 20557	MRS	DSMZ
Lactobacillus vaginalis	DSMZ 05837	MRS	DSMZ
Atopobium vaginae	DSMZ 15829	CBA	DSMZ
Enterococcus faecalis	ATCC 29212	MRS	W. Smoragiewicz, UQAM
Lactococcus lactis	ATCC 11454	MRS	W. Smoragiewicz, UQAM
Peptostreptococcus anaerobius	ATCC 27337	BHI	M. Alfa, SBGH
Streptococcus agalactiae	ATCC 13813	MRS	M. Alfa, SBGH
Streptococcus bovis	ATCC 49147	MRS	M. Alfa, SBGH
Gardnerella vaginalis	ATCC 14018	CA	M. Alfa, SBGH
Streptoccocus anginosus	ATCC 33397	BA	M. Alfa, SBGH
Candida albicans	ATCC 10231	SAB	D. Collins, HSC
Staphylococcus aureus	ATCC 25923	BA	D. Collins, HSC
Escherichia coli	BL-21	BA	G. McClarty, UM

[1] DSMZ: German culture collection, ATCC: American Type Culture Collection, [2] Media for routine culture; MRS: Mann-Rogosa Sharpe (EMD), CBA: Columbia Blood Agar (EMD), BHI: Brain-Heart Infusion (Difco), CA: Chocolate Agar, BA: Blood Agar, SAB: Sabouraud's agar (Clinical Microbiology Laboratory, Health Sciences Centre). [3] Strains purchased from DSMZ or kindly provided by the individual named; UQAM: Universite du Quebec a Montreal, SBGH: Clinical Microbiology Lab, St. Boniface General Hospital, HSC: Clinical Microbiology Lab. Health Sciences Centre, UM: University of Manitoba.

(continued…)

Table 2.1 (continued):

B. Probiotic strains

Organism	Strain	Media[1]	Source[2]
Lactobacillus acidophilus	8/4	MRS	W. Smoragiewicz, UQAM
Lactobacillus delbrueckii	151	MRS	W. Smoragiewicz, UQAM
Lactobacillus salivarius	AWH	MRS	W. Smoragiewicz, UQAM
Bifidobacterium animalis	B30	MRS	W. Smoragiewicz, UQAM
Lactobacillus rhamnosus	GG	MRS	G. Reid, UWO
Lactobacilllus rhamnosus	GR1	MRS	G. Reid, UWO
Lactobacillus fermentum	RC14	MRS	G. Reid, UWO

[1] Media used for routine culture, as for Table 1A, [2] Strains were kindly provided by the individual named; as for Table 1A, UWO: Lawson Health Research Institute, University of Western Ontario

C. Nairobi isolates

Organism[1]	Strain	Media[2]	Source[3]
Streptococcus sp.	S12	MRS	S. Iqbal, UM
Enterobacteria sp.	E7	BA	S. Iqbal, UM
Neisseria gonorrhea	G45	CA	S. Iqbal, UM
Lactobacillus sp.	L6	MRS	S. Iqbal, UM
Lactobacillus salivarius	N3	MRS	This study
Lactobacillus gasseri	N7	MRS	This study
Lactobacillus plantarum	N17	MRS	This study
Lactobacillus sp.	N19	MRS	This study
Lactobacillus sp.	N23	MRS	This study
Lactobacillus jensenii	N24	MRS	This study
Lactobacillus sp.	N27	MRS	This study
Lactobacillus sp.	N29	MRS	This study
Bifidobacterium infantis	N10	MRS	This study
Bifidobacterium pullorum	N33	MRS	This study
Streptococcus lutetiensis	N9	MRS	This study
Streptococcus bovis	N11	MRS	This study

[1] S12, E7 and G45 were identified phenotypically using routine procedures. Remaining organisms were identified by sequencing *cpn*60 UT (see Methods), [2] Media used for routine culture; as for Table 1A, [3] Strains were isolated as part of this study or kindly provided by the individual named; as for Table 1A.

Ribosomal interspacer analysis of cultured isolates – Initially, ISR profiles were generated for several reference strains (see previous section), followed by isolates from the pilot study and the Nairobi study. Genetic material was isolated from aliquots of isolate cells using Instagene Matrix (BioRad), according to the directions of the manufacturer. Briefly, a 200µl aliquot of washed sample in PBS was centrifuged, the supernatant discarded, and the pellet resuspended in 200µl Instagene Matrix. The suspension was vortexed, incubated at 56°C for 30min., vortexed, incubated at 100°C for 8min., vortexed and centrifuged. Supernatant containing isolate DNA was removed to a separate tube and used for subsequent molecular analyses.

Amplification of the ISR was carried out using published primers and procedures (40). Briefly, PCR is carried out using forward and reverse primers targeted to positions near the end of the 16S gene and beginning of the 23S gene respectively (ITS-F/R, Table 2.2) (40). All primers used in this study were synthesized in-house (DNA Core, National Microbiology Laboratory). PCR was carried out with 2µl template DNA in a reaction mixture containing 5µl 10X PCR buffer, 2.5µl 50mM $MgCl_2$, 1µl 10mM dNTP mixture, 1µl 25uM each primer and 0.5µl Taq polymerase (Invitrogen, Burlington, ON) and water to a final reaction volume of 50µl. Cycling conditions were as follows: 95°C/2 min., followed by 40 cycles of 95°C/45s, 55°C/1min., 72°C/2min., and a final extension of 7min./72°C. Resulting fragments include 99bp from 16S and 22bp from 23S.

In order to precisely size the ISR, the 5' end of the reverse primer is modified with the phosphoramidite dye HEX (6-carboxy-1,4-dichloro-2',4',5',7'-tetra-chlorofluorescein) for fragment detection by capillary electrophoresis in an ABI 3100 sequencer. Amplicons were purified by technical staff using an automated MagnaPure system (DNA Core,

National Microbiology Laboratory), diluted 20-fold in water, and 2µl mixed with 15µl HiDi formamide (ABI) and 3µl X-Rhodamine (ROX)-labelled MapMarker 1000 size standard (Bioventures), with 23 fragments covering a range of 50-1000bp.

The reaction mixture was denatured 5min. at 95°C, and immediately placed on ice until electrophoresis. Amplicon reactions were run on the ABI 310 genetic analyzer using POP7 polymer and standard denaturing electrophoresis settings for approximately 1.5h. The resulting electropherograms were processed and amplicons sized in relation to the internal size standard using GeneMarker (SoftGenetics, State College, PA).

Isolates with similar electropherograms were grouped together using Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm, source code by Michael Eisen) (55) and results visualized using Java TreeView v1.1.4r3 (http://jtreeview.sourceforge.net, created by Alok Saldanha) (173). Representatives from each cluster were selected for species identification by sequencing of the *cpn*60 universal target (see next section), and data used to address Hypotheses 7-13.

cpn60 UT-based sequencing and identification of isolates – Sequencing of the ~1.5kb 16S ribosomal gene (rDNA) is the most frequently reported method for identifying bacterial isolates. In this study, sequencing of the alternative chaperonin-60 universal target (UT) region was used instead. This alternative molecular target for microbial identification and phylogenetic analysis, a gene encoding the 60 kDa chaperonin or heat shock protein (also known as HSP60 and GroEL), is increasingly used for identification of isolates and to characterize complex microbial communities (87).

The *cpn*60 gene is nearly universal in Eubacteria and Eukaryotes, and the phylogenetically informative targetted region offers key advantages relative to 16S rDNA.

Table 2.2 – PCR primers used in this study

Purpose	Name	Sequence	Source
RISA[1]			
	ITS-F	GTC GTA ACA AGG TAG CCG TA	(40)
	ITS-R	GCC AAG GCA TCC ACC	
Isolates[2]			
	H729-F	**CGC CAG GGT TTT CCC AGT CAC GAC** GAI III GCI GGI GAY GGI ACI AC[3]	(86)
	H730-R	**AGC GGA TAA CAA TTT CAC ACA GGA** YKI YKI TCI CCR AAI CCI GGI GCY TT[3]	
Clone libraries			
	H279-F	GAI III GCI GGI GAY GGI ACI AC[3]	(85)
	H280-R	GAI III GCI GGI GAY GGI ACI AC[3]	
	H1612-F	GAI III CGI GGY GAC GGY ACS ACS AC[3]	(88)
	H1613-R	CGR CGR TCR CCG AAG CCS GGI GCC TT[3]	
qPCR, Target :			
L. crispatus	Lcp-F	TCC TTA CAT TTT GAT CAC TGA	(52)
	Lcp-R	GAG CTT CAC CAG TAA CGT C	
L. gasseri	Lga-F	ACA AGG TAA GCA CTA AGG ATG	(52)
	Lga-R	CCC TTT GAT TCC TCA ATT GTG	
L. iners	Lin-F	GCT TCT GTT TCT TCA GCA TCA	(53)
	Lin-R	TCG TGT CCT ACT CGT TCC ATT	
A. vaginae	Ava-F	ACA AGG TAA GCA CTA AGG ATG	(53)
	Ava-R	CCC TTT GAT TCC TCA ATT GTG	
G. vaginalis	Gva-F	GCT TCT GTT TCT TCA GCA TCA	(138)
	Gva-R	TCG TGT CCT ACT CGT TCC ATT	

[1] Ribosomal interspacer analysis, [2] Bold indicates M13 (–40)F and M13 (48)R sequencing primer landing sites, [3] Degenerate primer base definitions: I (inosine): all bases, Y: C or T, R: A or G, K, T or G, S, C or G.

These include shorter target length (549-567 bp), increased resolution to distinguish closely related species and subspecies, and a relatively uniform distribution of variability across the entire length of the target (73, 86). Further, an extensive, curated reference database is available to facilitate accurate identification of bacteria from which sequences are generated (cpnDB: http://cpndb.cbr.nrc.ca).

The *cpn*60 UT was amplified based on published primers and procedures (86). Briefly, target was amplified from isolate DNA using the degenerate sequencing primers H729-F/H730-R (Table 2.2, p.55), incorporating the M13 (–40)F and M13 (48)R sequencing primer landing sites. PCR was carried out with 2µl template DNA in a reaction mixture containing 5µl 10X PCR buffer, 2.5µl 50mM $MgCl_2$, 1µl 10mM dNTP mixture, 1µl 25uM each primer, 0.5µl Taq polymerase (Invitrogen) and water to a final reaction volume of 50µl. Cycling conditions were as follows: 95°C for 2 min., followed by 40 cycles of 95°C/30s, 42°C/30sec., 72°C/30sec., and a final extension of 2min. at 72°C.

Amplicons were purified using an automated MagnaPure system and sequenced in both directions using an ABI3700 genetic analyzer (DNA Core, National Microbiology Laboratory). Resulting sequences were assembled and trimmed using Lasergene 8 software (DNAStar, Madison, WI). Full-length UT (~552bp) were compared to cpnDB to determine nearest neighbour using FASTA queries (154).

Species-specific quantitative PCR – Bacteria in vaginal swab DNA extracts were quantified with species-specific UT-targeted PCR primers (Table 2.2, p.54), as described previously (52). Briefly, oligonucleotide sequences differentiating target *cpn*60 from related organisms were identified using Signature Oligo (LifeIntel Inc., Port Moody, BC). Compatible primers were determined using Beacon Designer (Sigma-Genosys Canada,

Oakville, ON). Specificity of primers was verified by comparison to cpnDB using BLASTn (5), followed by verification of no PCR product using template DNA derived from related strains (T. Dumonceaux).

Standard curves for determination of concentration of 5 phylotypes (*L. crispatus*, *L. gasseri, L. iners, G. vaginalis* and *A. vaginae*) in vaginal specimens were determined using serial dilutions of cloned UT plasmids, as described previously (51). *L. crispatus* and *L. gasseri* were enumerated using the Applied Biosystems plaform (2 μl of template DNA, 1 μl SYBR green master mix (Applied Biosystems, Foster City, CA) and 500 nM of each primer in a final volume of 25 μl). The reactions were quantified with an ABI 7500 Fast system (Applied Biosystems). Quantification of remaining targets was carried out on the Roche platform (2 μl template, 1μl LC480 SYBR green master mix (Roche, Laval, QC) and 500nM each primer in a final volume of 20μl), with reactions quantified on a LightCycler 480 (Roche). The results were expressed as the number of genomes detected per reaction, adjusted for sample weight as described for flow cytometry measures, and expressed as genomes per gram sample.

Construction of cpn60 UT-based clone libraries – All procedures used were based on published studies (85). Whole-sample DNA extract (2μl) was used as template in a PCR reaction containing a 1:3 mixture of two primer sets (H279/280 and H1612/1613) shown to increase amplification of high-GC organisms within the mixed commmunity, as described previously (88) (Table 2.2, p.54). PCR was carried out with 2μl template DNA in a reaction mixture containing 5μl 10X PCR buffer, 2.5μl 50mM $MgCl_2$, 1μl 10mM dNTP mixture, 0.25μl 25uM each H279/280 primers, 0.75μl 25uM each H1612/1613 primers, and 0.5μl Taq polymerase (Invitrogen) and water to a final reaction volume of 50μl.

Since members of a complex microbial community are likely to have different optimal annealing temperatures (85), amplifications from the same sample were run at 4 different annealing temperatures. A total of 8 reactions were performed for each sample, two each at 42°C, 46.5°C, 50.4°C and 60°C, in order to improve template representation. Cycling conditions were as follows: 95°C/2min., followed by 40 cycles of 95°C/30sec., annealing temp./30sec., 72°C/30sec., and a final extension of 2min./72°C. Amplicons were analyzed by gel electrophoresis in 2% agarose to ensure correct size (~600bp).

All reactions for a single sample were pooled and purified using the QIAquick PCR purification kit (QIAGEN, Germantown, MD) according to manufacturer's directions, followed by an additional purification step by gel electrophoresis in 2% agarose. Bands of the correct size were cut from the agarose slab and purified using the QIAEX II Agarose Gel Extraction Kit (QIAGEN) according to manufacturer's directions.

Concentrations of purified DNA were measured by Nanodrop and A-tailed in a reaction mixture including 50-60ng DNA with 1μl 10X PCR buffer, 0.6μl $MgCl_2$, 1μl 10mM dNTP mixture and 0.4μl Taq polymerase (Invitrogen) in a final reaction volume of 10μl, incubated at 72°C for 10min. A-tailed product (insert) was ligated into the cloning vector from the TOPO TA Cloning Kit for Sequencing (Invitrogen) according to the manufacturer's directions. Briefly, 4μl A-tailed product was incubated with 1μl TOPO vector and 1μl salt solution (from kit) at room temperature for 30min. Transformation of One-Shot Chemically Competent Cells (Invitrogen) was carried out by adding 2μl ligation reaction to one tube of thawed cells and incubating for 30min. on ice, followed by heat shock at 42°C for 30sec. and addition of 250μl SOC medium (from kit).

Transformed cells were allowed to recover for 2-3 hours by incubating at 37°C, followed by spread-plating of 50-100µl on LB-kanamycin agar (1% w/v tryptone, 0.5% w/v yeast, 1% w/v sodium chloride, 1.5% w/v agar, 0.05% w/v kanamycin, National Microbiology Laboratory) that had been previously spread with 100µl X-gal/IPTG solution (0.55g X-gal and 0.06g IPTG in 10ml dimethyl formamide) (all from Sigma). Plates were incubated overnight at 37°C and white colonies picked into 96-well plates containing 100µl LB-kanamycin broth (as above, excluding agar).

After incubation overnight at 37°C, 100µl of LB-kanamycin broth with 14% v/v DMSO (Sigma) was added to each well and plates stored at –80°C until sequencing. Preparation of clones for sequencing was carried out using the Illustra Templiphi Amplification Kit (GE Healthcare, Baie d'Urfe, PQ), using 1/4 of reagent volumes recommended by the manufacturer. Briefly, 0.5-1µl of thawed clone culture was added to 1.2µl of sample buffer (from kit) in 96-well sequencing plates, centrifuged briefly, incubated at 95°C for 3 minutes, and 1.2µl of enzyme/reaction buffer mixture (from kit) added to each well. Plates were centrifuged briefly and incubated overnight at 30°C, followed by enzyme inactivation at 65°C for 10min. and sequencing of cloned insert using T7 primers with the ABI3700 genetic analyzer at the DNA Core sequencing facility.

Resulting raw sequence data was quality-checked and assembled into contigs as previously described (85), using PreGap4 (v1.1) and Gap4 (v4.6) in the Staden software package (http://staden.sourceforge.net, release 2000.0, J. Bonfield, K. Beal, M. Betts, M. Jordan and R. Staden, 2000). Polished, trimmed contigs were compared to cpnDB and the nearest neighbour for each contig established as described above.

Clone library data was not used to directly address study hypotheses since only a small number of individuals were examined using this technique. However, clone library sequences were used to greatly enhance the *cpn*60 database for deep sequencing of a larger number of samples (next section).

Ultra-deep GS-FLX amplicon sequencing of cpn60 UT – Preparation of *cpn*60 UT amplicon from individual samples was carried out using a similar protocol to that used for clone libraries, except that primers (H279/280 and H1612/13) were modified so that GS-FLX emPCR primers were incorporated into amplicon upstream (5') of the forward primer and downstream (3') of the reverse primer. As well, a 4bp sequence tag unique to each sample was incorporated into forward primers (H279 and H1612) between the UT primer and the GS-FLX primer (Table 2.3).

This strategy allowed for multiplexing of 16 separate samples in each of 2 regions on the sequencing chip, avoiding the use of gaskets to separate the samples and dramatically increasing the number of wells available for sequencing reactions. PCR reactions were carried out as described for clone libraries, except that a total of 12 reactions were prepared for each sample, in each row of a 96-well plate, and amplification was carried out using a gradient from 42-60°C for the annealing temperature.

Amplicons were purified using an automated MagnaPure system, and then pooled, gel-purified and DNA extracted from gel slices as described for clone libraries. Sequencing libraries were prepared from purified amplicon using the GS DNA library preparation kit and emulsion PCR (emPCR) kit I, designed to sequence beginning from the forward primer only to capture the sample-unique 4bp tag in every read (Roche).

Table 2.3 – Tagged deep sequencing primers for multiplex GS-FLX

Primer	emPCR primer[1]	Tag	*cpn*60 UT primer
P279gatg:	GCCTCCCTCGCGCCATCAGAGATGC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279gctg:	GCCTCCCTCGCGCCATCAGAGCTGC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279gtga:	GCCTCCCTCGCGCCATCAGAGTGAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279ctct:	GCCTCCCTCGCGCCATCAGACTCTC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279ctga:	GCCTCCCTCGCGCCATCAGACTGAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279caga:	GCCTCCCTCGCGCCATCAGACAGAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279tctg:	GCCTCCCTCGCGCCATCAGATCTGC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279tgag:	GCCTCCCTCGCGCCATCAGATGAGC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279tgca:	GCCTCCCTCGCGCCATCAGATGCAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279tagt:	GCCTCCCTCGCGCCATCAGATAGTC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279caca:	GCCTCCCTCGCGCCATCAGACACAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279gaga:	GCCTCCCTCGCGCCATCAGAGAGAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279gtgt:	GCCTCCCTCGCGCCATCAGAGTGTC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279gaca:	GCCTCCCTCGCGCCATCAGAGACAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279gtct:	GCCTCCCTCGCGCCATCAGAGTCTC	GAIIIIGCIGGIGAYGGIACIACIAC	
P279tcta:	GCCTCCCTCGCGCCATCAGATCTAC	GAIIIIGCIGGIGAYGGIACIACIAC	
P280:	GCCTTGCCAGCCCGCTCAG	------	YKIYKITCICCRAAICCIGGIGCYTT
P1612gatg:	GCCTCCCTCGCGCCATCAGAGATGC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612gctg:	GCCTCCCTCGCGCCATCAGAGCTGC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612gtga:	GCCTCCCTCGCGCCATCAGAGTGAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612ctct:	GCCTCCCTCGCGCCATCAGACTCTC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612ctga:	GCCTCCCTCGCGCCATCAGACTGAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612caga:	GCCTCCCTCGCGCCATCAGACAGAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612tctg:	GCCTCCCTCGCGCCATCAGATCTGC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612tgag:	GCCTCCCTCGCGCCATCAGATGAGC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612tgca:	GCCTCCCTCGCGCCATCAGATGCAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612tagt:	GCCTCCCTCGCGCCATCAGATAGTC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612caca:	GCCTCCCTCGCGCCATCAGACACAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612gaga:	GCCTCCCTCGCGCCATCAGAGAGAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612gtgt:	GCCTCCCTCGCGCCATCAGAGTGTC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612gaca:	GCCTCCCTCGCGCCATCAGAGACAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612gtct:	GCCTCCCTCGCGCCATCAGAGTCTC	GAIIIIGCIGGYGACGGYACSACSAC	
P1612tcta:	GCCTCCCTCGCGCCATCAGATCTAC	GAIIIIGCIGGYGACGGYACSACSAC	
P1613:	GCCTTGCCAGCCCGCTCAG	------	CGRCGRTCRCCGAAGCCSGGIGCCTT

[1] emPCR primer: Sequence on left is primer for emulsion PCR, sequence in center is tag unique to each sample, and sequence on right is *cpn*60 UT-specific.

Amplicons were treated as "sheared" DNA so no nebulization step was required. Titration of amplicon concentration was carried out to ensure optimal ratio of DNA concentration to sequencing beads. Beads were loaded onto the sequencing chip and the run completed using standard kits and procedures (G. Peters, DNA Core, National Microbiology Laboratory).

Deep sequencing data was processed with default procedures from 454/Roche. Only filter-passed reads were used in subsequent analyses. Sequencing data (sff files) were imported and warehoused using the APED software package (http:/aped.sourceforge.net). Resulting sequence data was sorted based on the 4bp tag in order to distinguish individual samples in the multiplexed run.

Two full deep sequencing runs were completed during this project, resulting in libraries for 48 samples from the culture subset (Pyro subset), including 16 samples each from HIV-R, HIV-N and HIV+ groups. A number of control experiments were also conducted, including technical replicates of two samples, and 16S ribosomal DNA amplicons generated from a pool of four samples to compare to *cpn*60 UT deep sequencing libraries. 16S amplicons were generated using untagged PCR primers L27F and 355R targetting variable regions 1 and 2 (not shown) (194), pooled and purified as desribed above, and ligated to PCR linkers with a Roche multiplexing ID sequence prior to deep sequencing (G. Peters).

Taxonomic assignment – A non-redundant, customized database ("vagnr") of *cpn*60 UT sequences was created by J. Hill, comprising a single reference strain for each species in cpnDB as well as unique sequences from 25 cultured isolates from vaginal swabs of Nairobi individuals (1,373 sequences). In later analyses, the vagnr database was

augmented with 132 additional sequences derived from isolates and clone libraries ("vagnr+132" database, 1,505 sequences). For 16S analysis, a database of 66,304 near full-length sequences (>1,200bp) was extracted from the RDP (125)

Phylotype identity based on deep sequencing was established by combining a number of approaches (bioinformatics coordinated by M. Links). Initially, a "watered-BLAST" approach was used (176), combining BLAST (5) and Smith-Waterman alignments (191) to identify best matches between individual pyrosequencing reads and reference sequence databases. The output of this analysis is a list of phylotypes observed in each sample, the percent identity of each phylotype to its nearest neighbour in the vagnr+132 database, and the number of times it was observed in the sample (ie. number of deep sequencing reads). The percent of total reads per sample was calculated for each phylotype (reads/100). Total reads per sample were also adjusted to 50,000 reads and the relative abundance of each phylotype calculated (reads/50,000). Only reads longer than 150bp and >70% identical to database sequences were retained for downstream analyses.

Subsequently, all deep sequencing reads were assembled using Roche/454 software. Assemblies were compared to the vagnr+132 database using BLAST and a similar output to the previous analysis was generated, including a list of phylotypes present in each sample and the relative abundance of each phylotype, adjusted as described above. Assemblies not matching any database sequence at >70% were excluded from analysis.

Ecological analysis of deep sequencing data – Phylotype density (an estimator of species "richness") was defined as the total number of phylotypes within an individual sample (77). Proportion of total reads per sample represented by each phylotype were used to calculate overall diversity and evenness of a sample using the Shannon Diversity index

(*H'*, eq.1) and the Shannon Equitability Index ($E_{H'}$, eq. 2):

$$H' = -\sum_{i=1} (p_i)(\ln p_i) \qquad \text{eq.1}$$

$$E_{H'} = H' / \ln(S) \qquad \text{eq. 2}$$

where p_i is the relative abundance of the *i*th phylotype and S is the total number of phylotypes (77).

Shared community membership between samples and groups was assessed using the Jaccard index value (J_{clas} , eq. 3):

$$J_{clas} = \frac{S_{12}}{S_1 + S_2 - S_{12}} \qquad \text{eq. 3}$$

where S_1 and S_2 are the number of phylotypes in Group/Sample 1 vs. Group/Sample 2 respectively, and S_{12} is the number of phylotypes observed in both groups/samples (179).

Shared community structure between samples and groups, taking into account similarities in the relative abundance of each phylotype, was assessed using the Theta Index Value (θ, eq. 4):

$$\theta = \frac{\sum_{i=1}^{S_{12}} \frac{X_i}{n_{\text{total}}} \frac{Y_i}{m_{\text{total}}}}{\sum_{i=1}^{S_1} \left(\frac{X_i}{n_{\text{total}}}\right)^2 + \sum_{i=1}^{S_2} \left(\frac{Y_i}{m_{\text{total}}}\right)^2 - \sum_{i=1}^{S_{12}} \frac{X_i}{n_{\text{total}}} \frac{Y_i}{m_{\text{total}}}} \qquad \text{eq.4}$$

Where X_i and Y_i are the abundance of the *i*th phylotype in Group/Sample 1 and Group/Sample 2 respectively, and n_{total} and m_{total} are the total numbers of sequences

sampled in Group/Sample 1 and Group/Sample 2 respectively (179). Both Jaccard and Theta indices are on a scale of 0 to 1, with 0 indicating no shared phylotypes and 1 indicating all shared phylotypes.

Phylogenetic trees and bootstrapping – Phylogenetic trees of isolates, clone libraries and pyrosequencing reads were created using the neighbour-joining method (172) and bootstrap values calculated for each node based on 1000 assembled trees (59). Nodes with bootstrap values greater than 95% (robust) and 50% (valid) are indicated with filled circles and open circles respectively. Nodes with bootstrap values <50% are considered invalid. Trees are drawn to scale, with branch lengths proportional to evolutionary distances as calculated using the Maximum Composite Likelihood method (203). For pyrosequencing reads, phylogenetic trees were created based on full-length UT sequences of nearest neighbours in the vagnr+132 database, not on the reads themselves. All phylogenetic analyses were conducted using MEGA4 (www.megasoftware.net) (202).

Significance tests – In order to measure the statistical significance of associations between categorical, ordinal or continuous predictor and response variables, appropriate parametric and non-parametric significance tests based on linear regression, between group comparisons and contingency analyses were conducted in GraphPad Prism v5. Mean and standard deviation are reported only for variables with a normal distribution as assessed using the d'Agostino-Pearson omnibus normality test in Prism. Parametric tests (t test and ANOVA) were conducted only if data for each level were normally distributed and variances were not significantly different between levels (assessed using the F test in Prism). Otherwise, median values are reported and non-parametric Mann Whitney and Kruskal Wallis tests used to assess significance of differences between levels.

Hierarchical Linear Modelling – In order to analyze the effect of repeated sampling in the context of participant-level variables, multi-level logistic regression modelling was conducted using the HLM 6 Hierarchical Linear and Nonlinear Modelling software package (M. Cheang). In the pilot study, data were collected at multiple timepoints, describing participant attributes (age, birth control method) that did not vary at each timepoint (N=32) as well as participant-level sample attributes (ie. time since last menstruation) that did vary at each timepoint (N=103). This mixed-effect modelling accounts for the influence of participant variability on repeated sample-level measures, and has previously been described for analysis of vaginal microbial communities (169). One advantage of hierarchical linear modelling, as compared to other analytical techniques for repeated measures, is that there is no requirement for a balanced number of samples per participant or constant time interval between samples. This technique was re-applied in the retrospective Nairobi dataset (3538 samples from 1005 individuals collected over a 4 year period), in order to control for the effect of interpersonal variability in the context of repeated measures.

C. RESULTS

1. Description of study groups in Winnipeg and Nairobi

Characteristics of biological and behavioural descriptors with potential impact on vaginal microbiota were assessed through a self-completed survey in the pilot study and by extracting relevant information from the project database in the Nairobi study. The database consists of several hundred variables generated from biological samples and in-depth surveys collected at the time vaginal specimens were collected.

Profile of pilot study group – Characteristics of young women participating in the pilot study were elicited using a self-completed survey (Table 3.1). Most participants provided 3-4 samples over the six-week study period, allowing for longitudinal assessment of vaginal microbiology in this group. Characteristics included participant-level descriptors that did not change between sampling intervals, such as age and method of birth control, as well as sample-level participant descriptors specific to each sampling timepoint, such as length of time since last menstrual period (LMP) and last sexual intercourse.

The median age of the study group was 15 (range 13-18). Most reported using tampons as menstrual protection. Three quarters of participants reported previous sexual intercourse and all but one of these reported current hormonal contraceptive use, including combined hormonal (oral, transdermal, or intravaginal) or progesterone only (Depo-Provera). Many who used Depo-Provera reported longer than 4 weeks since LMP at all sampling timepoints, which is characteristic of delayed menstruation caused by Depo-Provera use. Less than 20% of the participants reported ever having douched, and all of those who did report douching were in the sexually active group.

Table 3.1 – Selected participant characteristics (Pilot study)

	Participants (N=32)	**Samples** (N=103)
Samples provided		
1-2	3 (9%)	5 (5%)
3-4	26 (81%)	82 (80%)
5-6	3 (9%)	16 (15%)
Age		
13-14	6 (19%)	22 (21%)
15-16	18 (56%)	52 (50%)
17-18	11 (34%)	29 (28%)
Type of menstrual protection		
Tampons	22 (69%)	70 (68%)
Pads	7 (22%)	23 (22%)
Both	3 (9%)	10 (20%)
Last menstrual period		
At time of sample		16 (16%)
In past week		20 (19%)
In past 2 weeks		21 (20%)
In past 3 weeks		15 (15%)
In past 4 weeks		10 (10%)
More than 4 weeks		20 (19%)
Missing		1 (1%)
Last sexual intercourse		
Never	8 (25%)	25 (28%)
In last 24h		21 (20%)
In last week		28 (27%)
In last month		18 (17%)
More than 1 month since		8 (8%)
Missing		2 (2%)
Birth control		
None	9 (28%)	30 (29%)
Oral (the Pill)	9 (28%)	23 (22%)
Depo-Provera	8 (25%)	27 (26%)
Transdermal	6 (18%)	16 (16%)
Intra-vaginal ring	1 (3%)	3 (3%)
Douching		
Never	26 (81%)	76 (74%)
In last 24h		12 (12%)
More than 24h since		9 (9%)
Missing		4 (4%)

Profile of Nairobi main study group and subsets – In-depth profiles of Majengo cohort members providing vaginal swab samples (Main group, N=242) were derived from a database of survey and clinical information collected at the time of sample collection by study staff at the Majengo clinic (Table 3.2).

Based on the main group, several subsets of individuals were chosen for in-depth analysis using a variety of techniques. First, a group of individuals was selected for in-depth culture-based analysis (Culture subset, N=96). Selection of this subset was biased in three ways. First, since the main study hypothesis addresses HIV– individuals, HIV+ individuals were purposefully under-represented in the Culture subset relative to the Main group (see below). Second, since microscopic diagnosis of BV was determined to be a critical parameter describing vaginal microbiota (see next chapter), an effort was made to balance representation of BV– vs. BV samples. Third, selection was biased towards younger individuals, especially from the HIV-R group, in order to reduce the significant age difference between HIV-R and HIV-N individuals (see below).

All other subsets were selected from the culture subset, including a group of 40 individuals selected for qPCR analysis, (qPCR subset, N=40), 10 individuals used for construction of clone libraries (Clone subset, N=10), and a group of 48 individuals selected for ultra-deep sequencing analysis (Pyro subset, N=48).

Nearly half of the main group is HIV+, but only one third of the Culture and Pyro subsets, 25% of the qPCR subsets and none of the Clone subset is HIV+. Of HIV– individuals, half are HIV-R in the Main group, Culture and Pyro subsets, while approximately two thirds are HIV-R in the qPCR and Clone subsets. These differences are due to intentional selection of individuals in these categories as described above.

Table 3.2 – Selected participant characteristics (Nairobi study)

	Main (N=242)	**Culture** (N=96)	**Pyro** (N=48)	**qPCR** (N=40)	**Clone** (N=10)
Serostatus					
HIV–	118 (48%)	64 (67%)	32 (67%)	30 (75%)	10 (100%)
HIV+	124 (52%)	32 (33%)	16 (33%)	10 (25%)	--
Serogroup					
(HIV– only)	*N=118*	*N=64*	*N=32*	*N=30*	*N=10*
HIV-N	57 (48%)	32 (50%)	16 (50%)	10 (33%)	4 (40%)
HIV-R	61 (52%)	32 (50%)	16 (50%)	20 (67%)	6 (60%)
Months in cohort					
Median	36	28	29	41	52
Range	0-259	0-236	0-209	0-207	9-185
Age					
Median	39	34	33	33	35
Range	21-60	21-56	21-55	24-50	25-45
Current CSW[1]					
Active	218 (90%)	87 (91%)	44 (92%)	35 (87%)	9 (90%)
Stopped	24 (10%)	9 (9%)	4 (8%)	5 (13%)	1 (10%)
CSW[1] clients per day					
(Active only)	*N=218*	*N=87*	*N=44*	*N=35*	*N=9*
Median	4	4	4	3	5
Range	1-18	1-18	1-10	1-10	1-10
Missing	7 (3%)	1 (1%)	--	--	--
Condoms for CSW[1]					
(Active only)	*N=218*	*N=87*	*N=44*	*N=35*	*N=9*
Always	202 (93%)	81 (93%)	42 (95%)	34 (97%)	8 (89%)
Not always	12 (6%)	5 (6%)	2 (5%)	1 (3%)	1 (11%)
Missing	4 (2%)	1 (1%)	--	--	--
Menopause					
Yes	17 (7%)	3 (3%)	1 (2%)	--	--
No	221 (91%)	92 (96%)	47 (98%)	40 (100%)	10 (100%)
Missing	4 (2%)	1 (1%)	--	--	--
Pregnant					
Yes	3 (1%)	2 (2%)	1 (2%)	2 (5%)	--
No	235 (98%)	93 (97%)	47 (98%)	38 (95%)	10 (100%)
Missing	4 (1%)	1 (1%)	--	--	--

[1] CSW: commercial sex work, [2] LMP: Last menstrual period, excluding those past menopause or pregnant.

(continued…)

Table 3.2 (continued):

	Main (N=242)	**Culture** (N=96)	**Pyro** (N=48)	**qPCR** (N=40)	**Clone** (N=10)
Days since LMP[2] (Excl. meno./preg.)	*N=222*	*N=91*	*N=46*	*N=38*	*N=10*
Median	12	12	12	11	20
Range	1-39	1-37	1-33	1-35	4-33
Missing	51 (23%)	16 (18%)	8 (17%)	6 (16%)	1 (10%)
Body Mass Index					
Median	28.5	28.9	29.6	29.3	32.9
Range	17.2-55.2	18.4-55.2	18.5-55.2	21.4-55.2	21.9-55.2
BMI>30	91 (38%)	37 (39%)	19 (45%)	13 (33%)	5 (50%)
Missing	12 (5%)	8 (8%)	2 (5%)	5 (13%)	1 (10%)
Intra-vaginal cleansing					
On day of sample	233 (96%)	91 (95%)	46 (96%)	38 (95%)	9 (90%)
Post-coitally	190 (79%)	79 (82%)	40 (83%)	32 (80%)	8 (80%)
Daily	46 (19%)	15 (16%)	8 (17%)	8 (20%)	2 (20%)
Water/Salt water	10 (4%)	5 (5%)	3 (6%)	2 (5%)	--
Soap and water	222 (92%)	86 (90%)	45 (94%)	37 (93%)	10 (100%)
Antiseptic soap	4 (2%)	3 (3%)	--	1 (2%)	--
Missing	6 (2%)	2 (2%)	--	--	--
Hormonal contraception					
No	178 (74%)	68 (71%)	33 (69%)	30 (75%)	7 (70%)
Oral (the Pill)	22 (9%)	10 (10%)	7 (15%)	3 (8%)	--
Depo-Provera	36 (15%)	16 (17%)	8 (17%)	6 (15%)	2 (20%)
Missing	6 (2%)	2 (2%)	--	1 (2%)	1 (10%)
Symptoms[3]					
Any	20 (8%)	10 (10%)	6 (13%)	6 (15%)	1 (10%)
Missing	4 (2%)	1 (1%)	--	--	--
Other diagnoses[4]					
Any	13 (5%)	2 (2%)	1 (2%)	1 (2%)	--
Missing	4 (2%)	1 (1%)	--	--	--
Current STI[5]					
Syphilis	7 (3%)	5 (5%)	3 (6%)	4 (10%)	1 (10%)
Gonorrhea	2 (1%)	2 (2%)	2 (4%)	2 (5%)	--
Trichomonas	10 (4%)	3 (3%)	1 (2%)	1 (2%)	--
Any	18 (7%)	10 (11%)	6 (13%)	7 (18%)	1 (10%)
Missing	4 (2%)	1 (1%)	--	--	--
Inclusion criteria[6]					
Meets all	158 (65%)	65 (68%)	31 (65%)	23 (58%)	7 (70%)

[3] Including discharge, itch and dysuria, [4] Including ulcers, candidiasis, cystitis, cervicitis, pelvic inflammatory disease, [5] STI: Sexually transmitted infection, [6] Including active CSW, not pregnant or past menopause, no symptoms, diagnoses or current STI.

Median follow-up for individuals in the Main group is three years, and differences between the main group and subsets were not significant (by Mann-Whitney, not shown). In contrast, the median age was significantly younger in all of the subsets relative to the Main group (by Mann-Whitney, not shown). Unsurprisingly, participant age is correlated with length of follow-up ($R^2 = 0.28$, $p<0.0001$ by linear regression, data not shown).

Approximately 90% of the Main group was currently actively engaged in CSW. Unsurprisingly, those who reported having stopped CSW were significantly older than those who reported active CSW ($p<0.0001$ by t test, data not shown). In the active CSW group, individuals reported a median of 4 clients per day and over 90% of individuals reported always using condoms for CSW. Information about LMP was available for 70% of the Main group, with the remainder either pregnant (1%), past menopause (7%) or with missing values (23%).

A surprising number of individuals have a body mass index (BMI) consistent with being overweight (30%) or obese (38%). Increased BMI is significantly associated with older age ($p=0.005$ by linear regression, data not shown).

All women in the main study reported regular douching (intra-vaginal cleansing or IVC), with all but three reporting IVC on the day of sample collection. Most individuals reported using a mixture of soap and water for IVC, with a small proportion using water/salt water only, or antiseptic soap (Dettol brand). Most reported post-coital IVC (ie. after each CSW client), while a minority indicated daily IVC. Consistent with this, active CSW are more likely to douche post-coitally compared to those who had stopped (129/149 or 87% vs. 7/23 or 30%, $p<0.0001$ by chi square test). Individuals douching post-coitally are younger on average than those douching daily (mean 38.3 vs. 40.8, $p=0.02$ by t test, not

shown), and report a higher median number of CSW clients per day (4 vs. 3, p=0.004 by Mann Whitney test, not shown).

Less than a quarter of the study group reported hormonal contraceptive use, either combination estrogen/progresterone (oral contraceptive therapy), or progesterone only (Depo-Provera). A small number of women reported vaginal symptoms such as vaginal discharge, itch, and dysuria (painful urination), were diagnosed with genital ulcers, candidiasis, cystitis, cervicitis, or pelvic inflammatory disease, or were diagnosed with a sexually transmitted infection (STI) at the time of sample collection (Table 3.2).

Based on this information, a number of inclusion criteria were established for statistical analyses comparing HIV-R and HIV-N individuals, including active CSW, not being pregnant or past menopause, and no vaginal symptoms, diagnoses or STI (Table 3.2). About two thirds of the study group meets all of these criteria. Further, these criteria were not considered when selecting subsets, meaning that the most detailed information in this study was collected from individuals not meeting inclusion criteria. Therefore, any observed differences between HIV-R and HIV-N individuals in study subsets must be considered in light of these factors.

Comparison of profile between HIV-R and HIV-N individuals – Since remaining HIV– for a longer period of follow-up defines HIV-R status, it is expected that HIV-R individuals have been followed for significantly longer than HIV-N (Table 3.3). Unsurprisingly, the HIV-R group is also significantly older on average than the HIV-N group, reflecting the longer time of follow-up that defines the HIV-R group (Table 3.3). Despite purposefully selecting younger individuals for study subsets, differences in length of follow-up remain significant between HIV-R and HIV-N in all subsets (not shown).

Table 3.3 – Comparison of study characteristics between HIV-R and HIV-N

	HIV-R (N=61)	**HIV-N** (N=57)	**Difference*** Main Group	Subsets[1]
Months in cohort				
Median	63	9	p<0.0001	All
Range	45-236	0-34		
Age				
Mean (SD[2])	42.2 (6.8)	34.5 (6.2)	p<0.0001	All
Range	26-60	21-48		
Current CSW[3]				
Active	37 (61%)	45 (74%)	p=0.02	--
Taking break	15 (25%)	11 (18%)		
Stopped	9 (14%)	1 (1%)		
Menopause				
Yes	8 (13%)	1 (2%)	p=0.03	--
No	52 (85%)	55 (96%)		
Missing	1 (2%)	1 (2%)		
Depo-Provera user				
Yes	6 (10%)	14 (25%)	p=0.05	--
No	54 (88%)	41 (71%)		
Missing	1 (2%)	2 (4%)		
Meets inclusion criteria[4]				
Yes	35 (57%)	20 (35%)	p=0.02	--
No	26 (43%)	37 (65%)		

* Significant difference in distribution between HIV-R and HIV-N, by chisquare/ Fisher's exact test (ordinal/categorical variables), Mann Whitney test (continuous variables with median reported) or t test (continuous variables with mean reported); variables with insignificant p values not shown, [1] Significant difference ($p<0.05$) in any subset, [2] SD: standard deviation, [3] CSW: commercial sex work, [4] Including active CSW, not pregnant or past menopause, no symptoms, diagnoses or current STI.

Consistent with observed age differences, individuals in the HIV-N group were more likely to be active in CSW and less likely to be past menopause compared to those in the HIV-R groups (Table 3.3). The observation that HIV-N individuals are significantly more likely to use Depo-Provera compared to HIV-R may also be due to younger individuals being more likely to use this form of birth control (Table 3.3). No differences between HIV-R and HIV-N individuals in terms of current STI, vaginal symptoms or other diagnoses were observed (not shown).

Profile of HIV+ individuals – In the main group, HIV+ individuals are less likely to be active CSW compared to HIV– individuals (67/124 or 54% vs. 82/118 or 69%, p=0.02 by Fisher's exact test) and also have lower median BMI compared to HIV– individuals (27.6 vs. 29.1, p=0.02 by Mann Whitney test).

Variables describing HIV+ participants include current use of antiretroviral (ARV) medication, antibiotic prophylaxis with Septrin (sulfamethoxazole-trimethoprim), as well as counts of CD4+ and CD8+ T cells (Table 3.4). ARV therapy was introduced less than two years before the study period, with 36% of participants currently taking ARV at the time of sample collection (Table 3.4). Most of those receiving ARV therapy also receive daily prophylaxis with the antibiotic Septrin (39/45 or 87%), with a further 13% of all HIV+ individuals receiving Septrin prophylaxis but not ARV therapy (Table 3.4).

Median CD4 counts are significantly lower in those receiving both Septrin and ARV therapy compared to those receiving neither (333 vs. 433, p=0.002). Interestingly, those receiving Septrin but not ARVs had lower median CD4 counts than those receiving both and those receiving neither (264, p=0.004 and p<0.0001 respectively, all comparisons by Mann-Whitney test).

Table 3.4 – Participant characteristics (HIV+ only)

	Main (N=124)	**Culture** (N=32)	**Pyro** (N=16)	**qPCR** (N=10)
Current ARV[1] therapy				
Yes	45 (36%)	10 (31%)	4 (25%)	3 (30%)
No	77 (62%)	22 (69%)	12 (75%)	7 (70%)
Missing	2 (2%)	--	--	--
Septrin prophylaxis				
Yes	55 (44%)	12 (38%)	7 (44%)	5 (50%)
No	67 (54%)	20 (62%)	9 (56%)	5 (50%)
Missing	2 (2%)	--	--	--
ARV[1]/Septrin use				
Neither	61 (49%)	16 (50%)	8 (50%)	3 (30%)
ARV only	6 (5%)	4 (13%)	1 (6%)	2 (20%)
Septrin only	16 (13%)	6 (19%)	4 (25%)	4 (40%)
Both	39 (31%)	6 (19%)	3 (19%)	1 (10%)
Missing	2 (2%)	--	--	--
CD4 count (cells/μl)				
Median	389	385	344	301
Range	35-1393	35-1164	52-1164	52-731
CD8 count (cells/μl)				
Median	982	897	937	923
Range	264-3072	264-2155	541-2085	639-1596
CD4/CD8 ratio				
Median	0.37	0.50	0.25	0.31
Range	0.06-1.58	0.06-1.33	0.16-1.33	0.06-0.81
Missing (CD4/8)	4 (3%)	1 (3%)	1 (9%)	--

[1] ARV: Antiretroviral

These findings may reflect increased likelihood that individuals with lower CD4 counts will be prescribed Septrin prophylaxis. Median CD8 counts are lower in those receiving both Septrin and ARVs compared to those receiving neither (907 vs. 1181, p=0.02 by Mann Whitney test), possibly indicating reduced immune activation in treated individuals.

Longitudinal analysis of BV diagnoses – In the retrospective study group, 3530 samples were collected from 1005 individuals over a four-year period (Table 3.5). As in the main study group, approximately half of individuals are HIV+, while ten individuals, providing a total of 47 samples, seroconverted to HIV during the study period. About 20% of all HIV– individuals were categorized as HIV-R at the beginning of the study period, with a further 10% categorized as HIV-R by the end. The median age of the study group is 36 years, with a significant difference in age between the HIV-R and HIV-N groups, as for the main study group (median 40 vs. 32, $p<0.0001$ by Mann-Whitney test, not shown).

For HIV+ individuals, median CD4 and CD8 counts were similar in the retrospective study and Main group. Since the database available for retrospective analysis contained fewer variables than the database available for the main study group, potentially influential factors such as LMP, CSW clients per day, current STI, ARV therapy and Septrin prophylaxis were not included in this analysis.

Summary – Two study groups are described: a longitudinal study over 6 weeks in a convenience sample of Winnipeg adolescents, and a large cohort-based study of CSW participating in an STD/HIV prevention project in Nairobi, Kenya. Profile of characteristics for Nairobi individuals is shown for the Main group and several subsets selected for analysis using different techniques (culture, deep sequencing, qPCR and clone libraries).

Table 3.5 – Characteristics of Nairobi participants and samples (Retrospective study)

		Participants (N=1005)		**Samples** (N=3530)
Samples provided				
1-2		375 (37%)		533 (15%)
3-4		314 (31%)		1084 (31%)
5-6		206 (20%)		1120 (32%)
7-8		110 (11%)		793 (22%)
Serostatus				
HIV–		550 (55%)		1834 (52%)
HIV+		421 (42%)		1615 (46%)
Seroconverters		10 (1%)		47 (1%)
Missing		24 (2%)		34 (1%)
Serogroup				
(HIV– only)	*N=550*		*N=1834*	
HIV-N		372 (68%)		945 (52%)
HIV-R		121 (22%)		547 (30%)
Transitioned to HIV-R[1]		57 (10%)		342 (19%)
Age				
Median		36		
Range		19-62		
Missing		104 (10%)		
HIV+ only			*N=1640*	
CD4 count (cells/μl)				
Median				288
Range				1-1807
CD8 count (cells/μl)				
Median				831
Range				1-4656
CD4/CD8 ratio				
Median				0.44
Range				0.009-60.5
Missing (CD4/8)				102 (6%)

[1] Individuals with <3 years follow-up at start of retrospective study period (Nov/04) and >3 years follow-up at end of study period (Jul/08).

Subsets were selected with intentional over-representation of HIV– and younger individuals (described in this chapter), as well as balanced representation of BV vs. BV– diagnoses (next chapter).

Several "inclusion" criteria were defined based on active CSW, pregnancy, menopause and vaginal symptoms/infections. However, somewhat less than 70% of individuals in the Main group and subsets meet all of these criteria, and they were not considered prior to selecting subsets. Therefore, these factors must be accounted for in all statistical comparisons undertaken in this study.

A major caveat in describing differences between HIV-R and HIV-N individuals in this study is the significantly older age of HIV-R. Although an attempt to mitigate this difference was made by selecting younger individuals for the Culture subset, it remains highly significant in all study groups and subsets, and likely explains other observed differences between the two groups, including CSW activity, menopause, Depo-Provera use and meeting inclusion criteria.

The profile of individuals in the retrospective study group (1005 individuals) is remarkably similar to the profile in the main study group, indicating that a representative cross-section of the larger study group was captured in the Main group. A total of 10 individuals or 2% of all HIV– individuals became HIV+ in the four-year study period. Nearly one third of HIV– individuals were considered to be HIV-R according to the three-year definition by the middle of 2008.

2. Diagnosis of BV by microscopy and assessment of vaginal pH

High-powered microscopic semi-quantitation of bacterial morphotypes in a Gram-stained smear from a vaginal swab and scoring based on the criteria of Nugent (146) or Hay and Ison (98), is currently the gold standard for laboratory diagnosis of bacterial vaginosis (BV). Since BV is a known risk factor for HIV acquisition and shedding, the main objective of this chapter is to determine if BV diagnosis and related factors, including vaginal pH, is associated with HIV resistance in the Majengo cohort.

Pilot study – All sample collection, processing and BV diagnosis procedures were optimized using 103 samples from 32 adolescents in the pilot study (Table 4.1). About one third of pilot study samples were diagnosed as BV+, with a further 20% diagnosed as BVI (intermediate BV) or BV0 (no visible bacteria on Gram stain). Half of the group was BV–. Nearly 60% of the group had *Lactobacillus* morphotypes observed on Gram stain, compared to only one third of the group with BV morphotypes observed on Gram stain.

Sample weight was assessed by weighing specimen collection devices to the nearest mg before and after sample collection. The mean weight of samples collected in the pilot study was 45mg. Nearly 60% of individuals provided at least one sample with a BV score, and nearly 30% always provided samples with BV scores.

Interestingly, decreased sample weight was significantly associated with having a BV sample. The mean weight of a BV– sample was 53mg, compared to 32mg for BVI samples ($p=0.02$), 39mg for BV+ samples ($p=0.006$), and 37mg for BV0 samples ($p=0.07$, all comparisons by t test).

Table 4.1 – BV diagnosis and related variables (Pilot study)

	Samples (N=103)
BV diagnosis[1]	
BV–	51 (50%)
BVI	8 (8%)
BV+	33 (32%)
BV0	9 (9%)
All BV	50 (49%)
Missing	2 (2%)
Abundance of *Lacto.* morphotypes[2]	
None	41 (40%)
1-30	19 (18%)
>30	41 (40%)
Missing	2 (2%)
Abundance of BV morphotypes[3]	
None	65 (63%)
1-30	23 (22%)
>30	13 (13%)
Missing	2 (2%)
Sample weight (mg)	
Mean (SD)	45 (24)
Range	4-108

[1] Bacterial vaginosis diagnosis by Gram stain: BV– (BV-negative), BVI (BV-intermediate), BV+ (BV-positive), BV0 (BV-none), All BV (BVI/BV+/BV0). [2] Abundance of *Lactobacillus* morphotypes on Gram stain (per high-powered field). [3] Abundance of BV morphotypes on Gram stain.

Increased *Lactobacillus* morphotypes and decreased BV morphotypes on Gram stain were also associated with increased sample weight (p=0.01 and p=0.05 respectively by oneway ANOVA, not shown). These findings indicate that BV is associated with reduced vaginal secretions or mucus of reduced density, a possible mechanism of increased vulnerability to HIV and STI in those with BV.

BV diagnosis and related variables (Nairobi study) – About one quarter of individuals in the Main group were diagnosed as BV+, with a further 20% diagnosed as BV0 (ie. no visible bacteria were observed on Gram stain) and 10% diagnosed as BVI (intermediate between BV+ and BV–) (Table 4.2). About half of the group was BV–. About 60% of the Main group had *Lactobacillus* morphotypes observed on Gram stain and over 40% had BV morphotypes observed on Gram stain. Surprisingly, these proportions are quite similar to what was observed in the pilot study, indicating a fundamental similarity in microbiological patterns observed in these very different groups. In the Pyro and qPCR subsets, more BV+ samples and less BV0 samples were selected relative to the Main group and Culture subset (Table 4.2).

About 20% of samples had yeast cells observed on Gram stain (Table 4.2). Although not significantly associated with BV diagnosis, observation of yeast was associated with increased *Lactobacillus* morphotypes and decreased BV morphotypes (p=0.006 and p=0.002 respectively by chi square test, not shown). Almost half of samples had leukocytes observed on Gram stain, an indicator of mucosal inflammation (44). Leukocytes were more likely to be observed in BVI and BV+ samples (14/22 or 63% and 41/58 or 71% respectively) compared to BV– samples (39/115 or 34%, p=0.02 and p<0.0001 respectively by Fisher's exact test).

Table 4.2 – BV diagnosis and related variables (Nairobi study)

	Main (N=242)	**Culture** (N=96)	**Pyro** (N=48)	**qPCR** (N=40)	**Clone** (N=10)
BV diagnosis[1]					
BV–	115 (48%)	42 (44%)	20 (42%)	20 (50%)	6 (60%)
BVI	22 (9%)	7 (7%)	4 (8%)	1 (2%)	--
BV+	58 (24%)	28 (29%)	22 (46%)	19 (48%)	4 (40%)
BV0	47 (19%)	19 (20%)	2 (4%)	--	--
All BV	127 (52%)	54 (56%)	28 (58%)	20 (50%)	4 (40%)
Lacto. morphotypes[2]					
None	100 (41%)	43 (45%)	20 (42%)	16 (40%)	4 (40%)
1-30	115 (48%)	43 (45%)	23 (48%)	19 (48%)	6 (60%)
>30	27 (11%)	10 (10%)	5 (10%)	5 (12%)	--
BV morphotypes[3]					
None	139 (57%)	51 (53%)	18 (38%)	18 (45%)	6 (60%)
1-30	62 (26%)	27 (28%)	15 (31%)	8 (20%)	--
>30	41 (17%)	18 (19%)	15 (31%)	14 (35%)	4 (40%)
Yeast observed[4]					
Yes	40 (17%)	18 (19%)	9 (19%)	6 (15%)	1 (10%)
No	202 (83%)	78 (81%)	39 (81%)	34 (85%)	9 (90%)
Leukocytes observed[5]					
Yes	104 (43%)	45 (47%)	23 (48%)	20 (50%)	4 (40%)
No	138 (57%)	51 (53%)	25 (52%)	20 (50%)	6 (60%)
Vaginal pH					
Median	5.0	5.0	5.0	5.2	4.2*
Range	4.0-7.0	4.0-7.0	4.0-7.0	4.0-7.0	4.0-5.8
Missing	1 (<1%)	1 (1%)	1 (2%)	--	--
Sample weight (mg)					
Median	50	51	52	56	66*
Range	7-147	22-136	29-105	27-93	40-85

* Significantly different from the main group ($p<0.05$ by Mann-Whitney test). [1] Bacterial vaginosis diagnosis by Gram stain: BV– (BV-negative), BVI (BV-intermediate), BV+ (BV-positive), BV0 (BV-none), All BV (BVI/BV+/BV0). [2] Abundance of *Lactobacillus* morphotypes on Gram stain (per high-powered field). [3] Abundance of BV morphotypes on Gram stain, [4] Yeast morphotypes observed on Gram stain, [5] Leukocytes observed on Gram stain.

Leukocytes on Gram stain were also associated with decreased *Lactobacillus* morphotypes and especially with increased BV morphotypes ($p=0.02$ and $p<0.0001$ respectively by chi square test, not shown). Individuals with yeast were also more likely to have leukocytes on Gram stain compared to those without (27/104 or 26% vs. 13/138 or 10%, $p=0.0008$ by Fisher's exact test). Since yeast is associated with increased *Lactobacillus* on Gram stain, these findings indicate that yeast and BV are inversely associated with each other and are both associated with mucosal inflammation as indicated by presence of leukocytes in vaginal secretions.

Median vaginal pH in the Main group was 5.0 (Table 4.2). As expected, a strong association was observed between BV diagnosis by Gram stain and vaginal pH in the main study (Figure 4.1). Interestingly, slides with intermediate BV scores (BVI) also had intermediate median pH values and slides with no bacteria (BV0) had the highest median pH, indicating that these BV categories have distinct biological properties with possible implications for vaginal health and susceptibility to infection. As expected, decreased vaginal pH was associated with increased *Lactobacillus* morphotypes and decreased BV morphotypes on Gram stain ($p<0.0001$ and $p=0.002$ respectively by Kruskal Wallis test, not shown).

Median sample weight was 50mg (Table 4.2). A significant inverse correlation was observed between sample weight and vaginal pH ($R^2 = 0.03$, $p=0.004$ by linear regression), suggesting that lower vaginal pH may be associated with increased volume and density of vaginal mucus. Range and central values for sample weight were similar in Nairobi compared to pilot study samples, although data for this variable were not normally distributed in Nairobi.

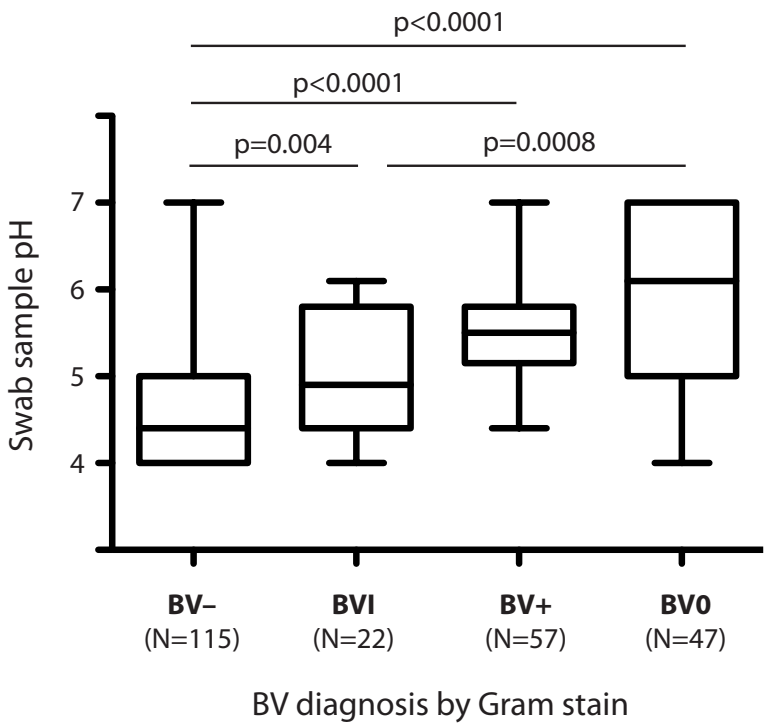

Figure 4.1 – Vaginal pH associated with BV diagnosis by Gram stain

In the main study group, samples with BV scores are associated with increased pH compared to BV– samples. Significance values calculated using the Mann-Whitney test in GraphPad Prism v5.

In contrast to the pilot study, no difference in sample weight was observed between BV– and BV+ individuals, however individuals with BVI samples had significantly reduced median sample weight (44mg) compared to both BV– and BV+ (50mg and 52mg, p=0.05 and p=0.04 respectively by Mann Whitney test, not shown). Although median sample weight was lowest in BV0 individuals (41mg), the difference between BV0 and other groups was not significant. These observations indicate that reduced sample weight may be a specific biological characteristic of BVI samples in terms of transition from BV– to BV+. Abundance of *Lactobacillus* and BV morphotypes on Gram stain is not associated with sample weight (not shown).

Comparison of BV-associated variables between HIV-R and HIV-N – No differences were observed between the two groups in terms of BV diagnosis, abundance of *Lactobacillus* morphotypes or vaginal pH, disproving Hypotheses 1, 2 and 4. Interestingly, HIV-R individuals had significantly less abundant BV morphotypes on Gram stain compared to HIV-N individuals, even when considering samples meeting inclusion criteria only (Table 4.3). Although decreased BV morphotypes were also associated with increased age (p=0.01 by Kruskal Wallis test, not shown), this association was not significant when only considering those meeting inclusion criteria. The fact that 25% of HIV-R individuals vs. 9% of HIV-N individuals had BV0 scores (ie. no visible bacterial morphoptypes) is likely to explain reduced BV morphotypes in HIV-R. Individuals with BV0 scores were likely to be older (mean 41.5) than BV– (38.9, p=0.03), BVI (38.5, p=0.07), or BV+ (36.8, p=0.0006, all by t test). However, these differences were not significant when considering those meeting inclusion criteria only. These data indicate that HIV-R individuals in the Main group are less likely to have BV morphotypes on Gram stain.

Table 4.3 – Comparison of BV-related variables between HIV-R and HIV-N

	HIV-R (N=61)	**HIV-N** (N=57)	**Difference*** All samples	Meets IC[1]
BV diagnosis[3]				
BV–	31 (51%)	29 (51%)	--	--
BVI	2 (3%)	5 (9%)		
BV+	13 (21%)	18 (31%)		
BV0	15 (25%)	5 (9%)		
All BV (I/+/0)	30 (49%)	28 (49%)		
Lacto. morphotypes[4]				
None	25 (41%)	22 (39%)	--	--
1-30	30 (49%)	28 (49%)		
>30	6 (10%)	7 (12%)		
BV morphotypes[5]				
None	43 (70%)	23 (40%)	p=0.001	p=0.005
1-30	7 (11%)	21 (37%)		
>30	11 (19%)	13 (23%)		
Yeast observed[6]				
Yes	7 (11%)	9 (16%)	--	--
No	54 (89%)	48 (84%)		
Leukocytes observed[7]				
Yes	21 (34%)	29 (51%)	--	--
No	40 (66%)	28 (49%)		
Vaginal pH				
Median	5.0	4.7	--	--
Range	4.0-7.0	4.0-7.0		
Missing	--	1 (2%)		
Sample weight (mg)				
Median	45	54	p=0.05	--
Range	7-147	22-136		

* Indicates significant difference in distribution between HIV-R and HIV-N, by chisquare/Fisher's exact test (ordinal/categorical variables) or Mann Whitney test (continuous variables), [1] Meets inclusion criteria, including active CSW, not pregnant or past menopause, no symptoms, diagnoses or current STI, [2] Significant difference ($p<0.05$) in any subsets. [3] Bacterial vaginosis diagnosis by Gram stain: BV– (BV-negative), BVI (BV-intermediate), BV+ (BV-positive), BV0 (BV-none), All BV (BVI/BV+/BV0). [4] Abundance of *Lactobacillus* morphotypes on Gram stain (per high-powered field). [5] Abundance of BV morphotypes on Gram stain, [6] Yeast morphotypes observed on Gram stain, [7] Leukocytes observed on Gram stain.

HIV-R individuals had significantly lower sample weight compared to HIV-N individuals, however this association was not significant when considering only samples meeting inclusion criteria or any subsets of individuals (Table 4.3). Reduced sample weight was also associated with older age ($R^2 = 0.07$, $p<0.0001$ by linear regression, not shown), likely explaining the difference observed between HIV-R and HIV-N.

Potential relationships between BV diagnosis and HIV-R status were also evaluated in a longitudinal data set covering the period from mid-2004 to mid-2008. This period covers 8 separate "re-surveys", ie. bi-annual survey timepoints in which cohort members are scheduled for research visits in study clinics. Data comparing all samples provided by HIV-R and HIV-N (ie. not considering repeated measures on the same individual) are shown in Table 4.4, however between-group statistical comparisons were carried out separately for each re-survey timepoint and no significant differences were observed for any measure at any timepoint. Therefore, the observation that HIV-R were less likely to have BV morphotypes in the main group was not observed at any timepoint in the retrospective study, indicating that this is unlikely to be a real difference (Table 4.4).

A repeated measures analysis using Hierarchical Linear Modelling (HLM) software was carried out on the whole dataset in order to account for repeated measures, controlling for individual variation in BV diagnosis. When comparing those with any BV diagnosis (BVI/BV+/BV0) to BV– individuals using the repeated measures technique, there was no difference between HIV-N and HIV-R individuals (Figure 4.2, A). Therefore, by comparing HIV-R and HIV-N at a single timepoint, at multiple timepoints and using longitudinal analytical techniques, it can be concluded that HIV-R are not more likely to be diagnosed with bacterial vaginosis, disproving Hypothesis 1.

Table 4.4 – BV-related variables in retrospective study

		HIV– only (N=1856)		
	All retrospective (N=3530)	**HIV-N** (N=1060)	**HIV-R** (N=796)	**Difference*** (at any timepoint)
BV diagnosis[1]				
BV–	1747 (49%)	555 (52%)	438 (55%)	--
BVI	520 (15%)	161 (15%)	100 (13%)	
BV+	912 (26%)	261 (25%)	182 (23%)	
BV0	351 (10%)	83 (8%)	76 (10%)	
All BV (I/+/0)	1753 (51%)	505 (48%)	358 (45%)	
Lacto. morphotypes[2]				
None	1227 (35%)	332 (31%)	249 (31%)	--
1-30	1624 (46%)	493 (47%)	375 (47%)	
>30	679 (19%)	235 (22%)	172 (22%)	
BV morphotypes[3]				
None	1596 (45%)	477 (45%)	391 (49%)	--
1-30	902 (26%)	282 (27%)	193 (24%)	
>30	1032 (29%)	301 (28%)	212 (26%)	
Yeast observed[4]				
Yes	472 (13%)	156 (15%)	79 (10%)	--
No	3058 (87%)	904 (85%)	717 (90%)	
Leukocytes observed[5]				
Yes	1806 (51%)	552 (52%)	375 (47%)	--
No	1724 (49%)	508 (48%)	421 (53%)	

* Indicates significant difference in distribution between HIV-R and HIV-N ($p<0.05$), by chisquare/Fisher's exact test at any of eight single timepoints between Nov/04 and Jul/08, [1] Bacterial vaginosis diagnosis by Gram stain: BV– (BV-negative), BVI (BV-intermediate), BV+ (BV-positive), BV0 (BV-none), All BV (BVI/BV+/BV0). [2] Abundance of *Lactobacillus* morphotypes on Gram stain (per high-powered field). [3] Abundance of BV morphotypes on Gram stain, [4] Yeast morphotypes observed on Gram stain, [5] Leukocytes observed on Gram stain.

BV diagnosis and serostatus – In the main group, HIV+ individuals had significantly higher median vaginal pH than HIV– (5.3 vs 4.7, p=0.03 by Mann Whitney test), however no significant differences in BV diagnosis, morphotypes, yeast or leukocytes on Gram stain were observed.

In the retrospective study, HIV+ individuals were more likely to have BV compared to HIV– individuals at multiple timepoints and overall by repeated measures analysis (Figure 4.2, B). HIV+ individuals also had lower abundance of *Lactobacillus* morphotypes on Gram stain ($p<0.001$ by HLM, not shown). Interestingly, while differences between HIV+ and HIV– individuals for these variables were significant or trending towards significance at most early timepoints, no significant differences were observed at the last two timepoints (late 2007 and mid-2008), possibly indicating a reduction in BV and an increase in *Lactobacillus* morphotypes among HIV+ individuals over time. Indeed, the overall proportion of individuals with BV appears to be declining in this cohort (Figure 4.2, C).

This decline is mirrored by a concomitant drop in the proportion of samples in which leukocytes are observed on Gram stain (from a high of 71% in Spring 2005 to a low of 32% in Spring 2008, $p<0.0001$ by chisquare analysis, data not shown). Further investigation will be required, however reduced BV and vaginal leukocytes may indicate an overall improvement in vaginal health, particularly among HIV+ women.

Interestingly, those diagnosed with any STI were more likely to have BV morphotypes on Gram stain (12/18 or 67% vs. 88/220 or 40%, p=0.005 by chisquare analysis), reinforcing the concept that BV organisms are associated with increased vulnerability to STI and HIV.

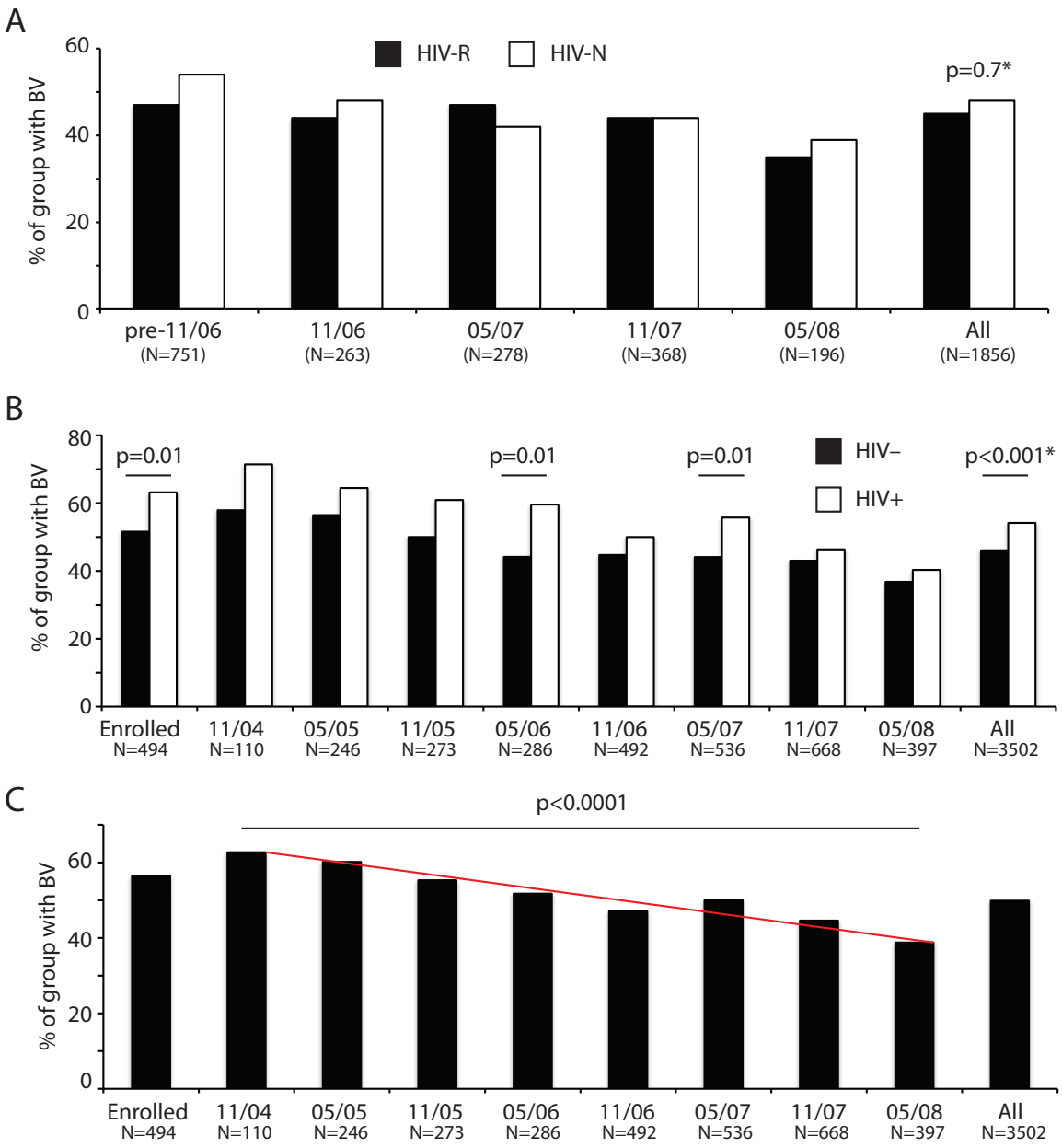

Figure 4.2 – BV diagnoses in Majengo cohort over 4 years

A. No differences in BV between HIV-R and HIV-N individuals at any timepoint or by hierarchical linear modelling (HLM*). B. More HIV+ have BV compared to HIV– individuals. The difference is significant at multiple timepoints (by Fisher's exact test, GraphPad Prism) and by HLM*. C. Significant reduction in those diagnosed with BV over time (p value calculated by chisquare test, GraphPad Prism).

Also in the retrospective study group, HIV+ individuals with BV (including BVI/BV+/BV0) have significantly increased CD8+ T cell counts and a decreased ratio of CD4+ to CD8+ T cells in peripheral blood compared to those who are BV– ($p<0.001$ by HLM for both measures, data not shown). This finding may indicate a potential influence of vaginal microbiology on profile of peripheral immune cell populations, or vice versa, or as a biological marker of another, undetected association.

Other biological and behavioural confounders – In the pilot study, no differences in BV diagnosis were observed between those who reported douching, current oral contraceptive or Depo-Provera use. However, a striking, stepwise association was observed between BV diagnosis and shorter self-reported time since last menstruation and last sexual intercourse, as well as sample weight (associated with BV as described above) (Table 4.5). All three variables were independently associated with BV by multivariate repeated measures analysis (175), indicating the importance of recent menstruation and sexual intercourse on profile of vaginal microbiology in this group of young women.

In contrast, no association between LMP and BV diagnosis was observed in the Nairobi study, although a weak correlation was observed between longer time since LMP and lower vaginal pH ($R^2 = 0.03$, $p=0.03$ by linear regression, not shown). This association may be masked by CSW and frequent douching in the Nairobi group.

Summary – By examining BV diagnosis and related variables at a single timepoint, at multiple timepoints and longitudinally accounting for repeated measures, this study found that HIV-R are no less likely than HIV-N individuals to be diagnosed with BV, contradicting Hypothesis 1.

Table 4.5 – Biological and behavioural correlates of BV (Pilot study)

	N	BV[1]	BV–	OR[2] (range)/Significance	
		N=41	N=51	Univ.	Multiv.
Sample weight (mg)				OR 1.08	OR 1.13
Mean (SD[3])		37 (22)	53 (21)	(1.05-1.12) p<0.001	(1.02-1.27) p=0.03
Time of last period					
Now	12	8 (67)	4 (33)	OR 2.25	OR 3.6
In past week	18	11 (61)	7 (38)	(1.6-3.3)	(1.2-10.5)
In past 2 weeks	17	8 (47)	9 (53)	p<0.001	p=0.02
In past 3 weeks	15	8 (53)	7 (47)		
In past 4 weeks	9	3 (33)	6 (67)		
More than 4 weeks	20	3 (15)	17 (85)		
Time of last SI[4]				OR 2.60	OR 14.6
In past 24h	19	13 (68)	6 (32)	(1.06-6.38)	(1.9-114.2)
In past week	23	13 (56)	10 (43)	p=0.04	p=0.01
In past month	17	3 (18)	14 (82)		
More than a month	8	3 (37)	5 (63)		

[1] Bacterial vaginosis, including BVI, BV+ and BV0. Intermediate/bacterial vaginosis, [2] Odds ratio, [3] Standard deviation, [4] Sexual intercourse

The observation that HIV-R individuals had increased BV morphotypes on Gram stain compared to HIV-N in the main group was not corroborated when considering multiple timepoints or in longitudinal analysis, calling into question the significance of this association. No differences in vaginal pH were observed between the two groups, contradicting Hypothesis 2. Therefore, BV diagnosis and vaginal pH are not associated with HIV resistance in the Majengo cohort.

An interesting association was observed between decreased sample weight and BV diagnosis in the pilot study group, indicating that BV results in a lower volume or lesser density of vaginal mucus, with possible implications for increased vulnerability to exogenous pathogens. Consistent with this observation, elevated pH (strongly associated with BV diagnosis) was associated with decreased sample weight in the Nairobi study. In contrast to the pilot study, sample weight was significantly decreased in BV intermediate (BVI) samples relative to both BV– and BV+ samples, indicating that BVI may have particular implications for increased susceptibility to infection in this group.

Although no association between BV diagnosis and HIV+ serostatus was observed in the main study group, HIV+ individuals had significantly higher vaginal pH and were more likely to have BV at multiple timepoints in the longitudinal study and overall when considering repeated measures. However, the difference between HIV+ and HIV– individuals was not significant at more recent timepoints. This finding is consistent with an overall drop in the proportion of cohort members diagnosed with BV over the four-year study period, possibly indicating an improvement in vaginal health over time in this group. A reduction in the proportion of individuals with vaginal leukocytes on Gram stain may also be an indication of reduced mucosal inflammation and improved vaginal health.

3. Flow cytometric quantification of bacteria in vaginal specimens

A simple method was developed to acquire flow cytometric profiles of bacterial and mucosal cell populations extracted from vaginal specimens. Although the full potential of this approach to quantify bacterial sub-populations using species-specific probes or antibodies was not realized in this project, quantitation of the concentration of bacterial cells in vaginal specimens is a novel application to better understand microbiological fluctuations associated with BV. Increased bacterial load associated with BV is hypothesized to increase stimulation of innate immune effectors and subsequent activation of T cell subsets, therefore decreased bacterial load may help to explain "immune quiescence" associated with HIV resistance. The goal of this chapter is to determine if HIV resistance is associated with a decreased concentration of bacteria in vaginal specimens.

Pilot study – Flow cytometry procedures were optimized during the pilot study (175). Nucleic acid staining revealed distinct cell populations within vaginal swab samples (Figure 5.1). The lower "bacterial cell-unit" (BCU) population strongly resembles cytometric profiles of pure cultures of common vaginal bacteria (Figure 5.1, A-B). The upper "mucosal cell-unit" (MCU) population strongly resembles cytometric profiles of vaginal epithelial cells (cultured using standard procedures) (Figure 5.1, C).

A third cell population was observed by flow cytometry in 3 of 103 samples, all from participants reporting sexual activity in the past 24h (Figure 5.2). Microscopic examination of Gram-stained swab smears revealed sperm cells in all three samples.

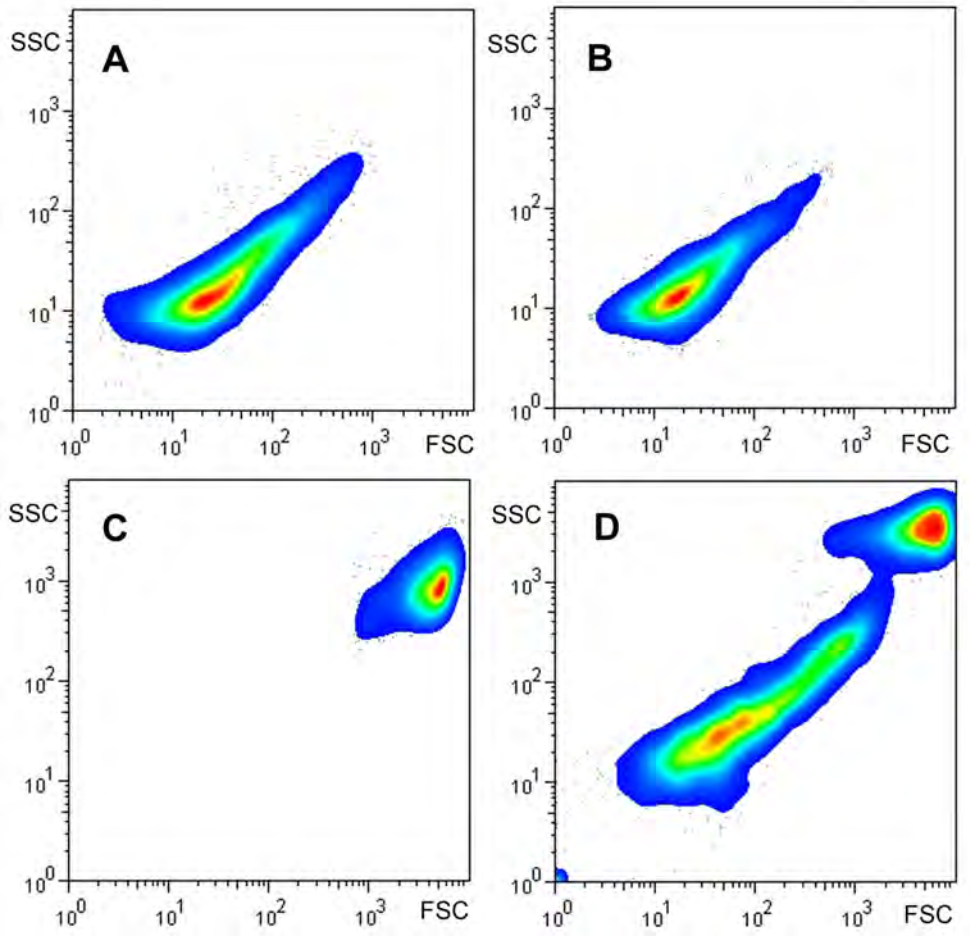

Figure 5.1 – Flow cytometric analysis of cell populations in vaginal specimens

Light-scatter profiles (smoothed scatterplot) of cultured bacteria, epithelial cells and vaginal specimen. A. Overnight culture of Lactobacillus crispatus DSMZ 14018, B. Gardnerella vaginalis ATCC 20584, C. Cultured vaginal epithelial cells VK2-E6E7, D. Typical light-scatter profile of a vaginal swab sample, showing lower (bacterial cell-unit or BCU) and upper (mucosal cell-unit or MCU) populations. FSC = forward scatter channel, SSC = side scatter channel.

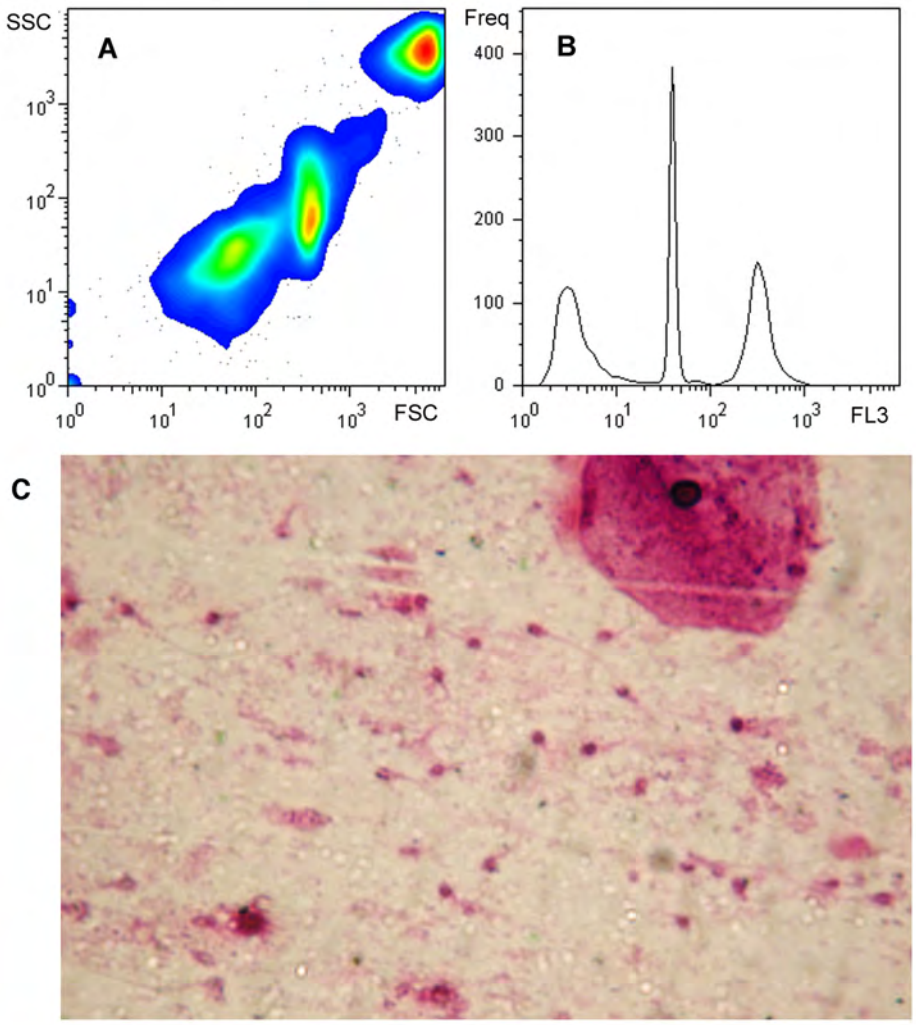

Figure 5.2 – Detection of sperm cells in vaginal specimens by flow cytometry

A distinct cell population was observed in three samples, provided by participants who reported sexual intercourse in the previous 24h. A. Light-scatter profile (smoothed scatterplot) of vaginal swab sample with third cell population, overlapping lower cell population. B. Red fluorescence (FL3) histogram showing sharply defined intermediate cell population. C. Gram-stained slide of sample shown in B, with visible sperm cells (below and to the left of epithelial cell). FSC = forward scatter channel, SSC = side scatter channel.

Since fresh specimens in the pilot study were used to optimize procedures, thawed batch-processed samples were used for quantitation experiments. In order to simplify presentation, cell-unit counts are $\log_{10}$ normalized (therefore, a value of 6.12 cell-units is equal to $10^{6.12}$ or ~1.3 million cell-units).

For 103 pilot study samples, median BCU is 6.64 (range 5.09-7.94) and median MCU is 6.56 (range 4.94-7.31). Both BCU and MCU counts were significantly correlated with sample weight (BCU, $R^2 = 0.22$, $p<0.0001$; MCU, $R^2 = 0.15$, $p<0.0001$ by linear regression, not shown). After adjusting for sample weight, median BCU/g is 8.05 (range 6.95-9.04) and median MCU/g is 7.88 (range 6.62-8.40).

Values for BCU and MCU are not correlated with each other. The ratio of BCU to MCU is highly variable with a median value of 9.3 BCU per MCU and ranging from 0.06 (ie. 6 BCU per 100 MCU) to 207. The biological significance of these data is unknown, however hypothetically they may reflect dynamics of the microbiota, such as overgrowth of bacteria relative to a constant rate of shed epithelial cells, or increased shedding of epithelials and leukocytes relative to a constant bacterial concentration in the context of mucosal inflammation.

Quantification of vaginal cell populations (Nairobi) – Flow cytometry was carried out with fresh specimens in Nairobi, in contrast to the use of previously thawed samples in the pilot study. In general, profiles of cell populations in Nairobi samples were identical to the pilot study. Several Nairobi samples with sperm detected on Gram stain also had detectable sperm populations by flow cytometry (not shown). These findings indicate that three cell populations in vaginal specimens from different study populations are accurately identified using this simple and rapid technique (175).

As in the pilot study, BCU and MCU were both significantly correlated with sample weight (BCU, R^2 = 0.06, p<0.0001; MCU, R^2 = 0.09, p<0.0001 by linear regression, not shown), while in contrast to the pilot study, BCU and MCU were significantly correlated with each other (R^2 = 0.30, p<0.0001 by linear regression).

Median values for BCU/g were nearly a log higher in the Nairobi study compared to the pilot study, while median MCU/g were half a log lower (Table 5.1, A). The median ratio of BCU:MCU was also substantially higher in the Nairobi study. The range of values for ratio was very wide, but did not range as low as in the pilot study. Although these findings may suggest actual differences in vaginal cell populations between these two groups, such as increased sexual activity and douching, they may also reflect experimental differences, such as the use of fresh vs. previously frozen samples for analysis.

Comparison of total BCU in HIV-R vs. HIV-N – HIV-R individuals in the Main group were found to have significantly reduced BCU/g compared to HIV-N individuals, although not when considering those meeting inclusion critieria only (Table 5.1, B). Reduced BCU/g was also correlated with increased age (R^2 = 0.03, p=0.009 by linear regression) and having stopped CSW (median 8.75 vs. 9.08 in active CSW, p=0.05 by Mann Whitney test), likely explaining why the association was no longer significant when considering only those meeting inclusion criteria. Reduced MCU/g in HIV-R individuals was only significant when considering the culture subset (Table 5.1, B). Therefore, HIV-R individuals do not have a reduced concentration of bacteria in vaginal specimens as measured by flow cytometry, disproving Hypothesis 3.

Table 5.1 – Quantitation of BCU/MCU and comparison of HIV-R vs. HIV-N

A. *Quantitation of BCU/MCU*

	Main (N=242)	**Culture** (N=96)	**Pyro** (N=48)	**qPCR** (N=40)	**Clone** (N=10)
BCU/g[1]					
Median	8.98	8.98	8.95	8.98	8.80
Range	7.19-10.13	7.32-10.13	7.77-10.13	7.97-10.13	8.51-9.81
MCU/g[2]					
Median	7.39	7.40	7.41	7.41	7.34
Range	5.90-8.82	5.90-8.10	6.77-8.10	6.77-8.10	6.96-7.72
Ratio (BCU:MCU)					
Median	49.0	31.5	34.3	36.4	49.9
Range	0.7-489	4.3-207	4.3-183	7.0-207	15.1-141
Missing	1 (<1%)	1 (1%)	1 (2%)	1 (2%)	1 (10%)

[1] BCU: Bacterial cell-units per gram sample, [2] MCU: Mucosal cell-units per gram sample

B. *Comparison of HIV-R vs. HIV-N*

	HIV-R (N=61)	**HIV-N** (N=57)	**Difference***	
			All samples	Meets IC[1]
BCU/g				
Median	8.88	9.17	p=0.03	--
Range	7.71-9.72	7.77-10.13		
MCU/g				
Median	7.36	7.51	--	--
Range	6.37-8.04	6.60-8.10		
Ratio (BCU:MCU)				
Median	28.9	40.9	--	--
Range	4-207	3-183		

* Indicates significant difference in distribution between HIV-R and HIV-N, by chisquare/Fisher's exact test (ordinal/categorical variables) or Mann Whitney test (continuous variables), [1] Meets inclusion criteria, including active CSW, not pregnant or past menopause, no symptoms, diagnoses or current STI.

Vaginal cell-unit populations, BV diagnosis and pH – Consistent with the concept of BV as bacterial overgrowth, individuals in the Nairobi study who are BV+ have the highest mean BCU/g (9.16), compared to BV– (8.93, p=0.005), BVI (8.90, p=0.05), or BV0 (8.59, p<0.0001, all comparisons by t test). Those with BV0 scores had significantly lower mean BCU/g compared to all other groups, indicating a clear correspondence between Gram stain characteristics and quantitation of bacteria by flow cytometry.

Since BV is associated with increased vaginal pH and also with higher BCU/g, it was surprising to observe that higher vaginal pH was correlated with lower BCU/g (R^2 = 0.07, p<0.0001 by linear regression, not shown). Further investigation revealed that BV– individuals with higher pH had lower BCU/g (R^2 = 0.14, p<0.0001) and BV0 individuals with higher pH had lower BCU/g (R^2 = 0.27, p=0.0002), but there was no association between BCU/g and pH in BV+ or BVI individuals (R^2 = 0.01 for both, p=0.3/0.5 respectively, all comparisons by linear regression).

These findings may indicate that reduced bacterial concentrations are associated with increased pH in individuals who are transitioning from BV– to BVI or BV+, ie. *Lactobacillus* populations drop and pH rises as part of the transition to BV. A similar dynamic may characterize BV0 samples, but below the limit of detection for Gram stain analysis. The distinct biological properties of the BV0 category in this analysis are intriguing, possibly indicating a "wash-out" phase in BV related to transition from one category to another.

Interestingly, individuals with leukocytes observed on Gram stain had significantly higher BCU/g compared to those without (median 9.12 vs. 8.93, p=0.01), as well as significantly higher MCU/g (7.51 vs. 7.33, p=0.02, both comparisons by Mann

Whitney test). This finding is likely to reflect increased mucosal inflammation in the presence of elevated bacterial concentrations, as quantitated by flow cytometry.

Other biological and behavioural factors – In contrast to the Nairobi study, BV– samples in the pilot study were found to have increased BCU/g. However, this association was insignificant in a repeated measures multivariate model controlling for time since last menstruation and last sexual intercourse (Table 5.2).

A strikingly linear association between increased BCU/g and longer self-reported time since last menstruation was observed. Reporting no sexual intercourse in the past week was also associated with increased BCU/g, and both of these variables remained independently associated with BCU/g in multivariate analysis. In contrast, no association between last menstruation and BCU/g was detected in the Nairobi study.

Since longer time since last menstruation and last sexual intercourse were themselves strongly associated with being BV– (see Table 4.5, p.94), it is difficult to reconcile these findings with the concept of bacterial overgrowth and opposite trends identified in the Nairobi study. Since previously frozen samples used in the pilot study, effects on cell-unit counts during freezing and thawing likely explain these discrepancies. Freezing buffer was originally designed for preservation of *Lactobacillus* organisms (109), so it is plausible that largely Gram-negative BV organisms may have been preferentially lysed. This is corroborated by the observation that BV– individuals had higher BCU/g in a subset of previously frozen Nairobi samples quantitated by flow cytometry (not shown), indicating that BV bacteria may be less likely to be detected after freeze-thaw.

Table 5.2 – Repeated measures correlates of bacterial cell-units

	N	log10 BCU[1]/g (Mean/SD[2])	Significance (p)	
			Univariate	Multivariate
BV diagnosis[3]				
BV–	51	8.18 (0.46)		
BVI/BV+/BV0	41	7.92 (0.56)	0.052	0.845
Time of last period				
Now	15	7.76 (0.44)		
In past week	19	7.91 (0.51)		
In past 2 weeks	22	7.94 (0.53)		
In past 3 weeks	15	8.15 (0.49)		
In past 4 weeks	11	8.17 (0.41)		
More than 4 weeks	20	8.35 (0.53)	0.002	0.004
SI[4] in past week				
Yes	49	7.85 (0.42)		
No	53	8.14 (0.51)	0.015	0.007

[1] Bacterial cell-units, [2] Standard deviation, [3] Bacterial vaginosis diagnosis: BV-negative (BV–), BV-intermediate (BVI), BV-positive (BV+), BV-none (BV0), [4] Sexual intercourse.

Summary – No age-independent differences in vaginal bacterial concentrations as measured by flow cytometry were observed in HIV-R vs. HIV-N individuals, contradicting Hypothesis 3.

Flow cytometry reliably detects and quantitates three cell populations (bacteria, mucosal cells and sperm) in vaginal specimens using a technique that is rapid and simple (175). A strong correlation between flow cytometry cell counts (BCU/MCU) and sample weight was observed in both pilot study and Nairobi samples.

The concept of BV as bacterial overgrowth was corroborated by the observation that BV+ individuals had significantly higher BCU/g. Interestingly, those with no bacteria observed on Gram stain also had the lowest counts of BCU/g. Therefore, the observation of an inverse correlation between BCU/g and vaginal pH was surprising. BV– and BV0 individuals with lower BCU/g also had higher vaginal pH, while this association was not observed in BVI or BV+ samples. These findings further demonstrate the complex dynamics of bacterial vaginosis and may specifically inform transition from BV– to BV+ states and vice versa.

Individuals with leukocytes on Gram stain also had increased BCU/g and MCU/g compared to those without, indicating that mucosal inflammation in the context of elevated bacterial concentrations can be measured by flow cytometry.

Further development of species-specific probes and antibodies will be required to realize the full potential of this simple quantitative method to address questions about absolute quantities of *Lactobacillus* and BV organisms in vaginal specimens.

4. Culture-based quantification and functional analysis of vaginal specimens

Since the presence of vaginal *Lactobacillus* producing H_2O_2 and scavenging mannose have been implicated in protection against HIV, the primary objective of this chapter was to determine if increased vaginal *Lactobacillus* colonization as measured on selective media, as well as increased H_2O_2 production and acid production on mannose-containing media, are associated with HIV resistance.

Pilot study – All pilot study samples were plated on selective Rogosa media to semi-quantitatively assess *Lactobacillus* colonization (Table 6.1). Absence of growth was observed in 13/103 or 13% of samples, while confluent growth was observed in 35% of samples. A subset of 48 samples from 16 individuals with a range of BV diagnoses was selected in order to optimize several culture-based assays, including quantification of *Lactobacillus* colony-forming units (CFU) on Rogosa with H_2O_2-detecting chromogens (Rogosa-TMB) and assessment of acid production on modified Brucella media with mannose as the only carbohydrate source (mBru-mannose) (Figure 6.1).

Growth on Rogosa-TMB media was observed in 36 out of 48 pilot study samples (75%). CFU counts correlated positively with sample weight ($R^2 = 0.12$, $p=0.02$ by linear regression) and with semi-quantitative results on Rogosa ($p<0.0001$ by Kruskal Wallis test, not shown). After adjusting for sample weight, median $\log_{10}$ CFU/g was 6.52 (range 0-8.92). All culture-positive plates were also positive for H_2O_2 production. In terms of acid production on mBru-mannose, 12/48 or 25% of samples produced no acid, 20/48 or 42% produced weak/medium acid, and 16/48 (33%) produced strong acid.

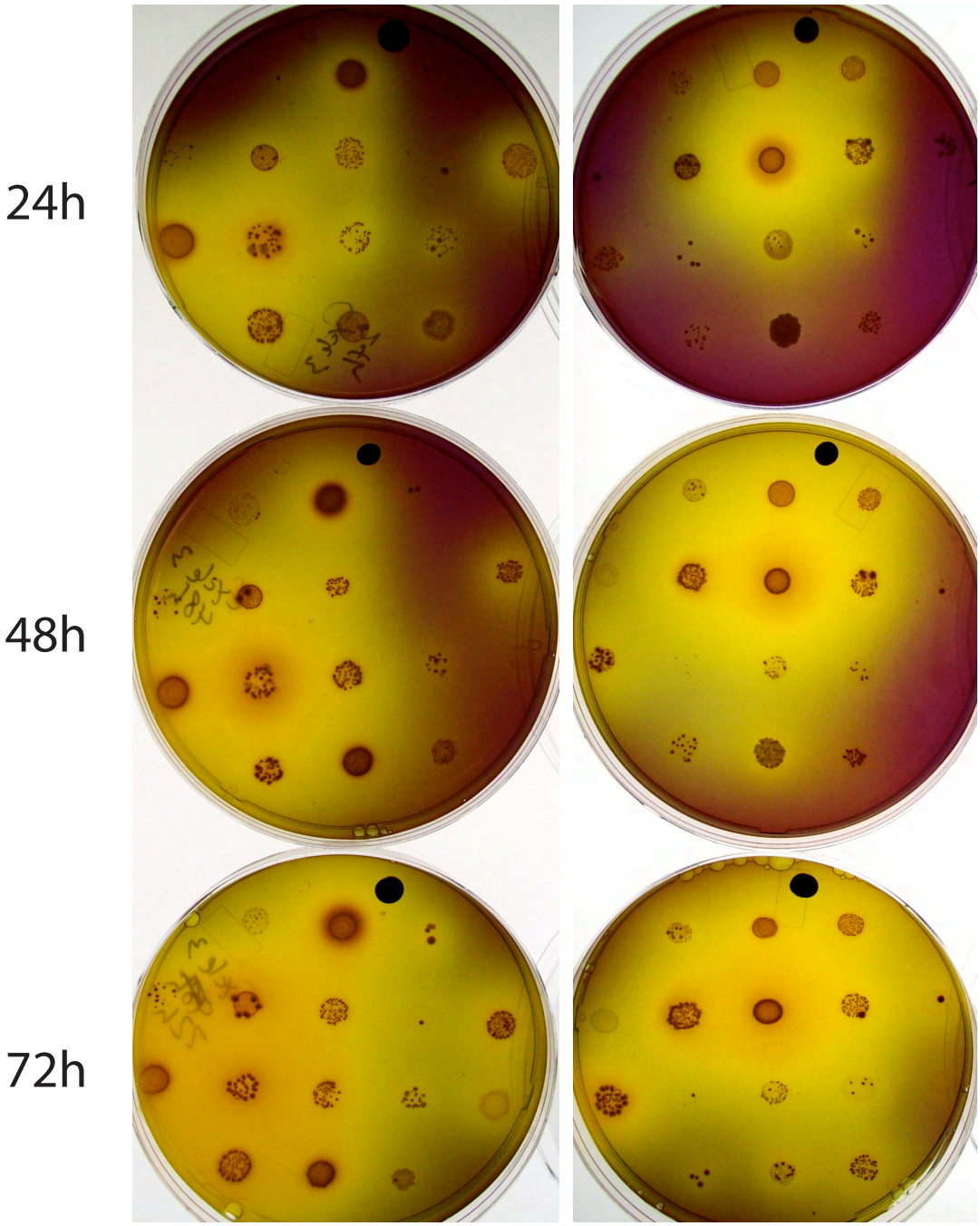

Figure 6.1 – Acid production on mBru-mannose

Two sets of three mBru-mannose plates, each showing results for 16 samples at 24h, 48h and 72h. Yellow colour indicates reduced pH of media surrounding culture. Positive and negative controls were done on separate plates under identical conditions (observed at 48h only, not shown). Black dot on left side of plate is India ink to facilitate plate orientation.

Interestingly, the appearance of Rogosa-TMB plates from sequential samples of the same individual was very similar, despite quantitative fluctuations, indicating short-term stability of culturable *Lactobacillus* populations (not shown). Lack of growth on Rogosa led to the decision to plate thawed Nairobi samples on mBru as well as Rogosa, in order to ensure recovery of isolates from all individuals. mBru is less acid than Rogosa and therefore is likely to be less selective for strict acidophiles.

Culturable Lactobacillus and functional characterization (Nairobi study) – Fresh specimens in Nairobi were plated onto Rogosa-TMB in order to assess *Lactobacillus* colonization and H_2O_2 production (Table 6.1). Most samples that were positive for growth on Rogosa were H_2O_2+, with striking variation in quantity and quality of blue colour (Figure 6.2). Based on these observed differences, Rogosa-positive samples (N=152) were classified in three groups: H_2O_2 negative, weak H_2O_2 and strong H_2O_2.

Of 96 Nairobi samples in the Culture subset, growth on Rogosa media was observed in all but 8 (8%) samples (Table 6.1). On pH-neutral mBru media, growth was observed in all but 1 sample, however CFU could not be quantified in 13 samples since there were more than 300 colonies even at the highest dilution carried out for that sample. These samples were assigned a CFU/g value of 9, slightly more than the highest value for samples with quantifiable CFU. Growth on mBru was observed for all samples in which there was no growth on Rogosa, and vice versa.

Rogosa and mBru CFU/g were significantly correlated (R^2 = 0.10, p=0.002 by linear regression, not shown). The difference in CFU/g between the two media was calculated, with positive values indicating increased growth on Rogosa, 0 indicating identical growth, and negative values indicating increased growth on mBru (Table 6.1).

Table 6.1 – Culture-based characteristics (Nairobi)

	Main (N=242)	**Culture** (N=96)	**Pyro** (N=48)	**qPCR** (N=40)	**Clone** (N=10)
Growth on Rogosa					
No growth	90 (37%)	35 (36%)	19 (40%)	15 (38%)	4 (40%)
>300 colonies	39 (16%)	13 (13%)	6 (12%)	6 (15%)	1 (10%)
Semi-confluent	40 (17%)	19 (19%)	8 (17%)	8 (20%)	--
Confluent	73 (30%)	29 (30%)	15 (31%)	11 (28%)	5 (50%)
H_2O_2 production (Rogosa+ only)	*N=152*	*N=61*	*N=29*	*N=25*	*N=6*
None	29 (19%)	13 (21%)	8 (28%)	5 (20%)	2 (33%)
Weak	89 (59%)	37 (61%)	15 (52%)	15 (60%)	2 (33%)
Strong	34 (22%)	11 (18%)	6 (21%)	5 (20%)	2 (33%)
Rogosa CFU/g[1]					
Median	--	6.05	6.24	6.09	5.57
Range		0-9.02	0-9.02	0-9.02	0-8.20
No growth		8 (8%)	3 (6%)	3 (8%)	1 (8%)
H_2O_2+		77 (80%)	41 (85%)	34 (85%)	9 (90%)
mBru CFU/g[1]					
Median	--	6.11	6.66	6.08	5.93
Range		0-9.00	0-9.00	0-9.00	0-9.00
No growth		1 (1%)	--	--	--
H_2O_2+		74 (77%)	37 (77%)	25 (63%)	6 (60%)
Difference in Rogosa[2]					
Median	--	–0.2	–0.3	–0.1	–0.3
Range		–9.0-3.8	–9.0-3.8	–9.0-3.8	–9.0-3.8
Acid production[3] (24h)					
None	--	40 (42%)	18 (38%)	20 (50%)	7 (70%)
Weak/medium		45 (47%)	24 (50%)	17 (43%)	2 (20%)
Strong		11 (11%)	6 (12%)	3 (7%)	1 (10%)
Acid production[3] (72h)					
None	--	12 (12%)	6 (13%)	6 (15%)	3 (30%)
Weak/medium		57 (59%)	27 (57%)	27 (68%)	6 (60%)
Strong		27 (28%)	15 (30%)	7 (18%)	1 (10%)

[1] Colony-forming units per gram sample on Rogosa or mBru media, [2] Difference of CFU/g between Rogosa and mBru media (ie. Rogosa CFU/g minus mBru CFU/g). [3] Acid production on mBru-mannose after 24h or 72h.

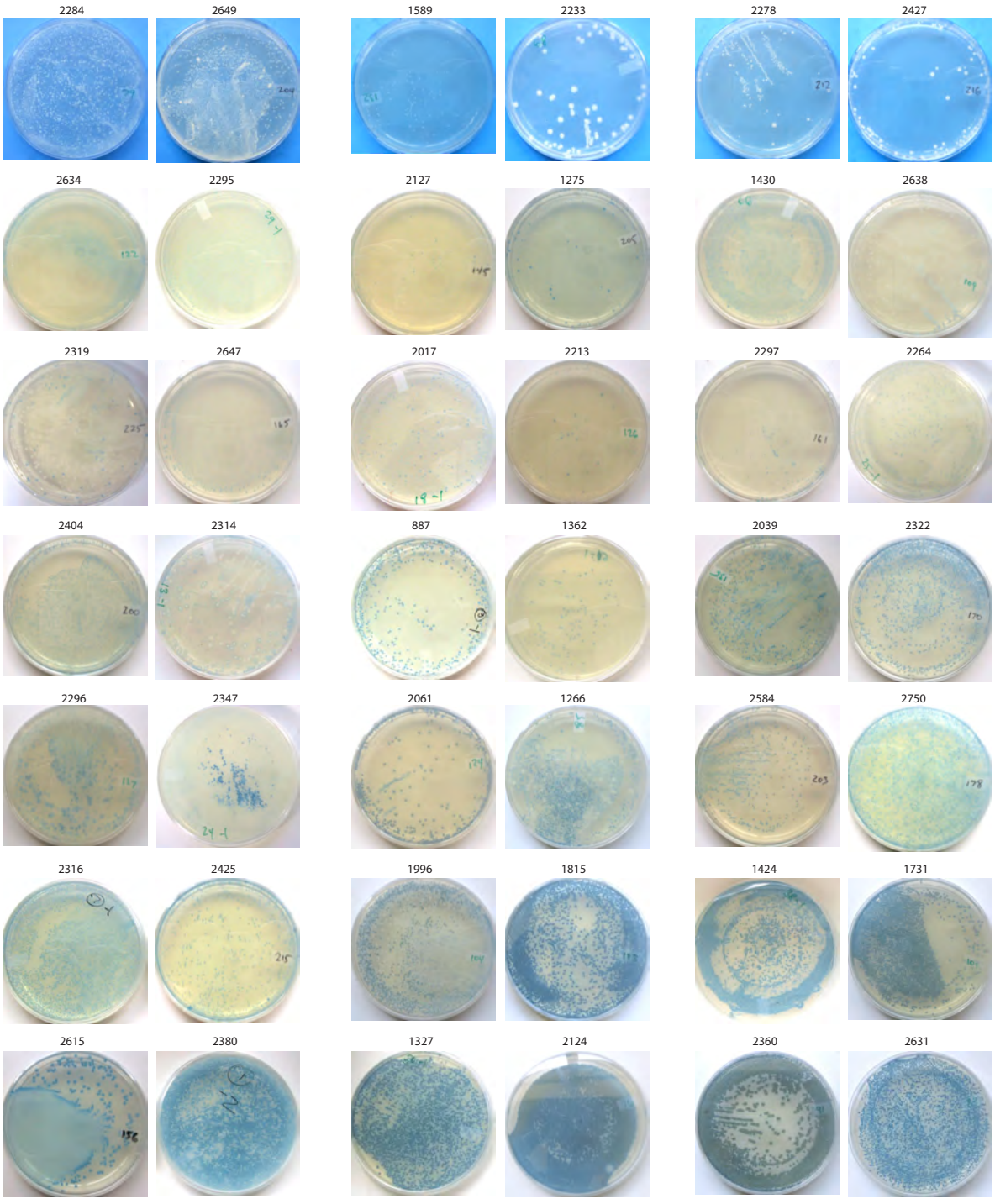

Figure 6.2 – H_2O_2 production on Rogosa-TMB

Striking variation in quantity and quality of blue colour indicating H_2O_2 production by bacteria in selected fresh samples. Top row shows culture positive, H_2O_2– samples, while other rows show increasing levels of H_2O_2 production.

Interestingly, there were significantly more CFU/g (ie. greater than 1 log difference) on Rogosa compared to mBru in 28/96 (29%) of samples. This observation was surprising since mBru is pH-neutral (therefore, less selective) and was designed to favour growth of a wider variety of *Lactobacillus* species (165). These findings indicate that the two media chosen for this study provide distinct "snapshots" of culturable bacteria in the samples. The profile generated is also unique to each individual – virtually no two samples are alike (Figure 6.3).

As in the pilot study, there was a strong, stepwise correlation between quantitative CFU and semi-quantitative abundance on Rogosa media ($p<0.0001$ by Kruskal Wallis test, not shown), indicating that semi-quantitation on a single plate is an acceptable alternative to serial dilution and multiple plates for assessing culturable *Lactobacillus* in vaginal specimens. Despite being correlated with Rogosa CFU/g, mBru CFU/g were not significantly correlated with semi-quantitative results, confirming that a distinct population of culturable bacteria are captured on this media.

H_2O_2 production on either Rogosa or mBru was detected in 89/96 samples (91%), including 5 samples with growth but no H_2O_2 production on either media, and 2 samples with no growth on Rogosa and no H_2O_2 production on mBru. Since mBru was designed for increased detection of H_2O_2production in isolates (165), it was surprising that there were 14 samples in which H_2O_2 was detected on Rogosa but not on Brucella media, but only 6 samples in which the reverse was true. Most samples (61/96 or 62%) were H_2O_2+ on both media. Several samples produced colonies with very dark blue colour and some with green or even yellow colour (eg. individual 2360, Figure 6.3), indicating an increased concentration of H_2O_2 in the surrounding media (102).

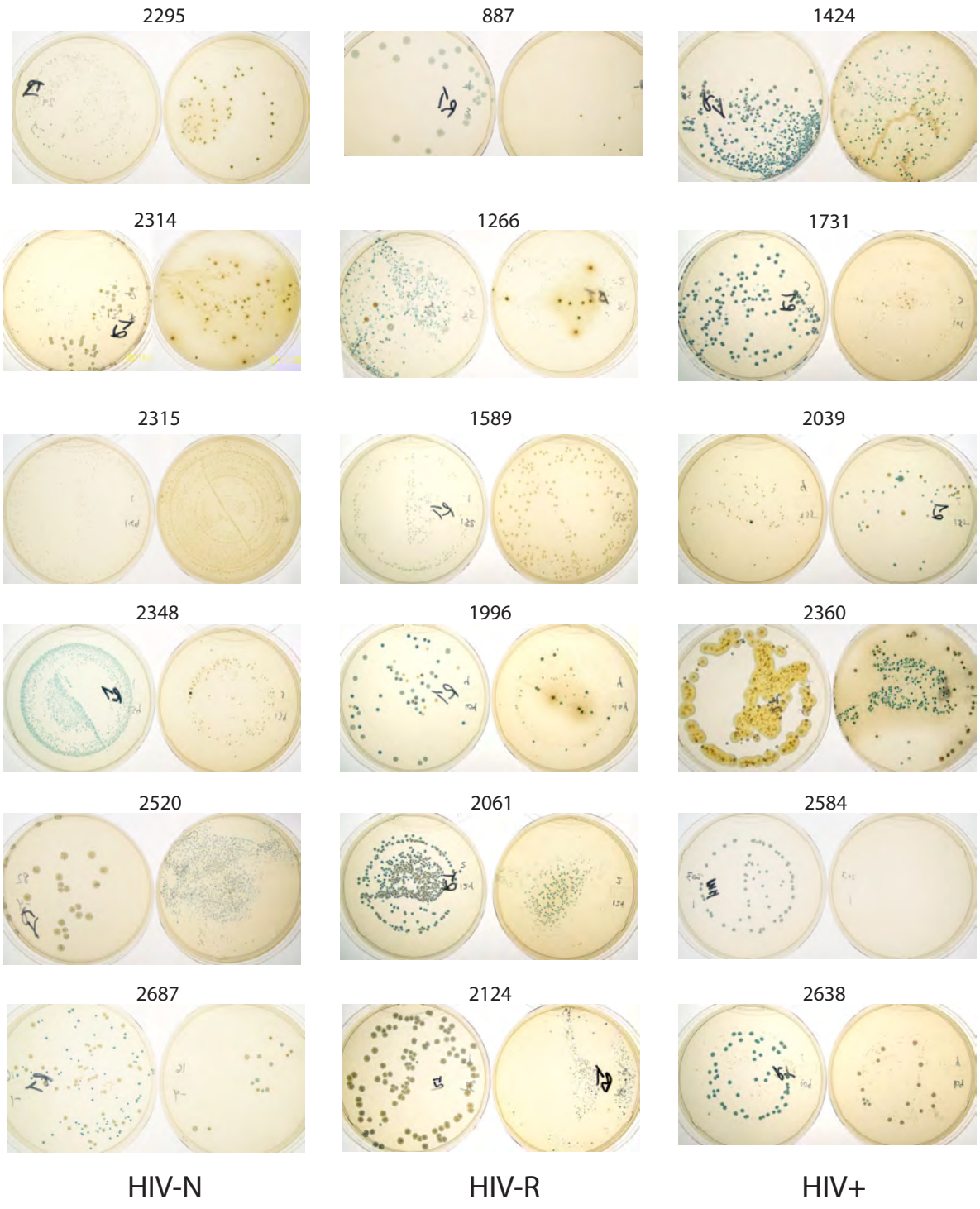

Figure 6.3 – Quantification of CFU on Rogosa-TMB and mBru-TMB

Appearance of selected Nairobi samples on Rogosa (left-hand plate) and mBru (right-hand plate), indicating that each media provides a unique snapshot of culturable bacteria in vaginal samples. Note bright yellow colonies in individual 2360, indicating very high concentrations of H_2O_2 in the surrounding media.

Most samples (58%) produced acid on mBru-mannose after 24h and almost all (88%) produced acid after 72h. It was somewhat surprising that acid production was not associated with sample weight at either timepoint, since greater sample weight implies greater concentrations of bacteria that can produce acid. However, production of acid was associated with increased Rogosa CFU/g ($p=0.002$ at 24h, $p=0.007$ at 72h) and increased mBru CFU/g ($p=0.0007$ at 24h, $p=0.001$ at 72h, both by Kruskal Wallis, not shown).

Comparison of culture-based characteristics in HIV-R vs. HIV-N – No significant differences in semi-quantitative abundance of culturable *Lactobacillus* or detection of H_2O_2 production were observed between HIV-R and HIV-N individuals (Table 6.2). In the Culture subset, no differences were observed in Rogosa or mBru CFU/g, detection of H_2O_2 production on either media, or acid production on mBru-mannose at either timepoint. These findings are inconsistent with Hypotheses 4, 5 and 6.

Culture-based characteristics, BV diagnosis and vaginal pH – In pilot study and Nairobi samples, semi-quantitative abundance of culturable *Lactobacillus* was decreased in those with any BV diagnosis compared to BV– individuals ($p<0.0001$ in the pilot study, $p=0.0003$ in the Nairobi study by chi square test, data not shown), indicating an expected association between BV and reduced culturable *Lactobacillus* in either study group.

In the pilot study, individuals with any BV diagnosis had median Rogosa CFU/g counts 3.5 logs lower than BV– (3.41 vs. 6.94, $p<0.0001$ by Mann Whitney test), while in Nairobi samples this difference was nearly 2 logs (5.09 vs. 6.94, $p=0.02$ by Mann Whitney test). In contrast, BVI and BV+ individuals in Nairobi both had significantly higher median mBru BCU/g compared to BV– individuals. (7.15 for BVI, 7.24 for BV+, compared to 5.97 for BV–, $p=0.02$ and $p=0.01$ respectively by Mann Whitney test).

Table 6.2 – Comparison of culture-based characteristics between HIV-R and HIV-N

	HIV-R (N=61)	**HIV-N** (N=57)	**Difference*** All samples	Meets IC[1]
Growth on Rogosa				
No growth	24 (39%)	24 (42%)	--	--
>300 colonies	10 (16%)	7 (12%)		
Semi-confluent	10 (16%)	8 (14%)		
Confluent	17 (28%)	18 (32%)		
H_2O_2 production (Rogosa+ only)	*N=37*	*N=33*		
None	8 (22%)	6 (18%)	--	--
Weak	23 (62%)	23 (70%)		
Strong	6 (16%)	4 (12%)		
Culture subset only	*N=32*	*N=32*		
Rogosa CFU/g[2]				
Median	6.10	6.29	--	--
Range	0-9.02	0-8.18		
No growth	1 (2%)	4 (7%)		
H_2O_2+	29 (91%)	24 (75%)		
mBru CFU/g[2]				
Median	5.99	6.55	--	--
Range	0-9.00	4.30-9.00		
No growth	1 (2%)	--		
H_2O_2+	23 (72%)	24 (75%)		
Difference in Rogosa[3]				
Median	0.20	–0.36	--	--
Range	–5.6-3.8	–9.0-2.5		
Acid production[4] (24h)				
None	15 (47%)	14 (44%)	--	--
Weak	13 (41%)	14 (44%)		
Strong	4 (12%)	4 (12%)		
Acid production[4] (72h)				
None	5 (16%)	3 (9%)	--	--
Weak	22 (68%)	18 (56%)		
Strong	5 (16%)	11 (35%)		

* Indicates significant difference in distribution between HIV-R and HIV-N, by chisquare/Fisher's exact test (ordinal/categorical variables) or Mann Whitney test (continuous variables), [1] Meets inclusion criteria, including active CSW, not pregnant or past menopause, no symptoms, diagnoses or current STI, [2] Colony-forming units per gram sample on Rogosa or mBru media, [3] Difference of CFU/g between Rogosa and mBru media (ie. Rogosa CFU/g minus mBru CFU/g). [4] Acid production on mBru-mannose after 24h or 72h.

This observation suggests that mBru media is more likely to grow bacteria associated with being BVI/BV+. Consistent with these observations, increased Rogosa CFU/g was correlated with decreased vaginal pH (R^2 = 0.17, $p<0.0001$ by linear regression, not shown), but mBru CFU/g was not (p=0.7 by linear regression, not shown). These findings indicate that Rogosa media reflects vaginal bacteria related to being BV–, while mBru media is more reflective of BVI and BV+ microbiology.

As expected, H_2O_2 production on semi-quantitative Rogosa plates was associated with decreased vaginal pH (p=0.008 by Kruskal Wallis test, data not shown), and BV+ individuals had reduced H_2O_2 production compared to BV– individuals (p=0.05 by chi square test, not shown).

Surprisingly, individuals with acid production on mBru-mannose had higher median vaginal pH and were more likely to have BV (Figure 6.4). Samples with no acid production at either timepoint had significantly lower median vaginal pH (Figure 6.4, A/B), while acid production was least likely to be observed in BV– individuals at 24h, and many of these were still acid-negative by 72h (Figure 6.4, C/D). In contrast, a smaller proportion of BV+ samples were acid-negative at 24h and all of these were acid-positive by 72h. Interestingly, all BVI samples were acid positive at both timepoints, possibly indicating increased acid-producing bacteria in BVI samples.

Similar results were observed in the pilot study, with 9/24 or 38% of BV– samples producing no acid on mannose after 48h incubation, compared to only 2/22 or 9% of BVI/BV+ samples (p=0.04 by Fisher's exact test). These data indicate that BV– samples contain bacteria that are less likely to produce acid from mannose than bacteria in BVI or BV+ samples.

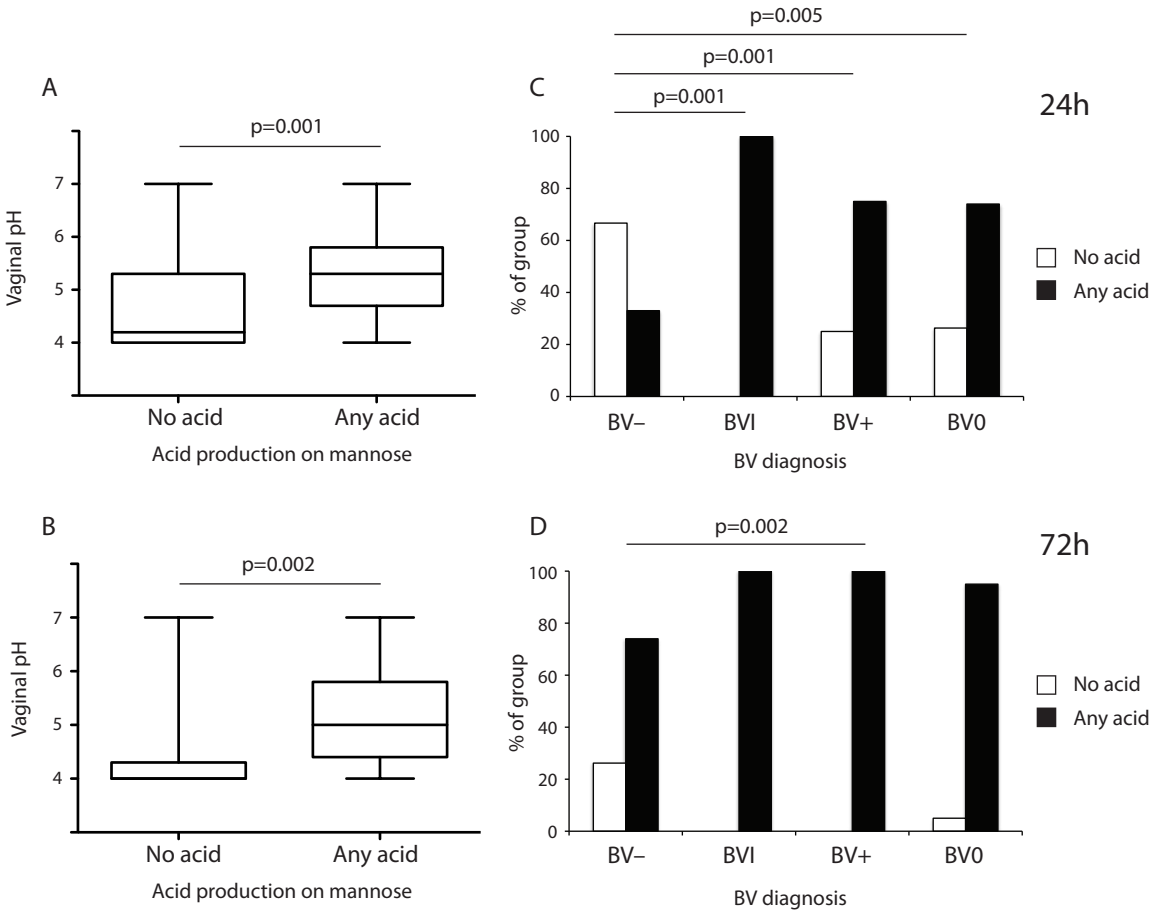

Figure 6.4 – Acid production on mannose in relation to BV diagnosis and pH

Significant association between acid production on mBru-mannose, increased vaginal pH and BV diagnosis. A. Increased vaginal pH in samples producing acid at 24h and, B. at 72h. C. Increased likelihood of acid production in BV samples compared to BV– samples at 24h and, D. at 72h. p values were calculated using the Mann Whitney test (A/B) and chi square test (C/D) in GraphPad Prism v5.

These observations are surprising considering that BV– samples are associated with decreased vaginal pH and therefore presumably with acid-producing bacteria, however the potential for biological significance is strengthened by similar observations in two socially and geographically distinct groups of women. One possible explanation for these findings is that the test media somehow favours acid production in bacteria associated with BVI or BV+ samples rather than bacteria in BV– samples. Another possibility is that bacteria that proliferate on the test media are more likely to produce acid when exposed to BV-related organisms. Increased acid production by bacteria in the context of BVI and BV+ microbiota may be in reaction to the presence of BV-related bacteria, with intriguing implications for the transition from BV+ to BV– status.

Summary – No significant differences in absolute quantity of culturable *Lactobacillus*, H_2O_2 production or acid production from mannose by vaginal samples were observed between HIV-R and HIV-N individuals, contradicting Hypotheses 4, 5 and 6.

Quantitative CFU results from multiple plates were strongly correlated with semi-quantitative results on a single plate in both the pilot study and in Nairobi, indicating that semi-quantitation is a valid approach for rapid characterization of culturable *Lactobacillus* concentrations in vaginal specimens.

Lack of growth on Rogosa for many samples in the pilot study led to the decision to use a second media recently reported to improve detection of a wider range of *Lactobacillus* species for quantification and strain isolation (mBru-TMB). This approach led to culturable isolates for all samples. Since mBru-TMB is less selective (neutral pH) it was expected that everything that grew on Rogosa-TMB would also grow on mBru-TMB, however for many samples there was increased CFU/g on Rogosa compared to mBru. This

indicates that these media provide distinct snapshots of culturable bacteria in vaginal specimens. Culture on both of these media with H_2O_2-detecting chromogens revealed the idiosyncracies of each sample. No two samples are alike, indicating that phenotypic profile of culturable *Lactobacillus* and other vaginal organisms may be completely unique to each individual.

Since mBru-TMB was also designed to enhance H_2O_2 production in *Lactobacillus* organisms, it was also surprising that many samples were H_2O_2+ on Rogosa-TMB but not mBru-TMB. Some organisms produced very high levels of H_2O_2 as evidenced by bright green or yellow colour on chromogenic media.

As expected, decreased CFU/g on Rogosa was strongly associated with BV in both the pilot study and in Nairobi, as well as with increased vaginal pH. In contrast, increased CFU/g on mBru was associated with BVI and BV+ samples, indicating that mBru captures bacteria that are characteristic of BV.

Unexpectedly, acid production on mBru mannose was associated with increased vaginal pH and BVI, BV+ and BV0 diagnoses in both the pilot study and Nairobi samples, indicating that BV samples contain organisms more likely to produce acid on test media compared to BV– samples. Since BV is associated with reduced vaginal pH, these observations require further elucidation. One intriguing possibility is that bacteria growing on test media are more likely to produce acid in the presence of BV-related organisms, perhaps as part of the transition from BV to BV– states.

5. Determination of acid and H_2O_2 production by vaginal isolates

Building on the results from the previous chapter, over 750 unique bacterial strains were isolated from 144 vaginal specimens and a phenotypic profile based on H_2O_2 production and acid production on mannose was generated for each isolate using chromogenic media. The goal of this chapter is to determine if HIV-R individuals are more likely to be colonized with H_2O_2/acid-producing isolates compared to HIV-N individuals.

Optimization of phenotypic assays using panel strains – Phenotyping experiments were optimized using a panel of relevant bacterial strains (see Table 1.1, p.50). Initially, 12 strains were cultured on Mann-Rogosa-Sharpe media for *Lactobacillus* (MRS), mBru-TMB for H_2O_2 detection and mBru-mannose for detection of acid production at 24h and 48h (Figure 7.1, A). Observed differences in H_2O_2 and acid production led to selection of *L. crispatus* and *E. faecalis* as experimental controls. *L. crispatus* was selected as positive control (blue) for H_2O_2 production and negative (weak) control for acid production, while *E. faecalis* was selected as negative control for H_2O_2 production (white) and the positive (strong) control for acid production (see EF and LC in Figure 7.1).

Striking differences in the quality of the blue colour generated by different strains on mBru-TMB media were observed; for example, the green phenotype displayed by *L. salivarius* AWH. This may be attributed to the different-coloured oxidation products of TMB by HRP in the presence of H_2O_2 (blue, yellow or a mixture of the two producing green). At higher H_2O_2 concentrations, the blue product is either partially degraded, resulting in green colour, or completely degraded, resulting in yellow colour (102).

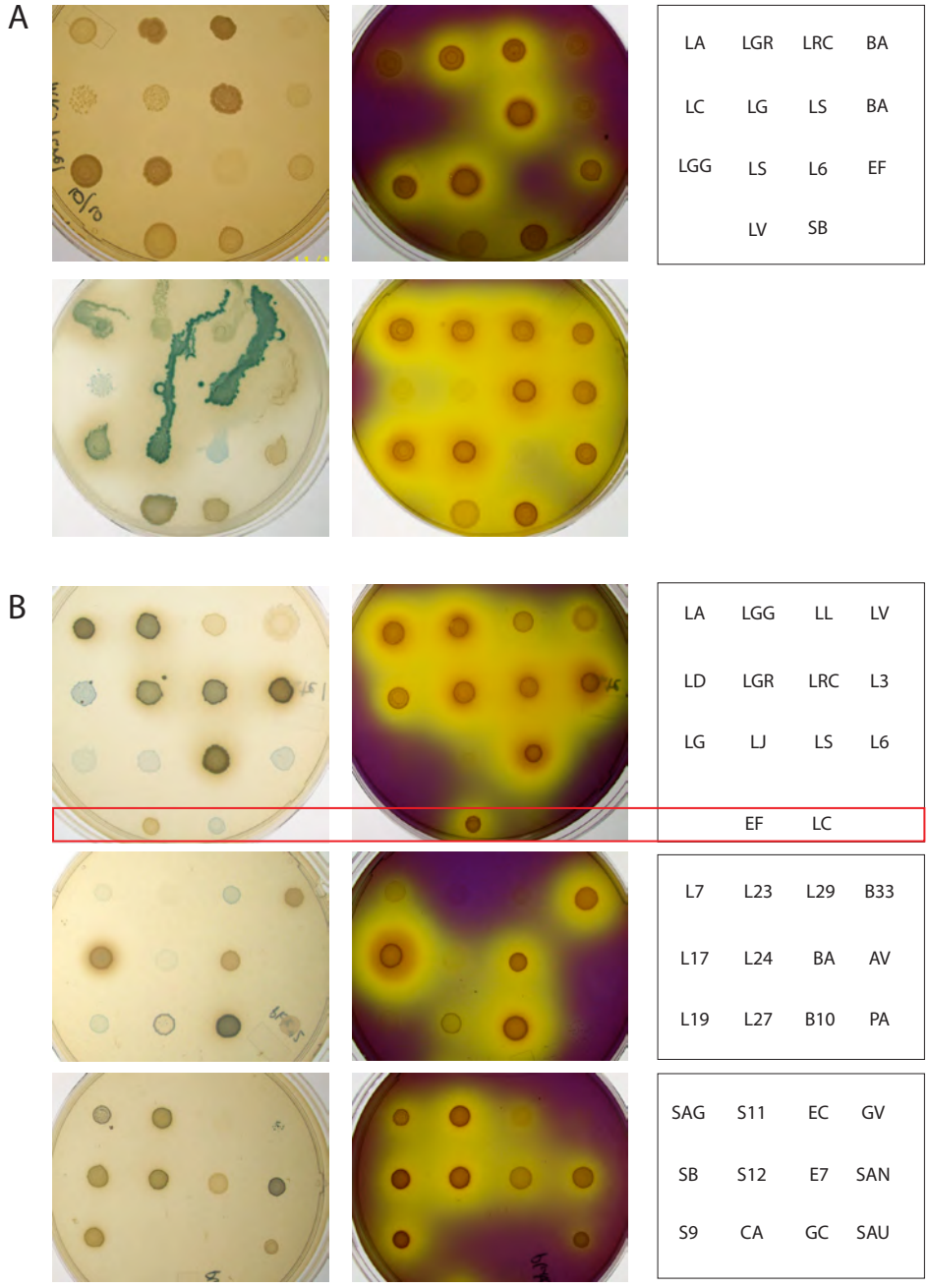

Figure 7.1 – Phenotyping of panel micro-organisms

Overnight cultures of panel strains were adjusted to 0.5 McFarland and spotted onto different media for phenotypic analysis using a replicator. Yellow colour indicates acid production and blue colour indicates H_2O_2 production. A. Initial experiment with 14 strains on MRS (top left), mBru-TMB (bottom left), and mBru-mannose (top right, 24h incubation; bottom right, 48h incubation). B. Follow-up with 36 strains on mBru-TMB (left) and mBru-mannose (right). The first experiment led to selection of E. faecalis (EF) and L. crispatus (LC) as positive and negative controls for future experiments (red box).

(continued…)

Figure 7.1 (continued):

Key to strain identifiers: **LA**: *Lactobacillus acidophilus* 8/4, **LC**: *L. crispatus* DSMZ, **LGG**: *L. rhamnosus* GG, **LGR**: *L. rhamnosus* GR-1, **LG**: *L. gasseri* DSMZ, **LS**: *L. salivarius* AWH, **LV**: *L. vaginalis* DSMZ, **LRC***: L. fermentum* RC-14, **L6**: *L.* sp. L6, **LD**: *L. delbrueckii* 151, **LI**: *L. iners* DSMZ, **LJ**: *L. jensenii* DSMZ, **L3***: L. salivarius* N3, **L7**: *L. gasseri* N7, **L17**: *L. plantarum* N17, **L19**: *L.* sp. N19, **L23**: *L.* sp. N23, **L24**: *L. jensenii* N24, **L27**: *L.* sp. N27, **L29**: *L.* sp. N29, **LL**: *Lactococcus lactis* ATCC, **SB**: *Streptococcus bovis* ATCC, **SAG**: *S. agalactiae* ATCC, **SAN**: *S. anginosus* ATCC. **S9**: *S. lutetiensis* N9, **S11**: *S. bovis* N11, **S12**: *S.* sp. S12, **EF**: *Enterococcus faecalis* ATCC, **BA**: *Bifidobacterium animalis* B30, **B10**: *B. infantis* N10, **B33**: *B. pullorum* N33, **AV**: *Atopobium vaginae* DSMZ, **PA**: *Peptostreptococcus anaerobius* ATCC, **CA**: *Candida albicans* ATCC, **EC**: *Escherichia coli* BL-21, **E7**: *E. coli* E7, **GC**: *Neisseria gonorrhea* G45, **GV**: *Gardnerella vaginalis* ATCC. See Table 1.1 (p.51) for full descriptiion of strains.

Interestingly, four probiotic strains (*Lactobacillus* GG, GR1, RC14 and *L. salivarius* AWH) were strongly H_2O_2+ and acid-positive ("double-strong"), indicating two potential mechanisms of how these strains exert their beneficial effects. In contrast, two classical vaginal strains, *L. crispatus* and *L. gasseri*, both from DSMZ, and a Nairobi isolate (L6), were pale blue on H_2O_2 media and were also weak acid producers.

Phenotypic experiments were repeated with an expanded panel of 36 organisms, incubated for 48h (Figure 7.1, B). Again, probiotic strains (*L. acidophilus* 8/4, GG, GR-1, RC-14 and AWH) all had a double-strong phenotype. Classical vaginal *Lactobacillus* strains, *L. crispatus*, *L. gasseri* and *L. jensenii*, as well as the Nairobi isolate L6, were all weakly acid-positive and H_2O_2+ (pale blue).

Selection strategy – The overall strategy to select isolates from CFU plates was to sample the greatest diversity possible based on colony appearance (size and morphology). The major benefit of this approach compared to random selection of isolates is that a greater diversity of bacteria is sampled by selecting a relatively small number of isolates, however non-identical colonies with similar appearance may not be selected. To address this, bacterial strains were selected from two different culture media, Rogosa and mBru, both with H_2O_2-detecting chromogens. The colour reaction greatly increased the number of

phenotypic criteria used to select colonies, including being able to distinguish between H_2O_2+ and H_2O_2– colonies, as well as variation in blue/green colour, blue-ringed vs. blue-centred colonies, blue-speckled colonies, etc. (Figure 7.2).

Although the biological significance of colour variation in primary isolates can only be speculated about based on current data, choosing a representative colony from each colony type based on appearance and colour variation resulted in good representation of isolates with different phenotypes, including selection of all H_2O_2+ and H_2O_2– colony types. Limitations of this approach include an unequal number of isolates selected from each sample and a lack of quantitative information about the proportion of total vaginal bacteria represented by each phenotype.

Phenotyping of primary isolates – A total of 192 isolates were selected from 42 pilot study samples and 576 isolates were selected from 96 Nairobi samples, with 3-10 isolates selected per sample. All pilot study isolates were spotted onto mBru-TMB to detect H_2O_2 production, as well as onto mBru-mannose and mBru-glucose (identical to mBru-mannose but using glucose instead of mannose as the only carbohydrate source). Since results for acid production were identical on mBru-mannose and mBru-glucose in the pilot study, it was decided to use only mBru-mannose to detect acid production in Nairobi samples.

In order to enhance phenotypic profile of Nairobi strains, isolates were spotted onto MRS media and mBru without chromogen as well as chromogenic media (mBru-TMB and mBru-mannose). A Gram stain of isolate material picked from either MRS or mBru phenotyping plates was also prepared and photographed (Figure 7.3).

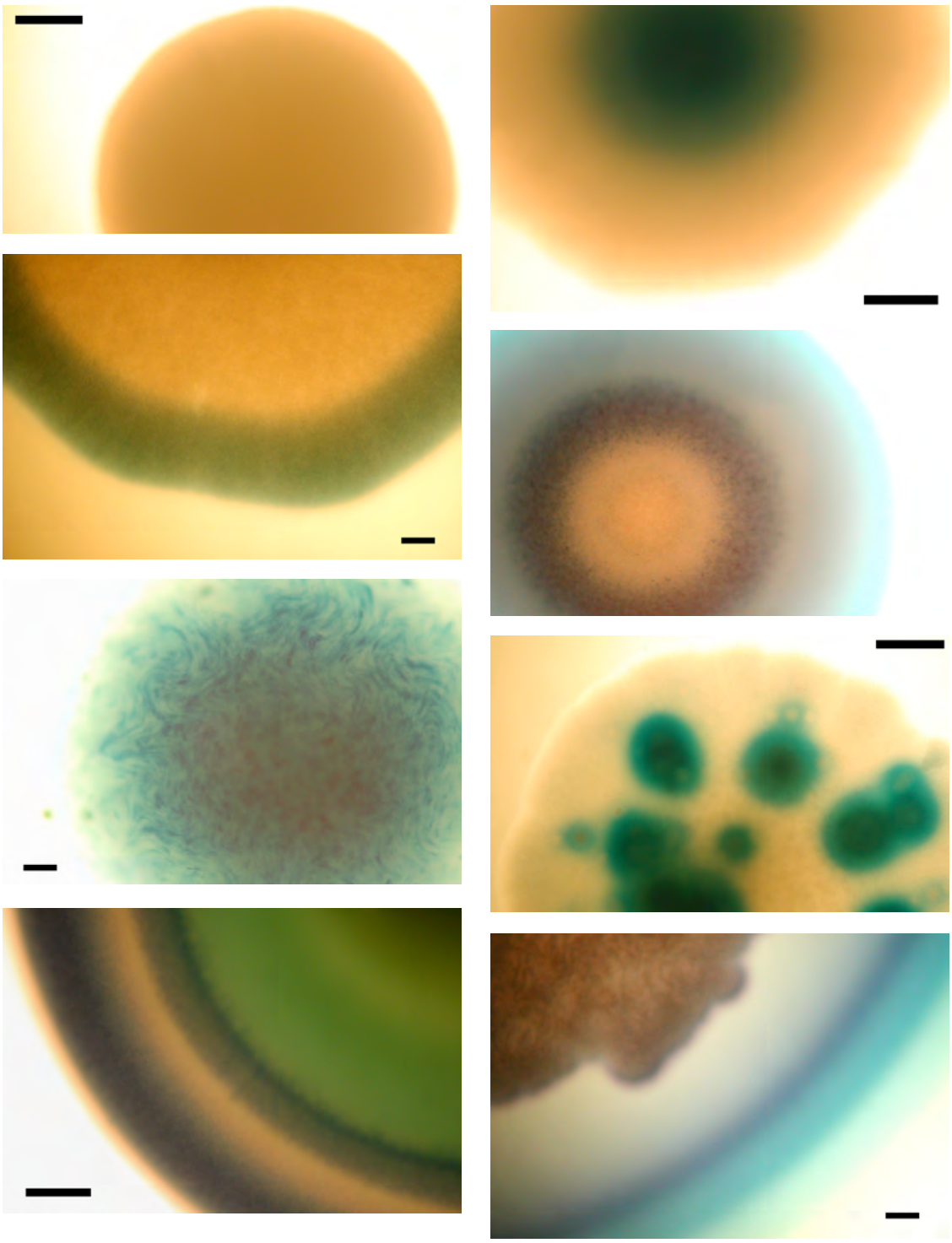

Figure 7.2 – Details of colonies selected for isolates

Representative colony phenotypes selected for isolation and phenotyping, including H_2O_2– (top left), and various H_2O_2+ phenotypes. Several often highly detailed patterns were observed (blue-ringed, blue-centred, blue-speckled, etc.), allowing for selection of a wide diversity of phenotypes. Bar = 100µm.

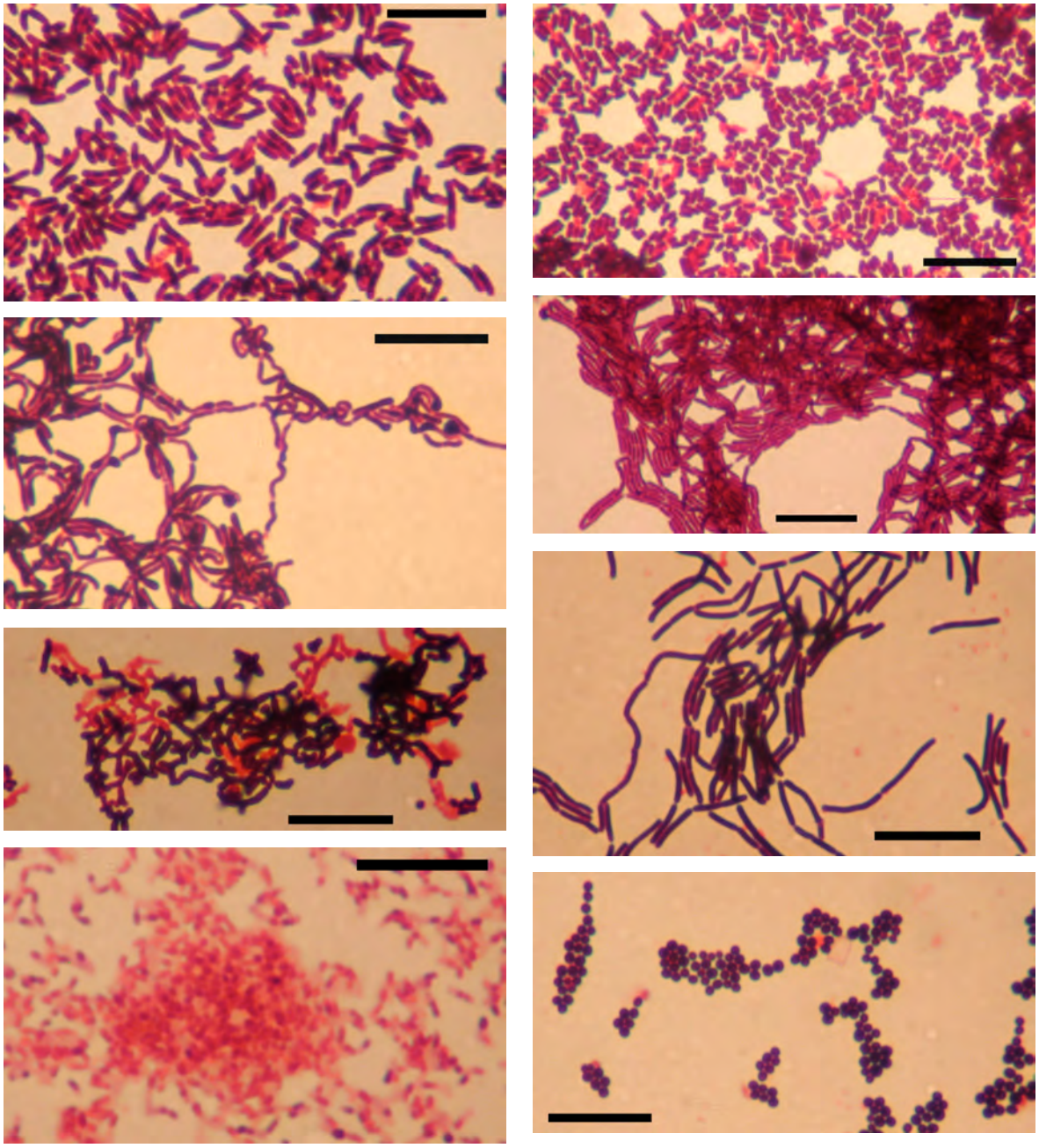

Figure 7.3 – Representative Gram stains of selected isolates

Shown are some of the major Gram stain patterns observed among isolates, including several types of Gram+ rods (top 6 panels; short, long, pleiomorphic, bifid), as well as Gram+ cocci (bottom right) and a unique Gram-variable phenotype isolated only from mBru media (bottom left). Bar = 10μm.

Approximately two-thirds of Nairobi isolates were acid positive, two-thirds were H_2O_2+, half were double positive, and 34 or 6% were double strong (Table 7.1, Figure 7.4). These double-strong isolates resemble bright yellow/dark blue phenotypic profile of probiotic strains examined during optimization experiments and therefore are considered "potential probiotics". Proportions of acid positive strains were similar in both studies, however significantly more Nairobi isolates were H_2O_2 producers compared to the pilot study (67% vs. 57%, p=0.01 by chi square test).

Most Nairobi isolates grew on MRS or mBru and two-thirds grew on both media. Only 5/576 isolates did not grow on either media during phenotyping experiments. Only 28% of double-negative isolates grew on both media, with 46% growing only on mBru media and a further 26% growing on MRS only. In contrast, 75/104 (71%) of H_2O_2+ isolates and 219/275 (80%) of double positive isolates grew on both MRS and mBru. These data suggest the existence of multiple subgroups of isolates as defined by growth on all four test media.

As expected, most of the Nairobi isolates were Gram+ rods, however one quarter of isolates were Gram+ cocci and 10% had a distinct Gram-variable aspect (Figure 7.3). A small number of isolates appear to be a mixture of Gram+ rods and cocci or other bacterial morphotypes, indicating that isolates picked from CFU plates did not always result in pure cultures. Therefore, an undetermined number of "isolates" likely consist of one or more bacterial strains. Identification and exclusion of data from mixed cultures was not attempted, since phenotypic data can be considered to validly represent an element that has been isolated from the microbiota, whether or not this element consists of one or more bacterial strains.

Table 7.1 – Phenotypic characteristics of isolates

	Pilot study (N=192)	**Nairobi study** (N=576)
Acid production		
None	66 (34%)	209 (37%)
Weak/Medium	85 (44%)	249 (43%)
Strong	41 (21%)	118 (20%)
H_2O_2 production		
Negative	83 (43%)	190 (33%)
Positive	109 (57%)	381 (67%)
Acid and H_2O_2 positive		
Double-negative	35 (18%)	103 (18%)
Acid positive only	48 (25%)	92 (16%)
H_2O_2 positive only	31 (16%)	106 (18%)
Double positive	78 (41%)	275 (48%)
Double-strong	2 (1%)	34 (6%)
Growth on MRS[1]		
Negative	--	124 (22%)
Positive		452 (88%)
Growth on mBru[2]		
Negative	--	77 (13%)
Positive		499 (87%)
Growth on MRS[1]/mBru[2]		
Neither	--	5 (1%)
mBru only		119 (21%)
MRS only		72 (13%)
Both		380 (66%)
Aspect on Gram stain		
Gram+ rods	--	337 (59%)
Gram+ cocci		150 (26%)
Gram-variable coccobacilli		58 (10%)
Gram– rods		5 (1%)
Mixture of morphotypes		22 (4%)
No cells		4 (<1%)

[1] MRS: Mann-Rogosa-Sharpe agar for Lactobacillus, [2] mBru: modified Brucella agar

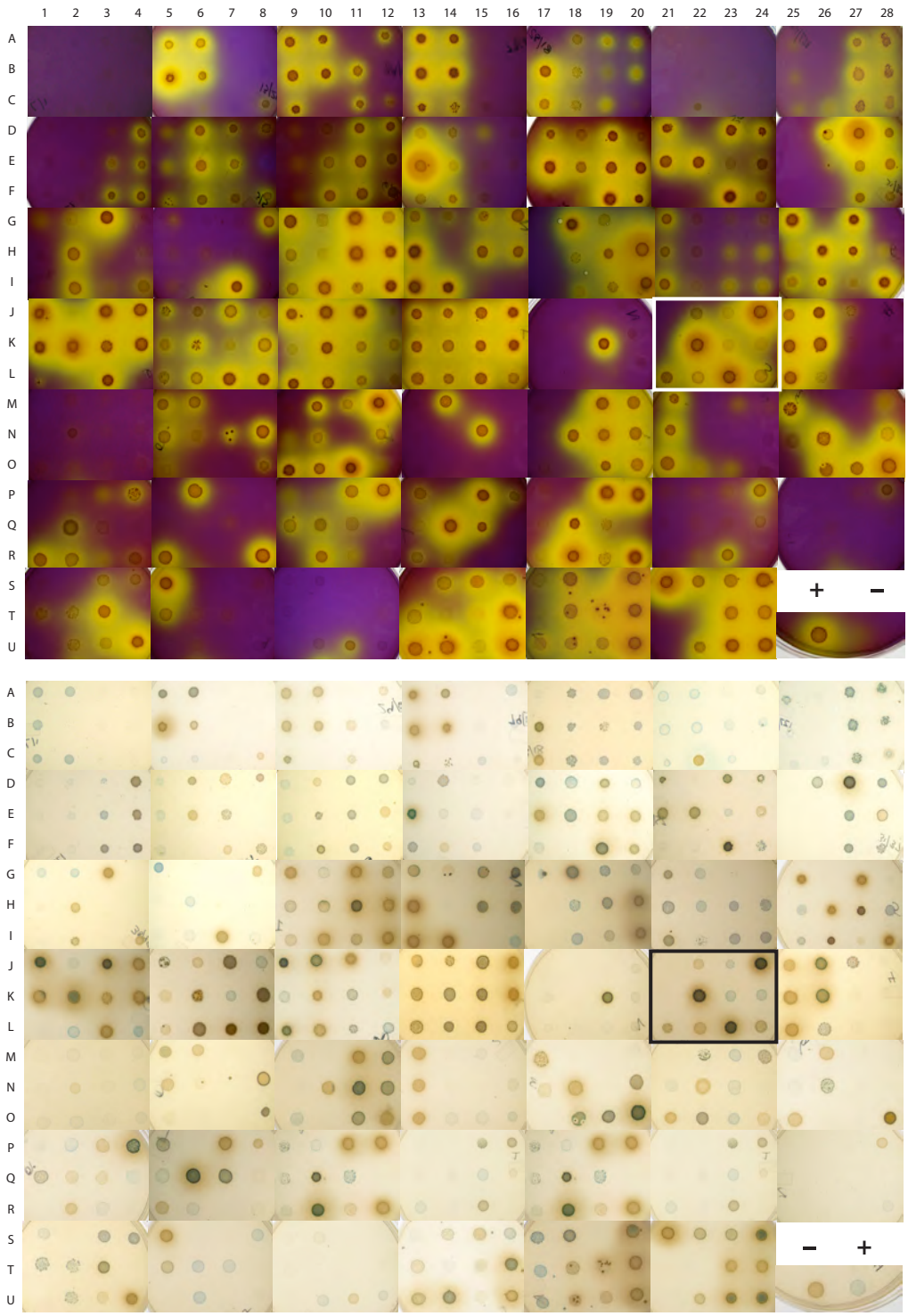

Figure 7.4 – Phenotypic analysis of primary Nairobi isolates

Acid production indicated by yellow colour (top) and H_2O_2 production indicated by blue colour (bottom) for 576 isolates from 96 individuals. Note "double strong" isolates in highlighted panel, resembling phenotype of probiotic strains. Positive and negative controls were run on each plate (examples shown in bottom right-hand corner).

Almost all Gram+ rods grew well on MRS but about 20% did not grow on mBru, while the opposite was seen for cocci (almost all grew well on mBru while about 30% did not grow on MRS). Almost all Gram-variable rods were isolated from mBru since hardly any grew on MRS. Less than 60% of Gram+ rods were acid positive, compared to nearly 90% of cocci and less than 30% of Gram-variable rods. However, a greater proportion of Gram+ rods were strong acid producers (25%) compared to cocci (20%). An equal proportion of Gram+ rods and cocci were H_2O_2+ (72%), compared to almost none of the Gram-variable rods. Just under half of the Gram+ rods were double positive for acid and H_2O_2, compared to two thirds of the cocci.

Phenotypic data from all test media and Gram stain analysis suggest several important subgroups of isolates, including Gram-variable rods that are double-negative and grow only on mBru, Gram+ rods that are double negative and grow only on MRS, Gram+ rods that are double positive and grow on both media and Gram+ cocci that are double positive and grow on both media. The significance of these subgroups will be defined further in subsequent chapters.

Isolate phenotype in individuals – A number of individual-level descriptors were derived from data about isolate phenotype, describing whether or not an individual was found to be colonized by isolates with defined phenotypic characteristics (Table 7.2).

Phenotypic diversity was defined by noting the total number of distinct colony phenotypes initially observed on Rogosa and mBru isolation plates (Table 7.2). The total number of isolate phenotypes per sample was defined by grouping isolates within a sample based on similar growth on all test media and Gram stain profile.

Table 7.2 – Phenotypes observed in individuals

	Culture (N=96)	**Pyro** (N=48)	**qPCR** (N=40)	**Clone** (N=10)
Initial phenotypes[1] (Rogosa)				
Median	3	3	3	3
Range	0-5	1-5	1-5	1-4
Initial phenotypes[1] (mBru)				
Median	2	3	3	3
Range	0-5	0-5	0-5	0-3
Total phenotypes[2]				
Median	3	4	3	4
Range	1-10	1-10	1-10	2-6
Acid positive isolates				
Any	84 (88%)	43 (90%)	36 (90%)	9 (90%)
Any strong	47 (49%)	26 (54%)	19 (48%)	5 (50%)
None	12 (12%)	5 (10%)	4 (10%)	1 (10%)
H_2O_2+ isolates				
Any	89 (93%)	44 (92%)	36 (90%)	9 (90%)
None	7 (7%)	4 (8%)	4 (10%)	1 (10%)
Double positive isolates				
Any	74 (77%)	38 (79%)	28 (77%)	8 (80%)
None	22 (23%)	10 (21%)	12 (30%)	2 (20%)
Any double strong	21 (22%)	15 (21%)	9 (23%)	2 (20%)
Double negative isolates				
Any	49 (51%)	22 (46%)	25 (62%)	5 (50%)
None	47 (49%)	26 (54%)	15 (38%)	5 (50%)
Aspect on Gram stain				
Any Gram+ rods	84 (88%)	43 (90%)	36 (90%)	10 (100%)
Any Gram+ cocci	58 (60%)	29 (60%)	24 (60%)	6 (60%)
Any Gram-variable rods	32 (33%)	17 (35%)	13 (33%)	1 (10%)

[1] Number of different phenotypes observed on Rogosa/mBru CFU plates, [2] Total number of unique phenotypes per sample, considering culture and Gram stain data.

Interestingly, the number of initial phenotypes observed on Rogosa does not correlate with number of phenotypes on mBru, but phenotypes on both media strongly correlate with the number of total phenotypes ($p<0.0001$ for both media by linear regression, not shown). This observation indicates that using both media provides greater insight into phenotypic diversity of isolates than each on its own, and that the selection strategy resulted in a set of isolates that reflects the breadth of phenotypic diversity observed on both media.

Individuals with increased total phenotypes were more likely to be colonized by acid positive isolates ($p=0.007$), strong acid producers ($p=0.008$), double positive isolates ($p=0.0005$) and double strong isolates ($p=0.0003$, all by Mann Whitney test, data not shown). In contrast, the likelihood of observing double negative isolates was not increased in samples with increased total phenotypes. These data indicate that isolates producing acid and H_2O_2 are more likely to be observed when more culture-based phenotypes are observed, suggesting that strains are more likely to produce acid and H_2O_2 within the context of a more diverse microbiota.

Comparison of isolate phenotype between HIV-R and HIV-N – HIV-R individuals were not more likely to be colonized with isolates that were acid positive, strong acid producers, H_2O_2+, double positive or double strong, disproving Hypotheses 5 and 6 (Table 7.3). The observation that 8/32 in the HIV-R group had double strong isolates compared to 3/32 in the HIV/N group may warrant further investigation in a larger study group. No significant differences were observed between HIV-R and HIV-N in terms of number of phenotypes observed, indirectly disproving Hypothesis 11.

Table 7.3 – Comparison of isolate phenotype between HIV-R and HIV-N

	HIV-R (N=32)	**HIV-N** (N=32)	**Difference***	
			All samples	Meets IC[1]
Phenotypes on Rogosa[2]				
Median	2	2	--	--
Range	0-4	0-5		
Phenotypes on mBru[2]				
Median	3	3	--	--
Range	1-5	1-5		
Total phenotypes[3]				
Median	3	3	--	--
Range	1-6	1-8		
Acid-positive isolates				
Any	31 (97%)	26 (81%)	--	--
Any strong	17 (53%)	13 (41%)	--	--
None	1 (3%)	6 (19%)		
H_2O_2+ isolates				
Any	30 (94%)	29 (91%)	--	--
None	2 (6%)	3 (9%)		
Double-positive isolates				
Any	26 (81%)	23 (72%)	--	--
None	6 (19%)	9 (28%)		
Any double-strong	8 (25%)	3 (9%)	--	--
Double-negative isolates				
Any	16 (50%)	15 (47%)	--	--
None	16 (50%)	17 (53%)		
Aspect on Gram-stain				
Any Gram+ rods	29 (91%)	27 (84%)	--	--
Any Gram+ cocci	19 (59%)	20 (63%)	--	--
Any Gram-variable rods	8 (25%)	13 (41%)	--	--

* Indicates significant difference in distribution between HIV-R and HIV-N, by chisquare/Fisher's exact test (ordinal/categorical variables) or Mann Whitney test (continuous variables), [1] Meets inclusion criteria, including active CSW, not pregnant or past menopause, no symptoms, diagnoses or current STI, [2] Number of different phenotypes observed on Rogosa/mBru CFU plates, [3] Total number of unique phenotypes per sample, considering culture and Gram stain data.

Isolate phenotype, vaginal pH and BV diagnosis and vaginal pH – Interestingly, increased phenotypic diversity was observed in BV-intermediate (BVI) samples. Those with BVI samples had a significantly higher median number of initial phenotypes on Rogosa compared to BV+ and BV0 samples (median 3 types in BVI vs. 2 in BV+/BV0, p=0.03 and p=0.0002 respectively by Mann Whitney test).

When considering total phenotypes, those with BVI samples had higher median phenotypes compared to BV– and BV0 samples (median 4 types in BVI vs. 3 in BV–/BV0, p=0.004 and p=0.007 respectively by Mann Whitney test). These findings suggest that BVI samples are likely to have a greater phenotypic density compared to other BV groups and BV– samples, possibly reflecting the presence of a broader complement of bacteria during the transition from BV– to BV+ or vice versa.

Those with any BV diagnosis are more likely to have acid positive isolates compared to BV– (48/51 or 94% vs. 33/42 or 79%, p=0.03 by Fisher's exact test). Similar differences were observed for strong acid producers (p=0.008) and double positive isolates (p=0.003). BV+ individuals were more likely to have double strong isolates when compared to BV– (10/28 or 36% vs. 6/42 or 14%, p=0.05 by Fisher's exact test).

Consistent with these observations, individuals with any acid positive isolates had higher median pH compared to those with none (5.2 vs. 4.2, p=0.02 by Mann Whitney test). Higher median pH was also observed in those with strong acid producers (5.3 vs. 4.7, p=0.03 by Mann Whitney test, not shown).

These findings build on observations described in the previous chapter concerning increased acid production in BV compared to BV– samples, indicating that bacteria in BV samples are more likely to produce acid compared to bacteria in BV– samples. Not

surprisingly, samples producing acid on mBru-mannose at 24h and 72h were also more likely to have acid positive isolates (p=0.04 at 24h/p=0.04 at 72h), strong acid producers (p<0.0001/p<0.0001), double positive isolates (p=0.0004/p=0.01) and double strong isolates (p=0.0005/p=0.0001, all comparisons by chi square test, data not shown).

Individuals with any Gram-variable rods had increased median vaginal pH compared to those without (5.3 vs. 4.6, p=0.01), as did individuals with any Gram+ cocci (5.3 vs. 4.4, p=0.008), while individuals with any Gram+ rods had decreased median vaginal pH (5.0 vs. 5.7, p=0.008, all by Mann Whitney test, data not shown). These findings demonstrate that phenotypic differences of vaginal isolates as determined on test media and by Gram stain are useful criteria to distinguish differences in BV diagnosis.

Other behavioural and biological factors – Individuals with any acid producing isolates had a higher median number of self-reported CSW clients per day (4 vs. 3, p=0.005). This difference was also observed in those with strong acid producers (5 vs. 3, p=0.008), double positive isolates (4 vs. 3, p=0.004) and double strong isolates (5 vs. 3, p=0.008). This observation may indicate that increased acid production by vaginal bacteria may occur in the context of frequent intercourse, possibly due to exposure to semen, condoms, increased douching, or other factors. Although increased clients per day was not associated with BV, the fact that presence of acid-producing isolates is associated with both suggests a link between increased production of acid and disturbance of vaginal homeostasis (ie. BV and frequent sexual intercourse).

Summary – HIV-R individuals were no more likely to be colonized with acid or H_2O_2 producing isolates compared to HIV-N individuals. Therefore, Hypotheses 5 and 6 were not supported by this data. No differences in phenotypic diversity were observed

between HIV-N and HIV-R, contradicting Hypothesis 11.

Primary isolation was carried out on two H_2O_2-detecting media (Rogosa-TMB and mBru-TMB). The strategy for strain selection was to sample the greatest phenotypic diversity possible within a single sample based on colony appearance. Striking differences in colour reaction due to oxidation of TMB in presence of HRP and H_2O_2 observed in panel experiments were also observed during primary isolation, greatly increasing phenotypic criteria for selection of colonies with different appearances. This approach minimizes the number of isolates per sample required to assess the range of phenotypes therein, although quantitative information about the representation of each phenotype within the sample cannot be assessed.

Nairobi isolates (N=576) were spotted onto four different culture media (two chromogenic and two non-chromogenic) and Gram stain analysis of each was also performed. Isolates were characterized as acid positive, strong acid producers, H_2O_2+, double positive or "double strong" (ie. bright yellow and dark blue, resembling probiotic strains analyzed in panel experiments). Phenotypic results in relation to Gram stain data indicate the existence of several subgroups of isolates with distinct phenotypic profiles.

Individual-level descriptors were derived from information about phenotype for each isolate, including the total number of phenotypes observed in each sample and whether or not an individual was colonized with specific types of isolates (acid positive, H_2O_2+, double positive etc.). Individuals with a higher number of distinct phenotypes were also more likely to be colonized with acid positive isolates, indicating that individual strains may be more likely to produce acid when isolated from a more diverse microbiota. Interestingly, individuals with BVI samples had a higher number of distinct phenotypes

compared to other BV groups and BV– individuals, possibly reflecting increased phenotypic diversity during this transitional state.

Similar to findings from the previous chapter, acid producing isolates were more likely to be observed in individuals with any BV diagnosis compared to BV– individuals. Those with acid-producing isolates were also more likely to have higher vaginal pH. Although this difference may not reflect the situation *in vivo*, these observations indicate that bacteria in BV samples are more likely to produce acid on test media.

Presence of acid-producing isolates was also associated with an increased number of CSW clients per day, indicating that acid production by bacteria on test media may be linked to factors associated with disturbance of vaginal homeostasis, including BV and frequent sexual intercourse.

6. Phylogenetic profiling of isolates by RISA and *cpn*60 UT sequencing

Building on results from the previous chapters, molecular profiling was carried out to determine species identity of vaginal isolates in order to assess differences in colonization with *Lactobacillus* phylotypes. Independent molecular targets were assessed: 1) hyper-variable interspacer regions (ISR) between 16S and 23S rDNA were precisely quantified (ribosomal interspacer analysis or RISA) and isolates with similar ISR profiles were assigned to groups, 2) *cpn*60 UT regions were sequenced in a subset of isolates, including representatives of each ISR group, and phylogenetic identity established in reference to the *cpn*60 UT database (cpnDB). RISA and sequencing results were compared and phylogenetic identity of unsequenced isolates was inferred based on ISR profile. The goal of this chapter was to determine if HIV-R individuals are more likely to be colonized with *Lactobacillus* isolates compared to HIV-N individuals.

Molecular profiling of isolates by RISA – Ribosomal interspacer analysis (RISA) results in an electropherogram that reflects the number and size of 16S-23S interspacer regions in each isolate (40). The size of the product(s) is highly variable across bacterial species and strains within species, creating a specific profile for that strain observable by gel electrophoresis (Figure 8.1, A). Observed bands are precisely sized by capillary electrophoresis in relation to an internal size standard (Figure 8.1, B). The amplicon with the greatest peak height within each isolate was defined as the "major" peak, while amplicons with smaller peak heights were defined as "minor" peaks.

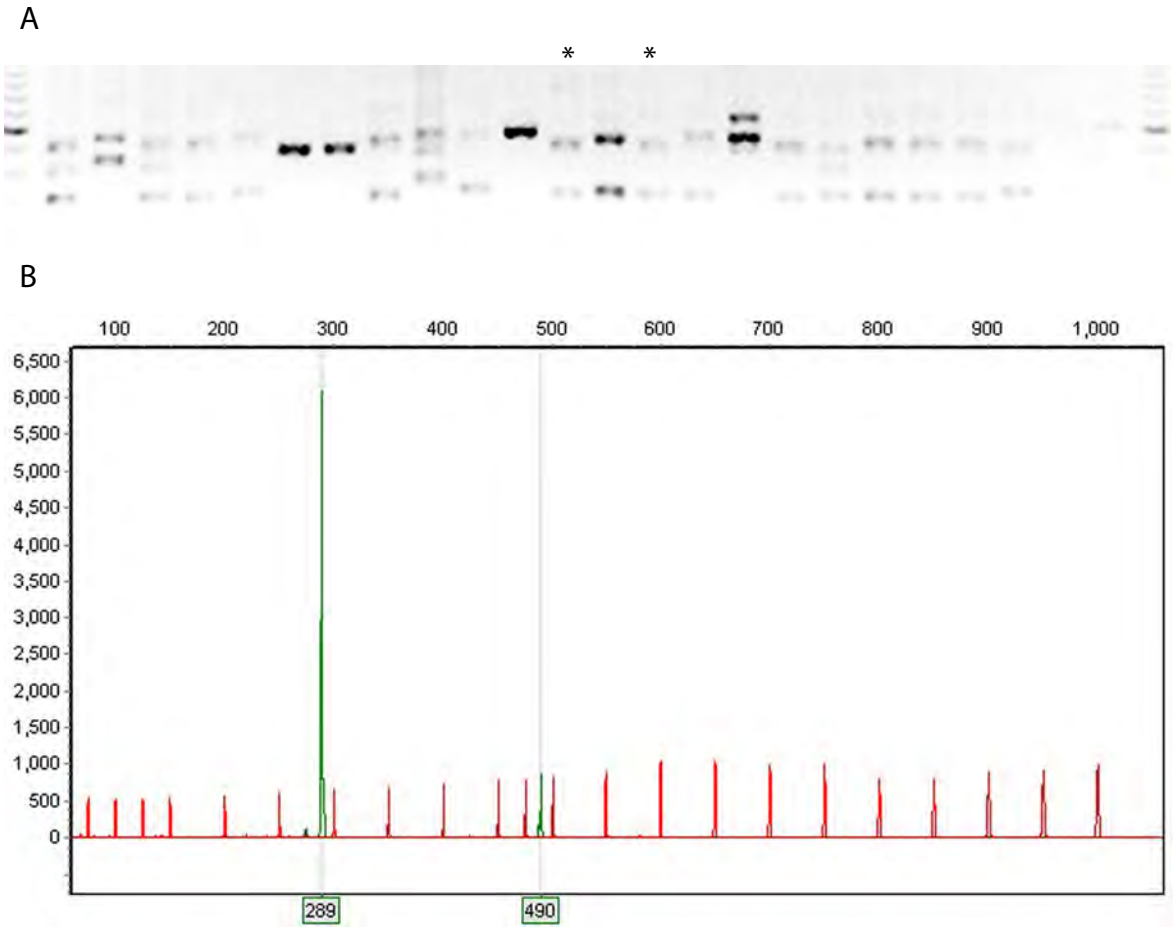

Figure 8.1 – Interspacer region (ISR) amplicons and electropherograms

A. Agarose gel showing ISR amplicons for 24 isolates (two images from a single gel, spliced, inverted and adjusted for brightness and contrast only). Lanes with asterisks indicate isolates with the profile shown in B. The darker band in the ladder on either end of the gel indicates 600bp. B. Electropherogram showing precisely sized amplicons (in green) in comparison to ROX-labelled size standard with 23 fragments from 50-1000bp (in red). Peak indicates the strength of the fluorescent signal during electrophoresis, corresponding to the amount of amplicon of that size in the PCR reaction. The isolate represented by this electropherogram would be interpreted as having a "major" peak at 289bp and a "minor" peak at 490bp.

Initially, ISR profiles were generated for 14 panel strains and 42 isolates (Figure 8.2). Most isolates selected were "double strong" acid/H_2O_2 (see previous chapter). All *Lactobacillus (L.)* panel strains had a similar profile, with a major peak ranging in size from 289-305bp and a minor peak ranging from 488-546 bp. *L. crispatus*, *L. gasseri* and *L. jensenii* had identical profiles, with one or two smaller minor peaks observed for these strains. One Nairobi isolate had a similar profile to this group.

Compared to these strains, the probiotic *L. salivarius* AWH strain had an identical major peak, but a smaller minor peak. Several Nairobi isolates had a profile that was very close to the probiotic *L. salivarius* panel strain, and most of these had a double strong phenotype. Two isolates in this cluster had a single major peak that was identical to both the *L. crispatus* and *L. salivarius* groups and no secondary peak to distinguish them.

Several strains were closest to *L. vaginalis* and the Nairobi isolate *L.* N6, but the only one with a double strong phenotype was from the pilot study. Two isolates were closest to the probiotic *L. delbrueckii* strain, and one of these was double strong. No isolates were close to the probiotic *L. rhamnosus* strains or the type *L. iners* strain. Several other isolates, including three groups with distinct profiles, had major/minor peaks resembling the *Lactobacillus* panel strains.

Streptococcus (S.) bovis had a single peak in its profile, as did one isolate with a similar profile. Two double strong isolates were quite close to the two-peak profile of *Enterococcus (E.). faecalis*. *Bifidobacterium (B.) animalis*, *Gardnerella (G.) vaginalis,* and *Atopobium (A.) vaginae* all had distinct 2-peak profiles, as did a number of isolates that did not resemble the *Lactobacillus* pattern.

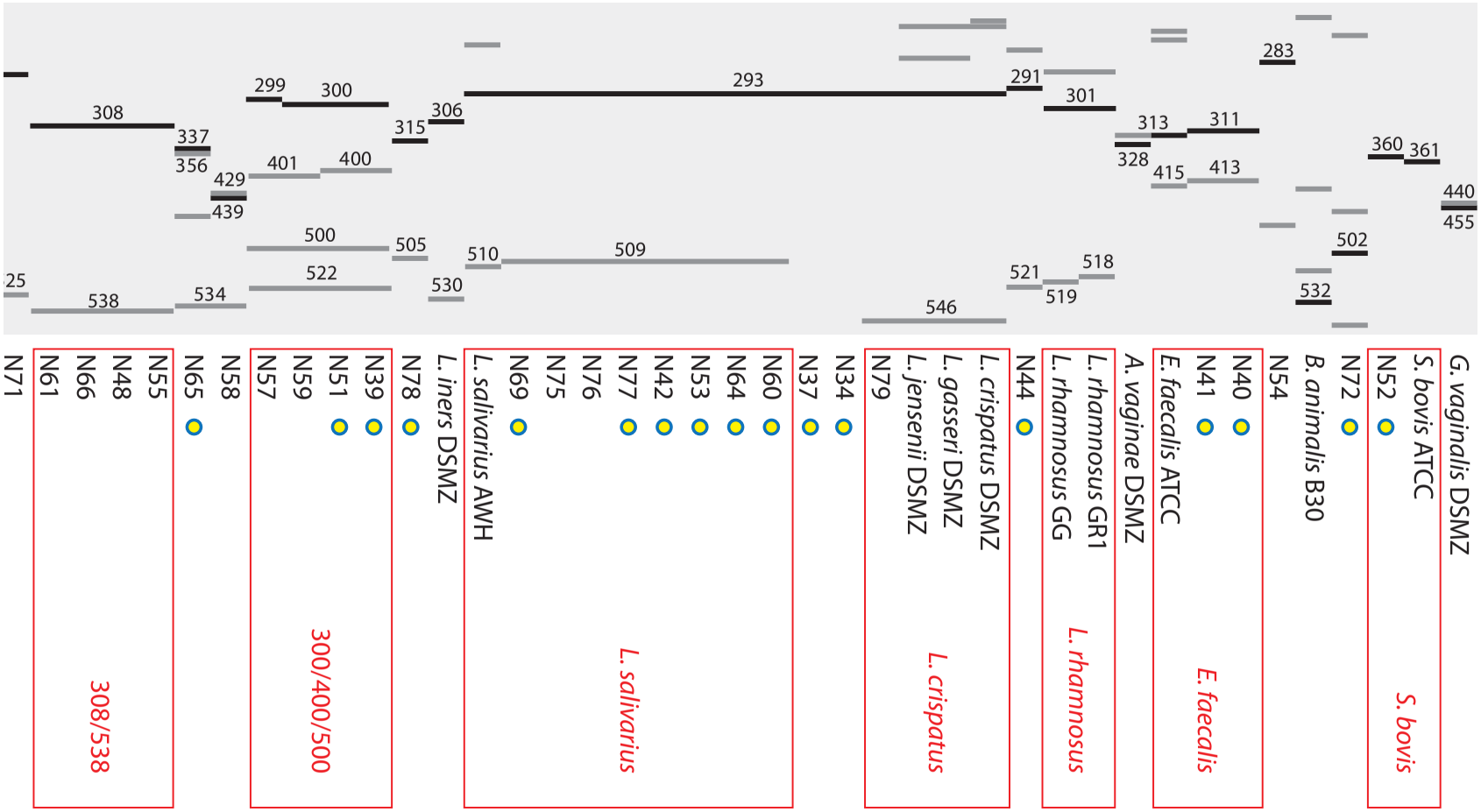

Figure 8.2 – ISR profile for selected panel strains and isolates
ISR profiles for panel strains and isolates, showing precisely sized peak. The size (number of base pairs) for most major and minor amplicons is shown. Red boxes show grouping of panel strains and isolate profiles based on similarities in electropherogram. Blue/yellow circles indicate double strong isolates. Band patternswere visualized using Java TreeView v 1.1.4r3. *L.=Lactobacillus, E.=Enterococcus, A.=Atopobium, S.=Streptococcus, G.=Gardnerella, B.=Bifidobacterium.*

Next, ISR profiles were generated for all remaining pilot study and Nairobi isolates. Interpretable electropherograms were obtained for all but 11/768 (1.4%) of isolates. Most isolates clustered unambiguously in groups defined above, for a total of 20 distinct groups representing 584/757 or 77% of total isolates. These groups ranged in size from 2-108 isolates, while 2 isolates had completely unique profiles. The remaining 173 isolates had ISR profiles that were not observed in the preliminary study. These fell into 18 distinct groups of 2-32 isolates, as well as 10 isolates with unique profiles. Representatives of each group were selected for *cpn*60 UT sequencing in order to confirm identity.

Sequencing of cpn60 UT for a subset of isolates – A total of 149 isolates were selected for full-length *cpn*60 UT sequencing and identification by comparison with cpnDB, including all 42 isolates described in the previous section, as well as representatives from all major groups of isolates defined based on ISR profile.

A phylogenetic tree describing evolutionary relationships of all isolates based on *cpn*60 UT sequence was constructed using the neighbour-joining method (Mega 4.0), revealing a profile consistent with what is already known of culturable vaginal microbiota (Figure 8.3). Since *Lactobacillus*-selective media was used for isolation, it is not surprising that related genera in order Lactobacillales (ie. *Streptococcus*, *Enterococcus, Pediococcus*) are well-represented (97/149 sequenced isolates or 65%). However, strains from a number of genera representing four phyla were isolated in this study, including other Firmicutes (*Staphylococcus*, *Bacillus*, *Gemella*; 15 isolates or 10%), Actinobacteria (*Gardnerella*, *Bifidobacterium*; 28 isolates or 19%), Bacteroidetes (*Bacteroides, Prevotella*; 5 isolates or 3%), and Proteobacteria (*Escherichia, Klebsiella*; 4 isolates or 3%).

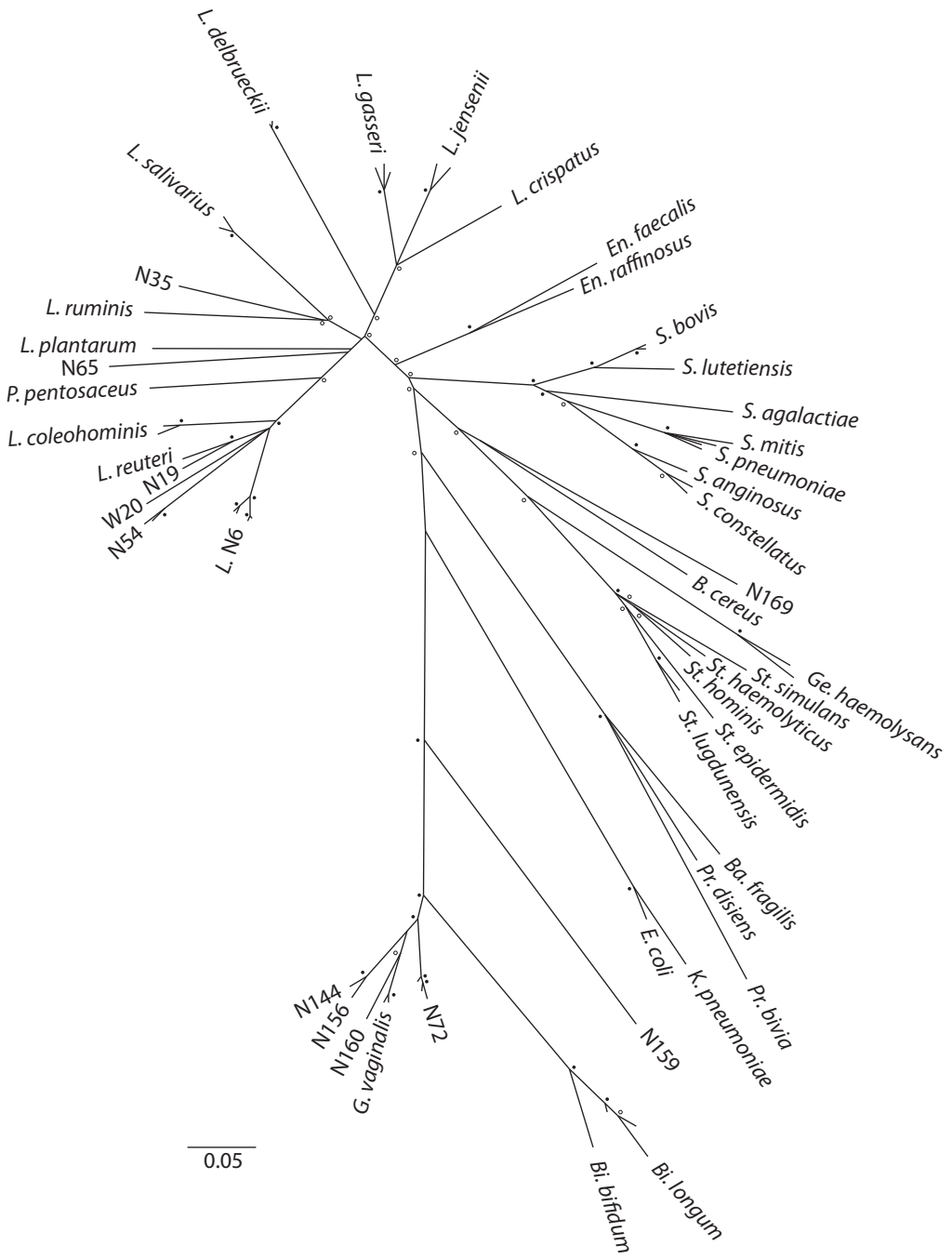

Figure 8.3 – Phylogenetic relatedness of 149 cultured isolates

Bootstrap consensus tree of 149 isolates based on *cpn*60 UT. Branch length indicates estimated phylogenetic distance by the Maximum Composite Likelihood Method (sum = 7.35, bar = 0.05). Bootstrap values greater than 95% (robust) and 50% (valid) are indicated with filled and open circles respectively. N=Nairobi isolate, *L=Lactobacillus, P.=Pediococcus, En.=Enterococcus, S.=Streptococcus, B.= Bacillus, Ge.=Gemella, St.=Staphylococcus, Ba.=Bacteroides, Pr.=Prevotella, E.=Escherichia, K.=Klebsiella, Bi.=Bifidobacterium, G.=Gardnerella.*

Several previously undescribed groups of vaginal bacteria were observed, including isolates similar to the N6 strain isolated in a previous study (S. Iqbal). This group has several branches with robust nodes, and a *cpn*60 UT sequence about 90% similar to both *L. vaginalis* and *L. reuteri* reference strains. Several other *L. reuteri*-related isolates distinct from *L.* N6 were also observed, including one from a pilot study sample. These observations indicate that novel *Lactobacillus* phylotypes, related but not identical to *L. vaginalis* and *L. reuteri*, were identified in both African and Canadian individuals.

A group of isolates phylogenetically linked to *L. salivarius* and *L. ruminis*, but with no near neighbour in cpnDB, was also identified (N35). Three single isolates of uncertain phylogeny (N65, distantly related to *Weissella confusa*; N169, linked to *B. cereus*; and N159, linked to phylum Actinobacteria) were also observed.

A surprising amount of sequence diversity was observed for *G. vaginalis* isolates. Although some are practically identical to the reference strain, the majority are less than 90% identical. Deep branching with several robust nodes was observed, indicating extensive phylogenetic diversification of this taxon (Figure 8.3).

Comparing cpn60 UT sequence and ISR profile – The correspondance between *cpn*60 UT sequence and ISR profile was excellent. In general, isolates with the same or similar *cpn*60 UT sequence also have similar ISR profiles and vice versa (Figure 8.4), indicating that both are sensitive and complementary for unambiguous species identification of bacterial strains.

In contrast to ISR analysis of DSMZ strains, *cpn*60 UT sequencing revealed that *L. crispatus*, *L. gasseri* and *L. jensenii* all have distinct ISR profiles, with major bands at 293, 308 and 288bp respectively (Figure 8.4).

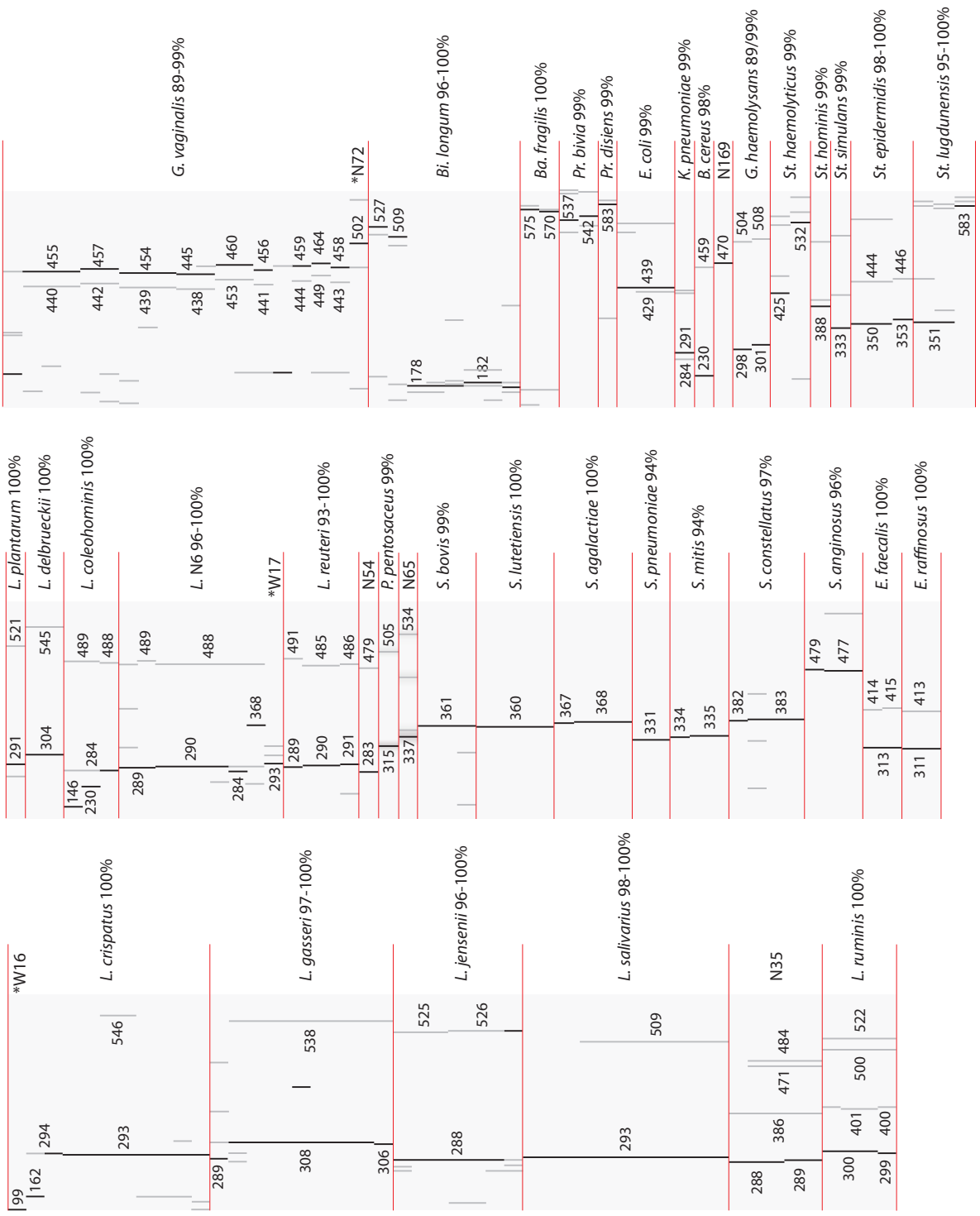

Figure 8.4 – Comparing UT sequences and ISR profiles of cultured isolates

Species identification of isolates by *cpn*60 UT sequencing is shown alongside ISR profile for each isolate (grey bars), demonstrating a close correspondence between the two targets. Black bands indicate major peaks and grey bands indicate minor peaks. Abbreviations of genera as in previous figure.

If this finding is confirmed, DSMZ strains labelled as *L. gasseri* and *L. jensenii* may actually be *L. crispatus* strains.

All *L. salivarius* isolates also had a major peak at 293bp but most could be distinguished from *L. crispatus* by the presence of a minor peak at 509bp (the type strain of *L. crispatus* and some sequenced isolates have a minor peak at 546bp).These data indicate that ISR profile may not be sufficient to distinguish between *L. crispatus* and *L. salivarius*. In this study, all isolates with a major peak at 293bp and no *L. crispatus*-like minor peak at 546bp were sequenced and determined to be *L. crispatus*.

Isolates in the *L.* N6/*L. reuteri* groups had peak profiles resembling the *L. vaginalis* reference strain, with a major peak around 290 and a minor peak around 488. Interestingly, all *L. coleohominis* isolates had a major peak at 284, while the N54 group had a major peak at 283 but a minor peak that was 10bp smaller than *L. coleohominis*. These findings indicate a clear correspondance between ISR profile and *cpn*60 UT sequence for *Lactobacillus* and related isolates. Further, variability in *cpn*60 UT sequence appears to be reflected in variability of ISR profile for closely related isolates.

In contrast to *Lactobacillus*, most isolates identified as *Streptococcus* had no minor peaks. Major peaks only a few bp apart were consistently associated with unique *cpn*60 UT sequences (eg. *S. bovis, S. lutetiensis, S. agalactiae*). *E. faecalis* and *E. raffinosus* had two-peak profiles that were distinct but very close to each other.

In parallel to observed *cpn*60 UT sequence diversity, *G. vaginalis* isolates had related but divergent ISR profiles. The *G. vaginalis* reference strain had major/minor peaks at 455/440bp, but other isolates had dual peaks that "wobbled" higher and lower, while always maintaining an interval of 15bp. Another example of this was observed in two

isolates from a single individual, one of which was 99% identical to *Gemella haemolysans* and had major/minor peaks at 298/504bp, and the other only 89% identical to *G. haemolysans* that had peaks 3bp larger at 301/507bp.

While wobbling of peaks may be an artefact of amplicon electrophoresis, this seems unlikely considering wobbling was not observed in all groups of isolates. This observation confirms that variability in *cpn*60 UT sequence is reflected in differences in ISR profile, although an understanding of the possible biological significance of these differences will require further investigation.

There were only three cases of an obvious discrepancy between *cpn*60 UT sequence and ISR profile (marked with asterisks in Figure 8.4). A single isolate closest to *G. vaginalis* by sequencing of the *cpn*60 UT (N72) had an ISR profile and Gram stain appearance (Gram+ bifid rod) inconsistent with the reference strain. This isolate had a major peak at 502bp (in the same range of some *Bifidobacterium* isolates), but also had a *G. vaginalis*-like peak at 458 bp. The most likely explanation is this "isolate" actually contains more than one strain: a *G. vaginalis* strain for the initial *cpn*60 UT sequence and a *Bifidobacterium*-like strain for the phenotypic profile, with peaks from both strains detected by RISA. In support of this, a closer examination of the Gram-stained slide by microscopy revealed that both Gram+ and Gram-variable cells were visible (not shown).

It is interesting to note that N72 had a double strong phenotype. Another double strong isolate, N65, also had a *Gardnerella*-like peak at 438bp, indicating the possibility that this isolate also contained both N65 and *G. vaginalis* strains. Given the observed increase in acid production by whole samples and isolates from individuals with BV (described in previous chapters), it is tempting to speculate that presence of *G. vaginalis*

increases strong acid and H_2O_2 production in co-isolates. Further investigation involving co-culture of isolates will be required to clarify this possible association.

The likelihood that an undetermined number of isolates are in fact mixtures of strains was addressed in the previous chapter. Many minor peaks were seen in ISR profiles of one *cpn*60 UT sequence group that were consistent with other *cpn*60 UT sequence groups. For example, a *Prevotella disiens* isolate resembled Gram+ cocci, and had peaks resembling other *Prevotella* isolates, as well as a peak characteristic of *S. epidermidis.*

In other cases where isolates had Gram stain characteristics inconsistent with molecular profile, the *cpn*60 UT sequence and ISR profile were always consistent with each other. Since DNA for molecular analysis was harvested from isolation plates prior to phenotyping, and Gram stained slides were prepared after incubation of phenotyping plates 48h later, it is likely that strains present at lower concentration at the time of DNA isolation may have overgrown the isolate during subsequent incubation. Therefore, phenotypic profile cannot be tied with certainty to phylogenetic profile.

In summary, these observations demonstrate that ISR profile and *cpn*60 UT sequence determine the same organism, even if it is not the only organism in the "isolate". Further, phenotypic profiling determines *ex vivo* functional characteristics of an organism or mixture of organisms in each "isolate", even if the relative influence of co-isolates cannot be determined with available data.

Summary of isolates – The identity of all unsequenced isolates was inferred based on the correspondence between ISR profile and *cpn*60 UT sequence, as described above. A total of 11 isolates (1.4%) did not generate an interpretable ISR profile, and 22 isolates (2.9%) belonged to ISR groups with no sequenced representatives(Table 8.1).

Table 8.1 – Summary of isolates

	Isolates		Study		Phenotype[2]		
***cpn*60 UT ID[1]**	Total	Inferred	Pilot	Nairobi	A+	H+	D+
L. crispatus	108	97	61	47	38%	63%	35%
L. N6	108	99	12	96	15%	66%	11%
L. gasseri	59	49	32	27	98%	46%	76%
L. jensenii	59	52	27	32	61%	51%	54%
L. coleohominis	30	27	--	30	23%	97%	20%
L. salivarius	26	15	--	26	100%	96%	96%
N27	24	23	7	17	100%	63%	63%
N35	24	19	--	24	100%	71%	71%
L. reuteri[3]	14	10	5	9	78%	93%	79%
L. ruminis	6	2	--	6	100%	100%	100%
P. pentosaceus	3	2	--	3	100%	100%	100%
L. delbruecki	2	--	--	3	100%	100%	100%
N65	1	--	--	1	+	+	+
L. plantarum	1	--	--	1	+	+	+
S. constellatus	32	28	7	25	97%	81%	81%
S. agalactiae	31	27	--	31	100%	94%	94%
S. pneumoniae[4]	27	22	1	26	93%	93%	89%
S. lutetiensis	20	16	6	14	100%	45%	45%
S. bovis	17	14	--	17	100%	76%	76%
En. faecalis	8	6	1	7	100%	63%	63%
S. anginosus	5	2	--	5	100%	80%	80%
En. raffinosus	2	--	--	2	100%	100%	100%
Staphylococcus sp.[5]	21	9	4	16	90%	10%	10%
Other Bacillales[6]	4	--	--	4	100%	100%	100%
G. vaginalis	68	49	5	63	12%	16%	7%
Bifidobacterium sp.[7]	23	15	6	17	100%	22%	22%
N159	1	--	--	1	+	–	–
Proteobacteria[8]	6	2	--	6	83%	33%	17%
Bacteroidetes[9]	5	--	--	5	100%	0%	0%
Unidentified/Failed	33	--	18	15	88%	27%	21%
Total	768	619	192	576	64%	64%	46%

[1] Species identity based on isolate *cpn*60 UT sequence in reference to cpnDB, [2] Phenotype: % of isolates that are acid-positive (A+), H_2O_2+ (H+) or double positive (D+), [3] Including N19 and W20, [4] Including *S. mitis*, [5] Including *St. epidermidis, St. simulans, St. haemolyticus, St. hominis, St. piscifermentans, St. lugdunensis*, [6] Including *B. cereus, G. haemolysans*, N169, [7] Including *Bi. longum, Bi. catenulatum*, [8] Including *E. coli*, *K. pneumoniae*, [9] Including *Ba. fragilis, Pr. bivia, Pr. disiens*.

When failed or unidentified sequences were excluded, remaining isolates (96%) were classified into categories based on similarity of ISR profile to isolates identified by *cpn*60 UT sequencing. Therefore, the vast majority of isolates were reliably identified based on *cpn*60 UT sequencing of 149/768 or 19% of isolates, representing a considerable savings in time and money spent on sequencing.

A total of 16 phylotypes were observed among 192 pilot study isolates compared to 47 phylotypes in 576 Nairobi isolates. N6 and *L. crispatus* were the most frequently observed phylotypes in Nairobi. In contrast, *L. crispatus, L. gasseri* and *L. jensenii* were most frequently isolates in the pilot study. These differences are likely to be explained by actual differences in *Lactobacillus* colonization between the two groups, but may also be affected by the use of two different media for culture of Nairobi isolates, revealing greater phylogenetic diversity in these samples.

Despite limitations in definitively linking phylotype to phenotype, it is interesting to note a correspondence between the two when considering all isolates. Of *Lactobacillus* phylotypes, *L. crispatus,* N6 and *L. coleohominis* isolates were least likely to be acid positive, while *L. gasseri, L. salivarius*, N27 and N35 isolates were virtually all acid positive. *L. gasseri* and *L. jensenii* were least likely to be H_2O_2+ while *L. salivarius* and *L. coleohominis* isolates were virtually all H_2O_2+. Most *L. salivarius* and several smaller groups of isolates were double positive. Almost all *Streptococcus* phylotypes were acid positive. *G. vaginalis* isolates were least likely to produce acid and H_2O_2. Most of these isolates grew only on mBru media and also had the appearance of Gram-variable coccobacilli. The majority of isolates in *Bifidobacterium* sp., and phyla Proteobacteria and Bacteroidetes were acid positive only.

Profile of isolate phylotype in individuals – Based on these data, a set of phylotype-related variables were derived to describe individual study participants, including colonization with a number of specific phylotypes (Table 8.2). The potential role of phylotype density (ie. total number of phylotypes observed per individual) was assessed by calculating the total number of phylotypes observed in each individual, as well as the number of *Lactobacillus*, *Streptococcus* and other phylotypes per individual.

About 80% of individuals in the Nairobi study were colonized with at least one *Lactobacillus* phylotype, 45% with at least one *Streptococcus* phylotype, and 62% with other phylotypes. Individuals were colonized most frequently with N6 (42%), G. vaginalis (41%), and *L. crispatus* (22%). In contrast, 8/16 (50%) individuals in the pilot study were colonized with *L. crispatus*, 5/16 (31%) each with *L. jensenii* and *L. gasseri* and only 3/16 (19%) each with N6 or *G. vaginalis*.

These findings indicate a remarkable similarity in major culturable phylotypes observed in two socially and geographically distant groups of women, as well as a distinct distribution of these phylotypes between the two groups.

Some individuals were colonized with a high number of culturable phylotypes (range 1-8), and up to 6 *Lactobacillus* phylotypes. Most *Lactobacillus* phylotypes were isolated from individuals who were colonized with two or more *Lactobacillus* phylotypes, indicating increased likelihood of isolating these phylotypes in the context of a more complex microbiota. In contrast, *L. crispatus* was the only *Lactobacillus* phylotype not associated with an increased density of *Lactobacillus* phylotypes, indicating that it is associated with a less complex microbiota.

Table 8.2 – Isolate phylotypes observed in individuals

	Culture (N=96)	**Pyro** (N=48)	**qPCR** (N=40)	**Clone** (N=10)
Total phylotypes				
Median	3	3	3	3
Range	1-8	1-8	1-7	2-5
Lactobacillus colonization				
L. crispatus	21 (22%)	8 (17%)	8 (20%)	3 (30%)
L. gasseri	15 (15%)	10 (21%)	7 (18%)	2 (20%)
L. jensenii	10 (10%)	6 (13%)	6 (15%)	1 (10%)
L. salivarius	14 (14%)	11 (23%)	6 (15%)	2 (20%)
N6	40 (42%)	23 (48%)	20 (50%)	5 (50%)
L. coleohominis	14 (14%)	8 (17%)	6 (15%)	2 (20%)
N27	8 (8%)	6 (13%)	3 (8%)	1 (10%)
N35	12 (12%)	9 (19%)	6 (15%)	3 (30%)
Any	78 (81%)	42 (88%)	33 (83%)	9 (90%)
None	18 (19%)	6 (13%)	7 (18%)	1 (10%)
Lactobacillus phylotypes				
Median	1	2	2	2
Range	0-6	0-5	0-3	0-3
Streptococcus colonization				
S. agalactiae	16 (17%)	6 (13%)	4 (10%)	2 (20%)
S. constellatus	17 (18%)	9 (19%)	7 (18%)	2 (20%)
S. pneumoniae	17 (18%)	6 (13%)	2 (5%)	--
Any	51 (53%)	24 (50%)	18 (45%)	4 (40%)
None	45 (47%)	24 (50%)	22 (55%)	6 (60%)
Streptococcus phylotypes				
Median	1	1	0	0
Range	0-3	0-3	0-2	0-2
Other colonization[1]				
G. vaginalis	39 (41%)	17 (35%)	17 (43%)	2 (20%)
Any	60 (62%)	20 (42%)	16 (40%)	5 (50%)
None	36 (38%)	28 (58%)	24 (60%)	5 (50%)
Other phylotypes[1]				
Median	1	1	1	1
Range	0-3	0-3	0-3	0-1

[1] Including *G. vaginalis, Bifidobacterium* sp., *Staphylococcus* sp., Bacillales, Proteobacteria or Bacteroidetes.

Several associations were observed between specific phylotypes in this study. For example, individuals colonized with *L. gasseri* were more likely to also be colonized with *L. salivarius* (6/15 or 40% vs. 8/81 or 10%, p=0.008) and *L. coleohominis* (5/15 or 33% vs. 9/81 or 11%, p=0.04). Those colonized with N6 were more likely to also be colonized with *L. jensenii* (9/14 or 64% vs. 31/86 or 36%, p=0.001) and less likely to be colonized by *L. salivarius* (1/14 or 7% vs. 39/82 or 48%, p=0.007, all by Mann Whitney test). These observations suggest possible metabolic interactions between *Lactobacillus* phylotypes.

Comparison of phylotype between HIV-R and HIV-N – No significant differences were observed between HIV-R and HIV-N individuals in terms of colonization with any *Lactobacillus* or specific *Lactobacillus* phylotypes, disproving Hypothesis 7, or in terms of *G. vaginalis* colonization, disproving Hypothesis 8 (Table 8.3). No differences in species density were observed between the two groups, disproving Hypothesis 11.

Phylotypes, BV and vaginal pH – Consistent with the concept that those with BV have an increased species density compared to those without BV, median phylotype density was higher in BVI and BV+ individuals compared to BV– individuals (4 vs. 3 for both, p=0.01 and p=0.05 respectively by Mann Whitney test). Interestingly, BVI individuals had a higher range of phylotype density (3 to 8 phylotypes) compared to either BV– or BV+ individuals (1 to 7 for both). This finding is consistent with the concept of increased species density as part of this transitional state.

Consistent with this observation, median density of *Lactobacillus* phylotypes was decreased in BV+ and BV0 individuals compared to BV– (1 vs. 2 for both, p=0.05 and p=0.03 respectively), indicating that more *Lactobacillus* phylotypes are likely to be observed in BV– individuals.

Table 8.3 – Comparison of isolate phylotype between HIV-R and HIV-N

	HIV-R (N=32)	**HIV-N** (N=32)	**Difference*** All samples	Meets IC[1]
Total phylotypes				
Median	3	3	--	--
Range	1-7	1-6		
Lactobacillus colonization				
L. crispatus	11 (34%)	6 (19%)	--	--
L. gasseri	5 (16%)	5 (16%)		
L. jensenii	5 (16%)	2 (6%)		
L. salivarius	6 (19%)	4 (13%)		
L. N6	15 (47%)	11 (34%)		
L. coleohominis	3 (9%)	8 (25%)		
N27	3 (9%)	3 (9%)		
N35	4 (13%)	4 (13%)		
Any	28 (88%)	24 (75%)		
None	4 (13%)	8 (25%)		
Lactobacillus phylotypes				
Median	2	1	--	--
Range	0-6	0-4		
Streptococcus colonization				
S. agalactiae	5 (97%)	8 (81%)	--	--
S. constellatus	6 (53%)	4 (41%)		
S. pneumoniae	4 (%)	3 (%)		
Any	15 (97%)	10 (81%)		
None	1 (3%)	6 (19%)		
Streptococcus phylotypes				
Median	1	0	--	--
Range	0-3	0-3		
Other colonization[2]				
G. vaginalis	11 (81%)	14 (72%)	--	--
Any	20 (19%)	18 (28%)		
None	12 (25%)	14 (9%)	--	--
Other phylotypes[2]				
Median	1	1	--	--
Range	0-1	0-1		

* Indicates significant difference in distribution between HIV-R and HIV-N, by chisquare/Fisher's exact test (ordinal/categorical variables) or Mann Whitney test (continuous variables), [1] Meets inclusion criteria, including active CSW, not pregnant or past menopause, no symptoms, diagnoses or current STI, [2] Including *G. vaginalis, Bifidobacterium* sp., *Staphylococcus* sp., Bacillales, Proteobacteria or Bacteroidetes.

In contrast, median density of *Streptococcus* phylotypes was significantly increased in BVI, BV+, and BV0 samples compared to BV– samples (1 vs. 0, p=0.02, p=0.006 and p=0.004 respectively by Mann Whitney test), indicating that colonization by one or more *Streptococcus* phylotypes is signifcantly associated with any BV diagnosis.

Not surprisingly, BV+ individuals were less likely to be colonized with any *Lactobacillus* phylotype compared to BV– (17/28 or 61% vs. 37/42 or 88%, p=0.01). This was also true for *L. crispatus* (4% vs. 25%, p=0.01), *L. jensenii* (0% vs. 21%, p=0.0009) and *L.* N6 (18% vs. 62%, p=0.0005). In contrast, BV+ individuals were more likely than BV– to be colonized with the *L. salivarius*-like strain N35 (32% vs. 2%, p=0.0008), *S. constellatus* (36% vs. 10%, p=0.01), or *G. vaginalis* (57% vs. 31%, p=0.05, all comparisons by Fisher's exact test).

The observation that a *Lactobacillus* phylotype is associated with BV is surprising, considering that absence of *Lactobacillus* defines BV. Most isolates for both N35 and *S. constellatus* were acid producers, compared to a lower proportion of *L. crispatus*, *L. jensenii* and *L*. N6 isolates. This is consistent with the observation of increased acid production by BV samples and isolates described in previous chapters.

Interestingly, 7/19 or 37% of individuals in the BV0 category were colonized with *L. crispatus*, significantly more than BV+ individuals (p=0.005 by chi square test) and greatly resembling BV– individuals in this respect. This observation may indicate that BV0 is not merely an artefact of Gram stain analysis, but rather a distinct biological entity with implications for vaginal health and vulnerability to STD/HIV infection. Increased *L. crispatus* in BV0 samples may indicate that BV0 is an intermediate stage between BV+ and BV– states. Further work will be required to confirm this hypothesis.

Increased density of *Lactobacillus* phylotypes was correlated with decreased vaginal pH (R^2 = 0.09, p=0.004 by linear regression), while increased density of *Streptococcus* or other phylotypes was associated with increased vaginal pH (R^2 = 0.11 for both, p=0.0009 and p=0.001 respectively by linear regression). *L. jensenii* and *L.* N6 were significantly associated with decreased vaginal pH (p=0.03 and p=0.0007 by linear regession, not shown), while *S. agalactiae* and *S. constellatus* were associated with increased vaginal pH (p=0.05 and p=0.004 by linear regression, not shown).

Other behavioural and biological variables – Interestingly, women with increased CSW clients per day also had increased overall phylotype density and increased *Lactobacillus* phylotype density (R^2 = 0.12, p=0.007 and R^2 = 0.13, p=0.004 respectively by linear regression), indicating that frequent sexual intercourse may result in colonization by an increased number of phylotypes. Having more clients per day and having an increased number of phenotypes was also associated with acid production by vaginal specimens and isolates (see previous chapters). Therefore, these observations build on previous findings linking increased complexity of microbiota to acid production, BV and frequent sexual intercourse.

Having *G. vaginalis* isolates was associated with increased acid production by whole samples at 24h and 72h (p=0.003 and p=0.0007 respectively by chisquare analysis, not shown). This is consistent with the hypothesis that vaginal bacteria growing on test media are more likely to produce acid in response to bacteria, such as *G. vaginalis*, that are present in BVI and BV+ samples. Those with any STI were also significantly more likely to have *G. vaginalis* isolates (8/39 or 21% vs. 2/56 or 4%, p=0.01 by Fisher's exact test), indicating that vulnerability to exogenous pathogens is enhanced in this context.

Summary – No significant differences were observed between HIV-N and HIV-R groups in terms of colonization with *Lactobacillus* overall, with any specific *Lactobacillus* phylotypes or with *G. vaginalis* (Hypotheses 7 and 8). No differences in phylotype density between the two groups were observed (Hypothesis 11).

A critical lesson from this study is the close correspondance between phylogenetic data from two independent molecular targets – the ribosomal 16S/23S interspacer and the *cpn*60 UT. These results confirm that capillary electrophoresis of amplicon from one target is a rapid, effective approach to characterizing the phylogeny based on sequencing of the other target. Further, variability in UT sequence appears to correspond to variability observed in ISR profile. Grouping of isolates based on ISR profile and UT sequencing of representatives of each group resulted in unambiguous identification to the species level and considerable savings in sequencing effort and cost.

All but two of 16 phylotypes in pilot study isolates were also observed among 47 phylotypes in Nairobi isolates, indicating increased phylotype density in Nairobi individuals. Individuals in the Nairobi study were most often colonized with the *L. reuteri*-like N6, *L. crispatus* and *G. vaginalis*, while pilot study individuals were most often colonized with *L. crispatus*, *L. jensenii* and *L. gasseri*. The overlap in phylotype profile is remarkable between these two very different study groups.

Individuals with BV had a higher phylotype density overall but lower *Lactobacillus* phylotype density compared to BV– individuals and were less likely to be colonized with several *Lactobacillus* phylotypes (*L. crispatus, L. jensenii*, N6). Surprisingly, BV+ individuals were more likely to be colonized with the *L. salivarius*-like N35, a phylotype with strong acid producing isolates. This observation is consistent with

increased acid production in BV samples described in previous chapters. Intriguingly, BV0 individuals resemble BV– individuals in terms of *L. crispatus* colonization. This observation strengthens the hypothesis that BV0 is a distinct biological entity, possibly as a transition phase from BV+ to BV– states, analogous to BVI as a transition from BV– to BV+ states.

The presence of a *Gardnerella*-like band in two isolates with distinct UT identities corresponded with the double strong phenotype, indicating that presence of *Gardnerella* may induce acid and H_2O_2 production in co-isolates. This was corroborated by the observation that having *G. vaginalis* isolates was strongly associated with increased acid production by whole samples at both timepoints.

Women with increased CSW clients per day also had increased phylotype density, building on previous observations linking BV, acid production and frequent sexual intercourse. Exposure to different sexual partners may also contribute to increased phylotype diversity in CSW.

7. Ultra-deep molecular characterization of vaginal microbiota

The goal of this chapter is to determine any differences in relative abundance of *Lactobacillus* and BV phylotypes between HIV-R and HIV-N individuals by in-depth culture-independent analysis of microbial community composition. Since clone libraries are considered the gold standard for culture-independent analysis of microbial community composition, this study began with the goal of completing extensive clone libraries for 20 individuals in each of the HIV-R and HIV-N groups. Briefly, *cpn*60 UT was amplified from vaginal samples, ligated into a cloning vector and competent *E. coli* transformed with the resulting plasmid. Transformants were plated onto selective media and colonies containing unique *cpn*60 amplicons were picked and sequenced.

Novel pyrosequencing technology became available while clone libraries were underway, offering the promise of a much deeper analysis of microbial communities than was possible using clones. Therefore, the focus of the study shifted to generation of amplicon for pyrosequencing studies. Clone libraries were curtailed to 10 individuals (4 HIV-N, 6 HIV-R), limiting their usefulness to determine differences between the two groups. However, information from the clone library study became instrumental in assessment of the feasibility of pyrosequencing (176) and greatly enhanced the database used to determine species identity of vaginal bacteria by pyrosequencing.

Construction of clone libraries – A total of 10,614 clones from 10 individuals were sequenced, resulting in 7,180 complete *cpn*60 UT sequences in 196 groups represented by a single sequence less than 99% similar to every other sequence. Sequences

were compared to cpnDB to determine species identity represented by each clone. Nineteen clones (8 contig groups) were excluded from further analysis since they were closest to human *cpn*60 UT sequences (likely derived from human cells present in vaginal samples). A phylogenetic tree of clone sequences was created using Mega software, showing that composition of vaginal microbiota in this study group is consistent with what is known of the vaginal microbiota using culture-independent techniques (Figure 9.1).

Sequences identical to *Lactobacillus iners*, which was only discovered in 1999 and is known to be a major "non-culturable" element of the vaginal microbiome (57), are almost half of total clones and not represented among isolates (Table 9.1). *L. iners* was observed in 7/10 individuals at proportions ranging from 1-100% of total clones.

Clones that were identical to isolate phylotypes were also observed, including *L. crispatus*, N6, and *L. salivarius*. These phylotypes differed in terms of proportion of total isolates vs. clone libraries. For example, N6 made up 17% of isolates but only 0.3% of clone libraries, while *L. crispatus* made up 18% of isolates and 14% of clone libraries. These differences may indicate unique biases associated with clone libraries compared to deep sequencing.

Only a single BV–, HIV-R individual had *L. crispatus* clones, at a low level (2%), while two BV–, HIV-N individuals had high levels, including one with 99% *L. crispatus* clones. One BV+ individual had 11% *L. salivarius* clones, while four BV– individuals had *L.* N6 clones at low levels (~1%).

Interestingly, these were the only *Lactobacillus* phylotypes identified through clone libraries, indicating that several phylotypes in this genus were detectable by selective culture but not in a clone library of this size.

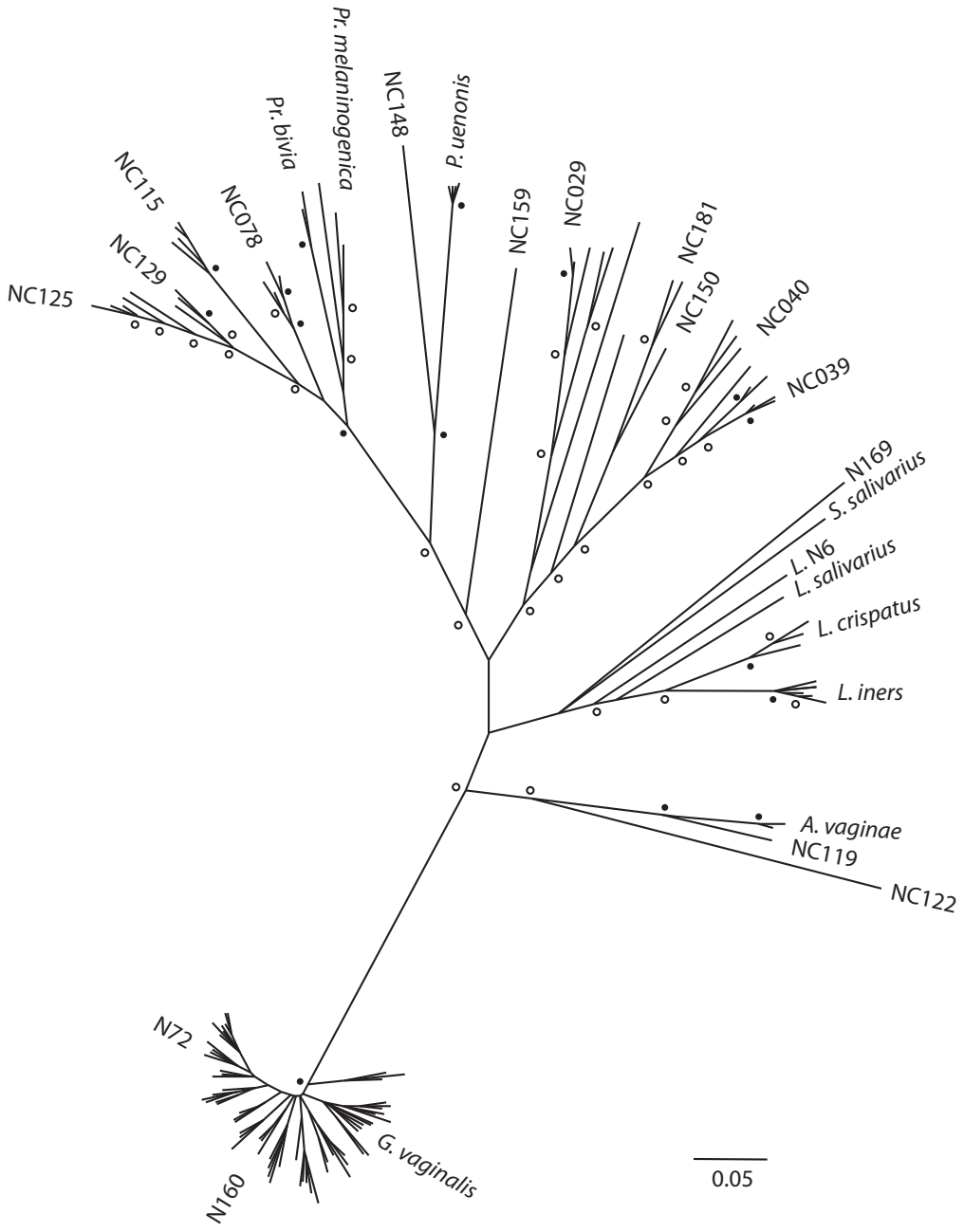

Figure 9.1 – Phylogenetic relatedness of sequences in vaginal clone libraries

Bootstrap consensus tree summarizing sequence information for 7,161 clones in 188 phylotypes with less than 99% similarity. Branch length indicates estimated phylogenetic distance by the Maximum Composite Likelihood method (sum = 8.36, bar = 0.05). Bootstrap values greater than 95% (robust) and 50% (valid) are indicated with filled and open circles respectively, based on 1000 replicates (Mega 4.0). Where clone sequence was <99% similar to isolate sequence, the isolate sequence label is shown. N=Nairobi isolate, NC=Nairobi clone, *G.=Gardnerella, A.=Atopobium, L.=Lactobacillus, S.=Streptococcus, P.=Porphyromonas, Pr.=Prevotella*.

Table 9.1 – Summary of clone libraries by individual

*cpn*60 UT ID[1]	Total	HIV-R						HIV-N			
		BV–				BV+		BV–		BV+	
		1327	**1529**	**1967**	**2017**	**2030**	**2231**	**2316**	**2347**	**2394**	**2445**
Total clones[2]	7161	245	996	551	687	618	922	920	452	893	877
L. iners	3178	97%	100%	>99%	>99%	--	48%	28%	--	1%	--
L. crispatus	1103	2%	--	--	--	--	--	71%	99%	--	--
L. salivarius	65	--	--	--	--	11%	--	--	--	--	--
L. sp. L6	20	1%	--	--	<1%	--	--	1%	1%	--	--
S. salivarius	19	--	--	--	--	2%	--	--	--	--	--
A. vaginae	123	--	--	--	--	--	<1%	--	--	13%	<1%
N160	534	--	--	--	--	83%	--	--	--	3%	--
N72	336	--	--	--	--	2%	27%	--	--	8%	--
G. vaginalis	27	--	--	--	--	--	--	--	--	3%	--
Other Actinobacteria	66	--	--	--	--	2%	1%	--	--	3%	2%
NC078	408	--	--	--	--	--	--	--	--	22%	24%
NC115	138	--	--	--	--	--	--	--	--	--	16%
Pr. melaninogenica	77	--	--	--	--	--	8%	--	--	--	--
P. uenonis	57	--	--	--	--	--	2%	--	--	--	5%
NC125	53	--	--	--	--	--	--	--	--	--	6%
Other Bacteroidetes	40	--	--	--	--	--	1%	--	--	--	5%
NC029	445	--	--	--	--	--	2%	--	--	29%	19%
NC181	200	--	--	--	--	--	6%	--	--	17%	--
NC039	127	--	--	--	--	--	2%	--	--	<1%	13%
NC150	70	--	--	--	--	--	--	--	--	--	8%
Other Clostridiales	46	--	--	<1%	--	--	2%	--	--	1%	2%

[1] Species identity based on isolate *cpn*60 UT sequence in reference to cpnDB, [2] Total clones per individual; data for phylotype is shown as % total.

Similarly, only a single *Streptococcus* phylotype (*S. salivarius*) was detected in clone libraries, in a single BV+ individual. This phylotype was not detected by culture.

Clones with sequences closest to *G. vaginalis* are a significant proportion of total clones (14%). Extensive sequence differences mirrored the culture-based study, with dramatic "bushiness" for this phylotype (Figure 9.1). As observed for isolates, only a few clones were identical to the reference strain sequence. Most clones in this group were identical to the cultured isolates N72 and N160, and were observed in 3/10 individuals, all of whom were BV+.

Atopobium vaginae is an important, recently described member of the vaginal microbiota, detected using culture-independent methods (221). It was observed in 3/10 individuals (BV+ only), including one at 11% of total clones.

Several Bacteroidetes phylotypes were observed, in contrast to only 5 isolates from 3 phylotypes in the culture-based study. Most of these phylotypes had no near neighbour in cpnDB, although some were identical to *Prevotella* or *Porphyromonas* phylotypes. Several Clostridiales phylotypes were also observed that were not seen in the culture-based study and were phylogenetically unrelated to database sequences. These phylotypes were observed in 3 BV+, each colonized by several phylotypes in these groups. As well, one BV– individual had a low frequency (<1%) of one Clostridiales phylotype.

These findings reveal a distinct picture of the vaginal microbiota relative to selective culture, including detection of *L. iners*, *A. vaginae* and several phylotypes in phylum Bacteroidetes and order Clostridiales. However, many *Lactobacillus, Streptococcus* and *Staphylococcus* phylotypes were detected by culture but were not detected in clone libraries. The micro-heterogeneity of *G. vaginalis* sequences was

observed in both clone libraries and in isolates. The most frequently observed clones in this group had identical sequences to cultured isolates. Therefore, clone libraries and selective culture provide complementary information about the composition of vaginal microbiota.

Deep sequencing libraries – A very large number of partial *cpn*60 UT sequences was generated by pyrosequencing, providing a very rich dataset to contrast with isolate sequences and clone libraries (Table 9.2). Once deep sequencing data was sorted into individual bins based on sequence tag, the total number of deep sequencing reads processed for downstream analyses was just under 1 million (992,974).

Sequences were assembled into 888 phylotypes using Roche/454 software, representing 86% of all reads. Remaining reads did not cluster with any other reads and were disregarded. Read sequences were compared to a customized, non-redundant version of the *cpn*60 UT database ("vagnr" database, designed by J. Hill). Subsequently, 132 sequences derived from isolates and clone libraries in this study were added to the vagnr database ("vagnr+132"). This greatly improved the resolution of the deep sequencing libraries. The proportion of total reads matching database sequences at less than 70% was reduced (from 31% to 16%) and the proportion matching at greater than 95% was dramatically increased (from 36% to over 73%).

This observation suggests that initial characterization of vaginal microbial communities using established and gold standard techniques such as culture and clone library construction will improve the quality of phylogenetic information generated by deep sequencing. As a result, the vagnr+132 database was used for all subsequent analyses.

A total of 576 assemblies representing a total of 66,508 reads (7% of total) did not match anything in vagnr+132 at greater than 70% identity and were disregarded.

Table 9.2 – Deep sequencing run metrics

	1st run (July 2008)		2nd run (Nov. 2008)		
	HIV-N	HIV-R	HIV+	Feasibility[1]	Total
Raw wells	474,716	482,670	447,489	493,587	1,898,462
Filter-passed	249,650	249,486	279,392	286,869	1,065,397
% Filter-passed	54%	54%	65%	60%	56%
Mean length (bp[2])	242	242	241	243	242
Total bases	6.0×10^7	6.0×10^7	6.7×10^7	7.0×10^7	2.6×10^8

[1] Experiments addressing feasibility of pyrosequencing, [2] bp: base pairs

The remaining 312 assemblies representing 784,278 sequences were retained for downstream analysis. Assemblies were identified to the species level based on sequence identity, resulting in 93 phylotypes. A phylogenetic tree was constructed based on nearest full *cpn*60 UT sequences (Figure 9.2).

All phylotypes identified through pyrosequencing were from four bacterial phyla: Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria. Deep sequencing reveals a taxonomic profile that is largely consistent with what is known about the vaginal microbiota. Most phylotypes detected in isolates and/or clone libraries were also observed in deep sequencing, indicating the increased resolution of this technique.

Several *Lactobacillus* phylotypes that were observed in isolates but not in clone libraries were also observed in deep sequencing, including *L. coleohominis*, N54 and *L. delbrueckii*. Interestingly, many culture-based phylotypes represented by only a few isolates were not seen by deep pyrosequencing, indicating either that their absolute levels are still below the resolution of deep sequencing or that there is some unknown bias against detecting these phylotypes using this technique. Similarly, several *Streptococcus* phylotypes identified in isolates or clone libraries were not observed by deep sequencing.

The micro-heterogeneity of *G. vaginalis* sequences was also observed by deep sequencing, with a total of 21 phylotypes also seen in clone libraries and among isolates. Deep sequencing was able to detect *Bifidobacterium* sp. (only seen in isolates) and *A. vaginae* (only seen in clone libraries), as well as five other Actinobacteria phylotypes not seen by other methods. Detection of *Mobiluncus* and *Corynebacterium* by deep sequencing, although at very low levels, is not surprising since these organisms are well known members of the vaginal microbiota.

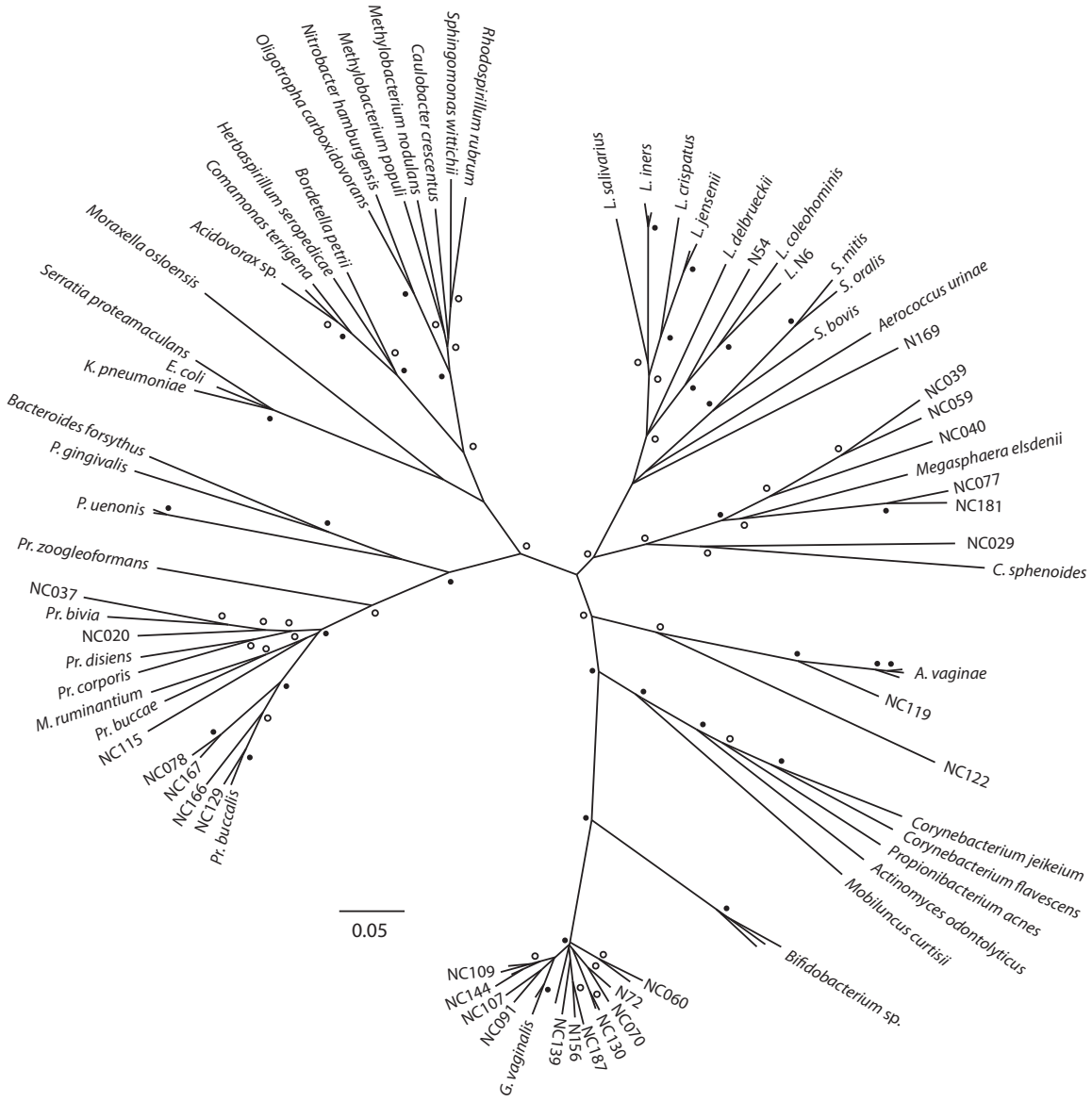

Figure 9.2 – Phylogenetic relatedness of deep sequencing phylotypes

Bootstrap consensus tree summarizing 784,278 partial UT sequences in 93 phylotypes. Tree is based on full-length UT of nearest neighbour in database, not on partial UT. Branch length indicates estimated phylogenetic distance based on the Maximum Composite Likelihood method (sum = 10.24, bar = 0.05). Bootstrap values greater than 95% (robust) and 50% (valid) are indicated with filled and open circles respectively, based on 1000 replicates (Mega 4.0). N=Nairobi isolate, NC=Nairobi clone, *G.=Gardnerella, A.=Atopobium,C.=Clostridium, L.=Lactobacillus, S.=Streptococcus, E.=Escherichia, K.=Klebsiella, P.=Porphyromonas, Pr.=Prevotella, M.=Micromonospora.*

Profile of Bacteroidetes and Clostridiales phylotypes is similar to what was observed in clone libraries, with several additional phylotypes in these groups that were only detected by deep sequencing. Two Proteobacteria phylotypes, *E. coli* and *K. pneumoniae*, were observed in isolates and deep sequencing, but not in clone libraries. Several phylotypes in phylum Proteobacteria were only observed by deep sequencing, including one about 90% identical to *Acidovorax facilis* and another about 90% identical to *Sphingomonas wittichii*, indicating that deep sequencing provides increased resolution to detect these phylotypes. A summary of the frequency that each phylotype was observed for the complete data set (N=62) and the pyro subset (N=48) is shown in Table 9.3.

In the culture subset, *G. vaginalis* and related phylotypes are a third of total reads, *Lactobacillus* sp. phylotypes are nearly one-quarter of total reads, with Bacteroidetes, Clostridiales and Proteobacteria each making up 10-20% of total reads.

As well as 600,000 reads for the Pyro subset, a further 125,000 reads were generated from 7 samples shipped in TRIzol reagent (Invitrogen) with DNA isolated according to manufacturer's directions (A. Meyers). This was done to address concerns about possible contamination of water in freezing buffer with environmental bacterial DNA, since storage in Trizol is water-free. A further 25,000 reads from 4 pilot study samples were also generated to compare the profile of Nairobi vs. Canadian individuals.

The overall profile of the Pyro subset is largely consistent with the TRIzol samples in terms of presence but not abundance of specific phylotypes. Interestingly, Proteobacteria phylotypes related to *Acidovorax* and *Sphingomonas* (thought to be possible environmental contaminants) were over-represented in the TRIzol samples compared to the Pyro subset, indicating that these phylotypes are not water-borne contaminants.

Table 9.3 – Summary of deep sequencing libraries

***cpn*60 UT ID (Phylotypes)[1]**	**Total Reads**	**% of Total**	**Pyro Subset**	**TRIzol samples[2]**	**Pilot Study**
Total Reads[3]	784,278		599,658	114,670	24,465
L. iners (3)	62,047	7.9	10.1	1.3	<0.1
L. crispatus	60,696	7.7	6.2	<0.1	96.9
L. jensenii (2)	21,305	2.7	3.6	--	<0.1
L. N6	13,828	1.8	2.2	0.3	<0.1
L. salivarius	2,171	0.3	0.3	0.4	--
N54	802	0.1	0.1	--	--
L. coleohominis	242	<0.1	<0.1	<0.1	--
L. delbrueckii	185	<0.1	<0.1	--	--
Streptococcus sp. (3)	3342	0.4	<0.1	2.8	<0.1
Aerococcus urinae	876	0.1	0.1	<0.1	--
G. vaginalis (21)	249,611	31.8	32.6	40.1	<0.1
A. vaginae (5)	8692	1.1	1.1	<0.1	--
Other Actinobacteria (5)	1207	0.2	0.1	0.3	0.4
NC078	42,280	5.4	4.0	15.0	<0.1
NC167	29,352	3.7	4.8	0.5	--
P. uenonis (2)	22,134	2.8	1.9	0.6	0.1
NC020	11,724	1.5	1.4	1.4	<0.1
Other Bacteroidetes (13)	33,038	4.2	3.0	7.3	0.3
NC181	52,189	6.7	8.2	<0.1	<0.1
NC029	45,673	5.8	5.6	0.8	<0.1
NC040	18,273	2.3	2.3	2.4	<0.1
Other Clostridiales (7)	10,258	1.3	1.2	1.4	<0.1
Sphingomonas wittichii	9,183	1.2	0.4	5.6	0.7
Other α-Proteobacteria (6)	885	0.1	<0.1	0.4	<0.1
Acidovorax sp. (2)	23,589	3.0	1.4	12.8	1.2
Other β-Proteobacteria (6)	1,176	0.1	0.1	0.4	<0.1
E. coli	51,865	6.6	8.5	0.5	<0.1
K. pneumoniae	2,094	0.3	0.3	<0.1	--
Other Proteobacteria (3)	105	<0.1	<0.1	<0.1	<0.1

[1] Species identity based on isolate *cpn*60 UT sequence in reference to cpnDB. Number in brackets refers to the number of phylotypes summarized in that row, [2] TRIzol samples : A set of 7 samples stored in Trizol and crude DNA isolated for deep sequencing, [3] Total Reads : Total number of reads for all deep sequencing libraries and subsets; phylotype data is shown as % of this total.

In contrast, *Lactobacillus* phylotypes appear to be under-represented in TRIzol samples. Although these were not paired samples, this observation indicates that storage buffer and/or DNA isolation method may have an important influence on deep sequencing results.

Pilot study samples were dominated by *L. crispatus*, however this is likely since most reads were from BV– samples. Interestingly, all major groups of phylotypes that were observed in Nairobi samples, including α- and β-Proteobacteria, were also observed at low abundance in pilot study samples, indicating a fundamental similarity in microbiota from these two very different groups of individuals.

Comparison of clone libraries and deep sequencing reads – Initially, *cpn*60 UT sequences in clone libraries were compared to sequences from deep sequencing of a single pooled sample (data not shown) (176). A total of 324 library clones and 4,410 reads were analyzed, representing 144 distinct *cpn*60 UT lineages, 72 of which were observed only by deep sequencing. Only one clone was completely distinct from anything seen by deep sequencing, indicating identification of novel sequences by deep sequencing (data not shown). Although taxonomic breadth was similar using the two methods, abundance of certain phylotypes was different, including over-representation of Actinobacteria and under-representation of Bacteroidetes and Lactobacillales by deep sequencing vs. clone libraries (not shown).

Next, profile of clones for four individuals with the largest clone libraries (877-996 clones/individual) were compared with matched deep sequencing read profiles (Table 9.4). The number of reads generated for each individual was approximately 1 log greater than the number of clones (9,061-21,302 reads/individual). As in the initial comparison, phylotypes for clone libraries were a subset of phylotypes identified by deep sequencing.

Table 9.4 – Comparison of clone libraries and deep sequencing reads

*cpn*60 UT ID[1]	**Total phylo.[2]**	1529 (BV–)		2316 (BV–)		2231 (BV+)		2347 (BV–)	
		Clones	Reads	Clones	Reads	Clones	Reads	Clones	Reads
Total clones/reads[3]		**996**	**9,782**	**920**	**9,061**	**922**	**21,302**	**877**	**9,994**
L. iners	**1**	100	95.1	28.0	1.0	47.8	0.3	--	0.2
L. crispatus	**1**	--	<0.1	70.9	92.5	--	<0.1	--	<0.1
L. salivarius	**1**	--	<0.1	--	--	--	--	--	--
L. jensenii	**1**	--	--	--	0.2	--	<0.1	--	<0.1
L. sp. L6	**1**	--	<0.1	--	<1%	--	<0.1	--	<0.1
Other Lactobacillales	**5**	--	--	--	--	--	<0.1	--	<0.1
G. vaginalis	**12**	--	1.1	0.1	2.2	28.3	33.8	1.9	20.1
A. vaginae	**2**	--	<0.1	--	--	0.5	<0.1	--	<0.1
Other Actinobacteria	**5**	--	<0.1	--	--	--	<0.1	--	<0.1
Any Bacteroidetes	**13**	--	0.5	--	0.5	12.4	32.4	56.2	64.8
Any Clostridiales	**5**	--	0.2	<1%	0.2	11.1	33.1	41.8	14.7
S. wittichii	**1**	--	0.1	--	--	--	<0.1	--	<0.1
Other α-Proteobacteria	**6**	--	0.3	--	<0.1	--	<0.1	--	--
Acidovorax sp.	**2**	--	2.3	--	0.1	--	<0.1	--	<0.1
Other β-Proteobacteria	**2**	--	0.1	--	--	--	<0.1	--	--
E. coli	**1**	--	--	--	<0.1	--	--	--	--
Phylotypes detected	**59**	**1**	**32**	**4**	**23**	**15**	**34**	**26**	**32**

[1] Species identity based on isolate *cpn*60 UT sequence in reference to cpnDB, [2] Total deep sequencing phylotypes for each UT ID, [3] Total clones per clone library or reads per deep sequencing library for each individual; data for phylotype by individual is shown as % total clones/reads, [4] Total phylotypes detected in clone library or deep sequencing read library for each individual.

Several more phylotypes were identified by deep sequencing compared to clone libraries, especially for the two BV– samples (1 vs. 32 and 4 vs. 23 respectively). These findings indicate that deep sequencing detects essentially everything detected by clone libraries and also generates information about many low abundance phylotypes not detected in clone libraries of this size.

As was observed in the initial analysis, a comparison of total clones vs. deep sequencing reads for each phylotype showed differences in abundances of specific phylotypes by each method. In two samples (2316 and 2231), *L. iners* made up a significant proportion of clone libraries but only a small proportion of reads. In one sample (2347), *G. vaginalis*-like phylotypes made up 20% of reads but only 2% of clone libraries. This is consistent with the initial analysis in which *Lactobacillus* was over-represented and *Gardnerella* under-represented in clone libraries, likely reflecting different biases inherent in cloning and ligation steps vs. emulsification PCR and bead layering. Despite these inconsistencies, presence of phylotypes was reliably detected by both methods.

Low abundances of BV-related bacteria were observed in BV– samples and low abundances of *Lactobacillus* phylotypes were observed in BV+ samples by deep sequencing but not in clone libraries, demonstrating persistence of small populations of bacteria typical of BV– microbiota in BV+ samples and vice versa. This observation reinforces the concept of BV as a different organization of co-existing microbial populations rather than acquisition of a specific pathogen.

Reproducibility of deep sequencing profiles – To determine the reproducibility of deep sequencing profiles, technical replicates for two samples were examined, with each sample amplified independently using different sequence-tagged primers (Table 9.5).

Table 9.5 – Comparison of technical replicates for deep sequencing

A. Comparison of phylotypes identified in replicates

*cpn*60 UT ID[1]	2648 (BV+)		2231 (BV+)	
	Original	Replicate	Original	Replicate
Total reads[2]	**10,800**	**34,112**	**21,302**	**3,365**
L. iners	0.2	--	0.3	0.1
L. crispatus	<0.1	--	<0.1	--
L. salivarius	--	--	--	--
L. jensenii	<0.1	--	<0.1	--
L. sp. L6	<0.1	--	<0.1	--
Other Lactobacillales	<0.1	--	<0.1	--
G. vaginalis	2.3	2.7	29.3	21.1
A. vaginae	2.7	4.4	<0.1	<0.1
Other Actinobacteria	<0.1	<0.1	<0.1	--
P. uenonis	34.1	30.0	0.4	1.3
Other Bacteroidetes	17.1	21.8	32.0	43.9
NC029	30.9	28.9	28.1	28.0
Other Clostridiales	12.0	12.4	5.0	5.2
S. wittichii	--	--	<0.1	<0.1
Other α-Proteobacteria	--	--	<0.1	<0.1
Acidovorax sp.	<0.1	<0.1	<0.1	0.2
Other β-Proteobacteria	<0.1	--	<0.1	<0.1
E. coli	--	<0.1	--	--

[1] Species identity based on isolate *cpn*60 UT sequence in reference to cpnDB, [2] Total reads per deep sequencing library for each individual; data for phylotype by individual is shown as % total reads.

B. Comparison of ecological measures between replicates

Ecological parameter[3]	2648 (BV+)		2231 (BV+)	
	Original	Replicate	Original	Replicate
Density (Total phylotypes)	32	29	35	22
Diversity (Shannon's index)	1.85	1.96	1.77	1.64
Evenness (Shannon's Index)	0.54	0.58	0.50	0.53
Shared membership (Jaccard's index)		0.65		0.54
Shared structure (Theta index)		0.98		0.94

[3] Index values for ecological measures calculated as described in Methods, p.64. Values for Jaccard's and Theta indices are comparisons of original and replicate for each sample.

For the first sample, three times as many reads were generated for the replicate compared to the original run. The overall profile was virtually identical in either replicate, although several low-abundance *Lactobacillus* phylotypes were not observed in the replicate, including *L. iners*, present at 0.19% of total reads in the original sample (Table 9.5, A).

In the second sample (2231), the replicate had a much smaller number of reads compared to the original sample. Despite this, overall profile was very similar in either replicate. As in the first sample, several *Lactobacillus* phylotypes detected at low abundance in the original sample were not seen in the replicate, although *L. iners* was detected in either replicate for this sample.

In terms of ecological similarity, phylotype density (ie. total number of phylotypes observed) was similar in replicates of the first sample (32 vs. 29) but a reduced number of phylotypes was observed in one replicate of the second sample (35 vs. 22) (Table 9.5, B). This difference is expected given the far smaller number of reads in the replicate compared to the original for that sample. Despite this, diversity and evenness indices were very similar for either replicate in both samples, indicating that community structure (ie. relative abundance and dominance of phylotypes) is very similar in either replicate regardless of the number of reads or phylotypes identified.

Indices describing shared community membership (Jaccard's) and shared community structure (Theta) were also calculated. Jaccard's index describes how similar two samples/groups are in terms of phylotypes in common, while the Theta index describes how similar two samples are in terms of the relative abundances of phylotypes in common. Both indices have possible values ranging from 0 to 1, with 0 indicating no shared phylotypes and 1 indicating all shared phylotypes (see Methods, p. 64).

Since many low abundance phylotypes were not shared between replicates, the shared membership index value is low for both samples and is lowest in the sample with the greatest discrepancy in numbers of reads and phylotypes identified. However, despite these differences, shared structure index values are close to 1 for both samples, indicating that replicates are nearly identical in terms of the relative abundance of phylotypes.

These findings indicate that *cpn*60 UT-based deep sequencing creates a highly reproducible profile of micro-organisms present at greater than 0.2% of total reads within an individual sample, despite large differences in the total number of reads generated in replicates (176). Further work will be required to establish a reliable cutoff for reproducibility of phylotype detection by deep sequencing.

Comparison of 16S to cpn60 UT – Since 16S rDNA is the most common target used for profiling of microbial communities, the performance of *cpn*60 UT in relation to 16S rDNA was evaluated. Profile of all deep sequencing UT reads for four individuals was compared to profile of 16S rRNA reads generated separately for the same individuals and pooled prior to tagging and deep sequencing (Figure 9.3) (176).

The overall profile is remarkably similar, with most phylotypes detected using either target, including *L. iners, L. crispatus, L. salivarius, G. vaginalis* and *A. vaginae*. However, differences in relative abundance of each phylotype was observed, particularly for *G. vaginalis* phylotypes, which made up 43% of UT reads but only 0.2% of 16S reads.

In contrast, Bacteroidetes and Clostridiales phylotypes made up comparable proportions of reads for both targets, however specific phylotypes were not comparable since most UT reads for these groups matched most closely to clone library sequences that could not be represented in the 16S database.

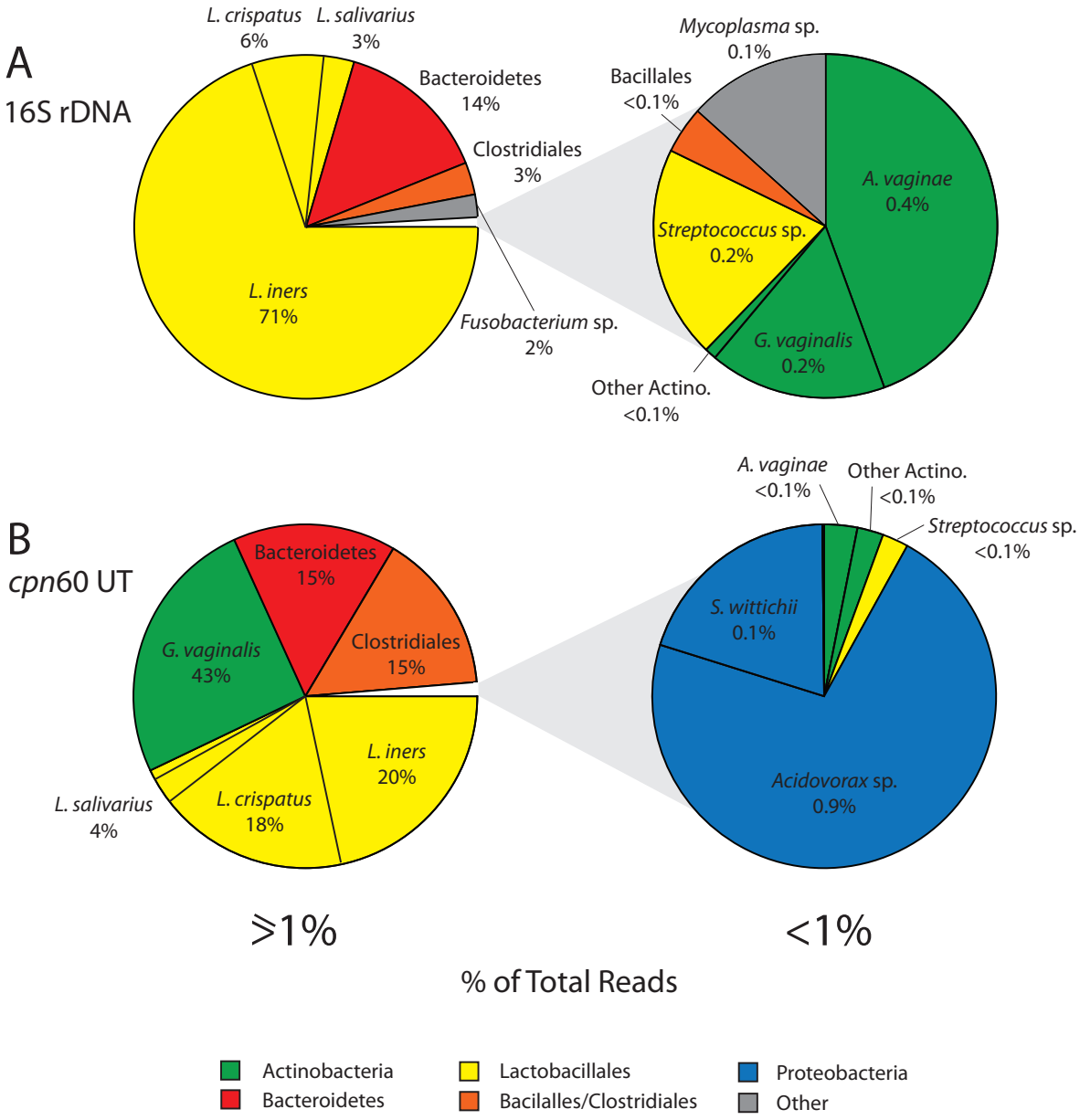

Figure 9.3 – Comparison of deep sequencing profiles for 16S rDNA vs. *cpn*60 UT

Profile of deep sequencing reads for four samples: A. amplified with 16S variable region targetted primers and pooled prior to deep sequencing, and B. the same four samples amplified using *cpn*60 UT targetted primers and pooled virtually after deep sequencing. Phylotypes or phylotype groups present at equal or greater than 1% of total reads are shown on left, while those present at <1% are shown on right. *L.=Lactobacillus, A.=Atopobium, S.=Sphingomonas*, Actino.=Actinobacteria.

Several low abundance phylotypes were identified by only one or the other target. *Fusobacterium, Mycoplasma* and Bacillales phylotypes were only observed by 16S and α/β-Proteobacteria phylotypes were only observed by *cpn*60 UT. These differences likely reflect biases for each target, including differences in amplification efficiency of different templates or ambiguities in taxonomic assignment for 16S vs. *cpn*60 UT databases.

Interestingly, analysis of phylotypes identified using either target showed that the number of distinct phylotypes was significantly higher for the *cpn*60 dataset, suggesting greater taxonomic resolution for this target (data not shown) (176). Consistent with this observation, Shannon's diversity index was higher for *cpn*60 UT compared to 16S (2.28 vs. 1.21) and Shannon's equitability index was also higher (0.57 vs. 0.33), indicating improved detection of a wider range of phylotypes using *cpn*60 compared to 16S.

These findings indicate that *cpn*60 UT is a robust alternative to 16S for phylogenetic analysis of vaginal microbial communities by deep sequencing. Shorter target length and increased resolution of closely related phylotypes are key benefits of *cpn*60 UT, particularly considering shorter average read lengths generated by deep sequencing. Further work will be required to determine biasesassociated with either target.

Tiers of abundance – Deep sequencing results in a very large and highly detailed dataset describing microbial communities at a very high resolution compared to clone libraries or culture-based analyses. Correspondance between the frequency that a phylotype is observed by deep sequencing (ie. number of reads) and its actual distribution in the microbiota are subject to a number of biases that limit quantitative analysis, however the sheer number of sampling units evaluated in this study is unparalleled compared to older techniques. Therefore, the biological significance of variation in frequencies of

phylotypes or groups of phylotypes by deep sequencing requires elucidation in order to evaluate the potential of this technique for quantitative or semi-quantitative analyses.

An examination of phylotype frequency as the percentage of total reads for the Pyro subset (599,658 reads in 48 samples) reveals that only 21/79 phylotypes are observed at a frequency of greater than 1% of total reads (Figure 9.4, A). Consistent with what was observed for clone libraries, the most frequently observed deep sequencing phylotype was *L. iners*, making up just over 10% of total reads. In contrast, most phylotypes were observed at <1% of total reads and just over half were observed at <0.1% of total reads.

Phylotypes present at low frequency overall were often present at high frequencies (>10%) in individual samples. Therefore, the term "tiers of abundance" was coined to describe the variation in read frequency observed by phylotype across samples.

An inverse relationship was observed between species density and abundance, with 43% of phylotypes and 82% of total reads represented at greater than 10% in any individual library, while 95% of phylotypes and 0.4% of reads were represented at less than 0.1% of any individual library (Table 9.4, B). Four "tiers" of abundance were defined, based on the proportion of total reads per phylotype per library: >10%, 1-9.9%, 0.1-0.99% and <0.1% of total reads. Most phylotypes were represented at all tiers of abundance and virtually all were represented at the lowest, suggesting omnipresence of most phylotypes and fluctuations in population size across several orders of magnitude (Figure 9.4, C).

Since vaginal microbiology has until now been defined in terms of phylotypes observable by culture or by clone libraries (ie. in higher tiers of abundance), these observations indicate that deep sequencing has the potential to fundamentally reorient our understanding of the composition and structure of vaginal microbial communities.

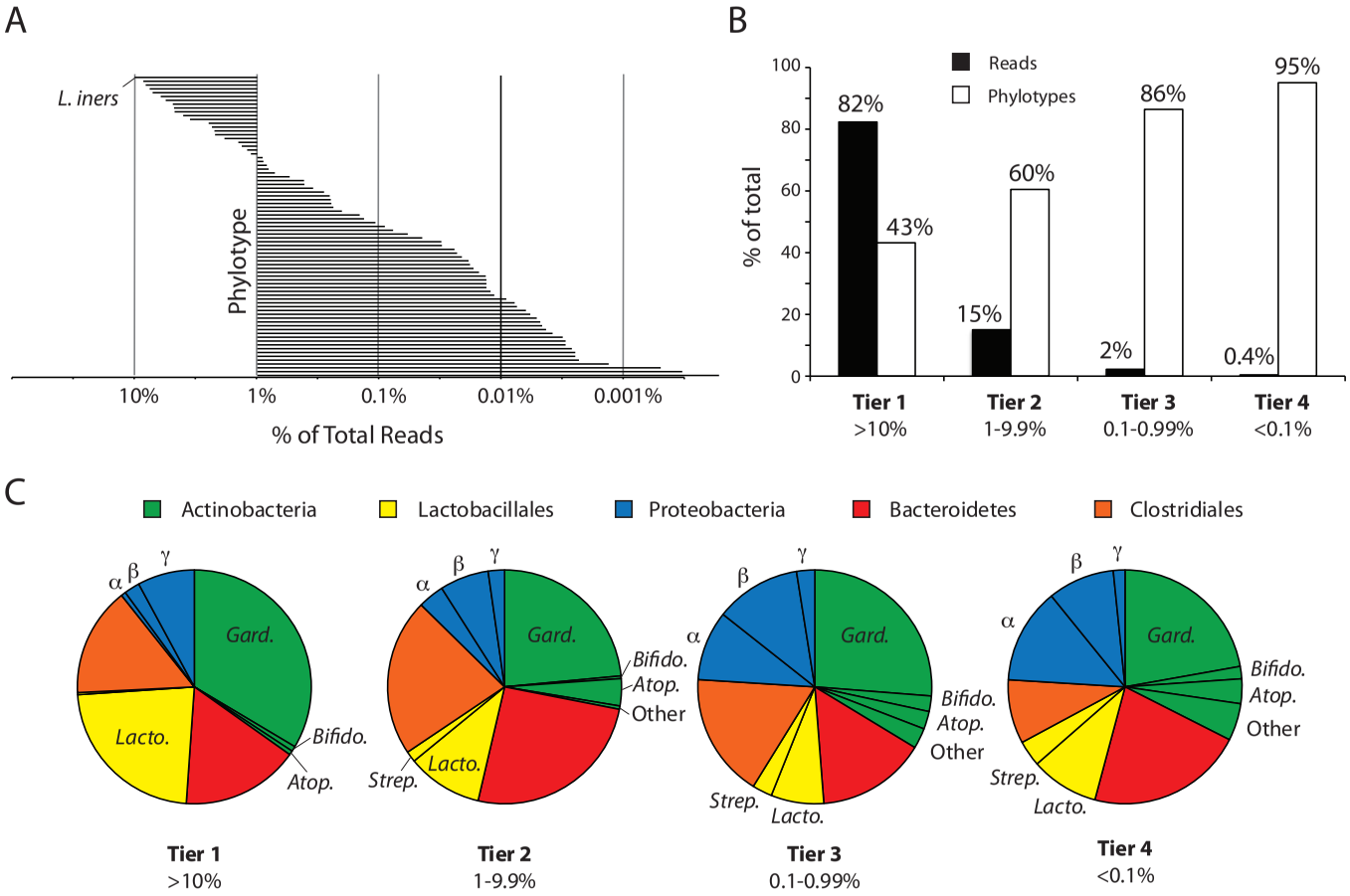

Figure 9.4 – Tiers of abundance

Variation in read frequency for phylotypes identified by deep sequencing. A. Frequency of 79 phylotypes observed in Pyro subset (599,658 reads) sorted by percentage of total reads (logarithmic scale) for each phylotype. *L. iners* was observed most frequently (over 10% of total reads. B. Inverse association between total reads and total phylotypes in four "tiers of abundance", based on percentage of reads for each phylotype per sample: Tier 1 – 10% or more, Tier 2 – 1% to 9.9%, Tier 3 – 0.1% to 0.99%, Tier 4 – less than 0.1%.C. Representation of phylotype groups at different tiers of abundance. *Gard.=Gardnerella, Strep.=Streptococcus, Lacto.=Lactobacillus, Bifido.=Bifidobacterium, Atop.=Atopobium,* α=alpha-, β=beta and γ=gamma Proteobacteria.

Phylogenetic characteristics of individual libraries – Based on these data, a number of variables were derived to describe individual samples. The mean number of reads per individual was just over 11,000, however the number of reads per individual varied greatly, particularly in the second deep sequencing run. Total reads ranged from 121 to over 45,000, with six samples under 500 reads and considered "failed" reactions. When "failed" reactions were excluded the range of total reads was 6,313 to 45,722.

The total number of sequence assemblies defined using Roche/454 software ranged from 20 to 98 per sample and correlated significantly with total reads ($R^2 = 0.40$, $p<0.0001$ by linear regression, not shown). The number of distinct phylotypes, based on identification of nearest neighbours for each assembled sequence using the vagnr+132 database, ranged from 10-39 per individual and was also significantly correlated with total reads ($R^2 = 0.21$, $p=0.001$ by linear regression, not shown). These correlations remained significant when excluding failed reactions, indicating that more assemblies and phylotypes are detected when more sequences are generated.

Calculation of diversity and evenness indices was carried out as described above (Table 9.6). The diversity index provides information about total phylotype density in a sample in relation to the relative abundance of each phylotype, while the evenness index provides information about how evenly represented each phylotype is in the sample. Lower index values for both measures are indicative of a smaller number of dominant phylotypes.

In terms of phylotypes detected, all individuals in the pyro subset had detectable *Lactobacillus*, *Gardnerella* and Bacteroidetes phylotypes, and all but one had Clostridiales phylotypes (Table 9.7). Nearly all individuals had *L. iners*, *G. vaginalis*, *S. wittichii* and *A. facilis*, while half of samples had *A. vaginae* or *E. coli*.

Table 9.6 – Ecological measures and comparison between HIV-R and HIV-N

	Pyro Subset (N=48)	**HIV-R** (N=16)	**HIV-N** (N=16)	**Diff.**[1]
Total Reads				
Median	11,088	10,318	11,533	--
Range	122-45,722	2,464-30,850	4,027-21,336	
Total Assemblies				
Mean (SD)	55.5 (19.7)	55.4 (17.7)	60.4 (16.3)	--
Range	20-98	21-89	30-93	
Phylotype Density				
Mean (SD)	26.6 (6.8)	26.3 (6.9)	28.3 (6.9)	--
Range	10-39	10-39	11-39	
Shannon's Diversity Index				
Mean (SD)	1.2 (0.6)	1.1 (0.5)	1.3 (0.6)	--
Range	0.2-2.2	0.3-1.9	0.4-2.1	
Shannon's Evenness Index				
Mean (SD)	0.38 (0.16)	0.35 (0.14)	0.39 (0.16)	--
Range	0.07-0.70	0.09-0.58	0.11-0.62	

[1] Significant difference between HIV-R and HIV-N by Mann-Whitney test (for Total Reads) or t test (other variables).

Table 9.7 – Phylotype presence/abundance and comparison of HIV-R/HIV-N

	Pyro Subset MR (%C)[1]	**HIV-R** MR (%C)	**HIV-N** (MR) %C	**Diff.**[2] MR/%C
All *Lactobacillus*	3,221 (100)	3,456 (100)	4,342 (100)	--/--
L. iners	827 (96)	781 (100)	491 (94)	--/--
L. N6	63 (60)	41 (69)	305 (81)	--/--
L. crispatus	15 (48)	34 (50)	12 (81)	--/--
L. jensenii	12 (54)	9 (44)	17 (81)	--/--
All *Gardnerella*-like	11,798 (100)	16,198 (100)	6,906 (100)	--/--
G. vaginalis	319 (96)	997 (100)	95 (100)	--/--
NC137	127 (92)	120 (94)	294 (100)	--/--
N72	75 (83)	85 (88)	29 (88)	--/--
NC187	61 (65)	61 (81)	11 (69)	--/--
NC088	29 (56)	9 (57)	15 (31)	--/--
NC144	20 (44)	31 (50)	15 (44)	--/--
NC070	12 (25)	8 (6)	15 (75)	p=0.0002*
A. vaginae	73 (54)	56 (63)	199 (56)	--/--
C. flavescens	16 (42)	22 (44)	6 (31)	--/--
M. curtisii	11 (23)	6 (6)	11 (56)	p=0.006*
All Clostridiales	1,242 (98)	740 (100)	2,360 (94)	--/--
NC029	83 (77)	56 (75)	136 (81)	--/--
NC181	67 (69)	38 (69)	96 (69)	--/--
NC040	758 (90)	237 (94)	550 (94)	--/--
NC039	116 (58)	131 (69)	314 (69)	--/--
All Bacteroidetes	332 (100)	288 (100)	994 (100)	--/--
NC167	50 (65)	38 (38)	88 (75)	--/--
NC078	86 (83)	90 (94)	49 (75)	--/--
P. uenonis	66 (50)	78 (38)	553 (75)	--/--
NC166	9 (35)	27 (31)	35 (50)	--/--
NC115	19 (48)	171 (38)	19 (69)	--/--
NC020	18 (38)	73 (63)	7 (38)	--/--
NC037	24 (52)	155 (44)	52 (50)	--/--
All α-Proteobacteria	150 (88)	150 (88)	46 (88)	--/--
S. wittichii	126 (85)	94 (88)	46 (81)	--/--
All β-Proteobacteria	340 (92)	726 (94)	233 (94)	--/--
Ac. facilis	331 (88)	609 (88)	213 (94)	--/--
E. coli	146 (52)	6 (31)	8 (31)	--/--

[1] MR: Median normalized reads for group, %C: % of group colonized, [2] Difference between HIV-N and HIV-R, as determined using Mann Whitney test (to compare median reads in those colonized) and Fisher's exact test (to compare % colonized in each group). * p value for % colonized in each group only (mean reads comparison not carried out since only 1 individual colonized in HIV-R group). N=Nairobi isolate, NC=Nairobi clone, *A.=Atopobium, C.=Corynebacterium, M.=Mobiluncus, P.=Porphyromonas, S.=Sphingomonas, Ac.=Acidovorax, E.=Escherichia.*

Comparison of HIV-R and HIV-N individuals – No differences were observed between HIV-R and HIV-N individuals in terms of phylotype diversity, disproving Hypothesis 11 (Table 9.6). Shared community membership between HIV-R and HIV-N individuals was 0.76 by Jaccard's index and shared community structure was 0.60 by Theta index. Since technical replicates of a single sample have index values of ~0.61 for membership and ~0.96 for structure, these findings indicate a relatively high level of shared membership and a relatively low level of shared structure between HIV-N and HIV-R. Further work will be required to determine the significance of these parameters.

No differences between HIV-R and HIV-N were observed for colonization with or relative abundance of any *Lactobacillus* or specific phylotypes, disproving Hypotheses 7 and 9 (Table 9.7). Interestingly, two BV-related phylotypes, *Gardnerella*-like NC070 and *Mobiluncus curtisii*, were less likely to be observed in HIV-R compared to HIV-N ($p=0.0002$ and $p=0.006$ respectively by Fisher's exact test), in support of Hypothesis 8. Since only one HIV-R individual was colonized with either of these phylotypes, relative abundance could not be compared, therefore Hypothesis 10 was not supported by this data.

NC070 sequences represent ~1% of total reads in the Pyro subset, while *M. curtisii* sequences represent ~0.1% of total reads, raising questions about the biological significance of low abundance phylotypes. Lack of replicability in detecting phylotypes present at less than 0.2% of total reads also limits the significance of these observations. Further, Bonferroni correction of p values for multiple comparisons reduces the significance of these observations substantially. Therefore, these findings do not constitute a robust difference between HIV-N and HIV-R individuals. Further work will be required to determine if these phylotypes are indeed differently distributed in these groups.

Phylotype presence/abundance by serostatus – Interestingly, HIV+ individuals were more likely to have detectable *E. coli* compared to HIV– individuals (15/16 or 94% vs. 10/32 or 31%, $p<0.0001$ by Fisher's exact test), and also had a significantly higher median abundance of *E. coli* (251 vs. 6 reads, $p=0.01$ by Mann Whitney test). HIV+ individuals were also less likely to have detectable *L. crispatus* compared to HIV– individuals (2/16 or 13% vs. 21/32 or 66%, $p=0.0006$ by Fisher's exact test), although median relative abundance for this phylotype could not be accurately calculated since only two HIV+ individuals had detectable levels.

Deep sequencing libraries were generated for two samples from a single individual (ML2039) collected six months apart. This individual tested HIV– at the first timepoint and HIV+ at the second timepoint, therefore these samples were compared in order to observe potential effects of seroconversion on vaginal microbiota (Figure 9.5).

This individual was diagnosed as BV+ at the first timepoint and BVI at the second timepoint, which appears to be reflected in an increase in *L. iners* reads at the second timepoint. Shared membership by Jaccard's index was 0.47, approaching the value observed in technical replicates (0.61), while shared structure by Theta index was quite low at 0.17, reflecting the presence of many phylotypes in both samples, although at different abundances. These observations indicate that deep sequencing is likely to be a robust tool for longitudinal analysis of vaginal microbiota in the same individual.

Consistent with observations in HIV+ individuals, *L. crispatus* was only observed pre-seroconversion and *E. coli* was only observed post-seroconversion. However, both of these phylotypes were present at very low levels (<0.1%). Further work will be required to determine if these phylotypes are in fact differently represented in HIV+ individuals.

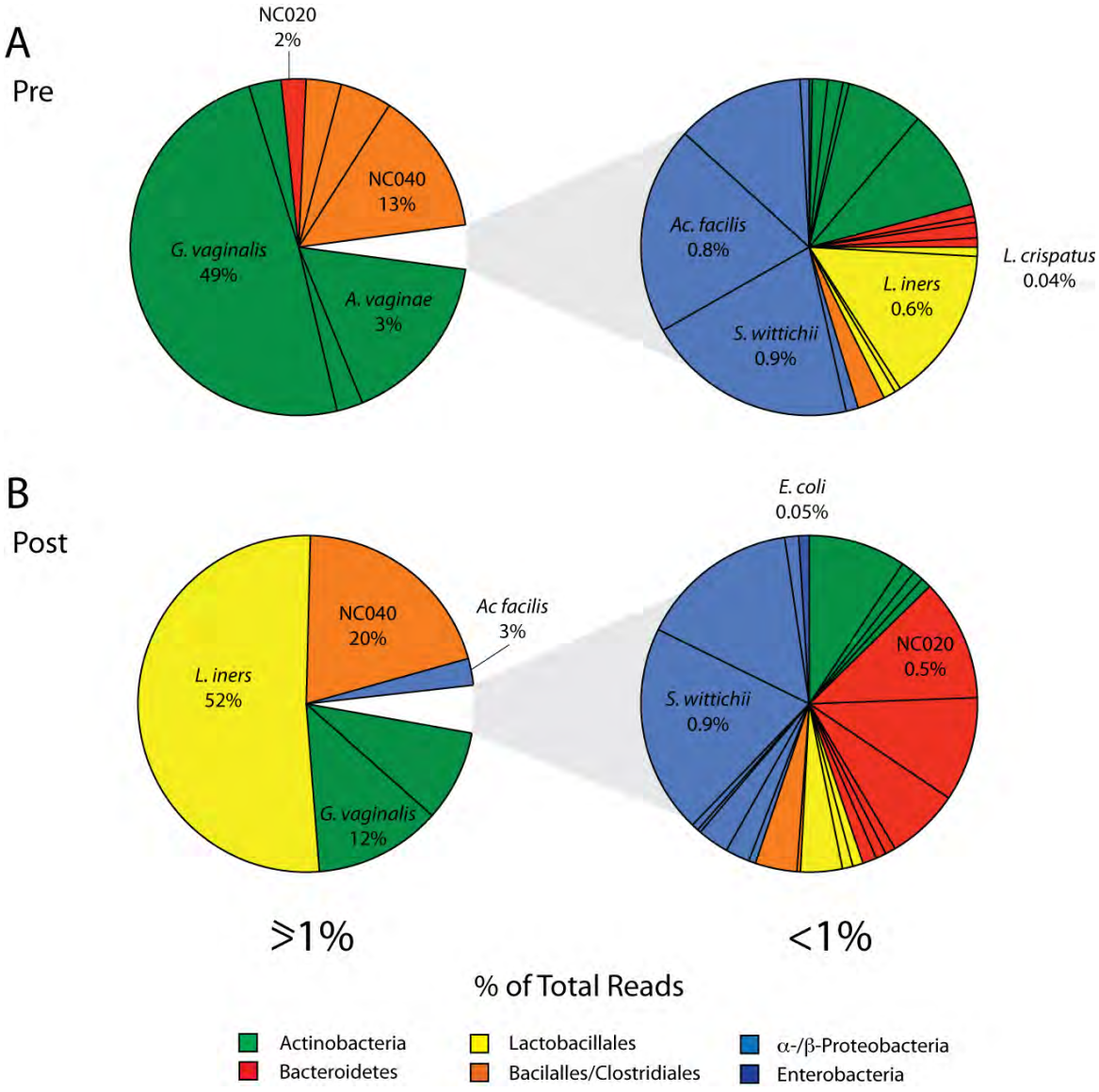

Figure 9.5 – Comparison of deep sequencing profiles pre-/post-seroconversion

Profile of deep sequencing reads for two samples from a single individual collected six months apart, with seroconversion occurring in the interval. A. First sample pre-seroconversion, and B. second sample post-seroconversion. Phylotypes present at equal or greater than 1% of total reads are shown on left, while those present at <1% are shown on right. NC=Nairobi clone, *G.=Gardnerella.=Lactobacillus, A.=Atopobium, Ac.=Acidovorax, S.=Sphingomonas*.

Phylotype presence/abundance by BV diagnosis – Analysis of presence and abundance of specific phylotypes and phylotype groups reveals a close correspondence between deep sequencing profile and expected clinical characteristics associated with BV (Table 9.8). Abundance (but not presence) of *Lactobacillus* phylotypes, *L. iners*, any Clostridiales phylotypes or any Bacteroidetes phylotypes was the strongest predictor of BV diagnosis. Interestingly, neither presence nor abundance of *L. crispatus* was significantly associated with BV diagnosis, surprising since this phylotype is most often characterized as a critical part of BV– microbiota.

Abundance of *Gardnerella* phylotypes was only weakly associated with BV, also surprising considering the importance of *G. vaginalis* in defining BV microbiota. Two species well-known to be associated with BV, *A. vaginae* and *M. curtisii*, were both observed, however only *A. vaginae* was (weakly) associated with BV diagnosis.

Most Bacteroidetes and Clostridiales phylotypes were found at significantly higher levels in BV samples and at very low levels in normal samples. Interestingly, several Proteobacteria phylotypes, especially *Acidovorax facilis*, were found at lower levels in BV compared to BV– samples. This is surprising considering that only *Lactobacillus* phylotypes have previously been reported to be more likely to be present or at greater abundance in BV– individuals.

These data demonstrate the biological significance of quantitative differences in read frequencies. Increased resolution of deep sequencing compared to older established techniques such as clone libraries indicates that BV– microbiota may be more phylogenetically complex than previously described.

Table 9.8 – Phylotype presence/abundance and BV diagnosis

	BV (N=28) MR (%C)[1]	**BV–** (N=20) MR (%C)[1]	**Difference**[2] MR/%C[1]
All *Lactobacillus*	226 (100)	32,370 (100)	**p<0.0001**/--
L. iners	101 (93)	7,414 (100)	**p<0.0001**/--
L. N6	11 (46)	416 (85)	p=0.003/p=0.008
L. crispatus	16 (50)	12 (45)	--/--
L. jensenii	8 (50)	151 (65)	p=0.03/--
Ae. urinae	11 (57)	31 (20)	--/p=0.02
All *Gardnerella*-like	15,107 (100)	1,361 (100)	p=0.01/--
G. vaginalis	872 (96)	182 (95)	--/--
NC137	244 (96)	33 (85)	p=0.04/--
N72	1072 (86)	32 (80)	p=0.02/--
NC187	152 (68)	7 (60)	p=0.04/--
NC088	50 (57)	26 (55)	--/--
NC144	22 (57)	15 (20)	--/p=0.04
NC070	7 (21)	23 (30)	--/--
A. vaginae	199 (68)	6 (35)	p=0.04/p=0.04
M. curtisii	94 (14)	6 (30)	--/--
All Clostridiales	5,869 (100)	210 (95)	**p=0.0001**/--
NC029	331 (82)	14 (70)	p=0.006/--
NC181	370 (68)	46 (70)	--/--
NC040	1,096 (96)	147 (80)	--/--
NC039	314 (75)	4 (35)	**p=0.001**/p=0.008
All Bacteroidetes	12,350 (100)	124 (100)	**p<0.0001**/--
NC167	174 (71)	24 (60)	p=0.02/--
NC078	181 (86)	39 (80)	p=0.04/--
P. uenonis	207 (57)	34 (40)	--/--
NC166	32 (39)	6 (30)	--/--
NC115	89 (61)	6 (30)	--/p=0.05
NC020	222 (43)	18 (30)	--/--
NC037	149 (64)	16 (35)	p=0.05/--
All α-Proteobacteria	101 (79)	211 (100)	p=0.03/p=0.03
S. wittichii	91 (79)	164 (95)	p=0.03/--
All β-Proteobacteria	202 (86)	886 (100)	p=0.008/--
Ac. facilis	209 (82)	937 (95)	p=0.004/--
E. coli	188 (39)	30 (70)	--/p=0.05

[1] MR: Median normalized reads for group, %C: % of group colonized, [2] Difference between BV and BV–, as determined using Mann Whitney test (to compare median reads in those colonized) and Fisher's exact test (to compare % colonized in each group). Bold indicates that p value remains significant after Bonferroni correction for multiple comparisons (p * 34 comparisons). Abbreviations as for Table 9.7, *Ae.=Aerococcus*.

Interestingly, distinct deep sequencing patterns were observed when visualizing all reads for an individual sample (Figure 9.6). Samples for HIV-R, HIV-N and HIV+ groups were sorted by decreasing proportion of reads for *Lactobacillus* phylotypes. For HIV– individuals, several samples were dominated by *Lactobacillus* phylotypes at >90% of total reads. All of these samples were BV– by Nugent score. When proportion of total reads for *Lactobacillus* is less than ~25% but greater than ~1%, *Gardnerella*-related phylotypes are dominant, making up greater than two thirds of total reads. These samples either had BV–, BVI or BV+ Nugent scores. When *Lactobacillus* proportion is less than ~1%, co-dominance of *Gardnerella*, Clostridiales and Bacteroidetes in roughly equal proportions is observed. Nearly all of these samples were BV+ by Nugent score.

These findings indicate that three deep sequencing patterns can be discerned in HIV– individuals that roughly correspond to BV– and BV+ Nugent scores. Further, a molecular definition for BVI microbiota is proposed, consisting of dominant *Gardnerella* phylotypes in the context of reduced *Lactobacillus* and low levels of Bacteroidetes and Clostridiales. Unfortunately, no BV0 samples in HIV– individuals were selected for deep sequencing, therefore the culture-based observation that *L. crispatus* is more likely to be present in these samples could not be confirmed. Future studies may provide a molecular definition for this category and confirm the hypothesis that BV0 is a transition phase between BV+ and BV– states, just as BVI is a transition between BV– and BV+ states.

Patterns of BV vs. BV– microbiota described for HIV– individuals were not clearly observed in most HIV+ individuals. In fact, no HIV+ individuals had >80% *Lactobacillus* reads and several had samples dominated by Proteobacteria (*Acidovorax, Sphingomonas, E. coli*).

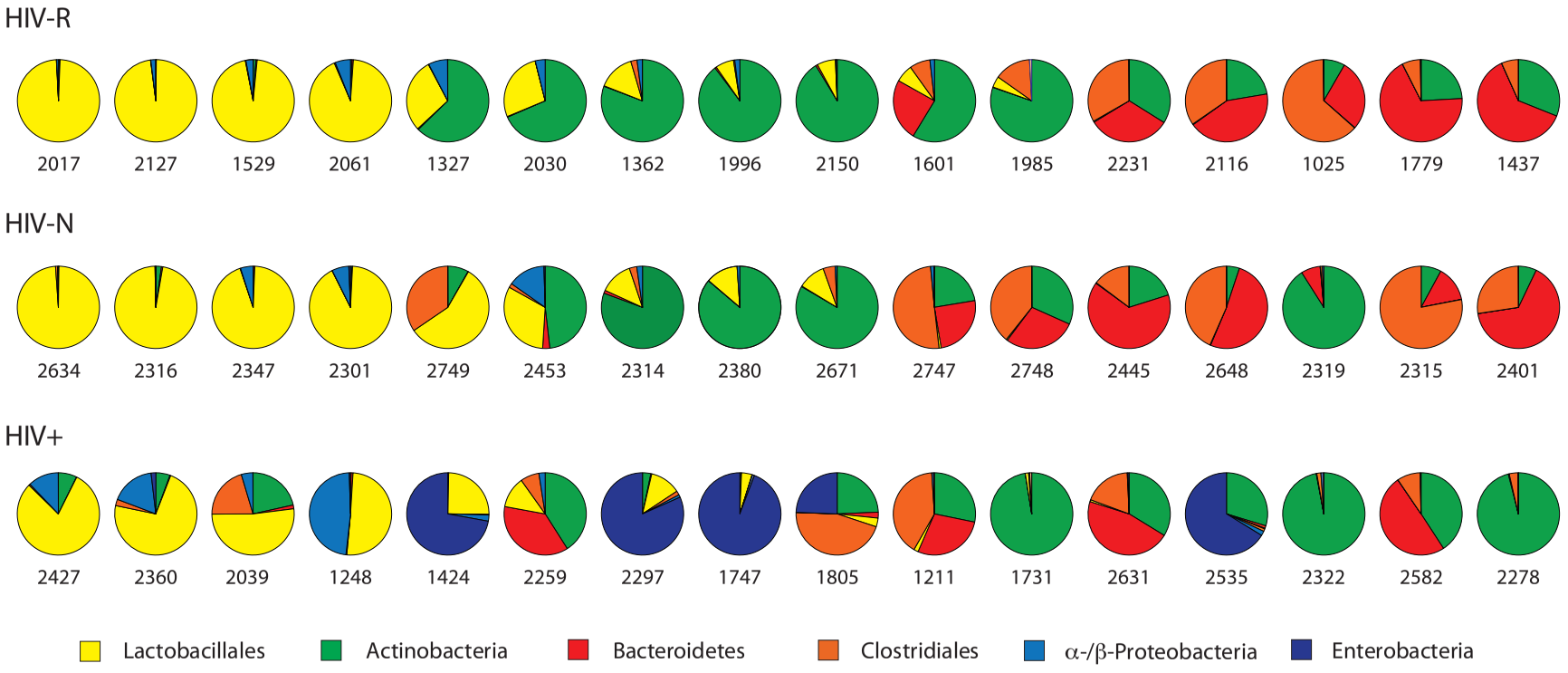

Figure 9.6 – Phylogenetic characteristics of individual deep sequencing libraries

Profiles for HIV-R, HIV-N and HIV+ individuals, ordered from left to right by decreasing proportion of *Lactobacillus* phylotypes. For HIV– individuals, decreased *Lactobacillus* is associated with dominant *Gardnerella*-like phylotypes, while lowest abundance of *Lactobacillus* is characterized by co-dominance of *Gardnerella*, Bacteroidetes and Clostridiales phylotypes. In HIV+ individuals, no samples have *Lactobacillus* at greater than 80% of total reads, while several samples have very high levels of Proteobacteria reads.

This observation is consistent with previous observations in HIV+ individuals, including increased BV, indicating that vaginal microbiota may be dysregulated in some way by HIV infection, ARV treatment or Septrin prophylaxis. Further work will be required to confirm this observation.

Ecological measures by BV diagnosis – BV microbiota is often described as more species rich and more diverse compared to BV– microbiota, suggesting that BV-related bacteria are acquired that displace dominant *Lactobacillus* phylotypes in BV– individuals. However, it is evident from analysis of deep sequencing data that many low abundance phylotypes are detectable that would be missed using other established techniques such as culture or clone library construction. Many phylotypes typical of BV microbiota can be observed at low abundance in BV– individuals and vice versa, confirming the concept that BV is a different organization of endogenous microbiota rather than the acquisition of exogenous bacteria.

This concept is formally demonstrated in this study by comparing phylotype density in BV vs. BV– individuals for all deep sequencing reads to density when considering only phylotypes present at >1% of total reads (Figure 9.7, A/B). Phylotype density in individuals with BV was significantly higher when considering only reads present at >1% (Figure 9.7, A), however density was identical when considering all deep sequencing reads (Figure 9.7, B). This observation indicates that BV and BV– individuals have an equivalent number of phylotypes, a fact likely to be overlooked when examining only the most abundant phylotypes. As expected, diversity and evenness indices were substantially increased in BV vs. BV– regardless of subset analyzed (Figure 9.7, C/D), indicating that BV– microbiota is dominated by a smaller number of phylotypes.

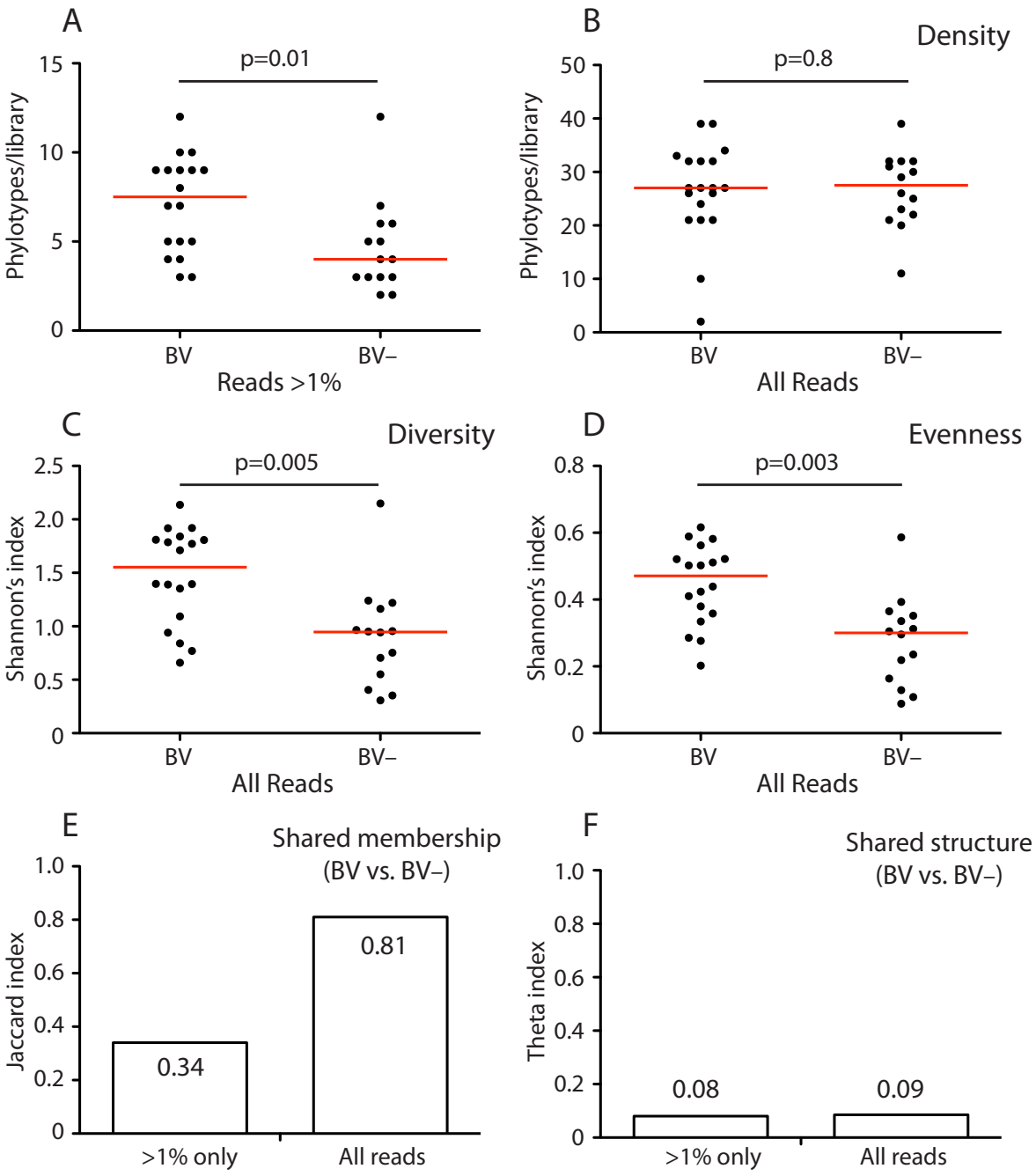

Figure 9.7 – Ecological measures by BV diagnosis

Deep sequencing reveals fundamental similarity in community membership between BV and BV– microbiota. A. Species density is increased in BV vs. BV– individuals when considering phylotypes present at highest abundance (>1% of total reads), but, B. Species density is identical when considering all phylotypes. C/D. Diversity and evenness, calculated using Shannon's indices, are significantly higher in BV vs. BV– libraries. E. Shared membership (Jaccard index) between BV and BV– is low when considering only the most abundant phylotypes, but much higher when considering all phylotypes, while, F. Shared structure (Theta index) is very low regardless of subset analyzed.

Further, shared community membership (Jaccard's index) between BV and BV– individuals is quite low when considering only reads present at >1% (0.34, compared to 0.61 in technical replicates), but much higher when considering all reads (0.81), indicating that increased depth reveals a fundamental similarity in composition of vaginal microbiota regardless of BV diagnosis (Figure 9.7, E). In contrast, shared community structure between BV and BV– individuals is very low regardless of whether all reads or only the most abundant reads are analyzed (0.08, compared to 0.94 in technical replicates), demonstrating that BV and BV– individuals differ in abundance rather than presence of specific phylotypes (Figure 9.7, F). Therefore, it can be concluded that deep sequencing has increased power to define fundamental aspects of BV microbiology compared to older techniques.

Summary – Two BV-related phylotypes in phylum Actinobacteria, the *Gardnerella*-like NC070 and *Mobiluncus curtisii*, were observed in significantly fewer HIV-R individuals compared to HIV-N individuals, in support of Hypothesis 8. However, these phylotypes represented only 1% and 0.1% of total pyrosequencing reads respectively, reducing the plausibility of biological significance for this observed difference. Further work should focus on functional aspects of low abundance bacterial populations as they relate to microbial ecology in general and bacterial vaginosis in particular.

Interestingly, HIV+ individuals were more likely than HIV– to have *E. coli* and less likely to have *L. crispatus* detected in vaginal specimens. Consistent with this, two samples from a single individual pre- and post-seroconversion showed that *L. crispatus* was only present pre-seroconversion and *E. coli* was only observed post-seroconversion, although only at very low levels for both phylotypes (>0.1%) This original observation

will require confirmation in a larger study group and with refined techniques. No associations were observed between microbiological differences in HIV+ individuals and Septrin prophylaxis or current ARV use, indicating that this difference, if confirmed, may be due to an altered relationship between host and commensal vaginal microbiota in the context of HIV infection.

Phylogenetic analysis of clone libraries revealed a complementary picture of the vaginal microbiota compard to cultured isolates. A vast improvement in matching of pyrosequencing reads to database sequences was observed when the database was enhanced by the addition of 132 sequences derived from isolates and clone libraries. Deep sequencing revealed an increased number of phylotypes compared to culture and clone libraries. These results may indicate a shift in purpose for clone libraries and isolates, from the current "gold standard" analytical tool in microbial ecology, to an important support role in improving database quality in preparation for deep sequencing.

Deep sequencing feasibility experiments demonstrated that reproducibility was very high between technical replicates of the same sample, but that lowest abundance reads reproduced least reliably. Comparison of results from *cpn*60 UT-based deep sequencing vs. the more well-established 16S target were very favourable. Virtually the same profile of major phylotypes emerged after sequencing of both targets from the same sample. Further, a greater number of phylotypes were identified in the *cpn*60 UT profile, confirming the improved resolution of this target relative to 16S.

Striking associations were observed between BV diagnosis and abundance of several phylotypes, including total *Lactobacillus*, *L. iners*, total Bacteroidetes, and total Clostridiales, reflecting what is known of BV microbiology. In contrast to current

literature, abundance of total *Gardnerella*-like and specific phylotypes were only weakly associated with BV, while abundance of several Proteobacteria phylotypes, particularly *Acidovorax facilis*, were associated with being BV–.

Several distinct molecular microbiological patterns were identified in individual deep sequencing profiles, corresponding roughly to BV–, BVI and BV+ categories as defined by Nugent score. BV– samples are dominated to very high levels by *Lactobacillus* phylotypes, while "intermediate" samples (defined in molecular terms) have reduced *Lactobacillus* and are dominated by *Gardnerella*-like phylotypes. In contrast, BV+ individuals have very low levels of *Lactobacillius* and co-dominant *Gardnerella*, Clostridiales and Bacteroidetes. Its intermediate position may explain why *Gardnerella* is only weakly associated with BV diagnosis, in contrast to current concepts of BV.

Ecological analysis of deep sequencing data provides a formal demonstration that phylotypes characteristic of BV microbiology are present at low levels in BV– individuals and vice versa, with equivalent species density and a high level of shared community membership observed in BV vs. BV– samples.

8. Enumeration of bacterial genomes for five *cpn*60 phylotypes by qPCR

qPCR is well-established for the quantification of specific bacterial targets in mixed microbial communities, and has frequently been applied to quantify specific targets of the vaginal microbiota (51, 53, 228). The goal of this chapter is to determine if HIV-R individuals have different absolute abundances of *L. iners*, *L. crispatus* and *L. gasseri* as well as the BV-related phylotypes *G. vaginalis* and *A. vaginae*, compared to HIV-N.

Optimization – Standard curves were prepared from serial dilutions of UT plasmids allowing for quantification of total genomes per qPCR reaction. These results were adjusted according for volume (genomes/ml) and sample weight (genomes/g).

First, pooled vaginal sample was divided into three aliquots and spiked with tenfold dilutions of overnight *B. animalis* culture. DNA was extracted using three different methods and spiked samples were measured by qPCR using *B. animalis*-specific primers (Figure 10.1). For DNA extracted using Instagene Matrix, all three dilutions were quantified precisely in relation to the standard curve and found to correspond exactly to flow cytometric measurements of the pure culture dilutions used to spike the pooled sample, indicating the precision of the qPCR technique (Figure 10.1, inset).

This experiment resulted in choosing Instagene Matrix to extract DNA from whole vaginal samples, since good representation of *Bifidobacterium* sp. is often limited by DNA extraction bias (J. Hill, personal communication).

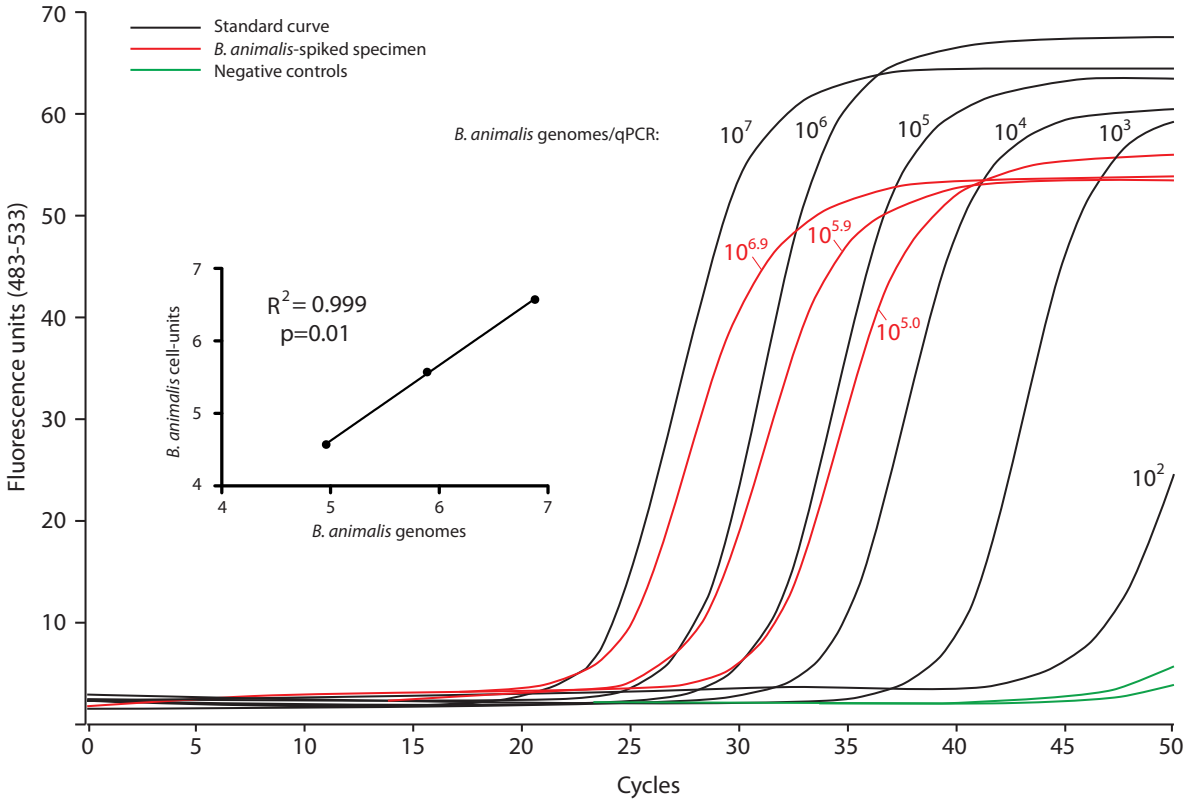

Figure 10.1 – Quantifying bacterial genomes by *cpn*60 UT-based qPCR

Estimating *Bifidobacterium animalis* genomes in vaginal samples spiked with three dilutions of cultured *B. animalis*. Standards derived from cloned *cpn*60 UT DNA for *B. animalis*. Precise quantification of *B. animalis* spiked into identical vaginal samples at three dilutions. Inset: Correlation between *B. animalis* genomes as quantified by qPCR and *B. animalis* cell-units as quantified using flow cytometry, indicating precise quantification by either method.

Enumeration of five bacterial targets by qPCR – Primers for specific amplification of *L. iners* (LI), *L. crispatus* (LC), *L. gasseri* (LG), *G. vaginalis* (GV) and *A. vaginae* (AV), were selected from the literature or designed and optimized by T. Dumonceaux (53). Quantitative results for specific bacterial targets are presented as $\log_{10}$ genomes per qPCR reaction (a value of 5.4 is equal to $10^{5.4}$ or ~250,000). All individuals had *G. vaginalis* detected, and almost all had *L. iners* (Table 10.1). Over 40% had *L. crispatus* and *L. gasseri*, and 25% had *A. vaginae*.

Comparison of HIV-R and HIV-N – No significant differences were observed in absolute levels or likelihood of being colonized by any of the organisms targetted by qPCR, disproving Hypotheses 4, 7 and 8 (Table 10.1). Some non-significant trends may be worth addressing in new studies with a larger number of individuals. qPCR assays should be developed to formally test whether NC070 and *Mobiluncus curtisii*, two BV-related phylotypes shown to be associated with HIV-R individuals, are in fact different between HIV-R and HIV-N individuals.

Comparison by serostatus – Consistent with observations by deep sequencing, *L. crispatus* was observed at low levels in only 2/10 HIV+ individuals, compared to much higher levels in 15/30 HIV– individuals. However, neither presence nor abundance of *L. crispatus* was significantly different between the two groups ($p=0.07$ by Mann Whitney test for abundance and $p=0.15$ by Fisher's exact test for presence). Further testing in a larger number of individuals may confirm a significant difference in L. crispatus between HIV+ and HIV– individuals. A similar assay should also be developed in order to test whether HIV+ individuals are in fact more likely to have *E. coli* compared to HIV– individuals.

Table 10.1 – Enumeration of bacterial genomes and comparison of HIV-R/HIV-N

	Pyro Subset (N=40)	**HIV-R** (N=20)	**HIV-N** (N=10)	**Diff.**[2] MG/%C[1]
L. iners				
Detected	33 (83%)	15 (75%)	9 (90%)	--/--
Median genomes	5.5	5.8	5.1	
Range	0.8-7.2	2.7-7.2	0.8-6.8	
L. crispatus				
Detected	17 (43%)	11 (55%)	4 (40%)	--/--
Median genomes	3.3	3.4	4.4	
Range	1.3-7.6	1.3-7.1	2.3-7.6	
L. gasseri				
Detected	18 (45%)	11 (55%)	3 (30%)	--/--
Median genomes	3.3	3.4	2.4	
Range	2.1-4.1	2.1-4.1	2.3-3.2	
	N=18	*N=10*	*N=5*	
G. vaginalis				
Detected	18 (100%)	10 (100%)	5 (100%)	--/--
Median genomes	3.8	3.3	4.6	
Range	1.4-7.6	1.4-6.1	2.1-7.6	
A. vaginae				
Detected	10 (25%)	7 (70%)	3 (60%)	--/--
Median genomes	5.0	4.0	7.2	
Range	1.2-7.4	1.2-6.2	5.6-7.4	

[1] MG: Median genomes detected per qPCR reaction, %C: % of group colonized, [2] Difference between HIV-R and HIV-N, as determined using Mann Whitney test (to compare median reads in those colonized) and Fisher's exact test (to compare % colonized in each group).

Correlations between qPCR, deep sequencing and culture – To assess deep sequencing as a quantitative or semi-quantitative tool, qPCR measurements were compared to deep sequencing read frequencies for several individuals (Table 10.2). Interestingly, *L. crispatus* genomes as measured by qPCR correlated most strongly with *L. crispatus* read frequency (Figure 10.2, A), as well as with total *Lactobacillus* reads and N6 reads. *L. crispatus* genomes also correlated inversely with a number of BV-related phylotypes.

L. gasseri genomes were observed at a much lower level and this phylotype was not detected at all by deep sequencing, indicating that deep sequencing may be less sensitive than qPCR for detecting this phylotype (Figure 10.2, B). Interestingly, *L. gasseri* genomes correlated positively with α-/β-Proteobacteria reads and inversely with all Clostridiales and all Bacteroidetes reads.

L. iners genomes for qPCR correlated most strongly with *L. iners* reads (Figure 10.2, C) and also correlated positively with some *G. vaginalis* and Clostridiales phylotypes. *G. vaginalis* genomes were not associated with reads for all *G. vaginalis*-like phylotypes or for the ATCC strain, but were most strongly correlated with NC139, indicating that the qPCR primers may target *G. vaginalis* sequences most closely related to this clone (Figure 10.2, D). Interestingly, *G. vaginalis* genomes were inversely correlated with another *G. vaginalis*-like sequence, NC088. In contrast to other targets, *A. vaginae* genomes did not correlate with deep sequencing reads for this phylotype (not shown).

These findings indicate that deep sequencing read frequencies correlate with quantitative measurements for some phylotypes but not others. Further work will be required to determine methodological biases and establish the quantitative basis of deep sequencing data.

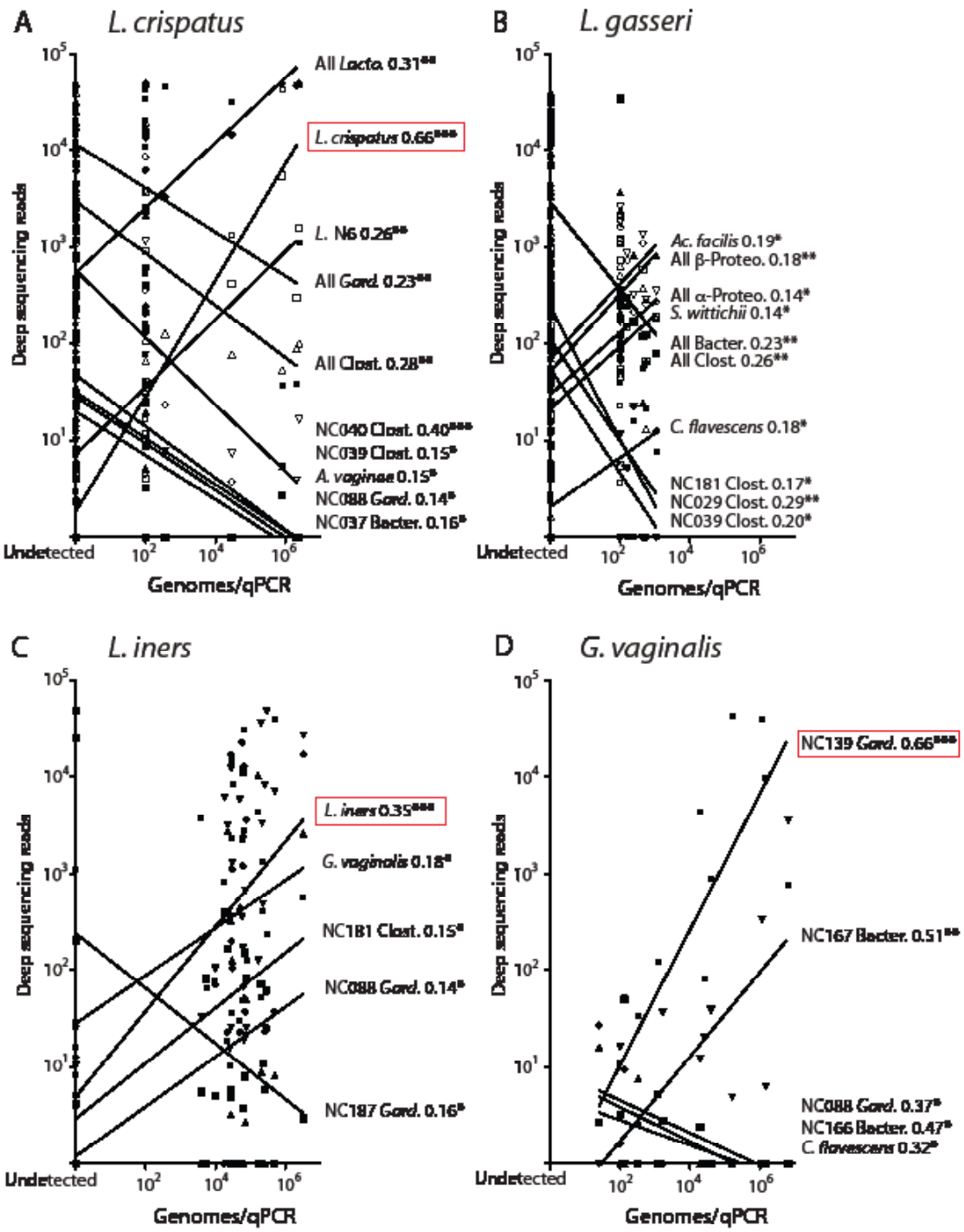

Figure 10.2 – Correlation between qPCR and deep sequencing reads

Plots showing linear regression between total genomes per sample as measured by qPCR for four targets (x axis) compared to total deep sequencing reads for selected phylotypes (y axis). Value following phylotype label indicates regression coefficient (R^2). Asterisks indicate significance level (*p<0.05, **p<0.01, ***p<0.001). A. *L. crispatus*, B. *L. gasseri*, C. *L. iners*, D. *G. vaginalis*.

Individuals with any cultured *L. crispatus* isolates or *L. gasseri* isolates had higher median genomes for these phylotypes by qPCR compared to those without (0 vs. 1.7 million, p=0.0002 for *L. crispatus*, and 0 vs. 3,462, p=0.002 for *L. gasseri*, both by Mann Whitney test, data not shown). In contrast, having any *G. vaginalis* isolates did not correspond to measures of this phylotype by qPCR. These data indicate that qPCR measures for *L. crispatus* correspond very well to L. crispatus as measured by culture-based isolation and deep sequencing.

Future prospects – While qPCR coupled with *cpn*60 UT-targetted primers is extremely powerful to detect and quantify specific phylotypes in natural samples, current protocols are limited to detecting one or two phylotypes per reaction. In contrast, bead-coupled flow cytometry assays using the Luminex platform can detect and provide semi-quantitative information for a very large number of phylotypes in a single run (53). A comparison of qPCR and Luminex assay (protocol designed and implemented by T. Dumonceaux) has demonstrated an excellent correspondance between profile of vaginal microbiota generated by either method (not shown) (53). Applying information about specific vaginal phylotypes observed in this study to generate a panel of Luminex probes tailored to the ML cohort has great potential to improve laboratory diagnosis of BV and track fluctuations of microbial populations of interest.

Summary – No significant differences were observed in terms of absolute abundance of 5 targetted phylotypes between HIV-R vs. HIV-N individuals, therefore Hypotheses 4, 7 and 8 are not supported. Correlation between absolute abundance of specific phylotypes as determined by qPCR and relative abundance as determined by deep sequencing was observed for some but not all phylotypes. Individuals colonized with

specific phylotypes also had higher genomes for the same phylotypes as measured by qPCR. These associations indicate that deep pyrosequencing provides quantitative information about bacterial concentration, at least for some phylotypes. Further work will be required to determine absolute quantitative discrimination of deep sequencing techniques.

D. DISCUSSION

In this thesis, a model describing potential microbiological influences on HIV resistance was proposed and several specific hypotheses arising from this model were evaluated using a combination of established and innovative techniques. The primary goal of this study was to determine if HIV-R individuals differ from HIV-N individuals when considering various measures of the vaginal microbiota. Several secondary goals were also addressed, including an understanding of effects of HIV infection on vaginal microbiology, defining BV in molecular terms and providing insight into aspects of vaginal health and prevention of STI/HIV in vulnerable groups such as CSW.

Vaginal microbiota and HIV resistance – This is the first study to address BV and *Lactobacillus* colonization specifically as possible factors in HIV resistance. Compared to other HIV-N individuals, those classified as HIV-R in the Majengo cohort are just as likely to be diagnosed with BV and have similar vaginal pH in a cross-sectional study. This finding was confirmed in a retrospective, longitudinal analysis of BV diagnosis over nearly 4 years in over 1,000 individuals. Therefore, BV is not a direct mechanism of HIV resistance in this cohort.

No differences in overall bacterial concentrations as determined by flow cytometry were observed between the two groups. HIV-R individuals did not have higher levels of culturable *Lactobacillus* as determined using selective culture media. Phenotypic characterization of 576 cultured isolates in terms of acid and H_2O_2 production revealed no differences in these parameters between the two groups. HIV-R individuals were not more likely to be colonized by culturable *Lactobacillus* isolates as determined by species identification using a novel combination of molecular targets (ISR and *cpn*60 UT). By

ultra-deep molecular characterization, no differences in richness, diversity or evenness were observed in HIV-R individuals and only two BV-related organisms, present at very low levels, were differently distributed between the two groups. Therefore, a determinative role for BV or phylogenetic and phenotypic characteristics of *Lactobacillus* organisms in HIV resistance was disproven by the present study.

A major limitation of this study is that the most in-depth analyses to address the study hypotheses were carried out in small groups at a single timepoint, including culture-based analysis (32 individuals each in HIV-R and HIV-N groups) and deep sequencing analysis (16 individuals per group). The striking idiosyncracy of microbial communities at the species level and in terms of presence/abundance of specific phylotypes creates several problems in terms of establishing significance of differences between groups.

Further work could expand on the number of individuals assessed, or improve the scope of microbiological parameters, including an expanded range of culture media, or combining species-specific probes with flow cytometry, etc. A critical next step should be an assessment of dynamics of microbial communities through assessment of individuals at multiple timepoints. Since functional differences between even very closely related strains are apparent, increased functional analysis of vaginal strains and consortia especially in terms of communicative processes with mucosal cells may also reveal differences between HIV-R and other groups.

Since BV is so common in the ML cohort, perhaps the way a person responds to its occurrence, or the cycle of occurrences that lead from one state to another may be a previously unrecognized factor contributing to HIV resistance. BV is known to induce a pro-inflammatory environment and is known to create vulnerability to HIV, so the ability

of an individual to modulate some aspect of the inflammatory process may determine her ability to block viral infection. Almost nothing is known about the transition from one microbiological state to another, but it seems likely that some aspect of this transition, perhaps inducing increased inflammation at a critical timepoint in HIV-N individuals, but not HIV-R individuals, might help to explain resistance.

Therefore, longitudinal studies examining dynamics of vaginal microbiota and mucosal responses over time will be necessary in order to definitively address the potential impact of vaginal microbiology on HIV resistance in the Majengo cohort. Application of qPCR and Luminex techniques developed in the context of this study have great potential to facilitate effective longitudinal analysis (53)

Effect of serostatus on vaginal microbiota – The present study is only the fifth to show a convincing association between BV and HIV infection using longitudinal data (130, 144, 200, 217). Due to the small numbers of individuals seroconverting during the study period, a causal relationship between BV and HIV was not established, however a recent meta-analysis on the topic points to a need for further prospective studies to enhance understanding of a causal role for BV in HIV infection (13). Therefore, the current study is likely to be a valuable addition to the literature.

HIV+ individuals with BV were found to have increased levels of CD8+ T cells in peripheral blood. It is difficult to imagine how peripheral cells could be influenced by bacteriological changes in the FGT, a relatively small mucosal compartment. One possibility is that microbiological changes in the gut occur in parallel with changes in vaginal microbiota associated with BV, triggering the overall pattern of CD8 expression on PBMC. Studies to profile oral-fecal bacteria simultaneously with vaginal bacteria could

address this hypothesis. Another related possibility is that BV might be a microbiological "symptom" of immunological events. For example, an overall inflammatory process resulting in increased CD8 levels triggers loss of *Lactobacillus* predominance by an unknown mechanism. Since almost nothing is known about the factors triggering BV or the transition from one state to another, it seems plausible that some immune mechanism involving changes in CD8 levels could be the cause of BV. Interestingly, those receiving both ARVs and Septrin prophylaxis have reduced CD8+ T cells, indicating an opposite effect on systemic immune cell profiles.

One of the most intriguing findings of the culture-independent studies was that HIV+ individuals were much more likely to have detectable *E. coli* and less likely to have detectable *L. crispatus* compared to HIV– individuals. Deep sequencing profiles in HIV+ individuals are suggestive of a dysregulated microbiota in at least some individuals. No differences in colonization by *L. crispatus* or *E. coli* was observed in those receiving ARV or Septrin prophylaxis, therefore, the possibility that this may reflect an altered crosstalk between the mucosal immune system and commensal microbiota in the context of HIV-induced depletion of Th cell populations will require further investigation.

Defining vaginal microbiota in Nairobi CSW – Only a few studies have characterized vaginal microbiology in African populations (9, 133, 136), and no previous studies have examined vaginal microbiology in CSW. Despite high frequency of sexual intercourse with multiple partners in this group, the range of microbiological outcomes observed in this group are consistent with what has been observed in other groups of women around the world (85, 201, 228). By combining multiple approaches to defining vaginal microbiology in this group of individuals, this study has generated an

unprecedented amount of data regarding detailed microbiological patterns associated with BV, greatly expanding upon current knowledge of this "enigmatic" condition, particularly in the African context (Figure 11.1). Although many aspects of the model presented here are hypothetical, it represents an attempt to synthesize information about BV gained in this study into a coherent, biologically relevant "story" about dynamics of vaginal microbiota.

Four major microbiological patterns were observed by microscopy of Gram-stained swabs in this study, consistent with numerous other studies (98, 146, 205). BV– samples are defined by predominance of Gram+ rods (*Lactobacillus* morphotypes) while BV+ samples are dominated by Gram– rods and coccobacilli (BV morphotypes). An intermediate phenotype (BVI) is often described that appears to be a transition phase between BV– and BV+ microbiota, ie. a gradual replacement of *Lactobacillus* morphotypes by BV morphotypes or vice versa. There has been very little description of the dynamics of the transition from one state to the other and hence many questions remain regarding the biological significance of transition microbiota and the sequence of events causing or curing BV. Until these questions are answered, hope for accurate diagnosis and treatment of this condition will remain elusive.

A fourth pattern was consistently observed in samples from the Majengo cohort – slides with no visible bacteria on Gram stain, called BV0 in this study. Under Nugent's scoring system, these slides are classified as Intermediate, while in the Hay-Ison scheme, they are classified simply as "0". A large proportion of individuals in the Nairobi study had slides with BV0 scores, indicating that concentrations of vaginal bacteria have fallen below the limit of detection for Gram stain. Several observations suggest that BV0 may hypothetically be a "wash-out" phase marking the transition between BV+ and BV– states.

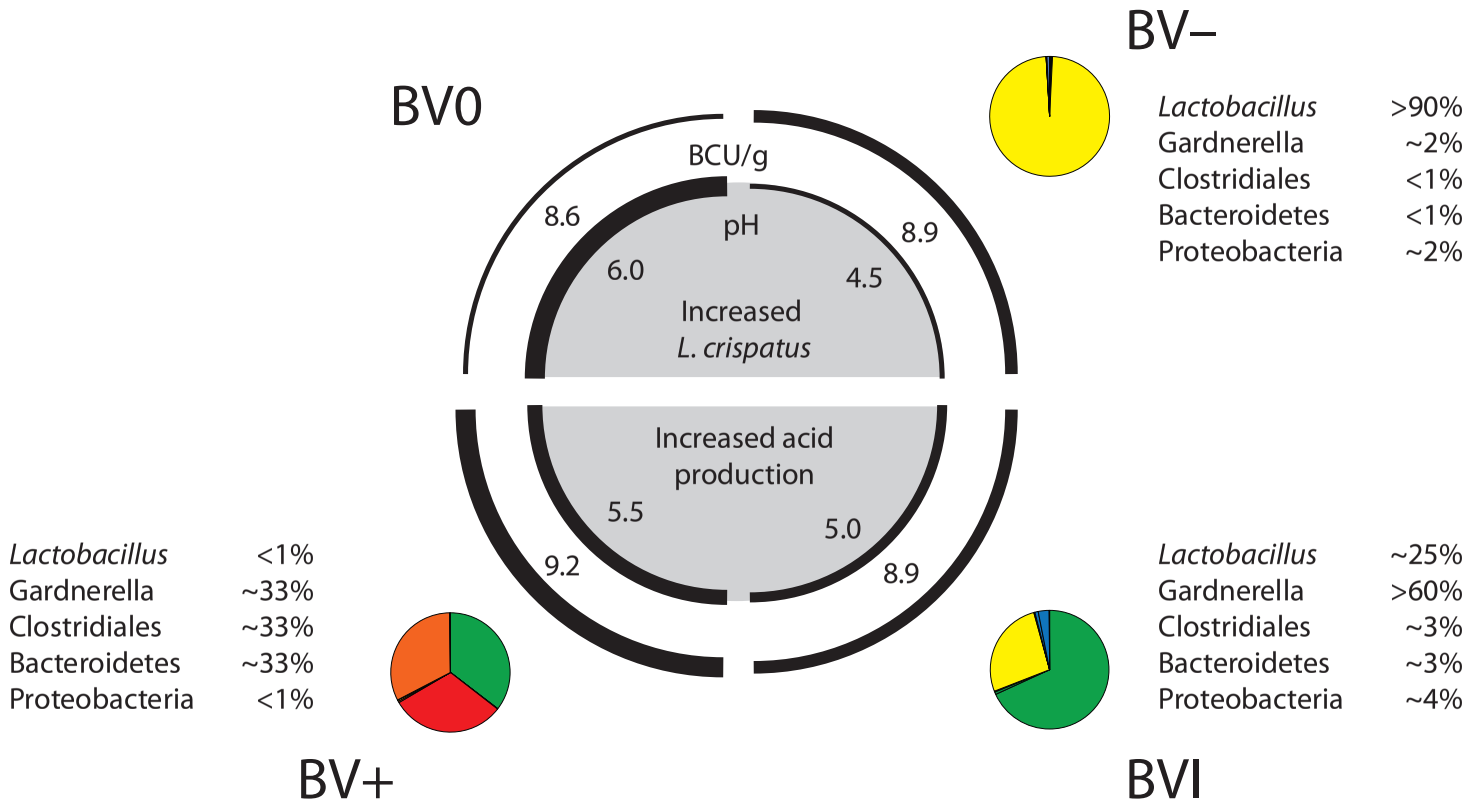

Figure 11.1 – Defining vaginal microbiota in the Majengo cohort
Synthesis of information about vaginal microbiota collected by microscopy, flow cytometry, culture-based analysis and deep sequencing. Four categories of vaginal microbiota are defined based on microscopy. BVI is proposed as the transition from BV– to BV+, while BV0 is proposed as the transition from BV+ to BV–. Vaginal pH (inner ring) is highest in BV0 and lowest in BV–. Bacterial concentration (BCU/g, outer ring) is highest in BV+ and lowest in BV0, indicating bacterial "wash-out". Increased acid-producing bacteria are observed in BVI/BV+ while increased *L. crispatus* is observed in BV0/BV–. By deep sequencing in HIV– individuals, BV– samples are characterized by dominant *Lactobacillus*, BVI by dominant *Gardnerella*, and BV+ by co-dominant *Gardnerella*, Bacteroidetes and Clostridiales.

As expected, a strong correspondance between BV diagnosis and vaginal pH was observed, further validating BV diagnosis results in this study. Interestingly, BVI individuals have median pH values that are intermediate between BV– and BV+ individuals, while those with BV0 scores had the highest median vaginal pH (~6). These observations indicate that BVI and BV0 categories may be biologically relevant with implications for vaginal health and vulnerability to exogenous pathogens.

This is the first study to show an association between BV diagnosis and sample weight. Although the amount of sample collected is probably stochastic to a certain degree, it seems plausible that denser mucus would yield a heavier sample. Many studies assess sample weight for the purpose of adjusting bacterial counts (132, 150), but none assessed how sample weight itself was correlated with other factors. Bacteria associated with BV are known to produce enzymes such as sialidise that lyse mucin and may lead to thinning of secretions (117, 167). In the pilot study, individuals with BV had decreased sample weight, but in Nairobi, individuals with BVI samples had reduced sample weight compared to both BV– and BV+. Reduced sample weight was also associated with increased pH. Therefore, both studies confirm the importance of BV diagnosis and pH on sample weight, indicating changes in the mucus layer that may help to explain why BV contributes to HIV vulnerability.

Flow cytometry – This is the first report of flow cytometric analysis of commensal bacterial populations in vaginal samples (175). "Total" bacterial cell-units (BCU) are rapidly quantified using our approach and correlate significantly with sample weight. Despite increasing use of flow cytometry in the detection and identification of bacteria in clinical samples (6), enumeration of total bacteria in complex samples is more common

among environmental and dairy microbiologists (78, 82). Flow cytometry has recently been used to enumerate total bacteria and sub-populations in the animal intestine (43) and in human feces (215). Enumeration of total bacteria and white blood cells in urine by flow cytometry has recently been evaluated as a diagnostic tool for urinary tract infection (69).

Estimates of total culturable bacteria in the vagina are quite similar across studies. For example, one study found an average total bacterial load of log10 8.1 CFU/g (range: 5-11.3) in 118 samples (21), while another study published the same year found an average log10 8.2 CFU/g (range: 5.5-11) in 102 samples (223). More recently, Onderdonk found an average concentration of 8.3 log10 CFU/g (SD=1.47) for anaerobic bacteria and 7.2 log10 CFU/g (SD=2.21) for aerobic bacteria in 229 samples (150). In this study, quantitation of bacteria by flow cytometry revealed a mean log10 cell count of 8.04 (SD=0.52, range 6.95-9.04) per gram in the Winnipeg study, and a mean of 8.90 (SD=0.60, range 7.26-10.16) per gram sample in Nairobi. Therefore, flow cytometry provides similar estimates of bacterial concentrations in vaginal fluids compared to well-established culture-based techniques with a fraction of the time and effort.

Many authors refer to an increase in total bacterial concentrations of 2 to 3 logs in bacterial vaginosis (63, 117, 164, 199). This increase should be observable by flow cytometry. As expected, Nairobi individuals with BV had reduced bacterial concentrations compared to BV– individuals, while the opposite was observed in the pilot study. Since previously frozen samples were used in the pilot study and fresh samples in the Nairobi study, this difference is likely to be due to effects of freeze-thaw on cytometric counts. The observation that previously frozen samples from Nairobi had higher bacterial concentrations in BV– compared to BV samples confirms this hypothesis.

BV+ individuals had the highest bacterial concentrations, while BV0 individuals had the lowest (Figure 11.1). If BV0 is indeed the transition phase from BV+ to BV– microbiology in this group, then this finding is consistent with the "wash-out" hypothesis. Surprisingly, increased bacterial concentration was associated with lower vaginal pH. This apparent contradiction is clarified by considering BV diagnosis and vaginal pH together. Within BV– and BV0 samples, bacterial concentrations are decreased at higher pH, but in BVI/BV+ samples these levels are about the same regardless of pH.

Although longitudinal studies will be required to fully address this question, these observations may reflect dynamics of the transition from BV– to BV+ states and vice versa. As pH rises in BV– individuals, for example during alkalinization after sperm deposition, the overall level of *Lactobacillus* bacteria may fall, resulting in lower bacterial cell counts at higher pH levels. Subsequent expansion of BV-related bacteria to high levels explains increased bacterial cell counts associated with BVI and BV+ samples regardless of pH level. After "wash-out" hypothesized to occur at the end of the BV+ phase, pH is highest and bacterial cell counts lowest in BV0, consistent with observed association of higher pH and lower cell counts in this category. As *Lactobacillus* and other BV– populations begin to expand, pH falls and cell counts increase.

Interestingly, only two previous studies have directly addressed the question of increased bacterial concentrations in BV, using quantitative anaerobic culture methods. A study comparing 22 BV patients and 22 normal controls found "quantitative polymicrobial changes", including slightly reduced mean levels of aerobic bacteria and a 1000-fold increase in anaerobic bacteria (132). In contrast, a second study found no differences in mean levels of total bacteria or anaerobic bacteria in a group of 14 women with non-

specific vaginitis compared to 20 normal controls (223). Further work will be required to determine overall bacterial concentrations at various stages of the BV "cycle" and the potential impact of this factor on vaginal inflammation, TLR stimulation and vulnerability to exogenous pathogens.

Culture-based studies – As expected, BV– individuals in this study had higher levels of growth and H_2O_2 production on *Lactobacillus*-selective media. Paradoxically, they were also less likely to have samples producing acid on mannose or have strong acid-producing isolates, double (acid/H_2O_2) positive isolates or strongly double positive ("double-strong") isolates. Although this observation may not be reflective of dynamics *in vivo*, one possible explanation is that increased acid production by vaginal isolates is a mechanism for how *Lactobacillus* bacteria become dominant again, competing with BV-related bacteria by up-regulating excretion of antagonistic molecules.

After identifying isolates based on ISR profile and cpn60 UT sequencing, it was determined that the *L. salivarius*-like phylotype N35 was found at higher levels in those with BV. These isolates all had double-strong isolates, indicating that these organisms may play a role in the shift back to BV– status. Similarly, *L. salivarius* and *Streptococcus* isolates were more likely to be double strong and were associated with BV samples. No other studies have reported observing a *Lactobacillus* phylotype associated with BV rather than BV– samples.

Interestingly, strong acid producers and double-strong isolates were also associated with an increased number of CSW clients per day. This would make sense in terms of the overall response of the vaginal microbiota to allogenic stimulation, alkalinization of vaginal environment by sperm or effects of exposure to condoms.

Lactobacillus and related phylotypes in the vagina may react to these physicochemical disturbances by upregulating acid/H_2O_2 production, just as they may have a greater propensity to produce acid in a BV/intermediate environment. Although further studies will be required to confirm these findings, it would appear from these results that bacteria can be induced to upregulate these metabolites in response to environmental stimuli in order to compete, rather than just passively pumping them out as a byproduct of metabolism. It is interesting that well-known probiotic strains, including two vaginal strains described by G. Reid, also have a "double-strong" phenotype. Perhaps one of their mechanisms of action is the concerted acidification/peroxidification that causes the BV consortium to unravel.

Further studies may indicate some role for this "probiotic" effect in HIV resistance – the phenomenon should be followed for more time in more individuals. Increased H_2O_2 production in women with BV might encourage reduced overall mucosal inflammation and contribute to "immune quiescence". Since reactive oxygen species have been shown to be critical in TGFβ activation and induction of Tregs, perhaps H_2O_2 of bacterial origin at very early timepoints in HIV infection might induce a tolerogenic milieu, suppressing immune activation at the identical point in the BV "cycle" where HIV vulnerability is highest. This hypothesis could be tested in cell culture.

Interestingly, BV0 samples were significantly more likely than BV+ samples to have *L. crispatus* detected by culture and greatly resembled BV– samples in this respect. This observation is consistent with the hypothesis that BV0 samples are a transition phase between BV+ and BV– states. Subsequent to bacterial "wash-out", small *L. crispatus* populations may begin to expand in a newly favourable environment.

Surprisingly, *Lactobacillus* phylotypes associated with BV– samples (L. crispatus, L. jensenii, N6) were also least likely to be acid producers on test media. Although this may reflect *ex vivo* experimental factors rather than *in vivo* dynamics, the hypothesis that BV-associated *Lactobacillus* and *Streptococcus* phylotypes produce acid in order to create conditions for BV– *Lactobacillus* phylotypes could be addressed in longitudinal culture-based studies.

Deep sequencing – This is the first description of the application of high throughput deep sequencing for analysis of amplicons derived from microbial communities using a target other than 16S rDNA. Only two other studies have used deep sequencing technology to describe culture-independent characteristics of the human vaginal microbiota, both using 16S rDNA (194, 196). Both of these studies examined far fewer numbers of reads per individual, meaning that the present study represents the most detailed analysis of vaginal microbiota undertaken to date.

Deep sequencing of *cpn*60 UT amplicons by GS-FLX compared very favorably to the clone library approach as a method of characterizing a complex microbial system. The huge number of reads generated by deep sequencing included essentially all sequences from the library method and greatly increased the number of distinct taxa identified. Deep sequencing of *cpn*60 UT amplicons results in a more highly detailed snapshot of the composition of microbial ecosystems than is feasible using clone libraries, making the detection of lower-abundance organisms possible.

Phylotypes identified were consistent with what would be expected from a human vaginal microbial community (11, 17). Virtually all phylotypes identified in clone libraries were also present in deep sequencing data, and several additional phylotypes present at low

abundance were revealed by deep sequencing. Increased depth was particularly notable in samples from BV– individuals.

Similar taxonomic distributions of 16S rRNA sequences in clone libraries and deep sequencing datasets have been found in samples taken from deep mines (54). Differences in abundance of certain taxa are likely related to different biases introduced by cloning and ligation versus emulsification PCR and bead layering. Taxonomic profile generated in technical replicates was highly reproducible, including the proportions of reads represented at the species level, although the normal range of variation within technical replicates should be further explored in order to establish a reliable cutoff for significance of low-abundance phylotypes.

This study has demonstrated that *cpn*60 UT-based deep sequencing compares favorably to 16S rDNA-based deep sequencing. Taxonomic profile of vaginal microbiota is very similar to that generated by the 16S rRNA molecular target but with a higher level of taxonomic resolution. Differences in presence and abundance of specific taxa between pyro read libraries from 16S and cpn60 targets might be explained by differences in the efficiency with which various species are amplified by the universal primers. Since taxa not observed for one target did not exceed 2% of the total reads for the other target, the taxonomic profiles within the two datasets were essentially in agreement with one another. Protein-encoding genes may provide an increased level of resolution compared to the structural 16S rRNA-encoding gene (13, 28), explaining the greater taxonomic depth observed for cpn60 pyro reads compared to 16S pyro reads.

Deep sequencing allows for a much more nuanced view of phylotype density and diversity to emerge. Three major patterns of vaginal microbiology were observed

corresponding to Nugent score categories, elucidating the biological significance of BVI category (Figure 11.1). Although longitudinal data will be required to confirm this hypothesis, reduced *Lactobacillus* is associated with increased *Gardnerella* initially, while very low levels of *Lactobacillus* are associated with co-dominant *Gardnerella*, Bacteroidetes and Clostridiales phylotypes. The finding that several Proteobacteria phylotypes, especially a phylotype with *cpn*60 UT 90% identical to that of *Acidovorax facilis*, are increased in BV– individuals is very intriguing, challenging current paradigms that define BV– vaginal microbiology as a "*Lactobacillus* monoculture". This concept, which dominates the current BV literature, may be an artefact of sampling methods that only capture the most abundant organisms in a vaginal sample. Several examples in this study indicate that a *Lactobacillus*-dominated sample would look like a monoculture if only phylotypes present at >1% are examined.

In some individuals, these phylotypes are present at >1% of total reads, indicating that clone libraries and other lower-resolution methods should be able to detect them, however these methods may not be able to detect enough of these low abundance organisms to distinguish BV and BV– individuals. An almost exclusive focus in the BV literature on 16S rDNA as phylogenetic target may also bias against detection of Proteobacteria phylotypes, since no Proteobacteria were detected by 16S rDNA analysis of four samples in this study, while several phylotypes in this phylum were detected using *cpn*60 UT.

One of the most intriguing phenomenon observed in this study is the extreme genetic variability observed in *G. vaginalis* by all study methods. In clone libraries, a huge amount of sequence diversity was observed even within single individuals. The

correspondence between ISR profiles and UT sequences and genetic variability in independent molecular targets was most striking in *Gardnerella* isolates. The older microbiology literature discusses *Gardnerella* "biotypes", including the association of some but not all biotypes with BV (159). Similarly, in this study, not all *Gardnerella* phylotypes were associated with BV.

Since *Gardnerella* was dominant in a number of samples with reduced levels of *Lactobacillus* (categorized as "intermediate" in this study), the hypothesis that the diversity observed in this phylotype group is due to evolutionary pressures related to its intermediate position should be explored further.

One of two phylotypes observed to be significantly different in HIV-R individuals is a *Gardnerella* organism, over-represented in "intermediate" libraries, it is conceivable that even a low-abundance phylotype like NC070 could play a role in altering the micro-environment during transition to/from BV, creating increased vulnerability to HIV infection. Since this phylotype was only observed in clone libraries, a cultured isolate derived from stored samples would be required to test this hypothesis in cell culture.

One area of concern is the complete absence of *Mycoplasma* and *Ureaplasma* phylotypes detected by any method used in this study. These genera have been previously shown to be important in BV (138, 206). Interestingly, several strains of *Mycoplasma* are among the very few organisms known to have no *cpn*60 gene (J. Hill, personal communication), indicating a limitation of the *cpn*60-based approach to detect these members of the vaginal microbiota.

Many questions remain: What are the primary triggers of the transition from BV– to BV+ states and vice versa? What is the typical time course of BV onset, duration of

intermediate phase and resolution? Is *Lactobacillus* depletion a cause or consequence of BV? What is the clinical significance of asymptomatic BV? Until longitudinal studies assessing some of the variability in bacterial concentration, function, composition and structure determined in this study are completed, an accurate representation of this dynamic ecology will remain elusive.

Synthesis: Isolates, clones and deep sequencing – Combining phylogenetic information from the three primary experimental approaches undertaken in this study provides what is likely to be the most comprehensive and detailed picture of the vaginal microbiota described to date (Figure 11.2). Overall, an excellent correspondence was observed, with all major groups of phylotypes and many individual phylotypes seen by all three methods. These approaches are also highly complementary, providing unique information about a wide range of phylotypes and the biases inherent in each approach.

Although culture-based studies might be expected to be the least informative of the overall composition and structure of the vaginal microbiota, several *Lactobacillus*, *Streptococcus* and *Staphylococcus* phylotypes were only picked up on selective media and not by culture-independent approaches. This finding may reflect the unique approach used to select isolates, ie. selecting the widest diversity of colony morphotypes on chromogenic media present in each sample, and indicates the importance of culture-based approaches for complete understanding of the phylogenetic structure of bacterial groups of interest.

For example, *L. gasseri* was not observed by deep sequencing although it is a common species in the vagina and well-represented among isolates. This suggests either that the actual quantitative distribution of *L. gasseri* is below the limit of detection for deep sequencing, or that some unknown factor biases its representation by deep sequencing.

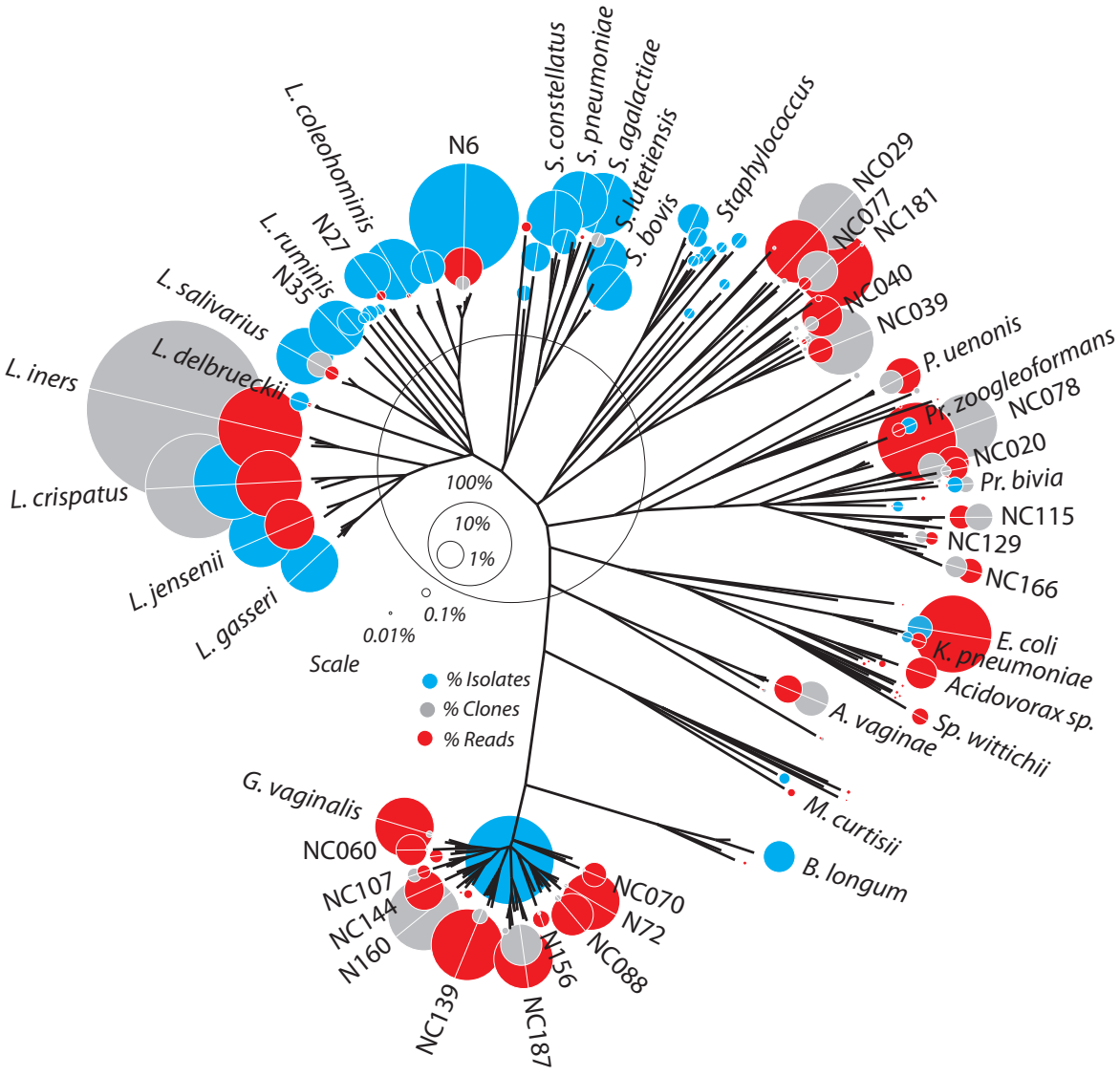

Figure 11.2 – Synthesis: Cultured isolates, clone libraries and deep sequencing

Tree of all phylotypes detected using three experimental approaches, combined with abundance information for each phylotype by each approach. Circles are proportional in size to the percentage of total sampling events for each method (isolates in blue, clones in grey, reads in red). Total area of all circles for each method are equal to the area of the large open circle in the centre. Blue circle at base of *G. vaginalis* group shows all isolates defined as *G. vaginalis* (ie. not assigned to individual branches). N=Nairobi isolate, NC=Nairobi clone, *L.=Lactobacillus, S.=Streptococcus, P.=Porphyromonas, Pr.=Prevotella, E.=Escherichia, K.=Klebsiella, Sp.=Sphingomonas, A.=Atopobium, M.=Mobiluncus, B.=Bifidobacterium, G.=Gardnerella*.

The observation that a small number of deep sequencing reads were detected for other isolated *Lactobacillus* phylotypes indicates that the first possibility is most likely, ie. that the limit of detection for culture on *Lactobacillus*-selective media is actually lower than the limit of detection for deep sequencing.

In contrast, several phylotypes were only seen by culture-independent means, including some known to be non-culturable or difficult to isolate (*L. iners*, *A. vaginae*). Clostridiales and α-/β-Proteobacteria phylotypes were only detected by culture-independent means, although these may have been detected by culture if different media had been used.

Interestingly, by deep sequencing, there is almost full coverage of phylotypes observed by both other methods, indicating that this approach may be less biased. All phylotypes represented by large numbers of deep sequencing reads were seen by at least one of the other approaches, again demonstrating the complementarity of these approaches. In contrast, several phylotypes represented by very small numbers of deep sequencing reads were only seen by this method. For example, besides a few *E. coli* and *K. pneumoniae* isolates, all Proteobacteria phylotypes were only detected by deep sequencing. Some of these, like *Acidovorax* and *Sphingomonas*, are actually present at quite a high percentage of total reads and theoretically should have been picked up in clone libraries, indicating that they are either not present in the actual samples used for clone libraries or that the cloning process is otherwise biased against these phylotypes.

The highly detailed species/strain-level information generated in this study is due to the very highly informative phylogenetic *cpn*60 UT. Other 16S rDNA-based deep sequencing studies present mostly genus-level information, providing a less coherent

picture of vaginal microbiota composition (194, 196). This is due to the clearly less informative nature of rDNA sequences, especially when considering the short reads generated by deep sequencing. This study demonstrates the superiority of cpn60 UT-based analysis in the context of short deep sequencing reads. Since average next-generation deep sequencing reads (Titanium) approach 500bp, covering nearly the entire length of the cpn60 UT, this is likely to be the ideal target for further deep sequencing studies.

This method of simultaneously visualizing phylogeny (composition) and abundance (structure) of a microbial community using multiple experimental approaches has not been previously described, however it represents a powerful approach to summarize very large amounts of information. Further development of this approach (ie. software development to allow automated generation of "abundance trees") will facilitate direct visual comparison of individuals, groups and techniques on a wider scale, and may represent an important development for microbial phylogenetics.

Vaginal health and HIV prevention – Interestingly, the proportion of individuals diagnosed with BV has steadily fallen since analysis of this parameter was introduced in 2004. Given the well-established association between BV and HIV in a number of prospective studies in sub-Saharan Africa (130, 144, 200, 217), observed reductions in BV may have significant implications for HIV incidence in this group. In fact, HIV incidence has declined steadily in the Majengo cohort between 1985 and 2005, concomitantly with decreases in gonorrhea incidence and increases in reported condom use (110).

This is the first report of declining BV incidence concomitant with reduced HIV infection in a high-risk group. These positive changes may be associated with an improvement in overall vaginal health in the cohort associated with better treatment for

STI, increased condom use, or other factors. A recent randomized control trial focussed on improvement of vaginal health showed reduced BV, increased presence of *Lactobacillus* and reduction in trichomoniasis in a cohort of Kenyan CSW (135).

Based on the results of this study, vaginal microbiology is predicted to influence susceptibility of mucosal surfaces to HIV infection via a number of direct and indirect mechanisms (Figure 11.3). This study has focussed on defining relatively few aspects of the microbiological environment in the female genital tract for a specific group of individuals, revealing extensive variability between individuals and groups related to BV diagnosis, overall bacterial concentrations, production of acid and H_2O_2, phylogenetic composition and structure.

The significance of this variability in terms of overall vaginal health, mucosal integrity, immune activation levels and susceptibility to viral infections remains to be determined in future studies. Clearly, however, variability in microbiological parameters determined in this study have until now largely been excluded from models of HIV transmission and pathogenesis.

One area of potential application for the results of this study is in the development of HIV microbicides, defined as topical agents applied vaginally to block HIV transmission (137, 143). Two potential microbicide candidates undergoing Phase III clinical trials (nonoxynol-9 and cellulose sulfate) have both been found to increase HIV transmission in those using these compounds intravaginally compared to those using placebo, likely by causing changes in mucosal integrity or inflammation (137, 143). However, few studies have comprehensively assessed the impact of microbicides on vaginal microbiota, examining only very limited culture-based parameters (187).

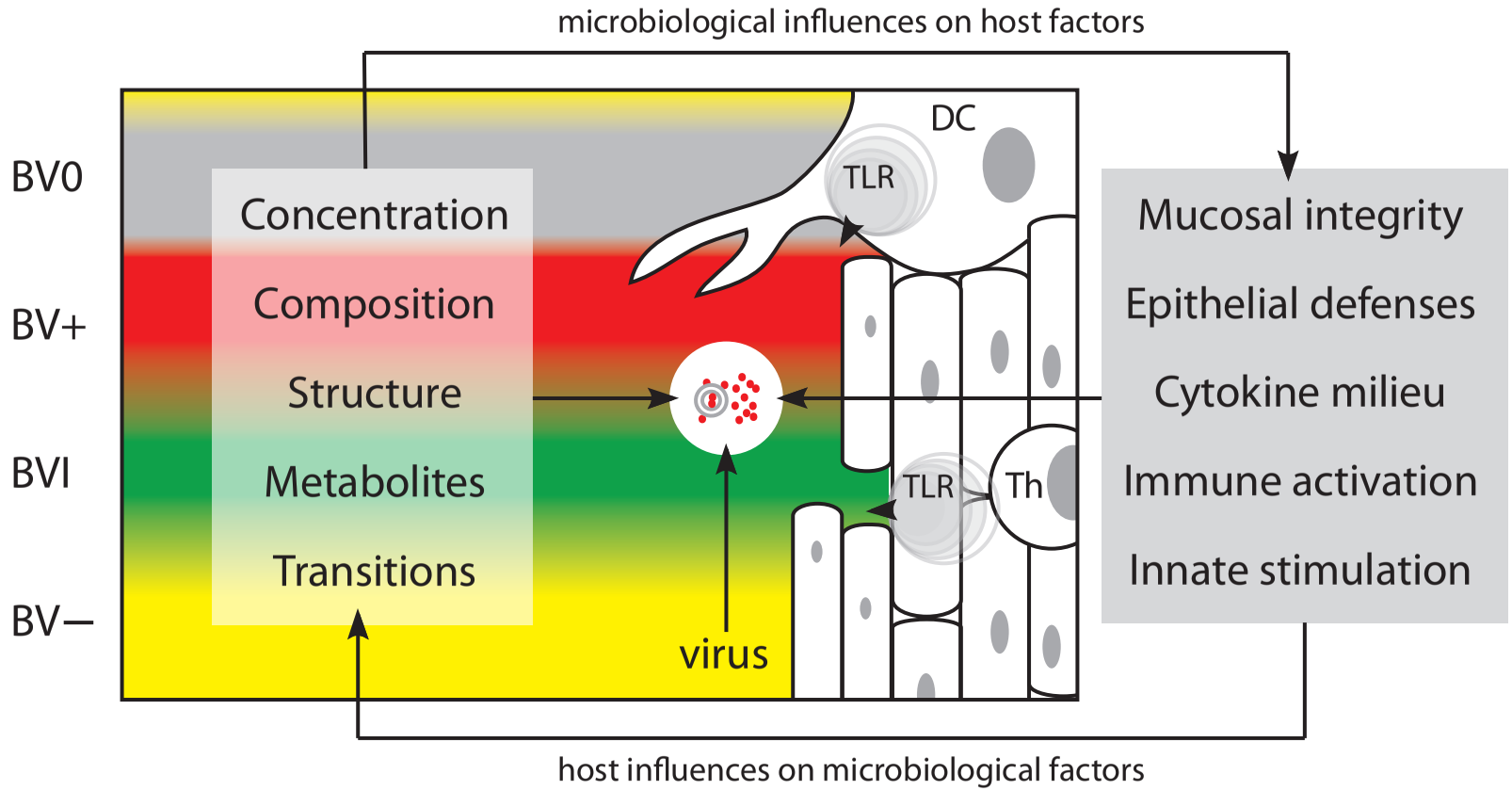

Figure 11.3 – An agenda for future HIV prevention-focussed research

HIV is shown at the interface of host and microbiota. This study has focussed on defining many aspects of vaginal microbiology, including the biological significance of different BV categories (BV–, BVI, BV+, BV0). Future work should focus on an understanding of how these factors influence and are influenced by host factors. The interaction of microbiological and host factors is predicted to determine whether HIV crosses the mucosal barrier and initiates a productive infection.

Therefore, the phenotypic and ultra-deep phylogenetic techniques applied in the current study may be extremely useful in assessing the impact of candidate microbicides on composition, structure and function of vaginal microbial communities. Further, an assessment of vaginal microbiota may help to explain why these recent trials failed. A concerted effort to generate detailed microbiological profiles and link this information to mucosal parameters, as described in Figure 11.3, may be a critically important component of upcoming clinical trials for "second-generation" microbicides as potentially effective tools for HIV prevention (137, 143).

E. CONCLUSION

In-depth microbiological characterization of vaginal swab specimens undertaken to address study hypotheses has provided a vast amount of information about vaginal microbiota in a group of African CSW. While the work presented in this thesis effectively disproved our original hypothesis, demonstrating no obvious biological role for bacterial vaginosis and *Lactobacillus* colonization in HIV resistance, the resulting data represent what is arguably the most in-depth analysis of vaginal microbiota undertaken to date. Using multiple techniques, including many previously unreported approaches, a coherent picture of the vaginal microbiota in this group has emerged.

By defining the "enigmatic" micro-ecological state of BV, this study has elucidated the microbiological context in which HIV transmission does or does not occur in highly HIV-exposed CSW much more clearly than in any previous studies. This knowledge is likely to be critical to a full understanding of the biology of the female genital tract and ongoing HIV prevention efforts worldwide.

This thesis began with an extensive description of the earliest events in HIV infection and the immunology of HIV pathogenesis. Resistance and susceptibility to HIV infection and disease progression can be seen as a continuum, with HIV-R individuals at one end and rapid progressors on the other. In natural infections, the best predictor of immunodeficiency and disease is not lentiviral replication within the host but T-cell activation and activation-induced cell death (AICD) leading to depletion of critical cell populations. Early immunosuppressive signals and upregulation of T-regulatory cells within 24 hours of infection are characteristic of non-pathogenic infections, despite

subsequent intense viral replication and initial depletion of CD4 in blood and lymphoid tissue in both pathogenic and non-pathogenic infections, while CSW who resist HIV infection entirely for long periods appear to have reduced levels of T cell activation and increased anti-inflammatory molecules in vaginal secretions. These findings indicate that modulating immune activation, rather than directly interfering with virus, may be an alternative strategy for stopping HIV infection or mitigating pathogenesis (18).

This thesis concluded with an in-depth microecological study of the function, composition and structure of vaginal microbiota, many aspects of which were previously undefined. These parameters do not predict HIV-R status in the Majengo cohort, however the likelihood that vaginal microbiota directly impacts protection against HIV transmission and pathogenesis has only been reinforced by the results of study. Whether or not vaginal bacteria might be manipulated to reduce mucosal immune activation, improve vaginal health and reduce susceptibility to infection, or if such a goal is desirable or even possible, is the subject of future studies.

REFERENCES

1. **Acosta-Rodriguez, E. V., G. Napolitani, A. Lanzavecchia, and F. Sallusto.** 2007. Interleukins 1beta and 6 but not transforming growth factor-beta are essential for the differentiation of interleukin 17-producing human T helper cells. Nat Immunol **8:**942-9.

2. **Al-Mushrif, S., and B. M. Jones.** 1998. A study of the prevalence of hydrogen peroxide generating Lactobacilli in bacterial vaginosis: the determination of H2O2 concentrations generated, in vitro , by isolated strains and the levels found in vaginal secretions of women with and without infection. J Obstet Gynaecol **18:**63-7.

3. **Alimonti, J. B., J. Kimani, L. Matu, C. Wachihi, R. Kaul, F. A. Plummer, and K. R. Fowke.** 2006. Characterization of CD8 T-cell responses in HIV-1-exposed seronegative commercial sex workers from Nairobi, Kenya. Immunol Cell Biol **84:**482-5.

4. **Alimonti, J. B., S. A. Koesters, J. Kimani, L. Matu, C. Wachihi, F. A. Plummer, and K. R. Fowke.** 2005. CD4+ T cell responses in HIV-exposed seronegative women are qualitatively distinct from those in HIV-infected women. J Infect Dis **191:**20-4.

5. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-402.

6. **Alvarez-Barrientos, A., J. Arroyo, R. Canton, C. Nombela, and M. Sanchez-Perez.** 2000. Applications of flow cytometry to clinical microbiology. Clin Microbiol Rev **13:**167-95.

7. **Amsel, R., P. A. Totten, C. A. Spiegel, K. C. Chen, D. Eschenbach, and K. K. Holmes.** 1983. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. Am J Med **74:**14-22.

8. **Antonio, M. A., L. K. Rabe, and S. L. Hillier.** 2005. Colonization of the rectum by Lactobacillus species and decreased risk of bacterial vaginosis. J Infect Dis **192:**394-8.

9. **Anukam, K. C., E. O. Osazuwa, I. Ahonkhai, and G. Reid.** 2006. Lactobacillus vaginal microbiota of women attending a reproductive health care service in Benin city, Nigeria. Sex Transm Dis **33:**59-62.

10. **Apetrei, C., N. W. Lerche, I. Pandrea, B. Gormus, G. Silvestri, A. Kaur, D. L. Robertson, J. Hardcastle, A. A. Lackner, and P. A. Marx.** 2006. Kuru experiments triggered the emergence of pathogenic SIVmac. AIDS **20:**317-21.

11. **Arthos, J., C. Cicala, E. Martinelli, K. Macleod, D. Van Ryk, D. Wei, Z. Xiao, T. D. Veenstra, T. P. Conrad, R. A. Lempicki, S. McLaughlin, M. Pascuccio, R. Gopaul, J. McNally, C. C. Cruz, N. Censoplano, E. Chung, K. N. Reitano, S. Kottilil, D. J. Goode, and A. S. Fauci.** 2008. HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. Nat Immunol **9:**301-9.

12. **Ascher, M. S., and H. W. Sheppard.** 1988. AIDS as immune system activation: a model for pathogenesis. Clin Exp Immunol **73:**165-7.

13. **Atashili, J., C. Poole, P. M. Ndumbe, A. A. Adimora, and J. S. Smith.** 2008. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. AIDS **22:**1493-501.

14. **Baele, M., M. Vaneechoutte, R. Verhelst, M. Vancanneyt, L. A. Devriese, and F. Haesebrouck.** 2002. Identification of Lactobacillus species using tDNA-PCR. J Microbiol Methods **50:**263-71.

15. **Bailes, E., F. Gao, F. Bibollet-Ruche, V. Courgnaud, M. Peeters, P. A. Marx, B. H. Hahn, and P. M. Sharp.** 2003. Hybrid origin of SIV in chimpanzees. Science **300:**1713.

16. **Ball, T. B., H. Ji, J. Kimani, P. McLaren, C. Marlin, A. V. Hill, and F. A. Plummer.** 2007. Polymorphisms in IRF-1 associated with resistance to HIV-1 infection in highly exposed uninfected Kenyan sex workers. AIDS **21:**1091-101.

17. **Baranger, K., M. L. Zani, J. Chandenier, S. Dallet-Choisy, and T. Moreau.** 2008. The antibacterial and antifungal properties of trappin-2 (pre-elafin) do not depend on its protease inhibitory function. FEBS J **275:**2008-20.

18. **Barré-Sinoussi, F.** 2008. HIV/AIDS pathogenesis: to fight or tolerate the enemy?, 25 Years of HIV, Institut Pasteur, Paris, France.

19. **Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220:**868-71.

20. **Bartlett, J. G., A. B. Onderdonk, E. Drude, C. Goldstein, M. Anderka, S. Alpert, and W. M. McCormack.** 1977. Quantitative bacteriology of the vaginal flora. J Infect Dis **136:**271-7.

21. **Bartlett, J. G., and B. F. Polk.** 1984. Bacterial flora of the vagina: quantitative study. Rev Infect Dis **6 Suppl 1:**S67-72.

22. **Barton, G. M., Pasare, C., Medzhitov, R.** 2004. Toll-like receptors and control of adaptive immunity, p. pp. 271-285. *In* S. H. E. Kaufmann, Medzhitov, R., Gordon, S. (ed.), The Innate Immune Response to Infection. ASM Press, Washington, DC.

23. **Beaumier, C. M., L. D. Harris, S. Goldstein, N. R. Klatt, S. Whitted, J. McGinty, C. Apetrei, I. Pandrea, V. M. Hirsch, and J. M. Brenchley.** 2009. CD4 downregulation by memory CD4+ T cells in vivo renders African green monkeys resistant to progressive SIVagm infection. Nat Med **15:**879-85.

24. **Begaud, E., L. Chartier, V. Marechal, J. Ipero, J. Leal, P. Versmisse, G. Breton, A. Fontanet, C. Capoulade-Metay, H. Fleury, F. Barre-Sinoussi, D. Scott-Algara, and G. Pancino.** 2006. Reduced CD4 T cell activation and in vitro susceptibility to HIV-1 infection in exposed uninfected Central Africans. Retrovirology **3:**35.

25. **Beigi, R. H., H. C. Wiesenfeld, S. L. Hillier, T. Straw, and M. A. Krohn.** 2005. Factors associated with absence of H2O2-producing Lactobacillus among women with bacterial vaginosis. J Infect Dis **191:**924-9.

26. **Bernstein, K. T., J. L. Marcus, G. Nieri, S. S. Philip, and J. D. Klausner.** 2009. Rectal Gonorrhea and Chlamydia Reinfection Is Associated With Increased Risk of HIV Seroconversion. J Acquir Immune Defic Syndr.

27. **Biasin, M., S. L. Caputo, L. Speciale, F. Colombo, L. Racioppi, A. Zagliani, C. Ble, F. Vichi, L. Cianferoni, A. M. Masci, M. L. Villa, P. Ferrante, F. Mazzotta, and M. Clerici.** 2000. Mucosal and systemic immune activation is present in human immunodeficiency virus-exposed seronegative women. J Infect Dis **182:**1365-74.

28. **Bobardt, M. D., U. Chatterji, S. Selvarajah, B. Van der Schueren, G. David, B. Kahn, and P. A. Gallay.** 2007. Cell-free human immunodeficiency virus type 1 transcytosis through primary genital epithelial cells. J Virol **81:**395-405.

29. **Boskey, E. R., R. A. Cone, K. J. Whaley, and T. R. Moench.** 2001. Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. Hum Reprod **16:**1809-13.

30. **Boskey, E. R., K. M. Telsch, K. J. Whaley, T. R. Moench, and R. A. Cone.** 1999. Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. Infect Immun **67:**5170-5.

31. **Bouschbacher, M., M. Bomsel, E. Verronese, S. Gofflo, Y. Ganor, C. Dezutter-Dambuyant, and J. Valladeau.** 2008. Early events in HIV transmission through a human reconstructed vaginal mucosa. AIDS **22:**1257-66.

32. **Braat, H., J. van den Brande, E. van Tol, D. Hommes, M. Peppelenbosch, and S. van Deventer.** 2004. Lactobacillus rhamnosus induces peripheral hyporesponsiveness in stimulated CD4+ T cells via modulation of dendritic cell function. Am J Clin Nutr **80:**1618-25.

33. **Brenchley, J. M., and D. C. Douek.** 2008. The mucosal barrier and immune activation in HIV pathogenesis. Curr Opin HIV AIDS **3:**356-61.

34. **Brenchley, J. M., M. Paiardini, K. S. Knox, A. I. Asher, B. Cervasi, T. E. Asher, P. Scheinberg, D. A. Price, C. A. Hage, L. M. Kholi, A. Khoruts, I. Frank, J. Else, T. Schacker, G. Silvestri, and D. C. Douek.** 2008. Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections. Blood **112:**2826-35.

35. **Brenchley, J. M., D. A. Price, and D. C. Douek.** 2006. HIV disease: fallout from a mucosal catastrophe? Nat Immunol **7:**235-9.

36. **Brenchley, J. M., D. A. Price, T. W. Schacker, T. E. Asher, G. Silvestri, S. Rao, Z. Kazzaz, E. Bornstein, O. Lambotte, D. Altmann, B. R. Blazar, B. Rodriguez, L. Teixeira-Johnson, A. Landay, J. N. Martin, F. M. Hecht, L. J. Picker, M. M. Lederman, S. G. Deeks, and D. C. Douek.** 2006. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nat Med **12:**1365-71.

37. **Buffa, V., D. Stieh, N. Mamhood, Q. Hu, P. Fletcher, and R. J. Shattock.** 2009. Cyanovirin-N potently inhibits human immunodeficiency virus type 1 infection in cellular and cervical explant models. J Gen Virol **90:**234-43.

38. **Burgener, A., J. Boutilier, C. Wachihi, J. Kimani, M. Carpenter, G. Westmacott, K. Cheng, T. B. Ball, and F. Plummer.** 2008. Identification of differentially expressed proteins in the cervical mucosa of HIV-1-resistant sex workers. J Proteome Res **7:**4446-54.

39. **Card, C. M., P. J. McLaren, C. Wachihi, J. Kimani, F. A. Plummer, and K. R. Fowke.** 2009. Decreased immune activation in resistance to HIV-1 infection is associated with an elevated frequency of CD4(+)CD25(+)FOXP3(+) regulatory T cells. J Infect Dis **199:**1318-22.

40. **Cardinale, M., L. Brusetti, P. Quatrini, S. Borin, A. M. Puglia, A. Rizzi, E. Zanardini, C. Sorlini, C. Corselli, and D. Daffonchio.** 2004. Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. Appl Environ Microbiol **70:**6147-56.

41. **Cauci, S., S. Guaschino, S. Driussi, D. De Santo, P. Lanzafame, and F. Quadrifoglio.** 2002. Correlation of local interleukin-8 with immunoglobulin A against Gardnerella vaginalis hemolysin and with prolidase and sialidase levels in women with bacterial vaginosis. J Infect Dis **185:**1614-20.

42. **Chow, Y. H., D. Yu, J. Y. Zhang, Y. Xie, O. L. Wei, C. Chiu, M. Foroohar, O. O. Yang, N. H. Park, I. S. Chen, and S. Pang.** 2002. gp120-Independent infection of CD4(-) epithelial cells and CD4(+) T-cells by HIV-1. J Acquir Immune Defic Syndr **30:**1-8.

43. **Collado, M. C., and Y. Sanz.** 2007. Quantification of mucosa-adhered microbiota of lambs and calves by the use of culture methods and fluorescent in situ hybridization coupled with flow cytometry techniques. Vet Microbiol **121:**299-306.

44. **Culhane, J. F., D. Desanto, R. L. Goldenberg, K. F. McCollum, F. King, and S. Guaschino.** 2005. Variation in Nugent score and leukocyte count in fluid collected from different vaginal sites. Obstet Gynecol **105:**120-3.

45. **Davey, H. M., and D. B. Kell.** 1996. Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. Microbiol Rev **60:**641-96.

46. **De Angelis, M., R. Di Cagno, G. Gallo, M. Curci, S. Siragusa, C. Crecchio, E. Parente, and M. Gobbetti.** 2007. Molecular and functional characterization of Lactobacillus sanfranciscensis strains isolated from sourdoughs. Int J Food Microbiol **114:**69-82.

47. **de Jong, M. A., and T. B. Geijtenbeek.** 2009. Human immunodeficiency virus-1 acquisition in genital mucosa: Langerhans cells as key-players. J Intern Med **265:**18-28.

48. **Delves, P. J., Martin, S.J, Burton, D.R., Roitt, I.M.** 2006. Essential Immunology, 11th edition. Blackwell Publishing, Malden, MA.

49. **Dethlefsen, L., M. McFall-Ngai, and D. A. Relman.** 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature **449:**811-8.

50. **Douek, D. C., M. Roederer, and R. A. Koup.** 2009. Emerging concepts in the immunopathogenesis of AIDS. Annu Rev Med **60:**471-84.

51. **Dumonceaux, T. J., J. E. Hill, S. A. Briggs, K. K. Amoako, S. M. Hemmingsen, and A. G. Van Kessel.** 2006. Enumeration of specific bacterial populations in complex intestinal communities using quantitative PCR based on the chaperonin-60 target. J Microbiol Methods **64:**46-62.

52. **Dumonceaux, T. J., J. E. Hill, S. M. Hemmingsen, and A. G. Van Kessel.** 2006. Characterization of intestinal microbiota and response to dietary virginiamycin supplementation in the broiler chicken. Appl Environ Microbiol **72:**2815-23.

53. **Dumonceaux, T. J., J. Schellenberg, V. Goleski, J. E. Hill, W. Jaoko, J. Kimani, D. Money, T. B. Ball, F. A. Plummer, and A. Severini.** 2009. Multiplex detection of bacteria associated with normal microbiota and with bacterial vaginosis in vaginal swabs by use of oligonucleotide-coupled fluorescent microspheres. J Clin Microbiol **47:**4067-77.

54. **Edwards, R. A., B. Rodriguez-Brito, L. Wegley, M. Haynes, M. Breitbart, D. M. Peterson, M. O. Saar, S. Alexander, E. C. Alexander, Jr., and F. Rohwer.** 2006. Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics **7:**57.

55. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A **95:**14863-8.

56. **El Aila, N. A., I. Tency, G. Claeys, H. Verstraelen, B. Saerens, G. L. Santiago, E. De Backer, P. Cools, M. Temmerman, R. Verhelst, and M. Vaneechoutte.** 2009. Identification and genotyping of bacteria from paired vaginal and rectal samples from pregnant women indicates similarity between vaginal and rectal microflora. BMC Infect Dis **9:**167.

57. **Falsen, E., C. Pascual, B. Sjoden, M. Ohlen, and M. D. Collins.** 1999. Phenotypic and phylogenetic characterization of a novel Lactobacillus species from human sources: description of Lactobacillus iners sp. nov. Int J Syst Bacteriol **49 Pt 1:**217-21.

58. **Favre, D., S. Lederer, B. Kanwar, Z. M. Ma, S. Proll, Z. Kasakow, J. Mold, L. Swainson, J. D. Barbour, C. R. Baskin, R. Palermo, I. Pandrea, C. J. Miller, M. G. Katze, and J. M. McCune.** 2009. Critical loss of the balance between Th17 and T regulatory cell populations in pathogenic SIV infection. PLoS Pathog **5:**e1000295.

59. **Felsenstein, J.** 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution **39:**783-791.

60. **Fleming, D. T., and J. N. Wasserheit.** 1999. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect **75:**3-17.

61. **Fontaine, E. A., and D. Taylor-Robinson.** 1990. Comparison of quantitative and qualitative methods of detecting hydrogen peroxide produced by human vaginal strains of lactobacilli. J Appl Bacteriol **69:**326-31.

62. **Forsum, U., A. Hallen, and P. G. Larsson.** 2005. Bacterial vaginosis--a laboratory and clinical diagnostics enigma. Apmis **113:**153-61.

63. **Forsum, U., E. Holst, P. G. Larsson, A. Vasquez, T. Jakobsson, and I. Mattsby-Baltzer.** 2005. Bacterial vaginosis--a microbiological and immunological enigma. Apmis **113:**81-90.

64. **Fowke, K. R., T. Dong, S. L. Rowland-Jones, J. Oyugi, W. J. Rutherford, J. Kimani, P. Krausa, J. Bwayo, J. N. Simonsen, G. M. Shearer, and F. A. Plummer.** 1998. HIV type 1 resistance in Kenyan sex workers is not associated with altered cellular susceptibility to HIV type 1 infection or enhanced beta-chemokine production. AIDS Res Hum Retroviruses **14:**1521-30.

65. **Fowke, K. R., R. Kaul, K. L. Rosenthal, J. Oyugi, J. Kimani, W. J. Rutherford, N. J. Nagelkerke, T. B. Ball, J. J. Bwayo, J. N. Simonsen, G. M. Shearer, and F. A. Plummer.** 2000. HIV-1-specific cellular immune responses among HIV-1-resistant sex workers. Immunol Cell Biol **78:**586-95.

66. **Fowke, K. R., N. J. Nagelkerke, J. Kimani, J. N. Simonsen, A. O. Anzala, J. J. Bwayo, K. S. MacDonald, E. N. Ngugi, and F. A. Plummer.** 1996. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. Lancet **348:**1347-51.

67. **Fredricks, D. N., T. L. Fiedler, K. K. Thomas, B. B. Oakley, and J. M. Marrazzo.** 2007. Targeted PCR for detection of vaginal bacteria associated with bacterial vaginosis. J Clin Microbiol **45:**3270-6.

68. **Furuta, Y., K. Eriksson, B. Svennerholm, P. Fredman, P. Horal, S. Jeansson, A. Vahlne, J. Holmgren, and C. Czerkinsky.** 1994. Infection of vaginal and colonic epithelial cells by the human immunodeficiency virus type 1 is neutralized by antibodies raised against conserved epitopes in the envelope glycoprotein gp120. Proc Natl Acad Sci U S A **91:**12559-63.

69. **Gassler, N., H. Paul, and M. Runge.** 2006. Rapid detection of urinary tract infection--evaluation of flow cytometry. Clin Nephrol **66:**331-5.

70. **Ghosh, M., Z. Shen, J. V. Fahey, S. Cu-Uvin, K. Mayer, and C. R. Wira.** 2009. Trappin-2/Elafin: a novel innate anti-human immunodeficiency virus-1 molecule of the human female reproductive tract. Immunology.

71. **Ghys, P. D., L. Belec, M. O. Diallo, V. Ettiegne-Traore, P. Becquart, C. Maurice, J. N. Nkengasong, I. M. Coulibaly, A. E. Greenberg, M. Laga, and S. Z. Wiktor.** 2000. Cervicovaginal anti-HIV antibodies in HIV-seronegative female sex workers in Abidjan, Cote d'Ivoire. AIDS **14:**2603-8.

72. **Gilbert, M. T., A. Rambaut, G. Wlasiuk, T. J. Spira, A. E. Pitchenik, and M. Worobey.** 2007. The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci U S A **104:**18566-70.

73. **Goh, S. H., S. Potter, J. O. Wood, S. M. Hemmingsen, R. P. Reynolds, and A. W. Chow.** 1996. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. J Clin Microbiol **34:**818-23.

74. **Gottlieb, M. S.** 1981. Pneumocystis pneumonia--Los Angeles. MMWR Morb Mortal Wkly Rep **30:**250-2.

75. **Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf, and A. Saxon.** 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med **305:**1425-31.

76. **Gravitt, P. E., J. V. Lacey, Jr., L. A. Brinton, W. A. Barnes, J. R. Kornegay, M. D. Greenberg, S. M. Greene, O. C. Hadjimichael, L. McGowan, R. Mortel, P. E. Schwartz, R. Zaino, and A. Hildesheim.** 2001. Evaluation of self-collected cervicovaginal cell samples for human papillomavirus testing by polymerase chain reaction. Cancer Epidemiol Biomarkers Prev **10:**95-100.

77. **Grice, E. A., H. H. Kong, S. Conlan, C. B. Deming, J. Davis, A. C. Young, G. G. Bouffard, R. W. Blakesley, P. R. Murray, E. D. Green, M. L. Turner, and J. A. Segre.** 2009. Topographical and temporal diversity of the human skin microbiome. Science **324:**1190-2.

78. **Gunasekera, T. S., P. V. Attfield, and D. A. Veal.** 2000. A flow cytometry method for rapid detection and enumeration of total bacteria in milk. Appl Environ Microbiol **66:**1228-32.

79. **Haase, A. T.** 1986. Pathogenesis of lentivirus infections. Nature **322:**130-6.

80. **Haase, A. T.** 2005. Perils at mucosal front lines for HIV and SIV and their hosts. Nat Rev Immunol **5:**783-92.

81. **Haase, A. T.** 2001. The pathogenesis of sexual mucosal transmission and early stages of infection: obstacles and a narrow window of opportunity for prevention. AIDS **15 Suppl 1:**S10-1.

82. **Hammes, F., M. Berney, Y. Wang, M. Vital, O. Koster, and T. Egli.** 2007. Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. Water Res.

83. **Hardie, R. A., E. Knight, B. Bruneau, C. Semeniuk, K. Gill, N. Nagelkerke, J. Kimani, C. Wachihi, E. Ngugi, M. Luo, and F. A. Plummer.** 2008. A common human leucocyte antigen-DP genotype is associated with resistance to HIV-1 infection in Kenyan sex workers. AIDS **22:**2038-42.

84. **Hay, P.** 2005. Life in the littoral zone: lactobacilli losing the plot. Sex Transm Infect **81:**100-2.

85. **Hill, J. E., S. H. Goh, D. M. Money, M. Doyle, A. Li, W. L. Crosby, M. Links, A. Leung, D. Chan, and S. M. Hemmingsen.** 2005. Characterization of vaginal microflora of healthy, nonpregnant women by chaperonin-60 sequence-based methods. Am J Obstet Gynecol **193:**682-92.

86. **Hill, J. E., A. Paccagnella, K. Law, P. L. Melito, D. L. Woodward, L. Price, A. H. Leung, L. K. Ng, S. M. Hemmingsen, and S. H. Goh.** 2006. Identification of Campylobacter spp. and discrimination from Helicobacter and Arcobacter spp. by direct sequencing of PCR-amplified cpn60 sequences and comparison to cpnDB, a chaperonin reference sequence database. J Med Microbiol **55:**393-9.

87. **Hill, J. E., S. L. Penny, K. G. Crowell, S. H. Goh, and S. M. Hemmingsen.** 2004. cpnDB: a chaperonin sequence database. Genome Res **14:**1669-75.

88. **Hill, J. E., J. R. Town, and S. M. Hemmingsen.** 2006. Improved template representation in cpn60 polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. Environ Microbiol **8:**741-6.

89. **Hillier, S. L.** 1998. The vaginal microbial ecosystem and resistance to HIV. AIDS Res Hum Retroviruses **14 Suppl 1:**S17-21.

90. **Hirbod, T., R. Kaul, C. Reichard, J. Kimani, E. Ngugi, J. J. Bwayo, N. Nagelkerke, K. Hasselrot, B. Li, S. Moses, K. S. MacDonald, and K. Broliden.** 2008. HIV-neutralizing immunoglobulin A and HIV-specific proliferation are independently associated with reduced HIV acquisition in Kenyan sex workers. AIDS **22:**727-35.

91. **Hladik, F., and M. J. McElrath.** 2008. Setting the stage: host invasion by HIV. Nat Rev Immunol **8:**447-57.

92. **Hooper, E.** 2001. Experimental oral polio vaccines and acquired immune deficiency syndrome. Philos Trans R Soc Lond B Biol Sci **356:**803-14.

93. **Horton, R. E., T. B. Ball, C. Wachichi, W. Jaoko, W. J. Rutherford, L. McKinnon, R. Kaul, A. Rebbapragada, J. Kimani, and F. A. Plummer.** 2009. Cervical HIV-specific IgA in a population of commercial sex workers correlates with repeated exposure but not resistance to HIV. AIDS Res Hum Retroviruses **25:**83-92.

94. **Huth, E. J.** 1989. Style notes: bacterial vaginosis or vaginal bacteriosis? Ann Intern Med **111:**553-4.

95. **Hyman, R. W., M. Fukushima, L. Diamond, J. Kumm, L. C. Giudice, and R. W. Davis.** 2005. Microbes on the human vaginal epithelium. Proc Natl Acad Sci U S A **102:**7952-7.

96. **Iqbal, S. M., T. B. Ball, J. Kimani, P. Kiama, P. Thottingal, J. E. Embree, K. R. Fowke, and F. A. Plummer.** 2005. Elevated T cell counts and RANTES expression in

the genital mucosa of HIV-1-resistant Kenyan commercial sex workers. J Infect Dis **192:**728-38.

97. **Iqbal, S. M., T. B. Ball, P. Levinson, L. Maranan, W. Jaoko, C. Wachihi, B. J. Pak, V. N. Podust, K. Broliden, T. Hirbod, R. Kaul, and F. A. Plummer.** 2009. Elevated elafin/trappin-2 in the female genital tract is associated with protection against HIV acquisition. AIDS **23:**1669-77.

98. **Ison, C. A., and P. E. Hay.** 2002. Validation of a simplified grading of Gram stained vaginal smears for use in genitourinary medicine clinics. Sex Transm Infect **78:**413-5.

99. **Jennes, W., D. Evertse, M. Y. Borget, B. Vuylsteke, C. Maurice, J. N. Nkengasong, and L. Kestens.** 2006. Suppressed cellular alloimmune responses in HIV-exposed seronegative female sex workers. Clin Exp Immunol **143:**435-44.

100. **Jennes, W., S. Sawadogo, S. Koblavi-Deme, B. Vuylsteke, C. Maurice, T. H. Roels, T. Chorba, J. N. Nkengasong, and L. Kestens.** 2003. Cellular human immunodeficiency virus (HIV)-protective factors: a comparison of HIV-exposed seronegative female sex workers and female blood donors in Abidjan, Cote d'Ivoire. J Infect Dis **187:**206-14.

101. **Jin, L., L. Tao, S. I. Pavlova, J. S. So, N. Kiwanuka, Z. Namukwaya, B. A. Saberbein, and M. Wawer.** 2007. Species diversity and relative abundance of vaginal lactic acid bacteria from women in Uganda and Korea. J Appl Microbiol **102:**1107-15.

102. **Josephy, P. D., T. Eling, and R. P. Mason.** 1982. The horseradish peroxidase-catalyzed oxidation of 3,5,3',5'-tetramethylbenzidine. Free radical and charge-transfer complex intermediates. J Biol Chem **257:**3669-75.

103. **Kader, M., X. Wang, M. Piatak, J. Lifson, M. Roederer, R. Veazey, and J. J. Mattapallil.** 2009. Alpha4(+)beta7(hi)CD4(+) memory T cells harbor most Th-17 cells and are preferentially infected during acute SIV infection. Mucosal Immunol **2:**439-49.

104. **Kane, C. T., S. Diawara, H. D. Ndiaye, P. A. Diallo, A. S. Wade, A. G. Diallo, L. Belec, and S. Mboup.** 2009. Concentrated and linked epidemics of both HSV-2 and HIV-1/HIV-2 infections in Senegal: public health impacts of the spread of HIV. Int J STD AIDS **20:**793-6.

105. **Kaul, R., C. Pettengell, P. M. Sheth, S. Sunderji, A. Biringer, K. MacDonald, S. Walmsley, and A. Rebbapragada.** 2008. The genital tract immune milieu: an important determinant of HIV susceptibility and secondary transmission. J Reprod Immunol **77:**32-40.

106. **Kaul, R., S. L. Rowland-Jones, J. Kimani, T. Dong, H. B. Yang, P. Kiama, T. Rostron, E. Njagi, J. J. Bwayo, K. S. MacDonald, A. J. McMichael, and F. A. Plummer.** 2001. Late seroconversion in HIV-resistant Nairobi prostitutes despite pre-existing HIV-specific CD8+ responses. J Clin Invest **107:**341-9.

107. **Kaul, R., D. Trabattoni, J. J. Bwayo, D. Arienti, A. Zagliani, F. M. Mwangi, C. Kariuki, E. N. Ngugi, K. S. MacDonald, T. B. Ball, M. Clerici, and F. A. Plummer.**

1999. HIV-1-specific mucosal IgA in a cohort of HIV-1-resistant Kenyan sex workers. AIDS **13:**23-9.

108. **Keane, F. E., C. A. Ison, and D. Taylor-Robinson.** 1997. A longitudinal study of the vaginal flora over a menstrual cycle. Int J STD AIDS **8:**489-94.

109. **Kilic, A. O., S. I. Pavlova, S. Alpay, S. S. Kilic, and L. Tao.** 2001. Comparative study of vaginal Lactobacillus phages isolated from women in the United States and Turkey: prevalence, morphology, host range, and DNA homology. Clin Diagn Lab Immunol **8:**31-9.

110. **Kimani, J., R. Kaul, N. J. Nagelkerke, M. Luo, K. S. MacDonald, E. Ngugi, K. R. Fowke, B. T. Ball, A. Kariri, J. Ndinya-Achola, and F. A. Plummer.** 2008. Reduced rates of HIV acquisition during unprotected sex by Kenyan female sex workers predating population declines in HIV prevalence. AIDS **22:**131-7.

111. **Klebanoff, S. J., and R. W. Coombs.** 1991. Viricidal effect of Lactobacillus acidophilus on human immunodeficiency virus type 1: possible role in heterosexual transmission. J Exp Med **174:**289-92.

112. **Koning, F. A., S. A. Otto, M. D. Hazenberg, L. Dekker, M. Prins, F. Miedema, and H. Schuitemaker.** 2005. Low-level CD4+ T cell activation is associated with low susceptibility to HIV-1 infection. J Immunol **175:**6117-22.

113. **Koprowski, H.** 2001. Hypotheses and facts. Philos Trans R Soc Lond B Biol Sci **356:**831-3.

114. **Kornfeld, C., M. J. Ploquin, I. Pandrea, A. Faye, R. Onanga, C. Apetrei, V. Poaty-Mavoungou, P. Rouquet, J. Estaquier, L. Mortara, J. F. Desoutter, C. Butor, R. Le Grand, P. Roques, F. Simon, F. Barre-Sinoussi, O. M. Diop, and M. C. Muller-Trutwin.** 2005. Antiinflammatory profiles during primary SIV infection in African green monkeys are associated with protection against AIDS. J Clin Invest **115:**1082-91.

115. **Kreiss, J. K., D. Koech, F. A. Plummer, K. K. Holmes, M. Lightfoote, P. Piot, A. R. Ronald, J. O. Ndinya-Achola, L. J. D'Costa, P. Roberts, and et al.** 1986. AIDS virus infection in Nairobi prostitutes. Spread of the epidemic to East Africa. N Engl J Med **314:**414-8.

116. **Lacap, P. A., J. D. Huntington, M. Luo, N. J. Nagelkerke, T. Bielawny, J. Kimani, C. Wachihi, E. N. Ngugi, and F. A. Plummer.** 2008. Associations of human leukocyte antigen DRB with resistance or susceptibility to HIV-1 infection in the Pumwani Sex Worker Cohort. AIDS **22:**1029-38.

117. **Larsson, P. G., M. Bergstrom, U. Forsum, B. Jacobsson, A. Strand, and P. Wolner-Hanssen.** 2005. Bacterial vaginosis. Transmission, role in genital tract infection and pregnancy outcome: an enigma. Apmis **113:**233-45.

118. **Legrand, F. A., D. F. Nixon, C. P. Loo, E. Ono, J. M. Chapman, M. Miyamoto, R. S. Diaz, A. M. Santos, R. C. Succi, J. Abadi, M. G. Rosenberg, M. I. de Moraes-Pinto, and E. G. Kallas.** 2006. Strong HIV-1-specific T cell responses in HIV-1-exposed

uninfected infants and neonates revealed after regulatory T cell removal. PLoS One **1:**e102.

119. **Levy, J. A.** 2009. HIV pathogenesis: 25 years of progress and persistent challenges. AIDS **23:**147-60.

120. **Lin, P. W., L. E. Myers, L. Ray, S. C. Song, T. R. Nasr, A. J. Berardinelli, K. Kundu, N. Murthy, J. M. Hansen, and A. S. Neish.** 2009. Lactobacillus rhamnosus blocks inflammatory signaling in vivo via reactive oxygen species generation. Free Radic Biol Med **47:**1205-11.

121. **Liovat, A. S., B. Jacquelin, M. J. Ploquin, F. Barre-Sinoussi, and M. C. Muller-Trutwin.** 2009. African non human primates infected by SIV - why don't they get sick? Lessons from studies on the early phase of non-pathogenic SIV infection. Curr HIV Res **7:**39-50.

122. **Livengood, C. H.** 2009. Bacterial vaginosis: an overview for 2009. Rev Obstet Gynecol **2:**28-37.

123. **Losikoff, P., R. Fichorova, B. Snyder, I. Rodriguez, S. Cu-Uvin, J. Harwell, and K. H. Mayer.** 2007. Genital tract interleukin-8 but not interleukin-1beta or interleukin-6 concentration is associated with bacterial vaginosis and its clearance in HIV-infected and HIV-uninfected women. Infect Dis Obstet Gynecol **2007:**92307.

124. **MacDonald, K. S., K. R. Fowke, J. Kimani, V. A. Dunand, N. J. Nagelkerke, T. B. Ball, J. Oyugi, E. Njagi, L. K. Gaur, R. C. Brunham, J. Wade, M. A. Luscher, P. Krausa, S. Rowland-Jones, E. Ngugi, J. J. Bwayo, and F. A. Plummer.** 2000. Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. J Infect Dis **181:**1581-9.

125. **Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, and J. M. Tiedje.** 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res **29:**173-4.

126. **Mares, D., J. A. Simoes, R. M. Novak, and G. T. Spear.** 2008. TLR2-mediated cell stimulation in bacterial vaginosis. J Reprod Immunol **77:**91-9.

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

128. **Marmor, M., K. Hertzmark, S. M. Thomas, P. N. Halkitis, and M. Vogler.** 2006. Resistance to HIV infection. J Urban Health **83:**5-17.

129. **Marrazzo, J. M., H. C. Wiesenfeld, P. J. Murray, B. Busse, L. Meyn, M. Krohn, and S. L. Hillier.** 2006. Risk factors for cervicitis among women with bacterial vaginosis. J Infect Dis **193:**617-24.

130. **Martin, H. L., B. A. Richardson, P. M. Nyange, L. Lavreys, S. L. Hillier, B. Chohan, K. Mandaliya, J. O. Ndinya-Achola, J. Bwayo, and J. Kreiss.** 1999. Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. J Infect Dis **180:**1863-8.

131. **Marx, P. A., P. G. Alcabes, and E. Drucker.** 2001. Serial human passage of simian immunodeficiency virus by unsterile injections and the emergence of epidemic human immunodeficiency virus in Africa. Philos Trans R Soc Lond B Biol Sci **356:**911-20.

132. **Masfari, A. N., B. I. Duerden, and G. R. Kinghorn.** 1986. Quantitative studies of vaginal bacteria. Genitourin Med **62:**256-63.

133. **Matu, M. N., G. O. Orinda, E. N. Njagi, C. R. Cohen, and E. A. Bukusi.** 2009. In vitro inhibitory activity of human vaginal lactobacilli against pathogenic bacteria associated with bacterial vaginosis in Kenyan women. Anaerobe.

134. **Mazzoli, S., D. Trabattoni, S. Lo Caputo, S. Piconi, C. Ble, F. Meacci, S. Ruzzante, A. Salvi, F. Semplici, R. Longhi, M. L. Fusi, N. Tofani, M. Biasin, M. L. Villa, F. Mazzotta, and M. Clerici.** 1997. HIV-specific mucosal and cellular immunity in HIV-seronegative partners of HIV-seropositive individuals. Nat Med **3:**1250-7.

135. **McClelland, R. S., B. A. Richardson, W. M. Hassan, V. Chohan, L. Lavreys, K. Mandaliya, J. Kiarie, W. Jaoko, J. O. Ndinya-Achola, J. M. Baeten, A. E. Kurth, and K. K. Holmes.** 2008. Improvement of vaginal health for Kenyan women at risk for acquisition of human immunodeficiency virus type 1: results of a randomized trial. J Infect Dis **197:**1361-8.

136. **McClelland, R. S., B. A. Richardson, W. M. Hassan, S. M. Graham, J. Kiarie, J. M. Baeten, K. Mandaliya, W. Jaoko, J. O. Ndinya-Achola, and K. K. Holmes.** 2009. Prospective Study of Vaginal Bacterial Flora and Other Risk Factors for Vulvovaginal Candidiasis. J Infect Dis **199:**1883-1890.

137. **McGowan, I.** 2010. Microbicides for HIV prevention: reality or hope? Curr Opin Infect Dis **23:**26-31.

138. **Menard, J. P., F. Fenollar, M. Henry, F. Bretelle, and D. Raoult.** 2008. Molecular quantification of Gardnerella vaginalis and Atopobium vaginae loads to predict bacterial vaginosis. Clin Infect Dis **47:**33-43.

139. **Mitchell, C. M., J. Balkus, K. J. Agnew, S. Cohn, A. Luque, R. Lawler, R. W. Coombs, and J. E. Hitti.** 2008. Bacterial vaginosis, not HIV, is primarily responsible for increased vaginal concentrations of proinflammatory cytokines. AIDS Res Hum Retroviruses **24:**667-71.

140. **Miyazawa, M., L. Lopalco, F. Mazzotta, S. Lo Caputo, F. Veas, and M. Clerici.** 2009. The 'immunologic advantage' of HIV-exposed seronegative individuals. AIDS **23:**161-75.

141. **Moreau, T., K. Baranger, S. Dade, S. Dallet-Choisy, N. Guyot, and M. L. Zani.** 2008. Multifaceted roles of human elafin and secretory leukocyte proteinase inhibitor (SLPI), two serine protease inhibitors of the chelonianin family. Biochimie **90:**284-95.

142. **Morison, L., G. Ekpo, B. West, E. Demba, P. Mayaud, R. Coleman, R. Bailey, and G. Walraven.** 2005. Bacterial vaginosis in relation to menstrual cycle, menstrual protection method, and sexual intercourse in rural Gambian women. Sex Transm Infect **81:**242-7.

143. **Morris, G. C., and C. J. Lacey.** 2010. Microbicides and HIV prevention: lessons from the past, looking to the future. Curr Opin Infect Dis **23:**57-63.

144. **Myer, L., L. Denny, R. Telerant, M. Souza, T. C. Wright, Jr., and L. Kuhn.** 2005. Bacterial vaginosis and susceptibility to HIV infection in South African women: a nested case-control study. J Infect Dis **192:**1372-80.

145. **Nam, H., K. Whang, and Y. Lee.** 2007. Analysis of vaginal lactic acid producing bacteria in healthy women. J Microbiol **45:**515-20.

146. **Nugent, R. P., M. A. Krohn, and S. L. Hillier.** 1991. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol **29:**297-301.

147. **O'Keefe, B. R., F. Vojdani, V. Buffa, R. J. Shattock, D. C. Montefiori, J. Bakke, J. Mirsalis, A. L. d'Andrea, S. D. Hume, B. Bratcher, C. J. Saucedo, J. B. McMahon, G. P. Pogue, and K. E. Palmer.** 2009. Scaleable manufacture of HIV-1 entry inhibitor griffithsin and validation of its safety and efficacy as a topical microbicide component. Proc Natl Acad Sci U S A **106:**6099-104.

148. **Oakley, B. B., T. L. Fiedler, J. M. Marrazzo, and D. N. Fredricks.** 2008. Diversity of human vaginal bacterial communities and associations with clinically defined bacterial vaginosis. Appl Environ Microbiol **74:**4898-909.

149. **Olmsted, S. S., K. V. Khanna, E. M. Ng, S. T. Whitten, O. N. Johnson, 3rd, R. B. Markham, R. A. Cone, and T. R. Moench.** 2005. Low pH immobilizes and kills human leukocytes and prevents transmission of cell-associated HIV in a mouse model. BMC Infect Dis **5:**79.

150. **Onderdonk, A. B., M. L. Lee, E. Lieberman, M. L. Delaney, and R. E. Tuomala.** 2003. Quantitative microbiologic models for preterm delivery. J Clin Microbiol **41:**1073-9.

151. **Onderdonk, A. B., G. R. Zamarchi, J. A. Walsh, R. D. Mellor, A. Munoz, and E. H. Kass.** 1986. Methods for quantitative and qualitative evaluation of vaginal microflora during menstruation. Appl Environ Microbiol **51:**333-9.

152. **Oyugi, J. O., F. C. Vouriot, J. Alimonti, S. Wayne, M. Luo, A. M. Land, Z. Ao, X. Yao, R. P. Sekaly, L. J. Elliott, J. N. Simonsen, T. B. Ball, W. Jaoko, J. Kimani, F. A. Plummer, and K. R. Fowke.** 2009. A common CD4 gene variant is associated with an increased risk of HIV-1 infection in Kenyan female commercial sex workers. J Infect Dis **199:**1327-34.

153. **Pandrea, I., D. L. Sodora, G. Silvestri, and C. Apetrei.** 2008. Into the wild: simian immunodeficiency virus (SIV) infection in natural hosts. Trends Immunol **29:**419-28.

154. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A **85:**2444-8.

155. **Peck, A., and E. D. Mellins.** Precarious balance: Th17 cells in host defense. Infect Immun **78:**32-8.

156. **Peng, B., R. Voltan, L. Lim, Y. Edghill-Smith, S. Phogat, D. S. Dimitrov, K. Arora, M. Leno, S. Than, R. Woodward, P. D. Markham, M. Cranage, and M. Robert-Guroff.** 2002. Rhesus macaque resistance to mucosal simian immunodeficiency virus infection is associated with a postentry block in viral replication. J Virol **76:**6016-26.

157. **Piot, P., F. A. Plummer, M. A. Rey, E. N. Ngugi, C. Rouzioux, J. O. Ndinya-Achola, G. Veracauteren, L. J. D'Costa, M. Laga, H. Nsanze, and et al.** 1987. Retrospective seroepidemiology of AIDS virus infection in Nairobi populations. J Infect Dis **155:**1108-12.

158. **Piot, P., T. C. Quinn, H. Taelman, F. M. Feinsod, K. B. Minlangu, O. Wobin, N. Mbendi, P. Mazebo, K. Ndangi, W. Stevens, and et al.** 1984. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. Lancet **2:**65-9.

159. **Piot, P., E. Van Dyck, M. Peeters, J. Hale, P. A. Totten, and K. K. Holmes.** 1984. Biotypes of Gardnerella vaginalis. J Clin Microbiol **20:**677-9.

160. **Plantier, J. C., M. Leoz, J. E. Dickerson, F. De Oliveira, F. Cordonnier, V. Lemee, F. Damond, D. L. Robertson, and F. Simon.** 2009. A new human immunodeficiency virus derived from gorillas. Nat Med **15:**871-2.

161. **Plummer, F. A., T. B. Ball, J. Kimani, and K. R. Fowke.** 1999. Resistance to HIV-1 infection among highly exposed sex workers in Nairobi: what mediates protection and why does it develop? Immunol Lett **66:**27-34.

162. **Plummer, F. A., J. N. Simonsen, D. W. Cameron, J. O. Ndinya-Achola, J. K. Kreiss, M. N. Gakinya, P. Waiyaki, M. Cheang, P. Piot, A. R. Ronald, and et al.** 1991. Cofactors in male-female sexual transmission of human immunodeficiency virus type 1. J Infect Dis **163:**233-9.

163. **Pudney, J., A. J. Quayle, and D. J. Anderson.** 2005. Immunological microenvironments in the human vagina and cervix: mediators of cellular immunity are concentrated in the cervical transformation zone. Biol Reprod **73:**1253-63.

164. **Pybus, V., and A. B. Onderdonk.** 1997. Evidence for a commensal, symbiotic relationship between Gardnerella vaginalis and Prevotella bivia involving ammonia: potential significance for bacterial vaginosis. J Infect Dis **175:**406-13.

165. **Rabe, L. K., and S. L. Hillier.** 2003. Optimization of media for detection of hydrogen peroxide production by Lactobacillus species. J Clin Microbiol **41:**3260-4.

166. **Robbins, K. E., P. Lemey, O. G. Pybus, H. W. Jaffe, A. S. Youngpairoj, T. M. Brown, M. Salemi, A. M. Vandamme, and M. L. Kalish.** 2003. U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. J Virol **77:**6359-66.

167. **Roberton, A. M., R. Wiggins, P. J. Horner, R. Greenwood, T. Crowley, A. Fernandes, M. Berry, and A. P. Corfield.** 2005. A novel bacterial mucinase, glycosulfatase, is associated with bacterial vaginosis. J Clin Microbiol **43:**5504-8.

168. **Ronnqvist, P. D., U. B. Forsgren-Brusk, and E. E. Grahn-Hakansson.** 2006. Lactobacilli in the female genital tract in relation to other genital microbes and vaginal pH. Acta Obstet Gynecol Scand **85:**726-35.

169. **Ross, R. A., M. L. Lee, M. L. Delaney, and A. B. Onderdonk.** 1994. Mixed-effect models for predicting microbial interactions in the vaginal ecosystem. J Clin Microbiol **32:**871-5.

170. **Rowland-Jones, S. L., T. Dong, K. R. Fowke, J. Kimani, P. Krausa, H. Newell, T. Blanchard, K. Ariyoshi, J. Oyugi, E. Ngugi, J. Bwayo, K. S. MacDonald, A. J. McMichael, and F. A. Plummer.** 1998. Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. J Clin Invest **102:**1758-65.

171. **Saez-Cirion, A., P. Versmisse, L. X. Truong, L. A. Chakrabarti, W. Carpentier, F. Barre-Sinoussi, D. Scott-Algara, and G. Pancino.** 2006. Persistent resistance to HIV-1 infection in CD4 T cells from exposed uninfected Vietnamese individuals is mediated by entry and post-entry blocks. Retrovirology **3:**81.

172. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406-25.

173. **Saldanha, A. J.** 2004. Java Treeview--extensible visualization of microarray data. Bioinformatics **20:**3246-8.

174. **Santiago, M. L., F. Range, B. F. Keele, Y. Li, E. Bailes, F. Bibollet-Ruche, C. Fruteau, R. Noe, M. Peeters, J. F. Brookfield, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 2005. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. J Virol **79:**12515-27.

175. **Schellenberg, J., T. Blake Ball, M. Lane, M. Cheang, and F. Plummer.** 2008. Flow cytometric quantification of bacteria in vaginal swab samples self-collected by adolescents attending a gynecology clinic. J Microbiol Methods **73:**216-26.

176. **Schellenberg, J., M. G. Links, J. E. Hill, T. J. Dumonceaux, G. A. Peters, S. Tyler, T. B. Ball, A. Severini, and F. A. Plummer.** 2009. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. Appl Environ Microbiol **75:**2889-98.

177. **Schellenberg, J., W. Smoragiewicz, and B. Karska-Wysocki.** 2006. A rapid method combining immunofluorescence and flow cytometry for improved understanding of competitive interactions between lactic acid bacteria (LAB) and methicillin-resistant S. aureus (MRSA) in mixed culture. J Microbiol Methods **65:**1-9.

178. **Schindler, M., J. Munch, O. Kutsch, H. Li, M. L. Santiago, F. Bibollet-Ruche, M. C. Muller-Trutwin, F. J. Novembre, M. Peeters, V. Courgnaud, E. Bailes, P. Roques, D. L. Sodora, G. Silvestri, P. M. Sharp, B. H. Hahn, and F. Kirchhoff.** 2006. Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. Cell **125:**1055-67.

179. **Schloss, P. D., and J. Handelsman.** 2006. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. Appl Environ Microbiol **72:**6773-9.

180. **Schwebke, J. R.** 1997. Bacterial vaginosis--more questions than answers. Genitourin Med **73:**333-4.

181. **Schwebke, J. R., S. C. Morgan, and H. L. Weiss.** 1997. The use of sequential self-obtained vaginal smears for detecting changes in the vaginal flora. Sex Transm Dis **24:**236-9.

182. **Serlin, M., M. A. Shafer, K. Tebb, A. A. Gyamfi, J. Moncada, J. Schachter, and C. Wibbelsman.** 2002. What sexually transmitted disease screening method does the adolescent prefer? Adolescents' attitudes toward first-void urine, self-collected vaginal swab, and pelvic examination. Arch Pediatr Adolesc Med **156:**588-91.

183. **Shafer, M. A., J. Moncada, C. B. Boyer, K. Betsinger, S. D. Flinn, and J. Schachter.** 2003. Comparing first-void urine specimens, self-collected vaginal swabs, and endocervical specimens to detect Chlamydia trachomatis and Neisseria gonorrhoeae by a nucleic acid amplification test. J Clin Microbiol **41:**4395-9.

184. **Shafir, S. C., F. J. Sorvillo, and L. Smith.** 2009. Current issues and considerations regarding trichomoniasis and human immunodeficiency virus in African-Americans. Clin Microbiol Rev **22:**37-45, Table of Contents.

185. **Shapiro, H. M.** 2003. Practical Flow Cytometry, 4th edition. John Wiley & Sons, Inc., New York, NY.

186. **Sharp, P. M., E. Bailes, R. R. Chaudhuri, C. M. Rodenburg, M. O. Santiago, and B. H. Hahn.** 2001. The origins of acquired immune deficiency syndrome viruses: where and when? Philos Trans R Soc Lond B Biol Sci **356:**867-76.

187. **Simoes, J. A., D. M. Citron, A. Aroutcheva, R. A. Anderson, Jr., C. J. Chany, 2nd, D. P. Waller, S. Faro, and L. J. Zaneveld.** 2002. Two novel vaginal microbicides

(polystyrene sulfonate and cellulose sulfate) inhibit Gardnerella vaginalis and anaerobes commonly associated with bacterial vaginosis. Antimicrob Agents Chemother **46:**2692-5.

188. **Simonsen, J. N., F. A. Plummer, E. N. Ngugi, C. Black, J. K. Kreiss, M. N. Gakinya, P. Waiyaki, L. J. D'Costa, J. O. Ndinya-Achola, P. Piot, and et al.** 1990. HIV infection among lower socioeconomic strata prostitutes in Nairobi. AIDS **4:**139-44.

189. **Sincock, S. A., and J. P. Robinson.** 2001. Flow cytometric analysis of microorganisms. Methods Cell Biol **64:**511-37.

190. **Smart, S., A. Singal, and A. Mindel.** 2004. Social and sexual risk factors for bacterial vaginosis. Sex Transm Infect **80:**58-62.

191. **Smith, T. F., and M. S. Waterman.** 1981. Identification of common molecular subsequences. J Mol Biol **147:**195-7.

192. **Snyder, M. L., and M. K. Clarke.** 1950. Evaluation of the colorimetric (Snyder) test; comparison of positive color reactions with the Lactobacillus counts of respective specimens of saliva routinely submitted for culture. J Dent Res **29:**298-303.

193. **Song, Y. L., N. Kato, Y. Matsumiya, C. X. Liu, H. Kato, and K. Watanabe.** 1999. Identification of and hydrogen peroxide production by fecal and vaginal lactobacilli isolated from Japanese women and newborn infants. J Clin Microbiol **37:**3062-4.

194. **Spear, G. T., M. Sikaroodi, M. R. Zariffard, A. L. Landay, A. L. French, and P. M. Gillevet.** 2008. Comparison of the diversity of the vaginal microbiota in HIV-infected and HIV-uninfected women with or without bacterial vaginosis. J Infect Dis **198:**1131-40.

195. **Steen, H. B.** 2000. Flow cytometry of bacteria: glimpses from the past with a view to the future. J Microbiol Methods **42:**65-74.

196. **Sundquist, A., S. Bigdeli, R. Jalili, M. L. Druzin, S. Waller, K. M. Pullen, Y. Y. El-Sayed, M. M. Taslimi, S. Batzoglou, and M. Ronaghi.** 2007. Bacterial flora typing with deep, targeted, chip-based Pyrosequencing. BMC Microbiol **7:**108.

197. **Sundquist, A., S. Bigdeli, R. Jalili, M. L. Druzin, S. Waller, K. M. Pullen, Y. Y. El-Sayed, M. M. Taslimi, S. Batzoglou, and M. Ronaghi.** 2007. Bacterial flora-typing with targeted, chip-based Pyrosequencing. BMC Microbiol **7:**108.

198. **Suy, A., P. Castro, M. Nomdedeu, F. Garcia, A. Lopez, E. Fumero, T. Gallart, L. Lopalco, O. Coll, J. M. Gatell, and M. Plana.** 2007. Immunological profile of heterosexual highly HIV-exposed uninfected individuals: predominant role of CD4 and CD8 T-cell activation. J Infect Dis **196:**1191-201.

199. **Swidsinski, A., W. Mendling, V. Loening-Baucke, A. Ladhoff, S. Swidsinski, L. P. Hale, and H. Lochs.** 2005. Adherent biofilms in bacterial vaginosis. Obstet Gynecol **106:**1013-23.

200. **Taha, T. E., D. R. Hoover, G. A. Dallabetta, N. I. Kumwenda, L. A. Mtimavalye, L. P. Yang, G. N. Liomba, R. L. Broadhead, J. D. Chiphangwi, and P. G. Miotti.** 1998. Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. AIDS **12:**1699-706.

201. **Tamrakar, R., T. Yamada, I. Furuta, K. Cho, M. Morikawa, H. Yamada, N. Sakuragi, and H. Minakami.** 2007. Association between Lactobacillus species and bacterial vaginosis-related bacteria, and bacterial vaginosis scores in pregnant Japanese women. BMC Infect Dis **7:**128.

202. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24:**1596-9.

203. **Tamura, K., M. Nei, and S. Kumar.** 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A **101:**11030-5.

204. **Tao, L., Pavlova, S., Anzinger, J., Carlson, S., Jacobs, A., Caffrey, M., Spear, G.** 2005. Fighting HIV with its natural enemy: Mannose-specific capture of HIV by Lactobacillus. Beneficial Microbes Conference, April 17-21, 2005, Poster Abstract #5B, accessed at http://www.asm.org/ASM/files/ccLibraryFiles/Filename/000000001988/Beneficial%20Microboes%20Program%20and%20Abstract%20Book.pdf, accessed on June 23, 2009.

205. **Thomason, J. L., S. M. Gelbart, R. J. Anderson, A. K. Walt, P. J. Osypowski, and F. F. Broekhuizen.** 1990. Statistical evaluation of diagnostic criteria for bacterial vaginosis. Am J Obstet Gynecol **162:**155-60.

206. **Thorsen, P., I. P. Jensen, B. Jeune, N. Ebbesen, M. Arpi, A. Bremmelgaard, and B. R. Moller.** 1998. Few microorganisms associated with bacterial vaginosis may constitute the pathologic core: a population-based microbiologic study among 3596 pregnant women. Am J Obstet Gynecol **178:**580-7.

207. **Tien, M. T., S. E. Girardin, B. Regnault, L. Le Bourhis, M. A. Dillies, J. Y. Coppee, R. Bourdet-Sicard, P. J. Sansonetti, and T. Pedron.** 2006. Anti-inflammatory effect of Lactobacillus casei on Shigella-infected human intestinal epithelial cells. J Immunol **176:**1228-37.

208. **Tomas, M. S., E. Bru, and M. E. Nader-Macias.** 2003. Comparison of the growth and hydrogen peroxide production by vaginal probiotic lactobacilli under different culture conditions. Am J Obstet Gynecol **188:**35-44.

209. **Tran, H. K., L. Chartier, L. X. Troung, N. N. Nguyen, A. Fontanet, F. E. Barre-Sinoussi, G. Pancino, and D. Scott-Algara.** 2006. Systemic immune activation in HIV-1-exposed uninfected Vietnamese intravascular drug users. AIDS Res Hum Retroviruses **22:**255-61.

210. **Trivedi, H. N., F. A. Plummer, A. O. Anzala, E. Njagi, J. J. Bwayo, E. N. Ngugi, J. E. Embree, and K. T. Hayglass.** 2001. Resistance to HIV-1 infection among African sex

workers is associated with global hyporesponsiveness in interleukin 4 production. FASEB J **15:**1795-7.

211. **Turk, W. J., X. J. Mao, T. Bielawny, R. Brunham, M. Luo, and F. A. Plummer.** 2008. Identification of novel human leukocyte antigen G alleles in an East African population by high-resolution sequence-based typing. Hum Immunol **69:**426-9.

212. **Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon.** 2007. The human microbiome project. Nature **449:**804-10.

213. **Tzianabos, A. O., Wetzler, L.M.** 2004. Cellular Communication, p. pp. 343-369. *In* G. B. Pier, Lyczak, J.B., Wetzler, L.M. (ed.), Immunology, Infection and Immunity. ASM Press, Washington DC.

214. **UNAIDS.** 2009. AIDS epidemic update: November 2009. World Health Organizatino.

215. **Vaahtovuo, J., M. Korkeamaki, E. Munukka, M. K. Viljanen, and P. Toivanen.** 2005. Quantification of bacteria in human feces using 16S rRNA-hybridization, DNA-staining and flow cytometry. J Microbiol Methods **63:**276-86.

216. **Van de Perre, P., D. Rouvroy, P. Lepage, J. Bogaerts, P. Kestelyn, J. Kayihigi, A. C. Hekker, J. P. Butzler, and N. Clumeck.** 1984. Acquired immunodeficiency syndrome in Rwanda. Lancet **2:**62-5.

217. **van de Wijgert, J. H., C. S. Morrison, J. Brown, C. Kwok, B. Van Der Pol, T. Chipato, J. K. Byamugisha, N. Padian, and R. A. Salata.** 2009. Disentangling contributions of reproductive tract infections to HIV acquisition in African Women. Sex Transm Dis **36:**357-64.

218. **Vasquez, A., T. Jakobsson, S. Ahrne, U. Forsum, and G. Molin.** 2002. Vaginal lactobacillus flora of healthy Swedish women. J Clin Microbiol **40:**2746-9.

219. **Veal, D. A., D. Deere, B. Ferrari, J. Piper, and P. V. Attfield.** 2000. Fluorescence staining and flow cytometry for monitoring microbial cells. J Immunol Methods **243:**191-210.

220. **Verhelst, R., H. Verstraelen, G. Claeys, G. Verschraegen, J. Delanghe, L. Van Simaey, C. De Ganck, M. Temmerman, and M. Vaneechoutte.** 2004. Cloning of 16S rRNA genes amplified from normal and disturbed vaginal microflora suggests a strong association between Atopobium vaginae, Gardnerella vaginalis and bacterial vaginosis. BMC Microbiol **4:**16.

221. **Verstraelen, H., R. Verhelst, G. Claeys, M. Temmerman, and M. Vaneechoutte.** 2004. Culture-independent analysis of vaginal microflora: the unrecognized association of Atopobium vaginae with bacterial vaginosis. Am J Obstet Gynecol **191:**1130-2.

222. **Westrom, L., G. Evaldson, and K. K. Holmes.** 1984. Taxonomy of vaginosis: bacterial vaginosis - a definition. Scand J Urol Nephrol Suppl **86:**259-64.

223. **Wilks, M., R. N. Thin, and S. Tabaqchali.** 1984. Quantitative bacteriology of the vaginal flora in genital disease. J Med Microbiol **18:**217-31.

224. **Wilson, M.** 2005. Microbial Inhabitants of Humans: Their ecology and role in health and disease. Cambridge University Press, Cambridge, UK.

225. **Worobey, M., M. Gemmel, D. E. Teuwen, T. Haselkorn, K. Kunstman, M. Bunce, J. J. Muyembe, J. M. Kabongo, R. M. Kalengayi, E. Van Marck, M. T. Gilbert, and S. M. Wolinsky.** 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature **455:**661-4.

226. **Yeaman, G. R., A. L. Howell, S. Weldon, D. J. Demian, J. E. Collins, D. M. O'Connell, S. N. Asin, C. R. Wira, and M. W. Fanger.** 2003. Human immunodeficiency virus receptor and coreceptor expression on human uterine epithelial cells: regulation of expression during the menstrual cycle and implications for human immunodeficiency virus infection. Immunology **109:**137-46.

227. **Zheng, J., Y. Xie, R. Campbell, J. Song, R. Q. Wang, R. Chiu, J. Berenson, M. Razi, S. Massachi, O. O. Yang, I. S. Chen, and S. Pang.** 2006. gp120-independent HIV infection of cells derived from the female reproductive tract, brain, and colon. J Acquir Immune Defic Syndr **43:**127-36.

228. **Zhou, X., S. J. Bent, M. G. Schneider, C. C. Davis, M. R. Islam, and L. J. Forney.** 2004. Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. Microbiology **150:**2565-73.

229. **Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp, and D. D. Ho.** 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature **391:**594-7.