


RESEARCH ARTICLE

Open Access



Gut microbiome profiling of a rural and urban South African cohort reveals biomarkers of a population in lifestyle transition

O. H. Oduaran^{1,3*} , F. B. Tamburini², V. Sahibdeen³, R. Brewster⁴, F. X. Gómez-Olivé^{5,6}, K. Kahn^{5,6}, S. A. Norris^{7,8}, S. M. Tollman^{5,6}, R. Twine⁵, A. N. Wade⁵, R. G. Wagner^{5,6}, Z. Lombard³, A. S. Bhatt^{2,4,9} and S. Hazelhurst^{1,10*}

Abstract

Background: Comparisons of traditional hunter-gatherers and pre-agricultural communities in Africa with urban and suburban Western North American and European cohorts have clearly shown that diet, lifestyle and environment are associated with gut microbiome composition. Yet, little is known about the gut microbiome composition of most communities in the very diverse African continent. South Africa comprises a richly diverse ethnolinguistic population that is experiencing an ongoing epidemiological transition and concurrent spike in the prevalence of obesity, largely attributed to a shift towards more Westernized diets and increasingly inactive lifestyle practices. To characterize the microbiome of African adults living in more mainstream lifestyle settings and investigate associations between the microbiome and obesity, we conducted a pilot study, designed collaboratively with community leaders, in two South African cohorts representative of urban and transitioning rural populations. As the rate of overweight and obesity is particularly high in women, we collected single time-point stool samples from 170 HIV-negative women (51 at Soweto; 119 at Bushbuckridge), performed 16S rRNA gene sequencing on these samples and compared the data to concurrently collected anthropometric data.

Results: We found the overall gut microbiome of our cohorts to be reflective of their ongoing epidemiological transition. Specifically, we find that geographical location was more important for sample clustering than lean/obese status and observed a relatively higher abundance of the *Melainabacteria*, *Vampirovibrio*, a predatory bacterium, in Bushbuckridge. Also, *Prevotella*, despite its generally high prevalence in the cohorts, showed an association with obesity. In comparisons with benchmarked datasets representative of non-Western populations, relatively higher abundance values were observed in our dataset for *Barnesiella* (\log_2 fold change (FC) = 4.5), *Alistipes* (\log_2 FC = 3.9), *Bacteroides* (\log_2 FC = 4.2), *Parabacteroides* (\log_2 FC = 3.1) and *Treponema* (\log_2 FC = 1.6), with the exception of *Prevotella* (\log_2 FC = -4.7).

(Continued on next page)

* Correspondence: ovokeraye.oduaran@wits.ac.za; scott.hazelhurst@wits.ac.za

¹Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: Altogether, this work identifies putative microbial features associated with host health in a historically understudied community undergoing an epidemiological transition. Furthermore, we note the crucial role of community engagement to the success of a study in an African setting, the importance of more population-specific studies to inform targeted interventions as well as present a basic foundation for future research.

Keywords: 16S, African microbiome, South African microbiome, Obesity, Transitional microbiome, Epidemiological transition

Background

There have been relatively few studies of the human gut microbiome in Africa, with most reported studies to date focusing on the extremes of non-Western traditional hunter-gatherer and agriculturalists African populations, as well as children with nutritional deficiencies [1–7]. A consistent finding of these studies is the inverse relationship in the relative abundance of *Bacteroides* and *Prevotella* genera of the *Bacteroidetes* phylum. *Prevotella* is associated with plant-based diets predominantly in non-Western populations, whereas increased relative abundance of *Bacteroides* is thought to result from animal fat- and protein-based diets [3, 7–11]. These studies have been vital in providing great insight into the microbiome of traditional African populations and pioneering the efforts of microbiome studies on the continent. It is important to note that across most of sub-Saharan Africa, although the lifestyle has been dominantly agricultural for at least 1000 years [12], relatively few people practice hunter-gatherer or pastoralist lifestyles. However, over the last 50 years in particular, there has been an epidemiological transition toward more industrialized and sedentary lifestyles, that has had significant impact on many Africans.

The role of the microbiome in areas of public health has also been a study focus area on the African continent. These include nutrition, vaccine response efficacy, the impact of antibiotics, mental health and human immunodeficiency virus (HIV) [13–16]. Obesity, a growing health burden [17] on the African continent, has received comparably less attention from microbiome researchers. In a ground-breaking effort, however, the first study on type 2 diabetes (T2D), a comorbidity of obesity, on a sub-Saharan African population [18], provided some insight into the association of gut microbial profiles to T2D in individuals in an urban African setting. The dramatic increase in the prevalence of obesity has been attributed, in part, to the ongoing shift on the continent towards more Westernized practices, such as the consumption of more animal-based and processed products with increasing physical inactivity [19–21], further complicating the existing challenge of malnutrition facing the continent [22, 23]. This is reflected in an analysis of demographic and health survey data from 24

African countries [17] where the prevalence of overweight and obesity among women increased in all 24 countries with either a doubling or tripling in the incidence of obesity reported in 50% of the surveyed countries. Pertinent to this study are the statistics indicating black South African women to have the highest prevalence of obesity (42%) within sub-Saharan Africa [24] with general continental body mass index (BMI) trends showing a decline in the underweight population with a concomitant increase in the overweight and obese population [25–27]. The implication of this is the potential increase in the prevalence of comorbidities including diabetes and other cardiometabolic diseases augmenting the health and economic burden in African societies [28–30]. Reports have also alluded to the influence of the growing globalization trend, its concurrent urbanization and consequent dietary implications on otherwise rural areas in South Africa [31–36]. This is reflected in the increasing numbers and proximity of supermarkets and fast food outlets in these areas [31, 33].

Globally, several studies have focused on understanding the apparent dysbiosis observed in obesity [37, 38]. African populations have, however, been understudied in these efforts. Consequently, there is a paucity of data within Africa comparing the gut microbiota of obese individuals to their leaner counterparts. This is crucial, as differences in dietary and environmental exposures may render findings in non-African populations poorly generalizable to the African context, especially with the ongoing epidemiological transition in Africa [4, 39, 40].

Here, we present a study that investigated the gut microbial composition of two South African cohorts with some insight into the microbial compositional differences between obese and lean individuals in the changing microbiota landscape. South Africa, with its diverse ethnolinguistic groups, presents a unique opportunity to study the effects of this continent-wide transition on the gut microbiome. With obesity being an established risk factor in cardiometabolic diseases, understanding the differences observed between obese and lean individuals in this setting could prove critical to improving our understanding of its association to the pathogenesis of the disease.

This pilot study was nested in the AWI-Gen project [41], a part of the Human Heredity and Health in Africa

(H3Africa) [42] initiative. AWI-Gen is a collaborative effort, with participants in six sites across four African countries, established to assess genomic and environmental factors that influence cardiometabolic diseases risk, with the aim of informing treatment and intervention strategies. The study focused on characterizing the gut microbiome of female adults, with body mass indices spanning the lean and obese range, from two cohorts comprising communities across two South African provinces, Gauteng and Mpumalanga, representative of relatively urban and transitioning rural lifestyles respectively. These cohorts are managed by established health and demographic surveillance sites (HDSS) in partnerships with the University of the Witwatersrand (Wits) and the Medical Research Council (MRC) of South Africa. The Agincourt HDSS [35] in Mpumalanga encompasses a collection of rural communities in the Bushbuckridge municipality undergoing rapid epidemiological changes which may allow for some of the areas to be classified as peri-urban. The Developmental Pathways for Health Research Unit (DPHRU) in Gauteng, on the other hand, is focused on Soweto, a highly urbanized area in the Johannesburg metropolitan area. Soweto has been urbanized for many generations even though in-migration remains at a high level.

In this study, we performed 16S rRNA gene analysis of the gut microbiome of 170 female individuals in Bushbuckridge and Soweto. We evaluated the overall microbial composition of the sampled data to improve our knowledge of the general microbiota landscape of these representative cohorts and assessed compositional differences in the microbiome between lean and obese individuals, using BMI values, within and between Bushbuckridge and Soweto. We also provide insight into the feasibility of such studies in rural communities whilst highlighting the importance of community engagement to this effort.

Results

Participant recruitment and study cohort

With ethics approval from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (M160121) and the Provincial Health Research Committee of the Province of Mpumalanga (MP2017TP22851), 132 female individuals from Bushbuckridge (24.8398° S, 31.0464° E) and 58 from Soweto (26.2485° S, 27.8540° E) were recruited for the study. However, only 170 participant samples (Bushbuckridge: 119, Soweto: 51) were included in the study due to confounding factors to the focus of this pilot (18 HIV-positive samples and two samples with collection irregularities were excluded). The age and BMI distribution of the cohorts are shown in Table 1.

Table 1 Age and BMI distribution of cohorts

	Mean ± SD	Median	Range
Age (years)			
Bushbuckridge	55.50 ± 7.77	55	43 - 72
Soweto	54.10 ± 5.86	54	43 - 64
BMI (Kg/m²)			
Bushbuckridge	32.6 ± 7.95	31.17	21.23 - 58.97
Soweto	36.05 ± 9.23	36.52	20.43 - 58.62

Pre-processing and quality control

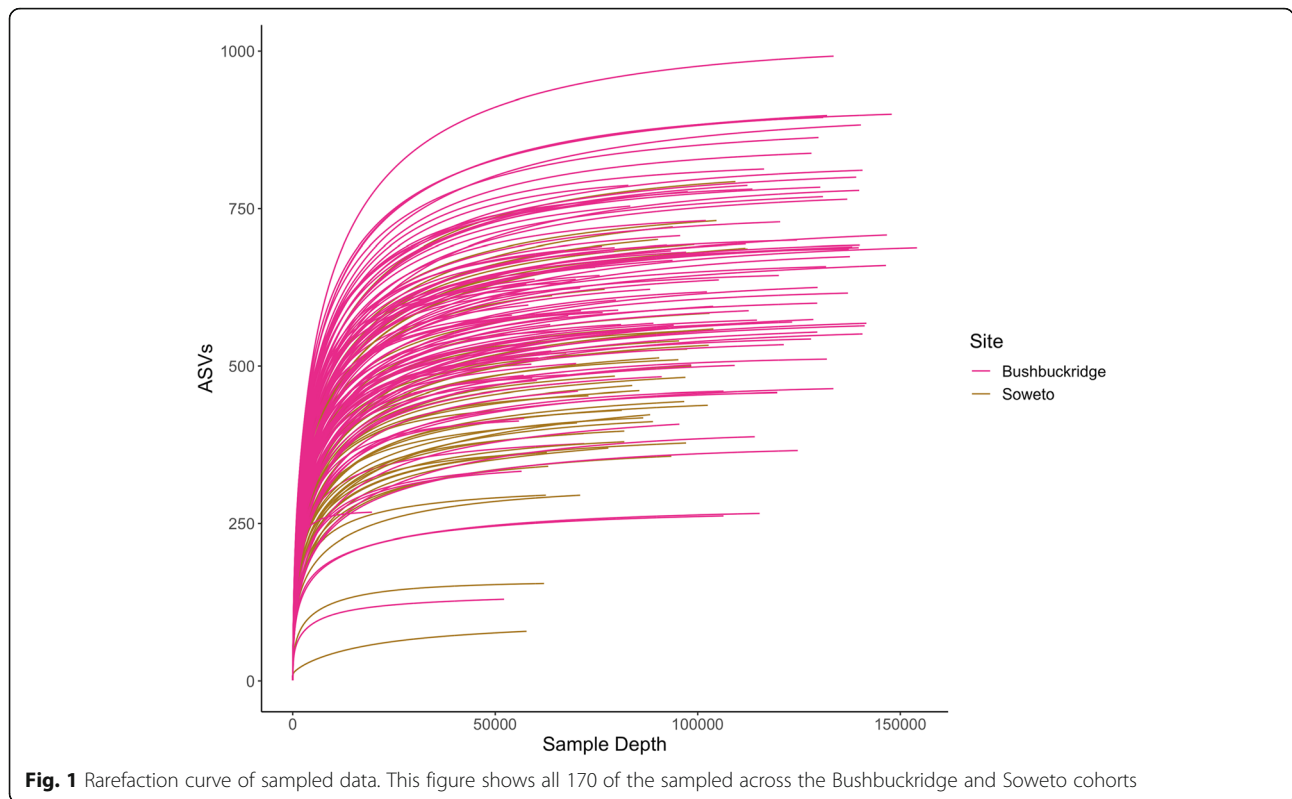
This was primarily done with the DADA2 pipeline [43]. 16S rRNA gene sequencing was performed with primers to the V3 and V4 regions. A total of 15,839,081 sequences were obtained from the 170 samples after quality control. The sequence depths ranged from 2 to 154,124 reads per sample (Supplementary Table 1), with a mean of $93,171.06 \pm 2275.40$ and a median of 93,066, resulting in a total of 10,088 unique amplicon sequence variants (ASVs) with redundant taxonomies. As a result of relatively low sampling depths, the spread of the read depths and the likelihood that the richness of the samples was not fully observed at their sequenced depths, three samples with fewer than 19,560 reads were excluded from downstream analyses (Fig. 1). The implication of this exclusion is an overall minimum sequence depth of 50,812 reads for the 167 samples. The dataset was further pruned to remove taxa not seen more than three times in at least 5 % of the 167 samples in order to protect against ASVs with small mean and trivially large coefficients of variation [44]. This resulted in 1688 ASVs being used as input for beta diversity and the differential abundance analysis implemented with DESeq2 [45]. The taxonomies associated with the corresponding ASVs accounted for two kingdoms (*Archaea* and *Bacteria*) resulting in 14 phyla, 25 classes, 30 orders, 54 families, 124 genera and 111 species, with unclassified ASVs also detected at all but the kingdom levels (Table 2). These numbers represent non-redundant taxa.

Microbial community richness estimates and differences

With the majority of diversity metrics being sensitive to varying sequencing depths across samples [46], rarefaction was done at a read depth of 50,800 to maximize the capture of the observed microbial taxa richness in the cohort. This cut-off was chosen based on the spread of the read depths as visualized in the rarefaction plot in Fig. 1. The rarefied dataset was used for the alpha diversity analyses.

Site differences

In a cohort-wide comparison to evaluate overall differences between the Bushbuckridge and Soweto sites irrespective of BMI status, statistically significant *p*-values



were observed for alpha diversity measures of both Shannon [47] ($p = 0.012$) and Chao1 richness ($p < 0.001$) [48] (Fig. 2), and the Bray-Curtis dissimilarity measure ($p = 0.001$), visualized in principal coordinate analysis (PCoA) [49] plots (Fig. 3). We find that geographical location was more important for sample clustering than lean/obese status. The PCoA plots also present a moving divide between rural Bushbuckridge and urban Soweto. This appears to reflect a transitional state possibly owing to gradual lifestyle and dietary changes.

BMI differences

In evaluating the potential diversity across BMI categories, Shannon diversity, a measure of richness and

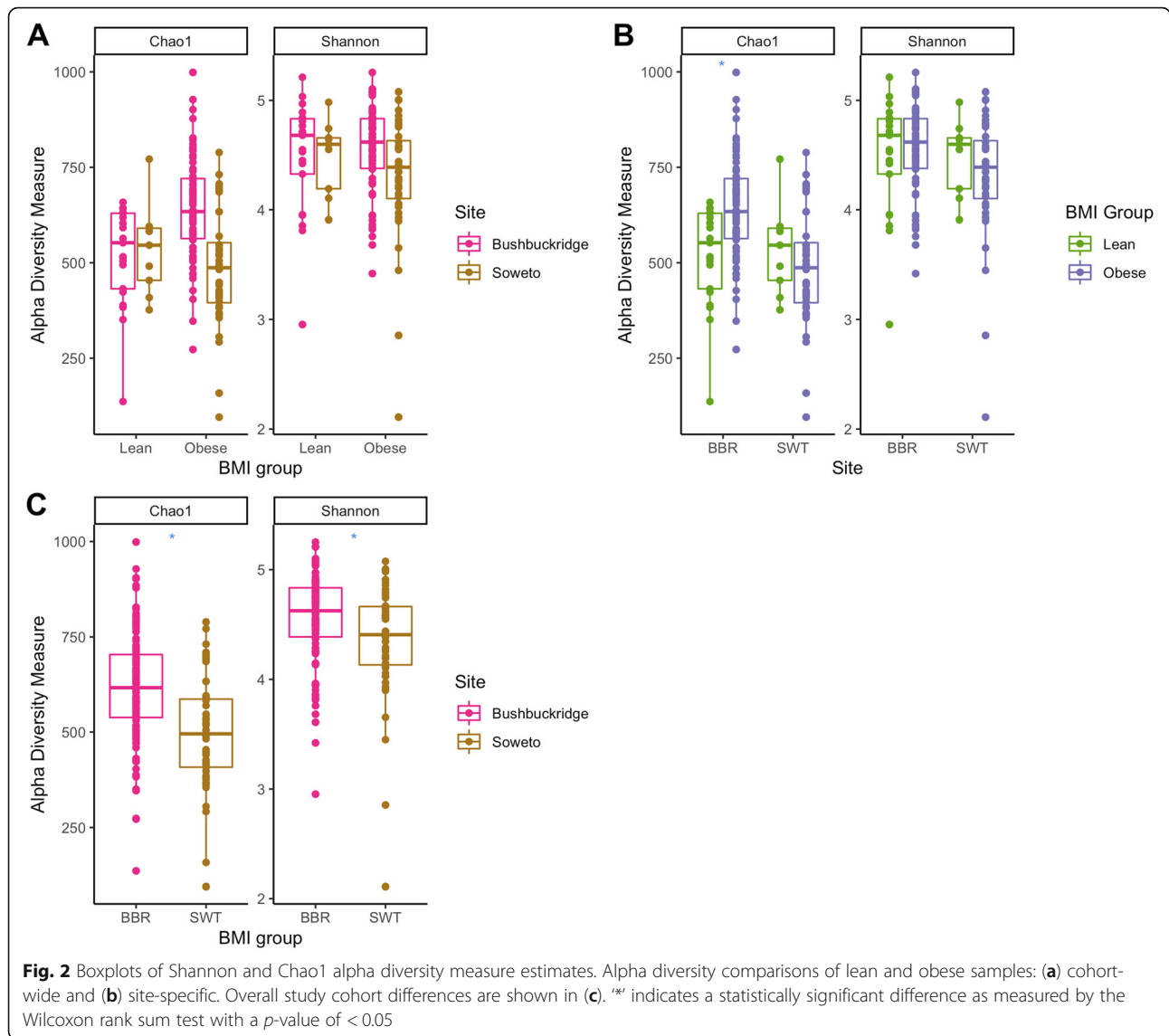
evenness, for the lean and obese groups in Bushbuckridge (Fig. 2b) were 4.49 ± 0.53 and 4.56 ± 0.39 , respectively. The exclusion of an apparent outlier in the Bushbuckridge lean group resulted in a Shannon index of 4.56 ± 0.41 in that group. The corresponding estimates for Soweto were 4.49 ± 0.34 (lean) and 4.30 ± 0.56 (obese). The differences between the lean and obese groups did not reach statistical significance as indicated by the non-parametric Wilcoxon rank sum test evaluating the Shannon diversity values between both groups ($p = 0.85$ and 0.45 for Bushbuckridge and Soweto respectively). Beta diversity measurements (Fig. 3), however, showed statistically significant differences between the lean and obese groups in Bushbuckridge with calculated Bray-Curtis distances using the permutational analysis of variance (PERMANOVA) test ($p = 0.02$ for Bushbuckridge and $p = 0.84$ for Soweto (Table 3).

Table 2 Distribution of taxonomic classification of filtered ASVs in sampled South African pilot dataset

Taxa Level	Classified ASVs	Associated Taxa	% Unclassified ASVs
Kingdom	1,688	2	0.00%
Phylum	1,668	14	1.18%
Class	1,638	25	2.96%
Order	1,628	30	3.55%
Family	1,444	54	14.45%
Genus	1,114	124	34.00%
Species	161	111	90.46%

Taxonomic analyses

Overall, *Firmicutes* ($43.7\% \pm 11.8\%$), *Bacteroidetes* ($40\% \pm 12.1\%$) and *Proteobacteria* ($12.5\% \pm 9.1\%$) were the dominant phyla observed in the combined gut microbiome data from these two South African cohorts (Fig. 4). Three phyla – *Actinobacteria* ($p < 0.001$), *Bacteroidetes* ($p = 0.001$), *Proteobacteria* ($p < 0.001$) and three genera – *Alistipes* ($p < 0.001$), *Bacteroides* ($p < 0.001$) and *Parabacteroides* ($p < 0.001$), showed significant differences in relative



abundance values between the two cohorts based on Kruskal-Wallis (KW) *p*-values.

The noticeably higher relative *Bacteroides*’ abundance (17.1% ± 15.%) observed in urban Soweto in comparison with Bushbuckridge (9.8% ± 11.4%) together with the presence of *Alistipes*, *Anaeroplasma* and *Barnesiella* amongst the most abundant genera is in line with the association of these taxa with non-Western populations in literature (Fig. 4d, e and f) [10, 40, 50]. These associations have been hypothesized to be driven by diet [3, 51, 52]. Of note, within-cohort taxonomic comparisons between lean and obese individuals did not reveal any significant differences at both phyla and genera levels.

Microbial compositional analyses

To better understand the contribution of lifestyle to microbiome composition in this pilot study, the DESeq2

[45] method was applied to further evaluate potential compositional differences in the South African cohorts. To accomplish this at site level, the data was first sub-setted to exclude the intermediate, overweight samples, while keeping only the lean (Bushbuckridge: 21, Soweto: 9) and obese samples (Bushbuckridge: 66, Soweto: 40).

Cohort-wide analysis

Differential abundance analysis revealed a general high prevalence of *Prevotella* in the South African dataset. Also present in the cohorts were *Phascolarctobacterium* and *Vampirovibrio*, which was observed primarily in the Bushbuckridge cohort (Fig. 5a and e; Supplementary Tables 2A and E). *Alistipes*, a genus associated with Western populations, showed significantly higher differential abundance in Bushbuckridge (Fig. 5a; Supplementary Table 2A). Some of the other taxa associated with

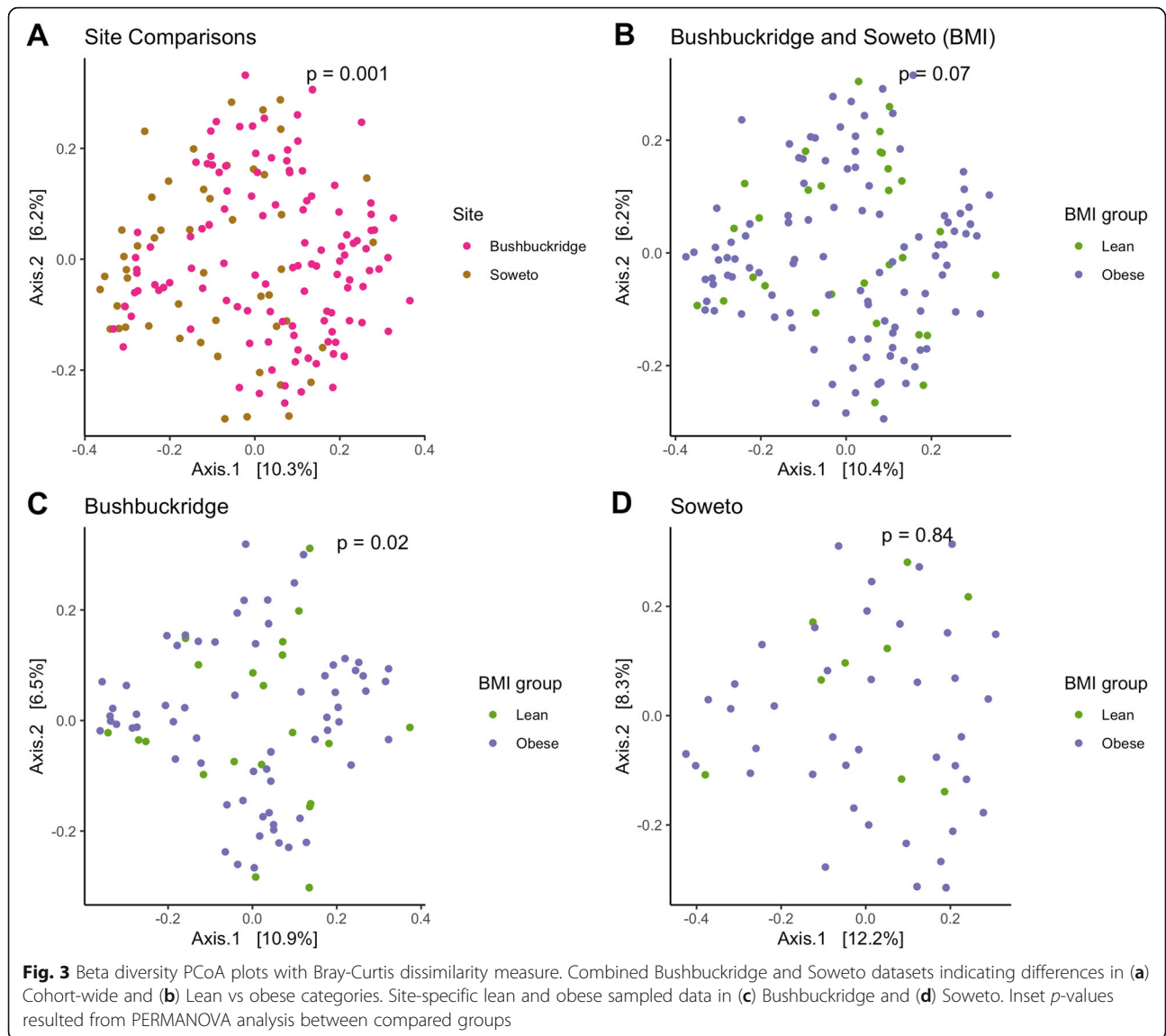
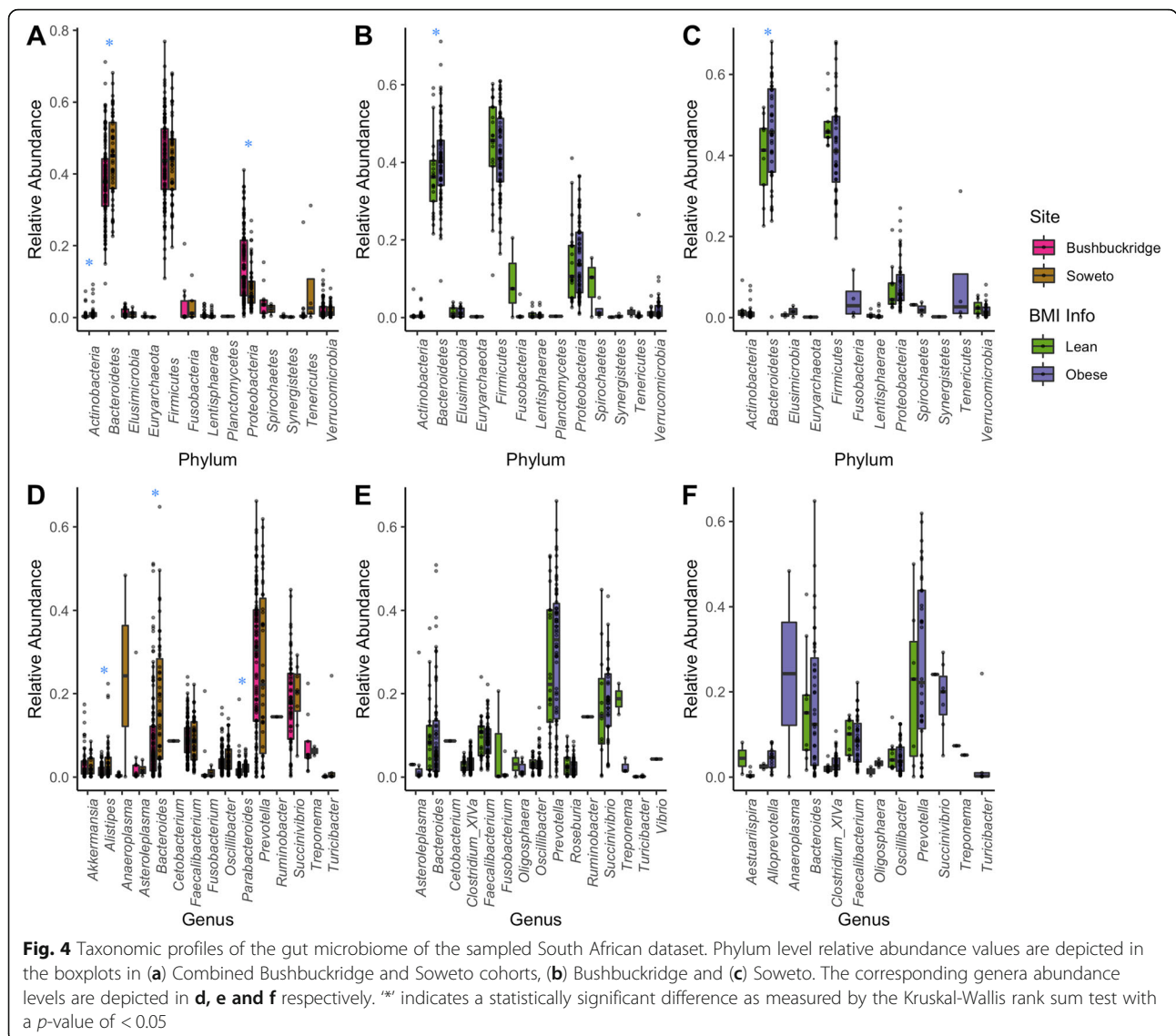


Table 3 Alpha and beta diversity significance of compared groups. Alpha diversity *p*-values were calculated with pairwise Wilcoxon rank sum test. Bray-Curtis diversity *p*-values were calculated with PERMANOVA

Sample Distribution	Group Comparisons	Shannon <i>p</i> -values	Chao1 <i>p</i> -values	Bray-Curtis <i>p</i> -values
Bushbuckridge and Soweto	Lean vs Obese	0.72	0.06	0.07
Bushbuckridge and Soweto	Lean vs Lean	0.69	0.72	0.01
Bushbuckridge and Soweto	Obese vs Obese	0.01	< 0.001	0.001
All samples	Bushbuckridge vs Soweto	0.01	< 0.001	0.001
Bushbuckridge	Lean vs Obese	0.85	0.001	0.02
Soweto	Lean vs Obese	0.45	0.33	0.84

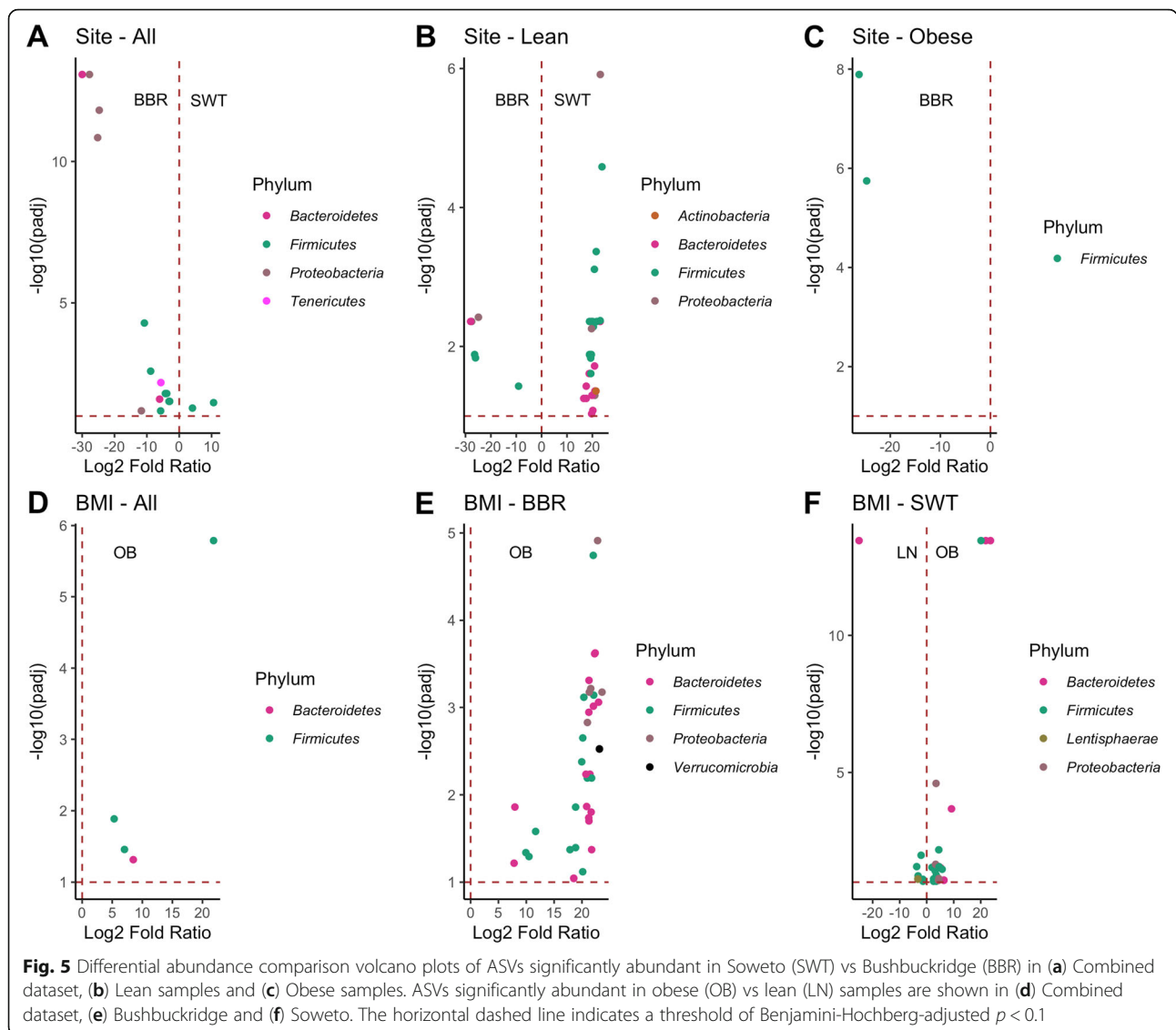


Bushbuckridge include the flavonoid-degrading *Flavonifractor*, *Parasutterella*, *Gemmiger*, and *Dialister* [48] (Fig. 5a and c; Supplementary Tables 2A and C). Soweto samples, on the other hand, showed a significant enrichment in *Bifidobacterium*, the oxalate-metabolizing *Oxalobacter* [53, 54], *Barnesiella*, *Acetanaerobacterium*, *Roseburia*, *Escherichia/Shigella* and *Streptococcus* (Fig. 5a, b and f; Supplementary Tables 2A, B and F).

Comparing the microbiomes of the combined obese groups (Bushbuckridge and Soweto) with their leaner counterparts revealed butyrate-producing *Intestinimonas* [55] and *Prevotella* to be more abundant in the obese category with log₂fold changes of 5.32 and 8.50 respectively (Fig. 5d; Supplementary Table 2D).

Site-specific analysis

Notably, *Prevotella* was found to be associated with obesity. This was clearly observed in Bushbuckridge, where *Prevotella* showed a higher relative abundance in the obese group (Fig. 5d, e and f; Supplementary Tables 2D, E and F). Also observed to be in higher abundance in the Bushbuckridge obese group were 36 ASVs representative of 11 unique genera which include *Prevotella* (12), unclassified genera (10), *Sutterella* (3), *Phascolarctobacterium* (2), *Ruminococcus* (1), *Clostridium_IV* (1), *Alistipes* (1), *Acetanaerobacterium* (1), *Parabacteroides* (1), *Catenibacterium* (1) and *Akkermansia* (1) (Fig. 5e; Supplementary Table 2E). The numbers in parenthesis are the associated ASVs. In Soweto, 24 ASVs, representative of 12 genera, were associated with the obese group while seven ASVs representative of four genera



presented higher abundance levels relative to their leaner counterparts. The obese group-associated genera are *Prevotella* (6), *Clostridium_XIVa* (3), *Haemophilus* (3), *Oscillibacter* (2), unclassified genera (2), *Clostridium_XIVb* (1), *Streptococcus* (1), *Escherichia/Shigella* (1), *Ruminococcus* (1), *Sporobacter* (1), *Oxalobacter* (1), *Intestinimonas* (1) and *Parabacteroides* (1). The genera associated with the lean group in Soweto are *Parabacteroides* (1), *Victivallis* (1), *Fusicatenibacter* (1) and unclassified genera (3) (Fig. 5f).

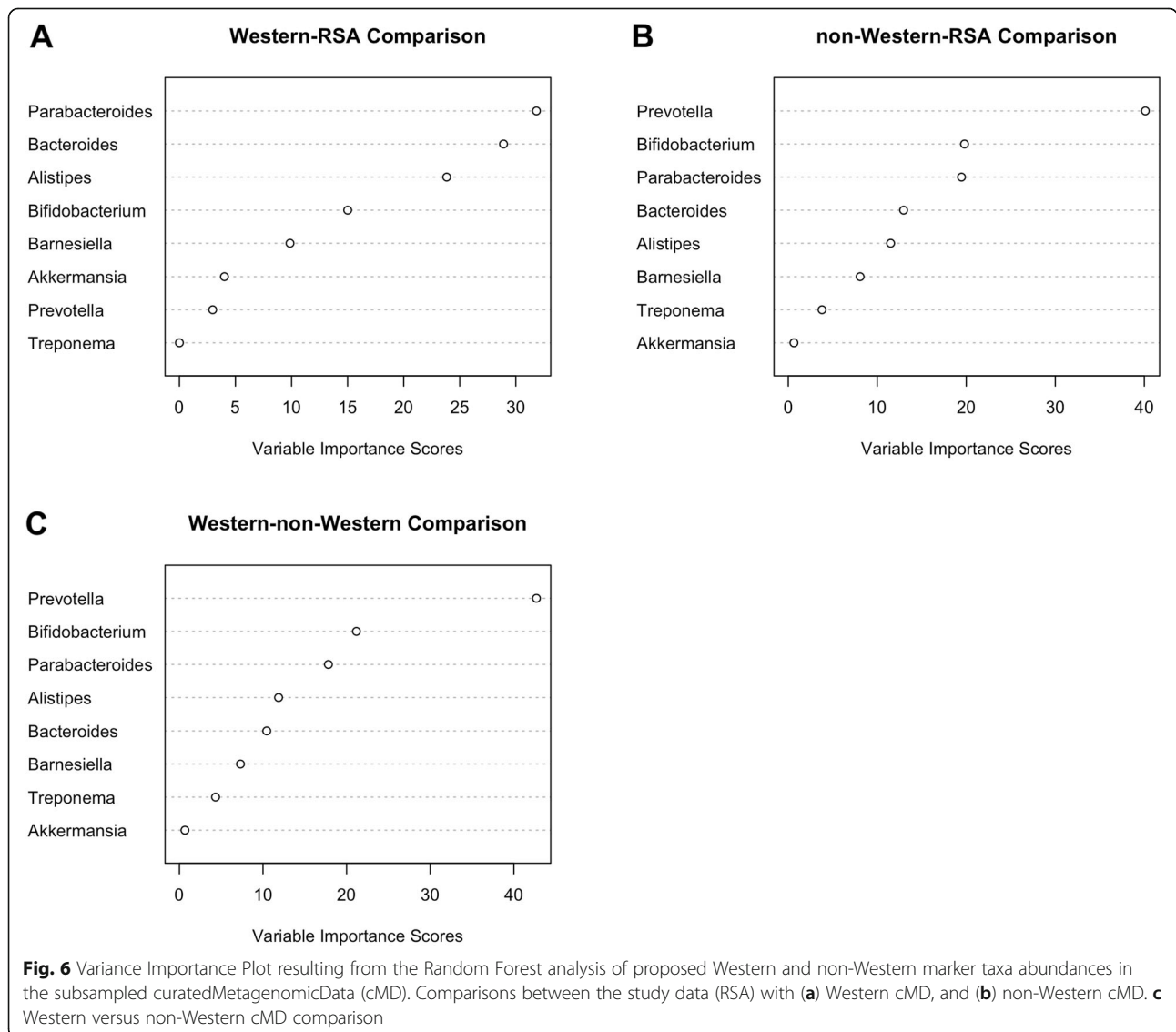
The apparent site-specific association of *Prevotella* to the obese group in Bushbuckridge is in line with literature linking the taxon to obesity [38, 56, 57], although there have also been some contradictory reports [1, 2].

Marker taxa analyses

A recent meta-analysis examined differences between the gut microbial composition of traditional, rural populations and their more industrialized counterparts from several studies with datasets encompassing 13 developed or industrialized societies and two traditional hunter-gatherer, pre-agricultural communities [3, 4, 8, 58, 59]. The study proposed a marker taxa list distinguishing Western and non-Western bacterial communities. This was corroborated by de la Cuesta-Zuluaga, et al. [60] by the analysis of 16 benchmark datasets with the Bioconductor package, curatedMetagenomicData (cMD) [61]. The cMD is a collection of processed data from whole-metagenome sequencing for thousands of human microbiome samples across different body sites.

To further evaluate the landscape of our study data with respect to the established population-dependent compositional expectations, we randomly selected 334 individuals from the cMD, 167 of whom were from populations of Western origin and the remaining 167 from traditional non-Western populations to match the number of samples in our dataset. The sampling was done from a total of 23 studies with 1763 samples (1433 Western and 330 non-Western) in the cMD. We compared the abundance values of Western-associated (*Alistipes*, *Akkermansia*, *Barnesiella*, *Bifidobacterium*, *Bacteroides* and *Parabacteroides*) and non-Western-associated (*Treponema* and *Prevotella*) marker taxa to their corresponding abundance profiles in our dataset. This was done by testing the null hypothesis that the mean ranks of the abundances of these marker taxa were the same in the subsampled cMD and our sampled

cohorts using the non-parametric Kruskal-Wallis test. Our results rejected the null hypotheses for all ($p < 0.001$) but three taxa, *Akkermansia*, *Barnesiella* and *Treponema* with $p > 0.1$ when compared to corresponding Westernized datasets. Comparisons with the non-Western dataset, on the other hand, resulted in the rejection of the null hypothesis for all but one taxon, *Treponema* ($p = 0.52$). We found the abundances of *Alistipes*, *Bacteroides*, *Prevotella*, and *Parabacteroides* in our data to be intermediate between the benchmarked Western and non-Western datasets, and the abundance of *Barnesiella* comparable to that in the Western microbiota (Table 5). In addition, Random Forest analysis comparing the South African cohorts to the subsampled cMD presented *Prevotella* and *Parabacteroides* as the most important discriminatory taxa in the non-Western and Western datasets comparisons respectively (Fig. 6a



and b). Interestingly, the importance scores associated with each taxon in the classification of the subsampled non-Western cMD with our dataset is comparable to the associated taxa scores in the classification of the cMD's Western and non-Western datasets (Fig. 6c). Altogether, these results reinforce the notion of a gradually changing microbial composition of the sampled cohort relative to the subsampled curated datasets.

Discussion

This study aimed to characterize the gut microbiome of two South African cohorts from two sites, about 483 km (300 miles) apart that represent relatively urban and transitioning rural lifestyle and diet-practicing populations, whilst exploring the microbial compositional differences observed in obese and lean individuals. To accomplish this, we collaboratively designed a study with active input from the community, in conjunction with a community advisory group (CAG) at Bushbuckridge. Although the community was familiar with the general research process, the concept of stool donation was relatively unfamiliar [62, 63]. Stool collection for microbiome research purposes had never before been carried out in this population. With prevailing traditional beliefs concerning stool carrying the soul, it was crucial to be sensitive and respectful whilst clearly presenting the importance and proposed usage of the stool samples as well as the aims of the research in understandable language. The recruitment process and sample collection for this study thus relied on extensive community engagement.

DNA extracted from the collected stool samples underwent 16S rRNA gene sequencing. We observed relative abundance levels of Western gut-associated marker taxon, *Barnesiella*, that were comparable to Western populations with intermediate abundance levels for *Alistipes*, *Bacteroides*, *Parabacteroides* and *Prevotella* when compared to the benchmarked datasets (Table 5). Within our cohorts, we found *Vampirovibrio*, a predatory *Melainabacteria* to be present with higher relative abundances in the rural samples and *Prevotella*, despite its generally high prevalence relative to all taxa present in the cohort, to be associated with obesity. Overall, we identified putative microbial features associated with host health and highlight the importance of population-specific considerations in microbiome research. Importantly, we also shed some light on the vital role of engaging the community of interest to the success of such studies in an African setting.

Within our cohorts, microbial composition reflected a transitional state comprising both Western- and non-Western-associated taxa. *Prevotella* and *Treponema* represented the traditional hunter-gatherer taxa. *Phascolarctobacterium*, a propionate and acetate producer that

has been shown to exert beneficial effects on its host [64–66], appears to be abundant across both sites. A recent study comparing various industrialized, urban populations to traditional rural societies identified *Phascolarctobacterium* to be the most significant contributing taxa to the non-Western population cluster [64]. A robust meta-analysis study that compared the gut microbiomes of urbanized and pre-agricultural populations also noted it to have relatively low abundance, and in some cases absence, in Western populations [67].

With global research findings on the apparent dysbiosis of the gut microbiome in obesity being inconclusive [38, 39, 68, 69], we sought to evaluate the differences between obese and lean individuals within and between the two study populations. The within site differences were moderate and did not reach statistical significance in Soweto. However, for Bushbuckridge, significant differences were observed for both alpha and beta diversity estimates between the lean and obese groups using Chao1 ($p = 0.001$) and Bray-Curtis ($p = 0.02$) measures. Log₂ fold changes ranging from 7.81 to 23.60 were observed in the differential abundance analyses of component microbial taxa of the obese samples relative to their leaner counterparts resulting in 11 classified genera. *Sutterella* and *Catenibacterium* which have been previously associated with obesity [70, 71], as well as *Clostridium_IV* were among the differentially abundant taxa in the obese samples in Bushbuckridge. *Oscillibacter* was associated with cohort-wide obesity irrespective of site. This association to obesity has been previously reported in a European cohort [72].

Overall, the lean comparisons showed slightly greater diversity than the obese groups with taxa representative of four different phyla and 14 genera (Fig. 5b and c). The PCoA plots comparing lean and obese individuals (Fig. 3b, c and d) appears to show a divide between samples that may not be entirely driven by BMI categories. It is, however, possible that associations with small effect sizes exist in our sampled cohort that could be detected with larger sampling. Also, as limited demographic and dietary data were collected for this pilot, further exploration is warranted.

Of great interest in the Bushbuckridge cohort was the predatory *Vampirovibrio*. Although not very well-studied in humans to date, *Vampirovibrio* is capable of invading and attacking other bacteria without harming human cells. It has been proposed for further studies in bioremediation [73] to reduce the use of antibiotics. *Melainabacteria*, the phylum to which *Vampirovibrio* belongs [74–76], is generally found to be present in aquatic habitats as well as associated with the guts of herbivorous mammals and humans with predominantly plant-based diets. They are also known to synthesize vitamins B and K, which in addition to their fiber-

digesting abilities posits them as beneficial bacteria to their hosts.

Several studies have identified obesity-associated taxa primarily in non-African populations [25, 77, 78] despite these reported connections being inconsistent [1, 2, 72]. The differential abundance, prevalence or presence of microbial taxa across populations may require population-specific associations for relevance, as universal classifications may not necessarily be generalizable. The seemingly ubiquitous presence of *Prevotella* in the sampled cohorts and its association with obesity in Bushbuckridge brings to the fore the role of some *Prevotella* strains as potential pathobionts involved in various human diseases by the promotion of chronic inflammation [79, 80]. Increased abundance of *Prevotella* species at mucosal sites have been linked to several diseases including metabolic disorders and low-grade systemic inflammation [38, 56, 81], a feature associated with obesity. *Prevotella* may thus present as a critical taxon in the obesity pandemic on the African continent. Further in-depth studies to ascertain the influence of its prevalence in a community undergoing such epidemiological transition will be insightful as the beneficial or detrimental effects of *Prevotella* may very likely be dependent on strain variations or its interaction with the prevailing lifestyle and environment [82].

Conclusions

This study provides us with a foundation to inform future microbiome studies in Africa. A clear outcome of this study was the statistically significant differences in microbial composition observed between the Bushbuckridge and Soweto cohorts with the Bushbuckridge cohort harboring relatively more diverse microbiota. This highlights the difference in stages of the cohorts along the continuum of transition, with the gradual lifestyle and dietary shifts towards more Western practices. Such clarity was not consistently achieved statistically for comparisons between the BMI categories considered. However, moderate differences were observed. This could possibly be attributed to the uneven and sparse sampling of the data especially with the lean category in Soweto. Notwithstanding, the core outcome of this analysis does not seem to have been affected as observed in comparisons between the lean populations of both cohorts. Similarly, a lack of inflated significance in differential abundances between the groups compared support the integrity of the study outcome.

We acknowledge that this study was limited by the unavailability of detailed dietary data at the time of sample collection that may have explained some of the observations and extended the scope of the study. No assumptions were made in this regard with the data presented as is. However, there are published reports on the

dietary changes accompanying the urbanization process across rural areas in South Africa [31–33, 36]. Another potential limitation of this study is the aforementioned uneven and sparse sampling of the data, which appears to have been inconsequential on the study outcome. It is important to note that this was a pilot exploratory study that has provided useful insights into the planning and execution of future studies in similar settings.

In broad summary, the compositional taxa of the gut microbiome of the collective ethnolinguistic groups in the cohorts are reflective of an epidemiologically transitional state, and the beneficial or detrimental effects of *Prevotella* are very likely diet- and lifestyle-dependent. Lastly, the largely intermediate abundances of the proposed Western and non-Western distinguishing marker taxa in our data set in comparison with benchmarked datasets substantiates the transitional state of our African cohorts with potential implications for disease pathogenesis and general health status. This accentuates the need for more population-specific studies as findings and translational applications in non-African populations may be poorly generalizable to the African context. Further studies with a larger sampled cohort will be very informative in this regard.

Methods

Community engagement

The research team engaged the community in two interactive sessions during this study - the planning phase and post-preliminary analyses on the data resulting from the collected stool samples. A survey was also conducted on the first 100 participants in Bushbuckridge to get their feedback on the process. Prior to the collection of stool samples for the study, there was interaction with the community in conjunction with a CAG at the Agincourt HDSS (Bushbuckridge), the rural site, which gave input into the process to ensure that sample collection methods were sensitive to the community beliefs and applicable to the existing toilet facilities in the area. This group comprised eight community representatives and indunas (village councillors). The meeting discussions were focused on creating awareness on what the project entailed and the importance of such research in the community, as well as on potential concerns and reactions of community members to stool sample collection and the practicality of such endeavor. Also deliberated on was the role of the trained fieldworker in the recruitment process and the available resources (graphical flyers) to clearly communicate the study aims and usage of the collected stool samples in understandable language to potential participants.

The interactive workshop that followed the preliminary data analysis aimed to reiterate the importance of the study, broadly present some of the initial results and

very importantly, solicit feedback from the community members and participants. As this was a pilot study, it was important to the research team to gauge the level of understanding of the study post-completion in order to inform future studies in this regard.

Recruitment and study cohort

This study is nested in the AWI-Gen project, which is a part of the Human, Heredity and Health in Africa (H3Africa) consortium. AWI-Gen explores genetic and environmental factors in cardiometabolic disorders in African populations with six sites across four countries. The recruitment of participants for this study was done at two of the South African sites – the Bushbuckridge area within the Agincourt HDSS, Mpumalanga (rural) and Soweto, Johannesburg, Gauteng (urban).

Participants were randomly selected from the AWI-Gen cohort within the BMI strata defined below and are in the age range of 43–72 years (Table 1). To minimize confounding effects, male and HIV+ participants were excluded. Participants were divided into three groups based on their BMI values – lean, overweight and obese. The lean group comprised participants with BMI < 25, the overweight group comprised participants with $25 \leq \text{BMI} < 30$ and the obese group had BMI ≥ 30 . Anthropometric (height and weight) and blood pressure measurements were taken at the time of collection, and a rapid HIV test done. We also had extensive other data about participants from previous engagements. The study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (M160121) and the Provincial Health Research Committee of the Province of Mpumalanga (MP2017TP22851).

To facilitate the participant recruitment and sample collection processes, comprehensive information sessions were held with the fieldworker on the study aims and its importance. This was crucial as the recruitment success could be reliant on the fieldworker's ability to effectively communicate these to prospective participants. The fieldworker was also aided by training videos and experience gained from self-collecting personal stool samples to facilitate reliability to the collection process.

Sample collection

Stool samples were collected from consented participants using DNA Genotek's OMNIGene microbial collection and stabilization kit and sent to the laboratory. The stool samples were subsequently aliquoted into cryovials and frozen at -80 degrees Celsius prior to DNA extraction.

DNA extraction and sequencing

Frozen stool samples were thawed on ice. Genomic (total) DNA was extracted using Qiagen's QIAmp

Powerfecal DNA kit and sent to a dedicated core facility for the sequencing of the V3 –V4 hypervariable region of the 16S rRNA gene on the Illumina MiSeq® platform using 341F 5'-CCTACGGGNGGCWGCAG-3' and 805R 5'-GACTACHVGGGTATCTAATCC-3' as forward and reverse primers respectively [83].

Sequence data analyses

The DADA2 (v1.10.1) pipeline [43] was used for pre-processing and performing quality control on the sequences. Briefly, the demultiplexed paired-end sequences were imported into DADA2. Based on the quality plots, the sequences were filtered with a maximum of expected errors of 2 and 4, and sequence lengths of 280 and 240 bases for the forward and reverse reads, respectively, with primers trimmed accordingly. The resulting reads were dereplicated and merged to obtain the full denoised sequence which was used in the creation of a count table containing the abundance values of sequence variants from the sampled data. Chimeras were subsequently removed, and the non-chimeric sequence table was utilized for downstream analyses.

Taxonomic classification

The DADA2 implementation of the naïve Bayesian classifier method was applied in the assignment of taxonomies to the amplicon sequence variants using the *RDP trainset 16* DADA2-formatted reference set from the Ribosomal Database Project (RDP) [84] and a minimum bootstrapping parameter of 50, with pseudo-pooling.

Alpha and Beta diversity analyses

The DADA2 output together with the sample metadata were imported into phyloseq [44] for diversity analysis. Based on the output from the pre-processing step, rarefaction was applied at a sampling read depth of 50,800 to allow for adequate capture of the observed microbial taxa richness in the cohort as diversity metrics are generally sensitive to sample read depths.

First, Shannon [47] and Chao1 [48] alpha diversity estimates for the samples were calculated. This measure was applied to a pairwise Wilcoxon rank sum (Mann-Whitney) test to assess whether the observed ASVs differed significantly ($p < 0.05$) between specified categories. Boxplots were generated to visualize the categorical differences based on the Shannon diversity values. Comparisons were done as indicated in Table 4.

Next, beta diversity between the samples was evaluated using Bray-Curtis dissimilarity distance matrices for PCoA [49] to generate relevant ordination plots. PERM ANOVA analysis was done to test for differences between specified categories (Table 4).

Table 4 Group comparisons evaluated in this study

Groups	Targeted Evaluation	Sample Data (No. of Samples)
Bushbuckridge vs Soweto	Overall site differences	All samples - lean, overweight and obese (167)
Bushbuckridge vs Soweto	Compositional differences in the lean category between sites	Bushbuckridge and Soweto lean samples only (30)
Bushbuckridge vs Soweto	Compositional differences in the obese category between sites	Bushbuckridge and Soweto obese samples only (106)
Lean vs Obese	Cohort-wide BMI compositional categorical differences	Bushbuckridge and Soweto lean and obese samples only (136)
Lean vs Obese	Site-specific BMI compositional categorical differences	Bushbuckridge lean and obese samples only (87)
Lean vs Obese	Site-specific BMI compositional categorical differences	SWT lean and obese samples only (49)

Differential abundance analyses

To evaluate differences in bacterial taxa abundance across BMI categories and sites, a negative binomial generalized linear model (DESeq2) [45] was used. Briefly, raw counts were modelled with a negative binomial distribution and internal adjustment done for “size factors”. This adjustment normalized for differences in sequencing depth between samples. Prior to analyses, the data

Table 5 Marker taxa analysis. Comparisons between the South African (RSA) cohorts data and benchmarked data sets from the curatedMetagenomicData (cMD). (a) cMD Western (W) data vs RSA data and (b) cMD non-Western (NW) data vs RSA data. The Kruskal-Wallis (KW) rank sum test was used in the calculation of the *p*-values

A				
Genus	RSA Median	W-Median	W-log ₂ FC	<i>p</i> -values (KW)
<i>Prevotella</i>	0.543	0.030	4.409	2.20E-16
<i>Treponema</i>	0.116	0.003	5.085	6.30E-01
<i>Bifidobacterium</i>	0.015	0.043	-4.139	3.09E-16
<i>Barnesiella</i>	0.049	0.033	-0.552	1.39E-01
<i>Akkermansia</i>	0.056	0.012	0.251	9.39E-01
<i>Alistipes</i>	0.040	0.118	-2.789	2.20E-16
<i>Bacteroides</i>	0.268	0.384	-1.183	1.08E-08
<i>Parabacteroides</i>	0.043	0.050	-0.719	7.21E-05
B				
Genus	RSA Median	NW-Median	NW-log ₂ FC	<i>p</i> -values (KW)
<i>Prevotella</i>	0.543	0.760	-4.685	1.02E-03
<i>Treponema</i>	0.116	0.001	1.638	5.18E-01
<i>Bifidobacterium</i>	0.015	0.097	-1.159	2.20E-16
<i>Barnesiella</i>	0.049	0.001	4.481	6.28E-14
<i>Akkermansia</i>	0.056	0.001	3.227	9.20E-03
<i>Alistipes</i>	0.040	0.008	3.917	3.59E-03
<i>Bacteroides</i>	0.268	0.020	4.243	2.20E-16
<i>Parabacteroides</i>	0.043	0.006	3.091	2.20E-16

was filtered to exclude taxa that was not observed more than three times in more than 5 % of the 167 samples. This cut-off was chosen with respect to the sample size and the general data sparsity to protect against ASVs with small mean and trivially large coefficients of variation across samples. This resulted in 1688 high abundance ASVs being included in this analysis. DESeq2 models were adjusted for potential batch effects, where applicable, and BMI for the overall site analysis. However, it is highly unlikely that substantial batch effects exist as 14 samples from the first batch that were re-sequenced and compared across the two sequence runs using Bray-Curtis measure indicate the absence of any potentially damaging batch effects (Supplementary Figure 2).

Statistical significance was determined by the Wald’s test with Benjamini-Hochberg corrected *p*-values and significant ASVs above a secondary alpha threshold of 0.1. The results are presented with Volcano plots (Fig. 5 and Supplementary Table 2).

Marker taxa analyses

To establish the status of our sampled cohorts along the continuum of westernization, we sought to compare the relative abundances of proposed Western and non-Western marker taxa as compiled by a recent meta-analysis [67] with the corresponding values in our dataset. The proposed taxa can be used as markers of lifestyle and geographical origin in the chosen public datasets as well as in the South African cohorts.

A total of 23 studies [5, 58, 85–105] with benchmarked Western and non-Western datasets comprising 1763 samples were downloaded from the curatedMetagenomicData [61] repository. The downloaded count data was converted to an ExpressionSet object and imported into phyloseq [44] for downstream analysis. The data was subsetted to include only the eight genera of interest – *Prevotella*, *Treponema*, *Bifidobacterium*, *Barnesiella*, *Akkermansia*, *Alistipes*, *Bacteroides* and *Parabacteroides*. The abundance counts were transformed to relative abundance

values and filtered to retain only ASVs with mean abundance greater than zero. The data was subsequently split by westernization and 167 samples were randomly selected from each of the two groups and merged with the South African (RSA) dataset to give two groups (Western-RSA and non-Western-RSA) of 334 samples each. These two sample groups were utilized for both comparisons between the subsampled cMD and our combined cohort data.

For each group of data, 70% (234) of the samples were used as the training set for Random Forest analysis to compare the two datasets, with the remaining 30% (100) as the test data. Variable Importance Plots were used to visualize the results (Fig. 6). Abundance levels of the selected taxa were also tested for significant differences using the Kruskal-Wallis test (Table 5).

Feedback from participants

The follow-up survey was done on the first 100 participants at Bushbuckridge about 3 months after collection. The survey was conducted telephonically – each person was phoned at least three times. One person refused to participate, and 65 people agreed. The community engagement process is detailed in the [Supplementary data section](#).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12866-020-02017-w>.

Additional file 1: Supplementary Figure 1. Beta diversity PCoA plots with Bray-Curtis dissimilarity measure. Combined Bushbuckridge and Soweto datasets indicating differences in (A) Cohort-wide and (B) Lean vs obese categories. Site-specific lean and obese sampled data in (C) Bushbuckridge and (D) Soweto. Ellipses represent a 0.95 confidence interval.

Additional file 2: Supplementary Figure 2. Batch-control test. To control for batch effects from different sequencing runs, 14 samples from the first batch were re-sequenced. Comparison of the samples from the two sequence runs using Bray-Curtis measure indicates the absence of any potentially damaging batch effects.

Additional file 3: Extended information on the community engagement process. **Supplementary Table 1.** Sample reads tracked through the pre-processing steps. **Supplementary Table 2.** Genera, associated *p*-values and log₂ fold changes corresponding to phyla on the volcano plots in Fig. 5.

Abbreviations

BBR: Bushbuckridge; BMI: Body Mass Index; CAG: Community Advisory Group; cMD: CuratedMetagenomicData; DPHRU: Developmental Pathways for Health Research Unit; H3Africa: Human Heredity and Health in Africa; HDSS: Health and Demographic Surveillance Sites; HIV: Human Immunodeficiency Virus; KW: Kruskal-Wallis; MRC: Medical Research Council; PCoA: Principal Coordinate Analysis; RSA: Republic of South Africa; SWT: Soweto; T2D: Type 2 Diabetes; Wits: University of the Witwatersrand

Acknowledgements

We thank our participants who generously gave of their time in support of African science. We relied on many people to make this happen, including Yusuf Ismail, Floidy Wafawanaka, Melody Mabuza, Michaela Hulley, Amanda Haye, and Daniel Ohene-Kwofie. We thank Michèle Ramsay for her wise counsel and leadership of the AWI-Gen consortium, DNA Genotek for the

initial donation of some sample collection kits and Whitehead Scientific for timely assistance with the delivery of kits and accessory items. We are also to grateful to Dylan Maghini for the critical feedback on the manuscript.

Authors' contributions

OHO analyzed and interpreted the data, drafted the paper and formed part of the team that extracted the DNA from stool samples with FBT and VS; FBT, VS and RB contributed substantially to the revision of the manuscript; SMT, ZL, SH, ASB conceptualized and received seed funding, and together with FXG, KK, SAN, ANW, RGW and OHO planned the project; RT facilitated the community engagement sessions at the Agincourt HDSS and contributed to the manuscript; ASB and SH contributed substantively towards the design and management of the project as well as extensively reviewed and contributed to the manuscript. All authors read through the manuscript prior to its submission. The author(s) read and approved the final manuscript.

Funding

This project was funded with a grant from the African Partnership for Disease Control, the South African National Research Foundation (CPRR160421162721), the National Human Genome Research Institute (U54HG006938) as part of the H3A Consortium, the Rosenkranz Prize (to ASB), a Stanford Center for Innovation in Global Health seed award (to ASB), a Fogarty Global Health Equity Scholar award (TW009338; to OHO.), the Center for Computational, Evolutionary and Human Genomics (to FT), and Stanford MedScholars program funding (to RB). ANW is supported by the Fogarty International Centre, National Institutes of Health under award number K43TW010698. The MRC/Wits Rural Public Health and Health Transitions Research Unit and Agincourt Health and Socio-Demographic Surveillance System, a node of the South African Population Research Infrastructure Network (SAPRIN), is supported by the Department of Science and Innovation, the University of the Witwatersrand, and the Medical Research Council, South Africa, and previously the Wellcome Trust, UK (grants 058893/Z/99/A; 069683/Z/02/Z; 085477/Z/08/Z; 085477/B/08/Z). This paper describes the views of the authors and does not necessarily represent the official views of the National Institutes of Health (USA).

Availability of data and materials

The nucleotide sequence data analyzed in this study can be accessed at the ENA under BioProject PRJEB40733. The corresponding phenotype data has been submitted to the EGA (study EGAS00001002482) in terms of the data sharing policy of the Human Heredity and Health in Africa consortium (H3A) and is available by request to the independent H3A Data and Biospecimens Access Committee which will consider each case in terms of H3A policies and to protect participants data. The R code to reproduce statistical analyses is available at <https://github.com/SBIMB/awimbipilot>

Ethics approval and consent to participate

The study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (M160121) and the Provincial Health Research Committee of the Province of Mpumalanga (MP2017TP22851). Written informed consent was obtained from all study participants before any sample collection was done.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa. ²Department of Genetics, Stanford University, Stanford, CA, USA. ³Division of Human Genetics, National Health Laboratory Service, and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. ⁴School of Medicine, Stanford University, Stanford, CA, USA. ⁵MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. ⁶INDEPTH Network, East Legon, Accra, Ghana. ⁷SAMRC Developmental Pathways for Health Research Unit, Department of

Paediatrics, University of the Witwatersrand, Johannesburg, South Africa. ⁸School of Human Development and Health, University of Southampton, Southampton, UK. ⁹Department of Medicine (Hematology), Stanford University, Stanford, CA, USA. ¹⁰School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa.

Received: 3 March 2020 Accepted: 19 October 2020

Published online: 31 October 2020

References

- Michail S, Lin M, Frey MR, Fanter R, Paliy O, Hilbush B, et al. Altered gut microbial energy and metabolism in children with non-alcoholic fatty liver disease. *FEMS Microbiol Ecol*. 2015;91:1–9. <https://doi.org/10.1093/femsec/fiu002>.
- Nakayama J, Yamamoto A, Palermo-Conde LA, Higashi K, Sonomoto K, Tan J, et al. Impact of westernized diet on gut microbiota in children on Leyte Island. *Front Microbiol*. 2017;8:197. <https://doi.org/10.3389/fmicb.2017.00197>.
- De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poulet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci*. 2010;107:14691 LP–14696 <http://www.pnas.org/content/107/33/14691.abstract>.
- Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M. Human gut microbiome viewed across age and geography. *Nature*. 2012;486.
- Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*. 2019;176:649–662.e20. doi:<https://doi.org/10.1016/j.cell.2019.01.001>.
- Iebba V, Santangelo F, Totino V, Pantanella F, Monsia A, Cristanziano V, Cave DD, et al. Gut microbiota related to *Giardia duodenalis*, *Entamoeba* spp. and *Blastocystis hominis* infections in humans from Côte d'Ivoire. *J Infect Dev Ctries*. 2016;10:09 SE-Brief Original Articles. <https://doi.org/10.3855/jidc.8179>.
- Brewster R, Tamburini FB, Asimwe E, Oduaran O, Hazelhurst S, Bhatt AS. Surveying gut microbiome research in Africans: toward improved diversity and representation. *Trends Microbiol*. 2019;27(10):824–35. <https://doi.org/10.1016/j.tim.2019.05.006>.
- Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, et al. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun* 2014;5. <http://dx.doi.org/10.1038/ncomms4654>.
- Gomez A, Petzelkova KJ, Burns MB, Yeoman CJ, Amato KR, Vlckova K, et al. Gut microbiome of coexisting BaAka pygmies and bantu reflects gradients of traditional subsistence patterns. *Cell Rep*. 2016;14:2142–53. <https://doi.org/10.1016/j.celrep.2016.02.013>.
- De Filippo C, Di Paola M, Ramazzotti M, Albanese D, Pieraccini G, Banci E, et al. Diet, environments, and gut microbiota. A Preliminary Investigation in Children Living in Rural and Urban Burkina Faso and Italy. *Front Microbiol*. 2017;8:1979 <https://www.frontiersin.org/article/10.3389/fmicb.2017.01979>.
- Ayeni FA, Biagi E, Rampelli S, Fiori J, Soverini M, Audu HJ, et al. Infant and adult gut microbiome and Metabolome in rural Bassa and urban settlers from Nigeria. *Cell Rep*. 2018;23:3056–67. <https://doi.org/10.1016/j.celrep.2018.05.018>.
- Holden CJ. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proc R Soc London Ser B Biol Sci*. 2002;269:793–9. <https://doi.org/10.1098/rspb.2002.1955>.
- Reyes A, Blanton LV, Cao S, Zhao G, Manary M, Trehan I, et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci*. 2015;112:11941 LP–11946. <https://doi.org/10.1073/pnas.1514285112>.
- Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, et al. Gut Microbiomes of Malawian Twin Pairs Discordant for Kwashiorkor. *Science* (80-). 2013;339:548 LP–554. <https://doi.org/10.1126/science.1229000>.
- Li SX, Armstrong A, Neff CP, Shaffer M, Lozupone CA, Palmer BE. Complexities of gut microbiome Dysbiosis in the context of HIV infection and antiretroviral therapy. *Clin Pharmacol Ther*. 2016;99:600–11. <https://doi.org/10.1002/cpt.363>.
- Vujkovic-Cvijin I, Dunham RM, Iwai S, Maher MC, Albright RG, Broadhurst MJ, et al. Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. *Sci Transl Med*. 2013;5:193ra91. <https://doi.org/10.1126/scitranslmed.3006438>.
- Amugsi DA, Dimbuene ZT, Mberu B, Muthuri S, Ezeh AC. Prevalence and time trends in overweight and obesity among urban women: an analysis of demographic and health surveys data from 24 African countries, 1991–2014. *BMJ Open*. 2017;7:e017344. <https://doi.org/10.1136/bmjopen-2017-017344>.
- Doumatey AP, Adeyemo A, Zhou J, Lei L, Adebamowo SN, Adebamowo C, et al. Gut microbiome profiles are associated with type 2 diabetes in urban Africans. *Front Cell Infect Microbiol*. 2020;10:63. <https://doi.org/10.3389/fcimb.2020.00063>.
- Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet*. 2012;380:247–257. doi:[https://doi.org/10.1016/S0140-6736\(12\)60646-1](https://doi.org/10.1016/S0140-6736(12)60646-1).
- Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev*. 2012;70. <https://doi.org/10.1111/j.1753-4887.2011.00456.x>.
- Patton GC, Coffey C, Cappa C, Currie D, Riley L, Gore F, et al. Health of the world's adolescents: a synthesis of internationally comparable data. *Lancet*. 2012;379:1665–75. [https://doi.org/10.1016/S0140-6736\(12\)60203-7](https://doi.org/10.1016/S0140-6736(12)60203-7).
- Popkin BM, Corvalan C, Grummer-Strawn LM. Dynamics of the double burden of malnutrition and the changing nutrition reality. *Lancet*. 2020;395:65–74. [https://doi.org/10.1016/S0140-6736\(19\)32497-3](https://doi.org/10.1016/S0140-6736(19)32497-3).
- Wells JC, Sawaya AL, Wibaek R, Mwangome M, Poullas MS, Yajnik CS, et al. The double burden of malnutrition: aetiological pathways and consequences for health. *Lancet*. 2020;395:75–88. [https://doi.org/10.1016/S0140-6736\(19\)32472-9](https://doi.org/10.1016/S0140-6736(19)32472-9).
- Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384. [https://doi.org/10.1016/S0140-6736\(14\)60460-8](https://doi.org/10.1016/S0140-6736(14)60460-8).
- Price AJ, Crampin AC, Amberbir A, Kayuni-Chihana N, Musicha C, Tafatatha T, et al. Prevalence of obesity, hypertension, and diabetes, and cascade of care in sub-Saharan Africa: a cross-sectional, population-based study in rural and urban Malawi. *Lancet Diabetes Endocrinol*. 2018;6:208–22. [https://doi.org/10.1016/S2213-8587\(17\)30432-1](https://doi.org/10.1016/S2213-8587(17)30432-1).
- Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet*. 2017;390:2627–2642. doi:[https://doi.org/10.1016/S0140-6736\(17\)32129-3](https://doi.org/10.1016/S0140-6736(17)32129-3).
- Kengne A, Bentham J, Zhou B, Peer N, Matsha T, Bixby H, et al. Trends in obesity and diabetes across Africa from 1980 to 2014: an analysis of pooled population-based studies. *Int J Epidemiol*. 2017;46:1421–32. <https://doi.org/10.1093/ije/dyx078>.
- Patton GC, Olsson CA, Skirbekk V, Saffery R, Wlodek ME, Azzopardi PS, et al. Adolescence and the next generation. *Nature*. 2018;554:458–66. <https://doi.org/10.1038/nature25759>.
- Sawyer SM, Afifi RA, Bearinger LH, Blakemore S-J, Dick B, Ezeh AC, et al. Adolescence: a foundation for future health. *Lancet*. 2012;379:1630–40. [https://doi.org/10.1016/S0140-6736\(12\)60072-5](https://doi.org/10.1016/S0140-6736(12)60072-5).
- Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet*. 2013;382:427–51. [https://doi.org/10.1016/S0140-6736\(13\)60937-X](https://doi.org/10.1016/S0140-6736(13)60937-X).
- Spires M, Delobelle P, Sanders D, Puoane T, Hoelzel P, Swart R. Diet-related non-communicable diseases in South Africa : determinants and policy responses. In: Padarath A, King J, Mackie E-L, Casciola J, editors. *South African Health Review*. 19th ed. Durban: Health Systems Trust; 2016. p. 35–42. <https://www.hst.org.za/>.
- Goedecke JH, Jennings CL, Lambert E V. Obesity in South Africa. 2006. <https://www.samrc.ac.za/sites/default/files/files/2016-07-14/cdl1995-2005.pdf>.
- Sedibe MH, Pisa PT, Feeley AB, Pedro TM, Kahn K, Norris SA. Dietary habits and eating practices and their association with overweight and obesity in rural and urban Black south African adolescents. *Nutrients*. 2018;10:145. <https://doi.org/10.3390/nu10020145>.
- Kruger HS, Puoane T, Senekal M, van der Merwe M-T. Obesity in South Africa: challenges for government and health professionals. *Public Health Nutr*. 2005;8:491–500. <https://doi.org/10.1079/PHN2005785>.

35. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt health and socio-demographic surveillance system. *Int J Epidemiol*. 2012;41:988–1001. <https://doi.org/10.1093/ije/dys115>.
36. Kabudula CW, Houle B, Collinson MA, Kahn K, Gómez-Olivé FX, Clark SJ, et al. Progression of the epidemiological transition in a rural south African setting: findings from population surveillance in Agincourt, 1993–2013. *BMC Public Health*. 2017;17:424. <https://doi.org/10.1186/s12889-017-4312-x>.
37. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JL. Obesity alters gut microbial ecology. *Proc Natl Acad Sci United States Am*. 2005;102:11070–5. <https://doi.org/10.1073/pnas.0504978102>.
38. Hu H-J, Park S-G, Jang HB, Choi M-K, Park K-H, Kang JH, et al. Obesity alters the microbial community profile in Korean adolescents. *PLoS One*. 2015;10:e0134333. <https://doi.org/10.1371/journal.pone.0134333>.
39. Qian L, Gao R, Hong L, Pan C, Li H, Huang J, et al. Association analysis of dietary habits with gut microbiota of a native Chinese community. *Exp Ther Med*. 2018;16:856–66. <https://doi.org/10.3892/etm.2018.6249>.
40. Senghob B, Sokhna C, Ruimy R, Lagier J-C. Gut microbiota diversity according to dietary habits and geographical provenance. *Hum Microbiome J* 2018;7–8:1–9. doi:<https://doi.org/https://doi.org/10.1016/j.humic.2018.01.001>.
41. Ramsay M, Crowther N, Tambo E, Agongo G, Baloyi V, Dikotope S, et al. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob Heal Epidemiol Genomics*. 2016;1:e20. <https://doi.org/10.1017/ghg.2016.17>.
42. Mulder N, Abimiku A, Adebamowo SN, de Vries J, Matimba A, Olowoyo P, et al. H3Africa: current perspectives. *Pharmgenomics Pers Med*. 2018;11:59–66. <https://doi.org/10.2147/PGPM.S141546>.
43. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3. <https://doi.org/10.1038/nmeth.3869>.
44. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One*. 2013;8:e61217. <https://doi.org/https://doi.org/10.1371/journal.pone.0061217>.
45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
46. Smith DP, Peay KG. Sequence depth, not PCR replication, Improves Ecological Inference from Next Generation DNA Sequencing *PLoS One* 2014;9:e90234. <https://doi.org/https://doi.org/10.1371/journal.pone.0090234>.
47. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27:379–423. <https://doi.org/https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
48. Chao A. Nonparametric estimation of the number of classes in a population. *Scand J Stat*. 1984;11:265–70 <http://www.jstor.org/stable/4615964>.
49. Gower JC. Principal Coordinates Analysis. *Encyclopedia Biostatistics*. 2005. <https://doi.org/10.1002/0470011815.b2a13070>.
50. Morton ER, Lynch J, Froment A, Lafosse S, Heyer E, Przeworski M, et al. Variation in rural African gut microbiota is strongly correlated with colonization by *Entamoeba* and subsistence. *PLoS Genet* 2015;11:e1005658. <https://doi.org/https://doi.org/10.1371/journal.pgen.1005658>.
51. Tomova A, Bukovsky I, Rembert E, Yonas W, Alwarith J, Barnard ND, et al. The effects of vegetarian and vegan diets on gut microbiota. *Front Nutr*. 2019;6:47 <https://www.frontiersin.org/article/10.3389/fnut.2019.00047>.
52. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505.
53. Allison MJ, Cook HM, Milne DB, Gallagher S, Clayman RV. Oxalate degradation by gastrointestinal Bacteria from humans. *J Nutr*. 1986;116:455–60. <https://doi.org/10.1093/jn/116.3.455>.
54. Cornick NA, Allison MJ. Assimilation of oxalate, acetate, and CO₂ by *Oxalobacter formigenes*. *Can J Microbiol*. 1996;42:1081–6. <https://doi.org/10.1139/m96-138>.
55. Bui TPN, Shetty SA, Lagkouvardos I, Ritari J, Chamlagain B, Douillard FP, et al. Comparative genomics and physiology of the butyrate-producing bacterium *Intestinimonas butyriciproducens*. *Environ Microbiol Rep*. 2016;8:1024–37. <https://doi.org/10.1111/1758-2229.12483>.
56. Moreno-Indias I, Sánchez-Alcoholado L, García-Fuentes E, Cardona F, Queipo-Ortuño MI, Tinahones FJ. Insulin resistance is associated with specific gut microbiota in appendix samples from morbidly obese patients. *Am J Transl Res*. 2016;8:5672–84 <https://www.ncbi.nlm.nih.gov/pubmed/28078038>.
57. Zhu L, Baker SS, Gill C, Liu W, Alkhoury R, Baker RD, et al. Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology*. 2013;57:601–9. <https://doi.org/10.1002/hep.26093>.
58. Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, et al. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun* 2015;6:6505. <https://doi.org/https://doi.org/10.1038/ncomms7505>.
59. Kane A V, Dinh DM, Ward HD. Childhood malnutrition and the intestinal microbiome. *Pediatr Res* 2014;77:256. <https://doi.org/https://doi.org/10.1038/pr.2014.179>.
60. de la Cuesta-Zuluaga J, Corrales-Agudelo V, Velásquez-Mejía EP, Carmona JA, Abad JM, Escobar JS. Gut microbiota is associated with obesity and cardiometabolic disease in a population in the midst of westernization. *Sci Rep*. 2018;8:11356. <https://doi.org/10.1038/s41598-018-29687-x>.
61. Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, et al. Accessible, curated metagenomic data through ExperimentHub. *Nat Methods*. 2017;14:1023–4. <https://doi.org/10.1038/nmeth.4468>.
62. Twine R, Hundt GL, Kahn K. The ‘experimental public’ in longitudinal health research: views of local leaders and service providers in rural South Africa. *Glob Heal Res Policy*. 2017;2:26. <https://doi.org/10.1186/s41256-017-0046-7>.
63. Wariri O, D’Ambruoso L, Twine R, Ngobeni S, van der Merwe M, Spies B, et al. Initiating a participatory action research process in the Agincourt health and socio-demographic surveillance site. *J Glob Health*. 2017;7:10413. <https://doi.org/10.7189/jogh.07.010413>.
64. Angelakis E, Bachar D, Yasir M, Musso D, Djossou F, Gaborit B, et al. *Treponema* species enrich the gut microbiota of traditional rural populations but are absent from urban individuals. *New Microbes New Infect*. 2018;27:14–21. <https://doi.org/10.1016/j.nmni.2018.10.009>.
65. Wu F, Guo X, Zhang J, Zhang M, Ou Z, Peng Y. *Phascolarctobacterium faecium* abundant colonization in human gastrointestinal tract. *Exp Ther Med*. 2017;14:3122–6. <https://doi.org/10.3892/etm.2017.4878>.
66. Li L, Su Q, Xie B, Duan L, Zhao W, Hu D, et al. Gut microbes in correlation with mood: case study in a closed experimental human life support system. *Neurogastroenterol Motil*. 2016;28:1233–40. <https://doi.org/10.1111/nmo.12822>.
67. Mancabelli L, Milani C, Lugli GA, Turrioni F, Ferrario C, van Sinderen D, et al. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ Microbiol*. 2017;19:1379–90. <https://doi.org/10.1111/1462-2920.13692>.
68. Fei N, Bernabé BP, Lie L, Baghdan D, Bedu-Addo K, Plange-Rhule J, et al. The human microbiota is associated with cardiometabolic risk across the epidemiologic transition. *PLoS One* 2019;14:e0215262. <https://doi.org/https://doi.org/10.1371/journal.pone.0215262>.
69. Andoh A, Nishida A, Takahashi K, Inatomi O, Imaeda H, Bamba S, et al. Comparison of the gut microbial community between obese and lean peoples using 16S gene sequencing in a Japanese population. *J Clin Biochem Nutr*. 2016;59:65–70. <https://doi.org/10.3164/jcbs.15-152>.
70. Hou Y-P, He Q-Q, Ouyang H-M, Peng H-S, Wang Q, Li J, et al. Human gut microbiota associated with obesity in Chinese children and adolescents. *Biomed Res Int*. 2017;2017:7585989. <https://doi.org/10.1155/2017/7585989>.
71. Zacarías MF, Collado MC, Gómez-Gallego C, Flinck H, Aittoniemi J, Isolauri E, et al. Pre-gestational overweight and obesity are associated with differences in gut microbiota composition and systemic inflammation in the third trimester. *PLoS One* 2018;13:e0200305. <https://doi.org/https://doi.org/10.1371/journal.pone.0200305>.
72. Thingholm LB, Rühlemann MC, Koch M, Fuqua B, Laucke G, Boehm R, et al. Obese Individuals with and without Type 2 Diabetes Show Different Gut Microbial Functional Capacity and Composition. *Cell Host Microbe*. 2019;26:252–64.e10. <https://doi.org/10.1016/j.chom.2019.07.004>.
73. Atterbury RJ, Hobbey L, Till R, Lambert C, Capeness MJ, Lerner TR, et al. Effects of orally administered *Bdellovibrio bacteriovorus* on the well-being and *Salmonella* colonization of young chicks. *Appl Environ Microbiol*. 2011;77:5794–803. <https://doi.org/10.1128/AEM.00426-11>.
74. Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC, et al. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to cyanobacteria. *Elife*. 2013;2:e01102. <https://doi.org/10.7554/eLife.01102>.
75. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An expanded genomic representation of the phylum cyanobacteria. *Genome Biol Evol*. 2014;6:1031–45. <https://doi.org/10.1093/gbe/evu073>.

76. Iebba V, Santangelo F, Totino V, Nicoletti M, Gagliardi A, De Biase RV, et al. Higher prevalence and abundance of *Bdellovibrio bacteriovorus* in the human gut of healthy subjects. *PLoS One* 2013;8:e61608. <https://doi.org/10.1371/journal.pone.0061608>.
77. Shankar V, Gouda M, Moncivaiz J, Gordon A, Reo NV, Hussein L, et al. Differences in Gut Metabolites and Microbial Composition and Functions between Egyptian and U.S. Children Are Consistent with Their Diets. *mSystems*. 2017;2:e00169–16. <https://doi.org/10.1128/mSystems.00169-16>.
78. Milani C, Ticinesi A, Gerritsen J, Nouvenne A, Lugli GA, Mancabelli L, et al. Gut microbiota composition and *Clostridium difficile* infection in hospitalized elderly individuals: a metagenomic study. *Sci Rep* 2016;6:25945. <https://doi.org/10.1038/srep25945>.
79. Maeda Y, Kurakawa T, Umamoto E, Motooka D, Ito Y, Gotoh K, et al. Dysbiosis contributes to arthritis development via activation of autoreactive T cells in the intestine. *Arthritis Rheumatol*. 2016;68:2646–61. <https://doi.org/10.1002/art.39783>.
80. Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. *Immunology*. 2017;151:363–74. <https://doi.org/10.1111/imm.12760>.
81. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyötyläinen T, Nielsen T, Jensen BAH, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*. 2016;535:376. <https://doi.org/10.1038/nature18646>.
82. Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, et al. The *Prevotella copri* complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe*. 2019. <https://doi.org/10.1016/j.chom.2019.08.018>.
83. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLoS One* 2014;9:e105592. <https://doi.org/10.1371/journal.pone.0105592>.
84. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y. Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42.
85. Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, et al. Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems*. 2017;2:e00164–16. <https://doi.org/10.1128/mSystems.00164-16>.
86. Brito IL, Yilmaz S, Huang K, Xu L, Jupiter SD, Jenkins AP, et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature*. 2016;535:435–9. <https://doi.org/10.1038/nature18927>.
87. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun*. 2015;6:6528. <https://doi.org/10.1038/ncomms7528>.
88. Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, et al. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol*. 2016;2:16180. <https://doi.org/10.1038/nmicrobiol.2016.180>.
89. Consortium THMP, Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, et al. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207. <https://doi.org/10.1038/nature11234>.
90. Karlsson F, Tremaroli V, Nielsen J, Backhed F. Assessing the human gut microbiota in metabolic diseases. *Diabetes*. 2013;62. <https://doi.org/10.2337/db13-0844>.
91. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500:541. <https://doi.org/10.1038/nature12506>.
92. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ-M, Quick J, et al. A culture-independent sequence-based Metagenomics approach to the investigation of an outbreak of Shiga-toxicogenic *Escherichia coli* O104:H4. *JAMA*. 2013;309(14):1502–10. <https://doi.org/10.1001/jama.2013.3231>.
93. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol*. 2014;32.
94. Rampelli S, Schnorr SL, Consolandi C, Turrioni S, Severgnini M, Peano C, et al. Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr Biol*. 2015;25:1682–93. <https://doi.org/10.1016/j.cub.2015.04.055>.
95. Raymond F, Ouameur AA, Déraspe M, Iqbal N, Gingras H, Dridi B, et al. The initial state of the human gut microbiome determines its reshaping by antibiotics. *ISME J*. 2016;10:707–20. <https://doi.org/10.1038/ismej.2015.148>.
96. Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, et al. Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell*. 2016;167:1897. <https://doi.org/10.1016/j.cell.2016.11.046>.
97. Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell*. 2016;165:1551. <https://doi.org/10.1016/j.cell.2016.05.056>.
98. Vincent C, Miller MA, Edens TJ, Mehrotra S, Dewar K, Manges AR. Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and *Clostridium difficile* colonization or infection. *Microbiome*. 2016;4:12. <https://doi.org/10.1186/s40168-016-0156-3>.
99. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Herczeg R, et al. Colorectal Cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS One*. 2016;11:e0155362. <https://doi.org/10.1371/journal.pone.0155362>.
100. Xie H, Guo R, Zhong H, Feng Q, Lan Z, Qin B, et al. Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. *Cell Syst*. 2016;3:572–84.e3. <https://doi.org/10.1016/j.cels.2016.10.004>.
101. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol*. 2014;10:766. <https://doi.org/10.15252/msb.20145645>.
102. Yu J, Feng Q, Wong SH, Zhang D, Liang Q, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*. 2017;66(1):70–8. <https://doi.org/10.1136/gutjnl-2015-309800>.
103. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490.
104. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature*. 2014;513:59–64. <https://doi.org/10.1038/nature13568>.
105. Soverini M, Turrioni S, Biagi E, Quercia S, Brigidi P, Candela M, et al. Variation of carbohydrate-active enzyme patterns in the gut microbiota of Italian healthy subjects and type 2 diabetes patients. *Front Microbiol*. 2017;8:2079. <https://doi.org/10.3389/fmicb.2017.02079>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

