DETERMINATION OF ABUNDANCE AND TYPES OF GUT MICROBIOTA AND ITS METAGENOMICS MARKERS IN TYPE 2 DIABETES MELLITUS, PREDIABETICS AND NORMAL PATIENTS VISITING SOUTH C MEDICAL CENTRE IN NAIROBI KENYA

DR. BINOD KUMAR (MBBS)

MASTER OF SCIENCE

(Molecular Medicine)

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2023

Determination of Abundance and Types of Gut Microbiota and Its Metagenomics Markers in Type 2 Diabetes Mellitus, Prediabetics and Normal Patients Visiting South C Medical Centre in Nairobi Kenya

Dr. Binod Kumar (MBBS)

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Molecular Medicine of the Jomo Kenyatta University of Agriculture and Technology

2023

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university

Signature……..Date……..

Dr. Binod Kumar (MBBS)

This thesis has been submitted for examination with our approval as university supervisors.

Signature……Date……..

Prof. Johnson Kinyua, PhD

JKUAT, Kenya

Signature……. Date……..

Dr. James Kimotho PhD

KEMRI, Kenya

DEDICATION

I dedicate this thesis to my late wife, Nita Sinha, my children Abhishek Sinha and Payal Sinha, Nidhi Sinha and Rajiv Srivastava, grandson Riyansh and granddaughter Vihika and the staffs of Bioinformatics Institute of Kenya specifically Mr. Augus Amollo for his kind inspiration and exceptional support and staffs of South C Medical Center Nairobi specifically management, medical record department and nurse named Mr. Joshua who selflessly supported me in data collection.

ACKNOWLEDGEMENTS

I wish to thank the almighty God for His goodness and gift of resources to go through this work. I also wish to express my gratitude to my supervisors, Prof. Johnson Kinyua, Dean SOBS Jomo Kenyatta University of Agriculture and Technology (JKUAT) and Dr. James Kimotho, Department of Innovation and Technology Kenya Medical Research Institute (KEMRI), for their guidance, encouragement and stimulating discussion during the course of the study and the preparation of this thesis.

Many thanks go to Mr. Augus Amollo of Bioinformatic Institute of Kenya for his immense advice and technical assistance.

I wish to extend my sincere appreciation and thanks to the staffs of South C Medical Centre Nairobi for allowing me to access their facility for data collection and medical record.

Many thanks go to Nurse Mr. Joshua of South C Medical Centre who provided me profound cooperation and support during data collection.

Last but not least, I am eternally grateful to my relatives for their financial support without which I could not have completed this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

F/B ratio	Firmicutes Bacteroides ratio
T2DM	Type 2 diabetes mellitus
2-AG	2-acylglycerol
2-OG	2-oleoyl glycerol
2-PG	2-palmitoylglycerol
ADA	American Diabetes Association
AHA	American Heart Association
AMOVA	Analysis of Molecular Variance
APCs	Antigen presenting cells
BMI	Body Mass Index
CDKN2A	Cyclin-dependent kinase inhibitor 2A
DCs	Dendritic cells
DNA	Deoxyribose Nucleic Acid
DPP-4	Dipeptidyl peptidase-4
ED	Erectile dysfunction
ERB	Ethical Review Board
FTO	Fat mass and obesity associated
GALT	Gut-associated lymphoid tissue
GDM	Gestational diabetes mellitus
GLP-I	Glucagon-like peptide-1
GWAS	Genome wide association studies
HDAC	Histone deacetylase
HDL	High density lipoproteins
HHEX	Haematopoietically expressed homeobox
HOMOVA	Homogeneity of molecular variance
IBD	Irritable bowel disease
IBS	Irritable bowel syndrome
IDF	International Diabetes Federation

IEC	Intestinal epithelial cells
IELS	Intraepithelial lymphocytes
IGF2BP2	Insulin-like growth factor two binding protein 2
ILCs	Innate lymphoid cells
KCNJ11	Potassium inwardly rectifying channel, subfamily J, member 11
KIBs	Bioinformatics Institute of Kenya
KNH	Kenyatta National Hospital
LDA	Linear discriminant analysis
LDL	Low density lipoproteins
LEfSe	Linear Discriminant Analysis (LDA) Effect Size
LPS	Lipopolysaccharide
MCT1	Monocarboxylate transporter 1
MDA	Mean decrease accuracy
MLNs	Mesentery lymph nodes
MTNR1B	Melatonin-receptor gene
MOTHUR	Analytical Tool
NACOSTI	The National Council of Science and Technology
NAFLD	Non-alcoholic fatty liver disease
OOB	Out of box
OSA	Obstructive sleep apnea
OTU	Operational taxonomic unit
PAMPs	Pathogen-associated molecular patterns
PCoA	Principal Coordinates
PCOS	Polycystic ovary syndrome
PCR	Polymerase chain reaction
PERMANOVA	Permutational Multivariate Analysis of Variance
PI3K	Phosphatidylinositol-3-kinase
PPARG2	Peroxisome proliferator-activated receptor gamma 2
PRRs	Pattern recognition receptors
QIIME	Quantitative Insights into Microbial Ecology

RF	Random forest
rRNA	Ribosomal RNA
RS1	Insulin receptor substrate 1
SILVA	Analytical Software
SCFA	Short chain fatty acid
SMCT1	Sodium-coupled monocarboxylate transporter 1
T1DM	Type 1 diabetes mellitus
TCF7L2	Transcription factor 7-like 2
TGF-b	Transforming growth factor beta
TLRs	Toll like receptors
Tregs	CD4+ T regulatory cells
WHO	World Health Organization

ABSTRACT

Type 2 Diabetes Mellitus (T2DM) has been progressing rapidly globally affects 347 Million people and one of the 10 leading cause of death. In Kenya prevalence of diabetes is projected to be 5.3% by 2025. There has been a myriad of factors responsible for unabated progression and organs damage and end result reflects the altered normal gut microbes. If altered microbiota is identified at earliest stage and treated then organs damage such as kidney, nerves and eyes can be avoided. This study aims at abundance of microbes and their metagenomics markers in diabetes in patients visiting South C Medical Centre Nairobi, Kenya. No metagenomics study on the role of the gut microbes in diabetes has been conducted in Kenyan population and there are no existing metagenomics markers of diabetes in Kenya. As such there is an existing knowledge gap on the type and abundance of gut microbiota and their metagenomics markers in diabetes, pre-diabetes and normal individual. This study fills the gap. This study adopted two basic objectives namely; type and abundance of bacteria colonizing the gut of diabetes, pre-diabetes and normal individual and their metagenomics markers based on the identified genera and abundance of bacteria. South C Medical Centre was chosen for study because of accessibility during pandemic period, no other hospital in Nairobi allowed for data collection. Total target population was 79 participants using Fisher's Statistical formula and stratified random sampling in cross-sectional study. Random sampling from each strata 33 diabetes, 13 pre-diabetic and 33 normal individual were selected by random blood sugar measurement then fecal samples were collected from all 3 strata of persons and subjected to 16SRNA and V5-V6 Gene Sequencing. Reads were analyzed using MOTHUR Microbial Pipeline Analysis, alignments were done using SILVA Reference Microbial dataset. Alpha and Beta diversity were annotated by using rarefaction curve which show marked microbial dysbiosis in diabetic group of individual. High abundance of proteobacteria, low level of bacteriods and high firmicutes and bacteriod ratio were obtained by analysis. Significant abundance of Escherichia_shigella (P value = 0.000588, FDR=0.004706) was reported in diabetes. It was observed that high level of Escherichia_shigella in diabetes patients contributes to progression of diabetes disease by upregulating the inflammatory pathway and downregulating the insulin receptor gene and causing diabetic complication. This study recommends that if in diabetic patient implantation of gut microbiota from normal gut microbes individual is done then life-threatening complications can be mitigated and alleviated.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Type 2 diabetes mellitus (T2DM) is a chronic debilitating condition whose main features include hyperglycemia, insulin resistance, abnormal lipid, carbohydrate, and protein metabolism, and progressive decline leading to microvascular complications such as retinopathy, neuropathy, and nephropathy (Kahn, et al., 2013). Type 2 diabetes mellitus is a major cause of morbidity and is the number one cause of adult blindness, lower limb amputations, and kidney failure (AHA, 2010). Type 2 diabetes mellitus is also a major cause of mortality with annual deaths of 1.6 million people directly attributed to the disease globally. Worldwide, T2DM affects 347 million people and is projected by the World Health Organization (WHO) to be one of the ten leading causes of death by 2030 (WHO, 2020) and to have a global prevalence burden of 615 million people by 2040 (Zheng, Ley, & FB., 2018). In Kenya, the prevalence of T2DM is estimated to be 5.3% (Ayah, et al., 2013)

Known risk factors for T2DM include sedentary lifestyle, poor diet, obesity, old age, family history, race, genetics, acanthosis nigricans, low HDL levels, and a history of heart disease and stroke (Martinez et al., 2019). Lifestyle changes such as physical activity, healthy body weight, diet, and avoidance of smoking can help prevent T2DM. Screening for early signs of prediabetes disease, control of blood cholesterol levels, and screening and treatment of retinopathy can also help prevent the condition and its complications (WHO, Diabetes., 2020). First line treatment with metformin is preferred where adequate glycaemia is not achieved with diet and exercise. Management with other drugs such as oral sulfonylureas and dipeptidyl peptidase-4 (DPP-4) inhibitors can be initiated later. Other drugs such as pioglitazone, alpha-glucosidase inhibitors, glucagon-like peptide-1 (GLP-I) receptor agonists, and insulin are also available (Kahn, Cooper, & Prato., 2013).

Several studies have linked gut microbiome dysbiosis to the onset and progression of health conditions such as colorectal cancer, irritable bowel disease (IBD), food allergies, Crohn's disease, and non-alcohol steatohepatitis (NASH). Emerging evidence also suggests that gut microbiome dysbiosis may play an important role in T2DM (FH Karlsson, 2013) (Larsen, et al., 2010) (Qin, et al., 2012); (Zhang, et al., 2013). Some studies suggest that the therapeutic activity of metformin in T2DM is largely based on its effect on the gut microbiome. According to Vallianou et al. (2019), metformin alters the composition of the gut bacteria by increasing the types and amounts of mucin-degrading Akkermansia muciniphila, as well as several short-chain fatty acid - producing microbiota (SCFA). Increase in these bacterial species leads to the increased production of butyrate and propionate which take part in glucose homeostasis (Vallianou,et al.., 2019).

Gut bacteria are also thought to contribute to T2DM by reshaping the intestinal barrier, changing host signaling pathways associated with insulin production, uptake, and resistance, and changing the host metabolism. Studies show that T2DM is characterize by a significant decrease in butyrate-producing bacteria (Vallianou, et al., 2019). Metagenomic markers of T2DM in the gut include specific populations of bacteria in the gastrointestinal tract that are differentially elevated in a number or type compared to bacterial populations in the gut of people without T2DM. Significantly reduced levels of bacteria such as Bifidobacterium, Bacteroides, Faecalibacterium, Akkermansia and Roseburia have been reported in diabetic patients and contribute to disease by increasing insulin resistance, reshaping the intestinal barrier, and altering host metabolism and signaling pathways.

Conversely, Ruminococcus, Fusobacterium, and Blautia species are positively associated with diabetes type 2 and abundance of these bacteria related to severity of diabetes type 2 and vice versa. The bacteroides/firmicutes ratio is also a known marker of gut dysbiosis and consequent disease process (Vallianou, et al.., 2019). Both the type and abundance of the gut bacteria is influenced by a variety of factors including diet, lifestyle, ethnicity, and genetics and varies across different geographical regions. This

means that prior findings of studies conducted in other geographical areas may not necessarily apply to the Kenyan context (Duffy & Wen., 2017). This study aims at determining the types and abundances of gut bacteria in diabetic, pre-diabetic, and normal individual also it aims at determining the metagenomic markers among them based on significant differences in types and abundances of microbes.

1.2 Problem Statement

Mounting evidence shows that gut bacteria play an important role in the onset and progression of T2DM (Karlsson, 2013; Larsen, et al., 2010; Qin, et al. 2012; Vallianou, et al., 2019; Zhang, et al., 2013). Gut bacteria can either precipitate or enhance T2DM disease or ameliorate it. Despite the important role of these bacteria and the widespread prevalence and effects of T2DM, no metagenomic study on the role of the gut microbiome in T2DM has been conducted in the Kenyan population and there are no existing metagenomic markers for T2DM in Kenya. As such, there's an existing knowledge gap on the types and abundances of gut bacteria in diabetic and pre-diabetic patients compared to normal healthy individuals.

Type 2 diabetes mellitus (T2DM) accounts for 90-95% of all diabetic cases globally representing 347 million people and is projected by the World Health Organization (WHO) to become one of the ten leading causes of death by 2030. According to the International Diabetes Federation (IDF), 80% of global T2DM cases manifesting highly elevated increases in T2DM occur in middle- and lower-income countries (IDF, 2013). In Kenya, for example, the prevalence of T2DM is estimated to be 5.3% (Ayah, 2013). 50% of all diabetic patients remain asymptomatic unless their blood sugar level crosses 10mmol/L by the time they realize there is something wrong in the body irreversible damages already occurred in the organ that is the reason prevalence of diabetes globally has been progressing unabatedly. Morbidity due to diabetes is so high as it causes lifelong renal replacement therapy like dialysis permanently or renal transplant that cripples the socio-economic fibres.

1.3 Justification

Type 2 Diabetes Mellitus (T2DM) once develops it progresses unabatedly and damages the vital organ permanently. Medicine and insulin replacement therapy have to be continued lifelong. However, organ damage at molecular level is inevitable due to cascading inflammatory pathway which upregulate genes of inflammatory pathway NF-Kb and downregulate genes of insulin receptors due to imbalance in proportion of gut microbes in diversity and density. This study identified microbial metagenomics markers in Type 2 Diabetes Mellitus (T2DM) which not only fills the gap as no gut microbial metagenomics markers of diabetes has not been conducted in Kenya patients before but also gives direction of management so that patients do not have to be dependent on lifelong medication and insulin injection. Secondly, a person with pre-diabetic whose blood sugar range between 7-8mmol/L remains asymptomatic and quite oblivious of their problem and sooner rather than later develops fully fledged diabetic and permanent organ damage. Through this study one can know the metagenomics markers of pre-diabetic and targeting in the management so that it can be nipped in the body before it blossoms into fully fledged diabetic and its complication like nephropathy, retinopathy and neuropathy.

1.4 Research Questions

1. What are the types and abundance of bacteria colonizing the gut of diabetic, pre-diabetic, and normal individuals?
2. What is the profile of diabetic metagenomics markers that is associated with the identified genera and abundance of bacteria?
3. What is the correlation between metagenomics markers of T2DM, prediabetic and normal (non-diabetic) subjects?

1.5 Aims and Objectives

1.5.1 General Objective

To determine the metagenomic markers of diabetic, prediabetic and normal patients obtained from types and abundance of microbiota colonizing the gut of the patients visiting South C Medical Centre Nairobi, Kenya.

1.5.2 Specific Objectives

1. To determine the types and abundance of bacteria colonizing the gut of diabetic, pre-diabetic, and normal, patients visiting South C Medical Centre.
2. To profile diabetic metagenomics markers based on the identified genera and abundances of bacteria.
3. To assess the correlation between metagenomics markers of T2DM, prediabetic and normal (non-diabetic) subject

CHAPTER TWO

LITERATURE REVIEW

2.1 Diabetes Classification

Diabetes mellitus is a chronic debilitating and heterogeneous metabolic disorder that is characterized by chronic hyperglycemia associated with insulin insufficiency and or sensitivity (Darwish., 2015). The key presenting features of diabetes mellitus include defects in insulin secretion, resistance, and or action; hyperglycemia, relative insulin insufficiency, and disturbances in the metabolism of proteins and fat (WHO, 2019)

Diabetes is commonly classified into four groups: type 1 diabetes mellitus (T1DM), type 2 diabetes mellitus (T2DM), other types, and gestational diabetes mellitus (GDM) (ADA, 2014). Type 1 diabetes mellitus can be classified into 3 groups, namely autoimmune type 1, idiopathic type 1, and fulminant type 1 diabetes (Darwish., 2015). Autoimmune TIDM results from the destruction of the pancreatic β cells and accounts for 5-10% of all diabetic cases with 80% to 90% of all diabetic cases in adolescents and children (Daneman, 2006) (Maahs, et al., 2010). Other types of diabetes include monogenic diabetes, exocrine pancreas disease, and genetic syndromes (Darwish., 2015). Gestational diabetes is characterized by hyperglycemia before or during pregnancy (Metzger, et al., 2008).

2.2 Economic Burden of Type 2 Diabetes

Type 2 diabetes mellitus (T2DM) accounts for 90-95% of all diabetic cases globally. Worldwide, T2DM affects 347 million people and is projected by the World Health Organization (WHO) to be one of the ten leading causes of death by 2030 (WHO, 2020) and to have a global prevalence burden of 615 million people by 2040 (Zheng, et al., 2018). Whereas the highest prevalence of T2DM is in the Middle East and North Africa region (MENA) with a prevalence rate of 10.9%, the Western Pacific region has the largest number of adults diagnosed with T2DM (IDF, 2013). According to the

International Diabetes Federation (IDF), 80% of global T2DM cases manifesting highly elevated increases in T2DM occur in middle- and lower-income countries (IDF, 2013). In Kenya, the prevalence of TDM is estimated to be 5.3% (Ayah, 2013).

Type 2 diabetes mellitus is associated with significant mortality and morbidity. According to the American Heart Association (AHA), T2DM is the number one cause of adult blindness, lower limb amputations, and kidney failure (AHA, 2010). People with T2DM are at an increased risk of nonalcoholic fatty liver disease (NAFLD), cataracts, obesity, and erectile dysfunction (ED). They also are at a higher risk of heart disease and peripheral arterial and cerebrovascular disease (WHO, Diabetes., 2020). Compared to non-diabetics, people with T2DM are at higher risk of infectious diseases such as tuberculosis, eye infections, bacterial pneumonia, urinary tract infections, skin and soft-tissue infections, and mucocutaneous and invasive fungal infections. Infectious diseases such as Klebsiella pneumoniae liver abscess, Fournier's gangrene, and rhinocerebral mucormycosis are almost always found only in diabetic patients (Crevel, et al., 2017).

Type 2 diabetes mellitus is also linked to more hospital visits, increased hospitalization rates, increased admissions to intensive care units, and increased amputations and death for necrotizing limb infections (Uehara, et al., 2014). According to Grijalva et al (2015), people with T2DM are also at heightened risk of increased mortality due to influenza, bloodstream infections, and pneumonia (Grijalva, et al., 2015). Data by Wu et al (2014) indicates that there is an elevated resistance to antimicrobials in T2DM patients (Wu, et al., 2014). Type 2 diabetes mellitus also highly elevates the risk of endogenous endophthalmitis in patients with bacteraemia or fungaemia (Vaziri, et al., 2015) and reduces the effectiveness of preventive vaccination (Castilla, Godoy, Domínguez, & etal., 2013). Annual global deaths directly attributable to T2DM total 1.6 million people. T2DM affects 347 million people and is projected by the World Health Organization (WHO) to be one of the ten leading causes of death by 2030 (WHO, Diabetes., 2020) and to have a global prevalence burden of 615 million people by 2040 (Zheng, et al., 2018).

2.3 Diagnosis of Type 2 Diabetes

Diagnostic criteria for diabetes are well established and diagnosis is based on the plasma glucose levels or hemoglobin A1C levels. Fasting plasma glucose (FPG) of ≥ 126 mg/dL (7.0 mmol/L), random plasma glucose ≥ 200 mg/dL (11.1 mmol/L), and HbA1c ≥ 6.5% (48 mmol/mol) is diagnostic of diabetes mellitus. The HbA1C levels are also used to monitor response to diabetes mellitus treatment (ADA, 2014).

2.4 T2DM Risk Factors and Associated Complications

Complications associated with T2DM include cardiovascular disease, diabetic nephropathy, diabetic neuropathy, diabetic retinopathy, and cancers. Cardiovascular disease is a leading cause of disease and death in both pre-diabetic and diabetic subjects. According to Chaturvedi (2007), cardiovascular disease in T2DM results from oxidative stress which contributes to atherogenesis and oxidation of low-density lipoproteins (LDL) (Chaturvedi., 2007).

Diabetic neuropathy is another common complication in T2DM and it leads to the formation of calluses, skin infections, ulcerations, foot and bone infections, and gangrene due to the loss of the protective sensation in feet. Diabetic neuropathy also contributes to erectile dysfunction, non-healing skin wounds, amputations, and foot ulcers (Sanghera & Blackett., 2012; Zatalia & Sanusi., 2013). Diabetic nephropathy is a leading cause of kidney disease. Diabetic retinopathy results from microvascular damage to retinal tissues caused by chronic hyperglycemia, leading to increased vascular permeability and fluid extravasation into the vitreous humor (Wu et al., 2014). Several studies demonstrate that the risk of colorectal, breast, liver, and kidney cancers is elevated in patients with T2DM (Donadon, et al., 2008) Diabetes mellitus, body size, and bladder cancer risk in a prospective study of Swedish men (Larsson, et al., 2008) Diabetes mellitus and risk of breast cancer: a meta-analysis., (Larsson, et al., 2007). Diabetes mellitus and incidence of kidney cancer: a meta-analysis of cohort studies. , (Larsson & Wolk, 2011)

Risk factors for T2DM include lifestyle, obesity, age, race, family history, acanthosis nigrans, low HDL level and or high triglyceride levels, hypertension, polycystic ovary syndrome (PCOS), and a history of stroke or heart disease. Lifestyle plays a major role in the onset and progression of T2DM. Lifestyle factors implicated in the development of T2DM include physical inactivity, consumption of alcohol, smoking, and a sedentary lifestyle. Diet is another important modifiable risk factor for T2DM. A higher risk of T2DM development is associated with a low fiber, high glycemic diet. Additionally, intake of large amounts of fatty acids is associated with insulin resistance and a higher T2DM risk. A negative correlation exists between consumption of total and saturated fat and T2DM independent of body mass index (BMI). Frequent consumption of processed red meat and soft drinks is also associated with a higher risk of T2DM development (Liu, et al., 2000; Willett, & Rimm, 2002; Schulze, et al., 2004; Dhingra, et al., 2007; Dam, et al., 2002).

Obesity is the most important risk factor for T2DM development. According to the World Health Organization (WHO), 90% of all diabetic patients develop T2DM due to excessive body weight (WHO, Classification Of Diabetes Mellitus, 2019). Age is also a risk factor for T2DM: people aged 45 years and above are at a higher risk. On race, the risk of T2DM is higher in African Americans, Asian Americans, Pacific Islanders, Alaska Natives, Hispanics and Latinos, American Indians, and Native Hawaiians (WHO, Classification Of Diabetes Mellitus, 2019).

Family history is yet another important risk factor for T2DM. Monozygotic twins have higher concordance rates than dizygotic twins. Kobberling (1982) demonstrated that 40% of first-degree relatives of T2DM may develop the disease and this rate is significantly higher than the 6% observed in the general population (Kobberling, 1982). Candidate genes uncovered by genome wide association studies (GWAS) and which are associated with T2DM development are listed include TCF7L2 (transcription factor 7-like 2), FTO (fat mass and obesity associated) gene, KCNJ11 (potassium inwardly rectifying channel, subfamily J, member 11), CDKN2A (cyclin-dependent kinase inhibitor 2A), RS1 (insulin receptor substrate 1), MTNR1B (melatonin-receptor gene),

IGF2BP2 (insulin-like growth factor two binding protein 2), HHEX (haematopoietically expressed homeobox) and PPARG2 (peroxisome proliferator-activated receptor gamma 2) (Wu, et al., 2014).

Acanthosis nigrans refers to the thick, dark, velvety skin behind the neck or under the armpits and has been strongly linked to T2DM, obesity, and insulin resistance. Due to this strong linkage, it has been severally proposed as a marker for T2DM (Bahadursingh, et al., 2014). More than 50% of all women with PCOS develop T2DM by age 40. Gambineri et al (2012) conducted a long-term prospective study of a large cohort of women with PCOS who were followed from youth to middle age to determine the association between PCOS and T2DM. Their results demonstrated that the risk of T2DM is significantly increased in women with PCOS (Gambineri, et al., 2012). Obstructive sleep apnea (OSA) may also influence the development of prediabetes and diabetes and is more common in diabetics than in non-diabetics ((Pamidi & Tasali., 2012). The role of gut bacteria in the onset and development of T2DM is increasingly being studied. Musso et al (2011) demonstrated that gut microbiome is a risk factor for T2DM development (Musso, Gambino, & Cassader, 2011)

2.5 Role of Gut Bacteria in T2DM

The role of gut bacteria in human health is fivefold. First, gut bacteria influence the host immune system. Structurally, the human gut consists of a mucosa made up of a single layer of epithelial cells (intestinal epithelial cells or IECs) and intraepithelial lymphocytes (IELS). The IECs have paneth and goblet cells. Paneth cells produce antimicrobial peptides while goblet cells produce mucus that provides coating for the epithelial layer. Below the mucosal layer is a gut-associated lymphoid tissue (GALT) known as lamina propria which contains Peyer's patches, T cells, B cells, CD8+ and CD4+ cells, innate lymphoid cells (ILCs), and antigen presenting cells (APCs). GALT is the largest immune system organ in the human body and exerts both local and systemic effects (Gopalakrishnan, et al., 2018) (figure 2.1).

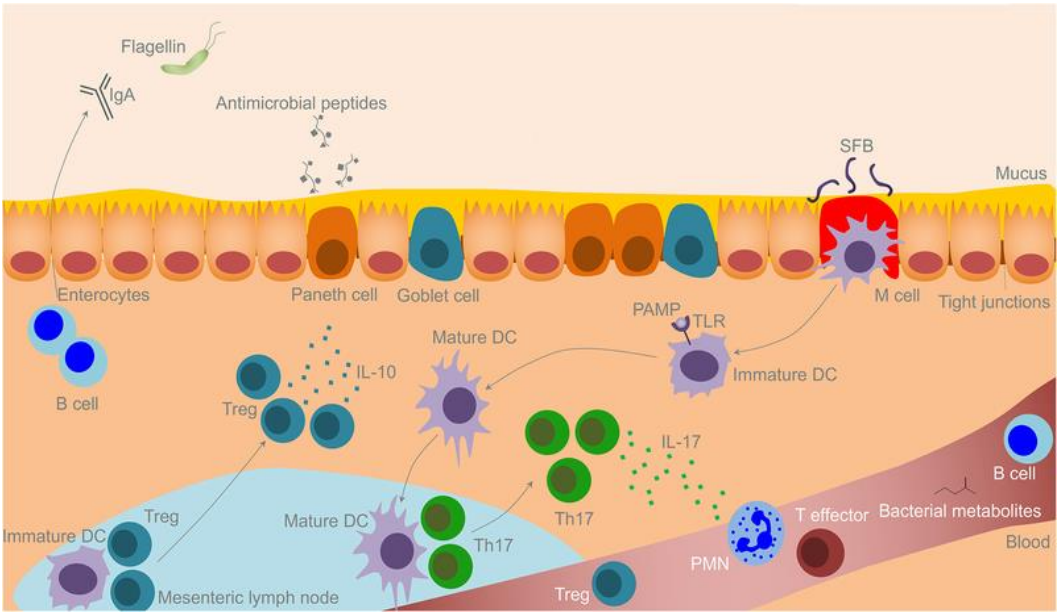


Figure 2.1: Structure of the human gut and the role of the gut microbiome in immunity. Source: Golapakrishnan et al. (2018)

Pattern recognition receptors (PRRs) such as toll like receptors (TLRs) detect the pathogen-associated molecular patterns (PAMPs) present on the invading organism surface and initiate an immune response. The toll like receptors are thirteen in number (TLR 1-13) and recognize PAMPs including flagellin, lipopolysaccharide (LPS) for TLR4, lipopeptides for TLR2/1 and TLR2/6 heterodimers, flagellin for TLR5, unmethylated CpG motifs in DNA for TLR9, profilin and Salmonella flagellin for TLR11 and various forms of RNA for TLRs 3, 7, 8 and 13. There is no known ligand currently identified for TLR 12 (Barbalat, Lau, Locksley, & Barton, 2009); (Mathur & Barlow, 2015).

2.5.1 Boosting host immunity

PAMPS induce the maturation of antigen presenting cells (APCs) such as Dendritic Cells (DCs) which travel to mesentery lymph nodes (MLNs) where they stimulate T cells to form CD4+ T regulatory cells (Tregs), and Th17 cells and CD8+ cells. Activation of the Nf-kB pathway by PAMPs leads to elevated levels of IgG, IgM, and

CD8 T cells thereby boosting host immunity (Dunkelberger & Song, 2010). The increased risk of T2DM patients to infections is due to their lower immunity (Hodgson et al., 2015) resulting from reduced levels of neutrophils, macrophages, T cells, and antibody-producing B cells; obesity, vitamin D deficiency, oxidative stress, and skin lesions and poor wound healing due to diabetic neuropathy, macroangiopathy, and microangiopathy (Crevel, Vijver, & Moore, 2017). Others are elevated levels of Staphylococcus aureus carriage in diabetics, overgrowth of urinary pathogens due to glycosuria, and nosocomial infections due to frequent attendance of hospitals (Crevel, Vijver, & Moore, 2017). Gut dysbiosis in T2DM patients has been established by several studies and is linked to the reduced immunity in T2DM patients.

2.5.2 Improvement in glucose metabolism

Gut bacteria such as Roseburia intestinalis, Bacteroides fragilis, Akkermansia muciniphila, Lactobacillus plantarum, and L. casei induce IL-10, leading to improvement of glucose metabolism. The improvement in glucose metabolism resulting from the over-expression of IL10 is based on the fact that IL10 is protective against ageing-related insulin resistance (Dagdeviren et al., 2017). Another mechanism through which bacteria protect the host from T2DM disease involves increase in IL-22 production. Bacteria such as R. intestinalis increase the production of IL-22 and IL-22 alleviates T2DM since it is an anti-inflammatory cytokine that restores insulin sensitivity, promotes the differentiation of T regulatory cells, induces TGF-b and suppresses intestinal inflammation (Dagdeviren, et al., 2017); (Hoffmann, 2016); (Zhu, 2018). According to Hoffman (2016), the bacteria Bacteroides thetaiotaomicron is protective against T2DM since it induces expression of T regulatory cell gene expression (Hoffmann, 2016). Bacteria negatively associated with T2DM increase the levels of several inflammatory cytokines that lead to disease (Gurung, et al., 2020).

Gut bacteria also ferment non-digestible dietary fibers such as resistant starch (RS), inulin, oat bran, wheat bran, cellulose, Guar gum, and pectin to produce short fatty chain acids (SFCAs) (Parada, et al., 2019) which may protect against T2DM. Structurally,

SCFAs consist of carboxylic acid and a small hydrocarbon chain and include butyrate, propionate, iso-butyrate, valerate, iso-valerate, hexanoate and acetate. Butyrate is produced by firmicutes bacteria while acetate, which is the most abundant SCFA in the gut, is produced by bacteroidetes. Bacteroidetes also produce propionate. In the human gastrointestinal tract, the highest SCFA concentration is found in colon at a molar ratio of approximately 60:20:20 for acetate:propionate:butyrate (Parada, et al., 2019). SCFAs receptors include GPR41 (free fatty acid receptor 3; FFAR3), GPR43 (free fatty acid receptor 2; FFAR2), and GPR109A (hydroxycarboxylic acid receptor 2; HCAR2).

Most abundant SCFAs are butyrate, acetate, and propionate as they form about 95% of total SCFAs (Cook & Sellin, 1998). In the colon, most of the SCFAs (~95%) are absorbed rapidly by large intestinal mucosal cells. The remaining amount is excreted in the feces (Topping & Clifton, 2001). Most of the absorbed SCFAs are used as energy sources by the body and acetate remains as the most abundant SCFA in the circulation with small amounts of butyrate and propionic acid in the periphery. SCFAs are transported by monocarboxylate transporter 1 (MCT1; encoded by SLC16A1) and the sodium-coupled monocarboxylate transporter 1 (SMCT1; encoded by SLC5A8). SCFAs are ligands for G-protein coupled receptors (GPCRs), including GPR109A, GPR43, and GPR41, thereby activating anti-inflammatory signaling cascades (Topping & Clifton, 2001).

Butyrate inhibits histone deacetylase (HDAC). Inhibition of HDAC leads to glycogen production, stimulation of insulin signaling, and glucose uptake all of which result in reduced hyperglycemia and reduction of inflammatory damage on β cells. Butyrate is known to reduce the levels of IFN-induced IFITM1, IFITM3, RIG-1, and IFIT2 in a dose-dependent manner. Acetate attaches to Gpr43, Gpr41, and Gpr109A receptors on the pulmonary epithelial cells, switching on IFN-beta and IFNAR leading to NF-kB activation. SCFAs boost the host immune system by modulating an increase in the production and expression of IgA (Davie, 2003).

The third mechanism through which gut bacteria protect against T2DM is through reducing insulin resistance and improving glucose homeostasis in organs like the muscles and liver. Gut bacteria also modulate sugar digestion and gut hormones involved in sugar digestion (Gurung, et al., 2020). According to Kim et al (2014), Bifidobacterium lactis reduces the expression of genes involved in hepatic gluconeogenesis, improves translocation of glucose transporter-4 (GLUT4), increases the synthesis of glycogen, and improves the uptake of insulin-stimulated glucose (Kim, et al., 2014). More evidence of the role of gut bacteria in the reduction of insulin is provided by Dang et al (2008) who reported that Akkermansia muciniphila reduces postprandial glucose by inhibiting alpha-glucosidase activity thereby preventing complex carbohydrates from being broken down (Dang, et al., 2018).

According to Plovier, reduction of the expression of hepatic flavin monooxygenase 3 (fmo3) by bacteria such as Akkermansia muciniphila and Lactobacillus plantarum results in the prevention of hyperlipidemia and hyperglycemia in mice (Plovier, 2017). Lactobacillus casei lowers insulin resistance through reduction of hepatic glycogenesis and reduction of gene expression of Akt, phosphatidylinositol-3-kinase (PI3K), AMPK, and insulin receptor substrate 2 (IRS2). In addition, it lowers hyperglycemia by upregulating genes such as CFTR, SLC26A3, SLC26A6, Bestrophin-3, ClC1-7, GABAAa1, and GlyRa1 (Zhang, et al., 2014); (Le, 2015). According to Singh (2017), insulin sensitization is further improved by L. rhamnosus which does this by enhancing the levels of adiponectin (Singh, 2017).

The fourth mechanism through which gut bacteria prevent T2DM is through the reduction of fatty acid synthesis and increase in fatty acid oxidation and energy expenditure. Bacteria which act in this manner include Akkermansia muciniphila, Bacteroides acidifaciens, and Lactobacillus gasseri (Houmard, 2008). Everard (2013) reports that Akkermansia muciniphila elevates oxidation of fatty acids and differentiation of adipocytes by increasing 2-acylglycerol (2-AG), 2-palmitoylglycerol (2-PG), and 2-oleoyl glycerol (2-OG) in the adipose tissue (Everard, 2013). Other bacteria that increase the oxidation of fatty acids are Bacteroides acidifaciens and

Lactobacillus gasseri (Kang, 2013); (Yang, 2017). Butyrate, propionate, and acetate also induce the oxidation of fatty acids. Butyrate does this by inhibiting HDAC in muscles leading to increased energy expenditure. The three SCFAs mentioned reduce PPAR-γ expression in adipose and hepatic tissues thereby leading to fatty acid oxidation (Gao, 2009); (den Besten, 2015).

The fifth mechanism through which the gut bacteria prevent T2DM is through potentiation of anti-diabetic therapy. Gurung et al (2020) have reviewed the role of anti-diabetic drugs on the gut microbiota and their findings are summarized in table 2.1 below. Anti-diabetic drugs promote the growth of "good bacteria" which in turn prevent T2DM disease (Gurung, et al., 2020).

Table 2.1: The effect of antidiabetic drugs on the gut microbiota (Gurung, et al., 2020)

Anti-diabetic Drug	Effectson of Microbiota Promotes	Reference PMID	Reduces	Reference PMID
Biguanides (Metformin)	Akkermansia muciniphila escherichia Bifidobacterium adolescentis, Lactobacilius Butyrivibrio, Bifidobacterium bifidum, Megasphaera, Prevetolia, Escherichi-Shigetta, Erysipelorichaceae incatatesedis, Fusobacterium, Flavanifactor, Lachnospiraceae incertaesedis and Clostridium XVIII and IV	23804561, 28530702, 25038099, 27999002, 29056513, 30261008, 30815546, 29789365	Intesnibacter, Romboutsia, Peptostreptococcaceae, unclassified Clostridiaceae-1, unclassified Asacharospora, Allistpes, Oscilibacter, Bacteroides, Parabacteroides, unRuminococaceae	28530702, 30261008, 29789365,
Alpha Glucosidese inhibitor (eg Acarbose, voglibose, miglitol	Lactobacillus, Faecalbacterium, Dailister, Subdoligranulum, Allisoella, Megasphaera, Bifidobacterium, Entercoccus faecalis	28130771, 29176714, 25327485	Btyriclococcus, Phasclarctobacterium, Ruminococcus, Eggerthela, Bacteroides, Oribacterium, Etysipelotichaceae, Coriobatctericeae, Bacteroides	28130771, 28349245, 29176714, 25327485
GLP-1Receptor agonist(eg. Liraglutide)	Akkrmansia muciniphila, Bateroides acidifacieus, Lachnoclostridium flavonifracter, Rumninococcus gnavius, Allbaculum, Turicibacter, Anaerostipes, Lactobacillus, Butirimonas, Desulfovibrio,	30815546, 30292107, 29171288, 27633081,	Helicobacterer, Prevotella, Ruminococcaceae, Chistensllacea, Roseburia, Canddatus Arthromitus, Marvinbryantia, Incertaesedis	30292, 107, 29171288, 27633081
Thiazolidinediones (eg. Pioglitazone) DPP-4 Inhibitor (eg. Vildagliptin, Silagliptinm, Saxogliptin	Lactobacillus streptococcus, Bacterides acidifaciens, Streptococcus hyoitestinalis, Eryspitotrichaceae, Allbaculu, Turibacter, Roseburia,	29797022, 29036231, 27633081, 276310113	Proteobacteria	27751827
SGLT2 inhibitors (eg. Dapagliflozin	Akkermansia Entrococcus	29703207	Oscillobacter, Ruminoclostridium_6, Anaerotrumcus, Kurthai, Christensenellaceae, Prevotellaceae, Bacteroides, Prevotella, Blautia, Oscillospira	29797022, 29036231, 27633081, 27631013

Specific metabolites or bacterial byproducts can alter the dendritic cell in a fashion that allows them to skew towards a Treg versus Th17 phenotype. Tregs function in secreting IL10, creating a local anti-inflammatory cytokine milieu. Th17 cells, meanwhile, produce IL17 which can increase Paneth cell production of antimicrobial peptides and can function in recruiting PMNs from the bloodstream. Some bacterial metabolites can enter the bloodstream directly further altering the systemic immune system (Gopalakrishnan, et al, 2018).

2.6 Metagenomic markers

Previous metagenomic studies have identified several bacteria that are significantly elevated in T2DM. Faecalibacterium, Akkermansia and Roseburia have been reported to be negatively co-associated with diabetic patients and contribute to disease by increasing insulin resistance, reshaping the intestinal barrier, and altering host metabolism and signaling pathways. Conversely, high abundances of Ruminococcus, Fusobacterium, and Blautia species are positively associated with T2DM (Vallianou et al., 2019). High firmicutes and bacteroidetes (F/B) ratio in T2DM is reported in previous studies by (Larsen et al., 2010). However, low firmicutes and bacteroidetes (F/B) ratio in T2DM has also been reported by (Schwiertz et al., 2009). Escherichia_Shigella metagenomic marker of T2DM has been reported by (Maskarinec et al., 2021). Proteobacterial species are known to be archetypal signatures of microbial dysbiosis (Shin et al., 2015). The normal gut microbiome of healthy people comprises of the following 4 phyla: Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria in that order. Verrucomicrobia and Fusobacteria are also present but in smaller quantities (Turnbaugh et al., 2007; Blaut, 2013). Polymorphism in the gene encoding Heat stock factor 1 (HSF) inreases the propensity of T2DM (Elen Klyosova et al., 2022) and NF-KB protein encoding gene in relation of T2DM and its microvascular complication (Romzova et al., 2006). In Africa a study conducted in Nigeria identified metagenomics markers known as Ruminococcus, Bifidobacterium and Collinsellen for the diabetes and Peptostreptocossuss for the normal (Afolayan et al., 2020).

Metagenomics microbial analysis is being done by using QIIME, MOTHUR and DADA2 software. MOTHUR gives high number of Operational taxonomic unit (OTU) than QIIME and DADA2. This study used MOTHUR and SILVA pipeline metagenomics microbial analysis. Although DADA2 has the advantage of giving single nucleotide polymorphism (SNP) and Amplicon sequence variant (ASV), this study did not use it due to its complexity and unavailability.

2.7 16S rRNA

The study of bacterial taxonomy and phylogeny is largely based on the 16S rRNA. The 16S ribosomal RNA (16S rRNA) is part of the 30s small subunit of the prokaryotic ribosome encoded by the 16S rRNA gene. Functionally, the 16S rRNA is involved in the binding of the 50S and 30S ribosomal units through its interaction with the 23S unit. It has the anti-Shine-Dalgarno sequence at its 3' end which attaches to the mRNA AUG start codon and to the S1 and S21 proteins thereby facilitating the synthesis of proteins. The 16S rRNA also facilitates the codon-anticodon pairing in the A site (Jay & Inskeep, 2015). The 16s rRNA has 9 hypervariable regions (V1-V9) that can be used to accurately fingerprint or identify bacteria. Fingerprinting of bacteria using 16S rRNA is useful because 16S rRNA is present in nearly all bacteria and exists as an operon or multigene family, it has a slow rate of evolution and is therefore largely conserved, and its genome size of 1.5 kbp is big enough for use in bioinformatics analysis (Coenye & Vandamme, 2003) (Patel, 2001).

Fingerprinting of bacteria is commonly based on specific hypervariable regions and not the entire 16S gene since these regions can provide high resolution at phylum level as accurately as the full gene. Studies demonstrate that the V3 region is best at identifying bacteria at genus level while the V6 is best at differentiating bacteria at species level (Patel, 2001). One of the weaknesses of the 16S fingerprinting is that it may not resolve species that are related very closely. For instance, species in the Clostridiaceae, Peptostreptococcaceae, and Enterobacteriaceae families share 99% homology and cannot be resolved accurately at species level. This is because the V4 sequences differ

by just a few nucleotide sequences and reference databases may not be able to classify such bacteria at species level. Limiting 16S studies to just a few hypervariable regions may fail to capture differences in taxa that are closely related and therefore underestimate the diversity (Jovel, et al., 2016). Long-read sequencing with Nanopore may overcome this weakness however because it can sequence the entire hypervariable region (Cuscó, et al., 2019).

To fingerprint bacterial communities using 16S rRNA typing, samples are first subjected to DNA extraction followed by PCR amplification on the desired hypervariable region using a set of universal primers. Several universal primers have been described in literature (Table 2.2).

Table 2.2: 16S rRNA Universal Primers (Greg, 2018)

Primer name	Sequence (5′–3′)
8F	AGA GTT TGA TCC TGG CTC AG
27F	AGA GTT TGA TCM TGG CTC AG
U1492R	GGT TAC CTT GTT ACG ACT T
928F	TAA AAC TYA AAK GAA TTG ACG GG
336R	ACT GCT GCS YCC CGT AGG AGT CT
1100F	YAA CGA GCG CAA CCC
1100R	GGG TTG CGC TCG TTG
337F	GAC TCC TAC GGG AGG CWG CAG
907R	CCG TCA ATT CCT TTR AGT TT
785F	GGA TTA GAT ACC CTG GTA
805R	GAC TAC CAG GGT ATC TAA TC
533F	GTG CCA GCM GCC GCG GTA A
518R	GTA TTA CCG CGG CTG CTG G
1492R	CGG TTA CCT TGT TAC GAC TT

Sequencing of the amplified products is then conducted followed by bioinformatics analysis. Bioinformatics analysis involves pre-processing of reads to remove low-quality features, alignment of processed reads to a reference dataset, and OUT-based or ASV-based analysis. Alignment of the sequences is done based on nucleotide sequence databases such as SILVA, Ribosomal Database Project (RDP), Green genes, Open Tree of Life (OTT), EzBio Cloud, and NCBI. Common software used in bioinformatics analysis of 16S rRNA data include QIIME, MOTHUR and DADA2.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Type

This was a cross sectional study. A cross sectional study is an observational study that analyzes data from a population, or a representative subset, at a specific point in time. A total of three groups were investigated with groups consisting of: (1) people with diabetes, (2) people without diabetes, and (3) people with pre-diabetes. This involved the use of blood sample for random blood sugar test and stool for microbiome study. Blood drawn by syringe injection is important for quick RBS test while stool is the most convenient way to study human microbiome where gastrointestinal tract (GIT) contents cannot be obtained in living participants.

3.2 Study Sites

The study was conducted at two (2) sites namely Diabetes Clinic in south C and the Bioinformatics Institute of Kenya (KIBs), located along Mombasa Road in Nairobi, Kenya. South C Health Centre was chosen because of convenience and accessibility management of this hospital was generously kind to a allow me to access their facility during Covid period in contrast to other hospitals like KNH, Agha Khan Hospital Nairobi and Nairobi Hospital just a few who declined to use their facility during Covid period. Random blood and stool samples were collected at a Diabetic clinic in South C while long read sequencing and bioinformatics analysis were conducted at the Bioinformatics Institute of Kenya. The Diabetic clinic in South C was chosen for sampling of participants because of convenience and flow of varied ethnic and gender of patients visiting the facility. This ensured a diverse participation from different backgrounds. In addition, KIBs has been chosen for this study for the diverse facilities and expertise needed to integrate genetic and clinical data for the improvement of human health, a potential learning platform.

3.3 Sampling Frame

The sampling frame was the list of patients at the South C Diabetic clinic between January 2020 and June 2020.

3.4 Sample Size

The sample size for this study was determined using the Fisher‘s statistical formula (Fisher, et al) .The formula is given as:

$$n = \frac{z^2 p(1-p)}{e^2}$$

Where:

n= required sample size.

z= Standard normal deviation from the required confidence interval.

p= the proportion of the target population estimated to have some particular characteristics.

e= the degree of accuracy desired.

To obtain an optimal sample for the study at 95% confidence interval, proportion set at 0.5and degree of precision set at 0.5 we obtain:

$$n = \frac{1.96^2 * 0.5(1-0.5)}{0.05^2}$$

$$= 384$$

Thus, a sample size of approximately 384 individuals was required for the study.

The individuals were selected by simple random sampling from the clinical records of South C Medical Centre Nairobi Diabetic clinic called and requested to partake in the study. They were tested and diagnosed for the study thoroughly. The results placed the study subjects into 3 sub-groups as required by the study objective. From the testing results,160 individuals were found to be diabetic while 64 were found to be pre-diabetic. The remaining 160 individuals were normal. The proportion of individuals with required characteristics was 75%.

The study was narrowed down to a smaller sample size for efficient analysis and due to feasibility and available cost resources. The sample was calculated at 95% confidence interval and degree of precision was set at 9.5%. The new sample size was calculated as:

$$n = \frac{1.96^2 * 0.75(1 - 0.75)}{0.095^2}$$

=79

Now, 79 individuals were required for the final study.

Proportional allocation scheme

Proportional allocation is a procedure for dividing a sample among the strata in a stratified sample survey. The stratified sampling scheme sets the sample size in each stratum equal to be proportional to the number of sampling units in that stratum. That is, nh/n = Wh. Proportional allocation therefore yielded a self-weighted sample (no additional weighting was required to estimate unbiased population parameters).

The study therefore utilized the simplistic Nyman's optimal allocation as;

$$n_h = \frac{n.N_h}{N}$$

Where;

nh =final required sample size for every stratum or group

N =total required secondary sample size

Nh =primary class sample size for every group

N =the total original sample size of 384

The sample was allocated to each of the groups in terms of proportional allocation optimally considering the sizes of the population in each of the three groups with respect to the total population under study. The sample was allocated as follows:

Normal=$\frac{160}{384}$*79=33

Diabetic=$\frac{160}{384} * 79$=33

Pre-diabetic=$\frac{64}{384}$*79=13

Thus, 33 diabetic individuals, 33 normal individuals and 13 pre-diabetic individuals were selected for the study as proved by the scientific sampling scheme using proportional optimal allocation.

3.5 Sample Selection

Random blood sugar (RBS) test was used to group participants into either normal, pre diabetic or diabetic depending on their sugar levels. The measurements were taken from a convenience sample 2 hours after eating from each participant. This was obtained from the blood collected from the participants who signed the consent form. Blood sample is necessary for quick analysis of random blood sugar. A small sample of blood (drop) was taken using a needle, often from the finger by skin puncture. This sample was then be measured by a glucometer machine The syringes were disposed of in a sharps container

and decontaminated before they were taken for sterilization and incineration. Participants with readings below 5-6 mm/l 2 hours post-prandial were classified as normal as per the American Diabetes Association (ADA) and were therefore assigned to the non-diabetic group (ADA,2014). Those with readings between 7-8 mm/l were classified as pre-diabetic and assigned to the pre-diabetic group while those having readings above 10 mm/l or higher were classified as diabetic and assigned to the diabetic group. Assignment to each group was based on their RBS readings. Each group was further stratified proportionately based on gender, age, and for the diabetic group, treatment and disease severity and presence of microvascular complications. The following criteria applied.

3.6 Inclusion Criteria

Aged 18 years and above, Kenyan citizen, male or female, consents to be in the study. For the non-diabetic group, participants with 2 hours post-prandial readings of 140 mg/dl or less. For the pre-diabetic group, participants with 2 hours post-prandial readings of 140-199 mg/dl. For the diabetic group, participants with 2 hours post-prandial readings of 200 mg/dl and above or HbA1c ≥7.0% and ≤10.5% either on diet and exercise alone or on a stable dose of metformin (≥1000 mg/day) for 3 months prior to screening and with a body mass index (BMI) ≥23 and ≤45 kilograms per meter squared at screening

3.7 Exclusion Criteria

Pregnant or lactating females. History of active uncontrolled gastrointestinal disorders or diseases including: inflammatory bowel disease (IBD) including ulcerative colitis (mild-moderate-severe), Crohn's disease (mild-moderate-severe), or indeterminate colitis, irritable bowel syndrome (IBS) (moderate-severe), persistent, infectious gastroenteritis, colitis or gastritis, persistent or chronic diarrhea of unknown etiology, Clostridium difficile infection (recurrent) or Helicobacter pylori infection (untreated) and Chronic constipation. People on antibiotic medication. People using any of the following within

the previous 6 months will be precluded from the study: systemic antibiotics, antifungals, antivirals or antiparasitic (intravenous, intramuscular, or oral), oral, intravenous, intramuscular, nasal or inhaled corticosteroids, cytokines, methotrexate or immunosuppressive cytotoxic agents, Large doses of commercial probiotics consumed (greater than or equal to 108 cfu or organisms per day) - includes tablets, capsules, lozenges, chewing gum or powders in which probiotic is a primary component. Ordinary dietary components such as fermented beverages/milks, yogurts, foods do not apply. For female subjects, combination hormone vaginal ring for contraception (due to unknown duration of local hormone effects). People with actively bleeding hemorrhoids. People with gastrointestinal disease. Vital signs outside of acceptable range at Screening Visit, i.e., blood pressure >160/100, oral temperature >100°F, pulse >100. Any condition that in the opinion of the Investigator precluded participation in the study.

3.8 Recruitment

Diabetic Clinic at South C requested patients meeting the defined inclusion and exclusion criteria to give their stool samples after getting informed consent from them. Those who agreed were shown how to collect the stool sample given a stool collection kit which included a toilet hat, gloves, collection tube, collection instructions, alcohol wipes, a biohazard mailing bag, and a postage-paid return mailing envelope, a form on collection date and time for filling, and a biohazard mailing bag.

3.9 Stool Collection Protocol

For those patients who agreed to participate in the study and who signed the consent form, the collection procedure was demonstrated to them with the collection materials in hand. This was done with the help of the staff recruiter who explained how the kit reagents and equipment's work and the best way to collect the samples. Participants were then allowed to carry home a collection kit that included a toilet hat, instructions for collecting the specimen, a collection tube, exam gloves, alcohol wipes. The tube included a non-toxic stabilizing reagent and mixing apparatus and was safe for home

use. After the sample was collected and the tube was capped, the user vigorously shook the tube for 30 seconds to homogenize and liquefy the sample. At that point the stool DNA was preserved for at least 60 days at ambient temperature. The remaining stool samples were stored in Ethanol at -200C fridge or cold room. The stool samples were then disposed by means of incineration once the study was complete and the retaining time for all the confirmations done.

3.10 Participant Data (Extraction Form)

This data was obtained from the participant records and some was generated before the initial collection of data. Table 3.1 below guides the data collection

Table 3.1: Data Collection Sheet

Participant code	Age	Height	Weight	BMI	Residence	Blood pressure	Glucose level	Any other symptoms	Date of first diagnosis

3.12 Ethical Considerations

Ethical approval was sought and obtained from the Jomo Kenyatta University of Agriculture and Technology (JKUAT) Ethical Review Committee and a research permit obtained from the National Council of Science and Technology (NACOSTI) thereafter. The participants were required to sign informed consent forms and participation was voluntary and well distributed for diversity and justice. The study ensured that the procedure for obtaining the sample was non-intrusive and did not cause harm to the participants. The skin prick for RBS measurement doesn't cause any harm to the participant and the stool sample collection is also harmless. The names of the participants were scrubbed and instead codes were used for privacy and confidentiality of the participants and the data obtained. The study aims and objectives, methods, and any other pertinent information was explained to the participants prior to their signing of the informed consent. The right of the participants to exit the study at any time was

upheld. To protect participant identity, anonymization by use of codes was used. The information obtained was confidential and the data kept under lock and key. The overall benefit generated from this study saw the participants given first priority.

3.11 Sample Collection

Faecal swabs were collected using sterile swab sticks. Participants were given sterile polypots and asked to deposit pea-sized amounts of stool inside the containers and to avoid contaminating the samples with urine or stool water. Hands were washed before and after collection. The peel pouch containing the swab was opened and patient details written on the tube containing the transport medium. Without touching the swab tip, the tip of the swab was inserted into the stool sample and rotated, taking note to collect bloody, slimy or watery areas of the stool. The swab was removed and examined to ensure enough fecal material was collected and the swab inserted again into the stool and rotated in case insufficient material had been swabbed and taking care not to overload the swab or scoop larger pieces of stool. The swab with the fecal material was then transferred into the tube with the transport medium, ensuring that the maximum filling line on the label was not exceeded.

Holding the swab shaft between thumb and finger, the swab was mashed and mixed gently against the side of the tube to ensure that the swab specimen was evenly dispersed and suspended in the transport medium. With the tube facing away from the collector's face, the end of the swab shaft was bent at an angle of 180 degrees to break it at the marked breakpoint and the broken upper part of the swab shaft discarded and the cap tightened. Vials were shaken gently until the sample appeared well mixed. Samples were kept in resealable plastic bags and transported within 6 hours to the lab in a cool box. The samples were stored at -20oC until DNA extraction.

3.12 DNA Extraction

One can isolate a few bacteria from culture but not all bacteria in fecal sample. Fecal sample contains thousands of bacteria so it was not tenable to isolate all fecal bacteria by culture using different media. Only possible for metagenomic analysis to extract DNA from all bacteria present in fecal sample.

DNA extraction from all sample diabetic (33 samples), pre-diabetic (13 samples), and normal (33 samples) bringing the total to 79 samples selected by a stratified random selection from 160 diabetic, 160 normal and 64 prediabetes based on the modified QIAamp® DNA Mini & Blood Kit method for the isolation of bacterial DNA from swabs. Briefly, the cotton swabs were cut using a sterile razor and placed in 1.2 ml sterile Eppendorf tubes. 400 ul of PBS (Phosphate Buffered Saline) was added to the Eppendorf tubes and subjected to vortexing to mix the solution. The solution was allowed to stand for 90 minutes. The cells were resuspended by pipetting. 200 ul of the bacterial solution was transferred into a 2.0 ml sterile Eppendorf tube. The enzymatic lysis buffer (ELB) containing 20 mM Tris-HCl, 2 mM EDTA, and 1.2% Triton was prepared. 5 mL of 1 M Tris-HCl, 1 mL of 0.5 M EDTA, 3 mL of Triton X-100 was mixed and molecular-grade water added to bring the volume up to 250 mL. Purpose of this process was to break down bacteria cell wall but gram (+) bacteria cell wall are toughest to breakdown its wall hence lysozyme enzyme was added in the solution and incubated for 30 minutes then to purify the DNA. Purification enzyme Proteinase K was added to the solution then solution was transferred in Mini spin column and centrifugation was done for 1 minute. The final process was to add wash buffered solution to wash any other contents in DNA then incubation at room temperature (15-250 C) for 5 min was followed by centrifugation at 8000rpm for 5 min. the eluate DNA was then stored at -20^0 C before sequencing.

3.13 DNA Sequencing

DNA sequencing was done using the Nanopore 16S Barcoding Kit (SQK-RAB204) (ONT, UK). This kit was used because it allows samples to be multiplexed in a single run hence reducing the price per sample and labour, intensity while allowing genus level bacterial identification. The Barcoding kit was used in combination with the FLO-MIN106 flow cells, the Flow Cell Wash Kit (EXP-WSH004) (ONT, UK), the Flow Cell Priming Kit (EXP-FLP002) and the MINION sequencing machine. The Barcoding Kit consists of 16S adapters (RAP), sequencing tether (SQT), loading beads (LB), sequencing buffer (SQB), and 24 barcode primers labelled from 16S01 to 16S024 (ONT, UK).

3.14 Library Preparation

The 16S Barcodes were thawed at room temperature, mixed by pipetting up and down, and spun down briefly then kept on ice until ready to use. The extracted DNA was prepared in nuclease-free water and 10 µl (10 ng) of the DNA dispensed into a 0.2 ml thin-walled PCR tube. 14 µl of nuclease-free water, 10 µM 1 µl of 16S Barcode and 25 µl of LongAmp Taq 2X master mix was added into the PCR tube to make a total of 50 ul. This was mixed gently by flicking the tube, spun down, and amplified using the following cycling conditions shown in Table 3.2.

Table 3.2: PCR cycling conditions for sequencing

Cycle Step	Temperature	Time	Number of Cycles
Initial denaturation	95 °C	1 min	1
Denaturation	95 °C	20 secs	25
Annealing	55 °C	30 secs	25
Extension	65 °C	2 mins	25
Final extension	65 °C	5 mins	1
Hold	4 °		

Samples were transferred to clean 1.5 ml Eppendorf DNA LoBind tubes and the AMPure XP beads resuspended by vortexing. 30 µl of the resuspended AMPure XP

beads were added to the reaction and mixed by pipetting then incubated on a rotator for 5 minutes at room temperature. 500 µl of fresh 70% ethanol was then prepared in nuclease-free water. The sample was spun down then pelleted on a magnet. With the tube on the magnet, the supernatant was pipetted off and the beads washed with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. The ethanol was then removed using a pipette and discarded. The tubes were spun down and placed back on the magnet. Any residual ethanol was pipetted off and the solutions allowed to dry for ~30 seconds, taking care not to dry the pellet to the point of cracking.

The tubes were removed from the magnetic rack and the pellet resuspended in 10 µl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. This was incubated for 2 minutes at room temperature. The beads were pelleted on a magnet until the eluate was clear and colourless. 1 µl of eluted sample was then quantified using a Qubit fluorometer. All the barcoded libraries were then pooled in the desired ratios to a total of 50-100 fmoles in 10 µl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. 1 µl of RAP was added to the barcoded DNA and the solution mixed gently by flicking the tube, and spinning it down. The reaction was then incubated for 5 minutes at room temperature and the library stored on ice until ready to load into the SpotON MinION flow cell.

3.15 Priming and Loading the SpotON MINION Flow Cell

The SQB, LB, FLT and FB were thawed at room temperature and the SQB, FLT, and FB tubes mixed by vortexing and spinning down at room temperature. The MinION Mk1B lid was then opened, the flow cell slid under the clip and pressed down firmly to ensure correct thermal and electrical contact. The priming port cover was slid clockwise to open the priming port and the SpotON flowcell primed and loaded.

Before loading, the contents of the Loading Beads (LB) tubes were mixed by vortexing and the library prepared as shown in Table 3.3.

Table 3.3: Library preparation schema

Reagent	Volume (µl)
SQB	34.0
LB	25.5
Nuclease-free water	4.5
DNA library	11.0

The flow cell priming was completed by loading 200 µl of the priming mix into the flow cell. The prepared library was mixed by pipetting up and down and 75 µl of sample added to the flow cell. The SpotON sample port cover was closed and the MinION Mk1B lid replaced.

The flowcell was connected to the laptop and the sequencing done using the MinKNOW software v 21.11.7 (ONT, UK, London).

3.16 Data Acquisition

The sequencing reads were saved in the local computer and loaded into the KENCLUST remote server using the WinSCP V 5.19.5. WinSCP is an open-source tool for the secure transfer of files between local computers and remote servers. WinSCP was used because it is free, secure, fast, portable, and supports secure copy protocol (SCP) client for Microsoft Windows, File Transfer Protocol (FTP), SSH File Transfer Protocol (SFTP), WebDAV, and Amazon S3 (WinSCP, 2020).

3.17 Bioinformatics Analysis

Bioinformatics analysis of the sequenced reads was performed using MOTHUR V 1.39.0 (Schloss et al., 2009), R v 4.0.3 (R Core Team, 2020) and Microbiome Analyst (Chong et al., 2020; Dhariwal et al., 2020). MOTHUR was selected as it has been shown to perform better than other metagenomic software such as the Quantitative Insights Into Microbial Ecology (QIIME) and other metagenome software. Not only does MOTHUR yield higher richness ($P < 0.05$) and better rarefaction curves than QIIME, but it also has a relatively larger analytic sensitivity (Lopez-Garcia et al., 2019). The Silva 1.38 release

reference dataset was used since it provides better outcomes than other databases and generates significantly higher number of OTUs in known taxa than QIIME ($P < 0.001$) (Lopez-Garcia et al., 2019).

Analysis of metagenomics data followed the standard MOTHUR pipeline. Briefly, the sequencing reads were first subjected to an initial pre-processing step to remove any sequencing and PCR errors present. Quality control involved dropping of low-quality reads, end trimming, and removal of adaptor sequences, ambiguous sequences, duplicates, and any other undesirable sequences including mitochondria, eukaryotes and other unknown sequences. After preparation of the contigs, ambiguous, duplicated, and long bases were removed. Alignment of the improved sequences to a Silva reference file was done (Quast et al., 2013; Yarza et al., 2014; Yilmaz et al., 2014; Glöckner et al., 2017).

The alignment was based on the V6-V8 regions. The alignment region was ascertained using a reference E. coli sequence obtained from NCBI and the primers used in sequencing. Sequences were filtered to remove overhangs and ensure that the sequences only overlapped the same alignment coordinates. The VSEARCH algorithm was used to remove chimeras followed by clustering using a Bayesian classifier and removal of undesirable sequences using the remove.lineages command. Error rates were assessed using a mock community.

OTU-based analysis included alpha and beta diversity. Alpha diversity indices such as Shannon, Caswell, Berger-Parker, Simpson, inverse simpson, Hill, Margalef, and McIntosh were determined. For beta diversity, species richness was assessed using chao1 chao2, jacknife1 and jackknife 2 plots. Statistical significances were determined using the AMOVA p-value & T-statistic and Mann-Whitney/Kruskal-Wallis p-value & Mann-Whitney statistic. Beta analysis will entail use of PERMANOVA statistics. Additionally, distance matrices were visualized using principal Coordinates (PCoA) and a plots. Statistical significance, p values, and R2 values between groups calculated using Analysis of MOlecular VAriance (AMOVA) and Homogeneity of molecular variance

(HOMOVA) (Excoffier, Smouse & Quattro, 1992). Abundance plots, rarefaction curves, principal component analysis (PCOA) and nmds plots were graphed using R. Phylogenetic trees were drawn using Phylip v.3.698 (Felsenstein, 2005). The correlation of the relative abundance of each OTU with the two axes in the NMDS dataset for different groups was determined using the Spearman method. Metastats is a non-parametric t-test used to determine differentially represented OTUs between the test and control groups in this study. Differential representation of OTUs was also assessed using LEfSe, a linear discriminant analysis effect size tool, and the LDA scores noted. All these tools were used for OTU-based, ASV-based, phylotype-based, and phylogeny-based analyses.

Clustering and correlation involved heatmap clustering and dendrogram and correlation analysis. Dendrogram analysis involved tree creation based on unweighted Unifrac distance while correlation analysis was done using SparCC correlation tables, network, and boxplots. Inter-group and within-group comparisons involve classical univariate analysis (AMOVA p-value & T-statistic; Mann-Whitney/Kruskal-Wallis p-value & Mann-Whitney statistic). It also involved MetagenomeSeq with zero inflated Gaussian fit: p-values and FDR values, Differential Abundance Analysis Methods with log2FC, logCPM, P-values, and FDR values, and Linear Discriminant Analysis (LDA) Effect Size (LEfSe) with outputs of LDA scores, p-values, FDR values, and analysis results table. Finally, random forest (RF) classification was done and outputs such as MDA Accuracy, OOB error rates, RF plots, important features computed and interpreted. Significant differences in types, abundances, and other features were correlated to presence or absence of T2DM.

Table 3.4 shows different modules used in analysis of data MORTHUR.

Table 3.4: Bioinformatics Data Analysis plan

Actual Annotation	Models of Analysis	Statistical Values
Visual Exploration	Abundance profiling	Stacked bar/area plot Interactive pie chart Rarefaction curves FB Ratio
Microbial community profiling	Alpha diversity analysis	Chao 1 index T-test/ANOVA p-value & T-statistic Mann-Whitney/Kruskal-Wallis p-value & Mann-Whitney statistic
	Beta diversity analysis	PCOA plot PERMANOVA statistics
Comparison and Classification	Classical univariate analysis	T-test/ANOVA p-value & T-statistic
	MetagenomeSeq	Zero inflated Gaussian fit: p-values and FDR values
	Differential Abundance Analysis Methods	log2FC, logCPM, P-values, and FDR values
	Linear Discriminant Analysis (LDA) Effect Size (LEfSe)	LDA scores, p-values, FDR values, analysis results table
	Random forest	MDA Accuracy, OOB error rates, RF plots, important features

CHAPTER FOUR

RESULTS

Microbial analysis was done through MOTHUR Pipeline. Before generating OTUs, sequencing reads were first subjected to pre-processing step to remove any sequencing and PCR error present. Quality control involved dropping off low-quality reads, end trimming, removal of adapter sequences, ambiguous sequences and all undesirable sequences then alignment of the improved sequences with SILVA Reference file was done. The alignment was based on V6-V8 regions and then it generates innumerable OTUs.

Then all OTUs of all groups diabetic, pre-diabetic and normal were analyzed on the basis of phylum of microbial and a stacked bar plots were annotated which is depicted in figure 4.1 which has two axis, horizontal axis shows total number of samples 79 including all groups diabetic, pre-diabetic and normal. Figure shows abundance and types of gut microbes Proteobacteria, Firmicutes and Actinobacteria were abundantly present in diabetic sample compared to pre-diabetic and normal. Although proteobacteria also abundantly present in normal sample. This finding of normal sample is abnormal and could be due to normal group participant had been suffering gastro intestinal pathology or it could be due to mixed up sample with diabetics. However, result of diabetic sample corelates with finding of other study done before (Shin et al., 2015). Table 4.1 shows Proteobacteria, Firmicutes and Actinobacteria and their percentage. Firmicutes was more than 20 times more abundant in the diabetes group compared to the prediabetes group and nine times more abundant in the diabetes group compared to the normal group. Actinobacteria was also more abundant in the diabetes than in the normal and prediabetic group. The abundance of Bacteroidetes was low in all the 3 groups. These microbes are incredibly higher in diabetes group in contrast to prediabetic and normal except proteobacteria higher in normal

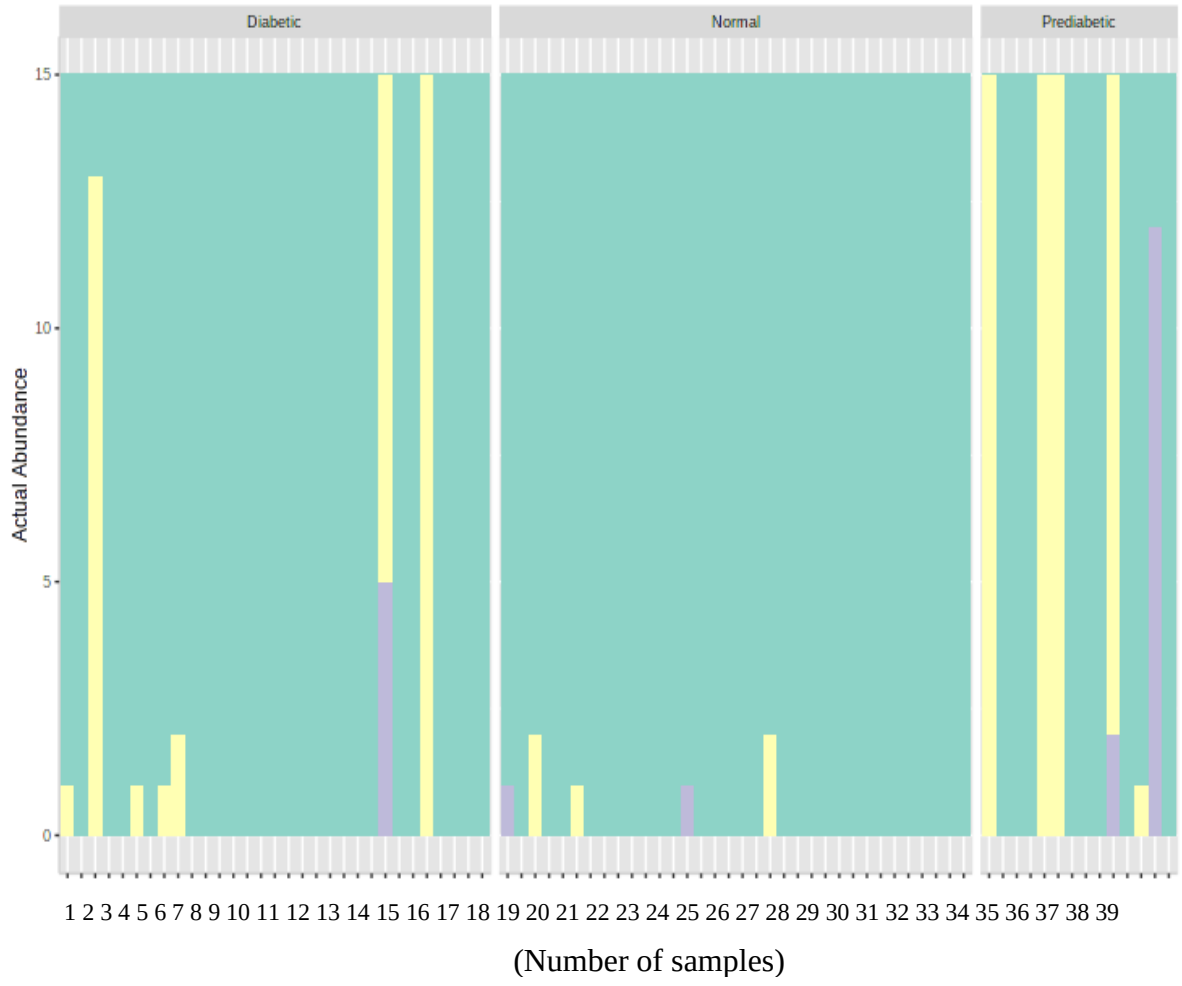


Figure 4.1: Stacked bar plot showing the abundance of gut bacteria phyla.

Table 4.1: Percentage abundance of gut microbiota phyla

	Percent Abundance (%)		
Gut Microbes Phyla	Diabetic	Normal	Prediabetic
Proteobacteria	89.68	98.63	0.65
Firmicutes	9.25	0.98	0.28
Actinobacteria	1.08	0.39	0.07

OTUs analysis on the basis of gut microbial genera was done which is depicted in Figure 4.3

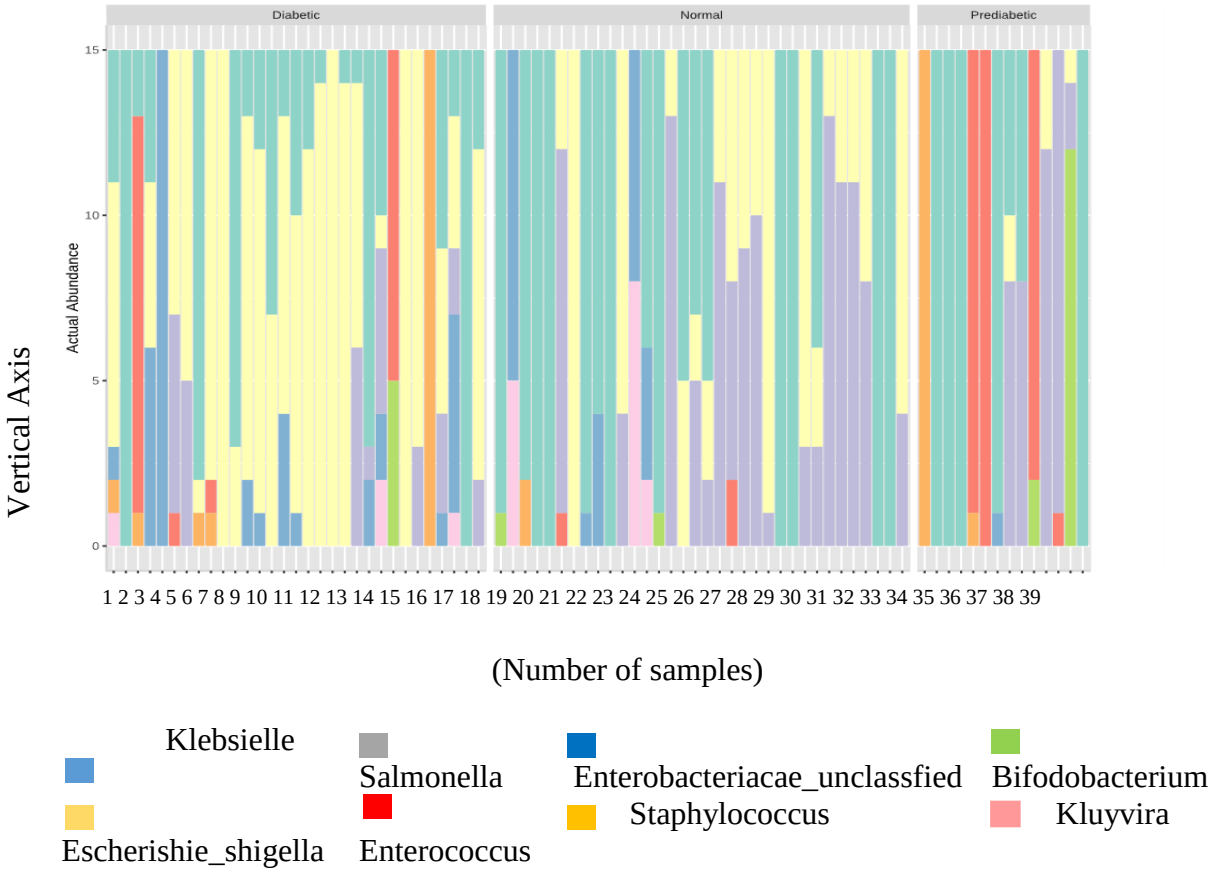


Figure 4.2: gut microbial in genera of all the three groups

Figure 4.3 is depiction of microbiota at genera level found in all three groups diabetic, pre-diabetic and normal individual. Horizontal axis of figure indicate number of samples and vertical axis indicates microbial abundance. In all three groups, abundance of bacteria in descending order Escherichia_Shigella (34.30%), Klebsiella (32.79%), Salmonella (14.24%), Enterobacteriaceae_unclassified (6.61%), Enterococcus (5.70%), Staphylococcus (3.39%), Bifidobacterium (1.58%) and Kluyvera (1.39%). This is depicted in Table 4.2

Table 4.2: Bacteria genera in all groups together.

Bacteria Genus	**Percent Abundance (%)**
Kluyvera	1.39
Bifidobacterium	1.58
Staphylococcus	3.39
Enterococcus	5.70
Enterobacteriaceae_unclassified	6.61
Salmonella	14.24
Klebsiella	32.79
Escherichia_Shigella	34.30
	100.00

Table 4.3 shows each group diabetic, pre-diabetic and normal OTUs percentage in which Escherichia_Shigella is most dominant in diabetic that is 63.6% followed by Enterobacteriaceae_unclassified which is 60.3%, and Staphylococcus (51.4%). In pre-diabetic dominant is Bifidobacterium (66.7%) and Enterococcus (61.4%). In normal dominant bacteria is Kluyvera (78.9%) and Salmonella (61.9%). These are the types and abundance of gut microbes in diabetic, pre-diabetic and normal individual. Reading of normal microbes is considered as abnormal and reason could be mixed sample or gastrointestinal pathology. Each group of microbes in percentage is depicted in table 4.3 and bar diagram of the same is depicted in figure 4.3

Table 4.3: Percent abundances of gut bacteria genera in diabetic, prediabetic, and non-diabetic (normal) sample in each group.

	Diabetic %	Normal %	Prediabetic %
Bifidobacterium	23.8	9.5	66.7
Enterobacteriaceae_unclassified	60.3	38.2	1.5
Enterococcus	34.3	4.3	61.4
Escherichia_Shigella	63.6	34.7	1.7
Klebsiella	28.2	51.4	20.4
Kluyvera	21.1	78.9	0.0
Salmonella	16.3	61.9	21.8
Staphylococcus	51.4	5.4	43.2

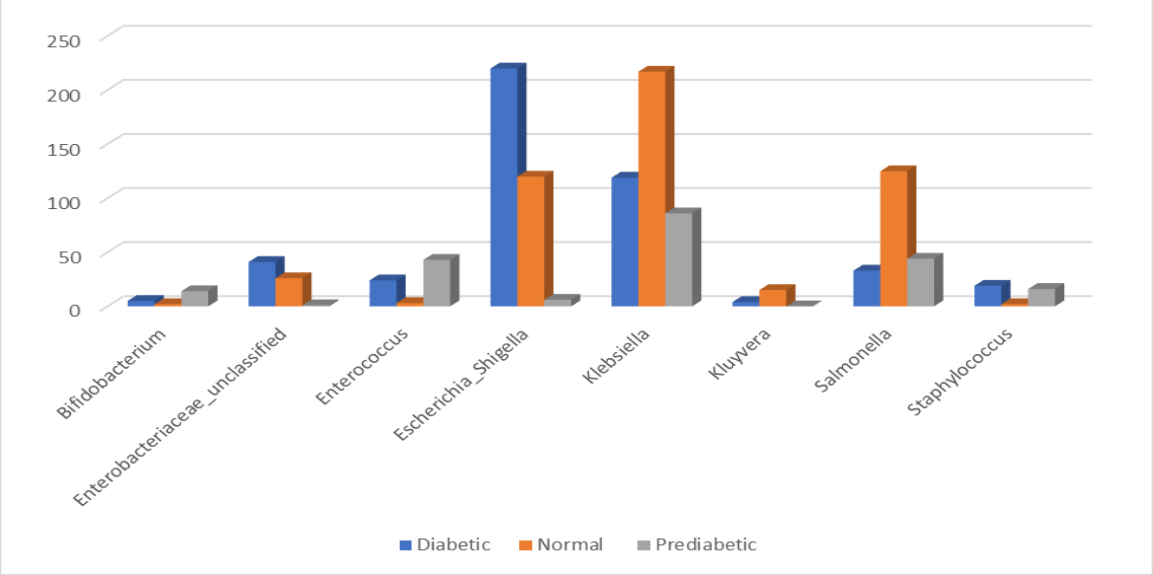
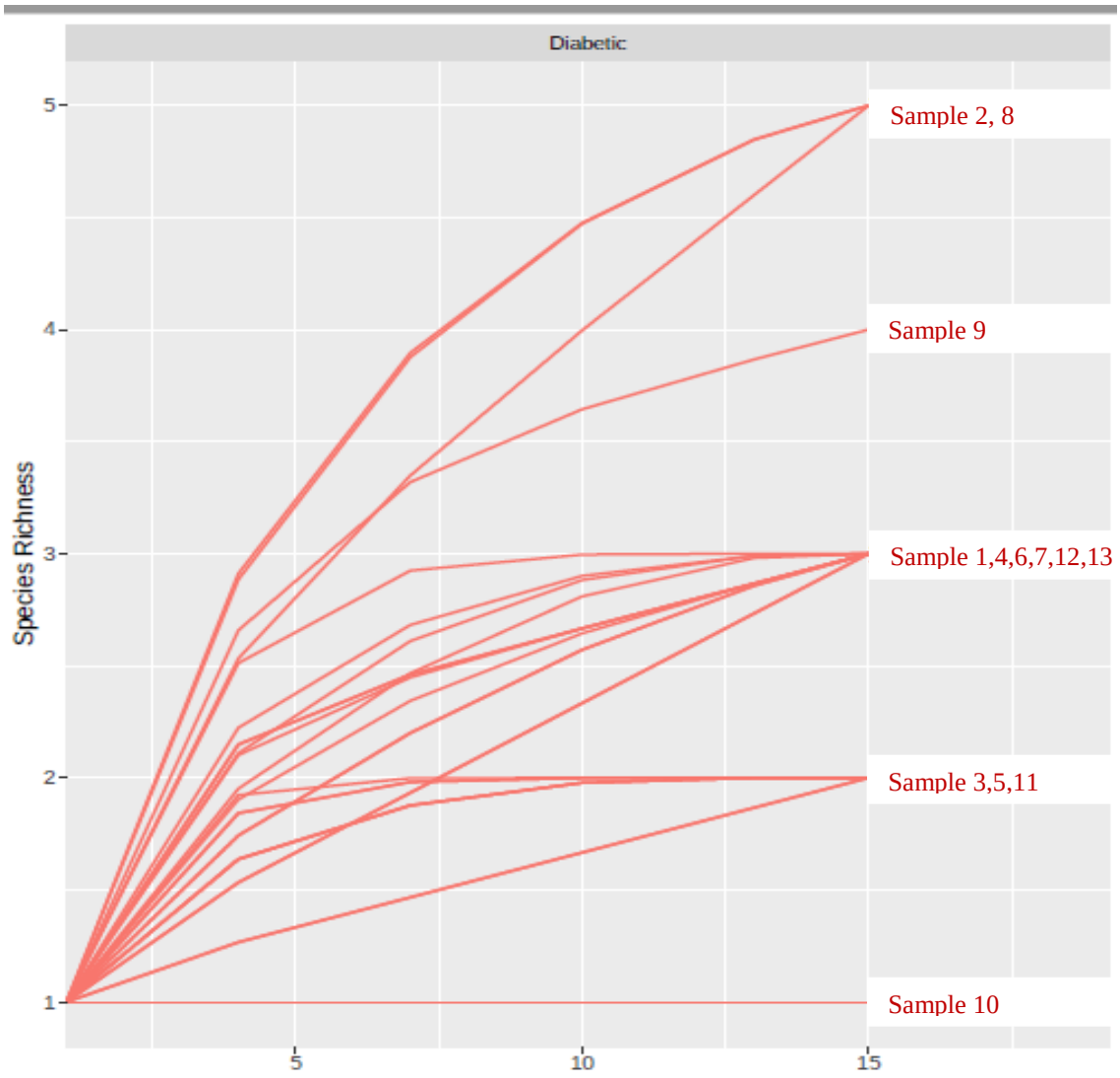


Figure 4.3: Bar chat of gut microbial in diabetic, pre-diabetic and normal

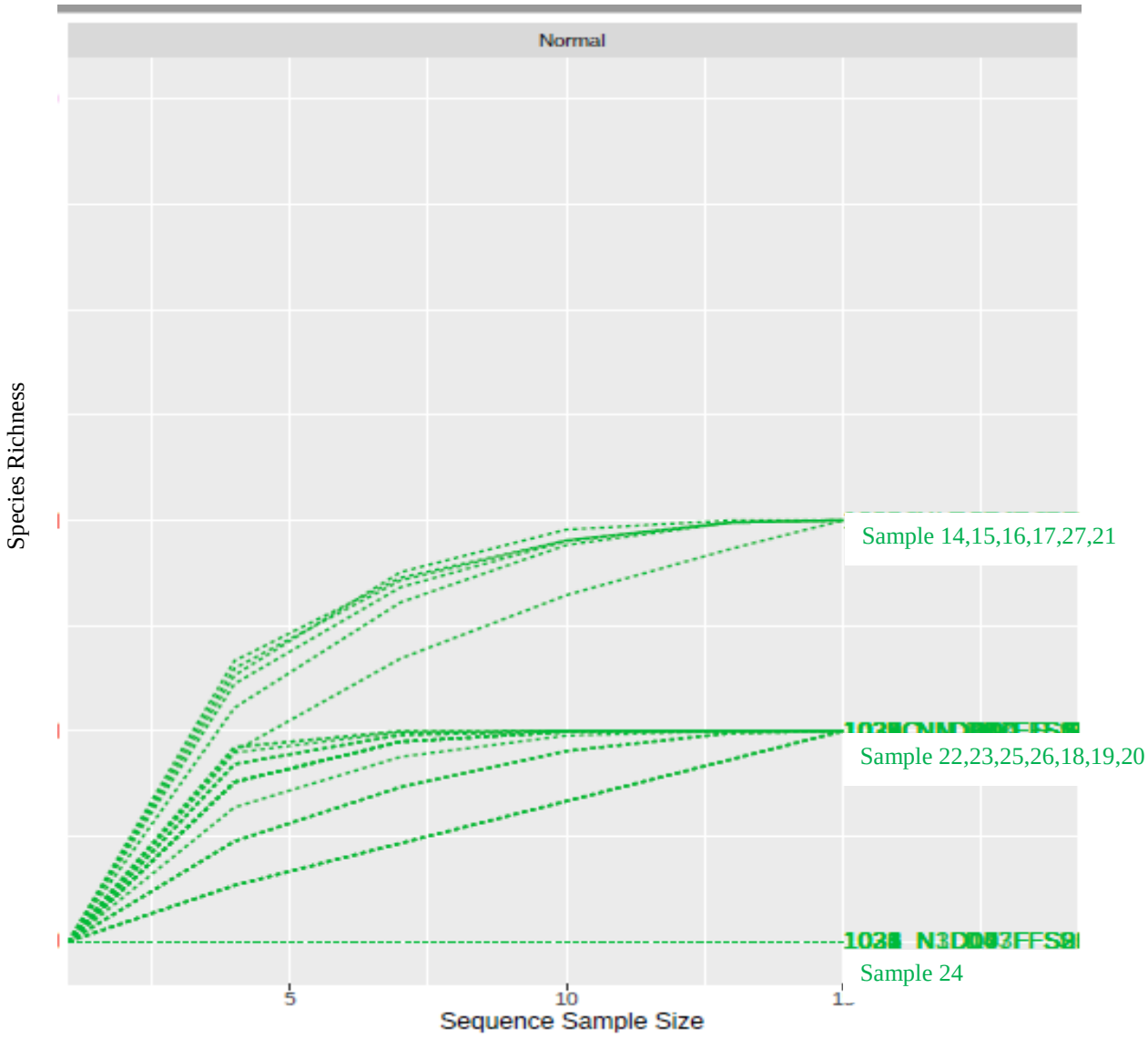
Next model of analysis of gut microbes abundance in diabetic, pre-diabetic and normal is alpha diversity. This model of analysis further confirm the abundance of microbes in diabetic patient. Alpha diversity is a species richness (how many microbes) in each sample of diabetic, pre-diabetic and normal and is drawn by Rarefaction curve which is shown in figure 4.4



Number of clusters (Nucleotides) of Microbes in diabetic

Figure 4.4: Rarefaction curve of diabetic

Figure shows horizontal axis number of cluster of Nucleotides of microbes in diabetic and vertical axis indicates species richness which shows increased diversity and density of gut microbes in diabetic in contrast to pre-diabetic and normal which are depicted in figure 4.5and 4.6.



Number of clusters (Nucleotides) of Microbes in normal

Figure 4.5: Rarefaction curve of normal sample

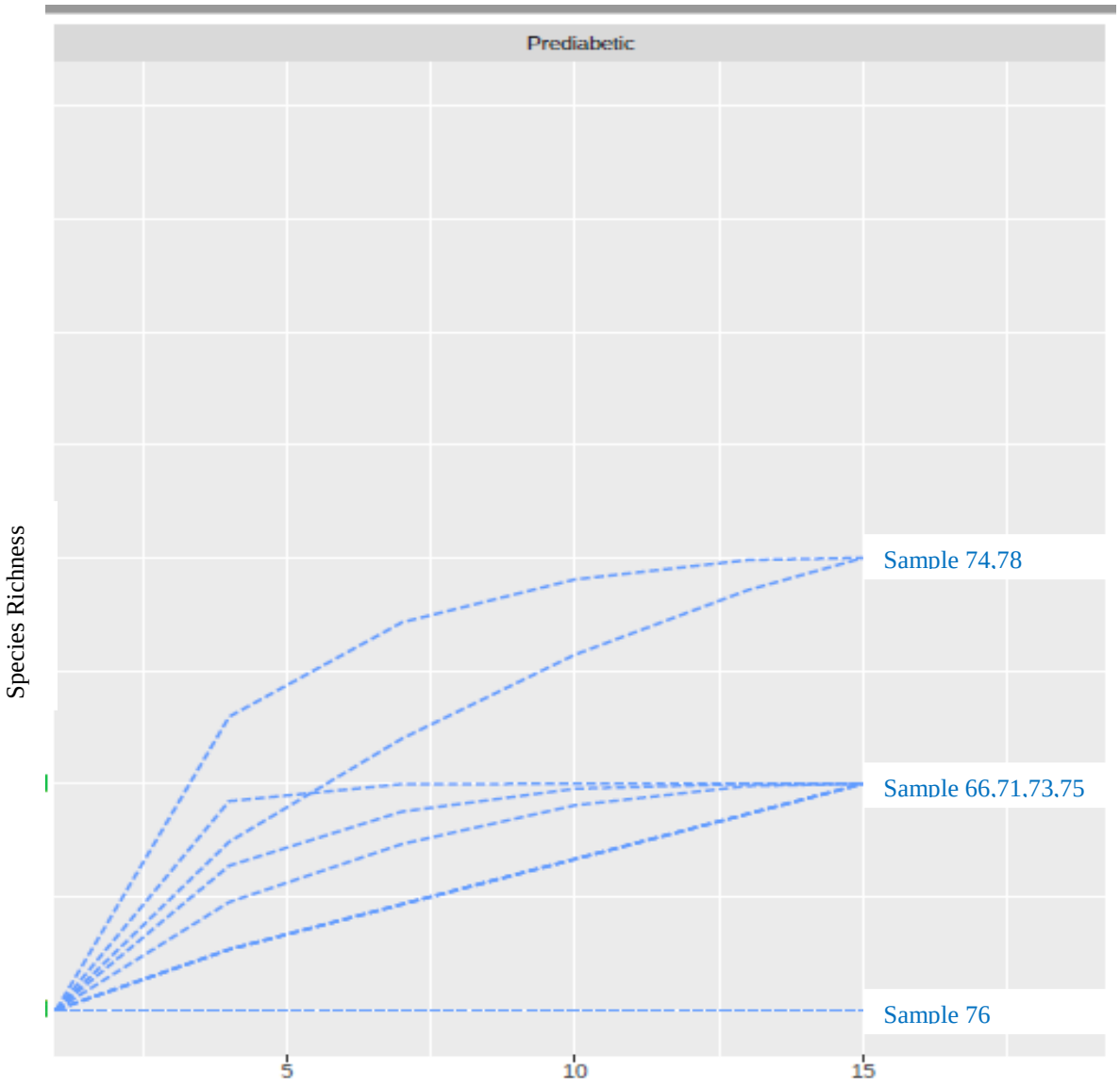


Figure 4.6: Rarefaction curve of pre-diabetic

Alpha Diversity was calculated using the Chao 1 index and significant differences ascertained using the non-parametric Kruskal Wallis/Mann Whitney test. The diabetic samples were significantly more diverse than non-diabetic and prediabetic samples

(Chao 1 -value: 0.019888; [Kruskal-Wallis] statistic: 7.8353) which is statistically significant.

Alpha diversity was confirmed by drawing the box plot figure 4.7 which also confirmed that diabetic sample was more diverse in both upper quartile and lower quartile whisker as compared to normal and pre-diabetic.

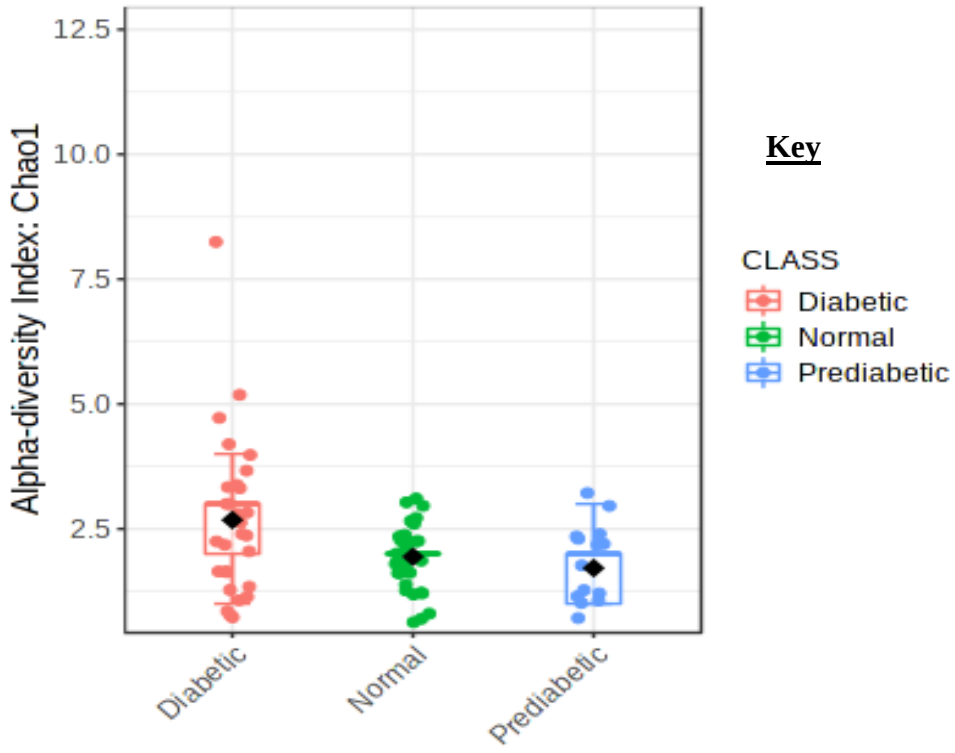
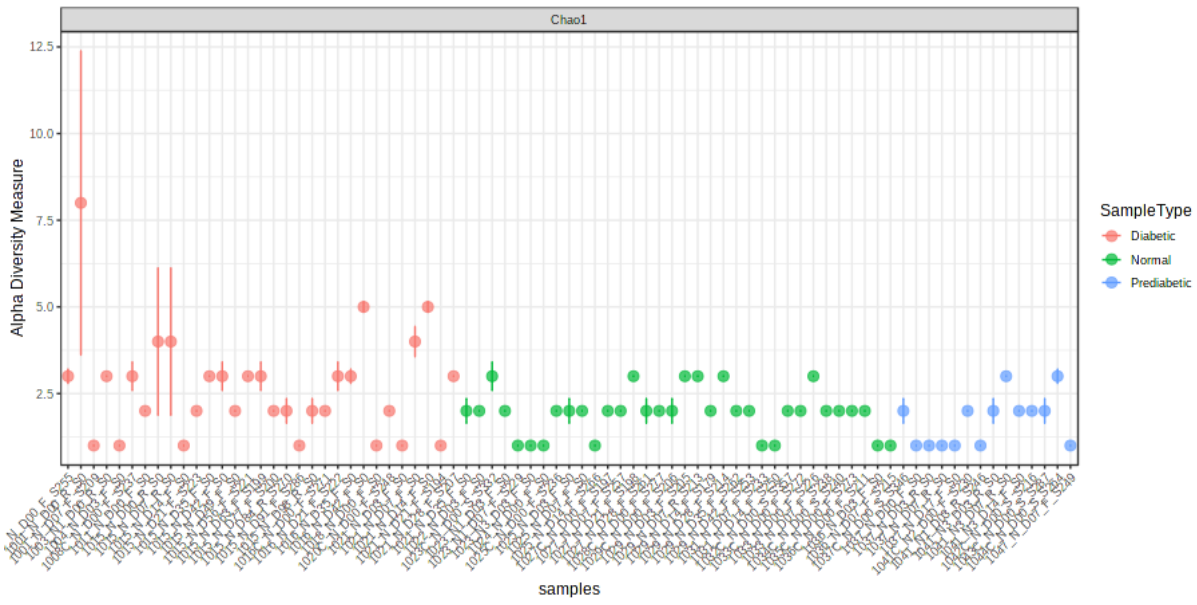


Figure 4.7: Box plot of diabetic sample Alpha Diversity



Vertical Axis

Horizontal Axis

Number of samples of diabetic, pre-diabetic and normal

Figure 4.8: 2 dimensional picture of Alpha Diversity plot.

2 dimensional picture of Alpha Diversity of three groups diabetic, pre-diabetic and normal, horizontal axis shows number of samples of each group and vertical axis shows diversity of Microbes, diabetics sample is more diverse in contrast to pre-diabetic and normal samples a shown in figure 4.8.

Another model of analysis is Beta diversity which tells us similarity or dissimilarity of microbes in OTUs between samples of diabetic, pre-diabetic and normal. The beta diversity was statistically significant [PERMANOVA] F-value: 4.1884; R-squared: 0.099279; p-value < 0.002). Beta diversity is depicted in 2 dimensional picture in figure 4.9 which shows Euclidean distance between samples of diabetic, pre-diabetic and normal. Diabetic samples are closer in different microbes of OTUs which is separated by distinct microbes

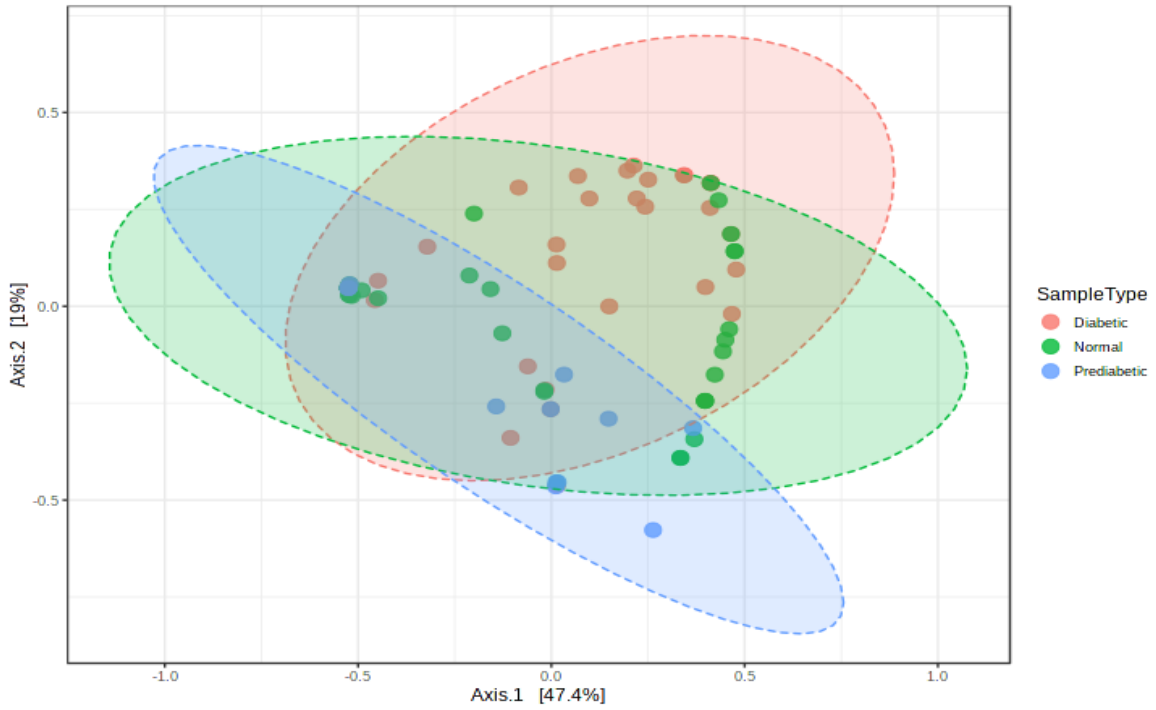


Figure 4.9: Beta diversity measured by Euclidean distance between samples of Diabetic, Pre-diabetic and Normal, 2 Dimensional picture of Beta diversity

Figure 4.10 which shows Principle Coordinate Analysis (PCOA) which is 3 dimensional picture of Beta diversity which separate similar microbes from each group.

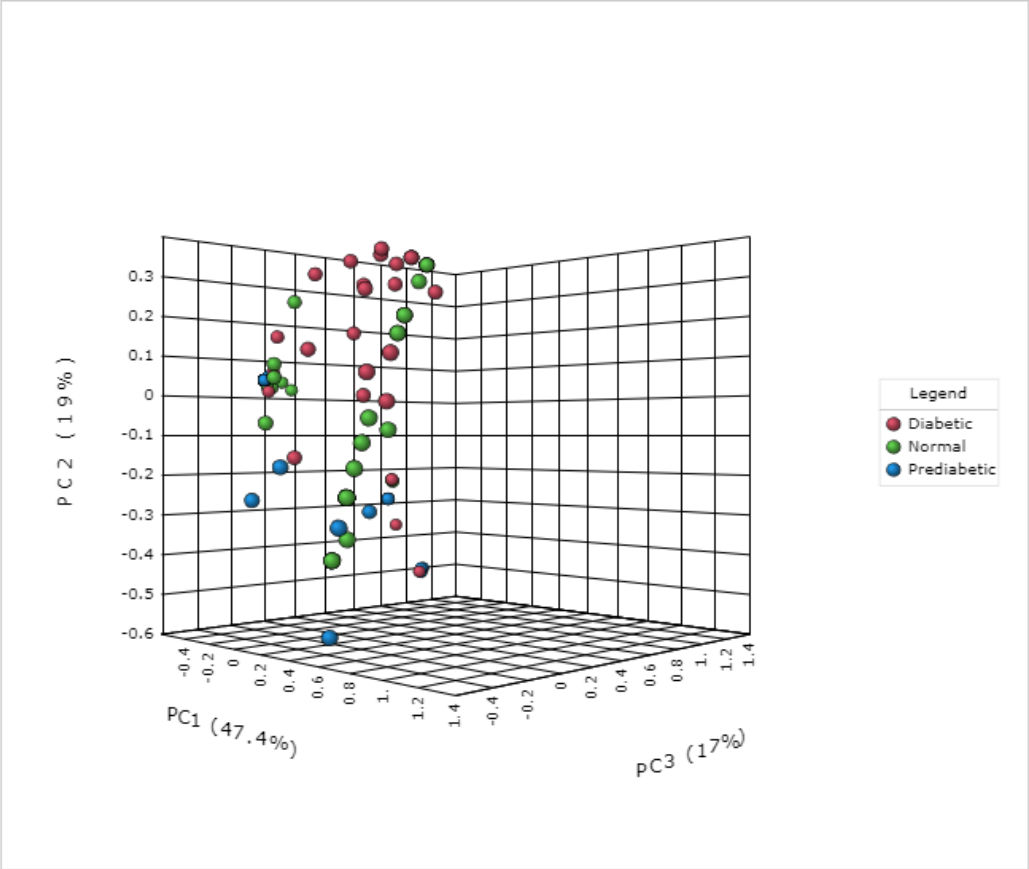


Figure 4.10: Beta diversity in 3 dimensional picture of three groups of samples.

Another model of microbial analysis is Linear Discriminative Analysis (LDA) and Linear Effective Size (LEFSE). This model of analysis give distinct OTU, which is OTU 000001 which contain statistically significant bacteria genera of Escherichia_Shigella which is a metagenomic marker in diabetic sample.

The LDA threshold for the score was 2. OTU000001 is significantly abundant in the diabetic group compared to the normal and pre-diabetic groups (p value =0.000588, FDR=0.004706). All other OTUs were not statistically significant we have left with 8 OTUs which are better p value than rest of OTUs All OTUs are depicted in Table 4.4 in which significant OTUs are arrange in descending order

Table 4.4: Lefse values testing significance for abundant OTUs

	P values	FDR	Statistics
Otu000001	0.000588	0.004706	14.877
Otu000004	0.089748	0.27498	4.8215
Otu000003	0.12563	0.27498	4.1488
Otu000007	0.13749	0.27498	3.9684
Otu000020	0.23729	0.37966	2.877
Otu000006	0.47369	0.63158	1.4944
Otu000002	0.78462	0.89671	0.48511
Otu000033	0.94045	0.94045	0.1228

Significant OTU000001 is Escherichia_Shigella which P value is less than 0.05 and box plot was drawn for OTU000001 which is depicted in figure 4.11. In diabetic box plot is symmetrical and median line is in center and distance of upper and lower whisker is higher compared to pre-diabetic and normal.

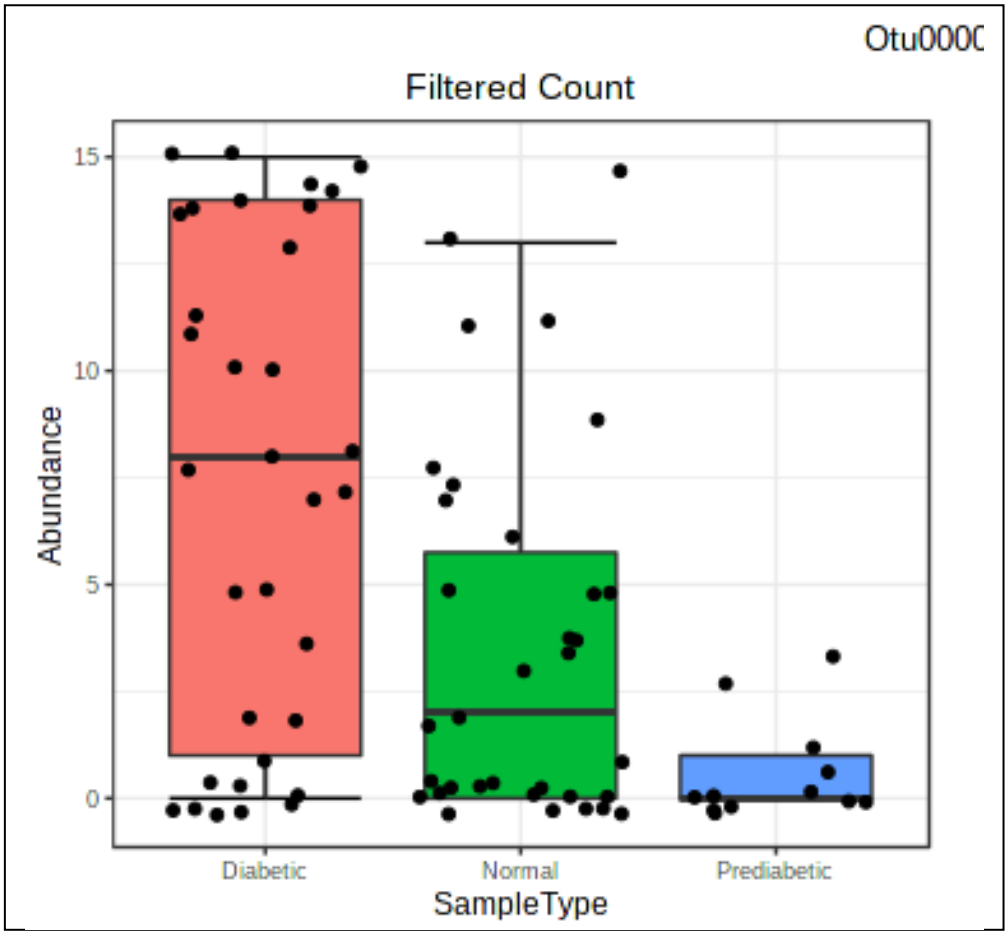


Figure 4.11: Box plot of OTU000001.

Another model of analysis is Random Forest. This model tells error in sample of diabetic, pre-diabetic and normal. Pre-diabetic group had relatively high error rate compared to other groups that is diabetic and normal. The normal group had the lowest error rate. This is depicted in figure 4.12.

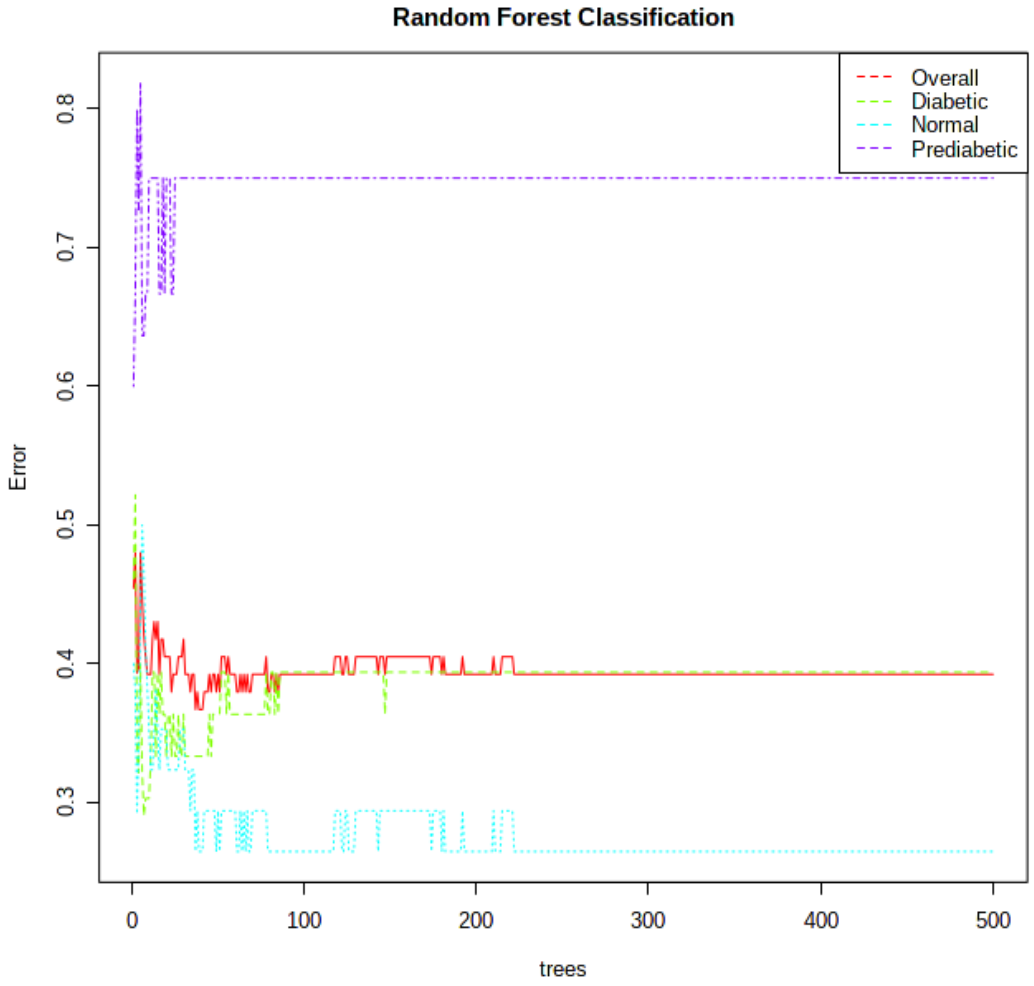


Figure 4.12: Random Forest Clasication

The robustness of Random Forest Classification was assessed using out of bag (OOB) error rate. Normal OOB rate is 0.392, OOB is depicted in Table 4.5. The class error for the pre-diabetic group was highest at 75% and lower at diabetic (39.4%) and Normal (26.5%)

Table 4.5: Out of bag (OOB) error.

	Diabetic	Normal	Pre-diabetic	Class.error
Diabetic	20	10	3.0	0.394
Normal	8	25	1.0	0.265
Pre-diabetic	1	8	3.0	0.75

Another model of analysis is Mean Decrease Accuracy (MDA)

The higher value of Mean Decrease Accuracy (MDA), the higher the importance in variable in model.

Important predictors of diabetes are OTU000001, OTU000006, OTU000003, and OTU000020. Important predictors of normal are OTU000004 and OTU000002. Important predictors of prediabetes are OTU000007. The Mean Decrease Accuracy (MDA) plot expresses how much accuracy the model losses by excluding each variable. Mean accuracy plot shown in figure 4.13 and table 4.5 shows important Predictors of the Diabetic, Pre-diabetic, and normal groups

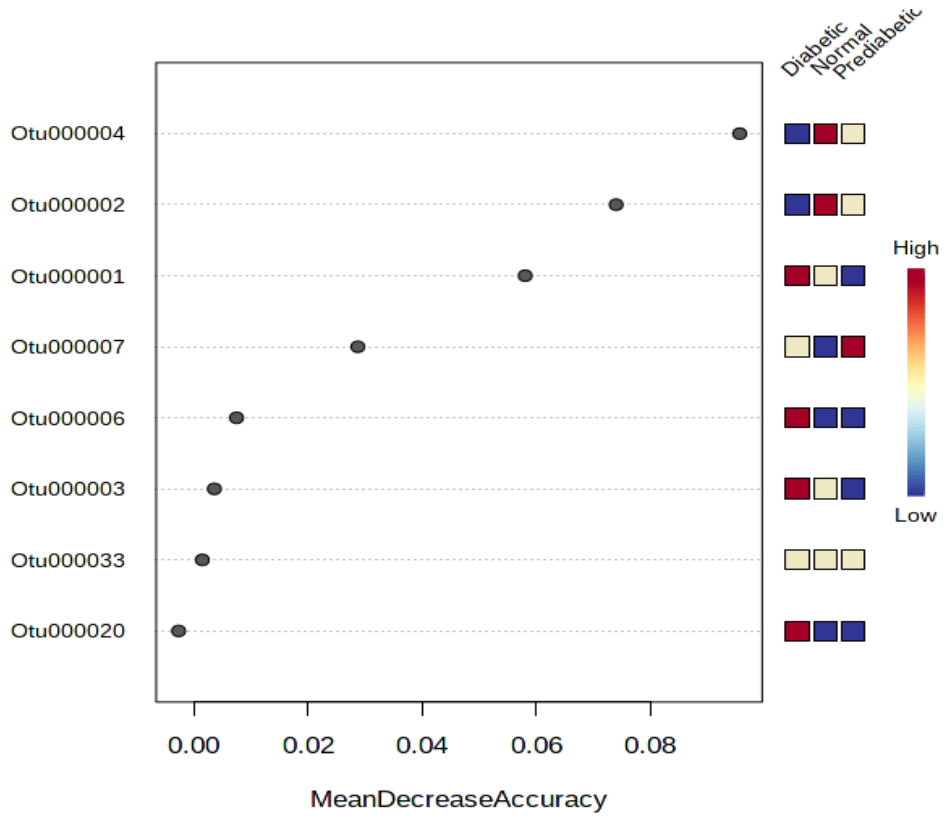


Figure 4.13: Mean Decrease Accuracy (MDA) plot

Figure 4.13 shows highest in the right side of bar is red and lowest point is blue , OTUs 1,OUT 6 ,OTU 3 , and OTU 20 starts with red that is highest point ,higher value of MDA, the higher the importance in variable in model that is also confirmed by p value and these OTUs are diabetic sample and Table shows the contents in the OTUs that is Microbiomes 0f phyla and genera Table 4.6 shows that

Table 4.6: Important Predictors of the Diabetic, Pre-diabetic, and normal groups

Predictor	Group	Phylum	Family	Genus
OTU000001	Diabetes	Proteobacteria	Enterobacteriaceae	Escherichia_Shigella
OTU000003	Diabetes	Proteobacteria	Enterobacteriaceae	Enterobacteriaceae_unclassified
OTU000006	Diabetes	Proteobacteria	Escherichia_Shigella	Kluyvera
OTU000020	Diabetes	Firmicutes	Staphylococcaceae	Staphylococcus
OTU000004	Normal	Proteobacteria	Enterobacteriaceae	Salmonella
OTU000002	Normal	Proteobacteria	Enterobacteriaceae	Klebsiella
OTU000007	Prediabetes	Firmicutes	Enterococcaceae	Enterococcus

Finding from microbial metagenomics analysis through MOTHUR Pipeline and different model of analysis answer the specific objective; 1. Types and abundance of bacteria colonizing the gut of diabetic, pre-diabetic and normal patients from automatic annotation of OTUs by drawing a stacked bar chart for phyla and genera of microbes like in diabetic proteobacteria, firmicutes and actinobacteria higher in diabetic contrast to pre-diabetic and normal. Although proteobacteria was also higher in normal. In genera of microbes in diabetic are Escherichia_Shigella, Enterobacteriaceae_unclassified and Staphylococcus which are low in pre-diabetic and normal.

Abundance of microbes in diabetic is confirmed by Alpha diversity and Beta diversity.

MDA modules confirm that predictor of diabetes OTU000001, OTU000006, OTU000003 and OTU000020 which have lowest P value. Predictor of Pre-diabetic is OTU000007 that is Enterococcus and predictor of Normal are OTU000002 and OTU000004 that is Salmonella and Klebsiella and specific objective two metagenomic marker of diabetes is OTU000001 which contain Escherichia Shigella which has very statistically significant value less than 0.05 and it is also confirmed by Box plot and MDA predictor.

CHAPTER FIVE

DISCUSSION

In this study, 16s rRNA sequencing was performed on fecal samples obtained from 33 diabetic, 33 Normal, and 13 prediabetic patients. The main objective was to profile the differential types and abundances of bacteria in each group and determine metagenomic markers of T2DM and prediabetes in patients. To determine the metagenomic markers of T2DM, we analysed the sequencing reads to determine the following four parameters within and between the groups: relative bacterial abundance, species diversity, significantly abundant genera, and important predictors.

5.1 Types and Abundance of Bacteria Colonizing the Gut of T2DM, Pre-Diabetic, and Non-Diabetic (Normal) Patients Visiting South C Health Centre

Relative bacterial abundance describes the percentages of specific bacteria making up the entire microbiome under study. The study findings show that the dominant phyla in the three groups under study were Proteobacteria, Firmicutes and Actinobacteria. Proteobacteria was most dominant in the normal group and least dominant in the prediabetic group. Firmicutes was more than 20 times more abundant in the diabetes group compared to the prediabetes group and nine times more abundant in the diabetes group compared to the normal group. Actinobacteria was also more abundant in the diabetes than in the normal and prediabetic group. The abundance of Bacteroidetes was low in all the 3 groups.

Proteobacterial species are known to be archetypal signatures of microbial dysbiosis (Shin et al., 2015).

Dysbiosis in the normal samples as proteobacteria is highest in contrast to diabetic and prediabetic samples, suggesting that such an interpretation should be used with caution. These unusual findings can be attributed to the use of patient samples. Unusual finding in normal samples could be due to mixed up samples or participant could be using

metformin medication for the diabetes and on data collection their sugar came under category of Normal sample. It could be also the participants of Normal sample had been suffering from some gastrointestinal disorder. Abundance of proteobacteria, firmicutes and actinobacteria in diabetic samples signified the dysbiosis in diabetic group. Diabetic is more diverse than normal and prediabetic groups.

High abundance of microbiota in diabetic group is confirmed by alpha diversity with a statically significant Chao 1 index and by using non-parametric Kruskal Wallis/Mann Whitney test. So high alpha density is also considered as metagenomic marker in diabetic group. Escherichia_Shigella, Enterobacteriaceae_unclassified and Staphylococcus are the Microbiota abundantly found in diabetic group. on the other hand, Salmonella and Kluyvera were abundantly in normal group. In the prediabetic, Bifidobacterium and Enterococcus were the dominant microbiota. This finding answered the study objectives 1 and 2

In our study it was found that high firmicutes and bacteroidetes ratios (F/B) which is considered as metagenomic marker in diabetes since firmicutes was higher in diabetis in contrast to normal and prediabetic groups and very low bacteroidetes. The same finding was reported in previous studies (Sedighi et al., 2017; Komaroff, 2017). High F/B ratio is a marker of elevated plasma glucose (Larsen et al., 2010).

5.2 T2DM Metagenomics Markers Based on the Identified Genera and Abundances of Bacteria

Elevated F/B ratio in diabetic samples was observed and may be used as a metagenomic marker of T2DM. Our finding reported high level of Firmicutes and very low level of Bacteroids so F/B ratio very high in diabetic samples. Another putative T2DM metagenomic marker was the high alpha diversity of T2DM - the diabetic samples are significantly more diverse than non-diabetic and prediabetic samples. High alpha diversity may be used as a metagenomic marker of T2DM. In our study high alpha

diversity was confirmed by Rarefaction curve which showed higher number of clusters (Nucleotides) of microbes in diabetic sample.

In diabetic group OTU000001 which has dominant genera of Escherichia_Shigella is metagenomic marker in our study which is statistically significant also confirmed by Box plot and other modules of analysis like linear discriminative analysis (LDA), linear effective size analysis (LEFSE) and mean decrease accuracy (MDA) which has important predictor OTU000001 in diabetes. This our observation of metagenomic markers in diabetes Escherichia_Shigella also support the finding by Maskarinec et al (2021) who recently reported elevated levels of E. shigella in patients with T2DM and associated this abundance with chronic systemic inflammation in T2DM disease (Mascarinec et al., 2021). The implication is that screening and treatment of Escherichia_Shigella infestation may perhaps slow down the low-grade inflammation hence assist in improving outcomes of T2DM disease.

Metagenomic marker in pre-diabetic group is OTU000007 that is Enterococcus and Bifidobacteria. This is confirmed by statistical analysis through Linear Discriminative Analysis (LDA) and Linear Effective Size (LEFSE). Although P value is 0.1379 in Normal group important OTUs is OTU000004 and OTU000002 that is microbial genus Salmonella and Kluyvera. This unusual finding in Normal sample could be due to participant may had some gastrointestinal problem or could be mixed up samples from pre-diabetic.

High level of Escherichia_Shigella in diabetes is responsible for upregulating of inflammatory pathway through genes of inflammatory that is NF-Kb transcription factor causing dysregulated immune response by releasing cytokines and chemokines through PAMPS (pathogen associated molecular pattern) ligating with PRR (Pattern Recognition Receptors) like Toll and Nod Receptors also Escherichia_Shigella causes opportunistic infection and organ damage (Grijalva, et al., 2015).

5.3 The correlation between metagenomic markers of T2DM, prediabetes and non-diabetic subjects and their clinical manifestations

The findings suggest that T2DM metagenomic markers that are reported in our studies can be correlated to clinical manifestation of T2DM in 2 ways. First, Escherichia_Shigella is known to produce pro-inflammatory endotoxins. A constant feature of T2DM disease is constant low-grade inflammation. The proinflammatory endotoxins associated with high abundance of Escherichia_Shigella in T2DM patients may be partly responsible for the chronic systemic inflammation in these patients. This observation seems to support findings by Maskarinec et al (2021) who recently reported elevated levels of E. shigella in patients with T2DM and associated this abundance with chronic systemic inflammation in T2DM disease (Mascarinec et al., 2021). The implication is that screening and treatment of Escherichia_Shigella infestation may perhaps slow down the low-grade inflammation hence assist in improving outcomes of T2DM disease.

In the second instance, we observed high levels of opportunistic pathogens such as Escherichia_Shigella, Enterobacteriaceae_unclassified and Staphylococcus in the T2DM group. Many studies have previously shown that people with T2DM are more susceptible to infections (Grijalva, et al., 2015) compared to normal persons. The clinical implication is that screening of the gut microbiome may be a useful strategy in monitoring and managing bacterial infections in people with T2DM.

5.4 Limitations of the study

- Variables and confounders like age, body mass index (BMI), blood pressure, medication and diet are not controlled in this study. These variables affect normal gut microbiota and causes imbalance in proportion, diversity and density of microbes and the end result is through PAMPS and DAMS which ligats with nods and tolls like receptors and causes dysregulation of immune response that is cytokine and chemokines which causes upregulation of transcription factors

NFKV and cuases chronic inflammation that down regulates the insulin receptors in cells and causes propensity of T2DM and organ damage

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This study sought out to find the metagenomic markers of diabetes in patients attending South C Medical Centre in Niarobi and also type and abundance of microbes colonizing the gut of diabeteic, pre-diabetic and Normal patients visiting South C Medical Centre Nairobi using 16s rRNA sequencing, 79 samples were investigated from diabetic, non-diabetic, and prediabetic patients. This study found out that:

- High abundance of microbiota in diabetic group and types of microbiota which are confirmed in our finding that microbiota in diabetic are Escherichia_Shigella, Firmicutes, Actinobacteria and Proteobacteria which answers our objectives 1 and 2. Therefore, Putative T2DM Metagenomic markers include

    1. High levels of Escherichia_Shigella
    2. Elevated levels of Firmicutes and Actinobacteria
    3. High FB Ratio
    4. Significantly high alpha diversity

High levels of Escherichia_Shigella in the T2DM patients contributes to progression of T2DM disease through dysregulated immune response through cytokines and chemokines by ligating PAMPS with PRR(pattern recognition receptors) causing inflammatory genes NF-Kb and increased catabolism cells and organs damage (Text book of medicine by Harrisons and.Molecular Biology by Lauren Pecorino)

High levels of opportunistic pathogens such as Escherichia_Shigella, and Kluyvera in the diabetic group contributes to susceptibility of infections in diabetes.(Grijlva,et al., 2015)

6.2 Recommendations and Future Prospects

1. From this study it has been established our diabetic sample is diversed and densed , and metagenomic marker of diabetes is Escherichia_Shigella, it is recommended that medical doctors and Diabotologist put consideration in their management of diabetes, metagenomic marker that is Escherichia_Shigella in treatment to avoid further progression and mitigation of complication.
2. From this study it is also recommended that consideration for fecal implantation on T2DM be made from Normal gut flora to abate and mitigate the complication from T2DM.
3. Future studies should be conducted to validate the findings of this study considering the variables like age, BMI, blood pressure, medication and diet which also affect the gut microbial normal flora and causes dysbiosis and increase propensity of diabetis.

REFERENCES

(ADA)., A. D. (2009). Diagnosis and Classification of Diabetes Mellitus. Diabetes Care., 32(1):S62-7.

(AHA)., A. H. (2010). Heart disease and stroke statistics 2010 update: a report from the American Heart Association. Circulation, 121(7):e46-e215.

Abdulfatai B. Olokoba, O. A. (2012). Type 2 Diabetes Mellitus: A Review of Current Trends. Oman Medcial Journal, 269-273.

ADA. (2014). Diagnosis and classification of diabetes mellitus. Diabetes Care, 42(Suppl. 1):S13–S28.

AHA. (2010). Heart disease and stroke statistics - 2010 update: a report from the American Heart Association. Circulation., 121(7):e46-e215.

Ayah, R., Joshi, M., Wanjiru, R., Njau, E., Otieno, C., & Njeru., E. (2013). A population-based survey of prevalence of diabetes and correlates in an urban slum community in Nairobi, Kenya. BMC Public Health., 13:371.

Bahadursingh, S., Mungalsingh, C., Seemungal, T., & Teelucksingh, S. (2014). Acanthosis nigricans in type 2 diabetes: prevalence, correlates and potential as a simple clinical screening tool - a cross-sectional study in the Caribbean. Diabetol Metab Syndr, 6: 77.

Bansal, D., Gudala, K., Muthyala, H., Esam, H., Nayakallu, R., & Bhansali., A. (2014). Prevalence and risk factors of development of peripheral diabetic neuropathy in Type 2 diabetes mellitus in a tertiary care setting. J Diabetes Investig., 5(6):714-21.

Barbalat, R., Lau, L., Locksley, R., & Barton, G. (2009). Toll-like receptor 2 on inflammatory monocytes induces type I interferon in response to viral but not bacterial ligands. Nat Immunol., 10:1200–7.

Blaut M. Ecology and physiology of the intestinal tract. Curr Top Microbiol Immunol. 2013;358:247–72. https://doi.org/10.1007/82_2011_192.

Castilla, J., Godoy, P., Domínguez, A., & etal. (2013). Influenza vaccine effectiveness in preventing outpatient, inpatient, and severe cases of laboratory-confirmed influenza. . Clin Infect Dis. , 57: 167-175.

Chaturvedi., N. (2007). The burden of diabetes and its complications: Trends and implications for intervention. . Diabetes Res Clin Pract. , 76(3): S3–S12.

Chong, J., Liu, P., Zhou, G., and Xia. J. (2020) "Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data" Nature Protocols (DOI: 10.1038/s41596-019-0264-1)

Coenye, T., & Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. FEMS Microbiology Letters., 228 (1): 45–9.

Cook, S., & Sellin, J. (1998). Review article: short chain fatty acids in health and disease. . Aliment Pharmacol Ther. , 12(6):499-507.

Crevel, R., Vijver, S., & Moore, D. (2017). The global diabetes epidemic: what does it mean for infectious diseases in tropical countries. The Lancet, 5(6): 457-468.

Cunningham, A.L., Stephens, J.W. & Harris, D.A. Gut microbiota influence in type 2 diabetes mellitus (T2DM). Gut Pathog 13, 50 (2021). https://doi.org/10.1186/s13099-021-00446-0.

Cuscó, A., Catozzi, C., Viñes, J., Sanchez, A., & Francino, O. (2019). Microbiota profiling with long amplicons using Nanopore sequencing: full-length 16S rRNA gene and the 16S-ITS-23S of the rrn operon [version 2; peer review: 2 approved, 3 approved with reservations]. F1000Research.

Dagdeviren, S., Jung, D., Friedline, R., Noh, H., Kim, J., Patel, P., . . . Kim, J. (2017). IL-10 prevents aging-associated inflammation and insulin resistance in skeletal muscle. FASEB J., 31(2):701-710.

Dam, R. v., Willett, W., & Rimm, E. (2002). Dietary fat and meat intake in relation to risk of 2 diabetes in men. Diabetes care., 25(3):417–424. .

Daneman, D. (2006). Type 1 diabetes. . Lancet., 367:847–858.

Dang, F., Jiang, Y., Ruili, P., Zhou, Y., Wu, S., Wang, R., . . . Man, C. (2018). Administration of Lactobacillus paracasei ameliorates type 2 diabetes in mice. Food Funct., 17;9(7):3630-3639.

Darwish., A. K. (2015). Diabetes mellitus: The epidemic of the century. WorldJ Diabetes., 6(6): 850–867.

Davie, J. (2003). Inhibition of histone deacetylase activity by butyrate. J Nutr. , 133:2485S−93S.

den Besten, G. (2015). Short-Chain fatty acids protect against high-fat diet-induced obesity via a PPARgamma-Dependent switch from lipogenesis to fat oxidation. . Diabetes. , 64(7):2398–2408.

Dhariwal, A., Chong, J., Habib, S., King, I., Agellon, LB., and Xia. J. (2017) "MicrobiomeAnalyst - a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data". Nucleic Acids Research 45 W180-188 (DOI:

10.1093/nar/gkx295) for 16S rRNA gene sequence analysis. Appl Environ Microbiol 77:3219. http://www.mothur.org/wiki/Cluster.split

Dhingra, R., Sullivan, L., & PFJacques. (2007). Soft drink consumption and risk of developing cardio-metabolic risk factors and the metabolic syndrome in middle-aged adults in the community. Circulation. , 116:480–488.

Donadon, V., Balbi, M., & Casarin, P. (2008). Association between hepatocellular carcinoma and type 2 diabetes mellitus in Italy: potential role of insulin. . World J Gastroenterol., 14(37): 5695–5700.

Duffy, A., & Wen., L. (2017). Factors Influencing the Gut Microbiota, Inflammation, and Type 2 Diabetes. . J Nutr., 147(7): 1468S–1475S.

Dunkelberger, J., & Song, W. (2010). Complement and its role in innate and adaptive immune responses. Cell Res , 20, 34–50.

Everard, A. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. . Proc Natl Acad Sci U S A., 110(22):9066–9071. .

FH Karlsson, V. T. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature, 498:99–103.

Gambineri, A., Patton', L., Altieri, P., Pagotto, U., Pizzi, C., Manzoli, L., & Pasquali, R. (2012). Polycystic Ovary Syndrome Is a Risk Factor for Type 2 Diabetes. Results from a Long-Term Prospective Study. . Diabetes, 61(9): 2369–2374.

Gao, Z. (2009). Butyrate improves insulin sensitivity and increases energy expenditure in mice. Diabetes, 58(7):1509–1517.

Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J, Westram R, Ludwig W (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. Opens external link in new windowJ. Biotechnol.

Gopalakrishnan, V., Helmink, B. A., Spencer, C. N., Reuben, A., & Wargo, J. A. (2018). The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy. Cancer Cell, 33(4): 570-580.

Greg, J. (. (2018). Universal Bacterial Identification by PCR and DNA Sequencing of 16S rRNA Gene. In M. Schuller, T. P. Sloots, G. S. James, C. L. Halliday, & I. W. Carter, PCR for Clinical Microbiology (pp. 209–214). Dordrecht: Springer.

Grijalva, C., Zhu, Y., Williams, D., & al, e. (2015). Association between hospitalization with community-acquired laboratory-confirmed influenza pneumonia and prior receipt of influenza vaccination. JAMA., 314: 1488-1497.

Gurung, M., Li, Z., You, H., Rodrigues, R., Jump, D. B., Morgun, A., & Shulzhenko, N. (2020). Role of gut microbiota in type 2 diabetes pathophysiology. EBioMedicine, 51: 102590.

H Wang 2019 et al. Fecal microbiota transplantation

Hoffmann, T. (2016). Microorganisms linked to inflammatory bowel disease-associated dysbiosis differentially impact host physiology in gnotobiotic mice. . ISME J. , 10(2):460–477.

Houmard, J. (2008). Intramuscular lipid oxidation and obesity. . Am J Physiol Regul Integr Comp Physiol., 294(4):R1111–R1116.

IDF. (2013). IDF Diabetes Atlas. Brussels, Belgium:: International Diabetes Federation.

Jay, Z. J., & Inskeep, W. P. (2015). The distribution, diversity, and importance of 16S rRNA gene introns in the order Thermoproteales. Biology direct, 10-35.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., & Mitchel, T. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. Frontiers in Microbiology. , 7: 459.

Kahn, S., Cooper, M., & Prato., S. D. (2013). Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present and future. Lancet, 383(9922): 1068–1083.

Kang, J. (2013). Anti-obesity effect of lactobacillus gasseri BNR17 in high-sucrose diet-induced obese mice. . PLoS ONE., 8(1):e54617. .

Karlsson, F., Tremaroli, V., Nookaew, I. et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature 498, 99–103 (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control - Nature

Kim, S., Huh, C., Choi, I., Jeong, J., Ku, H., Ra, J., . . . YT., A. (2014). The anti-diabetic activity of Bifidobacterium lactis HY8101 in vitro and in vivo. J Appl Microbiol., 117(3):834-45.

Knowler, W., Barrett-Connor, E., Fowler, S., Hamman, R., Lachin, J., Walker, E., & Nathan., D. (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. Diabetes Prevention Program Research Group. N Engl J Med., 346(6):393-403.

Kobberling, J. (1982). Empirical risk figures of first degree relatives of non-insulin dependent diabetes. London: Academic Press.

Komaroff AL. The microbiome and risk for obesity and diabetes. JAMA. 2017;317(4):355–6. https://doi.org/10.1001/jama.2016.20099.

Klyosova, Elen et al. (2022). Polymorphism in the gene coding heat stock factor (HSF).

Lambeth SM, Carson T, Lowe J, et al. Composition, diversity and abundance of gut microbiome in prediabetes and type 2 diabetes. J Diabetes Obes. 2015;2(3):1–7. https://doi.org/10.15436/2376-0949.15.031.

Larsen N, Vogensen FK, van den Berg FW, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS ONE. 2010. https://doi.org/10.1371/journal.pone.0009085.

Larsen, N., Vogensen, F., & Berg., F. v. (2010). Gut microbiota in human adults with type 2 diabetes differs from nondiabetic adults. PLoS One, 5:e9085.

Larsson, S., & Wolk, A. (2011). Diabetes mellitus and incidence of kidney cancer: a meta-analysis of cohort studies. . Diabetologia., 54: 1013–1018.

Larsson, S., Andersson, S., & Johansson, J. (2008). Diabetes mellitus, body size, and bladder cancer risk in a prospective study of Swedish men. Eur J Cancer., 44: 2655–2660.

Larsson, S., Mantzoros, C., & Wolk, A. (2007). Diabetes mellitus and risk of breast cancer: a meta-analysis. Int J Cancer., 121: 856–862.

Le, T. (2015). Bifidobacterium species lower serum glucose, increase expressions of insulin signaling proteins, and improve adipokine profile in diabetic mice. . Biomed Res. , 36(1):63–70.

Lee et al, 2018. Fecal microbiota transplantation is a method to treat T2DM.

Li, G., Zhang, P., Wang, J., Gregg, E., Yang, W., Gong, Q., . . . Bennett, P. (2008). The long- term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. . Lancet, 371(9626):1783-9.

Liu, S., Manson, J., Stampfer, M., Hu, F., Giovannucci, E., Colditz, G., . . . Willett., W. (2000). A prospective study of whole-grain intake and risk of type 2 diabetes mellitus in US women. Am J Public Health, 90(9):1409-15.

López-García, A., Pineda-Quiroga, C., Atxaerandio, R., Pérez, A., Hernández, I., García-Rodríguez, A., & González-Recio, O. (2018). Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. Front Microbiol. , 9: 3010.

Maahs, D., West, N., Lawrence, J., & Mayer-Davis., E. (2010). Epidemiology of type 1 diabetes. . Endocrinol Metab Clin North Am., 9:481–497.

Maskarinec G, Raquinio P, Kristal BS, Setiawan VW, Wilkens LR, Franke AA, Lim U, Le Marchand L, Randolph TW, Lampe JW, Hullar MAJ. The gut microbiome and type 2 diabetes status in the Multiethnic Cohort. PLoS One. 2021 Jun 23;16(6):e0250855. doi: 10.1371/journal.pone.0250855. PMID: 34161346; PMCID: PMC8221508.

Mathur, R., & Barlow, G. (2015). Obesity and the microbiome. . Expert Rev Gastroenterol Hepatol., 9(8):1087-99.

Metzger, B., Lowe, L., Dyer, A., Trimble, E., Chaovarindr, U., Coustan, D., . . . Sacks., D. (2008). Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, Hadden DR, McCance DR, Hod M, McIntyre HDHyperglycemia and adverse pregnancy outcomes. N Engl J Med., 358(19):1991-2002.

MOH. (2010). Kenya National Diabetes Strategy. First Edition. Retrieved from Ministry of public health and sanitation, Republic of Kenya: http://diabetes-communication.org/wordpress/wp-content/uploads/2012/09/Kenya-National-Diabetes-Strategy-2010-2015-Complete.pdf.

Musso, G., Gambino, R., & Cassader, M. (2011). Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annu Rev Med., 62():361-80.

Overall, J., Tonidandel, S., & Starbuck, R. (2006). Rule-of-thumb adjustment of sample sizes to accommodate dropouts in a two-stage analysis of repeated measurements. . Int J Methods Psychiatr Res., 15(1):1-11.

Pamidi, S., & Tasali., E. (2012). Obstructive sleep apnea and type 2 diabetes: is there a link? Front Neurol., 3():126.

Parada, V. D., De la Fuente, M., Landskron, G., González, M., Quera, R., Dijkstra, G., . . . Hermoso, M. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. Front Immunol., 28;10:1486.

Patel, J. B. (2001). 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. . Mol. Diagn. , 6:313–321.

Plovier, H. (2017). A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. Nat Med. , 23(1):107–113.

Qin, J., Li, Y., & Cai., Z. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature , 490:55–60.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Opens external link in new windowNucl. Acids Res. 41 (D1): D590-D596.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

RM, R. v., Willett, W., & Rimm, E. (2002). Dietary fat and meat intake in relation to risk of 2 diabetes in men. Diabetes care. Diabetes care., 25(3):417–424. .

Romzova, M. et al (2006). NF-KB protein encoding gene in relation ofT2DM

Sanghera, D., & Blackett., P. (2012). Type 2 diabetes genetics: beyond GWAS. J Diabetes Metab., 3(198): pii6948.

Schloss PD, Westcott SL (2011). Assessing and improving methods used in OTU-based approaches.

Schloss, P. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. . Appl Environ Microbiol, 75(23):7537-41.

Schulze, M., Manson, J., & Ludwig, D. (2004). Sugar-sweetened beverages, weight gain, and incidence of type II diabetes in young and middle-aged women. JAMA. , 292: 927–934.

Schwiertz A, Taras D, Schafer K, Beijer S, Bos NA, et al. (2009) Microbiota and SCFA in lean and overweight healthy subjects. Obesity. Advance online publication.

Sedighi M, Razavi S, Navab-Moghadam F, et al. Comparison of gut microbiota in adult patients with type 2 diabetes and healthy individuals. Microb Pathog. 2017;111:362–9. https://doi.org/10.1016/j.micpath.2017.08.038.

Shin NR, Whon TW, Bae JW. Proteobacteria: microbial signature of dysbiosis in gut microbiota. Trends Biotechnol. 2015 Sep;33(9):496-503. doi: 10.1016/j.tibtech.2015.06.011. Epub 2015 Jul 22. PMID: 26210164.

Singh, S. (2017). Lactobacillus rhamnosus NCDC17 ameliorates type-2 diabetes by improving gut function, oxidative stress and inflammation in high-fat-diet fed and streptozotocintreated rats. . Benef Microbes. , 8(2):243–255.

Topping, D., & Clifton, P. (2001). Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. . Physiol Rev., 81:1031–64.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804–10.

Uehara, K., Yasunaga, H., Morizaki, Y., Horiguchi, H., Fushimi, K., & Tanaka, S. (2014). Necrotising soft-tissue infections of the upper limb: risk factors for amputation and death. Bone Joint J. , 96-B: 1530-1534.

Vallianou, N., Stratigou, T., & Tsagarakis., S. (2019). Metformin and gut microbiota: their interactions and their impact on diabetes. . Hormones (Athens)., 18(2):141-144.

Vaziri, K., Pershing, S., Albini, T., Moshfeghi, D., & Moshfeghi., A. (2015). Risk factors predictive of endogenous endophthalmitis among hospitalized patients with hematogenous infections in the United States. . Am J Ophthalmol. , 159: 498-504. .

WHO. (2006). Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. Retrieved from World Health Organization: https://www.who.int/news-room/fact-sheets/detail/diabetes

WHO. (2019). Classification Of Diabetes Mellitus. Geneva: World Health Organization.

WHO. (2020, June 8). Diabetes. Retrieved from World Health Organization: https://www.who.int/news-room/fact-sheets/detail/diabetes

WinSCP. (2022). Free Award-Winning File Manager. https://winscp.net/eng/index.php

Wu, Y., Ding, Y., Tanaka, Y., & Zhang, W. (2014). Risk Factors Contributing to Type 2 Diabetes and Recent Advances in the Treatment and Prevention. Int J Med Sci., 11(11): 1185–1200. .

Yang, J. (2017). Gut commensal bacteroides acidifaciens prevents obesity and improves insulin sensitivity in mice. . Mucosal Immunol., 10(1):104–116. .

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Opens external link in new windowNature Reviews Microbiology volume 12, pages 635–645

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Opens external link in new windowNucl. Acids Res. 42:D643-D648.

Zatalia, S., & Sanusi., H. (2013). The role of antioxidants in the pathophysiology, complications, and management of diabetes mellitus. . Acta Med Indones., 45(2):141-7.

Zhang, X., Shen, D., & Fang., Z. (2013). Human gut microbiota changes reveal the progression of glucose intolerance. . PLoS One , 8:e71108.

Zhang, Y., Guo, X., Guo, J., He, Q., Li, H., Song, Y., & Zhang, H. (2014). Lactobacillus casei reduces susceptibility to type 2 diabetes via microbiota-mediated body chloride ion influx. Scientific Reports, (4):5654.

Zhao L, Lou H, Peng Y, Chen S, Zhang Y, Li X. Comprehensive relationships between gut microbiome and faecal metabolome in individuals with type 2 diabetes and its complications. Endocrine. 2019;66(3):526–37.

Zheng, Y., Ley, S., & FB., H. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its Complications. Nat Rev Endocrinol., 14(2):88-98.

Zhu, C. (2018). Roseburia intestinalis inhibits interleukin17 excretion and promotes regulatory T cells differentiation in colitis. . Mol Med Rep. , 17(6):7567–7574.

APPENDICES

Appendix I: Informed Consent Explanation

PROJECT TITLE: MICROBIAL METAGENOMIC MARKERS OF TYPE 2 DIABETES MELLITUS (T2DM) IN KENYA PATIENTS

PRINCIPAL INVESTIGATOR: BINOD KUMAR

INTRODUCTION

The study is to be carried out for metabolic Disease, adult-onset type 2 Diabetes Mellitus, in order to establish how some gut micro biota act as catalyst in exacerbating unabated progression of type 2 Diabetes Mellitus irreversibly and how some micro biota play protective role in cessation of progression of type 2 Diabetes Mellitus. The study will help us to find out early arrest of T2DM which when afflicted, person has a lifelong medication and its metabolic complication in the body leads to debilitating and life-threatening vital organ damage. In order to establish which bacteria causes this, we need to take stool sample from three groups of persons; persons with diabetes, persons with prediabetes and person who don't have diabetes. If you accept to take part in the study, then we will take permission from you and take stool sample. Please take time to read this information sheet about the study, and when you have read, feel free to ask question or to seek clarification on any issue related to this study or participation in it, or both.

PURPOSE AND BENEFITS

The purpose of the study is to see which bacteria in Diabetes aggravate the condition and in prediabetes progress unabated to full grown diabetes and the bacteria which has protective effects in cessation of progression. This study can latter serve to provide an alternative care. There will be free consultation with the Doctor and necessary assistance will be provided. The information obtained would contribute to overall improvement of

health status of diabetic patients and the participating volunteers are at forefront of any benefits achieved.

PROCEDURE TO BE USED

Participation in this study is voluntary. You will be required to understand the procedures involved. This will be made clear to you by the recruiting personnel. If you agree and you are comfortable with the procedure involved after having understood the work then you will be required to sign the consent form. Upon this agreement, you will be required to give your stool sample for study. This is purely voluntary and you can opt out even after signing in case you won't feel comfortable.

RISKS, HAZARDS AND DISCOMFORTS ASSOCIATED WITH THE PROCEDURES

Giving stool sample will not cause any harm to you and is less intrusive as the whole procedure will be performed at home. For random blood sugar test, only tiny drop of blood obtained by puncturing the fingertip is not harmful. This is usually less than 0.5mls of blood.

CONFIDENTIALITY

Your identity and test results will remain confidential. As a study participant, you will be assigned a number code, and yourself and the test results that will be carried out on the samples obtained from you, will remain confidential. All information and medical records will remain confidential, and will remain in lockable cabinet and will only be accessible to the people carrying out this study.

CONTACT OF THE SITE PRINCIPAL INVESTIGATOR: If you need more information about the study, please call: BINOD KUMAR _Cell phone: 0721510216

CONTACT OF JOMO KENYATA UNIVERSITYOF AGRICULTURE AND TECHNOLOGY ETHICS REVIEW COMMITTEE: If you have questions about your rights as a research participant, please contact: JKUAT ETHICAL COMMITTEE SECRETARY Telephone +254725996171, Email: ethics@jkuat.ac.ke

INFORMED CONSENT AGREEMENT FOR PARTICIPANTS

I, Mr./Mrs./Miss_____, being an adult aged 18 years and above do hereby give permission to Binod Kumar for new study entitled MICROBIAL METAGENOMIC MARKERS OF TYPE 2 DIABETES MELLITUS (T2DM) IN KENYA PATIENTS.I have been explained to about the test to be done on my stool, I have been given opportunity to ask question and to seek clarification of the issues I had not understood clearly and I am satisfied with the answer and the explanation given. I have also been informed that I if have additional question or concerns about the study later, I can contact the researcher in charge of the study or the Ethics Review Committee at Jomo Kenyatta University of Agriculture and Technology

I accept and I can provide stool sample for the test needed in this study.

_____ _____

Signature (or Thumb Print) of Participant Date

Witnessed by:

_____

Name of PI or study coordinating.

Appendix II: Kiswahili Version of Informed Consent

KICHWA CHA UTAFITI: MICROBIAL METAGENOMIC MARKERS YA TYPE 2 DIABETES MELITUS (T2DM) KWA WAGONJWA WA KENYA

MCHUNGUZI MKUU: BINOD KUMAR

UTANGULIZI: Utafiti huo unafanywa kwa Ugonjwa wa kimetaboliki, ugonjwa wa kisukari wa watu wazima aina ya 2, ili kujua jinsi baadhi ya utumbo wa biota hufanya kama kichocheo katika kuzidisha maendeleo yasiyopunguzwa ya aina ya 2 ya ugonjwa wa kisukari bila kubadilika na jinsi biota ndogo ndogo hucheza jukumu la kinga kukomesha maendeleo ya aina 2 ya ugonjwa wa kisukari

Utafiti huo utatusaidia kujua dalili za mapema za T2DM ambayo wakati mtu anaugua, ana dawa ya maisha na shida yake ya kimetaboliki mwilini husababisha kudhoofika na kutishia maisha kwa uharibifu muhimu wa viungo.

Ili kujua ni bakteria gani husababisha hii, tunahitaji kuchukua sampuli ya kinyesi kutoka kwa vikundi vitatu vya watu; watu wenye ugonjwa wa sukari, watu walio na prediabetes na watu ambao hawana ugonjwa wa kisukari.

Ikiwa unakubali kushiriki katika utafiti, basi tutachukua ruhusa kutoka kwako na kuchukua sampuli ya kinyesi. Tafadhali chukua muda kusoma karatasi hii ya habari juu ya utafiti, na wakati umesoma, jisikie huru kuuliza swali au kutafuta ufafanuzi juu ya suala lolote linalohusiana na utafiti huu .

KUSUDI LA UTAFITI: Kusudi la utafiti ni kuona ni bakteria gani katika ugonjwa wa sukari wanaongeza hali hiyo na katika ugonjwa wa prediabetes nini hufanya haswa kuendelea bila kukoma na bakteria ambayo ina athari za kinga katika kukomesha maendeleo.

UTARATIBU WA KUTUMIWA: unaweza kujumuishwa katika utafiti huu ikiwa utasaini fomu ya idhini inayopeana idhini ya kushiriki katika utafiti na kutoa ruhusa ya kutoa sampuli ya kinyesi chako kwa masomo.

HATARI, HATARI NA HASARA HUSHIRIKIANA NA TARATIBU: Kutoa sampuli ya kinyesi hakutasababisha madhara yoyote kwako.

USIRI: Utambulisho wako na matokeo ya mtihani yatabaki kuwa siri. Kama mshiriki wa utafiti, utapewa nambari, na wewe mwenyewe na matokeo ya mtihani ambayo yatafanywa kwenye sampuli zilizopatikana kutoka kwako, zitabaki kuwa siri. Rekodi zote za habari na matibabu zitabaki kuwa za siri, na zitabaki kwenye wodrobu zinzoweza kufungwa na zitapatikana tu kwa watu wanaofanya utafiti huu.

MAWASILIANO YA MCHUNGUZI WAKUU WA SEKTA: Ikiwa unahitaji habari zaidi kuhusu utafiti huu, tafadhali piga simu: BINOD KUMAR _Cell simu: 0721510216

MAWASILIANO YA KAMATI YA MAPITIO YA MAADILI YA KITETE KENYATTA: Ikiwa una maswali juu ya haki zako kama mshiriki wa utafiti, tafadhali wasiliana na: KATIBU WAJKUAT ERC, Simu: +25472599171 ethic@jkuat.ac.ke

Mimi, Bwana / Bi,Bibi._____, nikiwa mtu mzima mwenye umri wa miaka 18 na zaidi ninampa ruhusa Binod Kumar kwa utafiti mpya uitwao MICROBIAL METAGENOMIC markers OF TYPE 2 DIABETES MELLITUS (T2DM) kwa wagonjwa wa KENYA. Nimeelezwa kuhusu mtihani utakaofanyika kwenye kinyesi changu, nimepewa nafasi ya kuuliza swali na kutafuta ufafanuzi wa maswala ambayo sikuwa nimeelewa wazi na nimeridhika na jibu na ufafanuzi uliotolewa. Nimearifiwa pia kwamba ikiwa nina swali la ziada au wasiwasi juu ya utafiti baadaye, ninaweza kuwasiliana na mtafiti anayesimamia utafiti huo au Kamati ya Ukaguzi wa Maadili katika Hospitali ya Kitaifa ya Kenyatta.

Ninakubali na ninaweza kutoa sampuli ya kinyesi kwa jaribio linalohitajika katika utafiti huu.

_____ _____

Saini (au Chapisha Kidole gumba) ya Tarehe ya Mshiriki

Ameshuhudiwa na:

_____

Jina la PI au mratibu wa masomo