Trans-Domain Molecular Characterization of Gut Microbiota Reveals Eubacterial, Archaeal And Eukaryotic Dysbiosis in Indian Type-2-Diabetic Subjects

Shrikant Bhute[1], Mangesh Suryavanshi[2], Suyog Joshi[3], Yogesh S. Shouche[2*], Chittaranjan Yajnik[3*], Saroj Ghaskadbi[1*]

[1]Department of Zoology, Savitribai Phule Pune University, India, [2]National Centre for Cell Science, India, [3]Diabetes Unit, KEM Hospital and Research Centre, India

Submitted to Journal:
Frontiers in Microbiology

Specialty Section:
Microbial Symbioses

Article type:
Original Research Article

Manuscript ID:
232702

Received on:
24 Sep 2016

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

YSS, SSG and CSY contributed to conception, design and coordination of the study and to the critical revisions of the manuscript for important intellectual content. SMJ and CSY were involved in subject recruitment and sample collection. SSB acquired and processed the faecal samples for 16S rRNA amplicon sequencing. SSB performed detailed bioinformatics analysis. SSB and MVS performed archaeal, eukaryotic and fungal amplicon sequencing. SSB prepared the first draft of the manuscript and contributed to the critical revisions of the manuscript for important intellectual content. SSB and SMJ undertook statistical analysis and interpretation of results. SMJ contributed to the critical revisions of the manuscript for important intellectual content. All authors gave final approval of the version to be published.

Keywords

diabetes, Gut Microbiota, Eubacteria, Archaea, eukarya and amplicon sequencing

Abstract

Word count: 308

Diabetes in India has distinct genetic, nutritional, developmental and socio-economic aspects; owing to the fact that changes in gut microbiota are associated with diabetes, we employed semiconductor-based sequencing to characterize gut microbiota of diabetic subjects from this region. We suggest consolidated dysbiosis of eubacterial, archaeal and eukaryotic components in the gut microbiota of newly diagnosed (New-DMs) and long-standing diabetic subjects (Known-DMs) compared to healthy subjects (NGTs). Increased abundance of phylum Firmicutes and Operational Taxonomic Units (OTUs) of Lactobacillus were observed in Known-DMs subjects along with the concomitant graded decrease in butyrate-producing bacterial families like Ruminococcaceae and Lachnospiraceae. Eukaryotes and fungi were the least affected components in these subjects but archaea, except Methanobrevibacter were significantly decreased in them. The two dominant archaea viz. Methanobrevibacater and Methanosphaera followed opposite trends in abundance from NGTs to Known-DMs subjects. There was a substantial reduction in eubacteria, a significant decrease in Bacteroidetes phylum and an increased abundance of fungi in New-DMs subjects. Likewise, opportunistic fungal pathogens such as Aspergillus, Candida were found to be enriched in New-DMs subjects. Analysis of eubacterial interaction network revealed disease-state specific patterns of ecological interactions, suggesting the distinct behaviour of individual components of eubacteria in response to the disease. Further, eubacterial component was found associated with diabetes-related risk factors like fasting glucose, high triglyceride, low HDL, waist-to-hip ratio and fasting insulin. Metagenomic imputation of eubacteria depict deficiencies of various essential functions such as carbohydrate metabolism, amino acid metabolism etc. in New-DMs subjects. Results presented here shows that in the metabolic disorders like diabetes, the dysbiosis may not be just limited to eubacteria. Due to the inter-linked metabolic interactions among the eubacteria, archaeal and eukarya in the gut, the dysbiosis may extend into other two domains existing in the gut leading to trans-domain dysbiosis in microbiota. Our results thus contribute to and expand the identification of biomarkers in diabetes.

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: Yes

Please state the full name of the ethics committee that approved the study. If the study was exempt from this requirement please state the reason below.

The study and the experimental protocols followed were approved by Ethics Committee of KEM Hospital Research Centre, Pune, India (Study number 0847), and informed consent was obtained from each participant.

Please detail the consent procedure used for human participants or for animal owners. If not applicable, please state this.

The participants for the present study are the parents of PCS study cohort (PMC4472941).
The participants were called for group presentation during which they were briefed about the project and their role in the project.

After the satisfactory question-answer session, the written consent was obtained from those who wished to participate.

Please detail any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species. If not applicable, please state this.

Not Applicable

Trans-Domain Molecular Characterization of Gut Microbiota Reveals Eubacterial, Archaeal And Eukaryotic Dysbiosis in Indian Type-2-Diabetic Subjects

Shrikant S Bhute[a], Mangesh V Suryavanshi[b], Suyog M Joshi[c], Chittaranjan S Yajnik[c*], Yogesh S Shouche[b*], Saroj S Ghaskadbi[a*]

[a] Department of Zoology, Savitribai Phule Pune University, Ganeshkhind, Pune- 411 007, Maharashtra, India. Tel: +91-20-25601436, Fax: +91-20-25690617.

[b] Microbial Culture Collection-National Centre for Cell Science, Central Tower, Sai Trinity Building, Garware Circle, Sutarwadi, Pashan Pune-411021, Maharashtra, India. Tel: +91-20-25329000, Fax: +91-20-25329001.

[c] Diabetes Unit, KEM Hospital and Research Centre, Rasta Peth, Pune-411011, Maharashtra, India. Tel: +91-20-66405731, Fax: +91-20-26111958.

* Corresponding authors:
1. yogesh@nccs.res.in
2. csyajnik@hotmail.com
3. ssg@unipune.ac.in

Running Title: Diabetes and gut microbiota

Abstract

Diabetes in India has distinct genetic, nutritional, developmental and socio-economic aspects; owing to the fact that changes in gut microbiota are associated with diabetes, we employed semiconductor-based sequencing to characterize gut microbiota of diabetic subjects from this region. We suggest consolidated dysbiosis of eubacterial, archaeal and eukaryotic components in the gut microbiota of newly diagnosed (New-DMs) and long-standing diabetic subjects (Known-DMs) compared to healthy subjects (NGTs). Increased abundance of phylum Firmicutes and Operational Taxonomic Units (OTUs) of Lactobacillus were observed in Known-DMs subjects along with the concomitant graded decrease in butyrate-producing bacterial families like Ruminococcaceae and Lachnospiraceae. Eukaryotes and fungi were the least affected components in these subjects but archaea, except Methanobrevibacter were significantly decreased in them. The two dominant archaea viz. Methanobrevibacater and Methanosphaera followed opposite trends in abundance from NGTs to Known-DMs subjects. There was a substantial reduction in eubacteria, a significant decrease in Bacteroidetes phylum and an increased abundance of fungi in New-DMs subjects. Likewise, opportunistic fungal pathogens such as Aspergillus, Candida were found to be enriched in New-DMs subjects. Analysis of eubacterial interaction network revealed disease-state specific patterns of ecological interactions, suggesting the distinct behaviour of individual components of eubacteria in response to the disease. Further, eubacterial component was found associated with diabetes-related risk factors like fasting glucose, high triglyceride, low HDL, waist-to-hip ratio and fasting insulin. Metagenomic imputation of eubacteria depict deficiencies of various essential functions such as carbohydrate metabolism, amino acid metabolism etc. in New-DMs subjects. Results presented here shows that in the metabolic disorders like diabetes, the dysbiosis may not be just limited to eubacteria. Due to the inter-linked metabolic interactions among the eubacteria, archaeal and eukarya in the gut, the dysbiosis may extend into other two domains existing in the gut leading to trans-domain dysbiosis in microbiota. Our results thus contribute to and expand the identification of biomarkers in diabetes.

Keywords: Diabetes, gut microbiota, eubacteria, archaea, eukarya and amplicon sequencing.

Introduction

Eubacterial assemblage associated with the human body together with other microbes like archaea, eukaryotes and fungi are referred to as 'microbiota'. Trillions of these microbes that live in our distal gut are believed to be co-evolving with their hosts (Ley et al., 2008). Within the gut, these microbes are found to have good coordination amongst themselves and their host; together, their metagenomes contain genes that act as a repertoire of metabolic functions which influence human health (Clemente et al., 2012). Recent studies have revealed that the gut microbiota is subjected to variations in the host's diet (Turnbaugh et al., 2009), genotype (Spor et al., 2011) and health status (Cénit et al., 2014). Any perturbation in the delicate balance between microbial consortia and host results in 'dysbiosis', sometimes leading to severe ailments in the host. Thus, gastrointestinal disorders such as Inflammatory bowel disease (Frank et al., 2007) and colitis (Lucke et al., 2006); metabolic disorders such as obesity (Turnbaugh et al., 2006) and diabetes (Karlsson et al., 2013; Qin et al., 2012; Zhang et al., 2013) are found to be associated with the distinct pattern of gut microbiota in which certain OTUs/species are present in different proportions.

Although, studies on gut microbiota are largely dominated by eubacteria, in recent years, studies concerning archaea, (Gaci et al., 2014; Scanlan et al., 2008) fungi (Dollive et al., 2012; Wang et al., 2014) and eukaryotes (Grattepanche et al., 2014; Pandey et al., 2012) present in the human gut are being conducted to understand their distribution and possible role in human health. Thus, archaea such as genus *Methanobrevibacter* has been linked with human diseases like obesity (Million et al., 2012) and periodontitis (Lepp et al., 2004). Fungi residing in the gut are too associated with diseases such as colorectal adenomas (Luan et al., 2015) and, Crohn's disease (Li et al., 2014b). Similarly, eukaryotes in the gut are found to be very complex and are linked with human diseases (Gouba et al., 2014). Thus, besides the fact that reports on gut archaea, fungi and eukaryotes are lagging behind eubacteria, studies such as these are clear indication that these microbes together with eubacteria forms a very complex ecosystem in the gut and their functional role in human health and diseases needs to be evaluated thoroughly.

A study conducted in the Indian population (Patil et al., 2012) suggested compositional differences in gut microbiota and how it differs from the Western population; hence, efforts to define gut microbiota in the Indian population regarding different disorders such as diabetes, are likely to expand our understanding of the role of gut microbiota in aetiology of such disorders. Indeed, findings based on tagged amplicon sequencing and metagenomics have established a novel possible contribution of changes in gut microbiota to the aetiology of diabetes in European and Chinese populations (Karlsson et al., 2013; Qin et al., 2012; Zhang et al., 2013). But, comprehensive reports on variation in gut microbiome of diabetes-prone Indian population are lacking. India is one of the global capitals of diabetes with an estimated 69.1 million diabetic patients in year the 2015 (International Diabetes Federation, 2015). The explosive epidemic of diabetes in India is incompletely explained, although various contributing factors are suggested. Compared to diabetic patients in the Western world, Indian diabetic patients have unique and paradoxical characteristics. These include possible heightened genetic predisposition (Ramachandran et al., 2012), intrauterine undernutrition (thrifty phenotype) leading to epigenetic predisposition (Yajnik, 2001), the manifestation of diabetes at an earlier age and at a much lower body mass index (BMI) compared to white Caucasians (Yajnik, 2004). Diabetes seems to be precipitated in this population by rapid economic and nutritional transition and rural-urban migrations (Anjana et al., 2011).

Based on above facts, we hypothesized that the dysbiosis in gut microbiota may not be limited to just eubacteria but other two domains (Archaea and Eukarya) too are disturbed due to the

disease condition or vice-versa. In the present study, we investigated the composition of intestinal microbiota of newly diagnosed (New-DMs) and long-standing diabetic subjects (Known-DMs) and compared it with normal glucose tolerant subjects (NGTs). We used Ion torrent PGM sequencing technology, to analyse eubacterial and archaeal 16S rRNA gene, 18S rRNA gene from eukaryotes and fungal ITS tagged amplicon from faecal samples.

Material and Methods

Participants and Sample collection

We studied 49 adults, who are parents of children in the Pune Children Study (PCS) conducted by Diabetes Unit of KEM Hospital Research Centre (Yajnik et al., 1995). They have been followed up since 1995 along with their children with serial glucose tolerance testing. The present study refers to clinical and metabolic follow up in 2009. The study and the experimental protocols followed were approved by Ethics Committee of KEM Hospital Research Centre, Pune, India (Study number 0847), and separate written informed consent was obtained from each participant. Inclusion criterion in NGTs group was absence of any apparent acute or chronic disorders. New-DMs were the participants that were diagnosed with type 2 diabetes during the routine check-up, were not on anti-diabetic treatment until sample collection and free from any acute and chronic illness. Known-DMs subjects were known cases of type 2 diabetes in PCS cohort, were on anti-diabetic treatment at least for past one year and free from any acute and chronic illness. Whereas, general exclusion criterion for all three groups were subjects undergoing dietary intervention, use of antimicrobial in past three months and major surgeries of gastrointestinal tract. All participants were admitted to Diabetes Unit the evening before the investigations. Anthropometry was measured by trained observers according to standard protocols. The following morning, fasting blood specimens were assessed for plasma glucose, insulin and lipids. Sixteen known diabetic subjects underwent only fasting and post-breakfast glucose measurements. In the remaining subjects, an oral glucose tolerance test (75 g anhydrous glucose) was carried out according to the WHO 1999 protocol. Faecal samples were collected from all participants in a sterile container and preserved at -80 °C until DNA extraction.

Measurement of biochemical parameters

Plasma glucose, cholesterol, HDL-cholesterol, and triglyceride concentrations were measured using standard enzymatic methods (Hitachi 902, Germany). Between-batch coefficients of variation for all these assays were <3% in the normal range. Plasma insulin was measured using Delfia technique (Victor 2, Wallac, Turku, Finland). Overweight was defined as BMI ≥25 kg/m^2 and <30 kg/m^2, and obesity as BMI ≥30 kg/m^2. Diabetes mellitus was diagnosed if fasting plasma glucose ≥126 mg/dl or 120-minute plasma glucose ≥200 mg/dl. Hypercholesterolaemia was defined as plasma total cholesterol ≥200 mg/dl, hypertriglyceridaemia as plasma triglyceride concentration ≥150 mg/dl and low HDL-cholesterol as HDL-cholesterol concentration <40 mg/dl for men and <50 mg/dl for women. Hypertension was defined as systolic blood pressure (SBP) ≥130 mmHg or diastolic blood pressure (DBP) ≥85 mmHg.

Sequencing of 16S rRNA gene amplicons

Total community DNA was extracted from each faecal sample using QIAmp DNA Stool Mini kit (Qiagen, Madison USA) as per manufacturer's protocol. The PCR amplification and sequencing of resulting amplicons was performed as described earlier (Bhute et al., 2016). Briefly, the concentration of extracted DNA was measured using Nanodrop-1000, (Thermo Scientific, USA). DNA concentration was normalised to 100 ng/µl and used as template for

amplification of 16S rRNA gene. PCR was set up in 50 µl reaction using AmpliTaq Gold PCR Master Mix (Life Technologies, USA) and with 16S rRNA V3 region specific bacterial universal primers: 341F (5'-CCTACGGGAGGCAGCAG-3') and 518R (5'-ATTACCGCGGCTGCTGG-3') (Bartram et al., 2011). Following conditions were used for PCR: initial denaturation at 94°C for 4 min, followed by 20 cycles of 94 °C for 1 min, 56 °C for 30 s, and 72 °C for 30 s with final extension at 72 °C for 10 min. PCR products were purified using Agencourt AMPure XP DNA purification Bead (Beckman Coulter, USA). Resulting PCR products were end repaired and ligated with sample specific barcode adaptor as explained in Ion Xpress™ Plus gDNA Fragment Library Preparation user guide. Prior to sequencing, fragment size distribution and molar concentrations of amplicons were assessed on Bioanalyser 2100 (Agilent Technologies, USA) using High Sensitivity DNA Analysis Kit. All amplicons were diluted to the lowest molar concentration and pooled into sets of 10 samples. Emulsion PCR was carried out on Ion OneTouchTM System using Ion OneTouch™ 200 Template Kit v2 DL (Life Technologies) as explained in Ion OneTouchTM 200 Template Kit v2 user manual. The resulting template positive Ion Sphere particles were enriched using Ion OneTouch ES system and sequencing of amplicon libraries was carried out on 316 chips using Ion Torrent PGM system and Ion Sequencing 200 kit following the user guide: Ion PGM™ Sequencing 200 Kit v2.

Sequencing of archaeal 16S, eukaryotic 18S and fungal ITS genes

The archeal 16S, eukaryotic 18S and fungal ITS1 genes were PCR amplified using primers listed in Supplementary Table 1. The resulting PCR products were purified using Agencourt AMPure XP DNA purification Bead (Beckman Coulter, USA) and quantified using Nanodrop-1000 (Thermo Scientific, USA). Then, PCR products of all NGTs samples (n=19), all New-DMs (n=14) and all Known-DMs (n=16) were pooled by mixing equal quantities of concentration normalized PCR products. This way we obtained three pools for each archaeal 16S rRNA, eukaryotic 18S rRNA and fungal ITS1. All the pooled samples were then sequenced using Ion Torrent PGM. Since, fungal ITS amplicons varied in length, we fragmented 100 ng of it with Ion Shear Enzyme mix (Ion Xpress Plus Fragment Library preparation kit, Life Technologies) for 20 min and 200 bp size fragments were selected before adapter ligation step (Tang et al., 2015).

Sequence processing and bioinformatics analysis of Eubacterial 16S rRNA gene amplicons

All PGM quality-approved reads from 49 samples were exported as sample specific fastq files and pre-processed in Mothur pipeline (Schloss et al., 2009) with following conditions: 1) minimum length - 150bp, 2) maximum length - 200 bp, 3) maximum homopolymer - 5, 4) maximum ambiguity -0, and 5) average quality score - 20. This way we derived total of 2.1 million high quality amplicon reads from 49 samples; subsequently, these reads were pooled as single FASTA file for further analysis in QIIME: Quantitative Insights Into Microbial Ecology (Caporaso et al., 2010). Briefly, reads were binned into Operational Taxonomic Units (OTUs) at 97% sequence similarity using UCLUST algorithm and single sequence from each OTU was picked out for further analysis. The PyNAST algorithm was used to align representative sequences against Greengenes core set; all unaligned and chimeric sequences were excluded from alignment and downstream analysis. Then lane masking was applied to the alignment to retain conserved regions of 16S rRNA and phylogenetic tree was inferred using FastTree 2.1.3. Additionally, all reads were assigned to the lowest possible taxonomic rank by utilising RDP Classifier 2.2 with a confidence score of 80%. Alpha diversity measures such as Chao1 index (Chao A, 1984) and Shannon index (Shannon, 1948) were inferred. Phylum level abundance

data and alpha diversity indices were compared among the three groups using non-parametric test such as Wilcoxon sum rank test and Kruskal-Wallis rank sum test. To assess beta diversity among three study groups, we applied phylogenetic distance based UniFrac (both unweighted and weighted) analysis was used and the results are visualized as Principal coordinate plots. To determine differentially abundant OTUs among the three groups, OTU table was filtered such that at least 8 sample will have that OTU to be retained in the OTU table. Kruskal-Wallis rank sum test was then applied to filtered OTU table containing 1969 OTUs. We next applied supervised machine learning approach (Random Forest) to identify OTUs that were indicators of community differences in three groups. This was done by estimating the amount of error introduced if a particular OTU is removed from a group of indicator OTUs and assigning it an importance score. We considered only those OTUs as highly discriminative if its mean decrease in accuracy was greater than 0.002.

Clustering of samples into Enterotypes

To understand whether disease state has any effect on composition of enterotypes, we applied original measurements proposed by Arumugam et al. (Arumugam et al., 2011) and as detailed at http://enterotyping.embl.de (Arumugam et al., 2014) to partition the samples into distinct enterotypes clusters. Briefly, the genus level abundance data was segregated according to the three categories, imported in R and clustered using partitioning around medioid (PAM) algorithm followed by determination of optimal number of clusters by utilizing Calinski-Harabasz (CH) index. Finally results of between class (BC) analyses were visualized as principal component analysis. Additionally, taxa that influenced partitioning of samples into enterotypes (drivers of enterotype) were identified based on their abundance in a particular enterotype.

Bioinformatics analysis of archaeal 16S, eukaryotic 18S and fungal ITS genes

Most of the steps for analysis of pooled archaeal 16S, eukaryotic 18S and fungal ITS1 genes were similar as described in section 2.6, except for the fact that QIIME compatible SILVA_111 database (Quast et al., 2013) for archaeal 16S and eukaryotic 18S amplicons and QIIME compatible UNITE_12_11 database (Kõljalg et al., 2013) for fungal ITS amplicon was used during the OTU picking step.

Prediction of ecological relationships

To predict ecological relationships among gut microbiota, microbial association network showing co-occurrence and co-exclusion pattern was built as described before (Faust et al., 2012). Briefly, genus level abundance data was imported to CoNet plugin (version 1.0.4 beta) in Cytoscape 3.0.0 environment (Shannon et al., 2003). To produce association network, 100 top and bottom edges were used with two measures of similarity (Pearson and Spearman) and three measures of dissimilarity (Bray-Curtis, Hellinger and Kullback-Leibler). Spurious correlations due to compositional structure of relative abundances were avoided by bootstrapping and re-normalization and resulting networks were combined using Simes method followed by Benjamini-Hochberg-Yekutieli false discovery rate (FDR) correction with FDR cut-off of 0.05. Finally, all unstable edges outside the 95% confidence interval of bootstrap distribution score were removed and network was visualized and suitably edited.

Metagenomic imputation

For metagenomic imputation, amplicon sequences were binned into OTUs at 97% similarity using closed-reference OTU picking in QIIME. The resulting OTU table was filtered such that at least 8 samples will have that OTU to retain it in OTU table. Resulting OTU table was then

analysed using online tool PICRUSt (Langille et al., 2013) at *http://huttenhower.sph.harvard.edu/galaxy/*. PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states) is a computational tool that uses marker gene data for prediction of functional composition of metagenome. Briefly, OTU abundance table was first normalised for 16S rRNA copy number against known gene copy number for each OTU. Functional predictions were categorised into KEGG pathways and an annotated table of predicted gene family counts (KOs) for each sample using predict metagenome tab. Gene family table then categorised by function and further statistical analysis was performed in STAMP v2.0.1 (Parks and Beiko, 2010).

Additional Statistical Analysis

Biochemical and anthropometric parameters were expressed as mean (SD) and ANOVA test is used to compare differences among the study groups. Different type of data generated through QIIME was imported and analysed in ade4, vegan and ggplot2 packages within R software (The R Core Team, 2013) environment. In addition, relationship between biochemical parameters and microbiota were assessed using PERMANOVA: permutational multivariate analysis of variance test (Anderson MJ and Walsh DCI, 2013).Covariance between biochemical parameters dataset and genus abundance dataset was performed by using co-inertia analysis (Dray et al., 2003), these two datasets were connected to each other owing to the presence of same subjects.

Availability of data

Raw sequences generated in the present study are deposited to NCBI SRA under accession number SRP041693.

Results

Summary of biochemical parameters

Biochemical and anthropometric characteristics are shown in Table 1. Out of the 49 participants, 19 were NGTs, 14 were New-DMs and 16 were Known-DMs. In the total study group, 8 participants were obese and 28 were overweight. Twelve participants had hypercholesterolemia, 16 had hypertriglyceridemia, 45 had low HDL and 8 were hypertensive.

Altered eubacterial diversity and OTU composition of diabetic subjects

We obtained and analysed 4,111 eubacterial OTUs among the three study groups. Analysis of alpha diversity indices revealed that overall diversity in New-DMs was noticeably reduced and both expected (Chao1, $p<0.1$) and observed (Observed Species, $p<0.05$) species diversity indices were significantly lowered in New-DMs and Known-DMs subjects (Figure 1a). Out of eight bacterial phyla detected, Bacteroidetes ($p<0.1$) and Proteobacteria ($p<0.05$) were found significantly lowered in New and Known-DMs compared to NGTs. We noted that Firmicutes ($p<0.05$) tend to be progressively increased from NGTs to New-DMs and then to Known-DMs (Figure 1b). Kruskal-Wallis test revealed the presence of 83 significantly different OTUs ($p<0.01$) of which *Prevotella copri*, *Faecalibacterium prausnitzii* and Lachnospiraceae OTUs were enriched in NGTs whereas *Lactobacillus ruminis* OTUs were found enriched in Known-DMs (Figure 2a). Moreover, 2 OTUs belonging to genus streptococcus were abundant in New-DMs. Interestingly, the OTUs assigned to *Prevotella copri* and Lachnospiraceae were found to be negatively correlated with fasting glucose (Supplementary Table 2). Using UniFrac distance based PCoA biplots, we demonstrate substantial segregation of the subjects into three groups based on the presence/absence (Unweighted UniFrac, Figure 2b) and abundance of specific

bacterial taxa (Weighted UniFrac, Figure 2c).We thus suggest that the presence of discrete clusters of samples in PCoA biplot is an indication of unique bacterial community structure in the three study groups. We further observed that OTUs belonging to order Bacteroidales, family Lachnospiraceae and phylum Bacteroidetes and genus *Prevotella* were determinative taxa for segregation of NGTs from New-DMs and Known-DMs subjects on PCoA biplots. It was noted that *Lactobacillus* was the crucial contributor for segregation of Known-DMs from rest of the samples and thus confirms the findings of Kruskal-Wallis test (performed above) demonstrating enrichment of *Lactobacillus ruminis* in these subjects.

Disease state has profound effect of composition of enterotypes:

We were able to stratify the gut microbial communities of NGTs, New-DMs as well as Known-DMs subjects into three distinct enterotypes (E) (Figure 3 and Supplementary Figure 1). As observed earlier by Arumugam et al., healthy (NGTs) subjects (Figure 3b) grouped into three enterotypes (E1- Bacteroidetes, E2- *Prevotella* & E3- Ruminococcus). However, notable compositional changes were observed in enterotypes of both New-DMs (Figure 3c) and Known-DMs (Figure 3d) compared to enterotypes of NGTs subjects. Based on the abundance of the different genera we found that all three enterotypes in these subjects were found to driven by members of Firmicutes (New-DMs: E1- *Lachnospira*, E2- *Streptococcus* and E3- *Weissella* & Known-DMs: E1- *Veillonella*, E2- *Lachnospira* and E3- *Lactobacillus*). Notably, the E2 (five subjects) in New-DMs and E3 (eight subjects) in Known-DMs were dominated by taxa that were being enriched in these subjects.

Archaeal, Eukaryotic and fungal dysbiosis

We generated 109,561 good quality archaeal 16S rRNA amplicon reads from three pools of samples (NGTs, New-DMs and Known-DMs); which clustered into 65 OTUs belonging to Euryarchaeota and Thaumarchaeota phyla. The former being the most dominated phylum occupying more than 99% reads of all three groups. We noticed the gradual increase in *Methanobrevibacter* (which was also the most abundant taxa in all groups) and associated decrease in *Methanosphaera* abundance from NGTs to New-DMs to Known-DMs subjects. From the three pools of Eukaryotic sequence data, we obtained 41,959 good quality sequences that clustered into 383 OTUs and could be assigned to four phyla: Chloroplastida, Metazoa, Stramenopiles, and Metamonada. Members of Stramenophile especially members of genus *Blastocystis* were found abundant in all groups. Fungi, particularly members belonging to Saccharomycetales were abundant in New-DMs compare to NGTs and Known-DMs. For fungal ITS data, we could obtain 106,185 reads that clustered into 871 OTUs belonging to phyla Ascomycota being most dominant followed by Basidiomycota and Zygomycota to be least dominant. From the Ascomycota group; *Aspergillus* and *Emericella*, the two alternative forms of the same fungus predominated most of the sequences (Figure 4).

Altered microbial composition is associated with clinical parameters

To analyse the effect of different biochemical and anthropometric measurements on sampled microbiota among the three groups, we used PERMANOVA and Co-inertia analysis. After applying PERMANOVA test, we discovered that fasting glucose, HDL, triglyceride and waist-hip ratio to be associated with OTU diversity across all samples (Supplementary Table 3). In the case of Known-DMs, we found HDL, triglycerides and fasting insulin (FINS) and in the case of New-DMs, oral glucose tolerance test (OGTT) and waist-hip ratio to have an influence on distinct OTU diversity. Further, the covariance between genus abundance and clinical and anthropometric parameters were examined using co-inertia analysis (1000 permutations) of these datasets. This resulted in modest relationship (RV coefficient = 0.219, P-value = 0.196)

between these datasets (Figure 5). Similar and subsequent analysis were not performed on simulated datasets of Archaeal, Eukaryotic and Fungal datasets.

Eubacterial interaction network

Microbiome network containing a total of 108 nodes connected with 174 edges together representing 46% co-occurrence and 54% of mutual exclusion interactions were obtained. Further, to measure the scale-freeness of the network, we used fitted power law and obtained correlation of 0.6 with R-square value of 0.723 (Supplementary Figure 2). This network reveals that the patterns observed were disease state specific, i.e. majority of the edges were found clustering within one study group providing a clue that individuals in each group have distinctly interacting microbiome composition (Figure 6 and Supplementary Table 4). We then filtered the network to retain nodes positively interacting with each other, assuming that microbes represented by these nodes will stay together in a given community. In the filtered network of positively interacting genera, we noticed that a cluster of *Lachnospira*, *Ruminococcus*, *Faecalibacterium*, *Roseburia*, *Oscillospira*, *Parabacteroides* and *Bulleidia* to decompose from NGTs to New-DMs then to Known-DMs (Supplementary Figure 3, 4, 5). We also noted negative interactions of *Lactobacillus* in Known-DMs.

Deficient metabolic activities in New-DMs as revealed by imputed metagenome

Having identified the compositional changes in microbiota with respect to diabetes state we tested whether these changes are accompanied with selectively fostering or lacking particular functional capabilities of gut microbiota. Similarities and differences in metabolic capabilities in gut microbiota were evaluated by making the pair-wise comparison between the diabetes statuses using two-sided Welch's t-test. Compared to NGTs, metagenome of New-DMs was found augmented with glycerolipid metabolism, fructose and mannose metabolism, pentose phosphate pathway, galactose metabolism, glycolysis/gluconeogenesis and arginine and proline metabolism. Concurrently, these subjects were found to be deficient in many important metabolic activities such as carbohydrate metabolism (including carbohydrate digestion and absorption¸ TCA cycle, oxidative phosphorylation, glycan biosynthesis and metabolism¸ glycosyltransferases), amino acid metabolism (including metabolism of glycine, serine, threonine, histidine), vitamin B metabolism (including folate, biotin, pyridoxine metabolism), glutathione metabolism and other functions (Supplementary Figure 6). Compared to Known-DMs, New-DMs were deficient of carbohydrate digestion and absorption, glycosyltransferases and glutathione metabolism (Supplementary Figure 7). Conversely, they were enriched with functions unrelated to carbohydrate or amino acid or lipid metabolism compared to NGTs (Supplementary Figure 8).

Discussion

The present study is first to report perturbation in the gut microbiota of Indian diabetic subjects across the three domains of life. Considering the unique characteristics of Indian diabetic subjects, understanding their gut microbiota will be important to understand the possible role of gut microbiota in affecting these characteristics. Members of eubacteria such as *Prevotella copri*, Lachnospiraceae and Ruminococcaceae families were found significantly abundant in NGTs subjects. Known-DMs subjects exhibited increased abundance of Firmicutes and OTUs belonging to genus *Lactobacillus*. These organisms were seen to have an effect on the segregation of samples in both unweighted and weighted UniFrac based PCoA biplots. Fungi prevailed in New-DMs, especially genus *Aspergillus*, *Candida* and *Saccharomyces* were found enriched in these subjects. We also observed the progressive decline in butyrate-producing bacteria from NGTs to Known-DMs subjects. These variations in gut microbiota were

associated with diabetes risk factors such as fasting glucose, high triglycerides, low HDL and fasting insulin. Additionally, synergistic or antagonistic interactions occurring in gut microbiota were found specific to the stage of glucose intolerance. Using PICRUSt, we predicted that the gut microbiome of New-DMs subjects was metabolically disturbed and was lacking in many necessary functions.

Increased Firmicutes and proportionate decrease in Bacteroidetes is linked with more energy harvesting and storage in ob/ob animals (Turnbaugh et al., 2006). Analogous to animal studies, human obesity is also found to be linked with higher Firmicutes to Bacteroidetes ratio (Ley et al., 2006). Our finding of increased abundance of Firmicutes in known-DMs is in agreement with previous reports (Karlsson et al., 2013) but not with findings of Larsen and co-workers, who reported a decrease in the proportion of Firmicutes (Larsen et al., 2010). Association of Firmicutes with obesity and diabetes could operate through insulin resistance which is a common attribute of both the conditions (Pandolfi et al., 2000).

Analysis of differentially abundant OTUs revealed that NGTs were highly enriched with Prevotellaceae, Lachnospiraceae and Ruminococcaceae families. Members belonging to Prevotellaceae such as genus *Prevotella* contribute significantly to inter-individual variation in gut microbiota (Arumugam et al., 2011) and increased proportions of *Prevotella* are associated with the diet rich in plant-derived complex carbohydrates and fibres such as the diet in Indians (De Filippo et al., 2010). Additionally, a study in which subjects were kept of dietary interventions (barley kernel-based bread, which is considered as a rich source of fibres), showed that there was a significant increase in *Prevotella copri* and that it was found to be associated with improvement in glucose metabolism in these subjects (Kovatcheva-Datchary et al., 2015). Strikingly, several studies on type 1 diabetes, a pathophysiologically different disorder related to persistent hyperglycemia, are also reporting reduced levels of *Prevotella* in newly diagnosed as well as longstanding type 1 diabetic subjects (Alkanani et al., 2015; Mejía-León et al., 2014; Mejía-León and Barca, 2015). At this moment we could speculate that this could just be a coincidence or indeed it is linked with hyperglycemia *per se* which is a common attribute of type 1 and type 2 diabetes. Members of families Lachnospiraceae and Ruminococcaceae are known producers of short-chain fatty acids (SCFAs) such as acetate and butyrate. These SCFAs are known to confer many health benefits; individuals lacking bacterial families producing SCFAs suffer from many diseases (Morgan et al., 2012). Interestingly, we observed decreasing trends in the richness of these bacterial families with progressive deterioration of glucose tolerance (from NGTs to New-DMs to Known-DMs subjects). Presence of these families in the gut may be essential to foster a "healthy state", and their depletion might have a role in diabetes development (Remely et al., 2014). Thus, we hypothesise that the decreased abundance of *Prevotella copri* and concomitant loss of short chain fatty acids producers in New- and Known-DMs subjects could be linked with glucose intolerance in these subjects as these organisms were found to be negatively correlated with fasting glucose in our analyses.

We also found that Known-DMs were enriched with genus *Lactobacillus* consistent with previous studies on diabetic subjects in different populations of the world (Larsen et al., 2010; Lê et al., 2012). Karlsson and co-workers have also demonstrated enrichment of lactobacilli-derived metagenomic clusters (MGCs) in type 2 diabetic patients that they found positively correlating with fasting glucose and HbA_{1c}. Another large-scale study dealing with the characterization of over 170 *Lactobacillus* species from oral cavity showed a higher prevalence of lactobacilli in diabetic subjects (Teanpaisan et al., 2009) and this increase in *Lactobacillus* species has been linked with increased salivary glucose in children with diabetes (Karjalainen

et al., 1996). *Lactobacillus ruminis* that we have found to be significantly increased in Known-DMs subjects is a member of indigenous gut microflora (O' Donnell et al., 2015), and found to have as high as 16 carbohydrate utilization pathways including those for utilization of glucose, fructose, mannose, galactose, starch and sucrose (Forde et al., 2011). Thus, as reported earlier, the catabolic flexibility of this organism towards varied dietary carbohydrates is evident (O'Donnell et al., 2011). Above facts taken together, indicate that enrichment of the lactobacilli in gastrointestinal tract of diabetic subjects could be a consequence of higher than usual concentration of glucose, which needs to be confirmed.

Besides this, we also show the gradation of NGTs, New-DMs and Known-DMs samples on UniFrac biplots. These UniFrac biplots were plotted using phylogenetic distance which is calculated utilizing unique branch-lengths i.e. only those branches that lead to descendants from one or the other sample but not both samples in a phylogenetic tree were considered (Lozupone et al., 2011). Hence, we believe that segregation of the samples is robust and could be because of the above mentioned compositional differences in bacterial communities in these subjects. We next attempted to group study participants into distinct clusters based on the presence of unique and dominant gut microbial communities called 'enterotypes' (Arumugam et al., 2011). Currently, the concept of enterotype is generating a lot of debate; different groups have different opinions about the presence or absence of such discrete cluster in human gut microbiome (Knights et al., 2014; Moeller et al., 2015). Although, it has been shown earlier that during identification of enterotypes, various factors influence clustering of subjects into distinct enterotypes (Koren et al., 2013); we feel that it is beyond the reach of this article to deal with theories of formation of enterotypes and associated factors affecting their formation, hence, we performed this analysis as originally proposed (Arumugam et al., 2011). We find substantial changes in major contributors of enterotype in New- and Known-DMs subjects compared to NGTs subjects. Especially, we observed E2 in New-DMs and E3 in Know-DMs subjects to be driven by *Streptococcus* and *Lactobacillus* respectively. These findings are important because clustering of subjects based on the presence of unique and predominated taxa could help us in identifying disease-related biomarkers, thus it can find its implications in microbiome-based diagnostics (Knights et al., 2014).

We next looked into archaeal diversity in the three sample groups; *Methanobrevibacter* and *Methanosphaera* were the most prevalent genera. *Methanobrevibacter smithii* (*M. smithii*) and *Methanosphaera stadtmanae* are well adapted to the human gut environment, interestingly, the latter has acquired most of these adaptations through inter-domain lateral gene transfer (Lurie-Weinberger et al., 2012; Samuel et al., 2007). As perceived by us and reported in a previous study (Turnbaugh et al., 2006), *M. smithii* has been represented in large proportion along with increased Firmicutes; it was involved in increased energy harvest through polysaccharide degradation. Further, the same study noted that this attribute was transmissible such that microbiota transplantation from obese donor to lean germ-free mice lead to gain in body fat. Additionally, *M. smithii* directs polysaccharide utilization by gut inhabitants, leading to the formation of large pools of SCFAs which is later used by *M. smithii* for methanogenesis in the gut with consequent increase in host adiposity (Samuel and Gordon, 2006). Thus, *M. smithii* can be a therapeutic target to avoid obesity and associated complications such as diabetes (Samuel et al., 2007).

Based on the work we carried out and several other similar studies, gut eukaryotes and fungi appear to be important components of human gut. Such studies are crucial in the light of involvement of these organisms in human diseases both inside and outside of gastrointestinal

tract (Cui et al., 2013). Morphological and molecular phylogenetic-based classification of eukaryotes show that all eukaryotes originate from one of six super-groups and that most of them are microscopic in nature (Adl et al.). Although for decades human-associated eukaryotes are considered harmful to their host, recent examination of eukaryotic communities in the gut are amending our understanding of this generally neglected component (Hamad et al., 2012; Pandey et al., 2012; Parfrey et al., 2014). Studies such as these and our findings suggest that *Blastocystis* and fungi such as Ascomycota and Basidiomycota are predominant in the human gut. Fungi such as *Candida albicans*, *Aspergillus fumigatus* and *Saccharomyces* are opportunistic pathogens known to be exaggerated in immune-compromised people (Gouba and Drancourt, 2015; Li et al., 2014a). Fungal species mentioned above have also been associated with various diseases in type 1 (Soyucen et al., 2014) and type 2 diabetic subjects (Aly et al., 1991; Nowakowska et al., 2004) and are probably because of the high blood glucose level in these subjects. Thus, marked enrichment of fungi belonging to these and other genera in New-DMs subjects are likely due to the poor glycemic control in these subjects.

We investigated associations between clinical parameters and OTU richness using permutational multivariate analysis of variance (PERMANOVA). PERMANOVA is considered a powerful technique in detecting changes in community structure in response to environmental parameters (Anderson MJ and Walsh DCI, 2013). We observed that fasting glucose, HDL, triglyceride and waist-hip ratio as largest contributors to the observed variation in OTU richness. Such correlations between risk factors for diabetes and variation in microbes in the gut have been previously reported (Zhang et al., 2013) and are also reflected in our dataset. Thus, it could be relevant in the microbiome-phenotype associations, since, low HDL and high triglycerides are typical features of dyslipidaemia found in T2D and known risk factors for cardiovascular disease (Mooradian, 2009).

We used network analysis to capture specific ecological interactions among the eubacterial consortium in relation to diabetes status. Such interaction networks can predict the outcome of community alterations (Faust et al., 2012) and be helpful in designing intervention studies aimed at altering complex microbial communities to restore the healthy state. In essence, we are not demonstrating complete coverage of all microbial interactions in the gut; but analysing the interactions among microbes in the gut will help us understand how these communities develop or evolve in response to altered physiological and/or metabolic state such as diabetes. We thus highlight two characteristic features of this network: 1) the nature of the interactions observed were diabetes state specific and 2) the disintegration of the microbial cluster of genera: *Lachnospira*, *Ruminococcus*, *Faecalibacterium*, *Roseburia*, *Oscillospira*, *Parabacteroides*, *Bulleidia* from NGTs to New-DMs to Known-DMs. Almost all these genera include known beneficial species having the ability to produce SCFAs as mentioned earlier. Importantly, metagenomic linkage clusters (MLGs) belonging to these butyrate-producing genera were found enriched in non-diabetic controls in diabetes associated metagenomic study (Qin et al., 2012).

Finally, with the bioinformatics tool PICRUSt (Langille et al., 2013) which predicts functional composition using marker gene data, we had an opportunity to look into imputed metagenome-based discrete functional alteration in eubacterial component of our study subjects. We observed that New-DMs were severely depleted with metabolic functions involved in carbohydrate metabolism, amino acid metabolism, various cofactor synthesis and oxidative stress management. Although PICRUSt can accurately predict metagenomic functions, it is limited to those sequences that can be accurately mapped to existing Greengenes database and

does not consider sequences from novel microbial lineages (Langille et al., 2013). Thus, our explanation on imputed metagenome is limited and interpreted cautiously.

One of the strengths of our study is the comparison of gut microbiota of different grades of glucose intolerant subjects from a cohort which is has been followed for the past 20 years, this allowed a confident separation between newly diagnosed and known diabetic subjects. The participants are from the similar socioeconomic background and have a predominantly vegetarian diet. The age and gender distribution in the three groups were similar. One of the limitations of this study is that we were unable to describe sequential events in gut microbiota from healthy to diabetic state due to the cross-sectional design of this study. Another limitation of the study is the relatively small number of participants from one part of the country. Given the diversity in lifestyles, dietary habits, and social-economic status in the country, this study underscores a need for nationwide longitudinal studies. Our study is subject to inherent biases introduced by the use of high-throughput 16S rRNA amplicon sequencing. These include the region of 16S rRNA gene sequenced, set of primers used for gene amplification and use of sequence database for taxonomic assignments of the amplicon reads.

In conclusion, our results add to the growing literature suggesting an association between gut microbiota and diabetes. Broad similarities between our results and literature reports suggest that our measurements are reliable and support consistent association across populations. Additionally, we have broadened the boundaries of diabetes associated gut microbiota by providing the consolidated description on eubacterial, archaeal and eukaryotic dysbiosis in these subjects. Given the peculiarities of diabetes in Indians, these results suggest an important avenue be further explored for causality and possible interventions to prevent or modify the course of diabetes and related disorders. We anticipate the need for subsequent studies describing differences in gut microbial communities of diabetes patients from different populations and identification of relevant population specific biomarkers.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' contributions

YSS, SSG and CSY contributed to conception, design, and coordination of the study and to the critical revisions of the manuscript for important intellectual content. SMJ and CSY were involved in subject recruitment and sample collection. SSB acquired and processed the faecal samples for 16S rRNA amplicon sequencing. SSB performed detailed bioinformatics analysis. SSB and MVS performed archaeal, eukaryotic and fungal amplicon sequencing. SSB prepared the first draft of the manuscript and contributed to the critical revisions of the manuscript for important intellectual content. SSB and SMJ undertook statistical analysis and interpretation of results. SMJ contributed to the critical revisions of the manuscript for important intellectual content. All authors gave final approval of the version to be published.

Acknowledgements

The authors are thankful to Department of Biotechnology, Government of India, for funding Microbial Culture Collection Project (BT/PR10054/NDB/52/94/2007). Authors are thankful to team of Diabetes Unit and the Paediatric Department (KEM Hospital Research Centre, Pune) for their contribution in samples collection. Authors are also thankful to Dr Om Prakash

(Scientist C, Microbial Culture Collection, Pune, India); Dr Kamlesh Jangid, Scientist C (Microbial Culture Collection, Pune, India); Dr Dhiraj Dhotre, Scientist B (Microbial Culture Collection, Pune, India) and Mr Sudarshan Shetty (Laboratory of Microbiology, Wageningen University, The Netherlands) for helpful discussion. Pune Children Study was supported by The Wellcome Trust, London. SB and MS are supported by University Grant Commission (UGC) and CSIR for financial assistance respectively.

References

Adl, S. M., Simpson, A. G. B., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* 52, 399–451.

Alkanani, A. K., Hara, N., Gottlieb, P. A., Ir, D., Robertson, C. E., Wagner, B. D., et al. (2015). Alterations in Intestinal Microbiota Correlate With Susceptibility to Type 1 Diabetes. *Diabetes* 64, 3510–20.

Aly, F. Z., Blackwell, C. C., MacKenzie, D. A., Weir, D. M., Elton, R. A., Cumming, C. G., et al. (1991). Chronic atrophic oral candidiasis among patients with diabetes mellitus--role of secretor status. *Epidemiol. Infect.* 106, 355–63.

Anderson MJ, and Walsh DCI (2013). Anderson and Walsh (2013) PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions- What null hypothesis are you .pdf. *Ecol. Monogr.* 83, 557–574.

Anjana, R. M., Pradeepa, R., Deepa, M., Datta, M., Sudha, V., Unnikrishnan, R., et al. (2011). Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: phase I results of the Indian Council of Medical Research-INdia DIABetes (ICMR-INDIAB) study. *Diabetologia* 54, 3022–7.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–80.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., et al. (2014). Addendum: Enterotypes of the human gut microbiome. *Nature* 506, 516–516.

Bhute, S., Pande, P., Shetty, S. A., Shelar, R., Mane, S., Kumbhare, S. V., et al. (2016). Molecular characterization and meta-analysis of gut microbial communities illustrate enrichment of prevotella and megasphaera in Indian subjects. *Front. Microbiol.* 7, 1–14. doi:10.3389/fmicb.2016.00660.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat. Methods* 7, 335–336.

Cénit, M. C., Matzaraki, V., Tigchelaar, E. F., and Zhernakova, A. (2014). Rapidly expanding knowledge on the role of the gut microbiome in health and disease. *Biochim. Biophys. Acta* 1842, 1981–1992.

Chao A (1984). Nonparametric-estimation of the number of classes in a population.pdf. *Scandian J. Statisitcs* 11, 265–270.

Clemente, J. C., Ursell, L. K., Parfrey, L. W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell* 148, 1258–70.

Cui, L., Morris, A., and Ghedin, E. (2013). The human mycobiome in health and disease.

Genome Med. 5, 63. doi:10.1186/gm467.

Dollive, S., Peterfreund, G. L., Sherrill-Mix, S., Bittinger, K., Sinha, R., Hoffmann, C., et al. (2012). A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol.* 13, 1–13. doi:10.1186/gb-2012-13-7-r60.

Dray, S., Chessel, D., and Thioulouse, J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology* 84, 3078–3089.

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* 8, e1002606. doi:10.1371/journal.pcbi.1002606

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., et al. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14691–6.

Forde, B. M., Neville, B. A., O'Donnell, M. M., Riboulet-Bisson, E., Claesson, M. J., Coghlan, A., et al. (2011). Genome sequences and comparative genomics of two Lactobacillus ruminis strains from the bovine and human intestinal tracts. *Microb. Cell Fact.* 10 Suppl 1, S13. doi:10.1186/1475-2859-10-S1-S13.

Frank, D. N., St Amand, A. L., Feldman, R. a, Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13780–5.

Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., and Brugère, J.-F. (2014). Archaea and the human gut: New beginning of an old story. *World J. Gastroenterol.* 20, 16062–16078.

Gouba, N., and Drancourt, M. (2015). Digestive tract mycobiota: A source of infection. *Médecine Mal. Infect.* 45, 9–16.

Gouba, N., Raoult, D., and Drancourt, M. (2014). Gut microeukaryotes during anorexia nervosa: a case report. *BMC Res. Notes* 7, 33. doi:10.1186/1756-0500-7-33.

Grattepanche, J.-D., Santoferrara, L. F., McManus, G. B., and Katz, L. A. (2014). Diversity of diversity: conceptual and methodological differences in biodiversity estimates of eukaryotic microbes as compared to bacteria. *Trends Microbiol.* 22, 432–437.

Hamad, I., Sokhna, C., Raoult, D., and Bittar, F. (2012). Molecular detection of eukaryotes in a single human stool sample from senegal. *PLoS One* 7, e40888. doi:10.1371/journal.pone.0040888.

International Diabetes Federation (2015). Idf Diabetes Atlas. *Diabetes Atlas*. Available at: https://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf.

Karjalainen, K. M., Knuuttila, M. L., and Käär, M. L. (1996). Salivary factors in children and adolescents with insulin-dependent diabetes mellitus. *Pediatr. Dent.* 18, 306–11.

Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C. J., Fagerberg, B., et al. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498, 99–103.

Knights, D., Ward, T. L., McKinlay, C. E., Miller, H., Gonzalez, A., McDonald, D., et al. (2014). Rethinking "enterotypes". *Cell Host Microbe* 16, 433–7.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol.*

Ecol. 22, 5271–7.

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., et al. (2013). A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. *PLoS Comput. Biol.* 9, e1002863. doi:10.1371/journal.pcbi.1002863.

Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y. S., De Vadder, F., Arora, T., et al. (2015). Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. *Cell Metab.* 22, 971-82.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. a, et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–21.

Larsen, N., Vogensen, F. K., van den Berg, F. W. J., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., et al. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* 5, e9085. doi: 10.1371/journal.pone.0009085.

Lê, K.-A., Li, Y., Xu, X., Yang, W., Liu, T., Zhao, X., et al. (2012). Alterations in fecal Lactobacillus and Bifidobacterium species in type 2 diabetic patients in Southern China population. *Front. Physiol.* 3:496, doi: 10.3389/fphys.2012.00496.

Lepp, P. W., Brinig, M. M., Ouverney, C. C., Palm, K., Armitage, G. C., and Relman, D. A. (2004). Methanogenic Archaea and human periodontal disease. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6176–81.

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., et al. (2008). Evolution of mammals and their gut microbes. *Science* 320, 1647–1651.

Ley, R., Turnbaugh, P., Klein, S., and Gordon, J. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* 444, 1022–3.

Li, E., Hussein, H., Todiwala, A., and Kirby, R. (2014a). Primary gut aspergillosis in a patient with acute myeloid leukaemia: the importance of early suspicion and definitive treatment. *BMJ Case Rep.* 2014. doi:10.1136/bcr-2013-202316.

Li, Q., Wang, C., Tang, C., He, Q., Li, N., and Li, J. (2014b). Dysbiosis of Gut Fungal Microbiota is Associated With Mucosal Inflammation in Crohn's Disease. *J. Clin. Gastroenterol.* 48, 513–23.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–72.

Luan, C., Xie, L., Yang, X., Miao, H., Lv, N., Zhang, R., et al. (2015). Dysbiosis of Fungal Microbiota in the Intestinal Mucosa of Patients with Colorectal Adenomas. *Sci. Rep.* 5, 7980. doi:10.1038/srep07980.

Lucke, K., Miehlke, S., Jacobs, E., and Schuppler, M. (2006). Prevalence of Bacteroides and Prevotella spp. in ulcerative colitis. *J. Med. Microbiol.* 55, 617–24.

Lurie-Weinberger, M. N., Peeri, M., Tuller, T., and Gophna, U. (2012). Extensive inter-domain lateral gene transfer in the evolution of the human commensal Methanosphaera stadtmanae. *Front. Genet.* 3, 1–11. doi:10.3389/fgene.2012.00182.

Mejía-León, M. E., and Barca, A. M. C. de la (2015). Diet, Microbiota and Immune System in Type 1 Diabetes Development and Evolution. *Nutrients* 7, 9171–84.

Mejía-León, M. E., Petrosino, J. F., Ajami, N. J., Domínguez-Bello, M. G., and de la Barca, A. M. C. (2014). Fecal microbiota imbalance in Mexican children with type 1 diabetes. *Sci. Rep.* 4, 3814. doi:10.1038/srep03814.

Million, M., Maraninchi, M., Henry, M., Armougom, F., Richet, H., Carrieri, P., et al. (2012). Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii. *Int. J. Obes.* 36, 817–825.

Moeller, A. H., Peeters, M., Ayouba, A., Ngole, E. M., Esteban, A., Hahn, B. H., et al. (2015). Stability of the gorilla microbiome despite simian immunodeficiency virus infection. *Mol. Ecol.* 24, 690–7.

Mooradian, A. D. (2009). Dyslipidemia in type 2 diabetes mellitus. *Nat. Clin. Pract. Endocrinol. Metab.* 5, 150–9.

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V, et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13, R79. doi:10.1186/gb-2012-13-9-r79.

Nowakowska, D., Kurnatowska, A., Stray-Pedersen, B., and Wilczyński, J. (2004). Species distribution and influence of glycemic control on fungal infections in pregnant women with diabetes. *J. Infect.* 48, 339–46..

O' Donnell, M. M., Harris, H. M. B., Lynch, D. B., Ross, R. P., and O'Toole, P. W. (2015). Lactobacillus ruminis strains cluster according to their mammalian gut source. *BMC Microbiol.* 15, 80. doi:10.1186/s12866-015-0403-y.

O'Donnell, M. M., Forde, B. M., Neville, B., Ross, P. R., and O'Toole, P. W. (2011). Carbohydrate catabolic flexibility in the mammalian intestinal commensal Lactobacillus ruminis revealed by fermentation studies aligned to genome annotations. *Microb. Cell Fact.* 10 Suppl 1, S12. doi:10.1186/1475-2859-10-S1-S12.

Pandey, P. K., Siddharth, J., Verma, P., Bavdekar, A., Patole, M. S., and Shouche, Y. S. (2012). Molecular typing of fecal eukaryotic microbiota of human infants and their respective mothers. *J. Biosci.* 37, 221–226.

Pandolfi, C., Pellegrini, L., Sbalzarini, G., and Mercantini, F. (2000). Obesity and insulin resistance. *J. Clin. Invest.* 42, 473–481.

Parfrey, L. W., Walters, W. A., Lauber, C. L., Clemente, J. C., Berg-Lyons, D., Teiling, C., et al. (2014). Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. *Front. Microbiol.* 5, 1–13. doi:10.3389/fmicb.2014.00298.

Parks, D. H., and Beiko, R. G. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* 26, 715–21.

Patil, D. P., Dhotre, D. P., Chavan, S. G., Sultan, A., Jain, D. S., Lanjekar, V. B., et al. (2012). Molecular analysis of gut microbiota in obesity among Indian individuals. *J. Biosci.* 37, 647–657.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., et al. (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., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590-6. doi:10.1093/nar/gks1219.

Ramachandran, A., Snehalatha, C., Shetty, A. S., and Nanditha, A. (2012). Trends in prevalence of diabetes in Asian countries. *World J. Diabetes* 3, 110–7.

Remely, M., Aumueller, E., Merold, C., Dworzak, S., Hippe, B., Zanner, J., et al. (2014). Effects of short chain fatty acid producing bacteria on epigenetic regulation of FFAR3 in type 2 diabetes and obesity. *Gene* 537, 85–92.

Samuel, B. S., and Gordon, J. I. (2006). A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 10011–6.

Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., et al. (2007). Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc. Natl. Acad. Sci. U. S. A.* 104, 10643–8.

Scanlan, P. D., Shanahan, F., and Marchesi, J. R. (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiol.* 8, 79. doi:10.1186/1471-2180-8-79.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–41.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–504.

Soyucen, E., Gulcan, A., Aktuglu-Zeybek, A. C., Onal, H., Kiykim, E., and Aydin, A. (2014). Differences in the gut microbiota of healthy children and those with type 1 diabetes. *Pediatr. Int.* 56, 336–43.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.* 9, 279–90.

Tang, J., Iliev, I. D., Brown, J., Underhill, D. M., and Funari, V. A. (2015). Mycobiome: Approaches to analysis of intestinal fungi. *J. Immunol. Methods* 421, 112–21.

Teanpaisan, R., Hintao, J., and Dahlén, G. (2009). Oral Lactobacillus species in type 2 diabetic patients living in southern Thailand. *Anaerobe* 15, 160–163.

The R Core Team (2013). R : A Language and Environment for Statistical Computing. 1.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. a, Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–31.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* 1, 6ra14. doi:10.1126/scitranslmed.3000322.

Wang, Z. K., Yang, Y. S., Stefka, a. T., Sun, G., and Peng, L. H. (2014). Review article: Fungal microbiota and digestive diseases. *Aliment. Pharmacol. Ther.* 39, 751–766.

Yajnik, C. S. (2001). The insulin resistance epidemic in India: fetal origins, later lifestyle, or both? *Nutr. Rev.* 59, 1–9. doi: http://dx.doi.org/10.1111/j.1753-4887.2001.tb01898.x

Yajnik, C. S. (2004). A Critical Evaluation of the Fetal Origins Hypothesis and Its Implications for Developing Countries Early Life Origins of Insulin Resistance and Type 2 Diabetes in India and Other Asian Countries 1. *Aerican Soc. Nutr. Sci.* 134, 205–210.

Yajnik, C. S., Fall, C. H., Vaidya, U., Pandit, A. N., Bavdekar, A., Bhat, D. S., et al. (1995). Fetal growth and glucose and insulin metabolism in four-year-old Indian children. *Diabet. Med.* 12, 330–6.

Zhang, X., Shen, D., Fang, Z., Jie, Z., Qiu, X., Zhang, C., et al. (2013). Human gut microbiota changes reveal the progression of glucose intolerance. *PLoS One* 8, e71108. doi:10.1371/journal.pone.0071108.

Figures

Figure 1: Summary of diversity measurements. 1a. Assessment of alpha diversity indices in NGTs, New-DMs and Known-DMs subjects. 1b: Variation in phylum level abundance, the box depicts interquartile range between first and third quartiles and the line within box denotes median.

Figure 2: Differentially abundant OTUs & Beta diversity analysis. 2a. Heatmap of the differentially abundant OTUs in three study groups as determined by Kruskal-Wallis test. 2b. Unweighted UniFrac distance based and 2c. Weighted UniFrac distance based PCoA bi-plots; the grey coloured sphere represent a taxonomic group that influence clustering of samples (NGTs: green, New-DMs: yellow and Known-DMs: red) in particular area of the PCoA plot and its size demonstrate abundance of that taxonomic group

Figure 3: Clustering of NGTs, New-DMs and Known-DMs subjects into enterotypes (E). 3a. Clustering of all 49 subjects into enterotypes, NGTs, New-DMs and Known-DMs subjects are identified as different shapes. 3b. Clustering of NGTs subjects only. 3c. Clustering of New-DMs only and 3d. Clustering of Known-DMs only. Upper panel of each part are showing projection of first two principal components of between-class analysis and lower panel shows the driver genera in corresponding enterotypes (E1: green, E2: blue and E3: red).

Figure 4: Assessment of archaea, eukarya and fungi. Heatmap showing abundance of different members of archaeal, eukarya and fungal components of NGTs, New-DMs and Known-DMs subjects.

Figure 5: Co-inertia analysis of relationship of genus level abundance and clinical parameters. Upper panel shows positions of the site on the co-inertia axes using genus (origin of the arrow) and clinical parameter (arrowheads) co-inertia weights. The shorter the arrow, the better the concordance between the two projections. The numbers indicate the samples: NGTs - 1-19, New-DMs - 20-33, Known-DMs - 34-49. Lower pair of plot shows contribution of the two groups of variable to the canonical space; vectors pointing to the same direction are correlated.

Figure 6: Significant co-occurrence and co-exclusion relationships at genus level. Each node represents a bacterial genus; size of the node is proportional to the abundance of the genus and coloured according to diabetes status (Red: Known-DMs, Yellow: New-DMs and Green: NGTs). Each edge represents co-occurrence/co-exclusion relationships, edge width is proportional to the significance of supporting evidence, and colour indicates sign of the association (red: negative, green: positive).

Table1: Biochemical and Anthropometric parameters of the three study groups (Shown in the table mean±SD).

	NGTs	**New-DMs**	**Known-DMs**
N	19	14	16
Age	48.85±5.4	48.64±5.68	50.62±3.49
BMI kg/m^2	25.52±4.0	28.32±2.58[a]	27.41±3.53
Waist-hip ratio	0.92±0.088	0.99±0.071	0.96±0.061
% body fat	35.68±8.21	37.50±6.12	35.46±8.77
Fasting glucose mg/dl	93.8±8.16	138.07±47.35[a]	146.81±44.90[b]
120 min glucose mg/dl	110.50 (18.40)	250.86±77.76[a]	NA
PP glucose mg/dl	NA	NA	226.12±58.43
Fasting insulin IU/L	9.16±5.69	12.06±6.11	10.94±8.31
120 min insulin IU/L	71.39±36.60	127.75±183.76	NA
Systolic BP mmHg	115.66±12.77	114.07±37.81	110.69±31.64
Diastolic BP mmHg	73.53±10.74	73.43±23.31	70.22±20.34
Cholesterol mg/dl	166.63±24.06	194.57±44.15[a]	174.19±38.11
Triglycerides mg/dl	120.60±58	126.64±54.41	137.18±63.18
HDL cholesterol mg/dl	38.50±8.15	40.79±7.51	41.06±7.76

[a] p value <0.01 for New-DMs vs. NGTs
[b] p value <0.01 for Known-DMs vs. NGTs

Supplementary Material

Tables:

Supplementary Table 1: Primer used during the PCR amplification of archaeal, eukaryotic and fungal communities.

Supplementary Table 2: Showing top twenty negative correlations between OTU abundance and fasting glucose.

Supplementary Table 3: Summary of PERMANOVA test associating clinical parameters with the OTU abundance.

Supplementary Table 4: Summary number of edges observed among three groups.

Figures:

Supplementary Figure 1: Optimal number of clusters supporting formation of three enterotypes in NGTs, New-DMs and Known-DMs subjects.

Supplementary Figure 2: Node degree distributions of the network of co-occurrence, and co-exclusion associations. Node degree indicates the number of links that connect a node to others in the network. Power law degree distributions means that most nodes have only a few edges and are often connected by a few high-degree hub nodes.

Supplementary Figure 3: Significant co-occurrence relationships at genus level in NGTs subjects. Each node represents a bacterial genus; size of the node is proportional to the abundance of the genus. Each edge represents co-occurrence relationships; edge width is proportional to the significance of supporting evidence.

Supplementary Figure 4: Significant co-occurrence relationships at genus level in New-DMs subjects. Each node represents a bacterial genus; size of the node is proportional to the abundance of the genus. Each edge represents co-occurrence relationships; edge width is proportional to the significance of supporting evidence.

Supplementary Figure 5: Significant co-occurrence relationships at genus level in Known-DMs subjects. Each node represents a bacterial genus; size of the node is proportional to the abundance of the genus. Each edge represents co-occurrence relationships; edge width is proportional to the significance of supporting evidence.

Supplementary Figure 6: Extended error bar plot showing the differentially enriched KOs in NGTs subjects (green bars) as compared to New-DMs subjects (yellow bars).

Supplementary Figure 7: Extended error bar plot showing the differentially enriched KOs in NGTs subjects (green bars) as compared to Known-DMs subjects (red bars).

Supplementary Figure 8: Extended error bar plot showing the differentially enriched KOs in New-DMs subjects (yellow bars) as compared to Known-DMs subjects (red bars).

Figure 1.JPEG

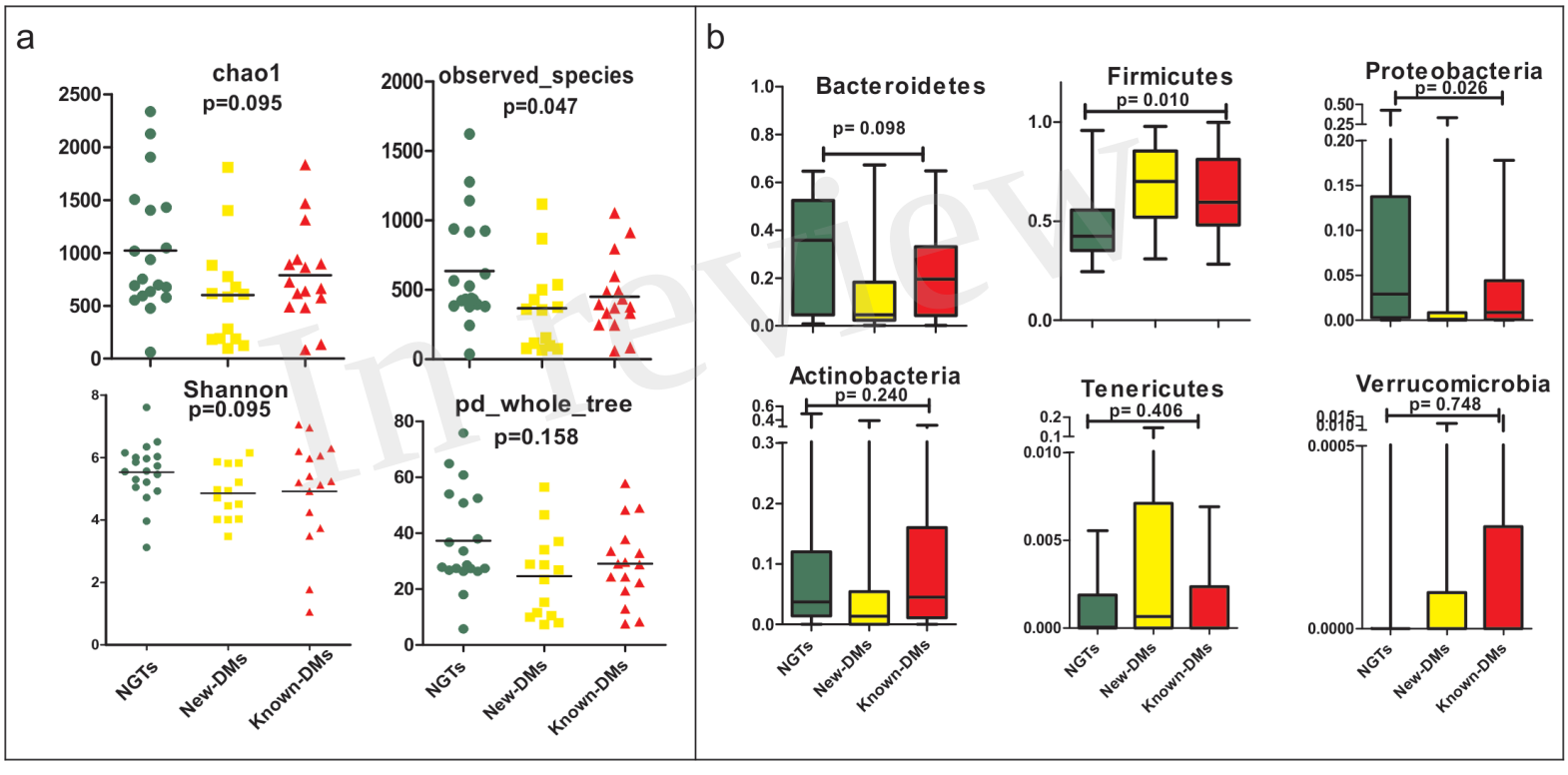


Figure 2.JPEG

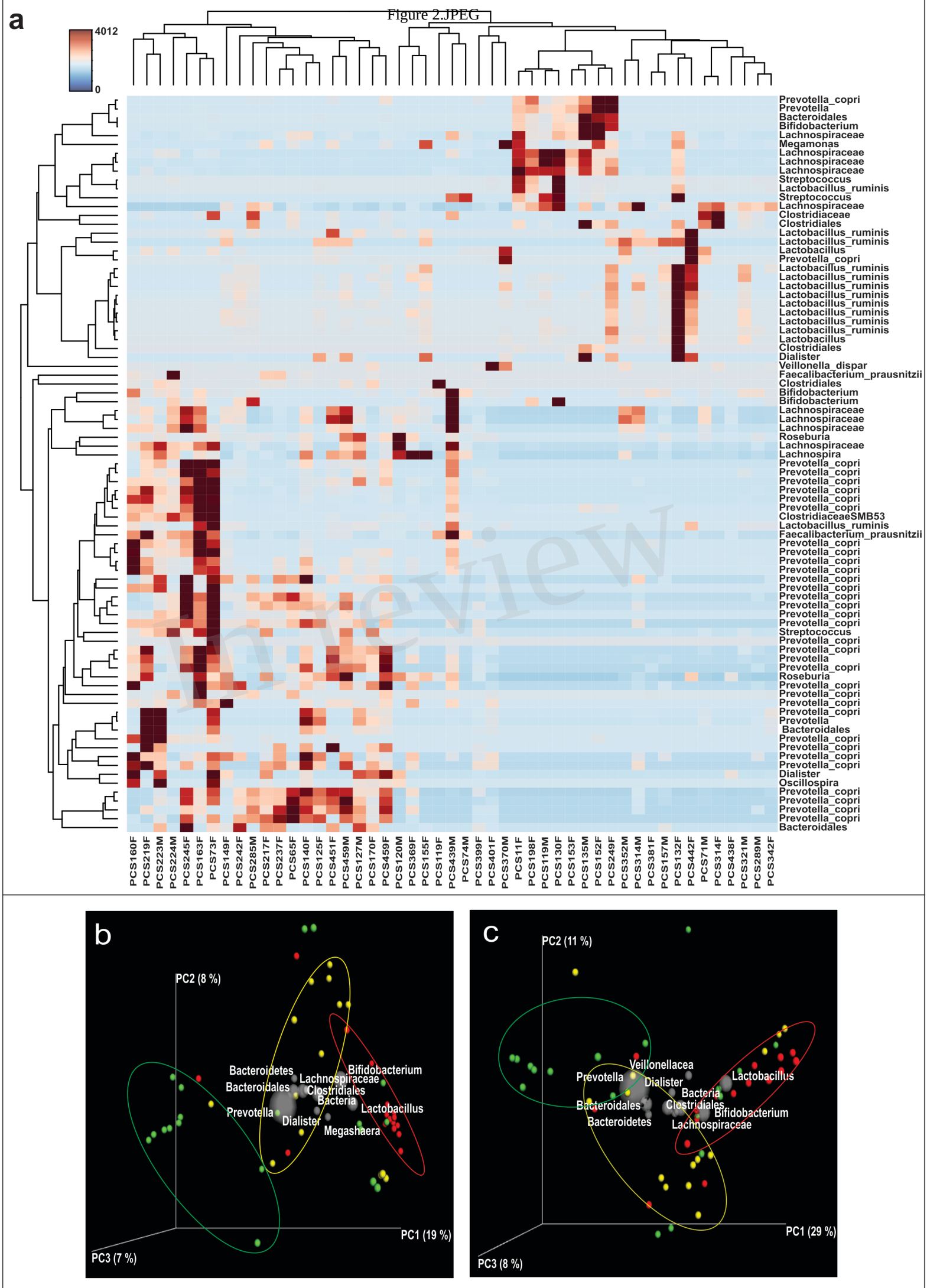


Figure 3.JPEG

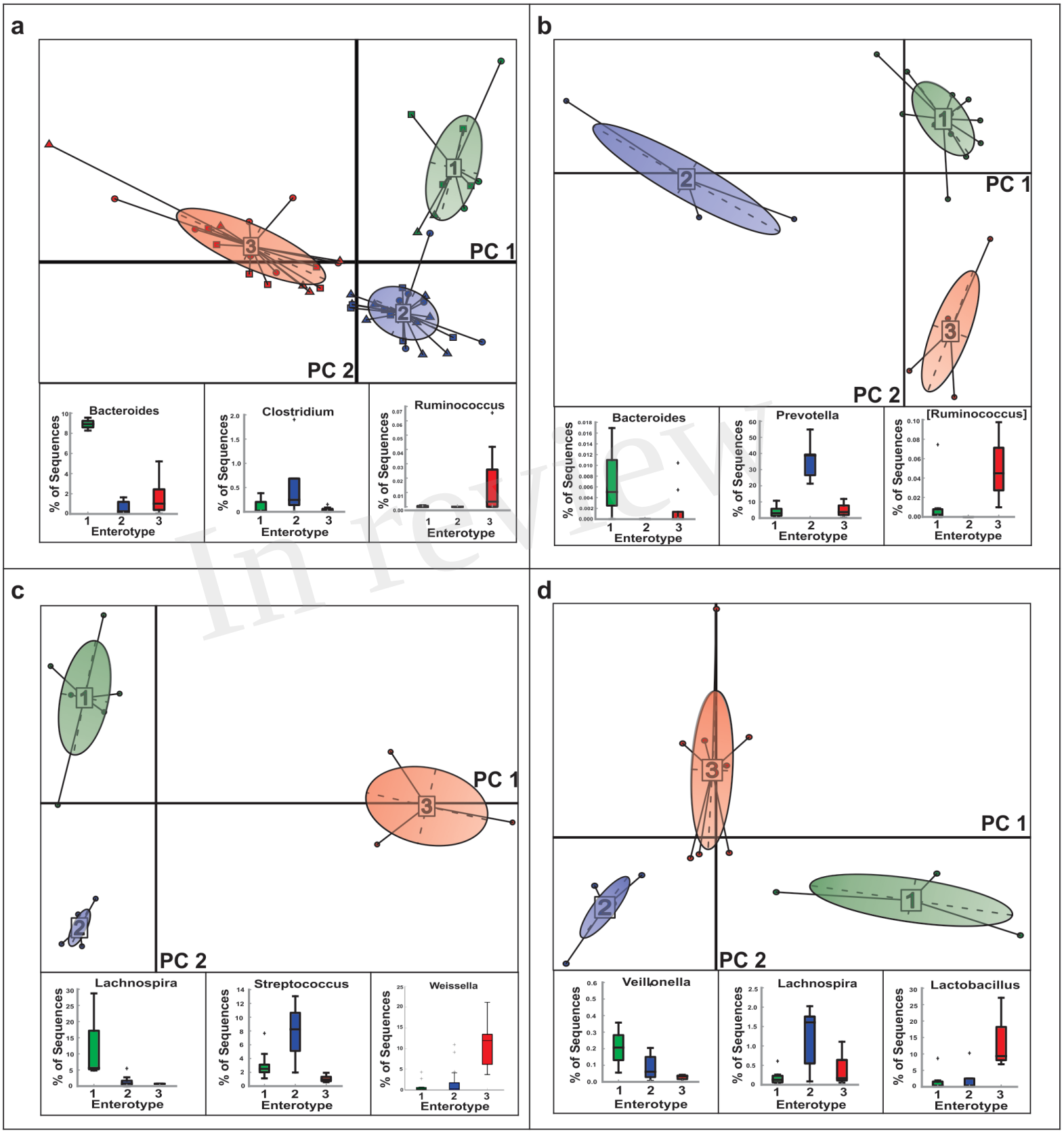


Figure 4.JPEG

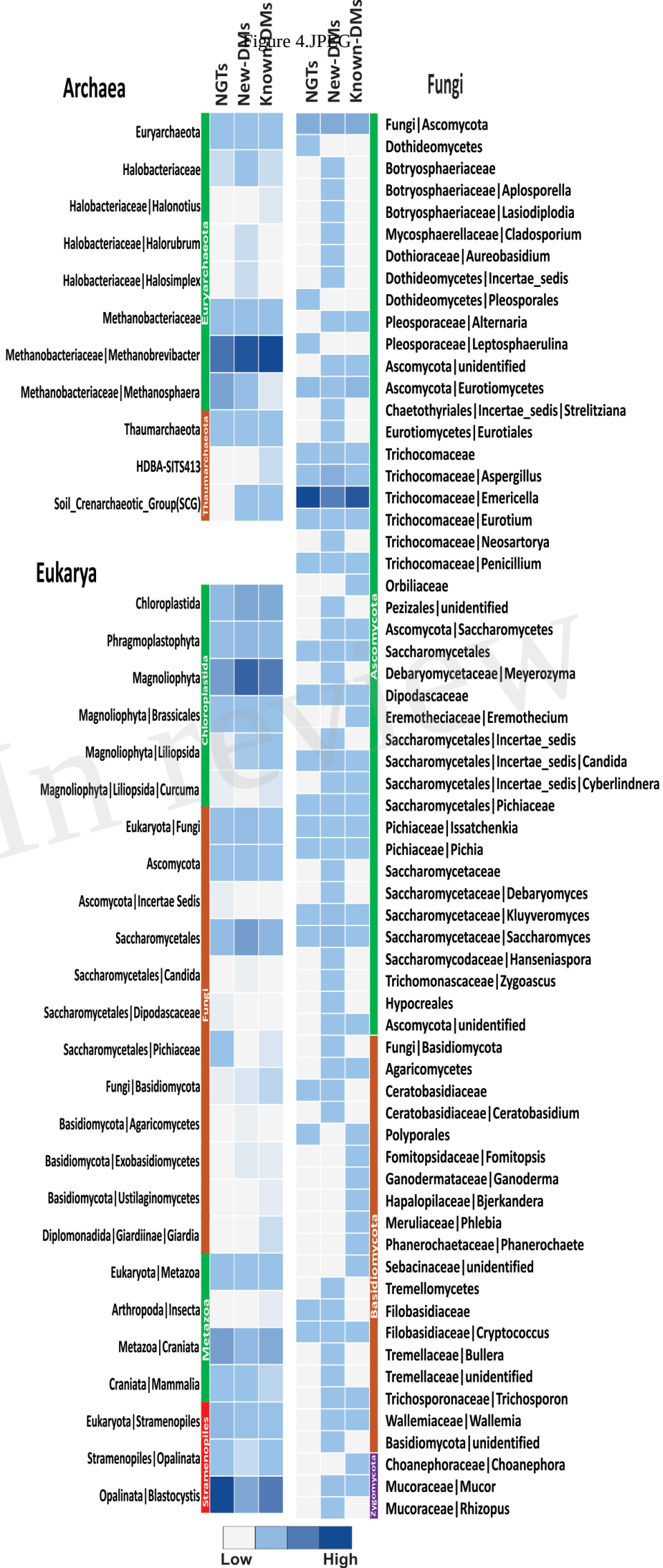
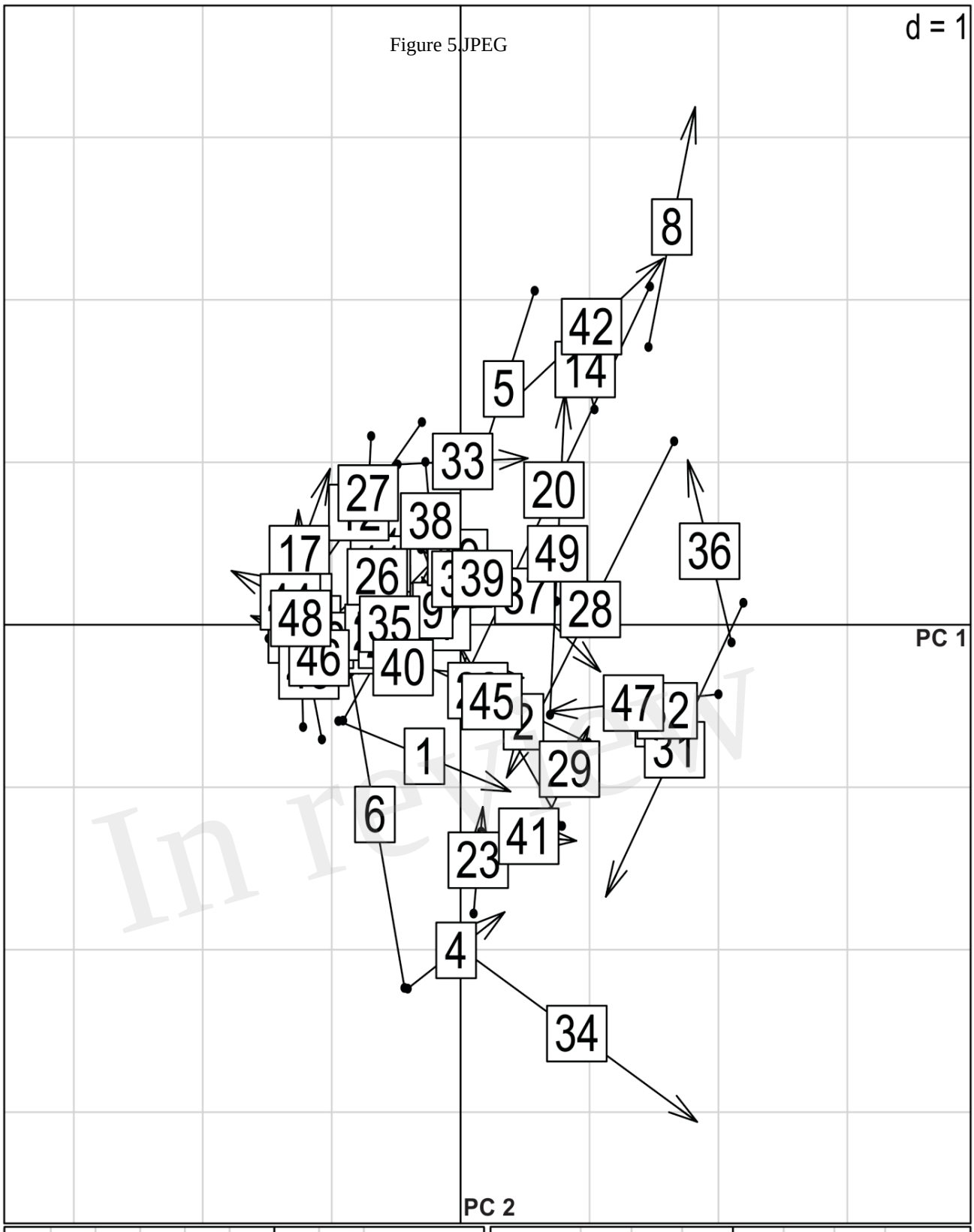
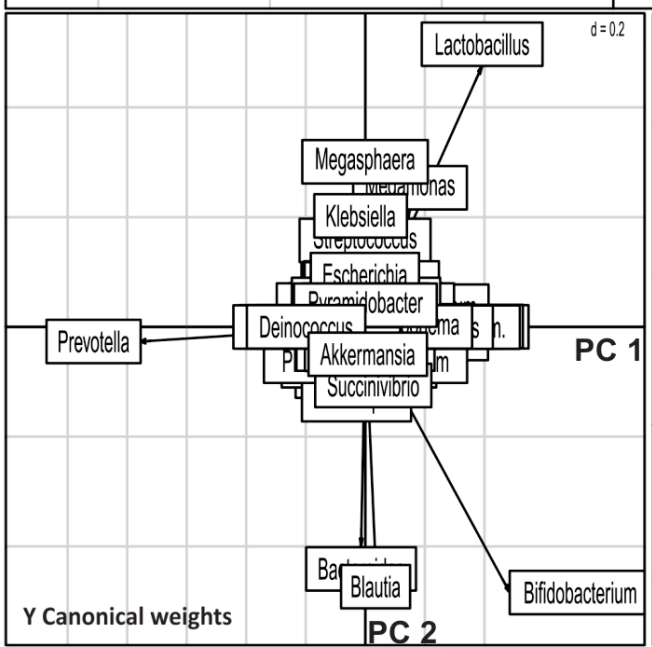


Figure 5.JPEG





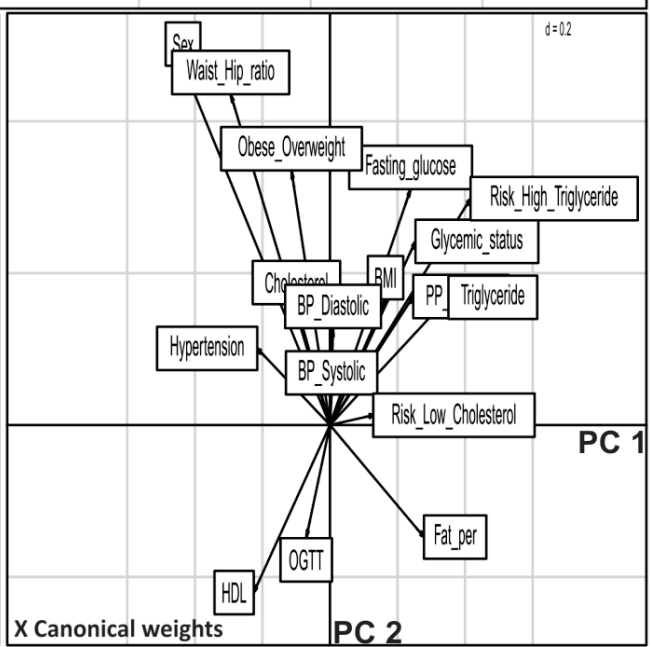


Figure 6.JPEG