Intestinal Microflora in the Diabetes-Prone Bio-Breeding Rat and its Possible Relationship to the Development of Type I Diabetes

by

Stéphanie McFadyen

A thesis submitted to

the Faculty of Graduate Studies and Research

In partial fulfillment of

the requirements for the degree of

Master of Science

Department of Biology

Carleton University

Ottawa, Ontario

December 1, 2009

Your file Votre référence
ISBN: 978-0-494-64430-0
Our file Notre référence
ISBN: 978-0-494-64430-0

Abstract

Bacteria play an important role in the development of the gut immune system and the development of gastro-intestinal disease, yet relatively few studies have focused on the role of the intestinal flora in the outcome of Type 1 diabetes. There is some evidence that exposing the Bio-Breeding diabetes-prone rat (BB rat) to food and bacterial antigens early in life delays and even suppresses later development of diabetes. This study used a culture-independent approach based on sequence variability in the 16S rDNA gene to analyse bacterial communities along different parts of the gastro-intestinal tract of the BB rat and their possible relationship to the development of diabetes. Several primers sets were evaluated to identify potential bias caused by primer length and PCR cycles.

Primers for variable regions 6-8 of the 16S rDNA gene were found to accurately represent the distribution of bacteria when compared to full-length sequence data. We observed a difference in the gut bacterial communities between control and diabetic-prone BB rats. An uncultured bacterium was predominant in the lymph nodes of the young diabetic-prone BB rat, but was absent from the lymph node of young control animals. Denaturing gradient gel electrophoresis profiling and non-metric multi-dimensional scaling (NMDS) showed age-related differences in the gut microbiota. NMDS and analysis of the shared operational taxonomic units suggest that the gut-associated lymphoid tissue is sampling bacteria from both the small intestine and the distal colon. The age and condition associated differences found between the experimental groups suggest that the intestinal flora has a role in the development of type 1 diabetes. Factors influencing composition of the intestinal flora could be a target for studies of therapeutic interventions.

Acknowledgements

This work could no have been completed without the support of many people.

Thank you to my thesis co-advisors Dr. Steve Brooks and Dr. Kenneth Storey for their valuable guidance, boundless enthusiasm and infinite patience.

I am very grateful to Judy Kwan and Morgan McAllister for skillfully teaching me the methods applied in this study. Their knowledge of laboratory technique is matched only by their impressive grasp of pop culture.

My thanks to my colleagues at Health Canada, Michèle Giddings, Will Robertson, and Dr. John Cooper, who gave me the confidence and the flexibe work schedule that allowed me to pursue my education.

I am deeply grateful to my mother, Darlene Powell, for her words of encouragement and the endless hours spent babysitting my children so that I could spend endless hours on this research.

To my husband Shawn, thank you for believing in me; there are no words...

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BBc	control BioBreeding rat
BBdp	diabetic prone BioBreeding rat
CFU	colony forming unit
CM	cow's milk
DAPI	4'-6-diamidino-2-phenylindole
DGGE	denaturing gradient gel electrophoresis
EMC	encephalomyocarditis
FBBG	filamentous brush border glycocalyx
Fc19	faeces control 19 day old
Fc55	faeces control 55 day old
Fd19	faeces diabetic prone 19 day old
Fd55	faeces diabetic prone 19 day old
Fiaf	angiopoeitin-like protein 4
FISH	fluorescent in-situ hybridisation
GAD	glutamic acid decarboxylase
GALT	gut-associated lymphoid tissue
GC	guannine, cytosine
GI	gastro-intestinal
IA-2	tyrosine phosphatase
IAA	insulin autoantibodies
ICA	islet cell autoantibody
Ig	immunoglobulin
Kda	Kilodalton
KRV	Kilham rat virus
LNc19	lymph nodes control 19 day old
LNc55	lymph nodes control 55 day old
LNd19	lymph nodes diabetic prone 19 day old
LNd55	lymph nodes diabetic prone 55 day old
Lyp	lymphopenia
MAC	microflora associated characteristics
MHC	major histocompatibiity complex
MLN	mesenteric lymph nodes
ng	nanogram
NMDS	non-metric multi-dimensional scaling
NOD mice	non-obese diabetic mice
OTU	operational taxonomic uni
PCR	polymerase chain reaction
PSA	polysaccharide A
rDNA	deoxyribonucleic acid
RNA	ribonucleic acid
SCFA	short chain fatty acid
SFB	segmented filamentous bacterium
SIc19	small intestine control 19 day old
SIc55	small intestine control 55 day old
SId19	small intestine diabetic prone 19 day old
SId55	small intestine diabetic prone 55 day old
SSU	small sub-unit
ST2	streptozotocin
T-RFLP	terminal fragment length polymorphism

Chapter I

Introduction and Objectives

Insulin-dependent diabetes mellitus, known as type 1 diabetes, affects approximately 0.2 to 0.6% of the population of developed countries (Onkamo et al 1999). Type 1 diabetes is an autoimmune disease that develops from complex interactions between the environment and several genes in susceptible individuals. This form of diabetes usually strikes children and young adults, although disease onset can occur at any age. The disease is the result of a chronic inflammatory process that progressively destroys most of the insulin-producing β-cells in the pancreatic islets of Langerhans. This autoimmune process involves progressive infiltration into the β-cell-containing core of the islets by mononuclear cells and macrophages, a process called insulitis. Insulin is the only known β-cell specific auto-antigen in type 1 diabetes, although other autoantibodies against islet antigens GAD65 (the M_r 65000 isoform of glutamic acid decarboxylase) and IA-2 (tyrosine phosphatase) serve as good predictors of the development of type 1 diabetes in humans (Hampe et al 1999, Roep and Atkinson 2004). Insulin autoantibodies (IAAs) are commonly found in children with newly diagnosed type 1 diabetes, and predict the disease when combined with islet cell autoantibodies (ICAs). In a prospective birth-cohort study, IAAs most frequently appeared as the first antibody in the ICA-positive offspring of diabetic parents, suggesting that immunization to insulin may be an early event in the autoimmune process leading to type I diabetes (MacFarlane *et al*, 2003). The severity and prevalence of insulitis or its sequelae (end stage islets) reflect the extent of the damage in the pancreas (MacFarlane *et al* 2003; Yoon and Jun 2004). To survive, people with type 1 diabetes must have insulin delivered by injection or a pump.

Genetic susceptibility to Type I diabetes is inherited. Human leukocyte antigens (HLA) genetics have a major role in the etiology of type 1 diabetes, and the HLA on chromosome 6 is believed to contribute about half of the familial basis for the disease (Gillespie, 2006). Those that carry the HLA-DQB *02/*0302 genotype or the *0302/x genotypes, where x stands for alleles other than *02, *0301, *0602, or *0603, are at an increased genetic risk of type 1 diabetes (Vaarala *et al*, 1999). Other genes also contribute to the genetic effect: the insulin gene on chromosome 11 has been identified through candidate studies as the second most important genetic susceptibility factor (Gillespie, 2006). It is interesting to note that children of mothers with type 1 diabetes are 50% less likely to develop the condition than children of diabetic fathers (Koczwara et al, 2004). While the genetic effect contributes 70-75% of the susceptibility to type I diabetes, there is evidence that environmental factors strongly influence expression of the disease.

Environmental factors may initiate or trigger the process that leads to the destruction of the β-cells and the onset of diabetes (MacFarlane et al 2003). The incidence of diabetes varies considerably among countries, which supports the theory that environmental factors have an important effect on disease development (Onkamo et al 1999). An interaction between susceptible genes and unknown environmental factors is also supported by estimates of increases in the incidence of type I diabetes. Onkamo et al. (1999) reported a 3% annual increase in global incidence of the disease from the years 1960-1996. The confidence intervals reported for these estimates are fairly narrow, however other factors that could influence reporting are not well controlled over the years. Despite this, the data suggests a significant trend. It is particularly interesting that the incidence is globally increasing in the age group 0-14 years. The authors calculated

that the incidence of type 1 diabetes will be 40% higher in 2010 than in 1998, and concluded that it is unlikely that a corresponding increase in the proportion of subjects with genetic susceptibility to type I diabetes would take place in the population in such a short time. Environmental influences on the susceptible genes are suspected (Onkamo et al 1999).

Studies in animals have shown that environmental factors can modify the development of type 1 diabetes (Malaisse et al, 2004, Guberski et al, 1991, Vaarala et al, 1999, Menser et al, 1978). Current thinking is that environmental factors, either alone or in combination, trigger an autoimmune process in a genetically susceptible individual. The autoimmune process is localized in the pancreas and leads to the destruction of the insulin-producing β-cells. The most commonly implicated environmental factors include dietary factors, viral infections, toxins, ante- and perinatal risk factors, exposure to chemicals, and stressful life events (Akerblom et al, 2002). There are several reasons for the difficulty in identifying specific environmental factors that may promote the development of the disease (MacFarlane et al 2003). These include:

a) Difficulty in linking past exposures to the development of the disease;

b) Lack of knowledge of the environmental antigens; and,

c) The large number of predisposing genes in individuals at risk.

Most environmental antigens enter the body through the mucosal interface of the gut. Antigens are carried from the mucosal surface to the draining lymph nodes, where specialized antigen presenting cells then present them to the immune system (Akerblom et al, 2002). Gut-associated lymphoid tissue must recognize harmless antigens present in food or commensal bacteria, and differentiate between these and pathogenic assault.

There is evidence that the regulation of the gut immune system is deviant in type 1 diabetes (Malaisse et al, 2004). Bacteria play an important role in the development of the gut immune system (Guarner and Malagelada, 2003; Macpherson and Harris 2004; Macpherson and Uhr 2004a, 2004b; Suzuki *et al,* 2007) and gastrointestinal disease, yet relatively few studies have focussed on the role of gut microbiota in diabetes outcome.

It is not known whether the absence of specific commensal bacterial species, or the presence of specific pathogens in the gut microflora predisposes an individual to long-term health problems such as type 1 diabetes (Mazmanian and Kasper, 2006). An increasing amount of data is showing that bacterial populations in patients with various gastro-intestinal (GI) related diseases are different from those in healthy subjects. The faecal microflora of Crohn's disease patients was found to contain an increased level of enterobacteria (*E. coli* group) and a high proportion of bacteria (30%) not belonging to defined phylogenic groups, compared to healthy volunteers (Seksik *et al,* 2003). One study in BioBreeding (BB) rats found that the faecal flora in the rat differs before the onset of diabetes in control versus diabetic prone animals, suggesting a role in the pathology of diabetes (Brugman et al, 2006). However, another study that compared the microflora-associated characteristics (MACs) in children with newly diagnosed diabetes with those of 27 healthy controls found no differences in MACS between diabetics and controls. Concentrations of 8 different short-chain fatty acids (SCFAs) were examined, looking for both differences between children and any familial patterns. Members of families with a diabetic child had a high concentration of acetic acid and lower concentration of several other SCFAs than control families, yet no statistically significant differences between concentrations of SCFAs in diabetic versus control children were

found (Samuelsson and Ludvigsson, 2003). More research is needed to determine what, if any, role gut microbiota may play in type I diabetes.

The objective of this study is to explore the hypothesis that there is a relationship between the development of type 1 diabetes and the composition of intestinal microflora. If this relationship exists, there should be a consistent difference in bacterial populations in Bio-Breeding control (BBc) and Bio-Breeding diabetic prone (BBdp) rats. The null hypothesis is that there is no consistent difference in bacterial populations in Bio-Breeding control (BBc) and Bio-Breeding diabetic prone (BBdp) rats.

The null hypothesis was tested by:

1. Surveying the composition and translocation of the intestinal microflora in control and diabetic-prone BioBreeding rats (animal model for type 1 diabetes) at two age points (shortly after weaning and just prior to onset of diabetes); and
2. Based on the above survey, determining microbial communities and structures and the phylogenic relationships between isolates and identifying similarities and differences.

During the course of the study, there was some concern that the primer sets were affecting the results. A third objective was added to assess this effect:

3. Evaluate different primer sets and identify bias that may be introduced by the selected primers.

Chapter II

Review of the Literature

2.1 Diabetes in Canada

In 2005-2006, approximately 1.9 million Canadians, or about one in 17 people had been diagnosed with diabetes (PHAC, 2008). In the Aboriginal communities of Canada, the disease prevalence is estimated to be 3 to 5 times that national average (PHAC, 2005). Currently, there are limitations in the physician billing data and the hospital discharge abstract data in identifying type 1 and type 2 diabetes. These limitations make it impossible to distinguish between diabetes types in the reported rates. However, the Public Health Agency of Canada (PHAC) estimates that in adults, type 1 diabetes accounts for 5 to 10 percent of all diagnosed cases of diabetes (PHAC, 2005). PHAC also estimates that by 2011, the number of Canadians with diagnosed diabetes will be approximately 2.6 million. This represents an average annual increase of almost 7% and an increase of about 33% from 2006. Older Canadians are more likely to have diabetes, with 22% of of the 75-79 age group diagnosed with diabetes, compared to about 2% of individuals in their 30s (PHAC, 2008). There is considerable geographic variation in the global annual incidence of diabetes; Canada has among the highest incidence of type 1 diabetes in the world (Figure 2.1).

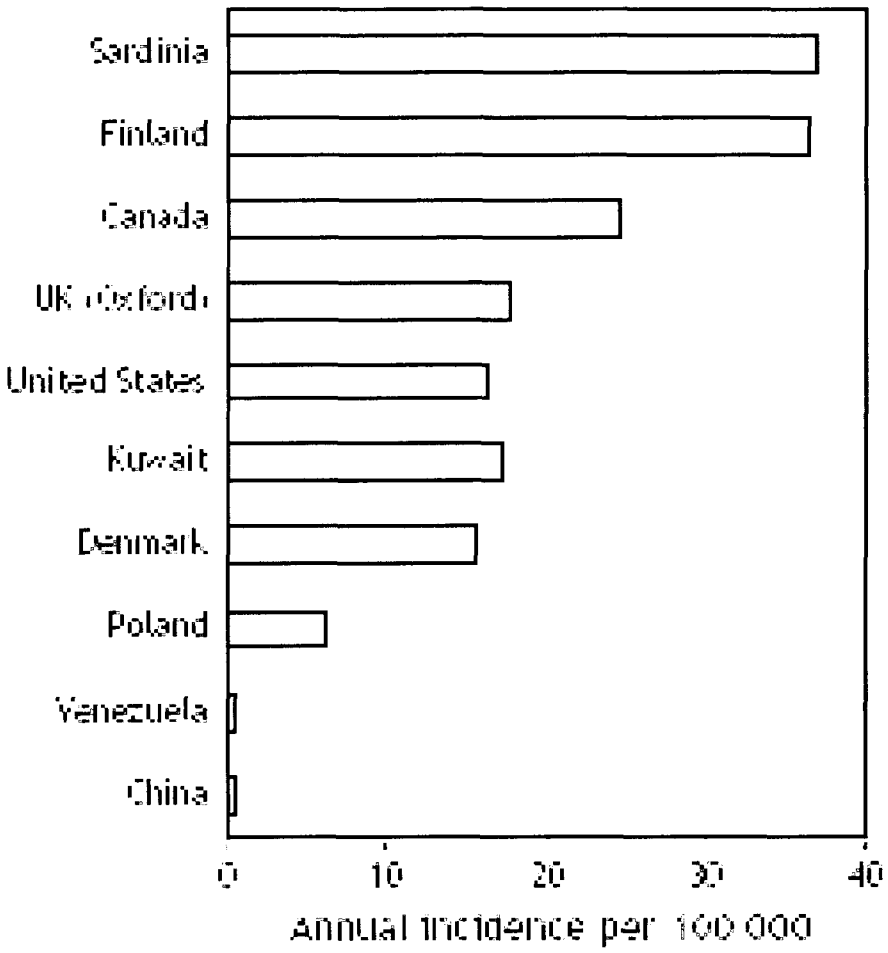


Figure 2. 1: Geographic variation in annual incidence of Type 1 diabetes

*(reproduced from Gillespie, 2006)

2.2 Diabetes and the gut immune system

It has been speculated that regulation of the gut immune system may be aberrant in type 1 diabetes and thus linked to the pathogenesis of the disease (Malaisse *et al,* 2004). It is known that antigens enter the body through the mucosal interface of the gut. The microenvironment in these mucosal barriers has a marked influence on the immune response. Gut-associated lymphoid tissue (GALT) must distinguish between pathogenic assault and harmless antigens that are present such as commensal bacteria and food antigens. There are several theories of how a change in some aspect of the immunological reaction of the body in response to the contents of the gut lumen may underlie the pathogenesis of certain diseases. These theories include:

- Inappropriate up-regulation or activation of a single component of the complex spectrum of gut immune cells and /or mediators;
- Failure of one or more of the inhibitory signals that normally down-regulate and switch off an appropriate, protective local inflammatory reaction when it is no longer necessary;
- An incomplete immune response (immunodeficiency), allowing persistence of a living or non-living antigen in the gut - even if this is a very small quantity of the organism, the ineffectual immune response will continue;
- Intestinal allergy, with tissue damaging immune reactions to harmless agents such as food constituents or additives. This may be IgE mediated, T-cell mediated (i.e delayed hypersensitivity) or multifactorial;

- Autoimmunity, the development of immune reactions to self-antigens – an example could be molecular mimicry, where there is a partial antigenic identity between a bacterial antigen and a tissue antigen.

The two most distinctive functions of the GALT are secretion of IgA and the generation of helper T cells (Akerblom *et al,* 2002). Immune responses mediated by T cells are dependent on the cell surface expression of the major histocompatibility complex (MHC) proteins. The normal phenotype of β-cells is low Class 1 MHC expression with no detectable Class II expression in humans, rats and mice. A hyperexpression of Class I MHC molecules is recognised as a key feature in newly diagnosed Type I diabetes patients, BBdp rats, and NOD mice (Scott, 1996).

IgA is produced in enormous quantity at the mucosal interface, and is a "non-inflammatory" form of immunoglobulin, in that it binds complement on the surface of the epithelium without inducing inflammation or tissue disease. IgA provides protection against bacterial, viral and parasitic mucosal pathogens, and the development of IgA-producing B cells is dependent on microbial colonization (Suzuki *et al*, 2007). The IgG class of antibodies, on the other hand, neutralize antigens within the tissue and form IgG immunocomplexes. As a result, an inflammatory response is initiated. This response can be harmful to the host tissue (Guarner *et al*, 2002). The predominance of IgA in the gut suggests that IgA plays a key role in selecting and maintaining a spatially diversified gut bacterial community (Suzuki *et al*, 2007).

Afferent lymphatics carry antigens from the mucosal surface to draining lymph nodes where specialized antigen-presenting cells present them to the immune system. The organised lymphoid tissues of the respiratory and gastrointestinal tracts contain the

largest numbers of lymphocytes. Although it consists of only a single layer of cells, the intestinal epithelium must both control the access of potential antigens (including highly immunologic and potentially immunostimulatory commensal bacterial flora) and pathogens, and function in the digestive absorption of dietary nutrients (Ferguson 1996; Nagler-Anderson, 2001). It is aided in this dual role by intercellular tight junctions that restrict the passage of even small (2 kDa) molecules. The apical surface of the enterocyte, which faces the intestinal lumen, is ideally suited for the terminal digestion of nutrients because of its dense coating with absorptive microvilli. At the tip of the microvilli is a layer of membrane-anchored glycoproteins, the filamentous brush border glycocalyx (FBBG). The FBBG will allow passage of digested nutrients but is relatively impermeable to macromolecules or bacteria (Nagler-Anderson, 2001).

Yet bacteria do cross this barrier; how does this occur? Specialized epithelial cells known as M cells are interspersed with the enterocytes in the epithelial monolayer. The M cells lack the FBBG, which is replaced by microfolds on the apical surface – these are more accessible to luminal antigens. M cells use transepithelial vesicular transport to carry bacteria and viruses to antigen-presenting cells in the underlying GALT. Macrophages and dendritic cells then take up the microbiota in the subepithelial dome. Invasiveness is not required for M-cell transport. For example, invasive (*Yersinia pseudotuberculosis, Yersinia enterocolitica, Salmonella typhirium* and *Shigella flexneri*) and non-invasive (*Vibrio cholerae* and enteropathogenic *E. coli*) bacteria and viruses (reovirus) are bound and transported by M cells (Nagler-Anderson, 2001). M cells are contained within discrete inductive sites known as Peyer's patches. Peyer's patches are organized follicular structures, consisting of aggregated lymphoid follicles found

primarily in the distal ileum of the small intestine. M cells are thought to act as gateways to the mucosal immune system, delivering antigen to the antigen presenting cells in the Peyer's patch subepithelial dome. There is also evidence that dendritic cells may extend dendrite-like processes through epithelial tight junctions and sample luminal antigen directly. The integrity of the barrier is maintained during this process because tight junctions are re-formed by proteins that are expressed on both enterocytes and the dendritic cells. Challenge studies have shown that while most commensal bacteria are killed within hours by macrophages, the bacteria can survive for up to 60 hours within dendritic cells, but do not penetrate futher than the mesenteric lymph nodes (Macpherson and Harris, 2004; MacPherson and Uhr, 2004b).

The Peyer's patches are the most discernible of the follicles dispersed throughout the small and large intestine. Together with isolated lymphoid follicles scattered throughout the intestine, Peyer's patches are the major inductive sites of the small intestine. Like lymph nodes, Peyer's patches have B-cell follicles and germinal centres surrounded by areas that contain predominantly T-cells. Intestinal IgA plasma cells are generated in organized structures such as Peyer's patches, isolated lymphoid follicles and mesenteric lymph nodes after antigenic stimulation and induction of germinal centres, as well as in diffuse gut tissues (the lamina propria) (Suzuki *et al*, 2007). The links between the gut immune response, dietary antigens and diabetes is explored further in the following pages.

2.3 Dietary factors in the incidence of diabetes

The two most heavily studied environmental factors in the etiology of Type I diabetes are diet and viruses. Spontaneous diabetes in animal models (the BioBreeding rat

and Non-obese diabetic mice) occurs most frequently when the animals are raised in ultraclean conditions; even gnotobiotic animals develop diabetes (Malaisse *et al*, 2004). Both major animal models of spontaneous type 1 diabetes, the BioBreeding (BB) rat and the non-obese diabetic (NOD) mouse show a strong link, (i.e at least half of all cases) to diet. This strongly supports diet and intestinal events as the major environmental factor in promoting the development of type 1 diabetes. The incidence of type 1 diabetes varies worldwide with dietary patterns; there is a significant positive correlation of diabetes incidence with the average daily intake of energy from products of animal origin (meat; $r = 0.55$, $P < 0.001$, dairy; $r = 0.80$, $P < 0.0001$) and an inverse correlation ($r = -0.64$, $P < 0.001$) with intake of energy from vegetable products (Muntoni *et al*, 2000). In healthy children systemic immune responses to food antigens develop after antigen exposure but these responses seem to decline with increased age. Thus it seems likely that a number of different food types might be diabetogenic, and that some susceptible people are unable to handle dietary antigens in a normal manner.

One way in which diet could influence diabetes is that an abnormal gut immune response to dietary antigens somehow leads to the automimmune response that develops against pancreatic β cells (Graham *et al,* 2004). This is supported by the fact that increased intestinal permeability has been reported in patients with type 1 diabetes, and in BBdp rats (Scott *et al,* 2002). In addition, some human patients with type 1 diabetes show high reactivity to the gluten component of wheat similar to that found in coeliac disease, with associated elevated inflammatory activity in the jejunum (Hardin *et al,* 2002, Graham *et al,* 2004). Abnormalities in the gut of BBdp rats include a greater percentage of goblet cells and greater mucosal depth than control animals. Both control

and diabetic-prone BB rats expressed less of the tight junction protein claudin, and showed greater intestinal permeability than did Wistar rats (Neu *et al,* 2005). Interestingly, Scott *et al* (2002) found that exposing BBdp rats to antigens (i.e. wheat-gluten or cereal based diets) in the first week of life, before gut closure and when the mucosa is immature, resulted in a delay or even avoidance of diabetes in approximately one-third of the animals.

Bovine proteins have been a central focus of research, but there are contradictory findings with some studies showing an association between early exposure to cow's milk, while others show no influence (Beales *et al*, 2002; Scott, 1996; Vaarala *et al*, 1999; Truswell, 2005). Enhanced humoral and cellular immune responses to several cow's milk proteins have been documented in patients with newly diagnosed type 1 diabetes. Exposure to cow's milk (CM) formulas induces IgG-class antibodies and T-cell immunity to bovine insulin. The first immunization to insulin takes place in the gut immune system, when infants are fed with CM formulas containing bovine insulin. The possibility that lymphocytes sensitized to bovine insulin in early infancy will later mature toward autoreactive insulin-specific lymphocytes and lead to the destruction of β-cells has been proposed as a hypothesis in the pathogenesis of type 1 diabetes. Vaarala *et al* (1999) found that oral exposure to cow's milk formulas induced bovine insulin-binding antibodies that cross reacted with human insulin, and concluded that dietary bovine insulin appears to be an environmental trigger of a primary immune response to a β-cell specific antigen in healthy children. Low levels of antibodies binding to insulin in exclusively breast-fed children indicate that human insulin in breast milk is less immunogenic than bovine insulin. Enhanced immune response to several cow's milk

proteins have been reported in type 1 diabetes and could be considered markers of oral tolerance to bovine antigens (Vaarala *et al,* 1999). A trial in which hydrolyzed formula was used during the first six to eight months of life in infants at increased genetic risk of type 1 diabetes suggest that the elimination of cow's milk proteins during infancy decreases the development of β-cell autoimmunity during the first years of life (Akerblom *et al,* 2002). However, both diabetic and healthy children showed a similar immune Th1- and Th2- like response to several cow's milk proteins (Bovine Serum Albumin, the ABBOS peptide and β-lactoglobulin) (Karlsson *et al,* 2001). A critical review of the case for cow's milk and type 1 diabetes concluded that "there is no convincing or even probable evidence that the A1 β-casein of cow milk has any adverse effect on humans" (Truswell, 2005). In addition, a multi-center study demonstrated that a milk-free, wheat-based diet produced the highest frequency of spontaneous diabetes in diabetes-prone BB rat and NOD mice in three separate locations (Beales *et al,* 2002). This suggests that even if milk protein does contribute to the development of diabetes, it is likely that there would still be an appreciable prevalence of the disease in human populations from the consumption of other foods, particularly wheat cereals.

Defined diets in which wheat is the sole protein source are potent inducers of diabetes in BB rats; the same results are seen in the NOD mouse (Scott, 1996) although this effect can vary with the age at which the diet is introduced and the BB rat colony. When BB rats are fed a mostly cereal-based or wheat-gluten diet from weaning, 65.3 ± 6.1% of animals develop diabetes, as compared to <20% (incidence of 18.8 ± 2.8%) if rats receive a hydrolysed casein diet (Scott, 1996; Scott *et al,* 2002). Adult human diabetes patients show a high prevalence of undiagnosed wheat gluten sensitive

enteropathy (coeliac disease) (Not *et al,* 2001), with a prevalence of 2-10%, a rate that is 10 to 33 times that found in the general population (Graham *et al,* 2004). However, a study of Italian school children with coeliac disease found that these children had a low prevalence of diabetes-related antibodies (d'Annunzio *et al,* 2009). In some patients, T-cell response to wheat proteins is increased, and high concentrations of wheat antibodies in blood have been reported (Malaisse *et al,* 2004). There are reports that increased peripheral blood T-cell reactivity to wheat gluten is more frequent in newly diagnosed patients than in controls, consistent with the involvement of dietary wheat proteins in diabetes pathogenesis (Beales *et al,* 2002). Diabetes incidence in the Beales *et al* (2002) study varied with the feedings of diets that differed mainly with the source of dietary amino acids. The milk-free, mainly wheat-based NTP-200 diet introduced at weaning (i.e at 23 days of age for BB rats and 17-21 days of age for NOD mice) consistently produced the highest incidence of diabetes in both BB rats and NOD mice. Diabetes-prone BB rats have high levels of antibodies against dietary wheat gliadin (solvent-extracted protein from wheat gluten) compared to control rats; this effect has also been shown in NOD mice, although not consistently (Scott *et al,* 2002).

2.4 Viruses and type 1 diabetes

Although diet seems to be the most significant environmental factor in the development of type 1 diabetes, it is important to note that more than 10 viruses have been associated with the development of type 1 diabetes-like syndromes in humans or animals. Congenital rubella is one example of virus-induced diabetes in humans (Menser *et al,* 1978). However, effective immunization programs have eliminated the condition from most Western nations. The most thoroughly studied diabetes related virus in animals

is encephalomyocarditis virus (EMC). Another well-studied virus is the Kilham rat virus (KRV); this small DNA virus has been found to induce diabetes in control BB rats by provoking autoimmune responses against the β-cell rather than direct β-cell cytolysis. Kilham rat virus was found to have contaminated a colony of diabetes resistant BB rats that suddenly turned diabetic (Guberski *et al,* 1991; Yoon and Jun, 2004). It is speculated that enteroviruses may be involved, although a specific diabetes-inducing enterovirus has not yet been identified. During the viremic phase, enteroviruses can spread from the lymphoid tissues of the pharynx and small intestine to various organs including the pancreas. One link between enterovirus infection and diabetes is a seasonal variation in the onset of diabetes that follows that of enterovirus infections (Gamble and Taylor, 1969). In addition, enteroviruses have been isolated from patients with newly diagnosed type 1 diabetes (Frisk *et al,* 1992; Helfand *et al,* 1995). A prospective study in Finland, a country with an unusually high incidence of diabetes, found enterovirus infections more frequently in siblings that progressed to clinical diabetes than in control siblings that remained non-diabetic. Infections were also clustered around the time period immediately preceding the appearance of autoantibodies (Hyoty *et al,* 1995, Hiltunen *et al,* 1997).

Further evidence that viruses may be implicated in the onset of the disease is that β-cells of diabetic patients express interferon-alpha, a cytokine that is induced during viral infection. This suggests the presence of some virus in the β-cells (Ray *et al,* 2008). In one study, antibodies against coxsackie B serotypes were found to be more frequent in newly diagnosed subjects than controls. However, as previously noted, the highest incidence of spontaneous diabetes in BB rats and NOD mice occurs when they are maintained in ultra-clean conditions. Even gnotobiotic animals will develop diabetes. The

finding that animals maintained in strict viral antibody-free conditions still develop diabetes supports the premise that diet is a major environmental stimulus in the development of type I diabetes, but underscores that the impact of a specific environmental trigger in initiating the autoimmune response may be different in various individuals depending on genetic susceptibility or protection (MacFarlane *et al*, 2003). The viral links to type 1 diabetes also raise the possibility that environmental organisms may induce diabetes in susceptible humans by stimulating effector cells or by disrupting a balanced network of autoreactive and regulatory cells, a theory that is relevant to the study of the gut microbial community (Guberski *et al,* 1991).

2.5 Approaches to examining microflora

Despite the wide use of rat models in disease modeling, the rat intestinal microbiota remains poorly characterized (Bonnet *et al,* 2002; Bernbom *et al,* 2006). Much of the current knowledge regarding this microbial community is based on culture-dependent methods, and has focused on the easily accessible faecal flora. Conventional analysis of faecal flora requires meticulous techniques for cultivation, various growth media, and an array of methods for the taxonomic identification of the isolates. Thus, while it is relatively easy (assuming that the bacteria can be grown) to obtain a total viable count of bacteria, enumerating individual species by culture is both labour and time intensive (Vaughan *et al,* 2000). It can take weeks to identify a single anaerobe. In addition, the culture conditions and selective media are simply not known for most of the strict anaerobic bacteria. It is now commonly accepted that gut microflora investigated in detail by anaerobic culture techniques provide an incomplete picture of the diversity of species present in the gut flora. The percentage of cultivated faecal flora range greatly,

from 14% based on DAPI (4'-6-diamidino-2-phenylindole) staining (Wilson and Blitchington, 1996) to 58% based on probe hybridization (Langendijk *et al,* 1995). Microscopic counts on human faeces also suggest that between 40 – 80% of bacterial cells can not be recovered by culture (Suau *et al,* 1999). While bacterial cultures are very useful for detecting individual groups of bacteria, they are difficult to use in the analysis of complex bacterial communities such as those found in environmental samples (Vaughan *et al,* 2000). Despite these shortcomings, culture-based analysis is still considered a good method for microbiological study in general. However, molecular approaches are rapidly becoming the standard for studies of gut microbiota.

More recently, culture-independent methods based on comparative sequence analysis of randomly cloned bacterial ribosomal RNA (i.e. 16S rRNA) or other bacterial genes has informed the field of microbiology. Culture-independent methods are increasingly being applied to this field, and have already made significant contributions to our knowledge (O'Sullivan, 1999; Tannock, 1999). Molecular methods used in studies of complex microbial communities include amplification by PCR of 16S ribosomal RNA genes from microbial DNA, denaturing gradient gel electrophoresis (DGGE) and terminal restriction fragment length polymorphism (T-RFLP) among others (Tannock, 1999; Brooks *et al,* 2003; Bernbom *et al,* 2006). Fluorescent in situ hybridization (FISH) with SSU rRNA-targeted oligonucleotide probes is another culture-independent technique used in studies of complex microbial communities of environmental (Amann et al 1990, Rigottier-Gois et al 2003, Lenaerts et al 2007) and clinical (Hogardt et al 2000) samples. This method combines hybridization of fluorescently labelled SSU rRNA probes with epifluorescent microscopy (Amann et al 1990), confocal laser microscopy or flow

cytometry to directly quantify individual bacteria (Zoetendal et al 2004). The 16S RRNA gene has been a common target (Lay et al 2007, Lenaerts et al 2007). FISH can be used to detect and enumerate individual cells, and with automation can be used for comparative analysis (Zoetendal et al 2004, Lenaerts et al 2007). FISH has been found to be highly specific, but only moderately sensitive, with detection limits of 10^5-10^6 CFUs/ml (Hogardt et al 2000, Zoetendal et al 2004). Since FISH targets only specifi species or groups, some preexisting knowledge about which species may be present, and which primers will target these species, is required. Thus, this technique may not capture the full diversity of the gut ecosystem.

Equally important to the molecular tools used in culture-independent approaches is the use of statistical tools that describe and compare microbial communities (Bonnet *et al,* 2002; Schloss and Handelsman, 2006a; 2006b).

The rRNA genes are ideally suited for determining phylogenic relationships, assessing environmental diversity, and detecting specific populations. This is because they are universally distributed in cellular life forms, as well as being highly constrained in function, with stretches of conserved sequence between more variable regions. These variable regions can be exploited by designing PCR primers to probe the rRNA genes with varying levels of taxonomic specificity (Acinas *et al,* 2004). The rRNA sequences can be clustered into relatedness groups based on their percent sequence identity, thus molecular methods address both detection and classification, often to the level of species (but not strain) (Brooks *et al,* 2003; Brooks *et al,* 2007). The ribosomal RNA genes encoding 16S, 23S and 5S rRNAs are often found linked together with tRNA molecules into operons (Klappenback *et al,* 2001). Sequencing of the 16S rRNA is the most

commonly used measure of environmental diversity. Consisting of about 1500 nucleotides, the 16S rRNA provides a large amount of information for a phylogenic study yet remains a reasonable size to sequence. Other avantages to the 16S gene is that horizontal transfer of the gene appears to be weak, thus making it a good monitor of species differences (Brooks *et al,* 2007). The 5S rRNA and the 23S rRNA have also been studied. The former is somewhat small thus offers a limited amount of information, while the latter, at approximately 3000 nucleotides, provides a great deal of information but requires more sequencing and increased risk of transcription error. Premature termination of transcription is frequent when transcribing stretches of longer than 400 bases (Vaughan *et al,* 2000). Based on genetic sequence information, a new species is considered to possess less than 97% sequence similarity with any known organism in a database (Stackebrandt and Goebel 1994). Comprehensive data sets of greater than 300,000 16S rRNA sequences (a number that is estimated to double every 7 months) are available through public internet-accessible databases such as GenBank, EMBL and the RDP II project (Maidak *et al,* 2001, Rudi *et al,* 2007). It must be noted that corrupt or chimeric sequences are known to be found in the public repositories. One estimate is that as many as 1 in 20 sequences have substantial errors, which could lead to a flawed picture of microbiological diversity (Ashelford *et al,* 2005). Despite these cautions, applying these approaches to microbial communities from the rumen, the human intestine, the pig intestine and mucosa, and the mouse intestine has demonstrated the diversity of the populations and indicated that the majority of the bacteria that make up these communities have yet to be cultured (Brooks *et al,* 2003). Overall, molecular techniques greatly improve the ability to analyze complex microbiological communities, and

promise a more full and accurate picture of the diversity, structure and dynamics of these communities. Table 2.1 summarizes the advantages and disadvantages of culture-based and molecular methods.

The polymerase chain reaction (PCR) is the basis for molecular sequence analysis. However, it is recognized that there are shortcomings to PCR-based analysis, as the process may introduce biases that affect the composition of clone libraries (Bonnet *et al,* 2002). The number of cycles carried out by PCR has been found to have a strong influence on the proportion of genes amplified. In comparing clone libraries generated by 10-cycle and 25-cycle PCR reactions, Bonnet *et al,* (2002) found that the OTU richness of the 25-cycle community was significantly lower. Increasing the number of PCR cycles can also lead to reannealing of major PCR products compared to template-to-product ratios (Suzuki and Giovannoni, 1996). It is therefore preferable to keep the number of PCR cycles as close to 10 as possible to reduce this bias. Other factors that can influence the results of a PCR reaction include variations in primer pairs, the melting temperature of the template DNA, exhaustion of substrates during the reaction, differing amplification efficiencies of amplicons, and chimera formation (Vaughan *et al,* 2000). It has also been noted that genome size as well as the number and the organization of *rrn* operons vary widely among prokaryotes. The *rrn* operon number can increase divergence between species by approximately 2%, which influences species cut-off when determining OTUs. As well, the number of genes coding for 16S rRNA can vary from 1 to 14; this has important implications for attempts to quantify the abundance of bacterial species using PCR methods (Farrelly *et al,* 1995).

The molecular analysis of samples can offer advantages over culture methods, but the purity of DNA extracted from faecal samples is a critical factor in the sensitivity and usefulness of biological analyses such as PCR for microbiological organisms (Vaughan *et al,* 2000). In comparison with blood or other clinical samples, the challenges inherent in DNA extraction from faecal samples include a complex microbial flora, as well as the variable consistency and endogenous and dietary components of faeces. Differences in the envelope morphology of Gram + groups of bacteria make many of them more difficult to disrupt than the Gram – component of the community (Brooks *et al,* 2007). Phenol-chloroform extraction is the traditional method of bacterial DNA extraction; this may require additional clean-up procedures to remove remaining phenols which inhibit PCR reactions. Other methods involve the lysis of bacteria and subsequent binding of released DNA onto a solid matrix, followed by washing and elution of the relevant components. These involve high labour costs and time constraints in processing numerous samples.

Table 2. 1: Pros and Cons of Culture and Molecular Methods

Method	Pros	Cons
Culture	Long history and experience	Labour and time intensive
	Basis of many studies	Anaerobic culture conditions unknown for many bacteria
	Accepted as a high standard	Multiple steps needed for reliable identification
	Relatively easy to obtain viable cell counts	Difficult to use for analysis of complex communities
Molecular	Useful for detecting individual groups of bacteria Very useful for analysing complex communities and their phylogenic relationships	Outcome of analysis can be influenced at multiple points in the molecular analysis (i.e. DNA extraction method, PCR conditions, primer selection, etc.)
	Provide a more complete picture of bacterial diversity Results can be achieved rapidly	Limited ability to infer function from gene sequence Multiple rrn operons and gene copies lead to overestimation of diversity

Commercial kits have been developed and are now widely used for DNA extraction specifically from faeces. Faeces contain several inhibitors of PCR reactions, such as bile salts, haemoglobin degradation products and complex polysaccharides. A comparison of 4 commercial kits found that QUIamp was the most effective extraction method for faecal samples (McOrist *et al,* 2002). Repetitive testing confirmed the superior performance of this kit over a range of spike concentrations. The extraction of bacterial DNA from faeces using this kit resulted in superior downstream performance in PCR compared to other methods under similar conditions. A similar comparison of faecal DNA kits in the Brooks laboratory also found that the QUIamp kit, in conjunction with three freeze-thaw/sterile sand, mortar and pestle grinding cycles, produced good results (Brooks *et al,* 2007).

Bacterial communities are commonly compared using denaturing gradient gel electrophoresis (DGGE), an electrophoretic method of seperating equal length DNA fragments. DGGE is based on sequence-dependent melting behaviours in a polyacrylamide gel containing a concentration gradient of increasing denaturant (Ercolini, 2004). DNA fragments of the same length but with different base-pair sequences can be separated based on the mobility of the partially melted DNA molecule (Muyzer *et al,* 1993). DGGE is a useful technique in which the diversity of PCR-amplified genes from a large number of samples can be compared in one gel, to evaluate changes in community structure over time or space. A 30 to 40 base GC clamp at the 5' end is necessary to prevent the two DNA strands from completely dissociating even under strong denaturing conditions (Muyzer *et al,* 1993; Ercolini, 2004). In theory, the technique is sensitive enough to separate DNA on the basis of a single point mutation; in

practice, the high similarity of rDNA sequences can make it difficult to resolve and identify closely related species. Often identification is limited to the genus (Vaughan *et al,* 2000; Acinas *et al,* 2004). A comparison of phylogenic distance (percent base difference) and the mobility differences by DGGE of sequences found that large differences in band mobility on the DGGE gel imply at least a minimal amount of sequence divergence. However, sequences with very similar mobility are not always identical (i.e sequences can still differ considerably despite identical mobility). Despite this, in most cases bands with equal mobility are found to have identical sequences. Multiple gene copies and heteroduplex formation during the last PCR cycles are potential problems when using DGGE band patterns to evaluate the biodiversity in a sample. Many bacteria have multiple heterogeneous rRNA operons. Bacterial genomes can contain between 1 and 14 operons that include the 16S gene, and the 16S rRNA sequences have been shown to differ up to several percent between operons (Farrelly *et al,* 1995; Acinas *et al,* 2004). This could lead to an overestimation of microbial diversity when using 16S rRNA approaches. The DGGE may be most useful in evaluating gradual changes in a community structure over time and space following a perturbation (Nicolaison and Ramsing, 2002).

2.6 Microbial diversity in the gastro-intestinal tract

In order to explore the potential impact of gut bacteria on the development of Type 1 diabetes, it is necessary to have an understanding of microbial diversity in the gastro-intestinal (GI) tract. At the level of the Operational Taxonomic Unit (OTU), where a species is defined as showing greater than or equal to 97% genetic similarity (Stackebrandt and Goebel 1994), progress with 16S rDNA–based enumeration has shown

that healthy individuals harbour unique microbial communities. Limits on the identification and understanding of the role of these communities exist due to an inability to culture the majority of species, biases introduced by preferential PCR amplification of 16S genes, and our limited ability to infer function from gene sequences (Gill *et al,* 2006).

Many studies cite the intestinal habitat of an individual adult human as containing 300-500 species of bacteria, with 30 to 40 species accounting for 99% of the total population (Gill *et al,* 2006). However, other researchers estimate that gastro-intestinal tract is home to approximately 500-1000 species, whose collective genomes contain 100 times more genes than the human genome (Bäckhed *et al,* al 2004). Colonisation of the mammalian GI tract starts immediately after birth, with numerous species of both facultatively anaerobic and anaerobic non-pathogenic bacteria, considered to be the "normal gut flora" becoming established in the first few days of life. The early pattern of colonisation in humans is initially affected by gestational age (pre-term or term) the type of delivery (e.g.vaginal vs. c-section) and diet (breast vs. formula) (Edwards and Parrett, 2003). Other environmental factors may have a major role as there are differences between infants in developing nations compared to industrialized nations, and even between hospital wards. These pioneer bacteria can modulate expression of genes in the host epithelial cells, thus creating a favourable habitat for themselves (Guarner and Malagelada, 2003). They can also prevent the growth of other bacteria introduced at a later date. For example, in breast-fed infants the reduced environment initially created by facultative anaerobes promotes the growth of anaerobes, particularly *bifidobacteria*,

while the facultative anaerobes decrease in number. Thus intestinal colonisation is very relevant to the final composition of permanent flora in adults.

Bacteria living in the human gut achieve very high cell densities, yet diversity at the division level (superkingdom or deep evolutionary lineage) is among the lowest found in an ecosystem: only 8 of 55 known bacterial divisions have been identified to date, and of these 5 are rare. The dominant divisions are the *Cytophaga-Flavobacterium-Bacteroides* (CFB) and the *Firmicutes* (e.g genus *Clostridium* and *Eubacterium*). The stomach and small intestine contain only a few species adhering to the epithelia and a few others in transit. The scarcity of bacterial species in the upper tract may be due to a combination of factors, including the luminal medium (acid, bile, and pancreatic secretion), and a more aerobic environment, which create a hostile environment. Add to this phasic propulsive motor activity towards the ileum which results in a faster transit time (approximately four hours in humans) and inhibits stable colonisation of the lumen (Dressman and Yomada, 1991).

Moving down the GI tract, one study found that the distal gut and fecal microbiota of three healthy adults surveyed was dominated by two bacterial divisions, the *Bacteroidetes* and the *Firmicutes*, which made up > 99% of the identified phylotypes, with greater diversity of the genera and subgeneric taxal in the *Firmicutes* than the *Bacteroidetes* (Dethlefsen *et al,* 2008).The numerically predominant genera inhabiting the large bowel of humans attain population levels of approximately 10^{10} cfus (shown by culture methods) to 10^{11} cells (shown by nucleic-acid based microscopy or flow cytometry) per g wet weight of contents (Kimura *et al,* 1997; Tannock, 2007).

In the large intestine, a complex, dynamic microbial ecosystem exists, with high densities of living bacteria ($10^{11} - 10^{12}$ cells/g of luminal contents) (Macpherson and Harris, 2004). Faecal solids consist of 60% bacteria, some of which are potential or opportunistic pathogens. Opportunistic pathogens are those that are harmless under normal, healthy conditions but can be a source of infection and sepsis under certain circumstances such as a breach in the integrity of the bowel barrier (Guarner and Malagelada, 2003). Anaerobes outnumber facultatively aerobic bacteria by 100 to 1000 times in faeces, compared to a factor of 10 to 100 in the colon, with *Firmicutes* and *Bacteroides* groups representing the dominant mucosal flora (Ferguson 1996; Tannock, 2007). The human dominant genera are listed in Table 2.2. Molecular hybridization procedures (dot-blot techniques using universal 16S rRNA probes) have shown that facultative aerobes including *E. coli* , *enterococci* and *lactobacilli* achieve high densities in the human ceacum, with 50% of ribosomal rRNA sequences in caecal contents corresponding to these species ; yet these species account for only 7% of bacterial ribosomal RNA found in faecal samples (Marteau *et al,* 2001). Over 62% of the cells in human feces that can be detected with a bacterium-specific probe belong to gram-positive groups (Breitbart *et al,* 2003). Suau *et al* (1999) found 95% of flora from a single human faecal sample could be assigned to *Bacteroides*, *C. coccoides* or *C. leptum* lineages. However, although the primer used in the Suau *et al* (1999) study is intended to be a universal 16S primer, when used as a bacterial probe in the RDP II release 9.60 database, it detected only approximately 10% of the genes in the target range. Faeces are subject to drying and loss of 85% of its fluid content in the descending colon and rectum, thus it is not surprising that there is a difference in the bacteria that are found when compared to

other gut tissues. Only 0.1 to 1% of the proportion of IgA and antibodies produced in the gut (as measured by whole-gut lavage) can be recovered from faeces, and this proportion is deeply affected by intestinal transit and faecal characteristics (Ferguson, 1996).

Table 2. 2: Dominant genera of bacteria found in the gastro-intestinal tract of adult humans*

Anaerobic bacteria	Facultative anaerobes
Bacteroides	*Escherichia*
Bifidobacterium	*Enterobacter*
Eubacterium	*Enterococcus*
Clostridium	*Klebsiella*
Peptococcus	*Lactobacillus*
Peptostreptococcus	*Proteus*
Ruminococcus	
Fusobacterium	

* Suau *et al,* 1999; Edwards and Parrett, 2003; Guarner 2003; Macpherson and Harris, 2004.

The diversity found within the human GI tract, which is represented by a few divisions of very tight clusters of related bacteria, may reflect strong host selection for specific bacteria whose collective impact is beneficial to the host. This implies that the bacterial community structure promotes functional stability to the gut ecosystem (Gill *et al,* 2006). An estimated 1200 viral genotypes in human feces suggest that bacteriophages (viruses that attack bacteria) may be a powerful force in shaping the diversity and population structure of the gut's microbial genetic landscape. Phages likely influence the composition of bacterial populations in the intestine through specific predation on microbial hosts, and may introduce new phenotypic traits, such as antibiotic resistance and the ability to produce exotoxins (Breitbart *et al,* 2003).

Only limited data are available on the natural variation of intestinal microbiota in animal models (Suau *et al,* 1999). Knowledge about variations in the intestinal microbiota over time and between individuals prior to intervention is an important prerequisite for design and interpretation of cause-and-effect studies. The composition of the fecal biota of mammals is known to be very stable at the level of genus, while changes do occur over time at species, subspecies or strain level; diet can also lead to changes in the dominant bacteria (Bernbom *et al,* 2006).

2.7 Host-bacterial mutualism in the gastro-intestinal tract

The distal human intestine houses an enormous population of bacteria dominated by relatively few divisions that are highly diverse at the strain/subspecies level. The inter-relationship among components of the microbiota may be commensal (one partner benefits while the other is unaffected) or mutualistic (both partners benefit). An example of mutualism exists in the guts of ruminants and termites, which contain anaerobic

bacteria that break down ingested polysaccharides and ferment the resulting monosaccharides to short-chain fatty acids. The host gains carbon and energy, while the microbiota is protected in an anoxic environment and a rich buffet of glycans. The structure and composition of the gut microbiota reflect natural selection at two levels: first, the ability of the microbiological organisms to adapt strategies such as growth rate, or substrate utilization patterns, affects the fitness of individual bacteria in a competitive ensemble; secondly at the host level, a microbiological community that is functioning at a suboptimal level can reduce host fitness (Gill *et al,* 2006). Bacteria in the mammalian gastro-intestinal tract have metabolic, trophic and protective functions.

The metabolic role of bacteria is one that provides hosts with genetic and metabolic attributes that the host has not evolved on its own, including the ability to harvest nutrients that would otherwise be inaccessible. This includes fermentation of non-digestible residues such as plant-derived pectin, hemicellulose and resistant starches, as well as endogenous mucus; salvage of energy as short–chain fatty acids; production of vitamin K; and absorption of ions.

The short-chain fatty acids (SCFAs - mainly acetate, propionate, and butyrate) produced by caecal bacteria in mammals are mainly absorbed by the hosts, and account for approximately 30% of calories extracted from a Western diet each day (Gill *et al,* 2006). The SCFAs may have a particularly important role in type 1 diabetes, as they are believed to act as modulators of glucose metabolism. Absorption of these SCFAs may result in a lower glycaemic response to oral glucose or standard meal , a response consistent with an improved response to insulin (Guarner and Malagelada, 2003; Bäckhed *et al,* 2005). The gut microbial community is essential for processing dietary

polysaccharides. Comparisons of germ-free mice to those that have a conventional microbiota acquired since birth shows numerous effects of indigenous microbiota on host biology. Young adult mice have 40% more total body fat than their germ-free counterparts fed the same polysaccharide-rich diet, even though conventional rats consume less chow per day. The effect can be explained by the presence of gut microbiota allowing energy to be salvaged from otherwise undigestible polysaccharides (Gill *et al,* 2006). Gut colonization produces substantial elevations in serum glucose and insulin (Bäckhed *et al,* 2004.)

Host signalling pathways regulated by the microbiota provide an opportunity to identify new therapeutic targets; for example, there is evidence that bacteria act through an integrated host signalling pathway to regulate energy storage in the host. Microbial suppression of intestinal *Fiaf* (angiopoietin-like protein 4, produced by brown and white fat, liver and intestine) promotes adiposity. It is speculated that changes in microbial ecology prompted by Western diets may act as a factor that affects predisposition toward energy storage and obesity. One study found that an increase in fat content in mice was accompanied by statistically significant elevation in fasting glucose and insulin levels, and an insulin-resistant state, as defined by insulin and glucose tolerance tests (Bäckhed *et al,* 2004).

Gut bacteria have a role in the development and homeostasis of the immune system, as well as control of epithelial cell proliferation and differential development. The intestinal mucosa is the main interface between the immune system and the external environment. Bacterial surface molecules include many carbohydrate-containing molecules, including lipopolysacccharides and teichoic acids, peptidoglycans and

glycoproteins, and offer a huge diversity of immunological epitopes that interact with the immune system (Mazmanian and Kasper, 2006). Gut associated lymphoid tissues contain the largest pool of immune competent cells in the human body; interactions between the host and bacteria at the mucosal interface play a part in the development of a competent immune system and recognition of commensal bacteria products through toll-like receptors (TLR) is critical to epithelial homeostasis (Guarner and Malagelada, 2003; Suzuki *et al,*, 2007). It has also been reported that commensal bacteria express an immunomodulatory molecule that is involved in the development of a normal mammalian immune system (Mazmanian *et al,* 2005). Conversely, germ-free animals have an undeveloped mucosal immune system: the Peyer's patches and isolated lymphoid follicles are small with few germinal centers; there are few IgA-secreting plasma cells in the lamina propria; and the numbers of lamina propria $CD4^+$ and intraepithelial $CD8\alpha\beta^+$ are reduced (MacPherson and Uhr, 2004a). The short-chain fatty acids (SCFAs) also have an important effect on the intestinal epithelium. All 3 major SCFAs stimulate epithelial cell proliferation and differentiation in the large and small bowel *in vivo* (Guarner and Malagelada, 2003). In particular, the rate of production of crypt cells is reduced in the colon of rats bred in a germ-free environment. Their crypts contain fewer cells than those of rats colonised by conventional flora, which suggests that the intraluminal bacteria affect cell proliferation in the colon. Differentiation of epithelial cells is affected by interaction with resident bacteria.

A study by McCartney et al (1996) found that the *bifidobacteria* and *lactobacillus* populations were characteristic of a particular human host. A follow-up to this study found that total *bifidobacterial* populations were similar for all 10 subjects (10^9 to 10^{10}

cfu g wet weight of faeces) while *lactobacillus* numbers varied greatly between subjects (approx 10^4 to $10^{8\ cfu}$/g ww of faeces). Each individual harboured unique collections of *lactobacillus* strains, but the composition of *lactobacillus* microflora was simpler than that of the *bifidobacteria* population. The predominant strains were used in lymphocyte transformation assays and serum antibody titrations to determine whether the human immune system is stimulated, and to what degree, by specific strains of intestinal bacteria. Serum antibody titers (IgG) against *bifidobacteria* were lower than those against *lactobacilli*, perhaps reflecting the view that truly indigenous members of the microflora, which have evolved while cohabitating with a specific animal species, do not elicit a marked immunological response because they have specific antigenic similarities with the tissues of their host (Kimura *et al,* 1997). Some species of animals (pigs, mice, rats, and fowl) have relatively large populations of *lactobacilli* as a component of the normal microflora. The relative paucity of IgA antibodies in the serum suggests (but is not conclusive evidence) that there would be an absence or low concentration of secretory IgA molecules reactive with these bacteria in the intestinal contents.

The existence of normal intestinal microflora, composed of enormous numbers of bacterial cells, does not induce a marked inflammatory response in the intestinal mucosa. However, it is likely that antigenic components of the normal microflora are accessible to the defense mechanisms of the body. The lack of inflammatory pathology in the majority of humans may be linked to the potential for the intestinal flora to modulate the expression of cytokine-encoding genes in the enterocytes (Kimura *et al,* 1997). A recent study found that *Bacteroides fragilis* produces an immonomodulatory molecule, polysaccharide A (PSA) that has an important role in the development of a mature

mammalian immne system. Mazmanian *et al* (2005) found evidence that PSA is sampled by dendritic cells of the GALT; these dendritic cells then migrate to the mesenteric lymph nodes and initate T-cell responses that contribute to a proper cytokine balance. This modulation could speculatively drive the immune system towards a humoral response rather than a cell-mediated response. The cell-mediated response might be more likely to result in an intestinal pathology such as inflammatory bowel disease (Kimura *et al*, 1997). Other studies comparing germ-free and conventional rodents support the theory that normal gut flora stimulate or play a regulatory role in the development of the immune system. This is supported by findings that animals bred in a germ free environment have low densities of lymphoid cells in the gut mucosa. These animals have small specialized follicle structures, and low circulating concentrations of immunoglobulins in the blood (Guarner and Malagelada, 2003; Bäckhed *et al,* 2005). Colonization of germ-free mice significantly increases serum concentrations of IgG1, IgG2a and IgA (Guarner *et al*, 2002). Overall, germ free animals are very susceptible to disease, while a complex microbial flora has been reported to raise concentrations of serum immunoglobulins in chickens, rats and mice (Klaasen *et al,* 1993; Butler *et al,* 2000). It has also been shown that colonization is necessary for the proper induction of tolerance to food antigens, which is believed to down-regulate T helper type 2 (Th2) cells in the GI tract. This may be an important factor in the development of diabetes in BB rats, which show a marked intolerance to a wheat-based diet. In mice and rats, a non-pathogenic and non-culturable segmented filamentous bacterium (SFB) that preferentially attaches to the ileal Peyer's patch epithelium apparently stimulates the development of mucosal immune architecture and function (Guarner and Malagelada, 2003). A study using specific pathogen-free and

germ-free CpB:SE (Swiss) mice found that SFBs strongly stimulate the mucosal immune response. SFBs produced a rise in the number s of IgA-SC in the lamina propria of the small intestine, increased the IgA titers in serum and intestinal secretions, and enhanced the ConA-induced proliferative responses of mesenteric lymph node cells (Klaasen *et al,* 1993).

Finally, it is speculated that the intestinal microflora offers a barrier effect against pathogen growth or attachment (Lievin-le Moal *et al,* 2006), although the evidence for this theory is speculative at this time. Indigenous gut flora such as *bifidobacteria* and *lactobacilli* are believed to play an important role in the health of the host by competitively inhibiting colonization by pathogenic forms, and certain keystone members of the microbiota operate to maintain the stability and functional adaptability of the intestine (Cummings and MacFarlane, 2002). Some bacteria contribute to a barrier effect by producing bacteriocins; e.g. a *Ruminococcus gnavus* strain produces a lantibiotic (ruminococcin A), which was shown in activity assays to inhibit growth of a range of pathogenic *clostridia* (Dabard *et al, 2002).* The segmented filamentous bacteria (SFBs) that occur in the distal small bowels of various animal species may also competitively exclude pathogens from the distal small intestine, in addition to enhancing host resistance by influencing gut-associated lymphoid tissue (Guarner and Malagelada, 2003). In the human gut, populations (i.e. types of speices and strains) are remarkably stable within individuals, implying that mechanisms exist to suppress blooms of subpopulations and/or promote the presence of desirable bacteria (Gill *et al,* 2006). Overall, although the concept of the barrier effect is biologically plausible, more evidence is needed to support the theory.

2.8 Bacterial tranlocation

Bacterial translocation is the passage of viable indigenous bacteria from the GI tract through intact epithelial mucosal surfaces to the mesenteric lymph nodes (MLN) and other organs (Ohsugi *et al,* 1991). As previously noted, the composition of GI microflora in an individual remains relatively constant, as long as the environmental and dietary conditions of the host are constant. Indigenous bacteria normally inhabit the GI tract almost exclusively and are not found in the MLN, kidney, liver, or spleen of specific pathogen-free mice (Imai and Kurihara, 1984). When healthy germ-free animals have been colonized in experiments using commensal bacteria, the live commensals penetrate the mucosa to reach the deeper tissues in the two to three weeks before B and T cell responses are mounted by the mucosal immune system. After this time period, the deeper tissues once again become sterile in healthy animals (MacPherson and Urh, 2004a).

Dysfunction of the gut mucosal barrier can result in translocation of many viable indigenous microorganisms, usually those belonging to gram-negative genera (e.g. *Escherichia, Klebsiella, Proteus*) (Schroff *et al,* 1995). This occurs notably in haemorrhagic shock, burn injury, trauma to the intestine, intestinal obstruction, severe pancreatitis, acute liver failure and cirrhosis. Translocation also occurs during various disease processes. While data suggest that the baseline rate of positive mesenteric lymph node culture could approach 5% in otherwise healthy people, in disorders such as those previously noted, rates of positive culture are 16-40% (Guarner and Malagelada, 2003). The outcome of bacterial translocations can be severe: septicemia in patients has been reported as an opportunistic infection. Overpopulation of certain microflora and translocation to the MLN and other organs are important steps in the pathogenesis of

infection and potentially in certain diseases. The increased caecal population level of particular bacterial strains has been found to parallel the numbers of viable bacteria of the strain translocating to the MLN. Aerobic gram-negative bacilli generally characterize the oral, pharyngeal or faecal flora of the immune deficient patient. The numbers of the organisms increase with the severity of the underlying illness and depression of the immune response (Imai and Kurihara, 1984).

In experimental animal models of diabetes, a very low level of bacterial translocation has been observed in healthy adult animals and the host immune defence system effectively eliminates bacteria that escape the GI tract. On the other hand, diabetic animals manifest greater susceptibility to infectious diseases. Mice injected with single large doses of streptozotocin (ST2) had indigenous GI tract bacteria translocated to the mesenteric lymph nodes and other organs. The rates of translocation in various mouse models were compared. ST2-injected (multiple low dose) mice, NOD (non-obese diabetic) mice, alloxan-injected mice and single large-dose ST2-injected mice were examined, and there was an attempt to determine if insulin would decrease translocation from the GI tract in these mouse models. *E.coli* was found to translocate to tested organs at a greater incidence in the ST2-induced diabetic mice than in NOD and alloxan-injected mice. ST2 itself depresses some cell-mediated immune responses independent of its diabetogenic actions. The authors concluded that in this study translocation was not induced by diabetes but by ST2 (Ohsugi *et al,* 1991). However, another study found changes in the indigenous flora of ST2-induced diabetic mice. After treatment, there was a significant increase in *Klebsiella pneumonia, Staphylococci micrococci, Streptococcus faecalis* and *Streptococcus faecium* in the caecum, while populations of *Bacteroides,*

Lactobacilli, and *fusiform* bacteria stayed the same (Imai and Kurihara, 1984). It should be noted that both the Ohsugi *et al* (1991) and Imai and Kurihara (1984) studies relied on culture methods to detect bacteria; molecular methods may have produced different results.

Chapter III
Materials and Methods

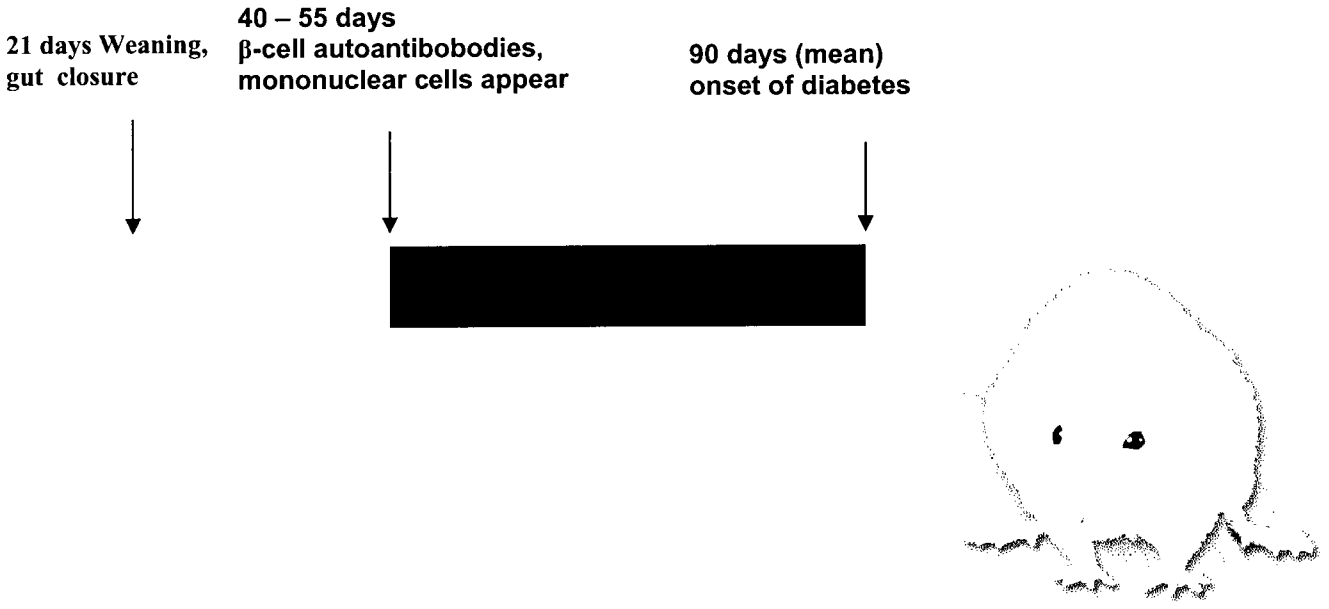
3.1 Animal model: the BioBreeding rat

This study of the BioBreeding rat gut microbiota was initiated before I joined the Brooks laboratory, thus the live animal protocol was completed and the necropsy performed by other students and staff from the laboratory. These same students and staff also completed the DNA extraction for the lymph node and small intestinal tissues at that time. Once I joined the project, I processed all of the BioBreeding rat faecal samples for DNA, and completed all other procedures and analysis as described below.

BioBreeding rats (BB rats) from the Animal Resources Division of Health Canada (Ottawa) were used in this study. The Health Canada Animal Care Committee approved this study, and housing and care for the rats followed the guidelines of the Canadian Council on Animal Care (CCAC, 1993).

The BB rat model is a closely bred diabetic-prone animal widely used for studying the relationship between diet and diabetes (Scott, 1996; Scott *et al,* 2002; Malaisse *et al*, 2004). The onset of diabetes in the BB rat is marked by spontaneous pancreatic insulitis as early as 40 days of age (see Figure 3.1). This is followed closely by destruction of β-cells and insulin deficiency, as well as hyperglycemia prior to puberty (Hardin *et al,* 2002). The mean incidence of diabetes in BB diabetic-prone rats (BBdp) fed a standard cereal-based diet has remained constant over the years 1998-2003 at 65.3% ± 14.9%(mean ± SD) (MacFarlane *et al,* 2003).

Figure 3. 1: Development of diabetes in the BioBreeding Rat



The original diabetic rats were discovered at BioBreeding laboratories near Ottawa in 1974 and transferred to Health Canada in 1977; the current colony is directly descended from the original group of animals. Although not completely inbred, the colony has remained closed for the past 30 years and about 90% of matings are between siblings, usually a non-diabetic female with a diabetic male (Scott, 1996). Genotyping for selected markers indicates that the animals are 80% identical at the DNA level. At least 3 genes are involved in diabetes in the BB rat: the lymphopenia (*Lyp*)gene on chromosome 4, *Iddm1;* a major histocompatibility complex-linked gene on choromosome 20, *Iddm2*; and *Iddm3*, an unmapped gene conferring resistance to diabetes in Fischer rats (Scott, 1996). The BBdp rats carry a mutation at the *Iddm1/Lyp* locus that is attributed to a frameshift deletion in a novel member of the immune-associated nucleotide related gene family *Ian5* (MacFarlane *et al*, 2003). The BBdp rat displays the insulitis lesion that is the key phenotype for destruction in type 1 diabetes, in a manner that closely mimics that found in the human disease (Roep and Atkinson 2004). The disease appears equally in males and females around puberty and adolescence, between 55 to 140 days of age, with a mean age at onset of about 90 days. The onset of diabetes is sudden, with associated weight loss or failure to gain weight, hyperglycaemia and loss of virtually all circulating insulin, glucosuria, and ketonuria (Scott, 1996). BBdp rats that are older than 70 days show significant inflammatory activity throughout the small intestine, inflammation that is not seen in younger (45-day old) animals (Hardin *et al,* 2002). They are severely lymphopenic, with a low percentage of T-helper cells and almost no cytotoxic or suppressor T-cells.

The non-diabetic BB rats (control, or BBc) have a normal range of T cells and are not lymphopenic (Hardin *et al,* 2002). The BBc rats are developed from an early subline of the colony that does not spontaneously develop diabetes, and represent a stable, closely bred isolated population fed a standardized diet (Scott, 1996).

For this study, BBc and BBdp rats were group-housed in polyethylene cages lined with woodchips. The animals were provided with glass balls, gnawing sticks and polyvinyl chloride pipe housing to improve their environment. The rats were on a 12h light: 12h dark cycle, at 21 °C and 40% humidity, and had free access to water (reverse osmosis treated to 95% purity). All rats were fed Rodent Laboratory Chow 5001 (Purina Mills, Inc., Pembroke, ON). One group of animals was sacrificed at 19 to 21 days. The control group had a total of six animals, as did the diabetic prone group. A second group was sacrificed at 55 to 60 days. This group included 6 control animals and 6 diabetic prone animals. Animals were killed individually by CO_2 asphyxiation. The cervical region was dislocated after 2 to 5 minutes to ensure that the animal was dead.

3.2 Collection of lymph node and small intestine tissue samples

After each animal was sacrificed, the abdominal fur was doused in 70% EtOH. An incision into the abdomen was made and the lymph node area of the abdomen (see Figure 3.1) was quickly removed and placed onto paper towel. Individual lymph nodes were picked out and placed on a glass plate previously cleaned with 70% EtOH, then divided into 2 portions which were approximately equal in size. Pieces of the small intestinal wall approximately 5 cm in size were excised, and cleaned using with peptone H_2O and a needle from the sacrificed animals. All samples were placed in 2 ml cryotubes, immediately frozen in liquid nitrogen and stored at -80 °C.

Figure 3. 2: Necroscopy of BB Rat.

The arrows indicate where the lymph node (a) and small intestine samples were collected (b).

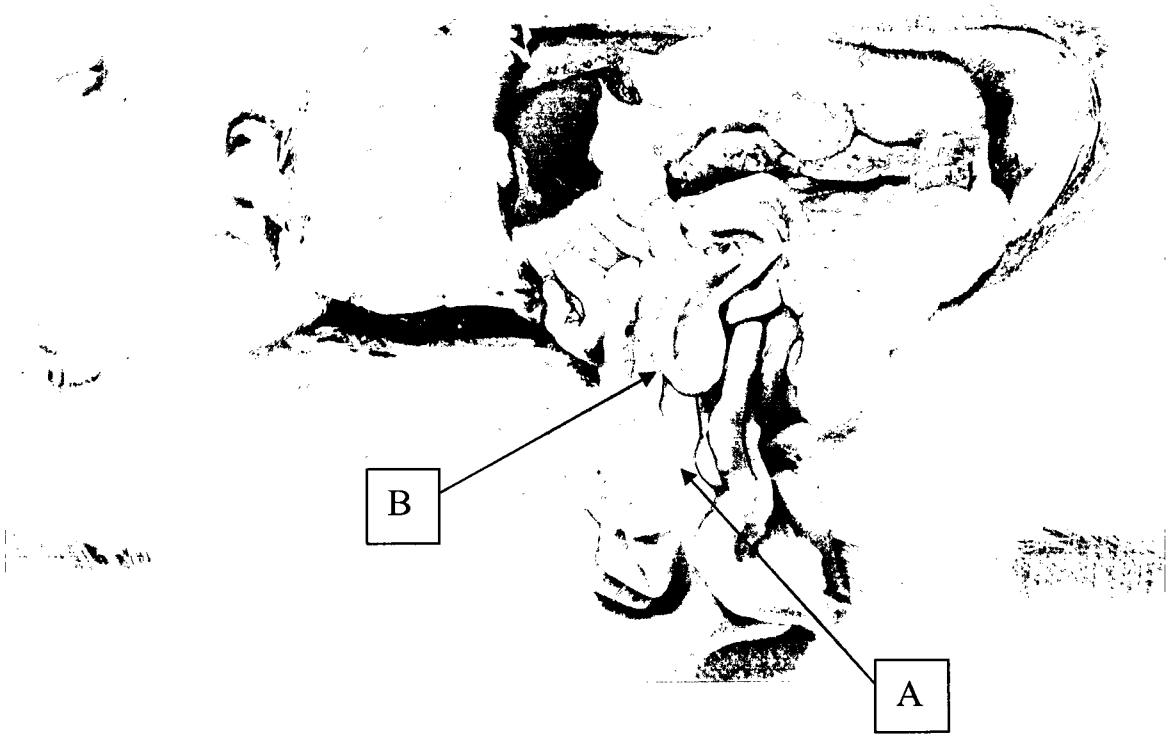


Figure 3. 3: Mesenteric lymph node mass excised from the gut of the BB Rat

3.3 Collection of rat faeces

Fecal pellets were removed from the last 2/3 of the colon and rectum and placed in 2 ml cryotubes with liquid nitrogen, then stored at -80 °C.

3.4 Extraction of DNA

Total genomic DNA was extracted from all samples (faecal, lymph node and small intestine) by the freeze-grind method (Hurt *et al,* 2001). Samples were weighed and kept on ice, then transferred to a cold mortar (pre-chilled at -80 °C) and frozen by adding liquid N_2. Sterile sand was weighed and the sample was ground to a fine powder. Each sample was refrozen and reground a total of three times, using a gentle pounding pressure with an ice-cold pestle. When grinding numerous samples, the mortar, pestle and spatulas were rinsed with 0.5M NaOH, followed by detergent, then pre-chilled with liquid nitrogen. DNA from samples weighing between 0.202 to 0.378 g was isolated with the QIAamp DNA Stool Mini kit (Quiagen Inc. Mississauga, ON). To prepare pooled DNA samples, DNA from each sample was quantified using the Nanodrop 1000 (Thermo Scientific, Wilmington DE U.S.A.), and all samples were adjusted to a concentration of 100 ng/uL. After adjusting the concentration, an equal volume of the DNA from each individual sample was pipetted into a sterile 2 ml tube. Pooled samples were used because in humans, bacterial composition varies among individuals within a given population (Hayashi *et al,* 2002). Although greater homogeneity was expected in the rat samples, the samples were pooled to produce maximum sampling diversity and minimize inter-animal differences. DNA quality was assessed by gel electrophoresis (0.8% agarose with ethidium bromide) and DNA concentration was determined spectrophotometrically

using the Nanodrop 1000 (Thermo Scientific, Wilmington DE U.S.A.). Aliquots were stored at 4°C or frozen at -20°C .

3.5 DNA amplification

Genomic DNA from faecal, lymph node and small intestine samples was used as a template for polymerase chain reaction (PCR) amplification of the full-length 16S rRNA eubacterial genes. The 16S, 23S and 5S rRNAa are often found linked together in one operon, which often contains an internal transcribed spacer and at least one tRNA (Acinas *et al,* 2004). In order to compare different universal primers, primer sets targeting specific regions of the 16S rRNA genes were also selected and amplified by PCR reaction as specified in Table 3.1. For each primer set, PCR reactions were optimized for annealing temperature, primer ratio and concentration, and target concentration. All PCR reactions were carried out in a total volume of 25 µl, using puReTaq Ready-To-Go beads (Amersham Biosciences, Baie d'Urfé, QC), using a BioRad DNA Engine Peltier Thermocycler (Hercules, California, U.S.A.).

Figure 3. 2: Mapping of the 16S rRNA positions and primer regions.

Helices (H) are numbered. The V3 and V6-V8 variable regiona are shown. V3 corresponds to positions 341 to 534 and is targeted by the HDA1/HDA2 primers. The V6-V8 region corresponds to positions 800 to 1400 and is targeted by the F940/R1422 primers. The primers are described in Table 3.1. (Adapted from Case *et al,* 2007)

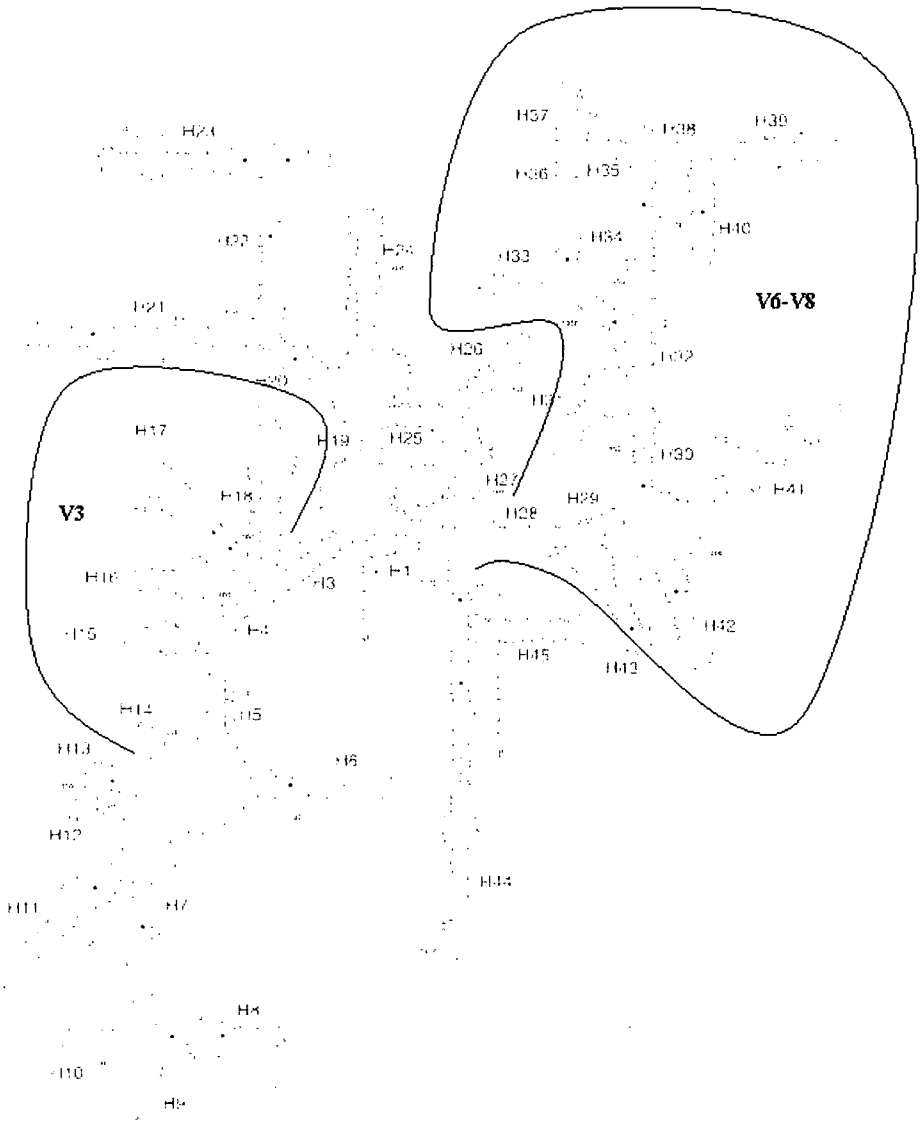


Table 3. 1: Primers Used to PCR-Amplify Genomic DNA

Target	**Sequence** (5′- 3′)	**Product length (bp = base pairs)**
Full-length 16S rRNA eubacterial genes	Sense (F27) [a] AGAGTTTGATCMTGGCTCAG Antisense (R1492) [a] TACGGYTACCTTGTTACGACTT	1500 bp
Full-length 16S rRNA eubacterial genes	Sense (F44)[b] RGTTYGATYMTGGCTCAG Antisense (R1543)[b] GGNTACCTTKTTACGACTT	1500 bp
V3 region of the *E. coli* 16S rRNA gene.	Sense (HDA1) [c] ACTCCTACGGGAGGCAGCAGT	193 bp
	with 5′ GC clamp: CGCCCGCCGCGCGCGGCGGGCGGGG CGGGGGCACGGGGGGCCTACGGGAG GCAGCAG	233 bp
	Antisense (HDA2) [c] GTATTACCGCGGCTGCTGGCA	
V6 / V8 (F940 / R1422) region of the *E. coli* 16S rRNA gene.	Sense (F940) [b] AAACTCAAAKRAATTGACGGGG Antisense (R1422) [d] GACGGGCGGTGTGTACAAG	470 bp

[a] Whitford *et al,*, 1998

[b] designed and validated in Brooks lab

[c] Muyzer *et al*, 1993

[d] Domann *et al*, 2003

In a previous study by researchers in the Brooks laboratory, the universal full-length 16S rRNA region primers F27 and R1492 (Whitford *et al* 1998; see Table 3.1) were used to obtain full-length faecal bacteria DNA sequences of the faecal flora of the BB rat (Brooks *et al,* 2003). These results are published in Brooks *et al* (2003). These full-length sequences are used for comparison in the present study, thus the PCR conditions are reported. The PCR conditions were one cycle of initial denaturation at 95 °C for 5 minutes; 10 cycles of denaturation (95 °C, 30s), annealing (50°C, 30 s), and elongation (72°C, 3 min), and a final cycle of extension for 5 min. This earlier research also produced full-length 16S rRNA sequence data from the BB rat small intestine and lymph node (unpublished data); I identified nearest relatives and aligned this sequence data using the ARB program (Ludwig *et al,* 1998; Ludwig *et al*, 2004). I attempted to use primers F44/R1543 (Table 3.1) to obtain full length sequence data. These primers were designed and validated in the Brooks laboratory, and were found to match 88.8 % and 90%, respectively, of bacterial sequences in the Ribosomal Database Project II , release 9.60 (http://rdp.cme.msu.edu/; Maidak *et al,* 2001). Note that primer F44 is a degenerate version of F27. PCR conditions for these primers were reported by the Brooks laboratory staff to be one cycle of initial denaturation at 94 C for 2 minutes; 10 cycles of denaturation (94 °C , 20s), annealing (54 °C, 30 s), and elongation (72 °C, 3 min), and a final cycle of extension 72 °C , 3 min.

Other researchers in the Brooks laboratory achieved good results using these primers with genomic DNA from pigs and from human (unpublished data). However, I was unable to amplify the BB rat DNA using this primer set, despite numerous attempts to modify and optimize the PCR conditions. I also attempted to repeat some of the

analysis using the F27/R1492 primers, and was not able to amplify the DNA with these primers. These primers match 68% and 50% respectively of the sequences in the RDP II database. The goal was to verify previous results obtained with the full-length 16S sequences, and compare the sequence information to the DGGE results. A possible explanation for my failure to amplify the samples is that the genomic DNA samples had degraded to a point where the full-length primers could not be used, as the DNA had been stored for two years before I attempted this analysis. Given this failure, the study was pursued using the shorter sequences as described in the following pages.

The V6-V8 region primer F940 primer is a degenerate primer designed by Dr. Brooks (Health Canada). It includes two degenerate bases. When compared to sequences corresponding to the *E.coli* region between 800-1400bp, this F940 primer was found to match 90.4% of all bacteria in the Ribosomal Database Project II (RDP II, release 9.60 (http://rdp.cme.msu.edu/; Maidak *et al,* 2001). The R1422 primer was designed by Domann *et al* (2003), and was found to match 91% of all bacteria in the RDP II. The PCR conditions were one cycle of initial denaturation at 95 °C for 2 minutes; 15 cycles of denaturation (95 °C, 20s), annealing (60 °C, 30 s), and elongation (72 °C, 45 s), and a final cycle of extension 72 °C for 44 s.

The V3 region is a variable 193 bp region corresponding to position 341 to 534 of the *E. coli* 16S rRNA gene. The V3 region was targeted with primers HDA1 and HDA2, which match 89.3% and 85% respectively, of the sequences in the RDP II database. When amplifying DNA for use in denaturing gradient gel electrophoresis (DGGE), the forward primer HDA1 included a G/C clamp, which is a G/C rich region at the 5′ end (Muyzer *et al,* 1993) . Primer concentrations were 0.1 micromols/L and genomic DNA

concentrations at approximately 100 ng per reaction. The PCR conditions were one cycle of initial denaturation at 96 C for 5 minutes; 15 cycles of denaturation (95 C , 30s), annealing (56 C, 30 s), and elongation (72 C, 1 min), and a final cycle of extension at 72 C for 5 min. For the primer comparison study, the PCR reaction was modified to include 35 cycles of denaturation, annealing and elongation. All other conditions were the same.

3.6 Random Cloning

PCR products were cloned into the vector pCR®2.1-TOPO (Invitrogen, Carlsbad, Calif.) and transformed into chemically competent TOPO® One Shot® *Escherichia coli*. Table 3.2 lists the volumes and reagents needed for the TOPO® Cloning reaction (TOPO TA Cloning Version R April 2004). The reaction was mixed gently and incubated at room temperature (22-23 °C) for 5 minutes. During this time, a vial of One Shot® Chemically Competent *Escherichia coli* was thawed on ice. After the 5 minute incubation, the cloning reaction was placed on ice; 2 µl of the cloning reaction was added into a vial of One Shot® Chemically Competent *Escherichia coli* and mixed gently. The vial was incubated on ice for 5 to 30 minutes (the length of incubation past 5 minutes is at the user's discretion). The cells were then heat shocked for 30 to 45 seconds at 42 °C without shaking. The tubes were immediately transferred to ice, and 250 µl of room temperature S.O.C. medium was added (2% Tryptone; 0.5% Yeast extract; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM glucose). The tube was tightly capped and shaken horizontally for one hour at 200 rpm and 37 °C. For each transformation, two pre-warmed plates of BHI agar containing ampicillin (100 µg/mL) and X-gal (20 µg/mL) were prepared. A volume of 40 µl of the transformation was spread one plate, and 20 µL spread on the second plate. Two different volumes were

spread from each transformation to ensure a sufficient number of well-spaced colonies. A hundred white colonies were picked from each transformation and inoculated into 2 mL of liquid Luria-Bertani (LB) medium (1.0% Tryptone; 0.5% Yeast Extract ; 1.0 % NaCL; pH 7.0) containing 100 μg/mL ampicillin. The colony picks were cultured overnight at 37 °C with horizontal shaking.

Table 3. 2: TOPO® Cloning Reaction

Reagant*	Chemically Competent *E.Coli*
Fresh PCR product	4 µl
Salt solution	1 µl
Sterile water	Add to a total volume of 5 µl
TOPO vector	1 µl
Final Volume	6 µl

* All reagents are stored at -20 °C when finished. Salt solutions and water can be stored at room temperature or 4°C.

3.7 Plasmid preparation

Plasmids were purified from the overnight liquid cultures using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO). The cells were harvested by centrifugation of the 2 mL cultures at approximately 12,000 X *g* for 1 minute. The supernatant was discarded, and the pellets completely resuspended by vortex in 200 µL of Resuspension Solution. The resuspended cells were lysed by adding 200 µL of the Lysis Solution. Contents were immediately mixed by inverting the tubes 6-8 times, until the mixture was clear and viscous. The mixture was not vortexed, as this harsh mixing will shear genomic DNA and result in chromosomal DNA contamination in the recovered plasmid DNA. The lysis reaction never exceeded 5 minutes, as a prolonged alkaline lysis may denature supercoiled plasmid DNA. The cell debris was precipitated by adding 350 µL of the Neutralization/Binding Solution to each tube. The tubes were inverted 4-6 times. Cell debris was pelleted using a centrifuge at approximately 12,000 X *g* for 10 minutes. During this time, GenElute Miniprep Binding Columns were placed into microcentrifuge tubes, and 500 µL of Column Preparation Solution was added to each miniprep column. The columns were centrifuged at approximately 12,000 X *g* for 30 seconds. The flow-through liquid was discarded. The cleared lysate from the Neutralization/Binding step was loaded into the prepared binding column and centrifuged at approximately 12,000 X *g* for 30 seconds. The flow-through liquid was discarded. The Wash Solution Concentrate was diluted with 95% ethanol prior to use; 705 µL of the diluted Wash Solution was added to the column and centrifuged at approximately 12,000 X *g* for 30 seconds. The column wash removes residual salt and other contaminants introduced when the column is loaded. The flow-through liquid was discarded, and the

column centrifuged again at maximum speed for 1 minute without additional wash solution to remove excess ethanol. The column was transferred to a fresh collection tube; 100 µL of molecular biology grade reagent water was added to the column. The column was centrifuged at approximately 12,000 X *g* for 1 minute. The eluted DNA was checked on a 0.8% agarose gel and stored at -20 C.

3.8 Sequencing

Initially, plasmid DNA concentration was determined using the Nanodrop ND-1000 Spectrophotometer (Thermos Fisher Scientific, Wilmington DE U.S.A), dehydrated and reconstituted with molecular biology grade water water (Hydropure Molecular Biology Water, Nuclease-Free dd 0.1 micron sterile filter, Hyclone, Utah, U.S.A) to a concentration of 100 ng/µL. However, this time-consuming step was eliminated when it was found that good results could be obtained from the sequencing reaction without concentrating the plasmid DNA. Sequencing reactions were prepared in a MicroAmp 96-Well Reaction Plate using the Big Dye Terminator v.3.1 Cycle Sequencing Kit (both from Applied Biosystems, Foster City, CA, U.S.A). Each reaction consisted of the following:

1 µL l BDT mix

3.5 µL 5X buffer

3 µL plasmid

1 µL of either F940 or R1422 primer (3.2 pmol/ul)

11.5 µL sterile water

The total reaction volume was 20 µL. The conditions for the sequencing reaction were an initial denaturation step at 96 °C for 1 min, followed by 25 cycles of denaturation

at 96 °C for 10 seconds, annealing at 50°C for 5 s, elongation at 60°C for 4 min. Cycle sequencing is used as it provides several opportunities to denature and extend the template, in order to produce a better signal in the sequencing reaction.

The sequencing reaction was cleaned using the Montage Seq96 Sequencing Clean-up Plate (Millipore Corp., Billerica, MA U.S.A) in order to remove excess dye. Injection solution (10 µL) was added to each well of the plate, and the sequencing reactions were added to their corresponding wells. The plate was vaccumed at approximately 23 HG until the wells were dry (2 minutes). The Seq96 plate was then blotted on paper towel , and 20 µL of injection solution added to each well. The plate was vacuumed at approximately 23 HG until all wells were dry (3 minutes). The plate was once again blotted, then a final volume of 20 µL of injection solution added. The plate was shaken horizontally for 10 minutes. Using a new MicroAMp Optical 96-well plate, 10 µL of Hi-Di formamide was added to each reaction well; the eluted samples from the clean-up plate were transferred to the reaction plate. The reaction plate was sequenced on a 16-capillary electrophoresis 3139xl Genetic Analyzer (Applied Biosystems/Hitachi, Life Technologies Corporation, Carlsbad, CA, U.S.A)

3.9 Primer comparison

Several primer sets were compared in this study in order to obtain the most accurate representation of the microbial community in the rat gut. The four primer pairs described in Table 3.1 were included in the comparison. The primers pairs were: two full length 16S rRNA primer sets F27/R1492 and F44/R1543, and two primers sets targeting variable regions of the 16S rRNA gene, HDA1/HDA2 and F940/R1422. For the primers HDA1/HDA2, the effect of additional cycles in the PCR reaction was also examined. For

this purpose, one reaction had 15 cycles and one had 35 cycles. For each pair of primers, the cloning reaction was performed as described above, using genomic DNA from the small intestine control samples from the 55 day old animals. Approximately 70 individual white colonies were picked and cultured from successful cloning reactions. Plasmid preparation and sequencing were as described above.

3.10 Denaturing Gradient Gel Electrophoresis

DGGE was performed with the Hoeffer DGGE System. PCR of genomic DNA was performed with HDA1-GC/HDA2 primers as described in the "DNA Amplification" section. PCR samples were mixed 1:1 by gentle pipeting with 2x loading dye (0.0% w/v bromopheol blue, 70% (v/v) glycerol in sterile water). A 15 µL volume of each PCR sample was applied directly into a running lane of 3.5% polyacrylamide stacking gel (wt/vol; 10X TBE buffer, 29:1 acrylamide/BIS, dH_2O) in 0.5X TAE (20 mM Tris acetate [pH 7.4], 10 mM sodium acetate, 0.5 mM Na_2-EDTA), overlying an 8% (w/v) denaturing polyacrylamide gel with a 40/60% denaturing gradient (see Table 3.3). The 1 Kb Plus™ DNA Ladder (Invitrogen, Carlsbad, California, U.S.A) was loaded every 5 lanes of the gels. Electrophoresis was performed at a constant voltage of 200V and a temperature of 60°C. All gel solutions were filtered and stored at 4°C. After electrophoresis, gels were fixed for 30 minutes in a solution of 10% ethanol, 0.5% acetic acid and dd H_20 (total volume of 600 ml), washed with a small volume of dd H_20. Gels were silver-stained for 5 minutes (0.2% silver nitrate, 0.02% formaldehyde, dd H_20 to 200 ml). The gels were developed for 10 to 20 minutes (3% NaOH, 0.1% formaldehyde, dd H_20 to 200 ml), followed by 2 washes for 20s each using ddH_20. Using the fixing solution described aboved, gels were fixed for an additional 5 minutes, washed for 10 minutes in ddH_20,

then soaked for 30 minutes in 20% ethanol, 10% glycerol, and ddH_20 to 140ml. Gels were stretched between cellulose sheets and dried in the dark (to prevent further development) for a minimum of 24 hours. Gels were photographed and aligned using Bionumerics (Bionumerics v. 35, Applied Maths 1998-2002).

Table 3. 3:Denaturing solutions used in DGGE

	60%	40%
40% acrylamide/BIS	20 ml	20 ml
50X TAE buffer	2 ml	2 ml
Formamide	24 ml	16 ml
Urea	25.2 g	16.8g
dH_20	To 100 ml	To 100 mL

A total of 20 bands of interest were identified and excised from the DGGE gels. To excise a band from a gel, both surfaces of the gel were wiped with 70% EtOH. The gel was then placed on a glass plate that had also been wiped with 70% EtOH. Using a sterilized scalpel, the band of interest was cut out and placed in a 2 ml Eppendorf tube and the piece of cellophane removed. Sterile water (10 μL) was added to the tube, and the tube was vortexed for 1-2 minutes. The water was removed with a pipet, the gel pieces were vortexed a second time for 1 minute, and any remaining water was removed. 100 μL of of sterile water was added, and the gel pieces were macerated using a pipet tip, then vortexed again for 1 min. The tube was then heated at 95°C for 20 minutes using a heating block and stored overnight at 4°C. After the overnight incubation, the tube was spun for 3 minutes at 15,000g to remove debris. The supernatant containing the DNA was transferred to a new Eppendorf, dried down in the Speed Vac for approximately 1 hour, then resuspended in 25 μL Sigma H_2O and stored at 4°C until needed.

3.11 Statistical analysis

DNA sequence data was analysed using Sequencher (Gene Codes Corporation, Ann Arbor, Michigan U.S.A.) to assemble the sequence fragments into contigs. The contigs were edited to remove vector sequence data, and then exported to text files. The text files were imported into the ARB program (Ludwig *et al,* 1998; Ludwig *et al*, 2004). Nearest relatives were identified and sequences were aligned in ARB, a process which involves viewing each sequence and manually correcting any mis-aligned or ambiguous bases. Aligned sequences were also compared with reference sequences in Ribosomal Database Project II; the RDPII Chimera Check program was used to identify chimeric sequences (RDPII, Release 8.1; Maidak *et al,* 2001). A chimera consists of 2 or more

phylogenetically distinct parental sequences, and is usually the result of a prematurely terminated amplicon reannealing to a different template DNA during PCR (Wang and Wang,1997; Gonzalez *et al,* 2005).

Using ARB (Ludwig *et al,* 1998; Ludgwig *et al*, 2004), distance matrices of the aligned species were calculated (using the Jukes-Cantor correction, Jukes and Cantor 1969) and exported to use as input files in the DOTUR program (Schloss and Handelsman, 2005). DOTUR was used to define operational taxonomic units (OTUs) in the small intestine, lymph node and faecal samples. The DOTUR program uses the genetic distance between sequences with a "furthest neighbour" approach to assign OTUs. This approach means that all sequences are within a known distance of the other sequences in the OTU; in the present study, a 97% similarity cut off is used to define OTUs (Stackebrandt and Goebel, 1994) Once sequences are assigned to OTUs, DOTUR uses the frequency data to contruct rarefaction curves for the number of species observed, as well as Shannon's and Simpson's diversity index, and Chao1, ACE, Jacknife and Bootstrap richness estimators. The Chao1 and ACE estimates are reported in this study, thus the equations for each of these are provided below (as seen in the DOTUR manual). DOTUR is available as free C++ source code and as a Windows executable (http://schloss.micro.umass.edu/software/dotur.html).

The Chao 1 richness estimator :

$$S_{Chao=} \frac{S_{obs} + n_1(n_1-1)}{2(n_2+1)} \quad \text{when } n_1 > 0 \text{ and } n_2 \geq 0 \text{ and when } n_1 = 0 \text{ and } n_2 = 0$$

$$S_{Chao=} S_{obs} + \frac{n_1^2}{2n_2} \quad \text{when } n_1 = 0 \text{ and } n_2 \geq 0$$

where :

S_{Chao} = Richness Estimate

S_{obs} = Observed number of species

n_1 = Number of OTUs with only one sequence

n_2 = Number of OTUs with only two sequences

ACE Richness Estimator:

$$N_{rare} = \sum_{i=1}^{10} i n_i$$

$$C_{ACE} = 1 - \frac{n_1}{N_{rare}}$$

$$\gamma^2_{ACE} = \max \left[\frac{S_{rare}}{C_{ACE}} \frac{\sum_{i=1}^{10} i(i-1)\, n_1}{N_{rare}\,(N_{rare}-1)} \right] - 1{,}0$$

$$S_{ACE} = S_{abund} \frac{S_{rare}}{C_{ACE}} + \frac{n_1}{C_{ACE}} \gamma^2_{ACE}$$

$$var(S_{ACE}) \approx \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial S_{ACE}}{\partial n_i} \frac{\partial S_{ACE}}{\partial n_j} \qquad \begin{array}{l} \text{cov}(f_i, f_j) = f_i (1 - f_i / S_{ACE}), \text{ if } i = j \\ \text{cov}(f_i, f_j) = -f_i f_j / S_{ACE}), \text{ if } i \neq j \end{array}$$

where :

n_i = the number of OTUs with i individuals

S_{rare} = the number of OTUs with 10 or fewer individuals

S_{abund} = the number of OTUs with more than 10 individuals

∫-LIBSHUFF (Schloss *et al,* 2004; Singleton *et al,* 2001), TreeClimber (Schloss and Handelsman, 2006a; Martin, 2002) and UniFrac (Lozupone and Knight, 2005) are all statistical tools used to compare community structure, based on the abundance of each member. As these tools take different statistical approaches to comparing the structures of communities, each one was used to analyse the data in this study .

Based on the ARB distance matrices that were created for the small intestine, lymph node and faecal samples, ∫-LIBSHUFF (Singleton *et al,* 2001) was used to test whether two or more 16S rRNA libraries are the same, different, or subsets of each other . If the libraries are significantly different from one another, it is assumed that they are derived from libraries of different composition. The ∫-LIBSHUFF analysis derives a homologous coverage curve that describes the extent to which the sequences in one clone library (library X) represent a given population based on sequence similarity (evolutionary distance) (Henriksen 2004). The curve is termed "homologous" because each sequence in library X is compared to other sequences within the same library using the equation:

$$\mathbf{C_X = 1-(N_X/n)}$$

Where:

C_X is the coverage for a single sample;

N_X is the is the number of unique sequences; and

n is the total number of sequences.

∫-LIBSHUFF also compares whether two or more clone libraries (libraries X and Y) share sequences by comparing each sequence in library X to all of the sequences in the second library, library Y. The comparison is made at several evolutionary distances

that define how distant two sequences can be before two sequences can be labelled as different from one another. The resulting values and curve are referred to as "heterologous" as defined by the equation :

$$\mathbf{C_{XY} = 1-(N_{XY} / n)}$$

Where:

C_{XY} is the heterologous coverage value

N_{xy} is the number of sequences in sample X that are not found in sample Y; and

n is the number of samples in X.

The difference between the homologous and heterologous curves is then calculated at each evolutionary distance and these are summed to calculate the Cramér-von Mises statistic:

$$\Delta C_{XY} = \sum_{D=0.00}^{0.5} C_X(D)-C_{XY}(D)^2$$

Where:

ΔC_{XY} is the difference between the heterologous and homologous curves;

$C_X(D)$ is the homologous curve;

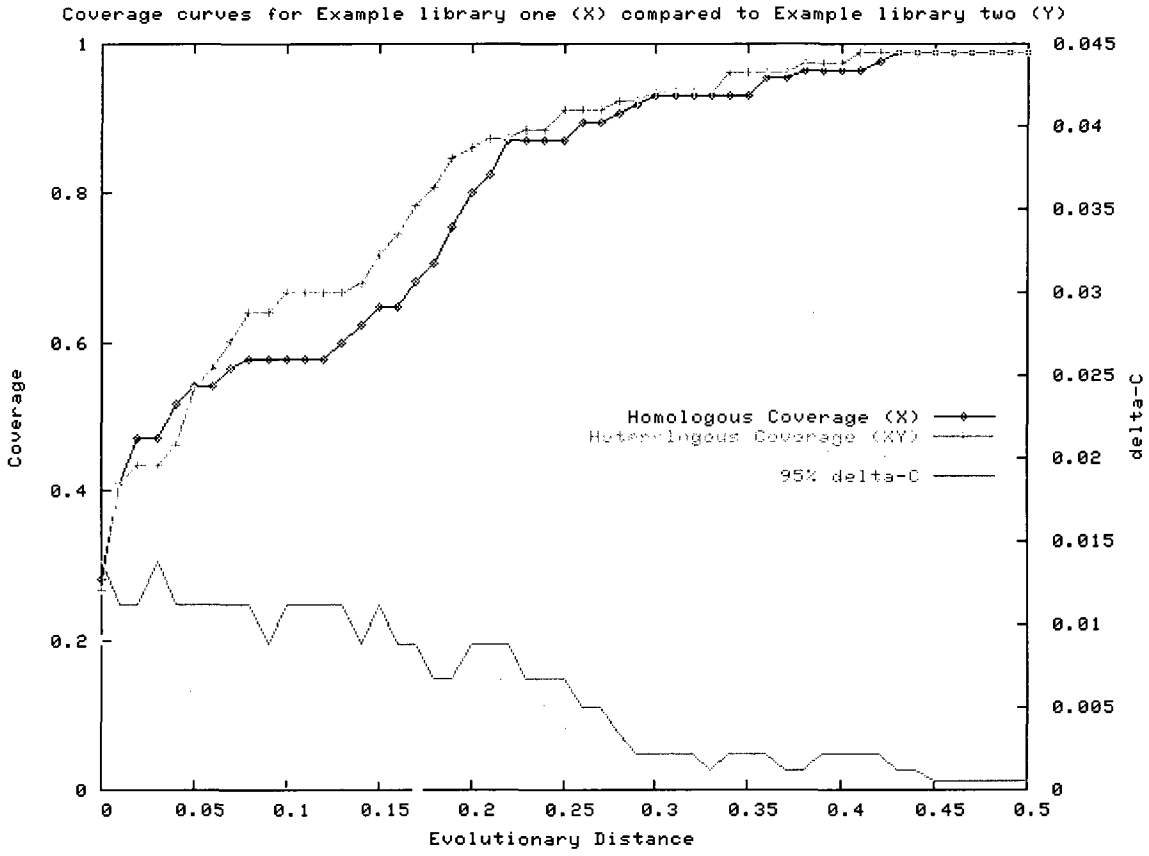
$C_{XY}(D)$ is the heterologous curve; and

D is evolutionary distance

Figure 3.4, reproduced from the documentation for webLIBSHUFF v. 0.96 (Henrikson, 2004), illustrates this concept.

Figure 3. 3: Coverage curves for example library 1 (x) compared to example library 2 (y).

In this test, evolutionary distance (D) increases by increments of 0.01. If the two libraries are identical, ΔC will be small; squaring the distance between $C_X(D)$ and $C_{XY}(D)$ makes ΔC sensitive to differences between the two curves (reproduced from Henrikson, 2004).



As a final step, ∫-LIBSHUFF applies a Monte Carlo simulation to the data to determine critical values for statistical significance; for this analysis, the iterations where set at 10 000. When comparing multiple libraries, as done in this experiment, a Bonferroni Correction is used to correct the P -value. It is important to correct for the experiment-wise error rate to account for the increased probability of detecting small P values due to chance when making multiple comparisons (Schloss *et al,* 2004). The ∫-LIBSHUFF P-value must be less than or equal to a critical value for any individual pariwise comparaison to insure that at least one library is difference with $P = 0.05$. Pre-calculated critical P-values are provide by Singleton *et al,* (2004) and are shown in Table 3. 4. For each of the pairwise comparisons, the result indicates a significant difference if the lower of the 2 P -values calculated by ∫-LIBSHUFF is equal to or lower than the critical value.

Table 3. 4: Corrected Critical *P* values.

A Bonferroni correction is used to correct for multiple comparisons. This correction is calculatd from the relationship $p = 1-(1-a)^{k(k-1)}$, where p is the experimentwise p-value of 0.05, a is the critical p-value, and k is the number of libraries. The table of pre-calculated p-values is reproduced from Singleton *et al,* 2004.

Critical P Values for an Experiment *P* Value of 0.05	
Number of Libraries being Compared	**Critical *P* Value**
2	0.025
3	0.0085
4	0.0043
5	0.0026
6	0.0017
8	0.00092
10	0.00057

The SONS (*S*hared *O*TUs and *S*imilarity) tool (Schloss and Handelsman, 2006b) was used to build on the DOTUR analysis in order to compare the membership and structures of the microbial communities at a specific OTU designation. Differences between the memberships of control and diabetes-prone gut microbial populations in the BB rat may contribute to the healthy or diabetic state. The SONS program compares the memberships and structures in communities by accounting for the abundance distribution of OTUs that are either endemic to one community or shared by two (or more) communities. SONS reads information from DOTUR-formatted file and designation for each sequence from a user-generated tab-delineated file (the same file is used in ∫-LIBSHUFF and TreeClimber). Then, SONS determines the number of individuals in each community that were sampled for each OTU, and calculates collector's curves for the fraction of shared OTUs between the two communities (with and without correcting for unsampled individuals), and the richness of OTUs shared between communities. In the present study, the iterations for calculating standard error values were set at 10,000. The output of the SONS program includes a file which reports the values of several community structure and richness esimates. A non-parametric estimator referred to as Shared Chao estimates the richness of shared OTUs between two communities (Schloss and Handelsman 2006b). It is analogous to the Chao 1 single-community estimator previously described in this report. Since the value for the Shared Chao estimate is reported in this study, the equation is provided below (as seen in (Schloss and Handelsman 2006b):

$$S_{1,2\,Chao} = S_{12(obs)} + f_{11}\frac{f_{1+}f_{+1}}{4f_{2+}f_{+2}} + \frac{(f_{1+})^2}{2f_{2+}} + \frac{(f_{+1})^2}{2f_{+2}}$$

where:

f_{11} = number of shared OTUs with one observed individual in A and B

f_{1+}, f_{2+} = number of shared OTUs with one or more individuals observed in A

f_{+1}, f_{+2} = number of shared OTUs with one or more individuals oberved in B

$S_{12(obs)}$ = number of shaed OTUs in A and B

SONS uses a non-parametric maximum likelihood estimator (Yue and Clayton, 2005) of similarity to measure the similarity of two community structures (Schloss and Handelsman, 2006b):

$$\theta_{YC} = \frac{\sum_{i=1}^{S_{12}} \frac{X_i}{n_{total}} \frac{Y_i}{m_{total}}}{\sum_{i=1}^{S_{12}} \left(\frac{X_i}{n_{total}}\right)^2 + \sum_{i=1}^{S_{12}} \left(\frac{Y_i}{m_{total}}\right)^2 - \sum_{i=1}^{S_{12}} \frac{X_i}{n_{total}} \frac{Y_i}{m_{total}}}$$

3.12 Phylogenetic Trees

Phylogenic trees for the microbial populations of the small intestine, lymph node and faecal samples were constructed using the neighbour-joining method in the PHYLO_WIN program (Galtier *et al,* 1996). A bootstrap analysis (500 replications) was used to evaluate the robustness of each node (branch) of the tree reconstructed from the whole data set, with a Jukes-Cantor correction (Jukes and Cantor, 1969).

TreeClimber (Schloss and Handelsman, 2006a) was used to analyse the phylogenic trees constructed in ARB for each of the samples. This program uses a parsimony-based test (Slatkin and Maddison, 1990) to detect any significant differences between the structures of two or more communities. The parsimony test considers a phylogenic tree where the taxa are labelled according to their original treatment (e.g. taxa

detected in 55-day old small intestine control samples are labelled S1, those from 55-day old small intestine diabetic prone samples are labelled S2, etc.). TreeClimber applies the parsimony algorithm to determine the minimum number of steps that would explain the covariation of phylogeny between the treatments, and scores are calculated for random trees generated by randomly linking taxa (Schloss and Handelsman, 2006a). A probability disribution is obtained by randomly joining trees in a Monte Carlo procedure (10 000 iterations are used in this study). This distribution describes the probability that each parsimony score would be by chance i.e a random process of combining sequences into groups. The user can then compare the parsimony score for their tree to the cumulative probability and assess the probability that a same or lower score might have been observed by chance. This tests the hypothesis that two (or more) bacterial communities share a common ancestral structure, and that the differences observed between the community structures can be explained by random variation. This hypothesis can be rejected if the populations in the community experience an effect such as perturbation or selection that results in either a) a gain or loss in the population (lower parsimony score) or b) reduced variation in the poplation (a higher parsimony score) (Schloss and Handelsman, 2006a).

Finally, UniFrac (Lozupone and Knight 2005) was also used to examine the phylogenic trees constructed in PhyloWin. This program measures the distance between taxa based on the branch length between lineages. UniFrac (*uni*que *frac*tion metric) measures the phylogenic distance between sets of taxa. This is measured as the fraction of the tree branch length that leads to descendant from either one condition or the other, but not both. Thus, the measure should reflect evolution (or adaptation) that is unique to one

environment (Lozupone and Knigth 2005). As with TreeClimber (Schloss and Handelsman, 2006a), Monte Carlo simulation is used to determine whether observed differences are greater than what would be expected by chance; the tree is kept constant and the condition assigned to each sequence in the tree is randomized.

Chapter IV
Results

4.1 Overview

The objective of this study was to explore the hypothesis that there is a relationship between the development of type 1 diabetes and the composition of the intestinal microflora. This hypothesis was tested by surveying the composition of the intestinal microflora in control and diabetic-prone BB Rats at two age points (shortly after weaning, and just prior to the onset of diabetes). Translocation of the intestinal microflora to the mesenteric lymph nodes was also assessed. The survey results were used to determine community structures and phylogenetic relationships on the microflora in an effort to understand changes along the digestive tract. An evaluation of the primers used in PCR analysis was added to the study in order to identify how the selection of primers may influence the results. The primers were assessed by comparative sequence analysis of random clone libraries, generated using different primer sets that target the 16s rDNA gene. The survey of the intestinal microflora was accomplished by a combination of DGGE and comparative sequence analysis of random clones.

Over the course of this study, 1531 clones were analyzed. One-hundred and sixty eight (168) of these clones (11%) either could not be sequenced despite repeated attempts, or were discarded as chimeras. The nearest relatives of the remaining 1363 clones were identified and used to align the clones; operational taxonomic units (OTUs) were defined by a 97% similarity cut-off (Stackebrandt and Goebel, 1994). At the beginning of this project, the phylogenic group and subgroup numbers corresponding to the listings found in Ribososmal Database Project (RDP) II, Release 8.1 listings (Cole *et*

al, 2003) were adopted. These numbers are used throughout the Results section of this paper. However, the most recent version of the RDP, Release 10, has adopted a revised, phylogenetically consistent higher-order taxonomy (Cole *et al,* 2009; Garitty *et al,* 2007; Wang *et al,* 2007). The RDP II 8.1 bacterial groups used in this report and their corresponding phylogenic assignement in RDP II 10 are shown in Table 4.1.

Table 4. 1: Names and numbers of bacterial groups and subgroups from RDP II 8.1 used in this study, and their corresponding phylogenic assignment in RDP II release 10

Phylogenetic group	RDP II 8.1 No.	RDP II, Release 10 Taxonomy (Phylum, Class, Order, Family, Genus)
Bacteroides and Cytophaga	2.15.1	*Bacteroidetes,Bacteroidales,Bacteroidaceae,Bacteroides*
Proteobacteria	2.28	Proteobacteria
Bacillus (a), *Lactobacillus* (b), *Streptoccus* (c) subdivision	2.30.7.17	ALL: *Firmicutes, Bacilli,* *(a) Bacillus: Bacillales, Bacillacea , Bacillus* (b) *Lactobacillales, Lactobacillacea, Lactobacillus.* (c) *Lactobacillales, Streptococcaceae, Streptococus*
Gram-positive bacteria	2.30	various
Clostridium lituseburense group	2.30.4.5	*Firmicutes,Clostridia,Clostridiales, Peptostreptococcaceae*
Clostridium leptum subgroup	2.30.9.13	*Firmicutes, Clostridia ,Clostridiales, Ruminococcaceae*
Clostridium coccoides group	2.30.4.1	*Firmicutes, Clostridia, Clostridiales, Lachnospiracea*
Clostridium aminobutyricum subgroup	2.30.4.4.4	*Firmicutes, Clostridia ,Clostridiales, Incertae sedis XIII*

4.2 Comparison and validation of primer sets

In the early stages of this project, we suspected that the shorter, variable region primer sets we had selected were preferentially amplifying certain DNA sequences. The sequence fragments did not seem to correlate with the sequences identified using longer, nearly complete 16S rRNA sequences from the small intestine and lymph node of the BB rat (refer to appendices A and B for the analyis of full-length sequences). For example, 28% of the full-length 16S small intestine clones aligned with *Clostridium lituseburense* M59107, yet none of the major bands from the DGGE aligned with this species. The full-length 16S clones from the lymph node were dominated by Acinetobacter junii (30% of clones), yet the corresponding band in the DGGE represented less than 5% of the total band density. In order to validate the primers, I compared four primer sets:

1. F27/R1492 primers for the full-length 16S sequences;
2. F44/R1453 primers for the full-length 16S sequences;
3. HDA1 / HDA2 primers targetting the V3 region of the 16S gene;
4. F940/R1422 primers targeting the V6-V8 region of 16S.

These primers were selected because they had either been successfully used in other studies (i.e. F27/R1492 used by Whitford *et al,* 1998; HDA1 / HDA2 used by Muyzer *et al,* 1993; R1422 used by Domann *et al,* 2003) or had been designed and validated in the Brooks laboratory (F44/R1453 and F940; upublished data). Most of these primers matched 85% to 91% of existing 16S gene sequences (RDP II, Release 9.6), indicating that they were good universal primers. The exception was the F27/R1492 pair, which matched only 50% and 68% respectively, of existing sequences. PCR conditions for each primer set had already been determined and optimized in the Brooks laboratory

(see Section 3.5, DNA amplification). It proved impossible to amplify the full-length 16S sequences with the F44/R1543 primers, possibly due to the DNA having degraded over time, thus this primer set was not included in the comparison. Only the 16S sequences from the F27/R1492 primers (unpublished data from the Brooks laboratory), the HDA1/HDA2 primers, and the F940/R1422 primers were assessed. In addition to potential differences between primer sets, the effect of PCR cycles may have an influence on the proportion of genes that are amplified (Bonnet *et al,* 2002). A higher number of PCR cycles can introduce bias by causing PCR products to reanneal to one another, rather than to the template (Suzuki and Giovanni, 1996). The effect of PCR cycle numbers was assessed for the HDA1/HDA2 primer set; because genomic primers (i.e. F940/R1422) are normally used in a 10-15 cycle PCR reaction, but DGGE reactions are normally run at 30 cycles, a 15 cycle library and a 35-cycle library were included in the analysis. This would allow us to find out if the differences in the major bands between the F27 survey and the DGGE bands were due to PCR bias. Genomic DNA from the BB rat 55 day-old small intestine (SIc55) was used in the primer comparison.

From the cloning reactions using the SIc55 samples and primers HDA1/HDA2 (15 cycles), HDA1/HDA2 (35 cycles) and F940/R1422 (15 cycles), 61, 68 and 100 clones respectively were picked. The full-length 16S clone sequences used in this comparison were from a database created by random cloning experiments before I entered the Brooks laboratory. This database contained 96 full-length sequences, amplified with the F27/R1492 primers (described in Table 3.1) from SIc55 samples. I imported these full-length 16S sequence data, together with the shorter V3 (HDA1/HDA2 - 15 cycles and HDA1/HDA2 - 35 cycles) and V6-V8 (F940/R1422 at 15 cycles)

sequence fragments, into ARB. All of the clones were aligned in ARB and their nearest neighbours were identified.

Figure 4.2.1 shows the distribution of clones among the groups of bacteria generated from each of the primers. The majority of clones aligned within the *Bacillus, Lactobacillus, Streptococcus* group. This group represents 83% of the V3 clones, 92% of the V6-V8 clones and 77% of full 16S library clones (Figure 4.2.1 A, B, and D). The rest of the clones aligned primarily within the *Clostridia* (17%) while some additional clones fell within the *Proteobacteria* group. A total of 17% of the full-length 16 clones, 14% of the V2-V3 clones and 7% of the V6-V8 clones fell within the *Clostridia* group.

Relative community similarity was assessed using abundance-based Jaccard (J_{abund}) index index and a non-parametric maximum likelihood estimator of similarity, θ_{YC} (Yue and Clayton, 2005; Figure 4.2.2). The J_{abund} estimator calculates the fraction of sequences that are found in shared OTUs at a given cutoff value. This estimator does not account for the similarity of the relative abundances of the OTUs that are shared between two communities (i.e., how similar is the distribution of the sequences among the OTUs). This is why the θ_{YC} estimator was also examined. The J_{abund} of the V6-V8 and full-length 16S libraries was 0.77 (se 0.17) (Figure 4.2.2) suggesting that these two libraries share a relatively high degree of community OTU similarity. The clone distribution among the OTUs within these two communities was relatively lower (θ_{YC} = 0.57, se 0.12). Figure 4.2.3 shows the proportion of OTUs that are shared between the V6-V8 and the full-length 16S clone libraries. At a 3% genetic cut-off, 84% of the OTUs found in the V6-V8 library are also found in the 16S library. The V6-V8 library appears to be a subset of the 16S library. The J_{abund} for the HDA1/2 – 15 cyle and 16S libraries was 0.55 (se 0.14); θ_{YC}

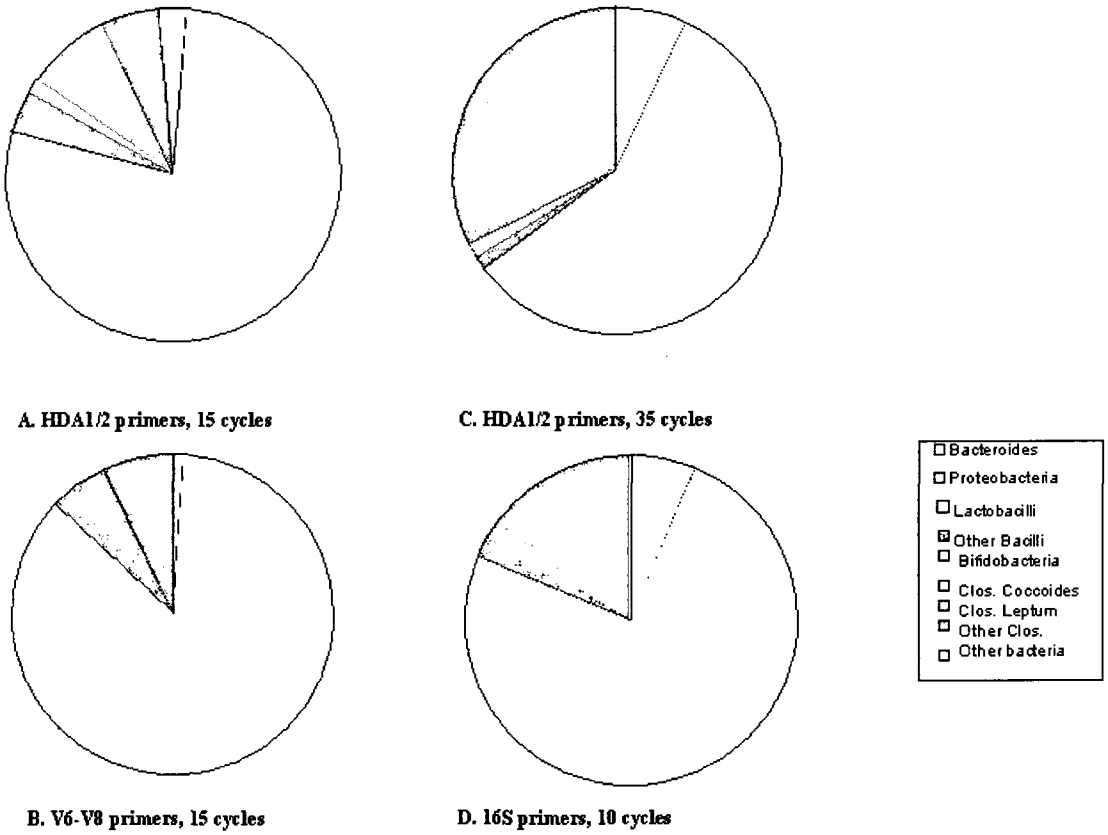
was 0.51, but with a large margin of error (se 0.21). The proportion of shared OTUs at a 97% cut-off was 41%-55%. A ∫-LIBSHUFF comparison detected no statistically significant differences ($P = 1.000$) between the V6-V8 and the 16S clone libraries.

Examination of the distribution of clones revealed differences. Six major groups of bacteria were represented in the HDA1/HDA2 - 15 cycle library as compared to 4 groups in the 35-cycle library (Figure 4.2.1 A and C). Proportionally less clones from the HDA1/HDA2 - 35 cycle primer library fell within the *Bacillus, Lactobacillus, Streptococcus* group as compared to the HDA1 / 2 – 15 cycle library (60% vs. 83%). The other major difference in the distribution of clones are the "*C. coccoides*" and "Other *Clostridia*" group, which represent a much larger proportion of the HDA1 / HDA 2 – 35 cycle library than of the 15 cycle library (33% vs. 14% respectively). The J_{abund} of the V6-V8 vs.16S libraries and that of the HDA1 / HDA 2 - 35 cycle vs 16S are both close to 0.80 (Figure 4.2.2) suggesting that these two libraries share a relatively high degree of community OTU similarity. However, the clone distribution among the OTUs was relatively lower in the HDA1 / HDA 2 – 35 comparison ($\theta_{YC} = 0.43$ se 0.12).

∫-LIBSHUFF (Singleton *et al,* 2001) was used to compare the HDA1/HDA2 (15 cycles), HDA1/HDA2 (35 cycles), and F27/R-1492 libraries and test whether they are the same, different, or subsets of each other. No differences were found between any of these libraries ($P = 1.000$).

Figure 4.2. 1: Clone distribution among bacterial groups from the 55-day old BB rat small intestine

Genomic DNA was cloned with primers A) HDA1/HDA2 at 15 cycles; B) HDA1/2 at 35 cycles; C) F940/R1422 at 15 cycles, and D) F27/R1492 at 10 cycles. All clones were aligned and identified in ARB.



As a final analysis, the estimated richness of the libraries (Chao1) created with the various primer sets was examined. The shared richness estimates are:

77 OTUs (95% CI: 58-128) for the HDA1/2 libraries (15 and 35 cycles);

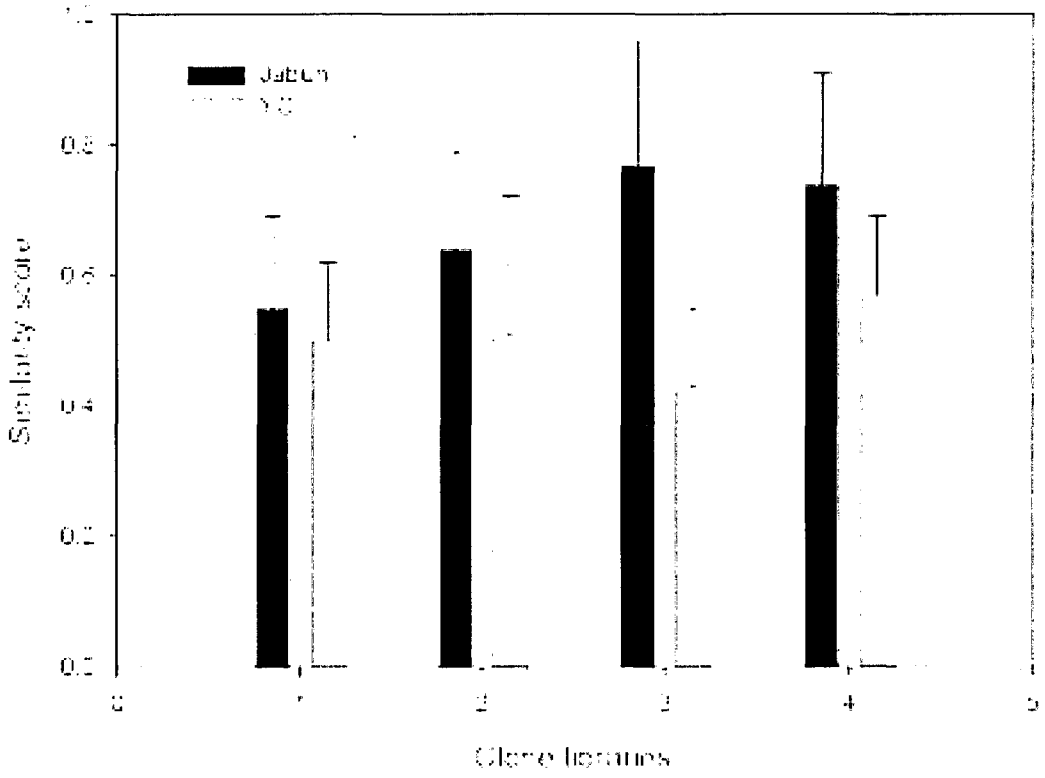
85 OTUs (95% CI: 50-198) for the F940/R1422 library; and

114 OTUs (95% CI : 66 – 252) for the F27/R1492 (full length) library

The lower richness estimate for the HDA1/2 library may be a result of including the 35-cycle library, as the additional PCR cycles would be expected to have a negative impact on species richness (Bonnet *et al,* 2002). The F940/R1422 (V6-V8 region) library has a lower average estimate of species richness when compared to the full length sequences. However, the average Chao1 estimate for the F940/R1422 library is within the 95% confidence interval of the full-length library. Youssef *et al* (2009) found that shorter sequence fragments can have an impact on species richness estimates (i.e. increase or decrease the estimate, depending on the primers) when compared to estimates for the full-length sequence. Suzuki and Giovannoni (1996) noted that shorter length primers (e.g. 15-mer versus an 18-mer or 20-mer) annealed at 55° C with less efficiency than longer primers. The authors also note that universal primers often include degenerate bases which may have an influence on the formation of primer-template hybrids (Suzuki and Giovanni 1996).

Figure 4.2. 2: Estimates of similarity in community membership (Jabund) and structure (θ_{YC}) between libraries generated with different primers.

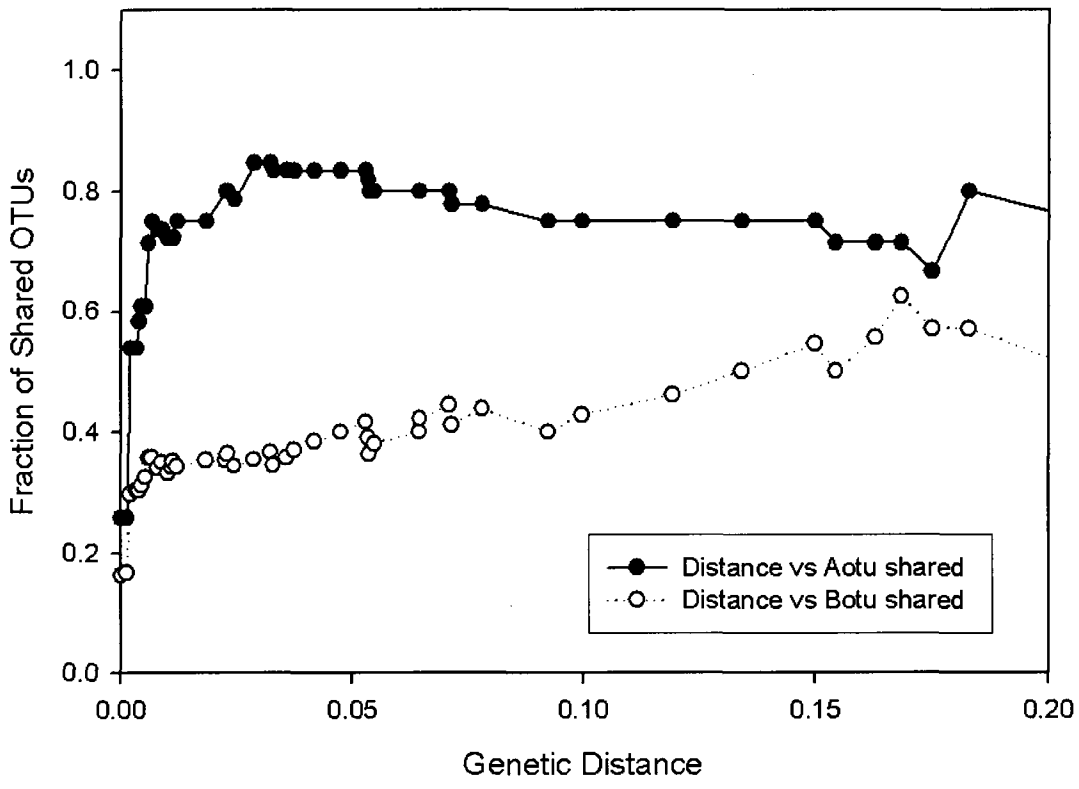
Standard error bars for 95% CI are shown.



Clone libraries
1 = HDA1/2 15 and 35 cycle
2 = HDA1/2 15 cycles and 16S
3 = HDA1/2 35 cycles and 16S
4 = V6-V8 and 16S

Figure 4.2. 3: Fraction of shared OTUs between clone libraries generated with V6-V8 region primers (Aotu shared) and full-length 16S primers (Botu shared).

All clones are from 55 day-old small intestine control samples. At a genetic distance of approximately 3%, Aotu shared is 0.8462, while Botu shared is 0.3548. This suggests that the V6-V8 library may be a subset of the 16S library



4.3 Analysis of Clones from DGGE

In order to investigate the bacterial communities of the lymph node and small intestine from 19 day-old control and diabetic and 55-day old control and diabetic-prone rates BB rats, a DGGE profile was created.

Twenty (20) major bands were selected from the DGGE gel for sequencing and subsequent identification. Bands represented either: (1) major, dense bands, evident in multiple lanes (for example, see R- 1619 and R-1610 in Figure 4.3.1); or 2) they appeared only rarely or in only one experimental group (eg R-1616 in the 19-day old control lymph node, Figure 4.3.1). The "major" bands were expected to concur with the major OTUs found in the appropriate full-length 16s clone libraries (Appendix A for small intestine clones; Appendix B for lymph node clones). The "rare" bands were expected to either correspond to less abundant bands in the 16S libraries, or represent unique OTUs. Appendix C, Table C1 contains a listing of all clones identified from the DGGE; the distribution of the clones in the major groups of bacteria is noted in Table 4.3.1.

Similarities among the banding patterns from different samples were assessed by quantifying band density (Bionumerics), transforming this data into fractional band densities (to sum to 1.0 per lane) and then analysing the fractional pattern using cluster analysis (Statistica 7.0; StatSoft, Tulsa OK, U.S.A). The cluster data, with city-block distances was plotted using the non-metric multi-dimensional scaling (NMDS) algorithm (Figure 4.3.2) to graphically express the relative relatedness of the different experimental conditions in two dimensions.

NMDS analysis (Fig. 4.3.2) showed a relatively tight clustering of the LNd55 and LNc55 samples as well as the SId55 samples. Visual inspection of the banding pattern from the DGGE gel (Fig. 4.3.1) illustrated this, showing alignment of the major bands

across these samples. The bands from the remaining conditions (SId55, SIc19, LNc19, LNd19) are interspersed throughout the scatterplot; the DGGE gel illustrates this with numerous rare bands visible in these groups. This supports the visual assessment of the DGGE profile, suggesting that there may be age-related differences in the gut microbiota. It can also be seen from the NMDS that the SId55 bands are widely scattered and the LNd55 bands are grouped together.

Most (4 of 5 bands) of the major bands that appeared consistently across the entire DGGE profiles aligned within the *Lactobacilli.* The band densities were used as a semi-quantitative estimate of the relative abundance of each OTU in the sample. In general, the *Lactobacilli* were found at approximately the same density in all of the experimental groups. The band corresponding to clone R-1619 was homologous (1.01% divergence) to *L. crispatus*, and was found across the DGGE profile at an average density of 11% (range of 7.5% -12.1%). *L. acidophilus* was also detected across all of the profiles (corresponding to clones R-1611 and R-1610) at a density of 8.2% (range of 5.6% - 11.3%) in the lymph node groups, and 9.1% (range of 8.3% – 10.6% in the small intestine groups. A major band (corresponding to clones R-1603, R-1604, R-1608 and R-1615) that appeared consistently across the DGGE profile was very prominent in the LNc55 (21.7%) and LNd55 (22.7%), and had a strong presence in the SId19 (12.5). However, this band represented only 3.2 – 6.8 % of the other groups. The band proved to be an isolate of *Lactobacillus acidophilus,* subsp. *johnsonii* (divergence <0.000). The final major band corresponding to a *Lactobacillus* (accession number AF157050) was clone R-1607. Overall, this band appeared at a slightly higher density (ave. 5.4%, range of 3.5% to 6.4%) in the small intestine groups than in the lymph node groups, (ave. 3.7.

range of 1.4% - 7.1%). However, the highest recorded density for R-1607 was 7.1% in LNc19.

The remaining major band, identified as clone R-1609, was homologous (divergence <0.000) to *Clostridium irregularis.* This band was prominent in the SIc55 at 20.5%, and was an important presence in the SId55 at 10.6%. This band was present in all other groups at between 4.4% – 7.6%.

Two bands were identified as *Enterococcus* species. One band was evident (average density of 5.3%, range of 2.7% – 8.3%) in all of the lymph node groups and in the 19 day-old small intestine (control and diabetic), but was absent from the 55-day small intestine. It was identified as *Enterococcus dispar* (clone R-1602, divergence of 1.01%). The other band, identified as *Enterococcus gallinarum* (0.5% divergence; clone R-1612), was also present across all groups, although it was barely detectable in the SIc19 (density of 0.4%). Its density in the SId55, LNc19 and LNd19 groups was also low (2.3%, 2.2% and 1.1% respectively). The band was found at between 3.6% - 5% in all other groups.

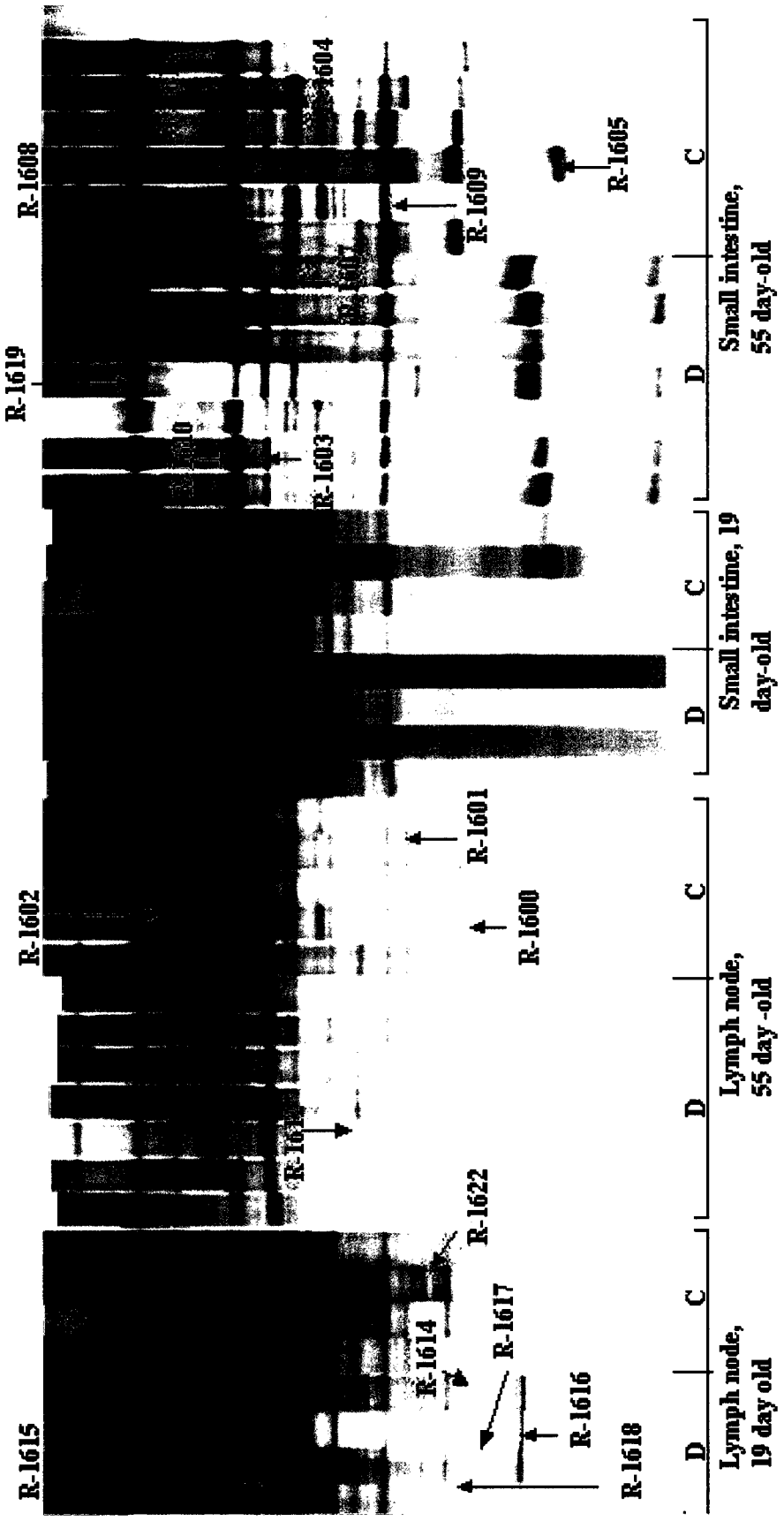


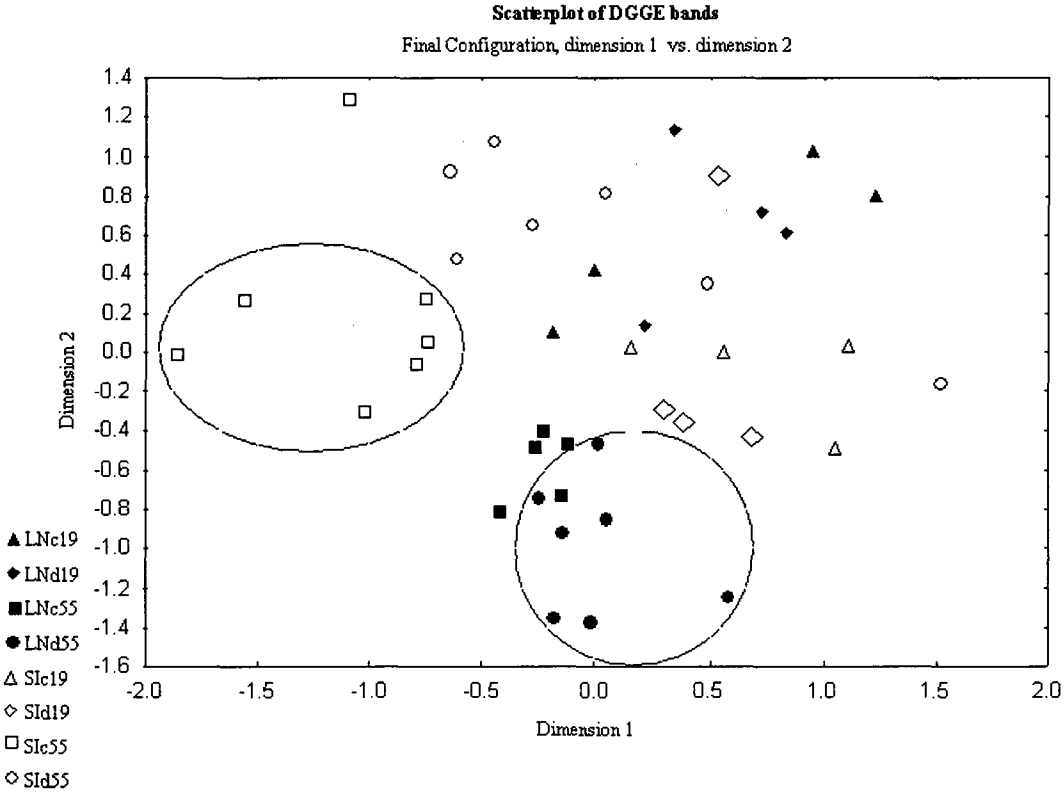
Figure 4.3. 1: DGGE profile of the gut bacterial community of the BB Rat (40/60% denaturing gradient). Gels were analyzed with Bionumerics v. 35, Applied Maths 1998-2002. Groups are identified by age and condition (Control.= C, diabetes-prone = D). The identified bands were excised, reamplified by PCR and cloned; clone numbers are shown by an "R-number". The identity of each clone can be found in Appendix C, Table C1.

Table 4.3. 1: Distribution of DGGE bands among the major groups and subgroups of bacteria

The proportion OTUs (%) in each group is noted. The corresponding proportions (%) of OTUs from the full-length 16S small intestine (SI) and lymph node (LN) are also shown.

Phylogenetic group	**RDPII No.**	**DGGE Clones (*n*)**	**% OTUs from DGGE**	**% OTUs from 16S (SI)**	**% OTUs from 16S (LN)**
Bacteroides and *Cytophaga*	2.15.1	3	20	2	0
Proteobacteria	2.28	2	13	10	50
Bacillus, Lactobacillus, Streptoccus subdivision	2.30.7.17	10	40	58	30
Clostridium lituseburense	2.30.4.5	5	20	13	10
Clostridium leptum	2.30.9.1	1	5	13	10

Figure 4.3. 2: Non-metric multi-dimensinal scaling analysis of DGGE bands.
A distance matrix of band densities was exported from Bionumerics. Analysis was performed in Statistica 7.0, using city-block distances (Beals *et al,* 1968) to calculate a non-metric, 2-dimensional scatterplot. Grey areas indicate one cluster of SIc55 bands, and a second cluster of LNd55 bands. All other bands are interspersed. Experimental condition is indicated by LN for lymph node, SI for small intestine, c for control or d for diabetic-prone and age (19 or 55 days old). Each point represents all of the bands in a single lane of the gel.



A grouping of faint bands was selected from the LNd19, as it appeared clearly in this group, but not in any other. Although they appeared to be three separate faint bands, bands R-1614, R-1616, and R-1617 were grouped into one OTU, related to *C. irregularis* (divergence of 6.6%). Band density analysis showed that this band accounted for 7.8% of the LNd19 community, and was also present at a low density in LNc19 at 1.9%. Clone R-1601, identified as *Acinetobacter junii* (divergence < 0.000), corresponded to a band in SIc55 (3.3%) and LNc55 (4.6%), but was very minor (less than 1.5%) or absent from other groups. R-1600, identified as a *Methylobacterium* species (divergence < 0.000) followed the same pattern: it was found in the SIc55 (density of 12.5%) and the LNc55 (density of 4.5%) but was negligible (< 1%) or absent in all other groups. Two "rare" bands were identified as *Bacteroides distasonig* (clone R-1622, divergence < 0.000%) and another *Bacteroides* species (AF139525; divergence 2.09%). The former appeared faintly in LNc55 (density of 1.1%) but nowhere else, while the latter appeared in the LNc55 (density of 4.4%), the LNc19 (density of 5.0%) and LNd19 (3.25%).

4.4 Descriptive analysis of clone libraries

A total of 1000 clones were obtained using the F940/R1422 primers. These clones were further divided into groups of diabetic-prone and control BB rats; two age groups (19 days and 55 days); and, three types of samples (lymph node, small intestine and faecal), as seen in Table 4.4.1.

When the 1000 clones from all twelve conditions were grouped together, they resolved into 177 OTUs. Very few OTUs were unique to the small intestine communities, which had 85% - 100 % of OTUs shared with either one or both the other two communities (faeces and/or lymph node). In contrast, the faeces community had a high proportion of unique OTUs, with the shared OTUs representing 40% – 60% of the total community. The lymph node community, which was expected to reflect a sampling of the small intestine community, shared 48% - 75% of OTUs in all groups save the 19-day old control, which shared 100% (no unique OTUs were found in LNc19).

Table 4.4. 1: Distribution of clones among experimental groups

BB Rat	Sample Type	Number of Clones (by age group)	
		19 days	55 days
Control	Small intestine	65	100
	Lymph node	42	102
	Faecal	96	73
Diabetic prone	Small intestine	81	98
	Lymph node	68	100
	Faecal	96	83

The small intestine libraries in both age groups and both control and diabetic-prone conditions were dominated by *Lactobacilli* species homologous to *L.acidphilus*, *L.reuteri*, *L. acidophilus* subspecies *johnsonii* and *L. murinus*. These 4 OTUs accounted for 73% of the SIc55 community, 50% of the SId55 community, 82% of the SIc19 community and 95% of the SId19 community.

In contrast, the Faeces libraries included many "rare" OTUs (i.e OTUs that represented only 1-3% of the total OTUs found in an experimental group), and only a few abundant OTUs. The most abundant OTU in the Fd19 library was identified as *Bacteroides distasonis* (divergence of 0.6%) and accounted for 25% of the OTUs in the community. This OTU was also found in the Fc19 community but represented only 12.7% of this group. The other dominant OTU in the Fd19 group, representing 15% of the total OTUs, was identifed as *Escherichia coli*. The two highest abundance OTUs in the Fd55 faeces library represented only 6% (clone R-3324) and 7% (clone R-3343) of the total OTUs in this group. Neither of these clones was closely related to any previously cultured species, and the nearest neighbour to both was a *Prevetolla* species from the *Bacteroides* group (9.2% and 9.11% divergence, respectively).

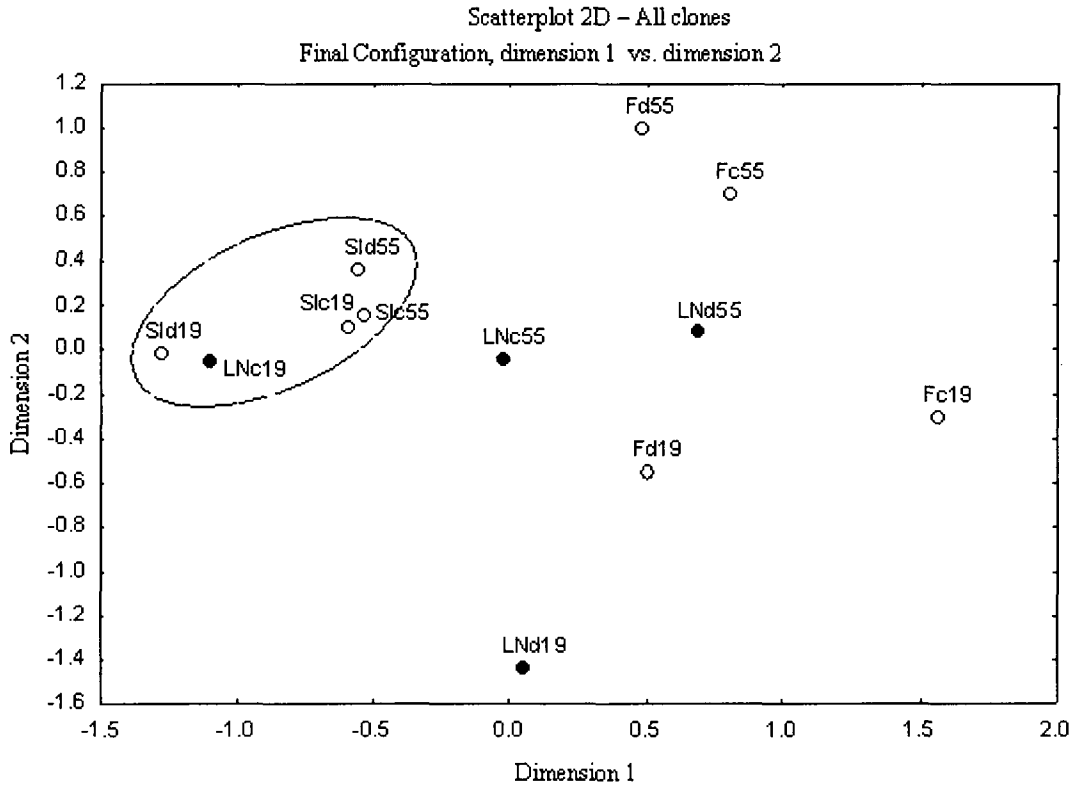
The clones and OTUs in the lymph node groups showed interesting differences. The LNc19 library consisted of only 4 OTUs, all of which where *Lactobacilli* species – in fact, 1 OTU, identified as *Lactobacillus reuteri* (divergence 0.2%), accounted for 77% of the community. The *Lactobacilli* species found in LNc19 were also detected as a core group of OTUs in all of the lymph node, small intestine and Faeces libraries. The LNd19 library included a very large component (75% of the total OTUs) of an uncultured bacterium from the *Clostridiales* family, distantly related to Eubacterium plautii

(divergence 9.72%). This OTU was unique to the LNd19 library. In the 55 day-old lymph node, a Gram-negative bacterium homologous to *Proteus vulgaris* (0.8% divergence) represented 31% of the OTUs in the LNc55 library and 18% of those in the LNd55 library. This OTU was detected in the SId55 group (3% of the OTUs) but was not found in any other library. Overall, the 55 day-old lymph node libraries had a higher diverstiy of OTUs than the 19 day-old lymph node libraries.

Non-metric multidimensional scaling (NMDS; Figure 4.4.1) was used to examine the overall relationship between the clone libraries from each of the experimental groups noted in Table 4.4.1. The SI libraries formed a relatively tight cluster with the LNc19 group. The fecal libraries were relatively distant from this cluster. The LNc55, LNd55 and LNd19 libraries were interspersed between the SI and LN libraries.

Figure 4.4. 1: Non-metric multi-dimensinal scaling analysis of clone libraries from the small intestine, faeces and lymph nodes of the BB Rat.

A distance matrix of band densities was exported from Bionumerics. Analysis was performed in Statistica 7.0, using city-block distances (Beals *et al,* 1968) to calculate a non-metric, 2-dimensional scatterplot. Experimental condition is indicated by LN for lymph nodes, SI for small intestine, F for faeces, c for control, d for diabetic-prone, and age (19 or 55 days old). Each point represents all of the OTUs found in each library. The small intestine OTUs cluster relatively together closely (grey area), while the faeces OTUs are relatively more widely dispersed. Three of the four lymph node OTUs fall in the distance between the two other libraries.



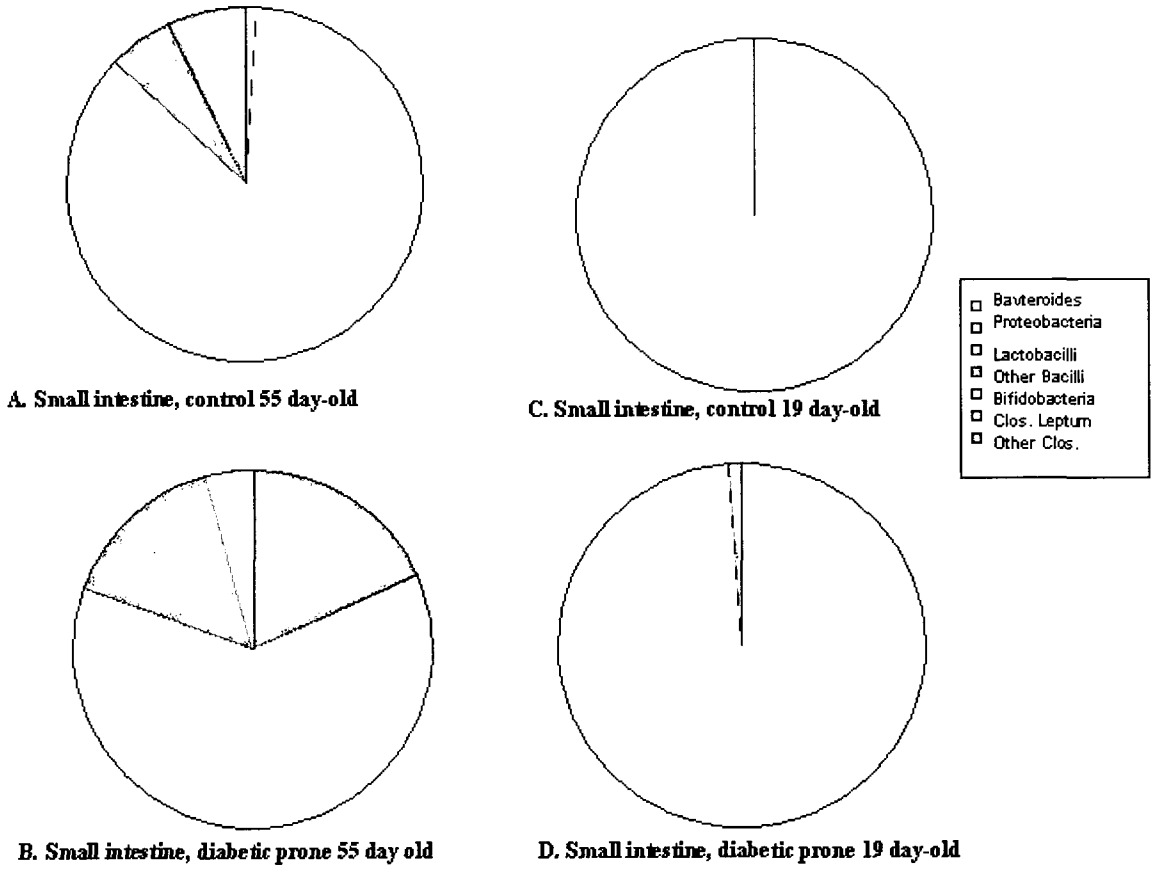
4.5 Library of clones from the small intestine

A total of 384 clones were obtained from the small intestines of control and diabetic prone BB rats. Forty (11%) of these could not be sequenced or were identified as chimeras. The remaining 344 clones were aligned within the ARB program. For each of the experimental conditions in the small intestine, the distribution of the clones among the major groups of bacteria is shown in Figure 4.5.1. The *Lactobacillus* group was the largest group represented.

Using DOTUR (Schloss and Handelsman, 2005) and a similarity cut-off of 97%, the 344 clones were found to fit into 25 distinct OTUs. Table 4.5.1 lists the distribution of clones and OTUs among the major groups and subgroups of bacteria (Appendix D, Table D1 lists all of the OTUs identified, and their nearest neighbours). The method of Chao and Lee (1992) was used to estimate species richness in each of the small intestine libraries. This method (the ACE estimate) predicted 6 OTUs in the SIc19, 3 OTUs in the SId19, 9 OTUs in the SIc55small intestine, and 10 OTUs in the SId55. Coverage for each of the libraries was estimated to be 100%. There was overlap between the OTUs found in each experimental group (Figure 4.5.2). With the exception of the SId55, each SI community overlapped by more than 60% with at least one other community. The SId55 shared fewer OTUs, with only 46% overlap. A core group of 4 OTUs is shared among the four conditions; three of these OTUs (corresponding to clones R-1813, R-1815 and R-1817) were homologous (divergence < 1%) to *Lactobacillus* species. The fourth OTU shared by all of the small intestine libraries was homologous to *Clostridium lituseburense* (divergence = 1.0%; clone R-1852).

Figure 4.5. 1: Distribution among the major groups of bacteria of clones from the small intestine of the BB rat.

Genomic DNA was cloned with F940/R1422 at 15 cycles. Experimental groups are shown under each pie chart.



A. Small intestine, control 55 day-old

C. Small intestine, control 19 day-old

B. Small intestine, diabetic prone 55 day old

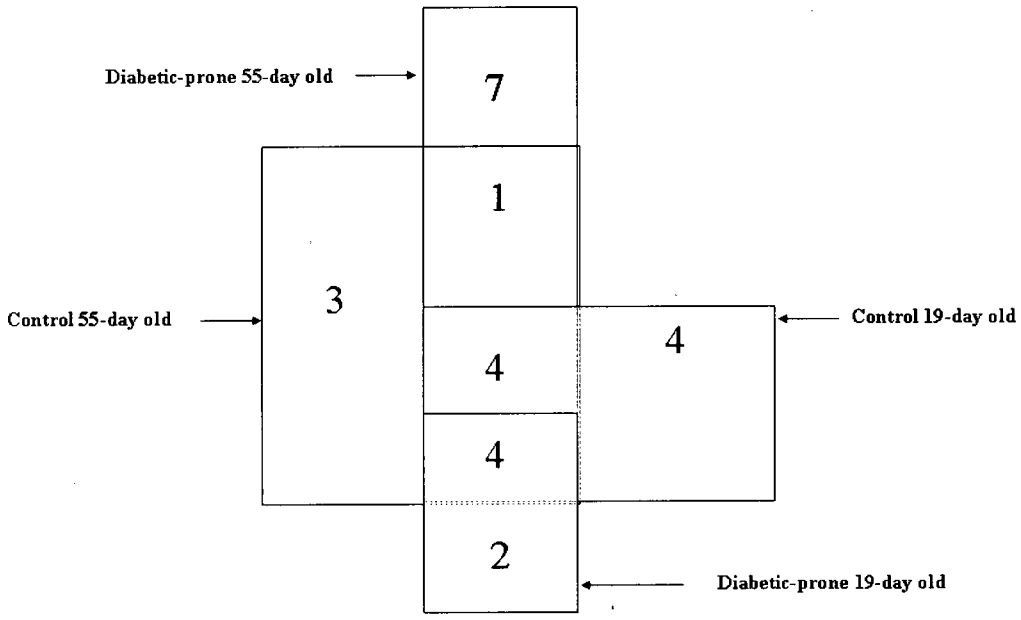
D. Small intestine, diabetic prone 19 day-old

Table 4.5. 1: Distribution of small intestine clones and OTUs among the major groups and subgroups of bacteria

Phylogenetic group	RDPII No.	Clones (*n*)	OTUs (*n*)
Bacteroides and *Cytophaga*	2.15.1	5	5
Proteobacteria	2.28	4	2
Bacillus, Lactobacillus, Streptoccus subdivision	2.30.7.17	284	8
Gram-positive bacteria	2.30	20	3
Clostridium lituseburense group	2.30.4.5	19	1
Clostridium coccoides group	2.30.4.1	2	2
Clostridium leptum subgroup	2.30.9.13	7	2
Clostridium botulinum subgroup	2.30.9.2	3	1

Figure 4.5. 2: Venn diagram comparing the OTU memberships at 0.03% divergence found in small intestine clone libraries

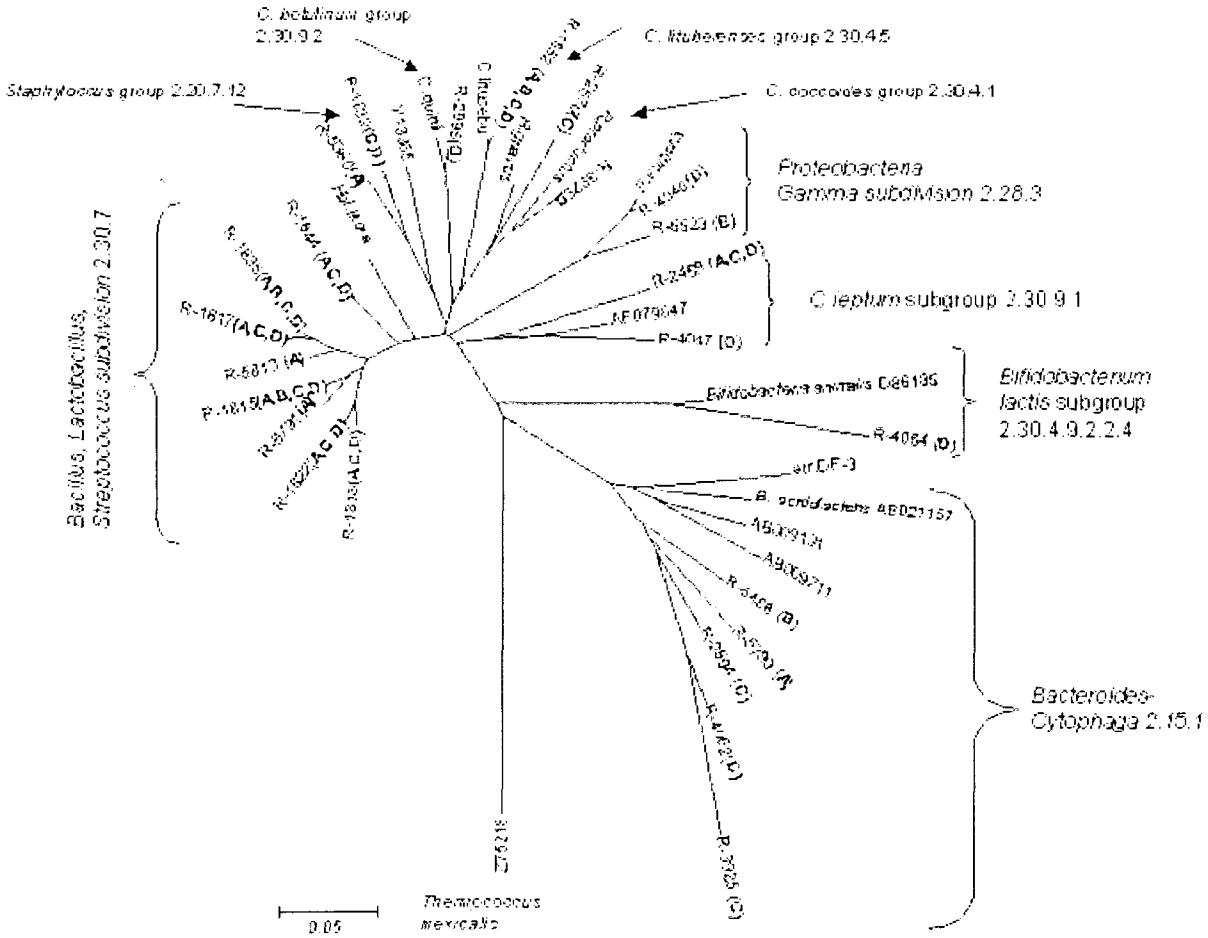
The libraries were 55 day-old control rats (SIc55, n = 100), 55 day-old diabetic prone rats (SId55 n = 98), 19 day old control rats (SIc19, n = 64) and 19 day-old diabetic prone rats (SIc19, n = 81).



A phylogenetic tree showing the relative positions of the OTUs from random clones and the associated experimental condition(s) is shown in Figure 4.5.3. Consistent with what is illustrated in the Venn diagram (Figure 4.5.2), the phylogenetic tree are no clusters of clones associated with any one condition; thus for the most part, the clones are present in every experimental condition. Clones that were homologous to *Proteobacteria* came from the diabetic prone condition (2 from the 55 day-olds, 2 from the 19 day olds). Most clones fell into the *Bacillus, Lactobacillus and Streptoccus* subdivision of bacteria. This large group of 284 clones resolved into only 8 OTUs (Table 4.5.1). All of these clones were homologous with, or very closely related to *Lactobacillus* species (divergence ranging from <0.000 to 3.5%; see Appendix D, Table D1). The clones within the order *Clostridiales* were grouped into several distinct lineages, including the *C.leptum*, *C.lituberense* and *C. coccoides* groups. Only two OTUs within this group were homologous to previously cultured species: R-1852 (picked 19 times) matched *C. lituseburense* (divergence of 1.0), while R-2995 (picked 3 times) matched *C. quinii* (divergence of 2.88%). None of the other clones were closely related to a previously cultured species, and were aligned against their nearest neighbours. The *Bacteroides-Cytophaga* group (2.15.1) was represented by a small number of clones (5 in all) , each of which was a distinct OUT. None of these clones were closely related to a previously culture species, and they form a distinct lineage within the *Bacteroides-Cytophaga* group see Figure 4.5.3).

Figure 4.5. 3: Rooted phylogenetic tree derived from the F940/R1422 DNA sequences

The tree shows the phylogenetic relationships among the small intestine random clones. Bar represents a 5% sequence divergence. The tree was constructed using a neighbour-joining method with a Jukes-Cantor correction (1969). The relative position of each clone was supported by bootstrap analysis for 500 trees. Some nearest neighbours are included and indicated by full name or accession numbers. Random clones are shown with and R-(number). The associated experimental condition is indicated in brackets following the clone number, with **A** = 19 day-old control, **B** = 19 day-old diabetic prone, **C** = 55 day-old control and **D** = 55 day-old diabetic prone.

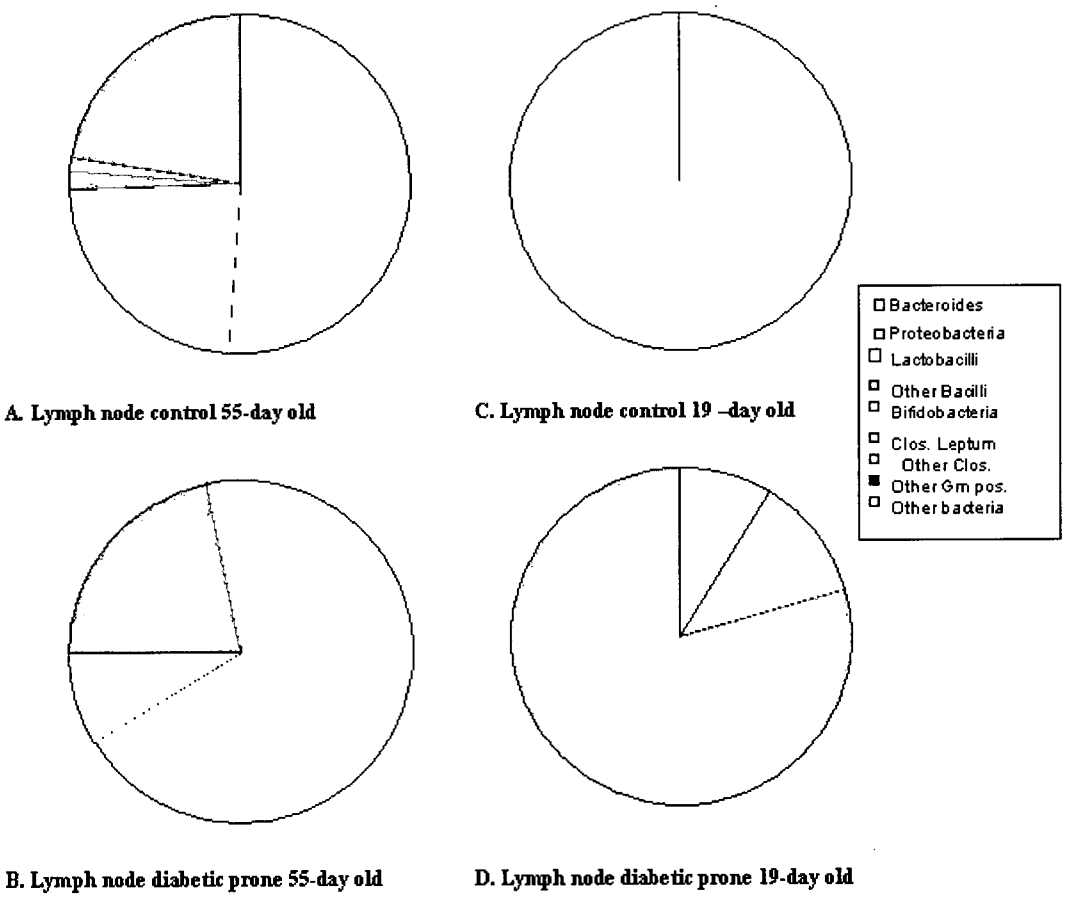


4.6 Library of clones from the lymph node

There were 312 clones obtained from the lymph node tissue of the control and diabetic BB rats. Of these, 18 clones (6%) either could not be sequenced or were identified as chimeras. The remaining 294 clones were aligned within the ARB program; the distribution of the clones among the major groups of bacteria for each of the experimental conditions is shown in Figure 4.6.1. The distribution of clones in each experimental group appears different. It should be noted that the control 19-day old library (LNc19) which consisted entirely of *Lactobacillus* species, was a smaller library than the others ($n = 42$). When I noticed that 93% of the selected clones were homologous to *Lactobacillus reuteri (*divergence 0.2%); and the remaining 7% of selected clones aligned with *Lactobacillus acidophilus*, I determined that further sampling was unlikely to add new information to the clone library. In contrast, the diabetic-prone 19-day old library (LNd19) was dominated by the *C.leptum* group. The 55-day old diabetic and control groups in the lymph node showed a predominance of Gram negative bacteria (*Bacteroides* and *Proteobacteria*).

Figure 4.6. 1: Distribution among the major groups of bacteria of clones from the lymph node of the BB rat.

Genomic DNA was cloned with F940/R1422 at 15 cycles. Experimental groups are shown under each pie chart.



A. Lymph node control 55-day old

C. Lymph node control 19 –day old

B. Lymph node diabetic prone 55-day old

D. Lymph node diabetic prone 19-day old

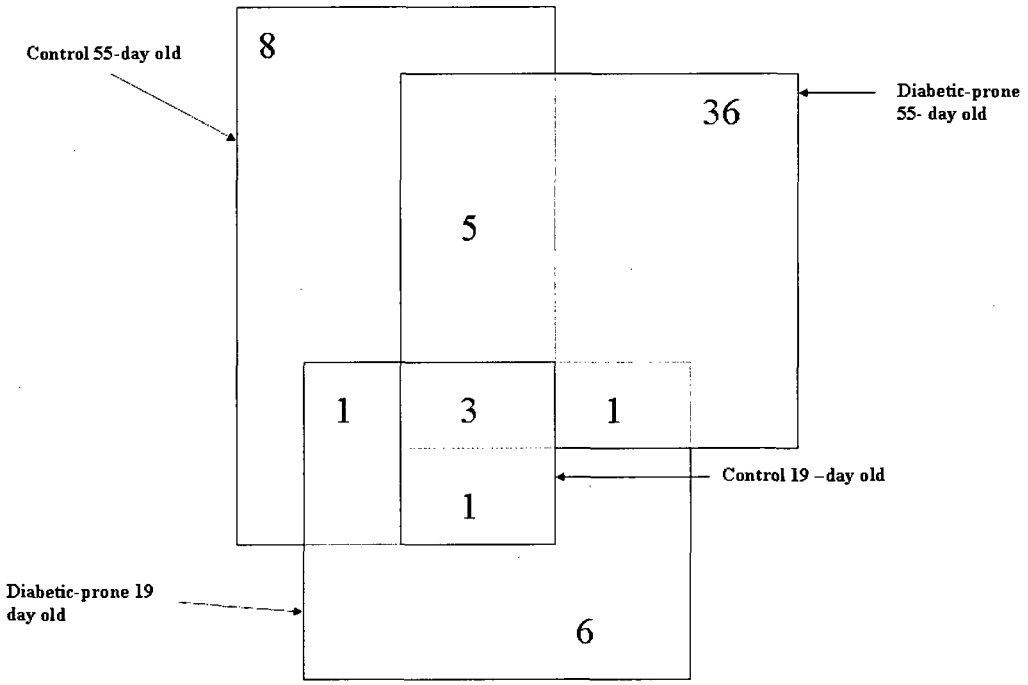
A total of 61 distinct OTUs (using a similarity cut-off of 97%) were identified from the 294 lymph node clones. Calulating the ACE estimate of species diversity (Chao and Lee 1992) predicted 13 OTUs in the LNc55, 37 OTUs in LNd55, 2 OTUs in the LNc19 and 10 OTUs in the LNd19; coverage was estimated to be 100% in each case. The distribution of the clones and OTUs among the major groups and subgroups of bacteria are shown in Table 4.6.1 (see Appendix E, Table E1 for a full list of OTUs). A Venn diagram ((Figure 4.6.2) illustrates the overlap between the communities. A small core group of 3 OTUs (all *Lactobacilli)*, was shared by all of the libraries. The LNc19 commmunity was a subset of the other groups, with no unique OTUs. Shared OTUs accounted for 50% of the LNd19 community, and 55% of the LNc55 group. The LNd55 was a much richer community than the other groups, and had a high proportion of unique OTUs.

Table 4.6. 1: Distribution of lymph node clones and OTUs among the major groups and subgroups of bacteria

Phylogenetic group	RDPII No.	Clones (*n*)	OTUs (*n*)
Bacteroides and *Cytophaga*	2.15.1	33	15
Proteobacteria	2.28	86	11
Bacillus, Lactobacillus, Streptoccus subdivision	2.30.7.17	70	5
Gram-positive bacteria	2.30	5	5
Clostridium lituseburense group	2.30.4.5	5	3
Clostridium coccoides group	2.30.4.1	31	11
Clostridium leptum subgroup	2.30.9.13	63	7
Clostridium botulinum subgroup	2.30.9.2	3	3

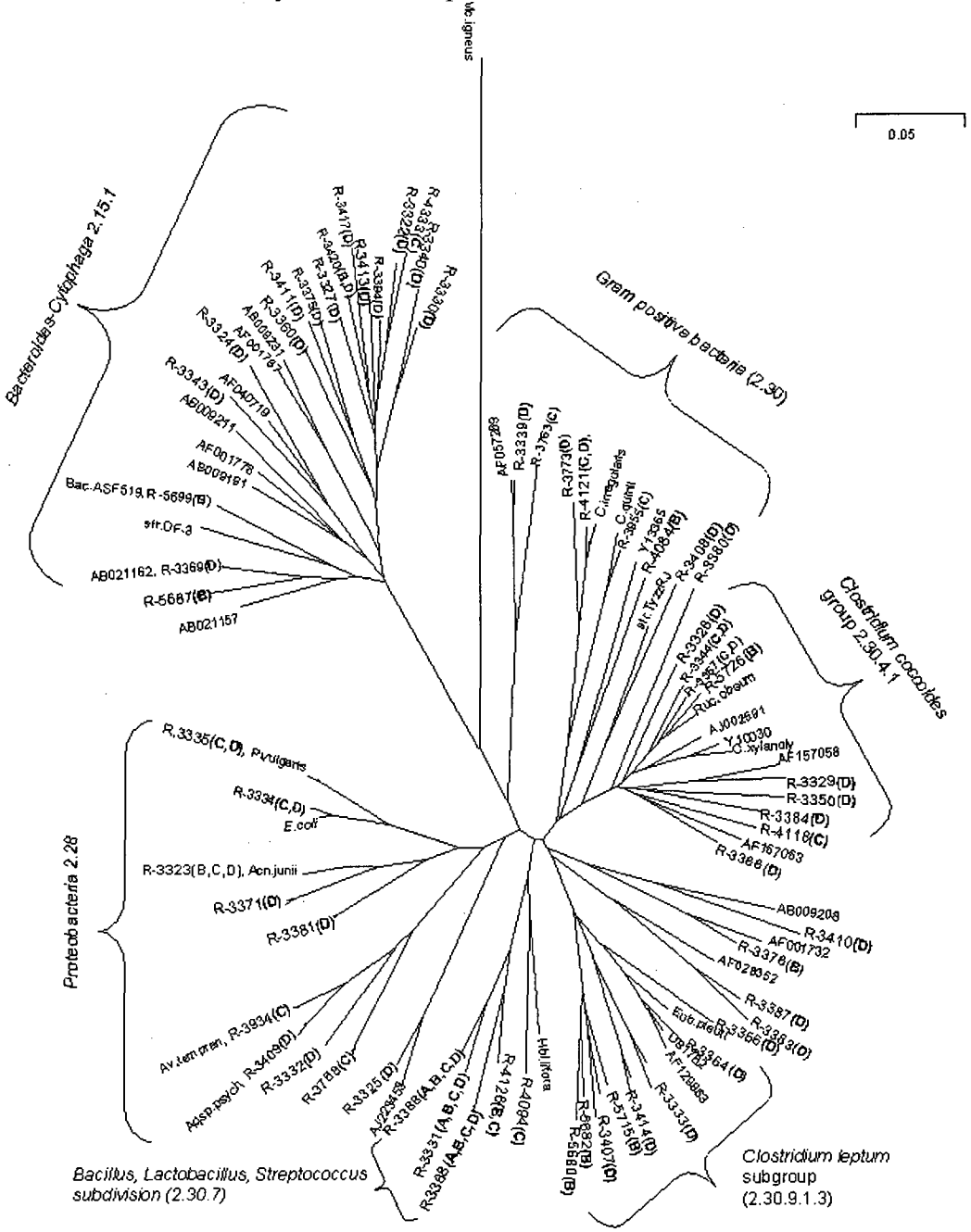
Figure 4.6. 2: Venn diagram comparing the OTU memberships at 0.03% divergence found in lymph node clone libraries

The libraries were (55 day-old control (LNc55, *n* =102), 55 day-old diabetic prone rats (LNd55 *n* = 100), 19 day old control rats (LNc19, *n* = 42) and 19 day-old diabetic prone rats (LNc19, *n* = 68).).



A phylogenetic tree showing the relative positions of the random clones and their nearest neighbours is shown in Figure 4.6.3. A high proportion of clones (29%) was found in the *Proteobacteria* group (86 clones in 5 OTUs); over half of these clones were homologous to *Proteus vulgaris* (divergence 0.8%). Almost all of the *Proteobacteria* OTUs were from the 55 year-old groups, and particularly the LNd55. Several unique OTUs were distantly related to members of the *Bacteroides-Cytophaga* group; the clones from the LNd55 group were again prominent. Both 19 and 55 day-old diabetic prone groups had a strong presence in the *C. leptum* group, while the 55 day-old diabetic-prone condition was evident in the *C. coccoides* group. Almost 25% of the clones were closely related to the *Lactobacilli;* these grouped into 5 OTUs, 3 of which were present in all experimental groups.

Figure 4.6. 3: Rooted phylogenetic tree derived from the F940/R1422 DNA sequences showing the phylogenetic relationships among the lymph node clones. Bar represents a 5% sequence divergence. The tree was constructed using a eighbour-joing method with a Jukes-Cantor correction (1969). The relative position of each clone was supported by bootstrap analysis for 500 trees. Some nearest neighbours are included and indicated by full name or accession numbers. Random clones are shown with and R-(number). The associated experimental condition is indicated in brackets following the clone number, with **A** = 19 day-old control, **B** = 19 day-old diabetic prone, **C** = 55 day-old control and **D** = 55 day-old diabetic prone.



4.7 Library of clones from the faeces

A total of 384 clones were analyzed from the faecal samples. Of these, 22 clones (6%) either could not be sequenced or were found to be chimeras. The nearest relatives of the remaining 362 clones were aligned and identified and similarity matrices were calculated using the ARB program (Ludwig *et al,* 1998). Figure 4.7.1 shows the distribution of the clones from each of the experimental groups, among the major groups of bacteria. The Gram-negative bacteria (particularly those in the *Bacteroides-Cytophaga* division) represented 67% of all clones in the Faeces libraries. In the Fc19 library, this group formed 79% of the total library. The *Bacteroides* comprised 53% of the Fd19 library; the *Proteobacteria*, namely *Escherichia coli*, were also significant in the Fd19 library (16%). The *Clostridiales* formed 10% (Fc19) and 14% (Fd19) of the Faeces libraries. The *Lactobacilli* were found to comprise only 10-14% of the faeces communities.

The proportion of *Bacteroides* was reduced in the 55 day-old groups, where they represented 37% of the Fc55 and 45% of the Fd55. In the SIc55, SId55 and LNc55, the *Bacteroides* clones formed 1-2% of the total clones; however, they represented 28% of the LNd55. The *Proteobacteria* were only 1% of the total number of clones in the 55 day-old faeces group, which was consistent with the proportion in SIc55,SId55 and LNc55, but not LNd55. In the latter group, 38% of clones matched the *Proteobacteria*. The *Clostridiales* family gained prominence in the 55-day old groups, with 41% (Fc55) and 35% (Fd55) of clones. The SIc55 clones included only 7% *Clostridia*, but the SId55,LNc55 and LNd55 had 18%-24% of clones in this bacterial group. Finally, the proportion of *Lactobacillus* clones changed very little in the Fc55 relative to the Fc19

(10%, up from 8%), but increased in the Fd55 (35%, up from 17%). Note that the *Lactobacilli* represented 61% - 86% of the small intestine clones for this age group, 9% of LNc55 and 24% of the LNd55 clones.

Using a 97% similarity cut-off, 127 distinct OTUs were identified from the 362 aligned clones. The majority of OTUs fell within the Gram-positive bacteria. Table 4.7.1 shows the distribution of faeces clones and OTUs among the major groups and subgroups of bacteria; a complete list of all OTUs can be found in Appendix F, Table F1. Estimates of species diversity using the method of Chao and Lee (1992) predicted 84 OTUs in the Fc55, 46 OTUs in the Fd55, 33 OTUs in the Fc19 and 18 OTUs in the Fd19; coverage ranged from 64% for the Fc55 to over 100% for the Fc19 and Fd19. Overall, only a relatively small number of OTUs were shared between the libraries, as can be seen in Figure 4.7.2. A very small core community of 2 OTUs (both *Lactobacilli*) was found in all four libraries; and each community had a high proportion of unique OTUs.

Figure 4.7. 1: Distribution among the major groups of bacteria of clones from the faeces of the BB rat.

Genomic DNA was cloned with F940/R1422 at 15 cycles. Experimental groups are shown under each pie chart.

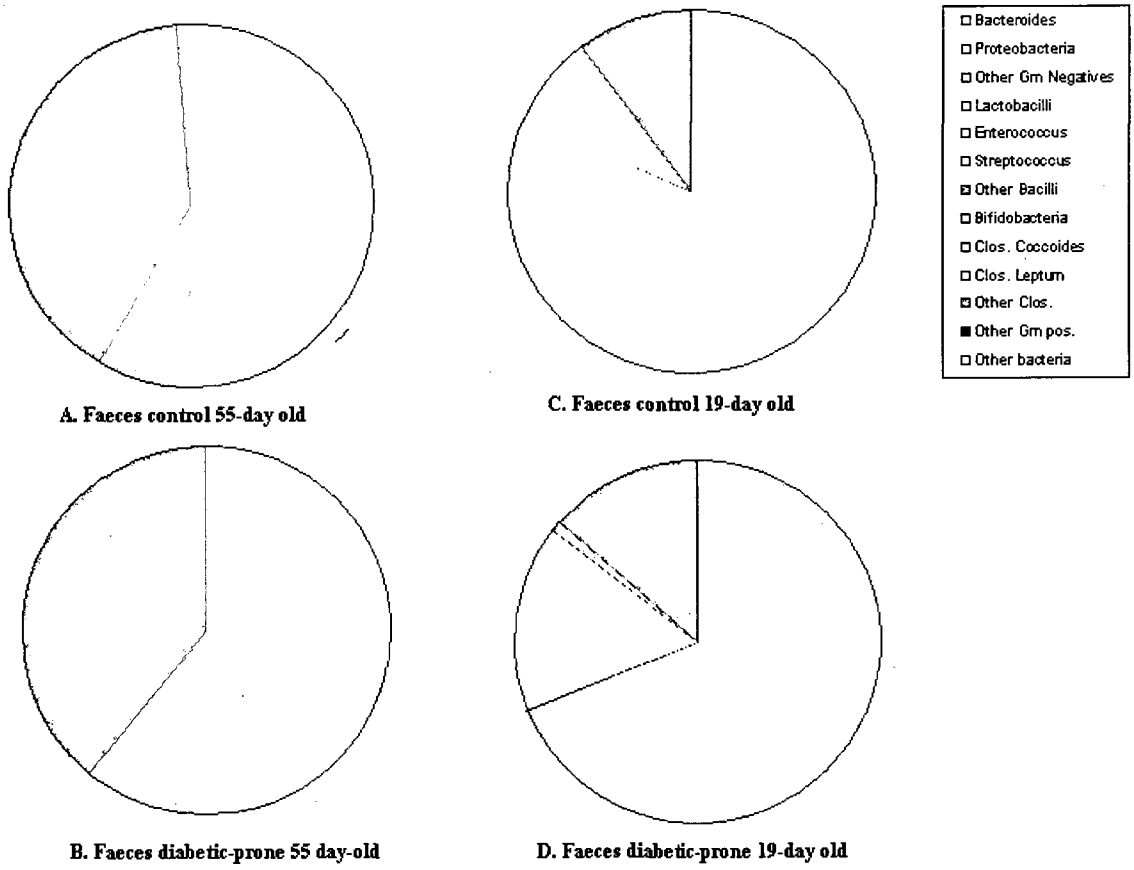
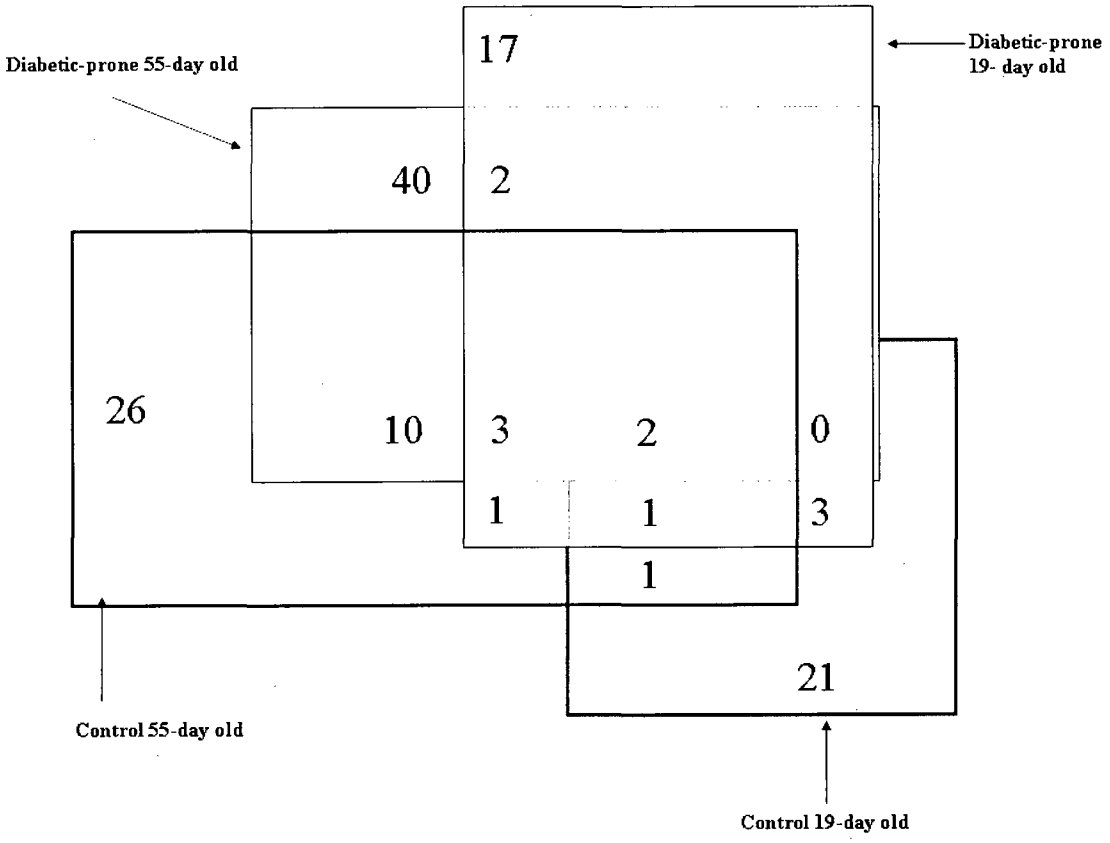


Table 4.7. 1: Distribution of faeces clones and OTUs among the major groups and subgroups of bacteria

Phylogenetic group	**RDPII No.**	**Clones (*n*)**	**OTUs (*n*)**
Bacteroides and *Cytophaga*	2.15.1	219	52
Proteobacteria	2.28	23	4
Bacillus, Lactobacillus, Streptoccus subdivision	2.30.7.17	57	10
Gram-positive bacteria	2.30	22	29
Clostridium lituseburense group	2.30.4.5	1	2
Clostridium coccoides group	2.30.4.1	6	6
Clostridium leptum subgroup	2.30.9.13	34	24

Figure 4.7. 2: Venn diagram comparing the OTU memberships at 0.03% divergence found in faeces clone libraries

Librarie are 55 day-old control (Fc55, n = 83), 55 day-old diabetic prone rats (Fd55 n = 79), 19 day old control rats (Fc19, n =) and 19 day-old diabetic prone rats (Fd19, n = 98).



A phylogenetic tree was derived using the faeces sequences derived from the V6-V8 variable region of the 16S rDNA (primer F940/R1422). Subtrees showing the relative positions of the faeces random clones and their associated experimental condition are shown in Figures 4.7.3 to 4.7.7. Figure 4.7.3 shows the Gram-negative clones related to the *Proteobacteria;* the *Bacteroides-Cytophaga* cluster is expanded in Figure 4.7.4. Only 5% of clones (3% of OTUs) were identified as *Proteobacteria.* Overall, these OTUs were most often associated with the 19-day control condition. Clone R-4671 was identified as an isolate of *E.coli* (divergence <0.000), and was identified almost exclusively from the 19-day diabetic-prone condition (16 of 17 picks). Figure 4.7.4 shows the phylogenetic subtree for the Gram-negative clones related to the *Bacteroides* and *Cytophaga.* This group dominated the faeces library, both in number of clones (61%) and unique OTUs (41%). Two large clusters of OTUs (identified as Cluster I and Cluster II in Figure 4.7.4) showed no close homology with any previously cultured bacteria and represent a new and separate lineage within the phylum. Cluster I contained 16 OTUs separated into several deeply branched lineages; Cluster II contained 21 OTUs, which were also separated into several deeply branched lineages. Although the OTUs in both clusters were found in every condition, the 55-day old group is slightly more evident. The phylogenetic subtree for the Gram-positive faeces clones, including the *Bacillus*, *Lactobacillus* and *Streptococcus* subdivison, are shown in Figure 4.7.5. Fewer *Lactobacillus* were found in the Faeces as compared to the small intestine and lymph node libraries (12% of clones in faeces compared to 24% in the lymph nodes and 85% in the small intestine) although a core group of *Lactobacilli* OTUs can be seen in each tree and are associated with all conditions (*L. acidophilus, L, acidphilus subsp.johnsonni, L. reuteri* and *L. murinus).* A

single clone R-4620 from the 19 day-old diabetic prone faeces library was very closely related (divergence of 3.04%) to *Streptococcus infantis* (Figure 4.7.5). The collapsed *Clostridium leptum* and *Clostridium coccoides* groups are expanded in Figures 4.7.6 and 4.7.7 respectively. Figure 4.7.6 suggests that although OTUs identified in the *C. leptum* group were identified with every condition, the control 55-yeal old condition predominated. The same is true of the *C. coccoides* group (Figure 4.7.7).

Figure 4.7. 3: Unrooted phylogenetic tree derived from the 16S rDNA variable region 6-8 faeces sequences (*Proteobacteria*).

The phylogenetic relationships among the *Proteobacteria* clones are shown. Bar represents a 5% sequence divergence. The tree was constructed using a neighbour-joining method (1000 boostrap iterations) with a Jukes-Cantor correction (1969). Some nearest neighbours are included and indicated by full name or accession numbers. Random clones are shown with and R-(number). The associated experimental condition is indicated in brackets following the clone number, with **A** = 19 day-old control, **B** = 19 day-old diabetic prone, **C** = 55 day-old control and **D** = 55 day-old diabetic prone. The *Bacteroides* and *Cytophaga* group of Gram –negative bacteria is collapsed (see Fig. 4.7.4).

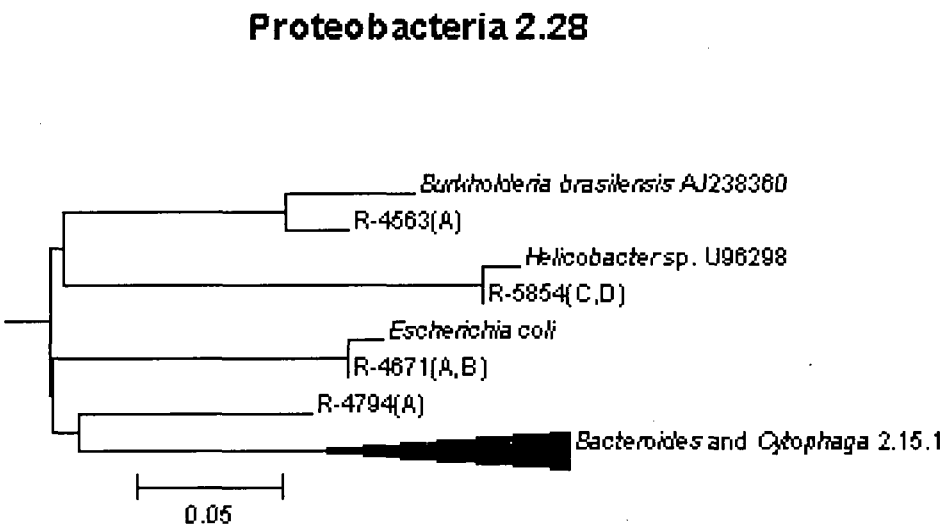


Figure 4.7. 4: Unrooted phylogenetic tree derived from the 16S rDNA variable region 6-8 faeces sequences (*Bacteroides-Cytophaga*).

Phylogenetic relationships among Gram-negative clones in the *Bacteroides-Cytophaga* group are shown. Bar represents a 2% sequence divergence. The tree was constructed using a neighbour-joining method (1000 boostrap iterations) with a Jukes-Cantor correction (1969). Some nearest neighbours are included and indicated by full name or accession numbers. Random clones are shown with and R-(number). The associated experimental condition is indicated in brackets following the clone number, with **A** = 19 day-old control, **B** = 19 day-old diabetic prone, **C** = 55 day-old control and **D** = 55 day-old diabetic prone.

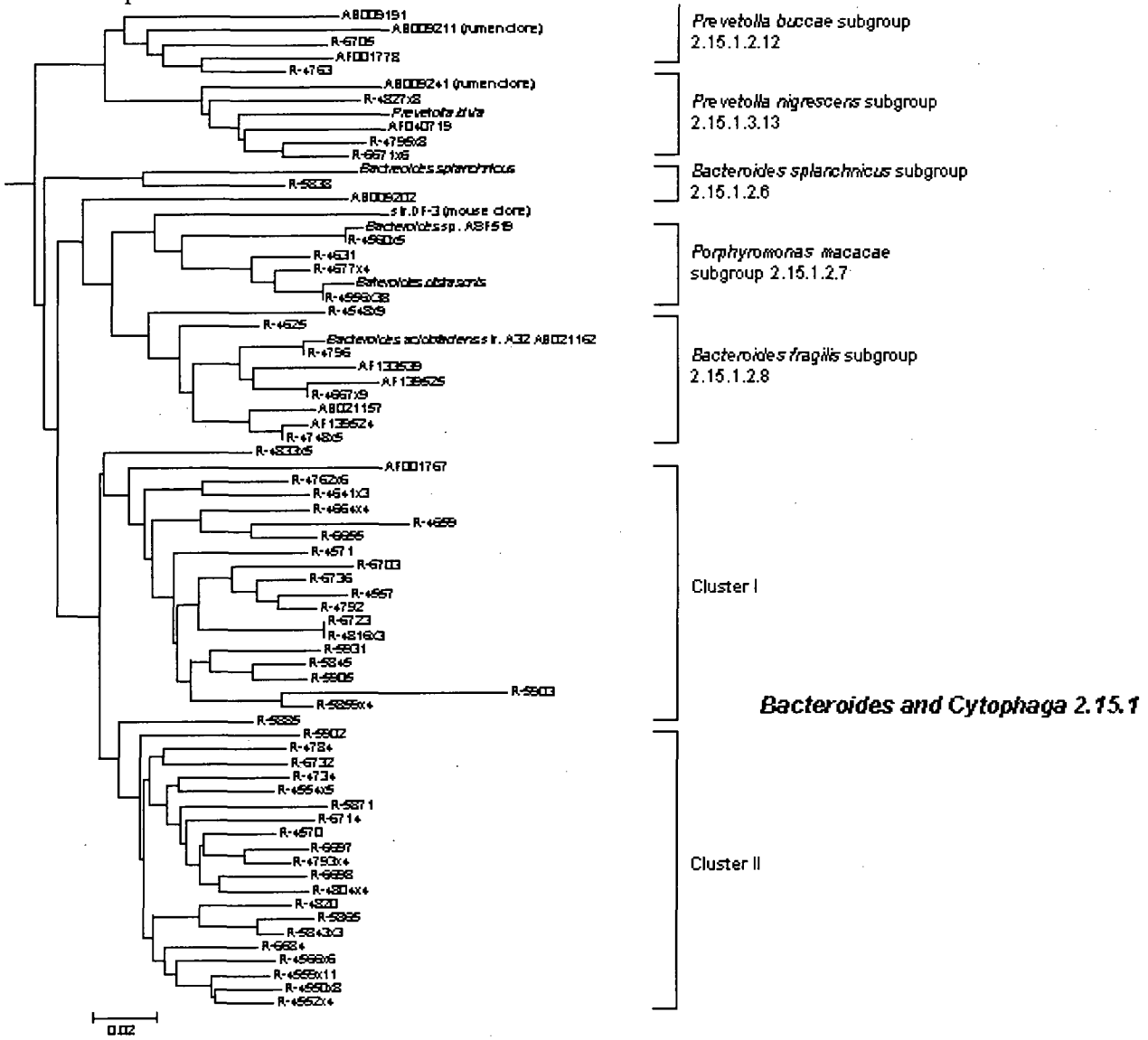


Figure 4.7. 5: Unrooted phylogenetic tree derived from the 16S rDNA variable region 6-8 faeces sequences (Gram positive clones).

Phylogenetic relationships among Gram-positive clones are shown. Bar represents a 2% sequence divergence. The tree was constructed using a neighbour-joining method (1000 boostrap iterations) with a Jukes-Cantor correction (1969). Random clones are shown as R-(number). Some nearest neighbours are included and indicated by full name or accession numbers. Random clones are shown with and R-(number). The associated experimental condition is indicated in brackets following the clone number, with **A** = 19 day-old control, **B** = 19 day-old diabetic prone, **C** = 55 day-old control and **D** = 55 day-old diabetic prone. The *Clostridium leptum* and *Clostridium cocccoides* groups are collapsed (see Figs. 4.7.6 and 4.7.7, respectively).

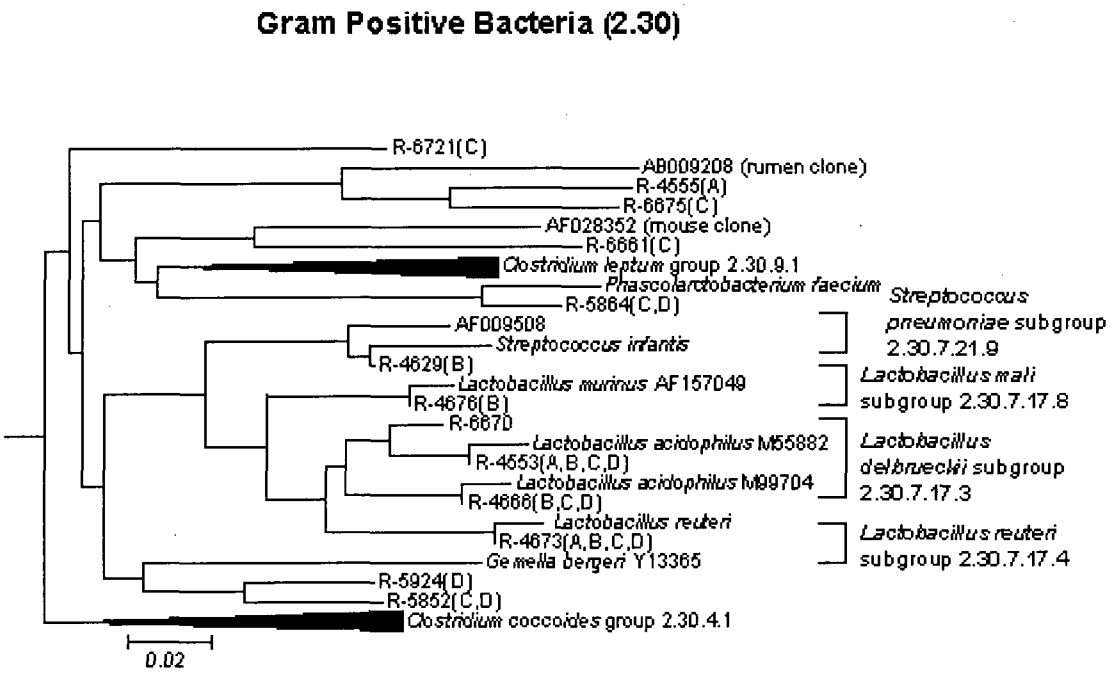


Figure 4.7. 6: Unrooted phylogenetic tree derived from the 16S rDNA variable region 6-8 faeces sequences (*C. leptum* group).

Phylogenetic relationships among bacteria within the *Clostridium leptum* group are shown. Bar represents a 2% sequence divergence. The tree was constructed using a neighbour-joining method (1000 boostrap iterations) with a Jukes-Cantor correction (1969). Some nearest neighbours are included and indicated by full name or accession numbers. Random clones are shown with and R-(number). The associated experimental condition is indicated in brackets following the clone number, with **A** = 19 day-old control, **B** = 19 day-old diabetic prone, **C** = 55 day-old control and **D** = 55 day-old diabetic
prone.

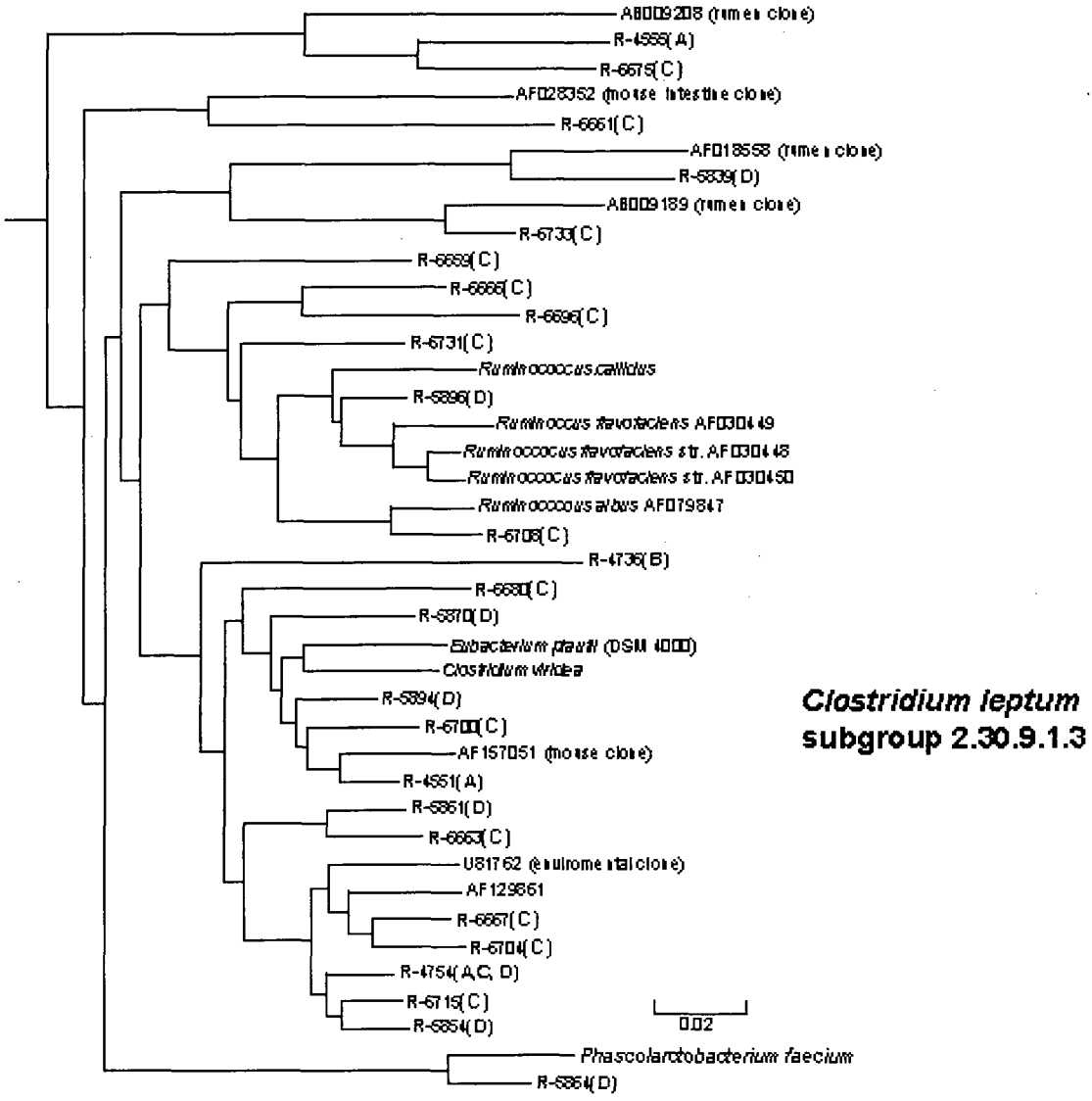
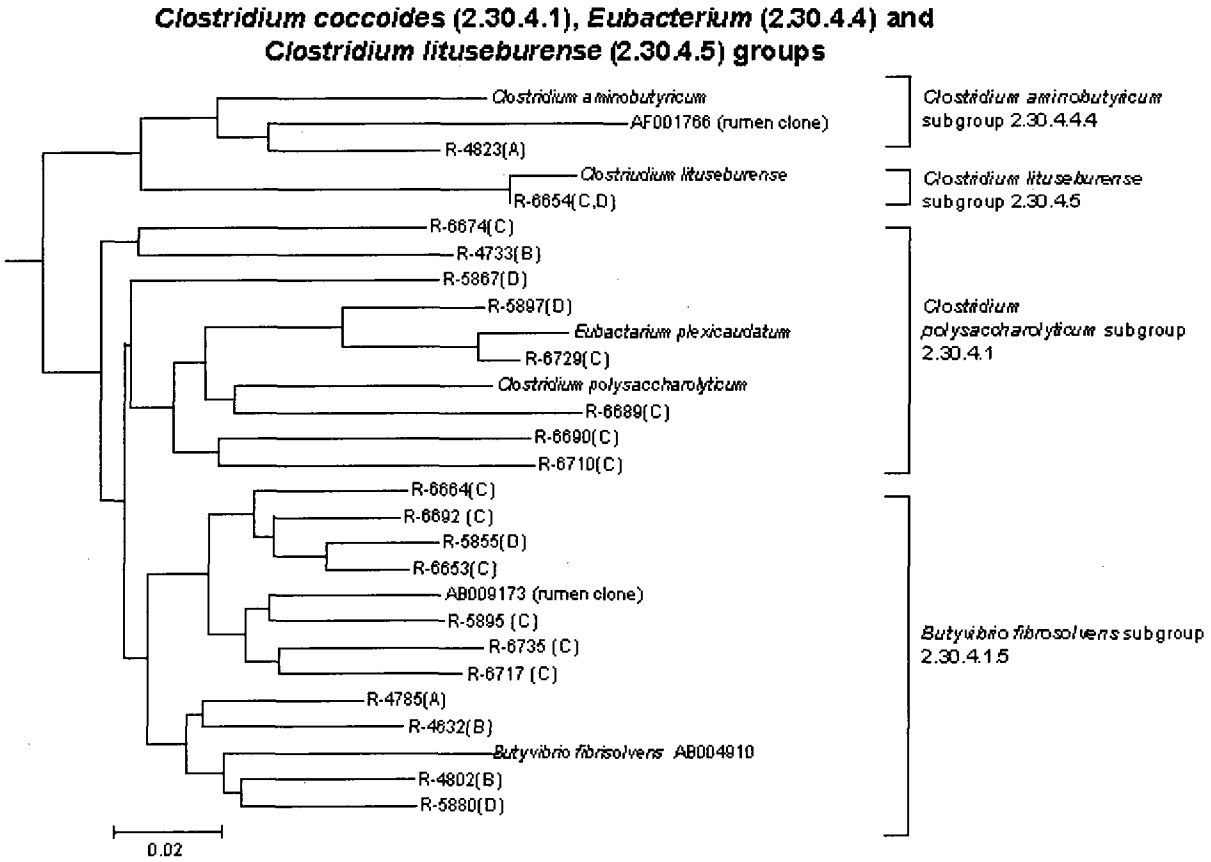


Figure 4.7. 7: Unrooted phylogenetic tree derived from the 16S rDNA variable region 6-8 sequences (other *Clostridia*).

Phylogenetic relationships among bacteria within the *Clostridium coccoides, Eubacterium* and *Clostridium lituseburense* groups are shown. Bar represents a 2% sequence divergence. The tree was constructed using a neighbour-joining method (1000 boostrap iterations) with a Jukes-Cantor correction (1969). Some nearest neighbours are included and indicated by full name or accession numbers. Random clones are shown with and R-(number). The associated experimental condition is indicated in brackets following the clone number, with **A** = 19 day-old control, **B** = 19 day-old diabetic prone, **C** = 55 day-old control and **D** = 55 day-old diabetic prone.



4.8 Statistical analysis of clone libraries

Various statistical tests were performed to detect differences in community membership and structure. For each of the samples (small intestine, lymph node and faeces), four smaller clone libraries were created. Each of the sets of 4 libraries was first compared using ∫-LIBSHUFF (Singleton *et al,* 2001) to the divergence of the homologous and heterologous curves at a given gentic distance.

SONS (Schloss and Handelsman, 2006b) was used to further characterize the similarities and differences between the four small intestine communities. TreeClimber (Schloss and Handelsman, 2006a) was used to see if there is a random distribution of OTUs from each library down each branch of the phylogenetic tree (to determine if there is on condition that is stimulating a distinct group of bacteria). Finally, the phylogenetic trees were assessed using Unifrac (Lozupone and Knight 2005) to examine the total genetic distance associated with each experimental condition and see if it is different.

4.8.1 Small intestine libraries: statistical analysis

The four small intestine libraries consisted of sequences from the 55 day-old control rats (SIc55, $n = 100$), 55 day-old diabetic prone rats (SId55 $n = 98$), 19 day old control rats (SIc19, $n = 65$) and 19 day-old diabetic prone rats (SId19, $n = 81$). The P-values from ∫-LIBSHUFF, Unifrac and TreeClimber are reported in Table 4 .8.1.

TreeClimber compared a PHYLOWIN-generated small intestine tree (constucted using a neighbour-joining, bootstrap method with a Jukes-Cantor correction) to the probability distribution of parsimony scores obtained from 10,000 random joining trees. The phylogenetic relationship demontrasted by the small intestine tree was found to significantly different than what would be expected by chance ($P = 0.0005$).

∫-LIBSHUFF analysis showed that the SId19 library was found to be significantly different from all other libraries ($P < 0.001$ in all cases). Conversely, a UniFrac comparison showed no significant diffrences ($P = 1$) between SId19 and any other small intestine library. The ∫-LIBSHUFF coverage curve for the SId19 compared to the SIc55 library indicates that the SId19 appears to be a subset of the SIc55 library (Figure 4.8.1). This is further supported by examining the fraction of shared OTUs between the communities (Appendix D, Figure D1), which shows that the SId19 library shares 60% of its membership with both the SIc55 and the SIc19 library at a genetic distance of 3%. The abundance-based Jaccard ($J_{abund)}$ indices and community structure similarity ($\theta_{YC)}$ were calculated and are shown in Figure 4.8.2. The indices ranged from 50% to 80%, with the exception of the values for the SId19-SId55.). Interestingly, the 19 and 55-day old diabetic prone communities showed differences in the relative abundance of their members ($J_{abun} = 0.46 \pm 0.28$) and community struture ($\theta_{YC} = 0.24 \pm 0.05$).

∫-LIBSHUFF analysis showed that the SIc19 and SId55 were significantly different (corrected $P = 0.004$); Unifrac showed suggestive, but not significant, differences ($P < 0.1$). SONS (Schloss and Handelsman 2006b) analysis revealed that a high proportion of OTUs (74-94%) were shared between the SIc19 and SId55 libraries. The J_{abund} and θ_{YC} indices for the SIc19 compared to SId55 were high (see Figure 4.8.2), suggesting a high degree of similarity both in abundance of shared OTUs, and distribution of clones within the OTUs, in the two communities. No significant differences were found between the SIc19 and SIc55 communities (see Table 4.8.1 for P values). The communities of the two control libraries, although not identical, were very

similar in both abundance of shared OTUs (J_{abun} = 0.93 ± 0.08) and distribution of clones among the OTUs (θ_{YC} = 0.87± 0.09).

Table 4.8. 1: P-values from small intestine libraries compared with ∫-LIBSHUFF, UniFrac and TreeClimber.

P-values are corrected using the Bonferroni correction and reported for a genetic distance of 0.03; siginifcant *P*-values are shown in **bold.** .

Clone libraries	∫-LIBSHUFF[1] *P*-value	Unifrac[1] *P*-value	TreeClimber *P*-value
SIc19 vs SId19	**< 0.001**	1	
SIc19 vs SIc55	0.0505	< 0.1	
SIc19 vs SId55	**0.004**	< 0.1	**0.0005**
SId19 vs SIc55	**< 0.001**	1	
SId19 vs SId55	**< 0.001**	1	
SIc55 vs SId55	0.2552	< 0.1	

[1] For each multiple library comparison, the corrected *P*-value (95% CI) is 0.0043 (Henrikson, 2004).

No significant differences were found between the SIc55 and the SId55 libraries ($P = 0.2552$; the UniFrac analysis showed suggestive (but not significant) differences between these communities ($P<0.1$). This is consistent with the ∫-LIBSHUFF comparison of the SIc55 and SId55 full-length 16S sequence library, which showed no significant differences between the two groups ($P = 1.00$; see Appendix A).

Figure 4.8. 1: Results of ∫-LIBSHUFF comparison .

Homologous and heterologous curves for 16S rRNA gene sequence libraries from small intestine samples (55 day old control and 19 day old diabetic prone) are shown. Distance is equal to the Jukes-Cantor (1969) corrected evolutionary distance.

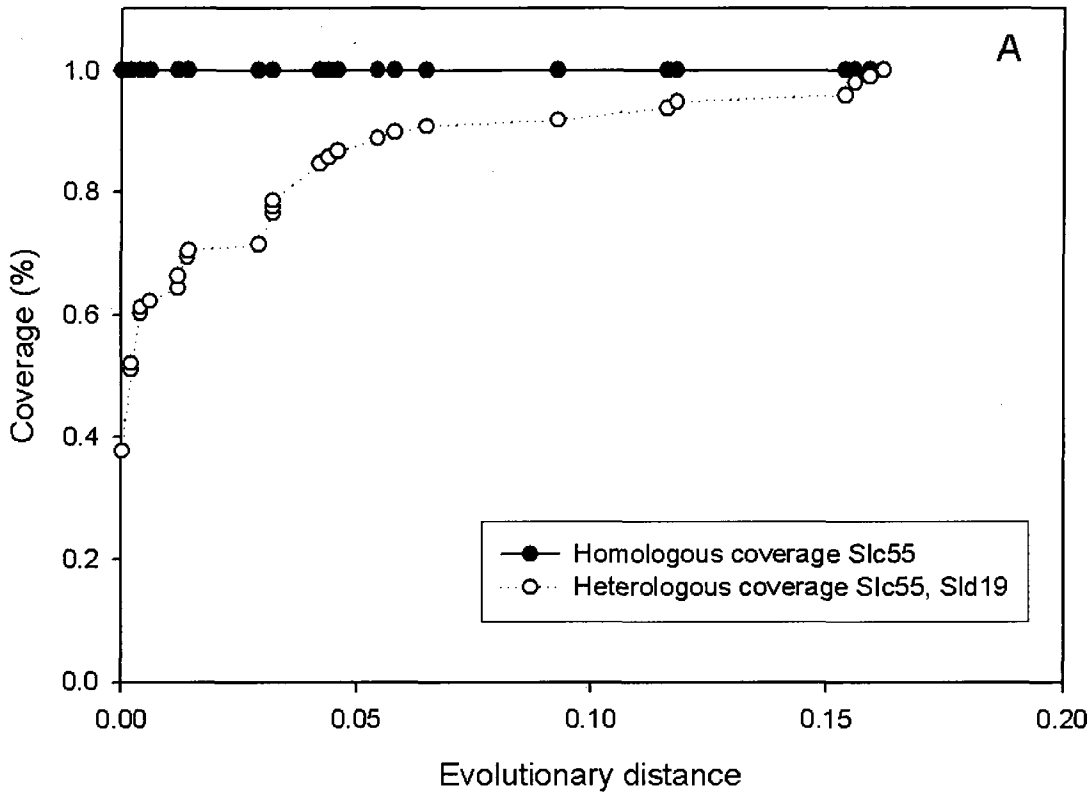
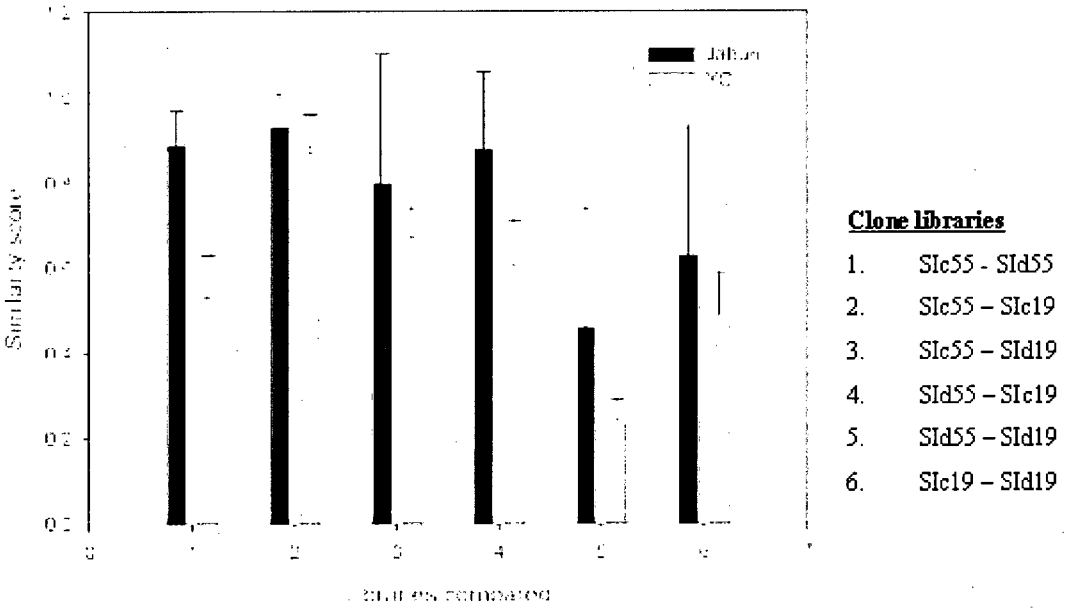


Figure 4.8. 2: Estimates of similarity in community membership (Jabund) and structure (θ_{YC}) between small intestine clone libraries from different conditions. Standard error bars for 95% CI are shown. SI indicates small intestine followed by an indication of condition (c for control, d for diabetic prone) and age (19 or 55 days).



4.8.2 Lymph node libraries: statistical analysis

The lymph node clone libraries were control 55-day old (LNc55, n = 96), diabetic prone 55-day old (LNd55, n = 100), control 19-day old (LNc19, n =30), and diabetic prone 19-day old (LNd19, n = 68). Once again, the statistical tests differed in their results. The ∫-LIBSHUFF analysis showed that all four lymph node clone libraries differed significantly ($P < 0.001$ in all cases). However, the results from the UniFrac analysis were not significant, although suggestive (corrected $P = 0.06$ in all cases). The TreeClimber parsimony score for the lymph node tree was highly significant (P = 0.0001).

The J_{abund} values for most of the lymph node library comparions were very low, as were the community structure similarity estimates (θ_{YC}), with the exception of the LNc55-LNd55 (J_{abund} = 0.67, se < 0.0); Figure 4.8.3). Assessing the fraction of sequences in each lymph node community that belonged to shared OTUs showed the LNc19 library to be a subset of the other libraries; LNc19 shared 100% of its membership with the other groups (illustrated in Figure 4.8.4 for the case of LNc19 vs. the LNc55 library).

Figure 4.8. 3: Estimates of similarity in community membership (J_{abund}) and structure (θ_{YC}) between lymph node clone libraries from different conditions.
Standard error bars for 95% CI are shown for θ_{YC}; the SE values for J_{abund} were < 0.000. LN indicates lymph node followed by an indication of condition (c for control, d for diabetic prone) and age (19 or 55 days).

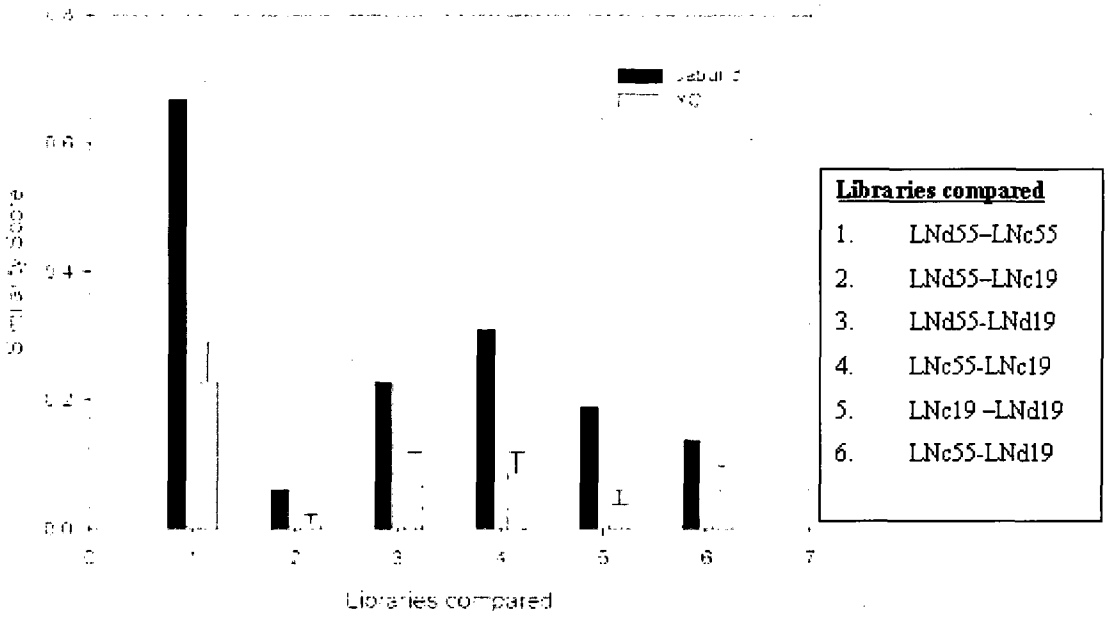
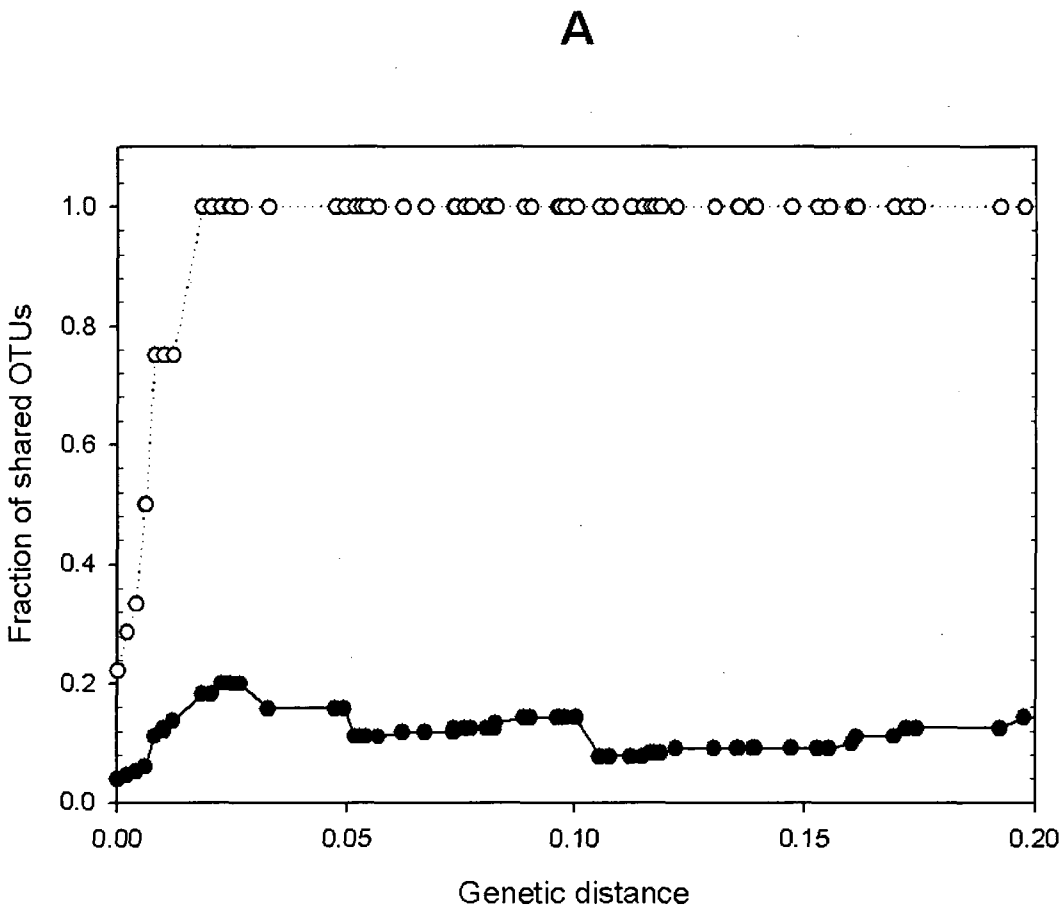


Figure 4.8. 4: Comparison of Lymph Node Libraries – Fraction of Shared OTUs

A. The estimated fraction of shared OTUs between LNc55 (Aotu shared) and LNc19 (Botu shared) libraries is shown. Aotu shared is less than 20% at a genetic distance of 0.03, and remains at this level even at a distance of 0.20. This means that there are many OTUs in LNc55 that are not found in the LNc19 library. Conversely, Botu shared rises very quickly, reaching 100% at a genetic distance of 0.018. it can be concluded that LNc19 is a subset of the LNc55 library, as all of the OTUs found in the former are shared with the latter.



4. 8.3 Faeces libraries: statistical analysis

The four faecal clone libraries were: control 55-day old (Fc55, n = 73), diabetic prone 55-day old (Fd55, n = 83), control 19-day old (Fc19, n = 96), and diabetic prone 19-day old (Fd19, n = 96). The results of library comparisons using ∫-LIBSHUFF, UniFrac and TreeClimber are presented in Table 4.8.2 . All libraries were found to be significantly different from one another by ∫-LIBSHUFF ($P \leq 0.0024$). Unifrac showed no signficant differences among faecal libraries. The parsimony score generated by TreeClimber for the faeces phylogenetic tree was highly significant ($P = 0.0004$).

The fraction of shared OTUs between Fc19 and all other libraries was low (less than 27% in all cases; see Appendix F, Figure F2 A and C). The J_{abund} index and θ_{YC} are shown in Figure 4.8.5. The coverage curve (Appendix F, Figure F1 B) of the Fd19 library (Fd19 was significantly different ,$P = 0.0004$, from the Fd55 library) showed that the Fd19 community was a subset of the Fd55 community. The OTU communities from Fd19 and Fd55 libraries had different abundances of shared OTUs ($J_{abun} = 0.11 \pm 0.06$) as well as differences in the distribution of clones among OTUs ($\theta_{YC} = 0.25 \pm 0.08$). Although Unifrac showed no significant differences between the Fc55 and Fd55 communities, the J_{abun} (0.45 ± 0.2) and θ_{YC} (0.25 ± 0.08) are relatively low, lending some weight to the significant P value from the ∫-Libshuff analysis.

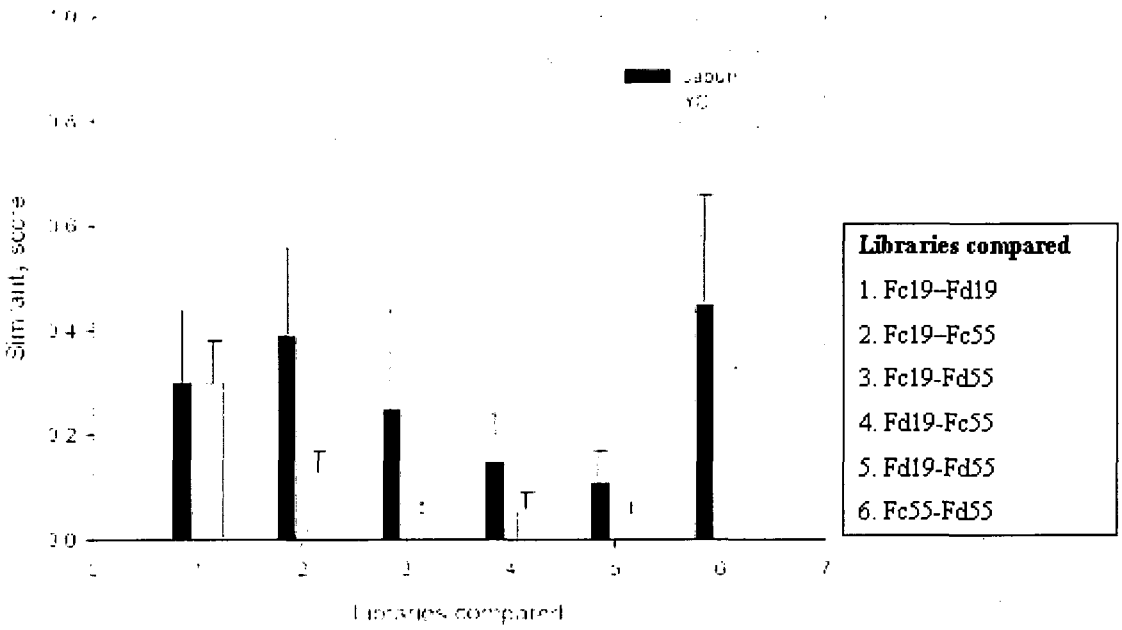
Table 4.8. 2: *P*-values from faeces libraries compared using ∫-LIBSHUFF, UniFrac and TreeClimber.

P-values are reported for a genetic distance of 0.03; siginifcant *P*-values are shown in **bold.** .

Clone libraries	∫-LIBSHUFF[1] *P*-value	Unifrac[1] *P*-value	TreeClimber *P*-value
Fc19 vs Fd19	**< 0.001**	≤ 0.06	**0.0004**
Fc19 vs Fc55			
Fc19 vs Fd55			
Fd19 vs Fc55			
Fd19 vs Fd55	**< 0.0004**		
Fc55 vs Fd55	**0.0024**	1	

1 For each multiple library comparison, the corrected *P*-value (95% CI) is 0.0043 (Henrikson, 2004).

Figure 4.8. 5: Estimates of similarity in community membership (J_{abund}) and structure (θ_{YC}) between lymph node clone libraries from different conditions. Standard error bars for 95% CI are shown for θ_{YC}; the SE values for J_{abund} were < 0.000. LN indicates lymph node followed by an indication of condition (c for control, d for diabetic prone) and age (19 or 55 days).



Chapter V

5.1 Discussion

A culture-independent approach based on sequence variability in the 16S rDNA gene was used to analyse bacterial communities along different parts of the gastro-intestinal tract of the BB rat, to study a possible relationship between gut microbiota and Type 1 diabetes. The null hypothesis was that there is no consistent difference in bacterial populations in BBc annd BBdp rats. Any inherent differences between individuals were expected to be minimal as the BB rats used in this study were all housed in close quarters, bred in the same rooms and were genetically similar (80% identical at the DNA level) Pozzilli *et al* 1993; Scott 1996). Feeding the animals the same diet from weaning and providing water treated to 95% purity by reverse osmosis, controlled any differences in microbiota that could have been introduced. As an additional part of this study, several primers sets were evaluated to identify any bias caused by primer length and number of PCR cycles.

Overall, results of the primer comparison suggested that the HDA1 / HDA 2 (V3 region) primers at both 15 and 35 cycles will amplify the variable regions of the 16s gene and produce clone libraries that are reasonably comparable to the full-length 16S F27 / R1492 primer library ($P = 1$), although at 35 cycles, there was a discernible influence on the species detected (Figure 4.2.1). Suzuki and Giovannoni (1996) developed a kinetic model that showed how in reactions with mixed templates and high amplification efficiency, the template with the higher initial concentration in the mixture reaches inhibitory concentrations sooner. The second template continues to be amplified efficiently, until eventually, the original concentration difference disappears and a 1:1

ratio exists. This bias is strongly influenced by the number of cycles, with lower cycle numbers reducing the bias. In this case, it can be speculated that the *Lactobacilli* is the template with the higher initial concentration, and the *Clostridiales* the template with a, lower concentration. The additional cycles in the 35-cycle PCR thus favoured the extra amplification of the *Clostridiales*.The F940/ R1422 (V6-V8 region) primer set was also found to be comparable to the 16S primers ($P = 1$). In this study, the F940/R1422 primers showed the highest correlation to the full-length 16S primers in terms of community membership and structure (Figure 4.2.1). In addition, these primers consistently and reliably amplified the samples for the comparison study. Difficulties with the full-length 16S primers were previously described, and may be due to degraded DNA. It is particularly challenging to amplify full-length sequences from the lymph nodes, due to degraded DNA. The F940/R1422 amplifies a shorter fragment of DNA ($\approx$ 500 bp with 3 variable regions), thus seeming to avoid this problem, while providing more sequence data for identifying clones than the HDA1/2 primers (higly variable, but only approximately 200 bp long). The HDA1/2 region will not distinguish among genus very well; the F940/R1422 offers better potential for identification at a finer level. Finally, the F940/R1422 primers performed marginally better than the HDA1/ HDA2 primers when matched to sequences in the RDP II database ($\approx$ 91% match compard to $\approx$ 87% match). Given these considerations, in this study the HDA1 / HDA2 primers were used in the DGGE analysis, and the F940/R1422 primers were selected for further random cloning, which allowed a comparison between the survey and the DGGE data, while accurately representing the distribution of bacteria.

The DGGE analysis (Figure 4.3.1) provided valuable insight into potential differences between microbial communities from control and diabetic-prone rats, and also highlighted age-related differences in the communities. Non-metric multi-dimensional scaling was used to graphically display differences among experimental conditions. NMDS reduces a DGGE fingerprint to a single point in a two dimensional space to represent relative distances among experimental conditions (Fromin *et al,* 2002). By observing the relationship among these points, an idea of the relative relatedness among bacterial libraries can be visualized. This can give a clearer picture of trends in community progression, for example over time (Diez *et al,* 2001) or to get better idea of the relatedness among the points (Widmer *et al,* 2001). This analysis showed that the DGGE bands from the lymph node of the 55 day-old animals (LNc55 and LNd55) are more similar to each other than to the bands from the other conditions. Other patterns are also discernable: both clusters represent the 55 day-old groups while the 19 day-old groups are scattered more widely. This is reflected in the DGGE profiles of Figure 4.3.1 and suggests that there are age-related differences in the gut microbiota. Similar differences have been identified during maturation of humans (determined by both viable bacterial counts and 16S sequencing) when children have been compared to elderly subjects or when young adults have been compared to elderly adults (Hopkins *et al,* 2001, Mueller *et al,* 2006). Maturation changes have also been followed over the first two years of life in breast fed and bottle fed infants (Morelli, 2008, Enck et al 2009).

Other trends are also apparent. It appears that the SId55 individuals were widely scattered but the LNd55 groups were more closely aligned. This raises the possibility that

specific bacteria were selectively sampled from the small intestine by the mesenteric lymph nodes in the diabetic-prone BB rat (see below for a complete discussion).

Profiling of bacterial communities with denaturing gradient gels has proven to be a powerful method for a culture-independent analysis of large numbers of inidividual samples (Smalla *et al,* 2001). However, a community fingerprinting method such as DGGE is not a useful way to assess the richness or diversity of a complex microbial community (Bent *et al,* 2007). In an effort to complement the DGGE analysis and explore community structure and richness more thoroughly, we decided to conduct further comparative analysis of clone libraries by random cloning of the 16S rRNA gene.

A total of 1000 clones were sequenced from twelve experimental conditions as noted in Table 4.4.1. Once again, non-metric multidimensional scaling (NMDS) was used to examine the overall relationship between the clone libraries. Because these libraries represent pools from each experimental condition, only a single point on the NMDS graph could be plotted. This analysis showed that the SI communities clustered closely together and included the LNc19 group. The Faeces communities, on the other hand, were far removed from the SI groups. This divergence reflects differences in OTU composition among the different locations along the GI tract. Other studies have also found (using random cloning of the 16S gene and DGGE analysis) that the dominant species in the faeces were different from those in the ascending, transverse and descending colon (Zoetandal *et al,* 2002), as well as in the cecum (Marteau *et al,* 2001). Bacterial colonization along the human GI tract varies greatly and there are significant differences in community composition and species abundance. Both the oral cavity and oesophagus are colonized by communities containing a wide variety of aerobic,

facultative anaerobes and anaerobic species, although a recent large scale molecular analysis suggests that only 8 genera account for 70% of phylotypes present in saliva samples. The stomach, duodenum and jejunum contain low numbers of microorganisms ($<10^4$ cfu/mL) with communities dominated primarily by lactic acid bacteria (Tannock, 2007). Further along the GI tract, the microbial load and species diversity significantly increases at the ileum ($>10^8$ cfu/mL) and throughout the colon ($>10^{11}$ cfu/mL). These communities are dominated by obligate anaerobes.

Finally, the LNc55, LNd55 and LNd19 groups are interspersed between the SI and LN groups. At the beginning of this study, it was thought that the microbiota detected in the lymph node would reflect sampling of the bacteria from the small intestine through the Peyer's patches, as these are the major inductive sites of the small intestine (Suzuki *et al*, 2007). However, a recent study measured substantial amounts of lymphoid tissue analogous to small intestinal Peyer's patches in the rat and mouse colon (McConnell *et al,* 2008). In this study, both the NMDS and analysis of the shared OTUs suggests that the gut-associated lymphoid tissue, particularly in the 55 day-old animals, is sampling bacteria from both the small intestine and the distal colon.

The distribution of the clones from the small intestine, lymph node and Faeces are shown in Fig. 4.51, Fig 4.61 and Fig 4.7.1, respectively. The predominant bacteria in the small intestine (98% of the SIc19 and SId19, 86% of the SIc55 and 62% of the Sd55) were identified as *Lactobacillus* species. In the SId55 *Clostridium* species had an important presence (18% of clones). The distribution of the Faeces clones is dramatically different from that of the small intestine clones (Figure 4.5.1) and the lymph clone nodes (Figure 4.6.1). The Gram-negative bacteria (particularly those in the *Bacteroides-*

Cytophaga division) represent 67% of all clones in the Faeces libraries. In the Fc19 library, this group is 79% of the total library, compared to less than 2% in the SIc19 and LNc19 groups. The *Bacteroides* comprised 53% of the Fd19 library, compared to 1% in the SId19 and 7% in the LNd19. *Escherichia coli* was also significant in the Fd19 library (16%). The *Clostridiales*, which are noticeably absent in the SIc19, SId19, and LNc19, make up 10% (Fc19) and 14% (Fd19) of the Faeces libraries. Interestingly, the LNc19 family consisted entirely of 4 *Lactobacillus* OTUs, yet 75% of the LNd19 library consisted of a previously uncultured bacterium distantly related to *Eubacterium plautii* (divergence 7.6%), a member of the *Clostridiales*. *Eubacterium plautii* has been shown in one case to cause severe infection after translocation in a transplant patient (Orlando *et al,* 2008). It is speculated that this relative of *E. plautii* could have a role in the etiology of Type 1 diabetes, and could be a useful target for future studies affecting the composition of the intestinal microflora. The *Lactobacilli*, which were 97%-98% of the 19-day old small intestine communities were found to be only 10-14% of the faeces communities. The 55-day old diabetic and control groups in the lymph node show a predominance of Gram negative bacteria (*Bacteroides* and *Proteobacteria*).

Using the parsimony test (Schloss and Handelsman, 2006a) determined that the phylogenetic relationships seen in the small intestine libraries ($P = 0.0005$), Faeces libraries ($P = 0.0001$) and the lymph node libraries ($P = 0.0004$) where unlikely to have occurred by random variation alone. A lower parsimony score than is expected by chance indicates that population is experiencing selection pressure (Schloss and Handelsman, 2006a). Significance for the parsimony test signals two important messages: 1) there is

less genetic diversity within each community than for the combined communities, and 2) the different communities harbour distinct phylogenetic lineages (Martin, 2002).

A pair-wise comparison conducted separately for each of the small intestine, Faeces and lymph node libraries found signifcant differences between some of the experimental groups. In the small intestine, noteworthy differences were found between the SId19 amd SId55 libraries. The SId19 is significantly different ($P < 0.0001$) in clone library composition, and has a different community abundance and structure, than the 55 year-old diabetic prone group. These changes in the small intestine gut microbiota could suggest either a symptom or a contributor to onset of Type 1 diabetes, given that 55 days of age is just prior to onset of the disease in the BB rat. Note that the SId19 appears to be a subset of the communities found in the control groups (SIc19 and SIc55). No significant differences were found between the 55 day-old control and diabetic prone groups, which share a common membership, abundance and community structure. Significant differences were found in the faeces libraries between the Fc19 and Fd19 libraries, as well as each of the 55 day –old (control and diabetic-prone) libraries were supported by apparent differences in abundance of shared OTUs and community structure. The control and diabetic prone 55 day-old libraries show some differences in community structure and membership. Because the significance is not as strong for the Fc55 vs Fd55 comparison ($P = 0.0024$), it is speculated that they are due to high species diversity rather than deep phylogenetic (i.e. lineage) differences. Given the differences seen in this study between the small intestine and faeces microbial communities of the BB rat, it may be difficult to infer the physiological or immunological function of specific bacteria to the intestinal ecosystem based on changes seen in the faecal communities.

However, all of the lymph node libraries were found to be significantly different from one another ($P < 0.001$ in all cases) (the LNc19 library proved to be a subset of the other libraries. The statistical analysis of the lymph node libraries found that the phylogenetic relationships in evidence could not be attibutated to random variation, suggesting that some other selective pressure is at work. The specialized antigen-presenting cells in the lymph nodes are the primary interaction between gut contents and the mucosal gut immune system (Nagler-Anderson, 2001). The distribution of OTUs seen in the lymph node in this study reflects both the small intestine and faeces populations, providing evidence that the gut-associated lymphoid tissue is presenting antigens from both areas of the GI tract. Thus, both the small intestine and faeces communities may have a role in aberrant gut immune system responses contributing to the onset of Type 1 diabetes.

It must be noted that the library comparisons using TreeClimber, ∫-Libshuff and Unifrac did not agree completely with each other: Where ∫-Libshuff and TreeClimber found significant differences, Unifrac identified only suggestive or no differences. The Unifrac analysis, which measured total genetic distance for each condition, showed that there were no unique sub-groups of bacteria related to only one condition; and that the distribution of clones down the branches of the phylogenetic tree is not different from random. The ∫-Libshuff analysis showed that it takes a large genetic distance before the libraries are similar, suggesting that there are different OTUs down these branches; however, they are not distributed into a unique group; the TreeClimber analysis supported this finding, based on OTUs rather than distribution of clones. Taken together, the statistical analyses suggest that the proportion of bacteria is probably not significantly

different, although the trend may be there; but the actual species distribution is different between conditions.

It is acknowledged that observations in animal models are less than perfect representations of human subjects. Although animal models such as the non-obese diabetic (NOD) mouse and the BB rat develop Type 1 diabetes with features resembling the disease in humans, there are important differences in the immune systems of these animals and the human subjects. Mesta and Hughes (2004) summarize these differences, which include the balance of leucocyte subsets, toll receptors, defensins, Ig subsets, T cell signalling pathway components, deviations in cytokines and cytokine receptors, Ag-presenting function of endothelial cells, and many others. Despite these differences, animal models can still teach us a great deal. There are specific similarities and differences in autoimmune diabetes of mice and men (reviewed by Roep, 2003; Roep and Atkinson, 2004). Genetic predisposition is one of the most striking similarities, with the MHC class II region as a major determinant of Type 1 diabetes susceptibility in both humans and the rodent models. The BB rat displays a degree of islet inflammation that closely resembles that seen in humans at the onset of the disease. In contrast, the NOD mouse displays a two-phase infiltrate characterized by a non-destructive infiltrate around the border of the islets, followed by infiltration into the islet core if the diabetes progresses (Roep and Atkinson 2004). Knowledge gained so far with regards to diabetes transmission and prevention in animal models may have long-term implications for prevention in humans. Animal models have also enabled the investigation of immune responses in the islets and pancreatic draining lymph nodes, areas that are not readily accessible for study in humans (Leiter and von Herrath, 2004). Although much can be

learn from animal models, there is clearly a strong need for caution. The strength of evidence is improved if more than one animal model is used (Roep and Atkinson 2004, Leiter and von Herrath, 2004); thus it would be helpful to characterize the intestinal microflora in other major models of type 1 diabetes, such as the NOD mouse and the Komeda diabetes-prone rat (Komeda *et al,* 1998). A recent study using NOD mice found that specific pathogen free NOD mice that lack the MyD88 protein (an adapter protein that is used by Toll-like receptors that recognize microbial stimuli), but are colonized with defined commensal bacteria, do not develop Type 1 diabetes. Germ-free NOD mice with the same MyD88 deficiency developed robust diabetes (Wen *et al,* 2008). These findings support the hypothesis that interactions of the intestinal microbiota are an important factor in the predisposition or development of Type 1 diabetes. Further reasearch could focus on identifying antibodies against gut bacteria and critical gut events associated with diabetes, which should coincide with differences in bacterial gut populations between diabetic prone and control animals.

This study focused largely on descriptive and frequency analysis of phylogroups (such as that performed by the DOTUR program) to study the microbiological clone libraries. Other statistical approaches including alignement-independent methods, may give different density distribution results. Rudi *et al* (2007) compared the human adult clone library using DOTUR and found 395 bacterial phylotypes from aligned sequences, while their alignement-independent analysis of the same data found 579 intervals with one or more taxa. In addition, exploring the diversity in a microbiological community not only considers all species, but their prevalence or scarcity, thus techniques to provide quantitative information must be applied to gain a complete picture (Stahl, 1997).

Examples of quantitative methods include quantitative PCR, fluorescent in-situ hybridization (FISH) and flow cytometry used in combination with FISH. Applying quantitative methods would improve our understanding of which bacterial species may contributing to the susceptibility or osent of Type 1 diabetes.

5.2 Conclusion

Overall, the analysis of the clone libraries showed significant differences in the bacterial populations of the small intestine, the Faeces and the lymph nodes of the control vs. diabetic prone BB rat. Age-related differences were also detected. The null hypothesis, that there is no consistent difference in gut bacterial populations in Bio-Breeding control (BBc) and Bio-Breeding diabetic prone (BBdp) rats, is rejected.

References

Acinas S.G., Marcelino L.A. , Klepac-Ceraj V., Polz M.F. **2004**. Divergence and Redundancy of 16S r RNA Sequence s in Genomes with Multiple *rrn* Operons. *J Bacteriology* 186(9): 2629-2635.

Akerblom H.K. Vaarala O., Hyoty H., Ilonen J., Knip, M. **2002.** Environmental Factors in the Etiology of Type 1 Diabetes. *Am. J. Med. Genet. (Semin. Med. Genet.)* 115:18-29.

Amann RI, Krumholz L, Stahl DA. **1990**. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J. Bacteriol. 172(2):762-70.

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

Bäckhed F., Ding H., Wang T., Hooper L.V., Koh G.Y., Nagy A., Semenkovich C.F., Gordon J.I. **2004.** The gut microbiota as an environmental factor that regulates fat storage. *PNAS* 101(44): 15718-15723.

Bäckhed F., Ley, R.E., Sonnenburg J.L., Peterson D.A., Gordon J.I. **2005**. Host-Bacterial Mutualism in the Human Intestine. *Science* 307: 1915-1920.

Beals R., Krantz D.H., Tversky A. **1968**. Foundations of multidimensional scaling. *Psychol Rev.* 75(2):127-42.

Beales P.E., Elliott R.B., Flohé S., Hill J.P., Kolb H., Pozzili P., Wang G.-S., Wasmuth H., Scott F.W. **2002**. A multi-centre, blinded international trial of the effect of A^1 and A^2 β-casein variants on diabetes incidence in two rodent models of spontaneous type 1 diabetes. *Diabetologia.* 45:1240-1246.

Bent S.J., Pierson J.D., Forney L.J. **2007.** Measuring Species Richness Based on Microbial Community Fingerprints: the Emperor Has No Clothes. *Appl. Environ. Microbiol.* 73: 2399-2401.

Bernbom N., Norrung B., Saadbye P., Molbak L., Vogensen F. K., Licht T.R. **2005**. Comparison of methods and animal models commonly used for investigation of fecal microbiota: Effects of time, host and gender. *J. Microbiol. Methods* 66:87-95.

Bessman A.N. and Sapico F.L. **1992**. Infections in the Diabetic Patient: the Role of Immune Dysfunction and Pathogen Virulence Factors. *J. Diabetes Complicat.* 6:258-262

Bonnet R., Suau A., Dore J., Gibson G.R., Collins M. **2002**. Differences in rDNA libraries of faecal bacteria derived from 10-and 25 cycle PCRs. *Int J Syst Evol Microbiol* 52: 757-763.

Breitbart M., Hewson I., Felts B., Mahaffy J., Nulton J., Salamon P., Rohwer F. 2003. Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J. Bacteriol.* 185(20): 6220-6223.

Brooks S.P.J, McAllister M., Sandoz M., Kalmokoff M.L. **2003.** Culture-independent phylogenic analysis of the faecal flora of the rat. *Can. J. Microbiol.* 49:589-601.

Brooks S.P.J., Kalmakoff M.L., Strudo-Kypke M., Inglis G.G., Selinge L.B. **2007.** Methods for monitoring the influence of diet on gut microbiota. *In*: Recent Research Developments in Analytical Biochemistry: Applications in Environmental Toxicology. *Ed.* P. Kumarathasan and R. Vincent. Research Signpost, India ISBN 978-81-308-1093-3

Brugman S., Klatter F.A., Visser J.T.J., Wildebooer-Veloo A.C.M, Harmsen, H.J.M., Rozing J., Bos N.A. **2006.** Antibiotic treatment partially protects against type 1 diabetes in the Bio-Breeding diabetes-prone rat. Is the gut flora involved in the development of type 1 diabetes? *Diabetologica* 49: 2105-2108

Butler J.E, Sun J., Weber P., Navarro P., Francis D. **2000.** Antibody repertoire development in fetal and newborn piglets III. Colonization of the gastrointestinal tract selectively diversifies the preimmune repertoire in mucosal lymphoid tissues. *Immunology* 100:119-130.

Case R.J., Boucher Y. Dahllof I., Holmstrom C., Doolittle W.F., Kjelleberg S. **2007**. Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Appl. Environ. Microbiol.* 73: 278-288.

CCAC (Canadian Council on Animal Care). **1993**. Guide to the Care and Use of Experimental Animals V.1 (2nd Edition). Ed.: E.O. Olfert, B.M. Cross, A.A. McWilliam. Canadian Council on Animal Care. ISBN 0-919087-18-3.

Chao A. and Lee S.-M. **1992.** Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* 87: 210-217.

Chao A. Ma M.C., Yang M.C. **1993**. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* 80:193-201/

Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM. **2003**. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res.* 31(1):442-3

Cole, J. R., B. Chai, R. J. Farris, Q. Wang, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, A. M. Bandela, E. Cardenas, G. M. Garrity, and J. M. Tiedje. **2007**. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res.* 35 (RDP 9, Update 59, Database issue 03/05/2008): D169-D172; doi: 10.1093/nar/gkl889 . Oxford University Press.

Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje. **2009**. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37 (RDP 10, Update 2, Database issue 08/01/2008): D141-D145; doi: 10.1093/nar/gkn879. Oxford University Press.

Cummings, J.H. and MacFarlane G.T. **2002**. Gastrointestinal effects of prebiotics. *Brit. J. Nutrit.* 87: Suppl. 2:S145-S151.

d'Annunzio G., Giannattasio A., Poggi E., Castellano E., Calvi A. Pistoria A. Barabino A., Lorini R. **2009**. β-Cell Autoimmunity in Pediatric Celiac Disease: The Case for Routine Screening? *Diabetes Care* 22: 254-256.

Dabard J., Bridonneau C., Phillipe C., Anglande P., Molle D., Nardi M., Ladiré M., Girardin H., Marcille F., Gomez A., Fons M. **2001**. Ruminocccin A, a new lantibiotic prodiced by a *Ruminococcus gnavus* strain isolated from human Faeces. *Appl. Environ. Microbiol.* 67:4111-4118.

Dethlefsen L, Huse S, Sogin ML, Relman DA. **2008**. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biol* 6(11): e280. doi:10.1371/journal.pbio.0060280

Diez B., Pedros-Alio C., Marsh T.L., Massana R..**2001**. Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques, *Appl.Environ.Microbiol.* 67: 2942-2951.

Domann E., Homg G., Imirzalioglu C., Turschner S., Kuhle J., Watzel C., Hain T., Hossain H., Chakraborty T. **2003**. Culture-Independent Identification of Pathogenic Bacteria and Polymicrobial Infections in the Genitourinary Tract of Renal Transplant Recipients. *J. Clin. Microbiol.* 41(12): 5500-5510.

Dressman J.B. and Yomada K. **1991**. Interspecies Scaling in Animal Models for Drug Absorption. In: *Pharmaceutical Bioequivalence* V.48, ed. P.G. Welling, F.S.S. Tse and S.V. Dighe. Informa Health Care.

Edwards C.A. and Parrett A.M. **2003**. Dietary fibre in infancy and childhood. *Proc. Nutr. Soc.* 62:17-23.

Enck P, Zimmermann K, Rusch K, Schwiertz A, Klosterhalfen S, Frick JS. **2009**. The effects of maturation on the colonic microflora in infancy and childhood. *Gastroenterol Res Pract.* 2009:752401. Epub 2009 Sep 16.

Ercolini D. **2004**. PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *J. Microbiol. Methods.* 56: 297-314

Farrelly V., Rainey F.A, Stackebrandt E. **1995**. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *App. Env. Microbiol.* 61(7): 2798-2801

Ferguson A. **1996**. Mucosal immunology: from bench to the bedside and beyond. *Immunol.* 89:475-482.

Frisk G., Friman G., Tuvemo T., Fohlman J., Diderholm H. **1992**. Coxsackie B virus IgM in children at the onset of Type I (insulin-dependent) diabetes mellitus: evidence for IgM induction by a recent or current infection. *Diabetologica* 35:249-253.

Fromin N, Hamelin J., Tarnawski S., Roesti D. Jourdain-Miserez K., Forestier N.,Teyssier-Cuvelle S., Gillet F., Aragno M., and Rossi P. **2002.** Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns, *Environ.Microbiol.* 4: 634-643.

Galtier, N., Gouy, M., and Gautier, C. **1996**. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *CABIOS* 12:543-548.

Gamble D.R., Taylor K.W. **1969.** Seasonal incidence of diabetes mellitus. *Br. Med. J.* 3:631-633.

Garrity, G.M., Lilburn, T.G., Cole, J.R., Harrison, S.H., Euzeby, J., and Tindall, B.J. **2007**. The Taxonomic Outline of Bacteria and Archaea. TOBA Release 7.7, March 2007. Michigan State University Board of Trustees. [http://www.taxonomicoutline.org/]

Gill S. R., Pop M., DeBoy R.T. Eckburg P.B., Turnbaugh P.J., Samuel B.S., Gordon J.I., Relman D.A., Fraser-Liggett C.M., Nelson K.E. **2006**. Metagenomic Analysis of the Human Distal Gut Microbiome. *Science* 312: 1355-1359.

Gillespie, K.M. **2006**. Type 1 diabetes : pathogenesis and prevention. *Can. Med. Assoc. J.* 175(2): 165-170

Gonzalez J.M., Zimmerman J., Saiz-Jimenez C. **2005**. Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics* 21(3):333-337

Graham S., Courtois P., Malaisse W.J., Rozing J., Scott F.W, Mowat A. **2004.** Enteropathy precedes type 1 diabetes in the BB rat. *Gut* 53:1436-1444.

Guarner F., Casellas F., Borruel N., Antolin M., Videla S., Vilaseca J., Malagelada J-R. **2002**. Role of microecology in chronic inflammatory bowel diseases. *Euro. J. Clin. Nutrit.* 56, Suppl 4: S34-S38

Guarner F. and Malagelada J-R. **2003**. Gut flora in health and disease. *The Lancet* 360:512-519.

Guberski D.L., Thomas V.A., Shek W.R., Like A.A., Handler E.S., Rossini A.A., Wallace J.L., Welsh R.M. **1991**. Induction of type 1 diabetes by Kilham's Rat Virus in diabetes resistant BB/Wor rats. *Science* 254: 1010-1013.

Hampe C.S., Ortqvist E., Rolandsson O., Landin-Olsson M., Törn C., Agren A., Persson B., Schranz D.B., Lernmark A. **1999**. Species-specific autoantibodies in type 1 diabetes. *J Clin Endocrinol. Metab.* 84: 643-648

Hardin J.A., Donegan R.C., Woodman C., Trevenen C., Gall D.G. **2002.** Mucosal inflammation in a genetic model of spontaneous type 1 diabetes mellitus. *Can J Physiol Pharmacol* 80: 1064-1070.

Hayashi, H., Sakamoto, M. & Benno, Y. **2002**. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol. Immunol.* 46, 535–548.

Helfand R.F., Gary H.E. Jr., Freeman C.Y., Anderson L.J., Pallansch M.A. **1995.** Serologic evidence of an association between enteroviruses and an onset of type I diabetes mellitus. J. *Infect. Dis.* 172:1206-1211.

Henrickson J.R. **2004**. webLIBSHUFF (http://libshuff.mib.uga.edu)

Hiltunen M., Hyoty H., Knip M, Ilonen J. Reijonen H, Vahasalo P., Roivainen M., Lonnrot M., Leinikki P., Hovi T., Akerblom H.K., Childhood Diabetes in Finland (DiMe) Study Group. **1997.** Islet cell autoantobody seroconversion in children in temporally associated with enterovirus infections. *J. Infect. Dis.* 175:554-560.

Hogardt M., Trebesius K., Geiger A.M., Hornef M., Rosenecker J., and Heesemann J. **2000**. Specific and Rapid Detection by Fluorescent In Situ Hybridization of Bacteria in Clinical Samples Obtained from Cystic Fibrosis Patients. J Clin Microbiol. 38(2): 818–825.

Hopkins M.J., Sharp R., MacFarlane G.T. **2001**. Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut* 48:198-205

Hurt R.A., Qui X., Wu L., Roh Y., Palumbo A.V., Tiedje J.M., Zhou J. **2001**. Simultaneous recovery of RNA and DNA from soils and sediments. *Appl. Environ. Microbiol.* 67: 4495-4503.

Hyoty H. Hiltunen M., Knip M, Laakkonen M., Vahasalo P. Karjalainen J., Koskela P., Roivainen M., Leinikki P., Hovi T., Akerblom H.K., Childhood Diabetes in Finland (DiMe) Study Group. **1995**. A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. *Diabetes* 44:652-657.

Imai A., and Kurihara Y. **1984**. Endogenous infection in mice with streptozotocin-induced diabetes: a feature of bacterial translocation. *Can. J. Microbiol.* (30) 110: 1344-1348.

Jukes, T.H. and Cantor S.R. **1969.** Evolution of protein molecules. *In:Mammalian protein metabolism.* Ed. by N. Munro. Academic Press Inc. New York. pp 81-132.

Karlsson M.G.E., Garcia, J., Ludvigsson L. **2001**. Cow's milk proteins cause similar Th1- and Th2-like immune response in diabetic and healthy children. *Diabetologia* 44(9): 1140-1147

Kimura K., McCartney A.L., McConnell M.A., Tannock, G. **1997**. Analysis of Fecal populations of *bifidobacteria* and *lactobacilli* and the investigation of the immunological responses of their human hosts to the predominant strains. *Appl. Environ. Microbiol.* 63(9): 3394-3398.

Klaasen H., Van der Heiden P.J., Stok W., Poelma F.G., Koopman J.P., Van den Brink M.E., Bakker M.H., Eling W.M., Beynen A.C. **1993.** Apathogenic, intestinal, segmented, filamentous bacteria stimulate the mucosal immune system of mice. *Infect. Immun.* 61(1) 303-306.

Klappenbach J.A., Saxman P.R., Cole J.R., Schmidt T.M. **2001.** rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Res.* 29:181-184.

Koczwara K., Bonifacio E. Ziegler A.G. **2004**. Transmission of maternal islet antibodies and risk of autoimmune diabetes in offspring of mothers with Type 1 diabetes. *Diabetes* 53: 1-4.

Komeda K. Noda M. Kanazawa M. Teroa K. Kanazawa Y. Kuzuya N. **1998**. Establishment of 2 substrains, diabetes-prone and nondiabetic, from Long-Evans Tokushima-Lean (LETL) rats. *J. Endocrinol* 45: 737-744.

Langendijk P.S., Schut F., Jansen G.J., Raangs G.C., Kamphuis G.R., Wilkinson M.H., Welling G.W. **1995.** Quantitative fluorescence in-situ hybridization of *Bifidobacterium* spp. with genus specific 16S rRNA-targeted probes and its application in fecal samples. *Appl. Environ. Microbiol.* 61: 3069-3075

Lay, C., Doré, J., Rigottier-Gois, L. **2007.** Separation of bacteria of the Clostridium leptum subgroup from the human colonic microbiota by fluorescence-activated cell sorting or group-specific PCR using 16S rRNA gene oligonucleotides. FEMS Microbiol Ecol . 60(3): 513-520

Lenaerts, J. , Lappin-Scott H.M., Porter J. **2007.** Improved Fluorescent In Situ Hybridization Method for Detection of Bacteria from Activated Sludge and River Water by Using DNA Molecular Beacons and Flow Cytometry . Appl Environ Microbiol. 73(6): 2020–2023.

Leiter E.H. and von Herrath M. **2004**. Animal models have little to teach us about type 1 diabetes: 2. In opposition to this proposal. *Diabetologia* (2004) 46: 1657-1660.

Ley R.E., Backhed F., Turnbaugh P., Lozupone C.A., Knight R.D., Gordon J.I. Obesity alters gut microbial ecology. PNAS 102(31)11 070 -11075

Lievin-le Moal and Servin. **2006.** The front line of enteric host defense against unwelcome intrusion of harmful microorganisms: Mucins, antimicrobial peptides and microbiota. *Clin. Microbial. Rev.* 19:315-337.

Lozupone C. and Knight R. **2005.** UniFrac: a new phylogenic method for comparing microbial communities. *Appl. Environ. Microbiol.*71: 8228-8235.

Ludwig W., Strunk O., Klugbauer S., Klugbauer N., Weizenegger M., Neumaier J., Bachleitner M., Schleifer K.-H. **1998**. Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* 19: 554-568.

Ludwig W., Strunk O., Westram R. Richter L., Meier H., Yadhukumar , Buchner A., Lai T. , Steppi S., Jobb G., Forster W., Brettske I., Gerber S., Ginhart A., Gross O., Grumman, S., Hermann S., Jost R., Konig A., Liss T., Lubmann R., May M., Nonhoff B., Reichel B., Strehlow R., Stamatakis A., Stuckmann N., Vilbig A., Lenke M., Ludwig T., Bode A., Schleifer K.-H. **2004**. ARB: A software environment for sequence data. *Nucleic Acids Research* 32(4) 1363-1371.

MacFarlane A.J., Brughardt K.M., Kelly J., Simell T., Simell O., Altosaar I., Scott F. W. **2003.** A type 1 diabetes-related protein from wheat *(Triticum aestivum)* cDNA clone of wheat storage globulin, G1b1, linked to islet damage. *J. Biol.Chem.* 278:54-63.

Macpherson A.J., Harris N.L. **2004.** Interactions between commensal intestinal bacteria and the immune system. *Nat Rev Immunol* 4: 478-485

Macpherson A.J., Uhr T. **2004a.** Compartmentalization of the Mucosal Immune Responses to Commensal Intestinal Bacteria. *Ann. N.Y. Acad. Sci.* 1029: 36-43

Macpherson A.J., Uhr T. **2004b.** Induction of Protective IgA by Intestinal Dendritic Cells Carrying Commensal Bacteria. *Science* 303: 1662-1665.

Maidak B.L., Cole J.R, Lilburn T.G., Parker Jr C.T., Saxman P.R., Farris R.J., Garrity G.M., Olsen G.J., Schmidt T.M., Tiedje J.M. **2001.** The RDP-II (Ribosomal Database Project). *Nucleic Acid Res.* 29(1): 173-174.

Malaisse W.J., Courtois P., Scott F.W. **2004**. Insulin-dependent diabetes and gut dysfunction : the BB Rat Model. *Horm. Metab. Res.* 36:585-594.

Marteau P., Pochart P., Doré J., Béra-Maillet C., Bernalier A, Corthier G. **2001**. Comparative Study of Bacterial Groups within the Human Cecal and Fecal Microbiota . *Appl. Environ. Microbiol.* 67: 4939-4942

Martin A.P. **2002**. Phylogenic approaches for describing and comparing the diversity of microbial communities. *Appl. Environ. Microbiol.* 68: 3673-3682.

Mazmanian S.K, Liu C.H., Tzianabos A.O. and Kasper D.L. **2005**. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122: 107-118.

Mazmanian S.K. and Kasper D.L. **2006.** The love-hate relationship between bacterial polysaccharides and the host immune system. *Nature Rev.Immunol.* 6:849-858.

McCartney A.L., Wang W. Tannock G.W. **1996**. Molecular analysis of the composition of the bifidobacterial and lactobacillus microflora of humans. *Appl. Environ. Microbiol.* 62: 4608-4613.

McConnell EL, Basit AW, Murdan S. **2008.** Measurements of rat and mouse gastrointestinal pH, fluid and lymphoid tissue, and implications for in-vivo experiments. *Pharm Pharmacol.* 60(1):63-70.

McOrist A.L., Jackson M., Bird A.R. **2002.** A comparison of five methods for extraction of bacterial DNA from human faecal samples. *J.Microbiol. Methods* 50:131-139.

Menser M.A., Forrest J.M., Bransby R.D. **1978**. Rubella infection and diabetes mellitus. *Lancet* 1(8055):57-60.

Mestas J., Hughes C.C.W. **2004**. Of mice and not men: differences between mouse and human immunology. *J Immunol* 172:2731-2738.

Morelli L. Postnatal development of intestinal microflora as influenced by infant nutrition. *J.Nutr.* 2008 Sep;138(9):1791S-1795S.

Muntoni S., Cocco P., Aru G., Cucca F., Muntoni S. **2000**. Nutritional factors and incidence of worldwide childhood type 1 diabetes. *Am. J Clin. Nutrit.* 71:1525-1529.

Mueller S., Saunier,K., Hanisch C., Norin E., Alm L., Midtvedt T., Cresci A., Silvi S., Orpianesi C., Verdenelli M.C., Clavel T. **2006**. Differences in Fecal Microbiota in Different European Study Populations in Relation to Age, Gender, and Country: a Cross-Sectional Study. *Appl. Environ. Microbiol.* 72: 1027–1033

Muyzer G., de Waal E.C., Uitterlinden A.G. **1993**. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction–amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59:695-700.

Nagler-Anderson C. **2001.** Man the barrier! Stategic defenses in the intestinal mucosa. *Nat. Rev. Immunol.* 1: 59-67

Neu J., Reverte C.M., Mackey A.D., Liboni K., Tuhacek-Tenace L.M., Hatch M., Li N., Caicedo R.A., Schatz D.A., Atkinson M. **2005**. Changes in intestinal morphology and permeability in the BioBreeding rat before the onset of type 1 diabetes. *J of Ped. Gastroenterol. Nutrit.* 40(5): 589-595

Nicolaison M.H. and Ramsing N.B. **2002**. Denaturing gradient gel electrophoresis (DGGE) approaches to study the diversity of ammonia-oxidizing bacteria. *J. Microbiol. Methods* 50:189-203.

Not T., Tommasini A., Tonini, G., Buratti E., Pocecco M., Tortul C., Valussi M., Crichiutti G., Berti I., Trevisiol C., Azzoni E., Neri E., Torre g., Martelossi S., Soban M., Lenhardt A.,Cattin L., Ventura A. **2001**. Undiagnosed coeliac disease and risk of autoimmune disorders in subjects with Type I diabetes mellitus. *Diabetologia* 44(9):151-155

Ohsugi T., Maejima K., Urano T. **1991**. Bacterial translocation from the gastrointestinal tract in various mouse models for human diabetes. *Microbiol. Immunol.* 35(9): 789-794.

Onkamo P., Väänänen S., Tuomilehto J. **1999**. Worldwide increase in incidence of Type I diabetes – the analysis of the data on published incidence trends. *Diabetologia* 42:1395-1403.

Orlando G., Pisani F. Mastrantonio P., Bonanni L. , Di Cocco P. , D'Angelo M., Tabilio A., Famulari. A. **2008.** . *Eubacterium plautii* infection in a kidney transplant recipient: a noteworthy case of pleural effusion and fever. *Clinical Transplantation* 22(4): 520-524

O'Sullivan D. **1999**. Methods for analysis of the intestinal microflora. *In* Probiotics, a critical review. *Edited by* G.W. Tannock. Horizon Scientific Press, Portland, Ore. U.S.A pp. 23-44.

PHAC (Public Health Agency of Canada). **2005**. Building a national diabetes strategy: synthesis of research and collaborations. Consultation findings, Volume 1. Minister of Health . ISBN 0-662-69322-1. Accessed on-line June 7, 2009 at http://www.phac-aspc.gc.ca/publicat/bnds-bsnd/pdf/bnds-bsnd-vol_1-eng.pdf

PHAC (Public Health Agency of Canada). **2008**. Report from the National Diabetes Surveillance System : Diabetes in Canada 2008. Accessed online February 5 2009.at http://www.phac-aspc.gc.ca/publicat/2008/ndssdic-snsddac-08/index-eng. Minister of Health. Cat. HP32-2/2006| 978-0-662-05874

Pozzilli, P., Signore, A., Williams, A. J., Beales, P. E. **1993**. NOD mouse colonies around the world–recent facts and figures. *Immunol. Today* 14, 193–196

Ray C.G. Palme J.P. Crossley J.R. Williams R.H. **2008.** Coxsackie B virus antibody responses in juvenile-onset diabetes mellitus. *Clin. Endocrinol.* 12: 375-378 .

RDP II Release 8.1. Ribosomal Database project http://rdp8.cme.msu.edu/html/

RDP II Release 9.6 . http://rdp.cme.msu.edu/

Rigottier-Gois, L., Rochet, V., Garrec, N., Suau, A. & Doré J. **2003.** Enumeration of Bacteroides species in human faeces by fluorescent in situ hybridization combined with flow cytometry using 16S rRNA probes. Syst. Appl. Microbiol. 26: 110–118.

Roep B.O. **2003.** The role of T-cells in the pathogenesis of type 1 diabetes: from cause to cure. *Diabetologia* 46: 305-321.

Roep B.O. and Atkinson M. **2004.** Animal models have little to teach us about type 1 diabetes: 1. In support of this proposal. *Diabetologia* 47:1650-1656.

Rudy K., Zimonga M., Kvenshagen B., Rugtveit J., Midtvedt T., Eggesbø M. **2007.** Alignement-independent comparisons of human gastrointestinal tract microbial communities in a multidimensional 16S rRNA gene evolutionary space. *Appl. Environ. Microbiol.* 73(8): 2727-2734

Samuelsson U. and Ludvigsson J. **2003.** The concentrations of short-chain fatty acids and other microflora-associated characteristics in faeces from children with newly diagnosed type 1 diabetes and control children and their family members. *Diabetic Med.* 21:64-67.

Schloss P.D. , Larget B.R., Handelsman J. **2004**. Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl. Environ. Microbiol.* 70(9): 5485-5492.

Schloss P.D. and Handelsman J. **2005**. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* 71(3):1501-1506.

Schloss P.D. amd Handelsman J. **2006a**. Introducing Treeclimber, a test to compare microbial community structures. *Appl. Environ. Microbiol.* 72: 2379-2384.

Schloss P.D. and Handelsman J. **2006b**. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structure. *Appl. Environ. Microbiol.* 72(10): 6773-6779

Schroff K.E., Meslin K. Cebra J. **1995.** Commensal enteric bacteria engender a self-limiting humoral mucosal response while permanently colonizing the gut. *Infect. Immun.* 63: 3904-3913.

Scott F.W. **1996**. Food-induced Type I Diabetes in the BB Rat. *Diabetes/Metabol. Rev.* 12(4): 341-359

Scott F.W, Rowsell P., Wang G-S., Burghardt K., Kolb H., Flohé S. **2002.** Oral Exposure to Diabetes-Promoting Food or Immunomodulators in Neonates Alters Gut Cytokines and Diabetes. *Diabetes* 51(1): 73-78.

Seksik P., Rigottier-Gois L., Gramet G., Sutren M., Pochart P., Marteau P., Jian R., Doré J. **2003.** Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 52(2):237-242

Singleton D.R., Furlong M.A., Rathbun S.L., Whitman W.B. **2001**. Quantitative comparaisons of 16S rRNA gene sequence librairies from environmental samples. *Appl. Environ. Microbiol.* 67: 4374-4376

Singleton D.R., Rathbun S.L., Diszynski G.E., Whitman W.B. **2004**. LIBSHUFF comparaisons of 16S rRNA gene clone librairies. In: *Molecular Microbial Ecology manual, 2nd Edition.* A.D.L. Akkermans, F.J. de Bruijn, G.A. Kowalchuk, I.M. Head, J.D van Elsas (eds.) Kluywer Academic Pub.

Slatkin M., Maddison WP. **1990**. Detecting isolation by distance using phylogenies of genes. *Genetics*. 126(1):249-60.

Smalla, K., G. Wieland, A. Buchner, A. Zock, J. Parzy, S. Kaiser, N. Roskot, H. Heuer, and G. Berg. **2001**. Bulk and rhizosphere soil bacterial communities studied by

denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. *Appl. Environ. Microbiol.* 67:4742-4751

Stackebrandt E. and Goebel B.M. **1994**. Taxonomic Note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44 :846-849.

Stahl D.A. **1997**. Molecular approaches for the measurement of density, diversity and phylogeny. In: *Manual of Environmental Microbiology*. G.C.J. Hurst, G.R. Knudson, M.J. McInerey, L.D. Stetzenbach and M.V. Walter, eds. ASM Press, Washington, DC. p. 102-114.

Suau A., Bonnet R., Sutren M., Goden J.-J., Gibson G., Collins M.D., Doré J. **1999.** Direct analysis of genes encoding 16S rRNA from complex communities reveal many novel molecular species within the human gut. *Appl. Environ. Microbiol.* 65(11): 4799-4804

Suzuki M.T., Giovannoni S.J. **1996**. Bias caused by template annealing in the amplification of mixtures of 16S rRNA Genes by PCR. *App. Environ. Sci.* 62(2): 625-630.

Suzuki K., Ha S.-A., Tsuji M., Fagarasan S. **2007**. Intestinal IgA synthesis: A primitive form of adaptive immunity that regulates microbial communities in the gut. *Seminars in Immunol.* 19:127-135.

Tannock, G.W. **1999**. Probiotics: a critical review. Horizon Scientific Press, Wymondham U.K. 164 pp.

Tannock G.W. **2007**. What immunologists should know about bacterial communities of the human bowel. *Sem. Immunol.* 19 (2007) 94–105

Truswell A.S. **2005**. The A2 milk case: a critical review. *Euro J of Clin Nut* 59:623-631.

Vaarala O., Knip M., Paronen J., Hämäläinen A.-M., Muona P., Väätäinen M., Iilonen J., Simell O., Äkerblom H.K. **1999**. Cow's milk formula feeding induces primary immunization to insulin in infants at genetic risk for type 1 diabetes. *Diabetes* 48:1389-1394.

Vaughan E.E., Schut F., Heilig H.G., Zoetendal E.G., de Vos W.M., Akkerman A.D.L. **2000.** A molecular view of the intestinal ecosystem. *Curr Issues Intest Microbiol* 1(1):1-12

Wang G.C.-Y., Wang Y. **1997.** Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *App. Environ. Microbiol.* 63(12) 4645-4650.

Wang Q.G.M., Garrity J., Tiedje M., Cole, J.R. **2007**. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73(16):5261-7.

Wen L., Ley R.E., Volchkov R.Y., Stranges P.B., Avanesyan L., Stonebraker A.C., Hu C., Won F.S., Szot G.L., Bluestone J.A, Gordon J.I, Chevonsky A.V. **2008**. Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature* 455: 1109 - 1114

Whitford M.F., Forster R.J., Beard C.E, Gong J. , Teather R.M. **1998**. Phylogenic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. *Anaerobe* 4:153-163.

Widmer F., Fliebach A., Laczko E., Schulze-Aurich J., Zeyer J. **2001**. Assessing soil biological characteristics: A comparison of bulk soil community DNA-, PLFA-, and biolog-analyses. *Soil Biol.Biochem.* 33:1029-1036.

Wilson K.H. and Blitchington R.B. **1996**. Human colonic biota studied by ribosomal DNA sequence analysis. *Appl Environ Microbiol* 62:2273-2278

Yoon J.-W. and Jun H.S. **2004**. Viruses in type 1 diabetes: brief review. *ILAR J* 45(3) 343-348.

Yue J.C and Clayton M.K. **2005**. A similarity measure based on species proportions. *Commun. Stat. Theor. M.* 2123-2131.

Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, Elshahed MS. **2009**. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Appl Environ Microbiol.* 75(16):5227-36.

Zoetendal E.G., Ben-Amor K., Akkermans A.D.L., Abee T., de Vos W.M. **2001.** DNA isolation protocols affect the detection limit of PCR approaches of bacteria in samples from the human gastrointestinal tract. *System. Appl. Microbiol.* 24: 405-410.

Zoetendal E.G., von Wright A.,Vilpponen-Salmela T., Ben-Amor K. , Akkermans A.D.L., de Vos W.M. **2002**. Mucosa-Associated Bacteria in the Human Gastrointestinal Tract Are Uniformly Distributed along the Colon and Differ from the Community Recovered from Feces. *Appl. Environ. Microbiol.* 68: 3401–3407

Zoetendal E.G., Collier C.T.,Koike S., Mackie R.I, . Gaskins H.R. **2004** . Molecular Ecological Analysis of the Gastrointestinal Microbiota: A Review J. Nutri 134(2):465-72.

Appendix A

Analysis of small intestine clones from full-length 16S Primers F27/R1492

Using the full-length 16S primers F27/R1492, ninety-six (96) clones were sequenced from the small intestine samples of 55 day-old BB rats (48 control and 51 diabetic prone clones). Using a similarity cut-off of 97%, the DOTUR program (Schloss and Handelsman, 2005) was used to define the 96 clones into 39 distinct operational taxonomic units (OTUs). DOTUR applies the method of Chao and Lee (1992) to estimate species diversity. This method predicted a total of 115 species in the small intestine (lower bound 95% confidence interval = 66, upper bound of 252). The ACE richness estimate (Chao and Lee 1992; Chao et al 1993) predicted 113 species (95% CI , 68 - 227). Species coverage is thus approximately 34%.

Two clone libraries consisting of control 55-day old animals (SIc55, $n = 45$), and diabetic prone 55-day old animals (SId55, $n = 52$) were built from the full-length 16S small intestine DNA sequences in this study. The distribution of clones among the major groups and subgroups of bacteria are shown in Table 6. The largest group is the *Bacillus, Lactobacillus, Streptoccus* subdivision (2.30.7.17) with 51% of clones (56% of OTUs) falling within this group. The *Clostridium lituberense* group (2.30.4.5) also had a large proportion of clones (33%) but these were grouped into a much smaller proportion of OTUs (12%). A complete listing of all OTUs identified fron the small intestine, and their nearest neighbours can be found in Appendix A, Table A2. Using ∫-LIBSHUFF (Singleton et al 2001), the small intestine control and diabetic prone libraries were compared to one another. No differences (p = 1.000) were found between the 16S small

intestine clone libraries. The critical p-value at 95% confidence interval for this library comparison is 0.025, with a margin of error of 0.0000 (Henrikson, 2004), and *P*-values for all libraries are reported for a genetic distance of 0.03.

Table A 1: Distribution of 16S small intestine clones and OTUs among the major groups and subgroups of bacteria

Phylogenetic group	**RDPII No.**	**Clones (*n*)**	**OTUs (*n*)**
Bacteroides and *Cytophaga*	2.15.1	2	1
Proteobacteria	2.28	6	4
Bacillus, Lactobacillus, Streptoccus subdivision	2.30.7.17	49	22
Gram-positive bacteria	2.30	1	1
Clostridium lituseburense group	2.30.4.5	31	5
Clostridium leptum subgroup	2.30.9.13	7	5

Table A 2: Nearest neighbour of full-length 16S clones from the small intestine of the BB Rat

OTUs from small intestine	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)
Gram negative clones			
Bacteroides-Cytophaga			
R-121	2	*Bacteroides mycoides* Z84591	12.31
Proteobacteria			
R-9	1	*Helicobacter rodentium* U96299	0.49
R-24	1	*Helicobacter rodentium* U96299	10.09
R-26	3	*Acinetobacter junii* D86305	1.93
R-110	1	*Acinetobacter junii* D86305	0.14
Gram-positive clones			
Bacillus, Lactobacillus* and *Steptococcus			
R-4	1	*Lactobacillus crispatus* Y17362 Y17362	8.20
R-5	1	*Lactobaccillus acidophilus* M99704	0.59
R-6	15	*Lactobacillus reuteri* X76328	1.12
R-12	1	*Lactobacillus reuteri* X76328	9.52
R-15	5	*Lactobacillus acidophilus* M55805	0.67
R-16	1	*Lactobaccillus acidophilus* M99704	5.62
R-17	2	*Lactobaccillus murinus* AF157049	0.40
R-34	4	*Lactobacillus* sp. AF157050	0.40
R-21	1	*Lactobacillus reuteri* X76328	10.54
R-22	1	*Lactobacillus panis* X94230	8.38
R-23	2	*Lactobacillus reuteri* X76328	4.9
R-25	4	*Lactobacillus panis* X94230	2.67
R-35	1	*Lactobacillus acidophilus* M99704	3.47
R-36	1	*Lactobacillus reuteri* X76328	3.83
R-37	1	*Lactobaccillus murinus* AF157049	2.97
R-40	1	*Lactobacillus reuteri* X76328	10.35
R-43	1	*Lactobacillus sp.* AF157050	9.68
R-50	1	*Lactobacillus panis* X94230	6.88
R-101	1	*Lactobacillus reuteri* X76328	5.33
R-104	2	*Lactobacillus murinus* AF157049	7.11

OTUs from small intestine	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)
R-165	1	*Lactobacillus acidophilus* M55802	6.10
R-31	1	*Lactobacillus reuteri* X76328	
R-159	1	*Streptococcus pleomorphus* M23730	14.49
***Clostridia* superfamily**			
R-18	1	*Clostridium lituseburense* M59107	5.07
R-47	1	Str.sfb3D86305	7.40
R-26	1	Str.sfb3 D86305	1.93
R-30	27	*Clostridium lituseburense* M59107	2.52
R-41	1	*Clostridium lituseburense* M59107	5.29
R-122	1	Str-sfb3 D86305	6.66
R-123	3	*Clostridium quinii* X76745	5.24
R-164	1	*Clostridium lituseburense* M59107	6.74
R-168	1	*Clostridium lituseburense* M59107	5.40
R-115	1	*Eubacterium plautii*	10.96

Appendix B

Analysis of lymph node clones from full-length 16S primers F27/R1492

A total of 97 clones were sequenced from the 55 day-old BB rat lymph node samples using the F27/R1492 primers (53 control and 49 diabetic prone clones). The DOTUR program (Schloss and Handelsman, 2005) defined 19 distinct operational taxonomic units (OTUs- similarity cut-off of 97%). The method of Chao and Lee (1992) was used to estimate species diversity and predicted a total of 43 species in the lymph node (95% CI 27 - 108). The ACE richness estimator (Chao and Lee 1992; Chao et al 1993) predicted 45 species (95% CI , 29 - 104) in the lymph node. Species coverage is thus approximately 43%.

Two clone libraries consisting of control 55-day old animals (LNc55, $n = 46$), and diabetic prone 55-day old animals (LNd55, $n = 51$) were built from the full-length 16S lymph node DNA sequences in this study. The distribution of clones and OTUs among the groups and subgroups of bacteria is shown in Table 8. No clones were identified as *Bacteroides* species. The majority of clones and OTUs belonged to either the *Proteobacteria* division (2.28) or the *Bacillus, Lactobacillus, Streptococcus* (2.30.7.17) subdivision. Appendix A, Table A2 lists all of the OTUs and their nearest neighbours.

Using ∫-LIBSHUFF (Singleton et al 2001), the lymph node control and diabetic prone clone libraries were compared to one another. No differences ($P = 1.000$) were found between the control and diabetic prone 16S lymph node clone libraries. The critical p-value at 95% confidence interval for this library comparison is 0.025, with a margin of error of 0.0000 (Henrikson, 2004), and p-values for all libraries are reported for a genetic distance of 0.03.

Table B 1: Distribution of 16S lymph node clones and OTUs among the major groups and subgroups of bacteria

Phylogenetic group	RDPII No.	Clones (*n*)	OTUs (*n*)
Bacteroides and *Cytophaga*	2.15.1	0	0
Proteobacteria	2.28	43	10
Bacillus, Lactobacillus, Streptoccus subdivision	2.30.7.17	49	6
Gram-positive bacteria	2.30	2	2
Clostridium lituseburense	2.30.4.5	0	0
Clostridium leptum group	2.30.9.1	1	1
Clostridium coccoides	2.30.4.1	2	1

Table B 2: Nearest neighbour of full length 16S clones from lymph node of BB rat

16S OTUs from lymph node	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)
Gram negative clones			
Proteobacteria			
R-51	3	*Acinetobacter junii* AB021321	0.55
R-52	29	*Acinetobacter junii* X81658	0.27
R-71	1	*Acinetobacter junii* X81658	3.10
R-72	1	*Acinetobacter junii* X81658	4.81
R-84	1	*Ralstonia pickettii* X67042	0.65
R-94	1	*Moraxella oreonsis* Z93987	2.04
R-74	2	*Acidovorax* sp. AJ130765	4.07
R-98	3	*Acidovorax. temprans* F078766	2.05
R-64	1	*Pseudomanas fluorecens* AB015251	0.27
R-67	1	*Bradyrhizobium* sp. X70404	0.14
Gram positive clones			
R-60	1	*Desulfotomaculum guttoideum* Y11568	3.59
R-182	2	*Clostridium* sp. AB009177	14.51
Bacillus, Lactobacillus and Streptococcus			
R-126	6	*Lactobacillus panis* X94230	2.53
R-128	11	*Lactobacillus reuteri* X76328	0.46
R-130	27	*Lactobacillus acidophilus* M58802	0.33
R-150	3	*Lactobacillus acidophilus* M99704	0.26
R-100	1	*Streptococcus pleomorphus* M23730	14.97
R-77	1	*Turicibacter* sp. AB021355	3.04
***Clostridia* superfamily**			
R-92	1	*Clostridium intestinalis* X76740 DSM 6191	1.98

Appendix C

Table C 1: Nearest neighbours of DGGE bands

OTUs from DGGE	Nearest neighbour (using ARB)	Divergence (%)	Source of picks *
Gram-negative clones			
Bacteroides-Cytophaga			
R-1622	*Bacteroides distasonig* M86695	< 0.000	19c Lymph node
R-1618	*Bacteroides* sp. AF139525	2.09	19d Lymph node
R-1605	*Tannerella sp.* AF018517	18.63	55c Small intestine
Proteobacteria			
R- 1600	*Methylobacterium sp.* AB024616	< 0.000	55c Lymph node
R-1601	*Acinetobacter junii* X81658	< 0.000	55c Lymph node
Gram-positive clones			
Bacillus, Lactobacillus and Streptococcus			
R-1603x4	*Lactobacillus johnsonii* AJ002515	< 0.000	55c Lymph Node (2) 19d Lymph Node 55d Small intestine
R-1610x2	*L. acidophilus* M58802	< 0.000	55d Small intestine
R-1619	*L. crispatus*	1.01	55d Small intestine
R-1607	*Lactobacillus* sp. AF157050	< 0.000	55d Small intestine
R-1602	*Enterococcus dispar* V18358	1.01	55c Lymph node
R-1612	*Enterococcus gallinarum* AF039900	0.5	55c Lymph node
***Clostridium* superfamily**			
R-1606	*Clostridium* Sp. Z94008	13. 16	55c Small intestine
R-1620	*Clostridium aurantibutyricum*	< 0.000	55d Small intestine
R-1609	*Clostridium irregularis.*	< 0.000	55d Small intestine
R-1614x3	*Clostridium irregularis.*	6.60	19c Lymph node 19d Lymph node

* 55c = 55 day-old control, 55d = 55 day-old diabetic prone, 19c = 19 day-old control, 19d = 19-day old diabetic-prone

Appendix D

Table D 1: Nearest neighbour of clones from the small intestine of the BB Rat

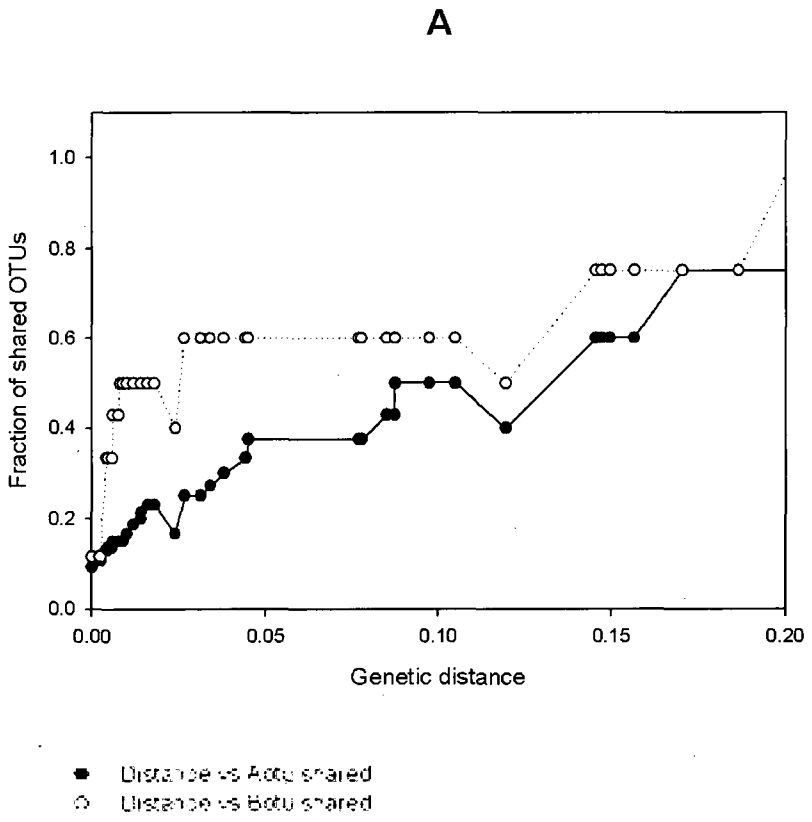
1 55c= 55 day-old control, 55d = 55 day-old diabetic prone, 19c = 19 day old control, 19d = 19 day-old diabetic prone

OTUs from small intestine	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks [1] sample(number of clones picked)
Gram Negative Clones				
Bacteroides-Cytophaga				
R-5790	1	*B. acidofaciens* AB021157	13.6	19c
R-2594	1	*B. acidofaciens* AB021157	12.6	55c
R-5486	1	*Tanerella* sp. AB009211	22	19d
R-3925	1	*Tanerella* sp. U4135	26.5	55c
R-4062	1	AB009191	14.6	55d
Proteobacteria				
R-4040	2	*Proteus vulgaris* X07652	1.4	55d
R-5523	2	*Escherichia coli* U00096	<0.000	19d
Gram-positive clones				
Bacillus, Lactobacillus* and *Streptococcus				
R-1813x47	47	*L. acidophilus* M58805	0.4	55d(24), 55c(10), 19c(13)
R-1815x20	20	*L. acidophilus* subspecies *johnsonii* M99704	< 0.000	55d (1), 55c(11) 19c(7), 19d(1)
R-5791	1	*L. acidophilus* subspecies *johnsonii* M99704	3.5	19c
R-1835x173	174	*L. reuteri* X76328	0.2	55d (21), 55c(49) 19c(40), 19d(32)
R-5813	1	*L. reuteri* X76328	3.1	19c
R-1817x15	15	*L. reuteri*, L23507	0.6	55c(3), 19c(8),

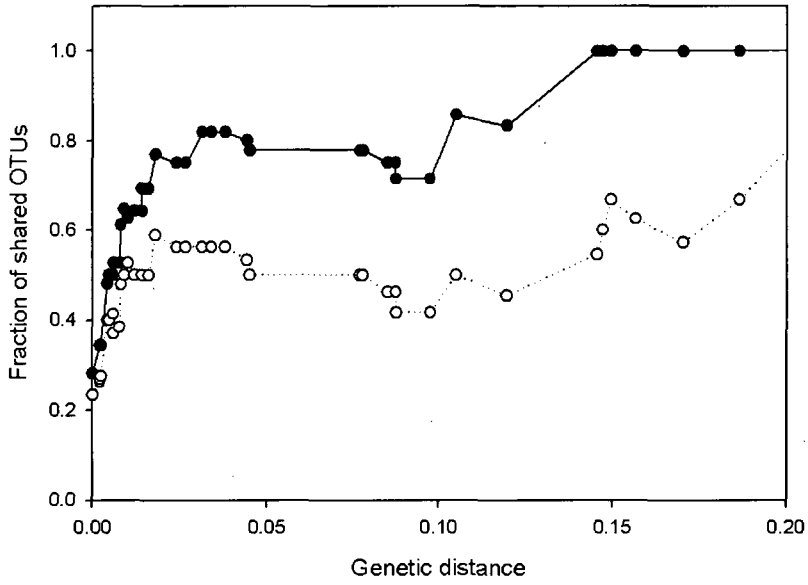
OTUs from small intestine	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks [1] sample(number of clones picked)
				55d(4)
R-1844x7	7	*L. murinus* AF157049	<0.000	55c(2), 55d(3), 19c(2)
R-1822x19	19	*L.* ASF360 sp. AF157050	<0.000	55c(8), 55d(7), 19c(4)
Other Gram-positive clones				
R-5380	1	*Gemella* sp. Y13365	13.2	19c
R-1832x18	18	*Gemella* sp. Y13365	13.3	55c(3), 55d(15)
R-4064	1	*Bidibacteria animalis* D86185	11.90	55d
***Clostridia* superfamily**				
R-2469x6	6	*Halobacillus litoralis* X94558	18.5	55c (4), 55d (1) 19c(1)
R-1852(19)	19	*Clostridium lituseburense* M59107	1.0	55c(5), 55d(12),19c(1), 19d(1)
R-2995(3)	3	*Clostridium quinii* X76745	2.88	55d(3)
R-4047	1	*Ruminococcus albus* AF079847	9	55d
R-2576	1	*Ruminococcus gnavus* L76597	9.6	55c
R-3929	1	*Ruminococcus productus*	16	55d

Figure D 1: Comparison of Small Intestine Libraries – Fraction of Shared OTUs

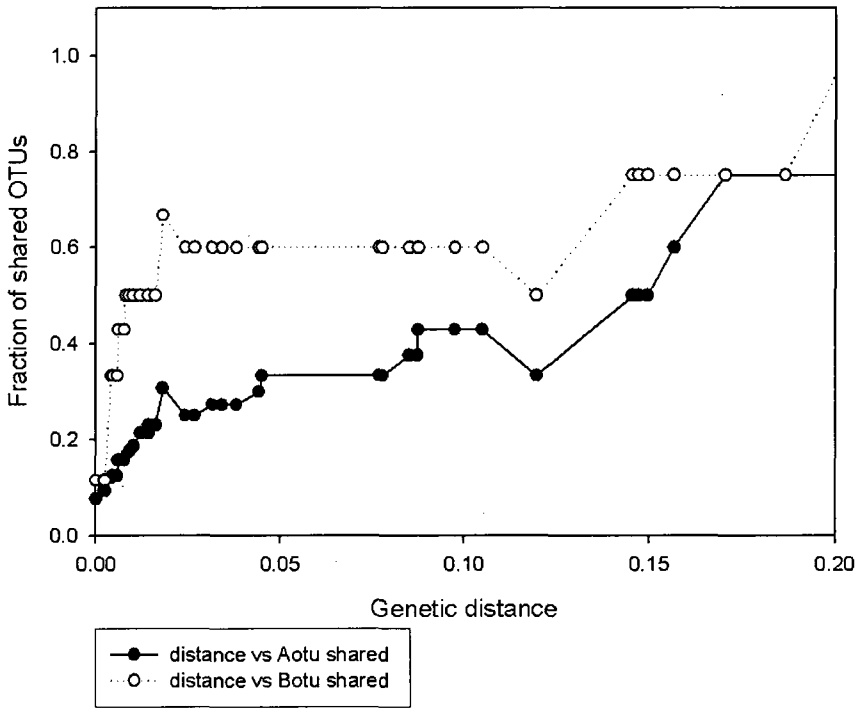
A. The estimated fraction of shared OTUs between SIc19 (Aotu shared) and SId19 (Botu shared) is shown. At a genetic distance of 0.03, Aotu shared is 25%, while Botu shared is 60%. **B.** The estimated fraction of shared OTUs between SIc55 (Aotu shared) and SId55 (Botu shared) is shown. At a genetic distance of 0.03, Aotu shared is 82%, while Botu shared is 56%. **C.** The estimated fraction of shared OTUs between SIc55 (Aotu shared) and SId19 (Botu shared) is shown. At a genetic distance of 0.03, Aotu shared is 27%, while Botu shared is 60%.



B


Fraction of shared OTUs
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.05
0.10
0.15
0.20
Genetic distance

C


Fraction of shared OTUs
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.05
0.10
0.15
0.20
Genetic distance
distance vs Aotu shared
distance vs Botu shared

Appendix E

Table E 1: Nearest neighbour of clones from the lymph node of the BB Rat

OTUs from lymph node	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks sample(# of clones picked)
Gram negative clones				
Bacteroides-Cytophaga				
R-3330x8	8	*Tannerella* sp. AB009191	12.63	55d
R-3340	2	*Tannerella* sp. AB009191	15.78	55d
R-3327x4	4	*Tannerella* sp. AF001767	16.03	55d
R-3420	2	*Tannerella* sp. AB009211	15.20	55d(1), 19d(1)
R-3417	2	*Tannerella* sp. AB009211	17.35	55d
R-3411	1	*Tannerella* sp. AB009211	17.53	55d
R-3394	1	*Tannerella* sp. AB009211	15.81	55d
R-4333	1	*Tannerella* sp. AB009211	19.06	55c
R-5687	3	*Bacteroides* sp. AF001778	14.83	19d
R-3369	1	*Bacteroides sp.* AB021162	10.99	55d
R-3360	1	*Bacteroides acidofaciens* AB021157	15.53	55d
R-5699	1	*Bacteroides* species ASF519	15.78	19d
R-3343	1	*Prevetolla sp.* AF40719	9.11	55d
R-3324	2	*Prevetolla sp.* AF040719	9.2	55d
R-3322	2	*Tannerella* sp. AB009211	18.1	55d
Proteobacteria				
R-3381	1	*Pseudomonas*	17.5	55d

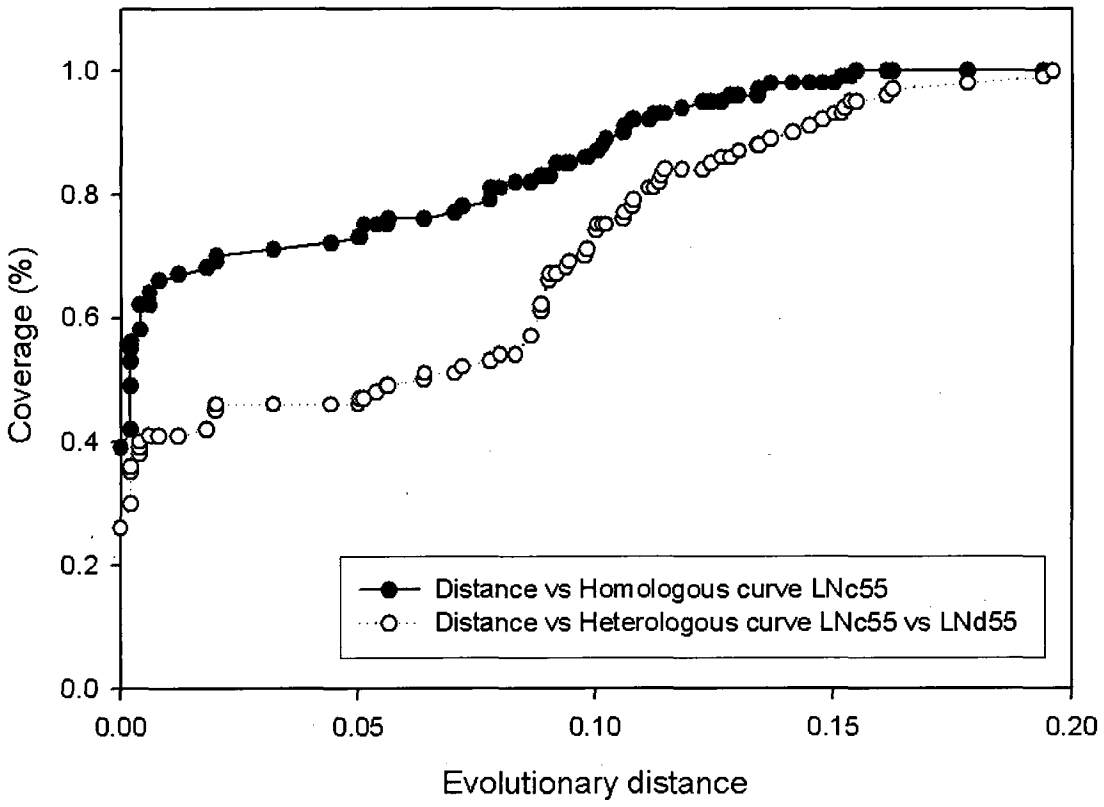
OTUs from lymph node	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks sample(# of clones picked)
		mosselii AF028352		
R-3335x48	48	*Proteus vulgaris* X07652	0.8	55c(21) 55d(19)
R-3375	1	*Proteus sp.* AB021162	12.6	55d
R-3371	2	*Moraxella* sp. AB021321	0.8	55d
R-3334x11	11	*Escherichia coli* U00096	1.82	55d(5), 55c(6)
R-3323x16	16	*Acinetobacter junii* X81658	2.25	55c(5) 55d(2) 19d(9)
R-3768	2	*Ralstonia solanacearum* X67036	0.4	55c(5)
R-3934	2	*Acidovorax temperans*	0.2	55c(5)
R-3332	1	*Dechlorimonas* sp. AF170354	< 0.000	55d
R-3409	1	*Aquaspirillum psychrophilum* AF078755	0.8	55d
R-3325	1	*Methylobacterium* sp. AJ223453	<0.000	55d
Gram-positive clones				
Bacillus, Lactobacillus and Streptococcus				
R-3368x39	39	*L. reuteri* X76328	0.2	19c(27), 19d(3), 55c(6), 55d(3)
R-3331x18	18	*L. acidophilus* M58802	4.32	55c(11), 55d(5), 19c(1), 19d(1)
R-3376	1	*L. acidophilus* M58802	16.6	55c
R-3388x9	9	*L. acidophilus* M99704	<0.000	55c(5), 55d(1) 19c(2), 19d(1)
R-4126	2	*L. murinus* AF157049.	<0.000	55c(1), 19d(1)
***Clostridium* superfamily**				
R-5682	2	*Eubacterium plautii* DSM 4000	9.52	19d

OTUs from lymph node	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks sample(# of clones picked)
R-5680x51	51	*Eubacterium plautii* DSM 4000	7.6	19d
R-3333	1	*Ruminococcus callidus* X85100	3.1	55d
R-3414	3	*Ruminocccus albus* AF079847	2.25	55d
R-3329	1	*Eubacterium plexicaudatum* AF157058	8.2	55d
R-3344x12	12	*Ruminococcus obeum* AJ408993	4.95	55c(9) 55d(3)
R-3357x8	8	*Ruminococcus obeum* AJ408993	4.53	55c(7) 55d(1)
R-3350	1	*Ruminococcus obeum* AJ408993	12.8	55d
R-5726	1	*Ruminococcus obeum* X85101.	3.89	19d
R-4121x4	4	*Clostridium lituseburense* M59107	1.41	55c(3), 55d(1)
R-3384	1	*Clostridium xylonalyticum*,	10.7	55d
R-4118	3	*Clostridium* sp. AJ002591	10.2	55c(3)
R-3955	1	*Clostridium quinii* X76745	3.1	55c
R-3356	2	*Clostridium* sp. U81762	9.4	55d
R-3326	1	*Clostridium* sp. Y10030	7.33	55d
R-3366	1	*Clostridium* sp. AJ002591	12.8	55d
R-3408	1	*Clostridium* sp.D14638	10.0	55d
R-3380	1	*Clostridium* sp. AF157053	12.8	55d
R-3773	1	*Clostridium irregularis*	6.27	55d
R-3410	1	AB009208	12.7	55d
R-3364	1	*Sporobacter* sp. AF129863	3.92	55d
R-3407	1	*Papillibacter* sp.	3.46	55d

OTUs from lymph node	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks sample(# of clones picked)
		AF157051		
Other Gram-positive clones				
R-4094	1	*Staphylococcus pasteuri* AF041362	0.2	55c
R-3763	1	*Bifidobacterium pseudocatenulatum* D86687	0.2	55c
R-3339	1	*Micrococcus luteus* AF057289	1.4	55d
R-3387	1	uncultured bacteria AF028352	12	55d
R-3383	1	*Halobacillus litoralis* X94558	12.6	55d

Figure E 1: Results of ∫-LIBSHUFF comparison.

Homologous and heterologous curves for 16S rRNA gene sequence libraries from the lymph node (LNc55; control 55-day old, $n = 96$ vs LNd55; diabetic prone 55-day old , $n = 100$) libraries. Comparisons were made using ∫-LIBSHUFF with 10,000 randomizations. The margin of error for the P value's 95% confidence interval near 0.05 was 0.004; Distance is equal to the Jukes-Cantor (1969) corrected evolutionary distance.



Appendix F

Table F 1: Nearest neighbour of clones from the faeces of the BB Rat

OTUs from faeces	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks
Gram negative clones				
Bacteroides-Cytophaga				
R-4556(38)	38	*Bacteroides distasonis* ATCC 8503	0.60	19c(15) 19d(23)
R-5902	1	*Bacteroides distasonis* M86695	13.31	55d
R-4677(4)	4	*Bacteroides distasonis* M86695	3.06	19c(2),19d(2)
R-4667(9)	9	*Bacteroides* spp. AF139525	1.83	19d(9)
R-4548(8)	8	*Bacteroides* spp. AF001778	9.17	19c(8)
R-4763	1	*Bacteroides* spp. AF001778	6.7	19c
R-5838	1	*B. splandinicus* L16496	11.17	55d
R-4625(2)	2	*Bacteroides fragilis* AF133539	5.21	19d
R-4763	1	*Bacteroides* sp. AF133539	6.70	19c
R-4559(10)	10	*B. acidofaciens* AB021162	13.82	19c
R-4804(4)	4	*B. distasonis* AF001767	14.58	19c(3), 55d(1)
R-4793(4)	4	*B. distasonis* AF001767	12.18	19c(3), 55d(1)
R-4664(4)	4	*B. distasonis* AF001767	15.07	19d
R-4833(5)	5	*B. distasonis* AF001767	11.95	55d(3), 19c(2)
R-4552(4)	4	*B. distasonis* AF001767	12.42	19c

OTUs from faeces	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks
R-4554(5)	5	*Bacteroides sp.* AF139524	13.11	19c(3), 55d(2)
R-5843(3)	3	*Bacteroides sp.* AF139524	12.39	55d(2), 55c(1)
R-4550(8)	8	*B. acidofaciens* AB021162	12.07	55d
R-5915	2	*B. acidofaciens* AB021157	11.32	55d
R-4734	1	*B. acidofaciens* AB021162	13.11	19d
R-6700	1	AF157051	5.28	55c
R-4659	1	*B. acidofaciens* AB021157	16.22	19d
R-6714	1	*B. acidofaciens* AB021157	12.63	55c
R-6697	1	*B. acidofaciens* AB021157	13.34	55c
R-6684	1	*B. acidofaciens* AB021157	14.31	55c
R-5865	1	*B. acidofaciens* AB021157	13.58	55d
R-4570	1	*B. acidofaciens* AB021157	12.43	19c
R-6698	1	*B. acidofaciens* AB021157	15.04	55c
R-4748x5	5	*Bacteroides* sp. AF139524	<0.000	19d(4), 55c(1)
R-4796	2	*B. acidofaciens* AB021162	<0.000	19c
R-4792	2	*Bacteroides Splandicus*	14.07	55c(1), 19c(1)
R-4560x5	5	*Bacteroides* sp. AF157056	<0.000	19c
R-6671x6	6	*Prevetolla bivia* L16475	7.77	55c
R-4795x8	8	*Prevetolla* spp. AB009241	7.55	55c(5), 19c(3)
R-4827x8	8	*Prevetolla* spp. AF040719	8.89	55c(5), 55d(2) 19c(1)
R-4566x6	6	*Tannerella* spp. AJ400254	11.92	19c
R-4784	2	*Tannerella* spp. AJ400254.	13.11	55c(1), 19c(1)

OTUs from faeces	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks
R-4762x6	6	*Tannerella* spp AB009191	12.18	55d(3), 55c(1) 19c(1), 19d(1)
R-4641	3	*Tannerella* spp AB009191	13.86	55d(2), 19d(1)
R-4631	1	*Tannerella* spp AB009191		19d
R-5859x4	4	*Tannerella* sp. AB009211	13.67	55c(2) 55c(2)
R-4557	3	*Tannerella* sp. AB009211		55d(1) 19c(2)
R-6736	2	*Tannerella* sp. AB009211	13.31	55c
R-6655	2	*Tannerella* sp. AB009211	14.31	55c
R-4571	1	*Tannerella* sp. AB009211	13.89	19c
R-4820	1	*Tannerella* sp. AB009211	13.58	19c
R-5845	1	*Tannerella* sp. AB009211	14.55	55d
R-5903	1	*Tannerella* sp. AB009211	17.77	55d
R-5931	1	*Tannerella* sp. AB009211	12.92	55d
R-5871	2	*Tannerella* sp. AB009211	13.11	55d
R-4816	3	*Tannerella* sp. AF001767	14.25	55c(1), 55d(1), 19c(1)
R-5885	1	*Tannerella* sp. AB009202	11.95	55d
R-5905	1	*Tannerella* sp. U81730	13.89	55d
Proteobacteria				
R-4671x17	17	*Escherichia coli* U0096	<0.000	19d(16), 19c(1)
R-5854	2	*Helibacter rodentium* U96298	0.4	55c(1), 55d(1)
R-4563	1	*Pseudomonas antimicrobica* AB021384	6.53	19c
R-4794	1	*Escherichia coli*	17.18	19c

OTUs from faeces	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks
		U0096		
Gram-positive clones				
Bacillus, Lactobacillus and Streptococcus				
R-4553x25	25	*L. acidophilus* M55802	0.40	55c(5), 55d(11) 19c(4), 19d(5)
R-6670	1	*L. acidophilus* M55802	3.48	55c
R-4666	10	*L. acidophilus* M99704	<0.000	55c(3), 55d(1), 19d(6)
R-4673x6	6	*L. reuteri* X76328	0.20	55c(1), 555d(1), 19c(1), 19d(3)
R-4676	2	*L. murinus* AF157049	<0.000	19d
R-6703	1	*Lactobacillus* sp. AF157050	14.31	55c
R-5852x4	4	*Gemella haemolysens* Y13365	13.15	55c(1), 55d(3)
R-4629		*Streptococcus infantis*	3.04	19d
R-5924		*Gemella haemolysens* Y13365	10.68	55d
***Clostridium* superfamily**				
R-5857	3	*Eubacterium plautii* DSM4000	4.55	55d(3)
R-4754	1	*Oscillospira* sp. U81762	5.2	55c(1), 55d(1), 19(1)
R-6680	1	*Oscillospira* sp. U81762	13.15	55c
R-6704	1	*Oscillospira* sp. U81762	4.98	55c
R-6714	1	*Eubacterium plautii* DSM4000	12.63	55c
R-6666	1	*Ruminococcus flaveciens* AF030450	8.80	55c
R-6721	1	*Ruminococcus flaveciens* AF030449	9.23	55c
R-6696	1	*Ruminococcus flaveciens* AF030449	9.73	55c

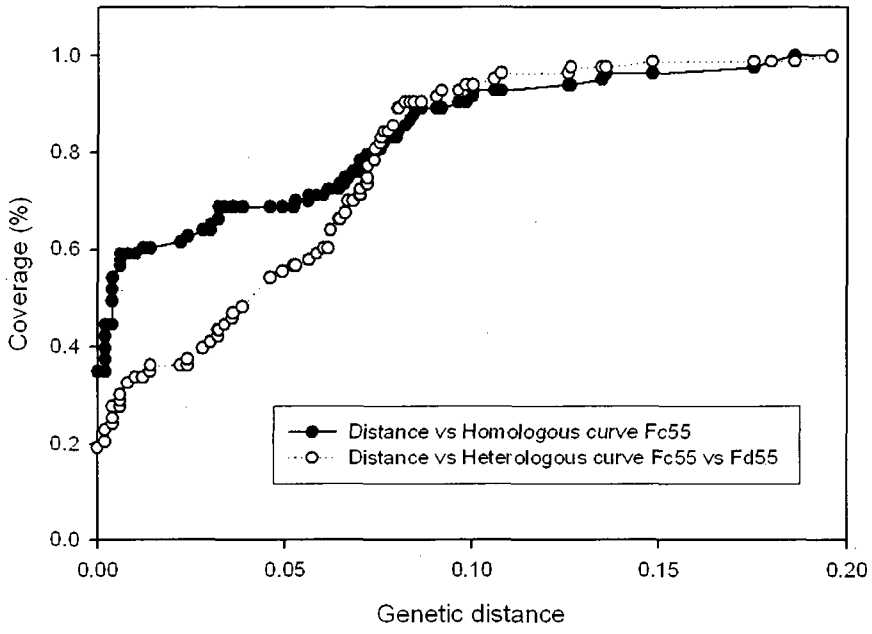
OTUs from faeces	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks
R-5896	2	*Ruminococcus flavofaciens* AF030448	3.30	55d
R-6731	1	*Ruminococcus callidis* X85100	7.44	55c
R-6708	1	*Ruminococcus albus* X85098	3.93	55c
R-4549	1	*Ruminococcus torques* L76604.	8.22	19c
R-6692	1	*Ruminococcus productus* L76595	7.10	55c
R-4669x4	4	*Ruminococcus* sp. 94964		19c
R-4723	2	*Eubacterium fissicatena* DSM3598	1	19d
R-5858	2	*Eubacterium ventriosum* DSM 3988	11.87	55d
R-6685	2	*Eubacterium ventriosum* DSM 3988	1.61	55c
R-5897	1	*Eubacterium plexicaudatum* AF157058	5.81	55d
R-6654x4	4	*Clostridium lituberenses* M59107	1.20	55c(3), 55d(1)
R-5907x4	4	*Clostridium xylanolyticum* X71855	5.40	55d
R-4828	1	*Clostridium xylanolyticum* X71855	7.55	19c
R-5855	2	*Clostridium sp.* AB009173	9.11	55d
R-6735	1	*Clostridium sp.* AB009173	6.89	55c
R-6653	2	*Clostridium* sp. AB009173	8.72	55c
R-6664	3	*Clostridium* sp. AB009173	6.91	55c
R-5846	1	*Clostridium* sp.	10.02	55c

OTUs from faeces	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks
		AB009173		
R-5895	1	*Clostridium* sp. AB009173	6.48	55c
R-6717	1	*Clostridium* sp. AB009173	6.89	55c
R-5864	3	*Phascoloribacterium faecium* X72867	4.72	55c(2), 55d(1)
R-5870	2	*Papillibacter* sp. AF157051	5.01	55d
R-5861	1	*Papillibacter* sp. AF157051	8.10	55d
R-5894	1	*Papillibacter* sp. AF157051	4.37	55d
R-4551	1	*Papillibacter* sp. AF157051	6.10	19c
R-4798	1	*Butyrivibrio crossotu* X89981	8.05	19c
R-5880	1	*Butyrivibrio fibrisolvens* AB004910	6.70	55d
R-4823	1	*Clostridium aminobutyricum* X676161 DSM 2634	7.84	19c
R-4555	1	*Clostridium* sp. AB009208,	13.86	19c
R-5839	1	*Clostridium* sp. AF018558	7.67	55d
R-6667	1	*Clostridium* sp. AF129861	5.20	55c
R-6663	1	*Clostridium viridae* X81125	7.06	55c
R-6668	1	*Clostridium* sp. AF087943		55c
R-6674	1	*Clostridium* sp.	8.46	55c
R-6675	1	*Clostridium* sp. Y10030	13.70	55c
R-5867	1	*Clostridium aminovalericum* M23929	10.48	55d
R-5928	1	*Clostridium* sp. Z95708		55c
R-5881	1	*Clostridium* sp. AJ002591	10.02	55d
R-4785	1	*Clostridium* sp.	6.25	19c

OTUs from faeces	Number of clones picked	Nearest neighbour (using ARB)	Divergence (%)	Source of picks
		AJ002591		
R-6689	1	*Clostridium polysaccharolyticum* X77839	11.17	55c
R-6690	1	*Clostridium polysaccharolyticum* X77839	10.94	55c
R-6705	1	*Clostridium* sp. AF001766	9.15	55c
R-6710	1	*Clostridium* sp Y10584	11.38	55c
R-4654	1	*Clostridium* sp Y10584	4.00	19d
R-4736	1	*Clostridium* sp. U81762	12.78	19d
R-4743	3	*C. fusiformis* AF028349		19d
R-4632	1	*C. fusiformis* AF028349	8.01	19d
R-4802	2	*C. fusiformis* AF028349	7.55	19d
R-6661	1	*Allobaculum stercoricanus* AF028352	14.66	55c
R-6729	1	*Eubacterium plexicaudatum* AF157058	2.43	55c
R-6733	1	AB009189	5.16	55c

Figure F 1: Results of ∫-LIBSHUFF comparison.

Homologous and heterologous curves for 16S rRNA gene sequence libraries from faeces samples are shown. A) shows the Fd55 vs Fc55 libraries, B) shows Fd19 vs Fd55. Comparisons were made using ∫-LIBSHUFF with 10,000 randomizations. The margin of error for the P value's 95% confidence interval near 0.05 was 0.004; Distance is equal to the Jukes-Cantor (1969) corrected evolutionary distance.



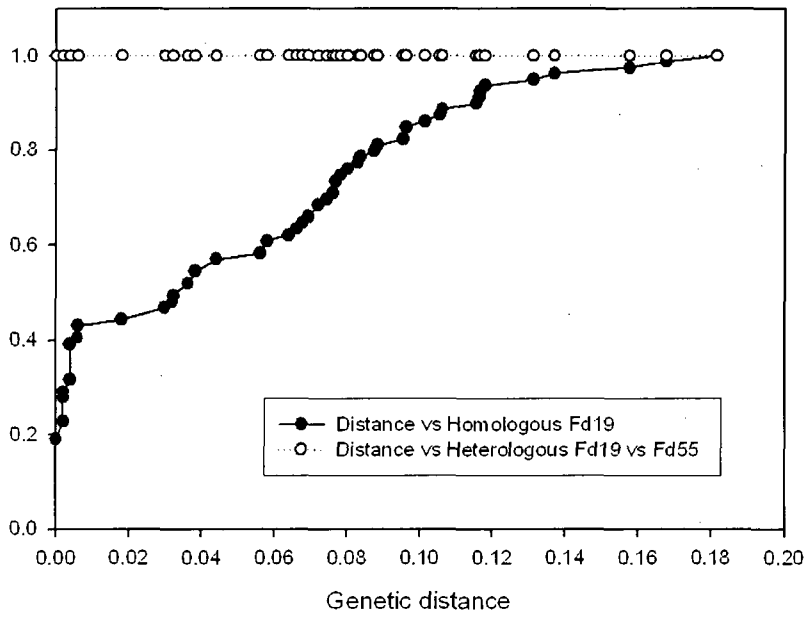
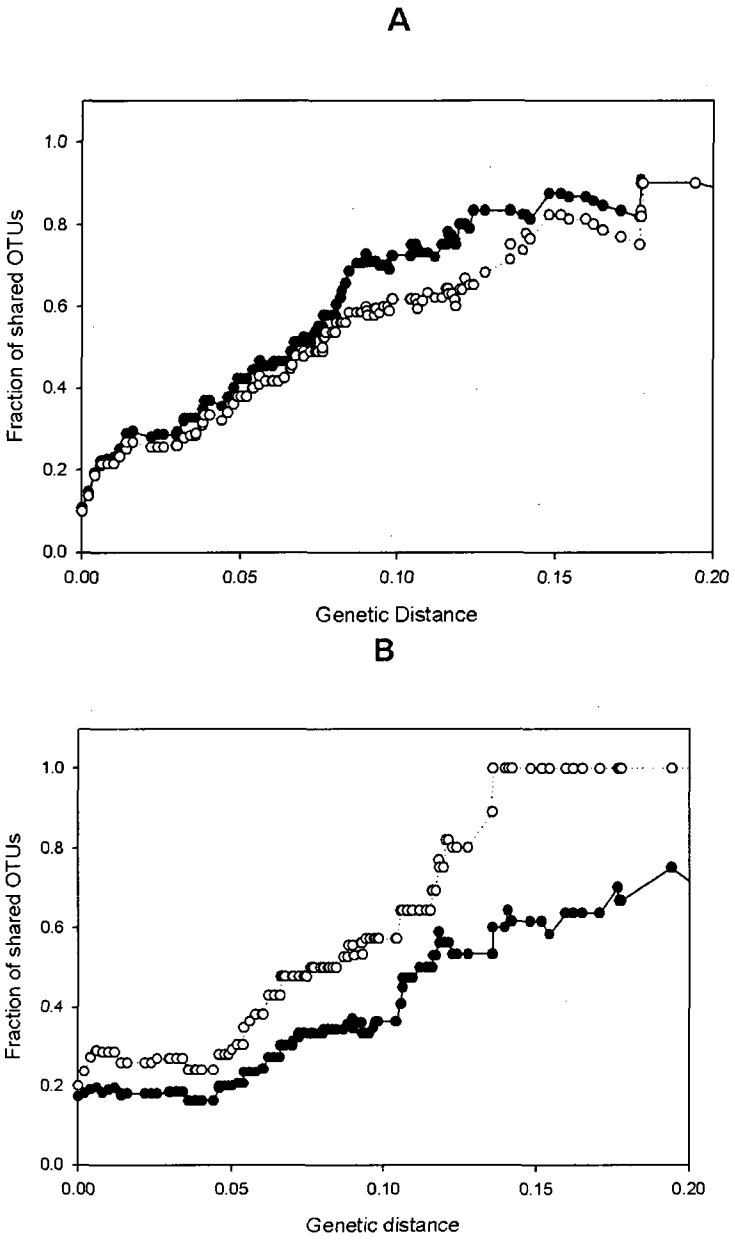
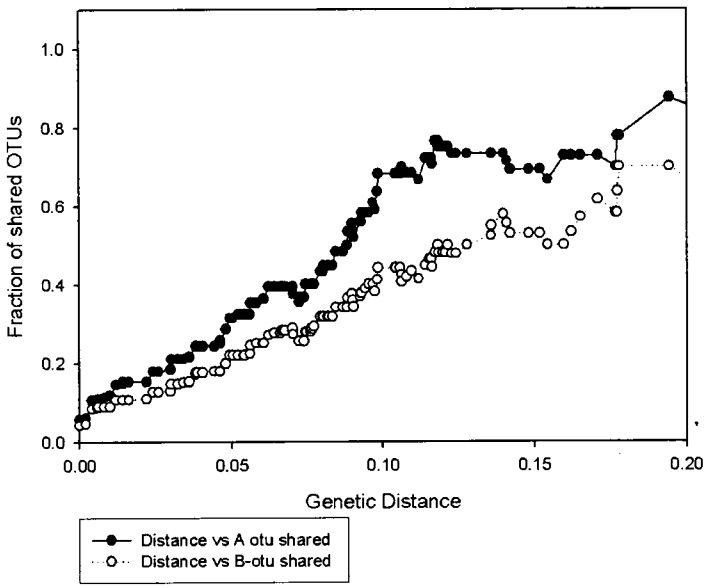


Figure F 2: Comparison of Faeces Libraries – Fraction of Shared OTUs

A. The estimated fraction of shared OTUs between Fc55 (Aotu shared) and Fd55 (Botu shared) is shown. At a genetic distance of 0.03, Aotu shared is 29%, while Botu shared is 26%. The curves follow each other closely as genetic distance increases. **B.** The estimated fraction of shared OTUs between Fc19 (Aotu shared) and Fd19 (Botu shared) is shown. At a genetic distance of 0.03, Aotu shared is 18%, while Botu shared is 27%. The curves follow each other closely as genetic distance increases. **C.** The estimated fraction of shared OTUs between Fc19 (Aotu shared) and Fd55 (Botu shared) is shown. At a genetic distance of 0.03, Aotu shared is 21%, while Botu shared is 15%. The curves follow each other closely as genetic distance increases.



C


Fraction of shared OTUs
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.05
0.10
0.15
0.20
Genetic Distance
Distance vs A otu shared
Distance vs B-otu shared