

The human microbiota is associated with cardiometabolic risk across the epidemiologic transition

Authors: Na Fei¹, Beatriz Peñalver Bernabé¹, Louise Lie², Danny Baghdan², Kweku Bedu-Addo³, Jacob Plange-Rhule³, Terrence E. Forrester⁴, Estelle V. Lambert⁵, Pascal Bovet⁶, Neil Gottel¹, Walter Riesen⁷, Wolfgang Korte⁷, Amy Luke², Stephanie A. Kliethermes⁸, Brian T. Layden^{9, 10}, Jack A. Gilbert¹, Lara R. Dugas^{2*}

¹ Microbiome Center, Department of Surgery, University of Chicago, Chicago, IL, US

² Public Health Sciences, Stritch School of Medicine, Loyola University Chicago, Maywood, IL, US

³ Department of Physiology, SMS, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

⁴ Solutions for Developing Countries, University of the West Indies, Mona, Kingston, Jamaica

⁵ Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa

⁶ Institute of Social & Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland & Ministry of Health, Republic of Seychelles

⁷ Center for Laboratory Medicine, Canton Hospital, St. Gallen, Switzerland

⁸ Department of Orthopedics and Rehabilitation, University of Wisconsin School of Medicine and Public Health, Wisconsin, US

⁹ Division of Endocrinology, Diabetes, and Metabolism, University of Illinois at Chicago, Chicago, IL, US

¹⁰ Jesse Brown Veterans Affairs Medical Center, Chicago, Illinois, US.

Author contact information:

Na Fei: feina405@gmail.com

Bea Penalver: bpenalver@gmail.com

Louise Lie: llie@luc.edu

Danny Baghdan: dbaghdan@luc.edu

Kweku Bedu-Addo: kwekuba@yahoo.com

Jacob Plange-Rhule: jprhule@gmail.com

Terrence E. Forrester: Terrence.forrester@uwimona.edu.jm

Estelle V. Lambert: Vicki.Lambert@uct.ac.za

Pascal Bovet: bovet.pascal@gmail.com

Neil Gottel: ngottel@surgery.bsd.uchicago.edu

Walter Riesen: wf.riesen@bluewin.ch

Wolfgang Korte: Wolfgang.Korte@zlmsg.ch

Amy Luke: aluke@luc.edu

Stephanie Kliethermes: Kliethermes@ortho.wisc.edu

Brian T. Layden: blayde1@uic.edu

Jack Gilbert: gilbertjack@uchicago.edu

Lara R. Dugas: ldugas@luc.edu

Short title: Human gut microbiota and cardiometabolic risk

Corresponding author: Lara Dugas, PhD, MPH

Full postal address: 2160 S. 1st Ave Maywood, IL, 60153, US

Phone: +1-708 327 9029; Fax: +1-708 327 9009, Email: ldugas@luc.edu

46 Abstract

47 Oral and fecal microbial biomarkers have previously been associated with cardiometabolic (CM) risk,
 48 however, no comprehensive attempt has been made to explore this association in minority populations
 49 or across different geographic regions. We characterized gut- and oral-associated microbiota and CM
 50 risk in 655 participants of African-origin, aged 25-45, from Ghana, South Africa, Jamaica, and the United
 51 States (US). CM risk was classified using the CM risk cut-points for elevated waist circumference,
 52 elevated blood pressure and elevated fasted blood glucose, low high-density lipoprotein (HDL), and
 53 elevated triglycerides. Gut-associated bacterial alpha diversity negatively correlated with elevated blood
 54 pressure and elevated fasted blood glucose. Similarly, gut bacterial beta diversity was also significantly
 55 differentiated by waist circumference, blood pressure, triglyceridemia and HDL-cholesterolemia. Notably,
 56 differences in inter- and intra-personal gut microbial diversity were geographic-region specific.
 57 Participants meeting the cut-points for 3 out of the 5 CM risk factors were significantly more enriched with
 58 Lachnospiraceae, and were significantly depleted of Clostridiaceae, Peptostreptococcaceae, and
 59 *Prevotella*. The predicted relative proportions of the genes involved in the pathways for
 60 lipopolysaccharides (LPS) and butyrate synthesis were also significantly differentiated by the CM risk
 61 phenotype, whereby genes involved in the butyrate synthesis via lysine, glutarate and 4-
 62 aminobutyrate/succinate pathways and LPS synthesis pathway were enriched in participants with greater
 63 CM risk. Furthermore, inter-individual oral microbiota diversity was also significantly associated with the
 64 CM risk factors, and oral-associated *Streptococcus*, *Prevotella*, and *Veillonella* were enriched in
 65 participants with 3 out of the 5 CM risk factors. We demonstrate that in a diverse cohort of African-origin
 66 adults, CM risk is significantly associated with reduced microbial diversity, and the enrichment of specific
 67 bacterial taxa and predicted functional traits in both gut and oral environments. As well as providing new
 68 insights into the associations between the gut and oral microbiota and CM risk, this study also highlights
 69 the potential for novel therapeutic discoveries which target the oral and gut microbiota in CM risk.

Introduction

Metabolic syndrome and cardiometabolic (CM) risk are associated with increased morbidity and mortality(1-3), and includes five risk factors: visceral obesity, elevated fasted blood glucose and elevated blood pressure, decreased high density lipoprotein (HDL) cholesterol, hypertriglyceridemia(4). In the US, as many as 35% of all US adults present with CM risk(5), when defined using the “Adult Treatment Panel criteria”(6). Between 1988 and 2012, the greatest increase in the prevalence of CM disease was seen among black men, estimated to be currently around 55%, and rose by 41% among black women(5). There is increasing evidence that the composition and metabolic function of the gut microbiota correlate with the progression and pathogenesis of CM disease, although the causality remains unclear(7).

One of the hallmarks of CM-related disease is the concomitant presence of low-grade systemic inflammation(8). Likewise, reduced intestinal bacterial diversity has also been reported in conditions with chronic inflammation and CM risk, with effects on the immune system(9, 10). The relative abundance of specific bacterial taxa has been shown to differ between individuals with CM risk to healthy participants. For instance, individuals with obesity, diabetes and cardiovascular disease, often have a significant reduction in fecal-associated short-chain fatty acid (SCFA)-synthesizing bacteria such as *Bifidobacteria*, *Roseburia*, *Faecalibacterium prausnitzii* and *Akkermansia muciniphilia*(11-13). All of these microorganisms have been shown to exhibit anti-inflammatory effects(13-16). Potential opportunistic pathogens, such as *Staphylococcus aureus* and those from the family Enterobacteriaceae, are often enriched in the stools of individuals with CM risk(17, 18). It has been suggested that these pro-inflammatory taxa contributes to the evolution of CM risk(17, 18).

There are several metabolic mechanisms through which the gut microbiota may contribute to the CM risk. Lipopolysaccharide (LPS) is a known precursor for the development of obesity and insulin resistance(19, 20) and an increase in the relative proportion of the bacteria that produce LPS in the gut has been associated with elevated systemic LPS concentration, likely resulting in high inflammation. Trimethylamine N-oxide (TMAO), which derived from microbial trimethylamine metabolism, is also increased in the stool of individuals with elevated CM risk(21, 22), although some find the inverse to be true(23) and one study found an inverse association with cardiac death in African Americans(24). Similarly, SCFAs such as propionate, acetate and butyrate produced by bacterial fermentation of indigestible fibers, are known to regulate host energy intake, expenditure and storage, decreases their concentrations have been associated with elevated CM risk(25, 26). Finally, specific gut bacteria have also been associated with altered bile acid composition, which also seem to play an important role in

diabetes, obesity, non-alcoholic fatty liver disease and other metabolic diseases via the farnesoid X receptor (FXR) and G protein-coupled bile acid receptor (GPCR) signaling pathway(27-29). These metabolic pathways represent mechanisms through which the gut microbiota may influence CM risk, and by which they might also serve as potential therapeutic targets for treating elevated CM risk.

Emerging data suggests that the oral microbiota may also be linked to elevated CM risk(30-32). Oral bacteria, especially potential pro-inflammatory pathogens like *Pseudomonas* and *Enterobacter*, have been detected in human atherosclerotic plaques, associated with CM risk (31, 33). Hence, systemic inflammation triggered by oral pathogens may be an important component in the pathogenesis of systemic disease(30, 34). However, to date, it is unclear whether there is an association between the oral and gut microbiota and elevated CM risk, and whether the oral microbiota exhibits similar associations as gut microbiota does with CM risk. If this is the case, then, the oral microbiota, which is substantially more convenient to collect, may be a proxy to determine gut microbiota-derived associations that may have a more specific direct mechanistic association with elevated CM risk. However, a lack of large-scale cohort studies limits our interpretation of CM risk-microbiota associations, particularly across diverse human populations.

In this study, we leveraged African-origin participants enrolled in the “Modeling the Epidemiologic Transition Study” (METS)(35) cohort to determine the association between the gut (stool-derived) and oral (saliva-derived) microbiota and elevated CM risk. We characterized the gut- and oral-associated microbiota and indices of CM risk in 655 participants of African-origin, aged 25-45, from Ghana, South Africa, Jamaica, and the United States (US). The central hypothesis is that the oral microbial composition is associated with the gut microbial composition, and that both are associated with elevated CM risk in blacks.

Results

Participant characteristics

Previously 2,506 adults from Ghana, South Africa, Jamaica, The Seychelles, and US were recruited in 2009, and prospectively followed on an annual basis in METS(35). The current study included a subsample of 655 participants, of which 196 were Ghanaian, 176 were South African, 92 were Jamaican, and 191 were from the US (**Table 1**). Approximately 60% of the participants were women. The average age among all the participants was 34.9 ± 6.4 years, and participants from South Africa and Jamaica were significantly younger than US participants ($p < 0.001$ and $p = 0.016$, respectively). Men and women

from Ghana, South Africa and Jamaica weighed significantly less ($p < 0.001$) than their US counterparts (63.2 ± 12.0 , 76.2 ± 20.5 , and 80.0 ± 21.1 kg, respectively), and there were significantly more US men and women that were overweight and obese (81.3%, $p < 0.001$ for all) compared with the other cohorts (i.e., Ghana, 33.2%; South Africa, 55.7%; Jamaica 65.2%) (**Table 1**). Americans also slept the least number of nightly hours (6.7 ± 1.4 hrs) compared to Ghanaians (7.9 ± 1.4 hrs, $p < 0.001$), Jamaicans (7.3 ± 2.1 hrs, $p < 0.033$), and South Africans (10.5 ± 1.7 hrs, $p < 0.001$), and reported significantly higher prevalences of cigarette smoking (37.7%, $p < 0.001$), and alcohol consumption (84.3%, $p < 0.001$) compared to the other 3 sites, after adjusting for age, sex and BMI (**Table 1**).

Cardiometabolic risk

We used the National Cholesterol Education Program's Adult Treatment Panel III (NCEP/ ATP III) criteria for metabolic syndrome(6) to indicate CM risk, and as follows; waist circumference >102 cm in men and >88 cm in females; elevated blood pressure ($\geq 130/85$ mm Hg), or receiving treatment for hypertension; hypertriglyceridemia (≥ 150 mg/dL) or receiving treatment; low high-density lipoprotein (HDL) cholesterol (<40 mg/dL in males and <50 mg/dL in females), or receiving treatment; and elevated fasting plasma glucose (≥ 110 mg/dL) or receiving treatment for type 2 diabetes. Individuals with a BMI ≥ 25 kg/m² were classified as overweight, and ≥ 30 kg/m² were categorized as obese(36). For the purpose of our analysis, we dichotomized CM risk creating a CM risk phenotype, whereby those with elevated CM risk had at least 3 of the five CM risk factors, compared to those with 2 or less. Because the Jamaican participants were missing HDL and triglyceride concentration measures in blood, they were excluded from the overall CM risk analysis. This resulted in a total of 563 participants with complete CM risk data. Overall, 46 of these 563 participants (8.2%) presented with elevated CM risk. Compared to the US participants (14.7%), the prevalence of elevated CM risk was significantly lower among the Ghanaians (1.5%, $p < 0.001$), and trended lower among the South Africans (8.5%, $p = 0.09$) (**Table 2**).

Gut-derived bacterial diversity associates with country of origin and CM risk factors

The associations between the gut-derived microbial diversity measured by 16S rRNA amplicon sequences using exact sequence variants (ESVs) and indices of elevated CM risk were calculated across the four cohorts. All analyses were adjusted for age, sex and BMI. Gut bacterial alpha diversity (intrapersonal gut diversity, measured by Shannon and Chao1 indexes)(37) was significantly greater among the Ghanaians and South Africans, compared to the US participants ($p < 0.001$; **Figure 1**, **Supplementary Figure 1**). The Shannon index, which incorporates both the richness and evenness of the community, was lower among participants with elevated blood pressure from South Africa ($p < 0.05$) and Ghana ($p < 0.01$), but showed no significant relationship among the US participants or the Jamaicans

(**Figure 1, Supplementary Figure 2**). Elevated fasting blood glucose was significantly associated with lower alpha diversity, but only among the Jamaicans ($p < 0.05$; **Figure 1, Supplementary Figure 2**). Individuals with elevated waist circumference, triglyceride concentration and HDL levels showed no significant associations for gut microbial alpha diversity (adjusted for age, sex and BMI, $p > 0.05$, **Supplementary Figure 2**). These results suggest that decreased gut microbial alpha diversity is associated some of the CM risk factors (i.e., elevated blood pressure and elevated fasting blood glucose), however, these associations are country specific.

Gut bacterial beta diversity was significantly different between countries [false discovery rate (fdr)-corrected $p < 0.01$; **Supplementary Figure 3**]. Therefore, as with alpha diversity, we derived associations of beta diversity against CM risk within each country independently (**Supplementary Figure 4 and 5**). Weighted and unweighted UniFrac distances, which compute differences between microbial communities based on phylogenetic information(38), were significantly different between high and low waist circumference, but only among the South Africans (fdr-corrected $p < 0.05$) and Ghanaians (fdr-corrected $p < 0.01$). The same was true for elevated blood pressure, whereby weighted UniFrac was significantly different among the South Africans (fdr-corrected $p < 0.05$) and Ghanaians (fdr-corrected $p < 0.05$), but interestingly, unweighted UniFrac was only significantly different in the South Africa cohort (fdr-corrected $p < 0.05$). This might suggest that the differences seen in the South Africans may be in part due to differences in the abundance of rare bacterial taxa, while in Ghana the differences may be due to more abundant bacterial taxa in the fecal samples. Similarly, for triglyceride concentrations, only weighted UniFrac was significantly different and only among the South Africans (fdr-corrected $p < 0.05$), while for HDL concentrations only weighted UniFrac was significantly different among US participants (fdr-corrected $p < 0.05$). Hyperglycemia was not significantly correlated with beta diversity in any of the cohorts. These results suggest that gut microbial structure is significantly associated with individual CM risk factors, and that these associations are geographically dependent, given that the contribution of the abundant versus rare bacterial taxa varied with country of origin. Taken together, contributions of CM risk as well as environmental factors, including country of origin, result in inter-individual dissimilarities in the gut microbial composition (**Supplementary Table 1**).

Gut microbial structure and function correlate with CM risk phenotype

We next investigated whether the gut microbiota correlated with an elevated CM risk phenotype in the participants from USA, Ghana and South Africa. Jamaican's were excluded from this analysis due to missing lipid data, and analyses were adjusted for age, sex and BMI. Participants with elevated CM risk phenotype (at least 3 risk factors) had significantly lower fecal alpha diversity (measured by Shannon

index), compared to those without elevated CM risk ($p < 0.001$; **Figure 2a**). Furthermore, weighted and unweighted UniFrac distances were significantly different between the two CM risk phenotypes (fdr-corrected $p < 0.05$; **Figure 2b, 2c**). We were able to identify the fecal bacterial ESVs which had significantly different abundance as a function of elevated CM risk. For these ESVs with a relative abundance $\geq 1\%$ in at least one group (adjusted for country, sex, age and BMI), participants with elevated CM risk were significantly enriched with 2 bacterial ESVs annotated to family Lachnospiraceae and genus *Bacteroides*, while participants without elevated CM risk had were enriched in 3 bacterial ESVs annotated to family Clostridiaceae, Peptostreptococcaceae, and genus *Prevotella* (fdr-corrected $p < 0.05$; **Figure 2d**).

Specific gut bacterial taxa were also associated with individual CM risk across the entire cohort (adjusted for country, age, sex and BMI, **supplementary Figure 6 and supplementary Table 2**) or when stratified by sex (adjusted for country, age and BMI, **supplementary Figure 6 and supplementary Table 2**). We found that the gut microbial taxa level was profoundly altered in participants for each of the CM risk factors. For example, among participants with an elevated waist circumference, hyperglycemia, elevated blood pressure or low HDL concentration, there were a significant enrichment of bacterial ESVs annotated to genera belonging to family Lachnospiraceae, Enterobacteriaceae and Clostridiaceae; genera *Streptococcus*, *Coprococcus* and *Blautia* (fdr-corrected $p < 0.05$). Conversely, bacterial ESVs annotated to genus *Prevotella* (family Prevotellaceae), and family Enterobacteriaceae, Clostridiaceae, and Peptostreptococcaceae were also significantly enriched among participants without CM risk (fdr-corrected $p < 0.05$; supplementary results). As expected, there were also several gut bacterial taxa that were differentially abundant between each country and each CM risk factor (**supplementary Figure 7 and Supplementary Table 2**).

The 16S rRNA amplicon data was used to predict gene abundance using Piphillin(39) to establish differences in the gut microbial metabolic potential in several well-known pathways associated with elevated CM risk, such as LPS biosynthesis, SCFA metabolism, TMA N-oxide (TMAO) biosynthesis, and genes associated with secondary bile acid biosynthesis. No statistically significant differences in the proportion of genes encoding components of TMA N-oxide (TMAO) biosynthesis, LPS biosynthesis, and secondary bile acid biosynthesis were observed among participants who had an elevated CM risk phenotype (at least 3 risk factors), compared to those without CM risk (adjusted for country, sex, BMI and age, $p > 0.05$). However, the relative abundance of the genes involved in the four different pathways for butyrate synthesis showed that several genes that encode for enoyl-CoA hydratase enzymes (K01692, K01782, K01825) in the acetyl-CoA pathway were enriched in participants without elevated CM

risk; while, crotonase (K01715) and β -hydroxybutyryl-CoA dehydrogenase (K00074) in the acetyl-CoA pathway were enriched in participants with elevated CM risk. Notably, the abundance of predictive genes that encode for glutaconyl-CoA decarboxylase (α , β subunits) (K01615) in the glutarate pathway and the 4-hydroxybutyryl-CoA dehydratase (K14534) in the 4-aminobutyrate/succinate pathway were enriched in participants with elevated CM risk (adjusted for country, sex, age and BMI, *fdr*-corrected $p < 0.05$; **Figure 2d**).

We also explored whether gut bacterial functions were associated with each of the 5 individual CM risk factor within each of the countries. Gut microbial predicted genes involved in LPS and SCFA biosynthesis pathways showed that genes involved in the butyrate synthesis via lysine, glutarate and 4-aminobutyrate/succinate pathways and LPS synthesis pathway were differentially enriched among the cohort with an elevated individual CM risk factor (**supplementary Figure 8-11**). Specifically, predicted genes involved in LPS biosynthesis pathways were enriched among participants with an elevated waist circumference in Ghana and US; participants with an elevated blood pressure in Ghana and Jamaica; participants with an elevated fasted blood glucose in Jamaica and the US, and those with low HDL in South Africa and the US. As the butyrate synthesis via lysine, glutarate and 4-aminobutyrate/succinate pathways, predicted genes involved in the glutarate pathway were significantly enriched in participants with a high waist circumference in South Africa and Ghana, and also participants with low HDL concentrations in the US and South Africa; predicted genes involved in the lysine pathway were enriched in participants with elevated blood pressure in Ghana; and predicted genes involved in the 4-aminobutyrate/ succinate pathway in South Africans were significantly associated with low HDL concentrations (see detail in supplementary results; **supplementary Figure 8-11**) .

Generalized linear models (GLM) were applied to explore the associations between alpha diversity, bacterial taxa and predicted functional genes and total CM risk (captured as the sum of CM risk factors in each participant). For instance, there was a significant positive correlation between the proportion of KEGG ID K01615 in the glutarate pathway and K14534 in the 4-aminobutyrate/succinate pathway and total CM risk score (i.e. a z-score transformation to generate continuous discrete measurements, controlling for country, age, sex, and BMI, $p < 0.05$, see methods for more details). However, there were no significant associations between diversity or differential bacterial taxa and total CM risk ($p > 0.05$). To explore whether the gut microbiota predicted elevated CM risk (three out of the five risk factors), random forest regression was used to determine the gut bacterial ESVs against the total CM risk (z-score transformed) pooling the 3 sites with CM risk data. Elevated CM risk could be predicted by gut-associated ESVs and explained almost 15% of the elevated CM risk variance among those with CM risk data.

273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305

Oral-associated bacterial diversity associates with country of origin and CM risk factors

On the other hand, oral microbiota diversity was associated with the country of origin, but not with the individual CM risk factors. Shannon diversity was significantly greater among the Ghanaians compared to the US participants (adjusted for age, sex and BMI, $p < 0.05$; **Supplementary Figure 12**). As oral microbial alpha diversity was significantly different between countries, we correlated CM risk against alpha diversity for each country independently, and found that none of the CM risk factors were significantly associated with oral microbial alpha diversity ($p > 0.05$, **Supplementary Figure 13**). However, oral microbial beta diversity was significantly different between participants from different countries (fdr-corrected $p < 0.01$, **Supplementary Figure 14**), and therefore, tests of association were performed separately for each country. In the US sample, weighted UniFrac distance was significantly different between individuals with either high and normal waist circumference (fdr-corrected $p < 0.01$) and similarly, between participants with either elevated and normal glucose concentrations (fdr-corrected $p < 0.05$, **Supplementary Figure 15**). Unweighted UniFrac distances were significantly different between individuals with either elevated and normal waist circumference in the US (fdr-corrected $p \leq 0.01$), Ghana (fdr-corrected $p < 0.05$), and South Africa (fdr-corrected $p < 0.05$) (**Supplementary Figure 16**). In the US alone, unweighted UniFrac was significantly different by HDL risk (fdr-corrected $p < 0.05$) (**Supplementary Figure 16**). Therefore, oral beta diversity, and hence microbial structure, was significantly associated with several CM risk factors, but similarly differed by country, suggesting that environmental factors are critical for inter-individual dissimilarities in oral microbial composition (**Supplementary Table 3**).

Oral-associated bacterial community structure correlates with the elevated CM risk phenotype

We next investigated whether the oral microbiota correlated with the elevated CM risk phenotype (3 out of five risk factors) and CM risk score in the US, Ghanaian and South African participants. Alpha diversity was not significantly different between low and elevated CM risk (adjusted for age, sex and BMI, p (fdr-corrected) < 0.05 , **Supplementary Figure 17**). However, both weighted and unweighted UniFrac distance were significantly differentiated by CM risk (fdr-corrected $p < 0.01$, **Figure 3a, 3b**). The differential taxa analysis (with relative abundance $\geq 1\%$ in at least one group) showed that participants with elevated CM risk had a significant enrichment of 4 oral-bacterial ESVs annotated to the genera *Prevotella*, *Veillonella*, and *Streptococcus*; while participants without elevated CM risk had a significant enrichment of 5 bacterial ESVs annotated to genus *Haemophilus*, *Neisseria*, *Fusobacterium* (adjusted for country, sex, BMI and age, p (fdr-corrected) < 0.05 ; **Figure 3c**). Several oral-bacterial ESVs annotated to the genus *Rothia*

presented opposite behaviors (i.e., some were enriched, and some were depleted in participants with elevated CM risk compared with low CM risk)

At a more granular level, specific oral bacterial taxa were also associated with each CM risk factor across the entire cohort (adjusted for country, age, sex and BMI; **supplementary Figure 18 and supplementary Table 4**) or when stratified by sex (adjusted for country, BMI and age; **supplementary Figure 18 and supplementary Table 4**), separately. The results indicate that the oral microbiota of participants who have greater CM risk, e.g. a higher waist circumference, elevated blood pressure or low HDL concentration, are significantly enriched for potentially pro-inflammatory taxa, including *Streptococcus*, *Prevotella*, and *Veillonella* (fdr-corrected $p < 0.05$; supplementary results). There were also specific oral bacteria that were differentially abundant between each country by each CM risk factor (fdr-corrected $p < 0.05$; supplementary results; **supplementary Figure 19 and Supplementary Table 2**).

To determine whether the taxa could identify participants with elevated CM risk, we once again used GLM to determine the association between the differential proportional oral bacterial taxa and total CM risk (sum of the number of CM risk factors for each participant). There was a significantly positive association between the proportion of a *Streptococcus* ESV and total CM risk (z-score transformation), controlling for country, age, and sex ($p < 0.05$). To test whether the oral microbiota can identify participants with the elevated CM risk phenotype, random forest regression was used to examine the association of oral bacterial ESVs against CM risk (z-score transformed). Among participants with elevated CM risk, oral bacterial ESVs, accounted for almost 8% of the variance.

Correlation of beta diversity between gut and oral microbiota

Finally, to explore whether the beta diversity trends seen in the oral microbiota correlated with those in the gut microbiota, we applied a Mantel test using both unweighted and weighted UniFrac distance matrices. Pooling all countries, there was a significant correlation with both unweighted ($r = 0.094$, $p = 0.001$) and weighted ($r = 0.082$, $p = 0.001$) UniFrac distance between the gut and oral microbiota (**Supplementary Figure 20 and 21**). However, within each country, only the weighted UniFrac metric had a significant correlation between oral and gut microbial beta diversity, and only among the US and South African participants ($p = 0.034$, $r = 0.073$ and $p = 0.021$, $r = 0.09$ respectively) (**Supplementary Figure 20 and 21**).

Discussion

In our study of African-origin adults from Ghana, South Africa, Jamaica and the US, we performed a comprehensive analysis exploring the relationship between the gut and oral microbiota and elevated CM risk, representing 4 geographically diverse countries spanning an epidemiologic transition. Overall, our results provide evidence that the gut and oral microbiota may potentially be both predictive as well as a therapeutic target for elevated CM risk, and is in line with previous evidence that the human microbiota are associated with CM risk(40).

Consistent with previous studies(9, 41), we found that the gut microbial alpha diversity was significantly lower in participants with elevated CM risk. However, these associations are mostly geographical and dependent on the type of CM risk factor, e.g. for elevated blood pressure, the association was only found among South Africans and Ghanaians, while only Jamaicans had differences for elevated fasted blood glucose. Similarly, gut bacterial beta diversity was also significantly different between participants with an elevated waist circumference, elevated blood pressure, hypertriglyceridemia or low HDL concentration, compared with their healthy counterparts. Although again, the associations differed by country, e.g. for an elevated waist circumference, the associations were only found among the South African and Ghanaian participants, and for low HDL concentration, only among the US participants.

Previous studies have reported conflicting results as to which specific bacterial taxa associate with CM risk factors(11-13, 17, 18). In our study, gut bacterial ESVs annotated to family Lachnospiraceae were significantly enriched in participants with the elevated CM risk phenotype, along with individual risk factors, including elevated waist circumference and hypertriglyceridemia. Previously, a number of studies have found that an enrichment of Lachnospiraceae was associated with the development of obesity, insulin resistance and other metabolic disorders(42-46). Notably, participants with an elevated waist circumference also showed an enrichment for the genus *Streptococcus*, which has previously been found to be enriched in some CM diseases(17). In our study, participants with healthy waist circumferences, and blood pressure and a normal fasting blood glucose exhibited a significant enrichment of the Ruminococcaceae family (supplementary results), which includes several kinds of beneficial bacteria known to be negatively correlated to metabolic disease and associated with lower CM risk, e.g. genus *Faecalibacterium*(47, 48) and genus *Oscillospira*(49). Similarly, gut bacterial ESVs annotated to the Clostridiaceae and Peptostreptococcaceae families and *Prevotella* genus were enriched in participants with a lower total CM risk as well as individual CM risk factors, e.g. healthy waist circumferences and normal fasting blood glucose levels, consistent with previous research, reporting a negative association with obesity or other CM risk factors(18, 50, 51). Although it should also be noted that there are a handful

of studies that suggest that Clostridiaceae and *Prevotella* are significantly enriched in patients with metabolic diseases(51-53).

Our predicted metagenomic functional analysis showed that participants with elevated CM risk harbored a more pronounced inflammatory phenotype. Indeed, predicted genes involved in LPS biosynthesis pathways were significantly enriched in participants presenting with individual CM risk factors across diverse geographic settings. Similar findings have been reported previously and suggest increased LPS synthesis potential in the gut microbiota in people with obesity, diabetes and other related metabolic diseases(19, 20). Of note, the predicted relative proportion of genes that encode for enzymes involved in butyrate synthesis pathway suggested that the lysine pathway, glutarate pathway, and 4-aminobutyrate/succinate pathway were significantly enriched in individuals with elevated CM risk and also individual risk factor. Butyrate can be synthesized via different substrates, driven by enzymes that are produced and secreted by different bacteria(54). While the most common pathway for butyrate synthesis is via pyruvate and acetyl-coenzyme A, other less-dominant pathways include amino-acids (lysine, glutarate and 4-aminobutyrate/succinate) as substrates via the 4-aminobutyrate pathway, which can produce pro-inflammatory byproducts(54) and are related with obesity. Together, these results suggest that the gut microbial metabolic functional potential of participants with elevated CM risk had a marked inflammation-driving capacity, which may influence host systemic inflammatory levels and may ultimately lead to the CM disease consequence.

While the impact of ethnicity on the core of gut microbiota has been demonstrated by several studies(55, 56), exploring the impact of ethnicity and the associations between gut microbiota and CM risk has not been performed to date. Some studies find that there is a decrease in the gut microbial alpha-diversity that is associated with CM risk from different ethnicities, e.g. Danish obese individuals when compared to non-obese individuals⁹ and obese compared with lean in a mid-western US female adolescent twin cohort study⁽⁴¹⁾. Similarly, previous studies find that there is a shared alteration in the gut microbial composition of the individuals with elevated CM risk and adiposity derived from different ethnicities. For example there was an adiposity-related enrichment of Lachnospiraceae in the gut microbiota of participants enrolled in the Twins UK cohort study(46), as well as among British obese individuals(45), and obese, Colombian subjects (57) with increased cardiometabolic risk, and finally among individuals in the midst of Westernization(58) when compared to the lean individuals, respectively. Interestingly, the association between TMAO and cardiovascular disease has been found to be significantly greater in whites compared to blacks⁽²⁴⁾. In our study of African-origin adults, there was no significant association between predicted genes involved in TMAO biosynthetic pathway and total CM risk or individual CM risk

factors, consistent with previous research, finding a negative association between TMAO concentration and cardiac death in black participants⁽²⁴⁾.

Numerous reports have implicated a close linkage between oral infections, particularly periodontitis, and several systemic diseases, e.g. atherosclerosis, type 2 diabetes mellitus, atherosclerotic cardiovascular disease(30-32). In our study, we found that the oral microbial structure (measured by beta diversity) was associated with indices of elevated CM risk, e.g. high waist circumference, elevated fasted blood glucose, and low HDL concentration. However, as with gut microbiota, these associations varied according to the geographic region. There were also several ESVs significantly associated with total elevated CM risk as well as the individual CM risk factors. For example, participants with total elevated CM risk or individual CM risk factors, e.g. elevated waist circumference, low HDL concentration, elevated blood pressure or hypertriglyceridemia, exhibited enriched oral bacterial ESVs annotated to genus *Streptococcus*, *Prevotella*, or *Veillonella*, which were previously reported to be associated with obesity and other cardiometabolic disease(59-62). Systemic inflammation is an important component of the role of oral bacteria in the pathogenesis of systemic diseases(30, 34). Some strains in genera *Streptococcus*, *Prevotella*, or *Veillonella* are opportunistic pathogens and have been previously reported to be associated with inflammatory diseases(63-65), even though, we also found that ESVs in genus *Neisseria* were higher in participants without individual CM risk, e.g. elevated waist circumference, elevated blood pressure or hypertriglyceridemia, inconsistent with previous reports as described above. Importantly, we found a number of instances whereby different ESVs of the same genus were associated with differential CM risk. For example, different ESVs belonging to the genus *Rothia* were significantly enriched in the oral microbiota of participants both with and without elevated CM risk. While it is possible that these are spurious relationships, it is also possible that these two unclassified *Rothia* strains have different genetic makeup in non-genera conserved regions or different transcriptional and/or translational profiles in genus conserved regions associated with CM risk (e.g., distinct immunological and/or metabolic properties) that could result in different host-associated impact. Indeed, this may explain the conflicting results with certain studies(66). The oral microbiota may significantly impact the CM risk through triggering of inflammatory processes. Consequently, an analysis of the oral microbiota can be imagined to be used as a pre-screening test for CM risk, despite differences between the gut and oral bacterial diversity and taxa associated with elevated CM risk. Notably, in this study, we also found a weak but significant association between the oral and gut microbiota, within the structure and also the functional taxa. Thus, it is conceivable that the oral microbiota might be used as a proxy to determine gut-derived associations with CM risk.

In conclusion, our findings extend our insights into the relationship between the human microbiota and elevated CM risk at the structural and functional level, pointing to possible future therapeutic modalities for CM risk targeting the gut and oral microbiota. Our findings identify previously unknown links between gut and oral microbiota alterations, and CM risk, suggesting that gut and oral microbial composition, structure and predicted functional potential may serve as predictive biomarkers for identifying CM risk, as well as therapeutic targets. The features of association between human microbiome and CM risk is diverse in different populations and indicates that different interventions and potential individualized treatment methods targeting the microbiome need to be developed to control the development of CM risk across the world. Indeed, we are continuing our exploration of these relationships in our large international cohort of African-origin adults spanning the epidemiologic transition(35).

Methods

Study participants

Participant selection

Previously, 2,506 African-origin adults (25-45yrs), were enrolled in METS between January 2010 and December 2011 and followed on a yearly basis. A detailed description of the METS protocol has previously been published(35). For the current study, fecal samples from 655 men and women from Ghana (N=196), South Africa (N=176), Jamaica (N=92) and the US (N=191) were collected in 2014. In addition, 620 of them also supplied saliva samples. Participants were excluded from the original study if they self-reported an infectious disease, including HIV-positive individuals (South Africa), pregnant or lactating women, and unable to participate in normal physical activities(35). METS was approved by the Institutional Review Board of Loyola University Chicago, IL, US; the Committee on Human Research Publication and Ethics of Kwame Nkrumah University of Science and Technology, Kumasi, Ghana; the Research Ethics Committee of the University of Cape Town, South Africa; the Board for Ethics and Clinical Research of the University of Lausanne, Switzerland; and the Ethics Committee of the University of the West Indies, Kingston, Jamaica. All study procedures were explained to participants in their native languages, and participants provided written informed consent after being given the opportunity to ask any questions.

Lifestyle and biochemical measurements

All measurements were made at research clinics located in the respective communities. Weight and height were measured. Participants were asked to provide an early morning fecal sample, using a standard collection kit (EasySampler stool collection kit, Alpco, NH) at their home. Fecal samples were

immediately brought to the site clinics and stored at -80 °C. Participants were requested to fast from 8 pm in the evening prior to the clinic examination, during which fasting capillary glucose concentrations were determined using finger stick (Accu-check Aviva, Roche).

Cardiometabolic risk

We defined CM risk using the National Cholesterol Education Program's Adult Treatment Panel III NCEP/ATP III criteria for cardiometabolic disease(6). These 5 risk factors include: 1) waist circumference >102 cm in males and >88 cm in females; 2) elevated blood pressure ($\geq 130/85$ mm Hg), or receiving treatment; 3) hypertriglyceridemia (≥ 150 mg/dL), or receiving lipid-lowering treatment; 4) low high-density lipoprotein (HDL) cholesterol (<40 mg/dL in males and <50 mg/dL in females), or receiving lipid-lowering treatment; and 5) elevated fasting plasma glucose (≥ 110 mg/dL) or receiving glucose-lowering treatment.

DNA isolation and 16S ribosomal RNA (rRNA) gene sequencing

Microbial genomic DNA was extracted from the human stool samples using the DNeasy PowerSoil DNA Isolation Kit (Qiagen). The V4 region of 16S rRNA gene was amplified and sequenced using the Illumina MiSeq platform(67). The primers used for amplification (515F-806R) contain adapters for MiSeq sequencing and single-end barcodes allowing pooling and direct sequencing of PCR products(68).

16S rRNA gene pyrosequencing data preprocessing and analysis

Raw sequences were pre-processed, quality filtered and analyzed using the next-generation microbiome bioinformatics platform (QIIME2 version 2018.6 pipeline) according to the developer's suggestion(69, 70). We used the DADA2 algorithm(71), a software package wrapped in QIIME2, to identify exact sequence variants (ESVs). Quality control, filtering low quality regions of the sequences, identification and removal of chimera sequences, merging paired-end reads, which yielded the ESV feature table (ESV table). Alpha and beta-diversity analyses were performed in R using the *phyloseq* package(72). Alpha diversity was calculated by Shannon's diversity index, observed OTUs, and Chao1 diversity(37). Results were adjusted for BMI, age, sex and country. Principal coordinate analysis (PCoA) was performed based on weighted and unweighted UniFrac distances, a method for computing differences between microbial communities based on phylogenetic information(38). Weighted UniFrac considered both ESVs presence and absence and abundance distances, and unweighted UniFrac only considered ESVs presence. Permutational multivariate analysis of variance (PERMANOVA, R function *adonis* (vegan, 999 permutations)) was used to analyze statistical differences in beta diversity(73). Benjamini–Hochberg false discovery rate (fdr) correction was used to correct for multiple hypothesis testing(74).

The contribution of CM risk factors and environmental factors (sex, age, BMI, sleep, smoke and alcohol consumption) to the overall weighted and unweighted UniFrac dissimilarities in gut and oral microbiota composition was also assessed using PERMANOVA (R function `adonis` (vegan), 999 permutations), which decomposes the dissimilarity matrix into ‘variance’ explained by each covariate. The obtained R^2 gives the proportion of variability observed in the entire dissimilarity matrix that can be independently attributed to the studied variables.

For taxonomic comparisons, relative abundances based on all obtained reads were used. We used the QIIME2 plugin “q2-feature-classifier” and the Naïve Bayes classifier(75) that was trained on the metagenome annotation package Greengenes13.8(76) 99% operational taxonomic units (OTUs) full-length sequences to obtain the taxonomy for each ESV. Significantly differential ESVs were determined using the statistical framework called analysis of composition of microbiomes (ANCOM)(71) for two group comparisons. FDR correction was used to correct for multiple hypothesis testing. Results were adjusted for BMI, age, sex and country.

A two-side Mantel test using Spearman correlation coefficients (999 permutations) was applied to identify the correlation between the beta diversity of the oral and the gut microbiota, with both unweighted and weighted UniFrac distance matrices in software R with the function “`mantel.test`”(77).

To test for correlations between oral microbiota and the Shannon index or ESVs, which their relative abundances was greater than 1% and are also significantly correlated with CM risk in gut microbiota, random forest regression and generalized linear models (GLM) were performed. Random forest regression was done with 1,000 regression trees based on 10-fold cross-validation and performed with QIIME2 plugin “qiime sample-classifier regress-samples” and the Random Forest regressor in the R programming environment. A randomly drawn 80% of samples were used for model training and the remaining 20% were used for validation.

Metagenome functional predictions of the microbial pathways

We used Piphillin an algorithm to predict the functional profiles of the microbiome (39). Briefly, this algorithm uses direct nearest-neighbor matching between 16S rRNA gene sequencing datasets and microbial genomic databases to infer the metagenomic content of the samples. Gene prediction was performed on ESVs table using online Piphillin (<http://secondgenome.com/Piphillin/>), with KEGG database version 2017 and 97% identity cut-off. Predicted gene content and gene copy numbers within each genome were retrieved and classified in terms of the enzymes code by each gene as KEGG orthology

541 (KOs)(78). Results were adjusted for BMI, age, sex and country. Statistical analyses were performed in
542 R. Student's t-test (normally distributed) or Mann-Whitney U test (not normally distributed) was used for
543 to detect differentially abundant KOs between two groups. FDR correction was used to correct for multiple
544 hypothesis testing.

ReferenceUncategorized References

1. Wu SH, Liu Z, Ho SC. Metabolic syndrome and all-cause mortality: a meta-analysis of prospective cohort studies. *Eur J Epidemiol*. 2010;25(6):375-84.
2. Fan J, Song Y, Chen Y, Hui R, Zhang W. Combined effect of obesity and cardio-metabolic abnormality on the risk of cardiovascular disease: a meta-analysis of prospective cohort studies. *Int J Cardiol*. 2013;168(5):4761-8.
3. Li Z, Yang X, Yang J, Yang Z, Wang S, Sun F, et al. The Cohort Study on Prediction of Incidence of All-Cause Mortality by Metabolic Syndrome. *PLoS One*. 2016;11(5):e0154990.
4. Grundy SM, Cleeman JJ, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation*. 2005;112(17):2735-52.
5. Moore JX, Chaudhary N, Akinyemiju T. Metabolic Syndrome Prevalence by Race/Ethnicity and Sex in the United States, National Health and Nutrition Examination Survey, 1988-2012. *Prev Chronic Dis*. 2017;14:E24.
6. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*. 2002;106(25):3143-421.
7. Aron-Wisnewsky J, Clement K. The gut microbiome, diet, and links to cardiometabolic and chronic disorders. *Nat Rev Nephrol*. 2016;12(3):169-81.
8. Foster M, Samman S. Zinc and regulation of inflammatory cytokines: implications for cardiometabolic disease. *Nutrients*. 2012;4(7):676-94.
9. 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(7464):541-6.
10. Nowak P, Trosied M, Avershina E, Barqasho B, Neogi U, Holm K, et al. Gut microbiota diversity predicts immune status in HIV-1 infection. *AIDS (London, England)*. 2015;29(18):2409-18.
11. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, et al. Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A*. 2013;110(22):9066-71.
12. Cui X, Ye L, Li J, Jin L, Wang W, Li S, et al. Metagenomic and metabolomic analyses unveil dysbiosis of gut microbiota in chronic heart failure patients. *Sci Rep*. 2018;8(1):635.
13. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A*. 2008;105(43):16731-6.
14. Gareau MG, Sherman PM, Walker WA. Probiotics and the gut microbiota in intestinal health and disease. *Nat Rev Gastroenterol Hepatol*. 2010;7(9):503-14.
15. Schneeberger M, Everard A, Gomez-Valades AG, Matamoros S, Ramirez S, Delzenne NM, et al. *Akkermansia muciniphila* inversely correlates with the onset of inflammation, altered adipose tissue metabolism and metabolic disorders during obesity in mice. *Sci Rep*. 2015;5:16643.
16. Machiels K, Joossens M, Sabino J, De Preter V, Arijis I, Eeckhaut V, et al. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut*. 2014;63(8):1275-83.
17. Jackson MA, Goodrich JK, Maxan ME, Freedberg DE, Abrams JA, Poole AC, et al. Proton pump inhibitors alter the composition of the gut microbiota. *Gut*. 2016;65(5):749-56.
18. Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun*. 2017;8(1):845.
19. Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, et al. Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes*. 2007;56(7):1761-72.

- bioRxiv preprint doi: <https://doi.org/10.1101/595934>; this version posted April 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.
20. Fei N, Zhao L. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *Isme J.* 2013;7(4):880-4.
 21. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med.* 2013;19(5):576-85.
 22. Tang WH, Wang Z, Shrestha K, Borowski AG, Wu Y, Troughton RW, et al. Intestinal microbiota-dependent phosphatidylcholine metabolites, diastolic dysfunction, and adverse clinical outcomes in chronic systolic heart failure. *J Card Fail.* 2015;21(2):91-6.
 23. Kaysen GA, Johansen KL, Chertow GM, Dalrymple LS, Kornak J, Grimes B, et al. Associations of Trimethylamine N-Oxide With Nutritional and Inflammatory Biomarkers and Cardiovascular Outcomes in Patients New to Dialysis. *Journal of renal nutrition : the official journal of the Council on Renal Nutrition of the National Kidney Foundation.* 2015;25(4):351-6.
 24. Shafi T, Powe NR, Meyer TW, Hwang S, Hai X, Melamed ML, et al. Trimethylamine N-Oxide and Cardiovascular Events in Hemodialysis Patients. *J Am Soc Nephrol.* 2017;28(1):321-31.
 25. Kimura I, Ozawa K, Inoue D, Imamura T, Kimura K, Maeda T, et al. The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. *Nat Commun.* 2013;4:1829.
 26. Gao Z, Yin J, Zhang J, Ward RE, Martin RJ, Lefevre M, et al. Butyrate improves insulin sensitivity and increases energy expenditure in mice. *Diabetes.* 2009;58(7):1509-17.
 27. Sayin SI, Wahlstrom A, Felin J, Jantti S, Marschall HU, Bamberg K, et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab.* 2013;17(2):225-35.
 28. Watanabe M, Houten SM, Matakai C, Christoffolete MA, Kim BW, Sato H, et al. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature.* 2006;439(7075):484-9.
 29. Thomas C, Gioiello A, Noriega L, Strehle A, Oury J, Rizzo G, et al. TGR5-mediated bile acid sensing controls glucose homeostasis. *Cell Metab.* 2009;10(3):167-77.
 30. Kholy KE, Genco RJ, Van Dyke TE. Oral infections and cardiovascular disease. *Trends Endocrinol Metab.* 2015;26(6):315-21.
 31. Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci U S A.* 2011;108 Suppl 1:4592-8.
 32. Mitra S, Drautz-Moses DI, Alhede M, Maw MT, Liu Y, Purbojati RW, et al. In silico analyses of metagenomes from human atherosclerotic plaque samples. *Microbiome.* 2015;3:38.
 33. Serra e Silva Filho W, Casarin RC, Nicolela EL, Jr., Passos HM, Sallum AW, Goncalves RB. Microbial diversity similarities in periodontal pockets and atheromatous plaques of cardiovascular disease patients. *PLoS One.* 2014;9(10):e109761.
 34. Inaba H, Amano A. Roles of oral bacteria in cardiovascular diseases--from molecular mechanisms to clinical cases: Implication of periodontal diseases in development of systemic diseases. *J Pharmacol Sci.* 2010;113(2):103-9.
 35. Luke A, Bovet P, Forrester TE, Lambert EV, Plange-Rhule J, Schoeller DA, et al. Protocol for the modeling the epidemiologic transition study: a longitudinal observational study of energy balance and change in body weight, diabetes and cardiovascular disease risk. *BMC Public Health.* 2011;11:927.
 36. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: executive summary. Expert Panel on the Identification, Evaluation, and Treatment of Overweight in Adults. *Am J Clin Nutr.* 1998;68(4):899-917.
 37. Peet RK. The measurement of species diversity. *Annual review of ecology and systematics.* 1974;5(1):285-307.

- bioRxiv preprint doi: <https://doi.org/10.1101/595934>; this version posted April 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.
38. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*. 2005;71(12):8228-35.
 39. Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, Dabbagh K, et al. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. *PLoS One*. 2016;11(11):e0166104.
 40. Brown JM, Hazen SL. The gut microbial endocrine organ: bacterially derived signals driving cardiometabolic diseases. *Annu Rev Med*. 2015;66:343-59.
 41. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009;457(7228):480-4.
 42. Meehan CJ, Beiko RG. A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol Evol*. 2014;6(3):703-13.
 43. Cho I, Yamanishi S, Cox L, Methe BA, Zavadil J, Li K, et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature*. 2012;488(7413):621-6.
 44. Poroyko VA, Carreras A, Khalyfa A, Khalyfa AA, Leone V, Peris E, et al. Chronic Sleep Disruption Alters Gut Microbiota, Induces Systemic and Adipose Tissue Inflammation and Insulin Resistance in Mice. *Sci Rep*. 2016;6:35405.
 45. Salonen A, Lahti L, Salojärvi J, Holtrop G, Korpela K, Duncan SH, et al. Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME J*. 2014;8(11):2218-30.
 46. Beaumont M, Goodrich JK, Jackson MA, Yet I, Davenport ER, Vieira-Silva S, et al. Heritable components of the human fecal microbiome are associated with visceral fat. *Genome Biol*. 2016;17(1):189.
 47. Balamurugan R, George G, Kabeerdoss J, Hepsiba J, Chandragunasekaran AM, Ramakrishna BS. Quantitative differences in intestinal *Faecalibacterium prausnitzii* in obese Indian children. *Br J Nutr*. 2010;103(3):335-8.
 48. Lukovac S, Belzer C, Pellis L, Keijser BJ, de Vos WM, Montijn RC, et al. Differential modulation by *Akkermansia muciniphila* and *Faecalibacterium prausnitzii* of host peripheral lipid metabolism and histone acetylation in mouse gut organoids. *mBio*. 2014;5(4).
 49. Konikoff T, Gophna U. Oscillospira: a Central, Enigmatic Component of the Human Gut Microbiota. *Trends in microbiology*. 2016;24(7):523-4.
 50. Karlsson FH, Fak F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun*. 2012;3:1245.
 51. Neyrinck AM, Possemiers S, Druart C, Van de Wiele T, De Backer F, Cani PD, et al. Prebiotic effects of wheat arabinoxylan related to the increase in bifidobacteria, Roseburia and Bacteroides/Prevotella in diet-induced obese mice. *PLoS One*. 2011;6(6):e20944.
 52. Larsen N, Vogensen FK, van den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One*. 2010;5(2):e9085.
 53. Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, et al. Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A*. 2009;106(7):2365-70.
 54. Vital M, Howe AC, Tiedje JM. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio*. 2014;5(2):e00889.
 55. Zhang J, Guo Z, Xue Z, Sun Z, Zhang M, Wang L, et al. A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *The ISME journal*. 2015.
 56. Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, Herrema H, et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nature medicine*. 2018;24(10):1526.

- bioRxiv preprint doi: <https://doi.org/10.1101/595934>; this version posted April 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.
57. De La Cuesta-Zuluaga J, Corrales-Agudelo V, Carmona J, Abad J, Escobar J. Body size phenotypes comprehensively assess cardiometabolic risk and refine the association between obesity and gut microbiota. *International Journal of Obesity*. 2017.
 58. 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. *Scientific reports*. 2018;8(1):11356.
 59. Casarin RC, Barbagallo A, Meulman T, Santos VR, Sallum EA, Nociti FH, et al. Subgingival biodiversity in subjects with uncontrolled type-2 diabetes and chronic periodontitis. *Journal of periodontal research*. 2013;48(1):30-6.
 60. Palleja A, Kashani A, Allin KH, Nielsen T, Zhang C, Li Y, et al. Roux-en-Y gastric bypass surgery of morbidly obese patients induces swift and persistent changes of the individual gut microbiota. *Genome medicine*. 2016;8(1):67.
 61. Ramos E, Gaudier FL, Hearing LR, Del Valle GO, Jenkins S, Briones D. Group B streptococcus colonization in pregnant diabetic women. *Obstetrics and gynecology*. 1997;89(2):257-60.
 62. Zeigler CC, Persson GR, Wondimu B, Marcus C, Sobko T, Modeer T. Microbiota in the oral subgingival biofilm is associated with obesity in adolescence. *Obesity (Silver Spring)*. 2012;20(1):157-64.
 63. Echchannaoui H, Frei K, Schnell C, Leib SL, Zimmerli W, Landmann R. Toll-like receptor 2-deficient mice are highly susceptible to *Streptococcus pneumoniae* meningitis because of reduced bacterial clearing and enhanced inflammation. *The Journal of infectious diseases*. 2002;186(6):798-806.
 64. Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. *Immunology*. 2017;151(4):363-74.
 65. Mogensen TH, Paludan SR, Kilian M, Ostergaard L. Live *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis* activate the inflammatory response through Toll-like receptors 2, 4, and 9 in species-specific patterns. *Journal of leukocyte biology*. 2006;80(2):267-77.
 66. Bhat M, Pasini E, Copeland J, Angeli M, Husain S, Kumar D, et al. Impact of Immunosuppression on the Metagenomic Composition of the Intestinal Microbiome: a Systems Biology Approach to Post-Transplant Diabetes. *Sci Rep*. 2017;7(1):10277.
 67. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*. 2016;1(1).
 68. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *Isme J*. 2012;6(8):1621-4.
 69. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335-6.
 70. Hall M, Beiko RG. 16S rRNA Gene Analysis with QIIME2. *Methods in molecular biology (Clifton, NJ)*. 2018;1849:113-29.
 71. Mandal S, Van Treuren W, White RA, Eggesbo M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis*. 2015;26:27663.
 72. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217.
 73. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral ecology*. 2001;26(1):32-46.

74. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002;64(3):479-98.
75. Murphy KP. Naïve bayes classifiers. University of British Columbia. 2006;18.
76. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*. 2006;72(7):5069-72.
77. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer research*. 1967;27(2):209-20.
78. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*. 2012;40(Database issue):D109-14.

Acknowledgments

The authors would like to acknowledge the original METS study participants and site-specific clinic staff members. The current work is funded in part by the National Institutes of Health R01DK080763, R01DK090360 and R01DK111848. BTL is supported by the National Institutes of Health under award number, R01DK104927-01A1 and Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, VA merit (Grant no. 1101BX003382-01-A1). BPB is supported by the Arnold O. Beckman Postdoctoral Award.

Author information:

Affiliations

Microbiome Center, Department of Surgery, University of Chicago, Chicago, IL, US

Na Fei, Beatriz Penalver Bernabe, Neil Gottel, and Jack A. Gilbert

Public Health Sciences, Stritch School of Medicine, Loyola University Chicago, Maywood, IL, US

Louis Lie, Danny Baghdan, Amy Luke, and Lara R. Dugas

Department of Physiology, SMS, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Kweku Bedu-Addo, Jacob Plange-Rhule

Solutions for Developing Countries, University of the West Indies, Mona, Kingston, Jamaica

Terrence E. Forrester

Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa

Institute of Social & Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland & Ministry of Health, Republic of Seychelles

Pascal Bovet

Center for Laboratory Medicine, Canton Hospital, St. Gallen, Switzerland

Walter Riesen and Wolfgang Korte

Department of Orthopedics and Rehabilitation, University of Wisconsin School of Medicine and Public Health, Wisconsin, US

Stephanie A. Kliethermes

Division of Endocrinology, Diabetes, and Metabolism, University of Illinois at Chicago, Chicago, IL, US

Brian T. Layden

Corresponding author:

Lara Dugas, PhD, MPH

Full postal address: 2160 S. 1st Ave Maywood, IL, 60153, US

Phone : +1-708 327 9029 ; Fax : +1-708 327 9009, Email : ldugas@luc.edu

Contributions

LRD, BTL and JAG supervised the project. LRD, NF, BTL, and JAG conceived the idea. LRD, LL, JPR, and AL, all collected the human data; NF performed the statistical, and computational analyses. NF, LRD, BTL, JAG, BPB, PB, and WK wrote the manuscript. NG (University of Chicago) performed the microbiome sequencing of the human gut microbiota. All authors reviewed the manuscript.

Competing interests

The authors declare no competing financial interests.

Supplementary Materials:

Supplementary results

Supplementary Figures 1 to 21

Supplementary Table 1 to5

Table 1. Participant characteristics by site. The US is the reference site.

	Ghana	South Africa	Jamaica	US	Overall
N	N=196	N=176	N=92	N=191	N=655
Age (y)	35.8 ± 6.6	33.3 ± 5.9*	33.9 ± 6.2*	36.0 ± 6.3	34.9 ± 6.4
Weight (kg)	63.2 ± 12.0**	76.2 ± 20.5**	80.0 ± 21.1**	94.2 ± 25.0	78.1 ± 23.3
Height (cm)	161.7 ± 7.6**	163.7 ± 7.7**	166.8 ± 10.9**	169.6 ± 8.4	165.2 ± 9.0
BMI (kg/m²)	24.3 ± 5.0**	28.7 ± 8.2**	29.1 ± 8.9**	32.8 ± 8.8	28.6 ± 8.4
Normal weight (BMI < 25 kg/m²), N, %	131, 66.8%**	78, 44.3%**	32, 34.8%**	36, 18.9%	277, 42.3%
Overweight (BMI ≥ 25 kg/m² - < 30 kg/m²)	41, 20.9%	28, 15.9%**	28, 30.4%	52, 27.2%	149, 22.8%
Obese (BMI ≥ 30 kg/m²)	24, 12.2%**	70, 39.8%**	32, 34.8%**	103, 53.9%	229, 35.0%
Fat-free mass (kg)	44.3 ± 7.4**	44.6 ± 7.2**	-	55.1 ± 11.2	48.1 ± 10.2
Fat mass (kg)	19.0 ± 9.2**	31.7 ± 16.2**	-	39.3 ± 18.4	29.9 ± 17.3
% Body fat	29.2 ± 9.84**	39.3 ± 11.0	-	39.9 ± 10.8	36.0 ± 11.6
Sleep hours (hrs/night) *	7.9 ± 1.4**	10.5 ± 1.7**	7.3 ± 2.1*	6.7 ± 1.4	8.2 ± 2.2
Smokers*	4, 2.0%**	48, 27.3 %**	10, 10.9%**	72, 37.7%	134, 20.5%
Drinkers*	59, 30.6%**	75, 42.9%**	-	161, 84.3%	295, 52.8%

*p<0.05, **p<0.01 compared to US.

*Adjusted for age, sex and BMI

bioRxiv preprint doi: <https://doi.org/10.1101/595934>; this version posted April 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 2. Cardiometabolic risk by site, and overall. The US is the reference site, and comparisons are adjusted for age, sex and BMI.

	Ghana	South Africa	Jamaica	United States	Overall
	N=196	N=176	N=92	N=191	N=655
Fasted plasma glucose (mg/dL)	100.7 ± 12.2	83.3 ± 18.1**	99.6 ± 15.5	106.2 ± 40.4	97.5 ± 26.8
Elevated fasted plasma glucose, N, %	2, 1.0%**	8, 4.6%**	18, 19.6%	45, 23.9%	73, 11.2%
Systolic blood pressure (mmHg)	112.9 ± 13.5**	124.4 ± 20.5*	112.7 ± 12.5**	123.3 ± 17.6	119.0 ± 17.5
Diastolic blood pressure (mmHg)	66.6 ± 11.0**	79.4 ± 12.8	72.0 ± 34.0**	80.5 ± 13.1	74.9 ± 18.1
Elevated blood pressure, N, %	6, 3.1%**	26, 14.8%	1, 1.1%*	25, 13.1%	58, 8.9
Triglycerides (mg/dL)	82.1 ± 36.5	88.2 ± 55.1	-	98.5 ± 62.1	89.5 ± 52.5
Hypertriglyceridemia, N, % (N=563)	12, 6.2%	16, 9.1%	-	29, 15.7%	57, 10.3%
HDL-cholesterol (mg/dL)	46.1 ± 13.6**	49.9 ± 15.0*	-	52.1 ± 15.1	49.3 ± 14.7
Low HDL-cholesterol, N, % N=563	104, 53.6%**	79, 45.4%*	-	64, 34.8%	247, 44.8%
LDL-cholesterol (mg/dL)	100.7 ± 29.1	92.9 ± 32.6**	-	110.1 ± 35.5	101.4 ± 33.1
Total cholesterol (mg/dL)	163.3 ± 34.3**	160.1 ± 36.1**	-	183.0 ± 38.6	168.9 ± 37.7
High waist circumference N, % N=563	37, 18.9%**	78, 44.3%**	37, 40.2%**	116, 60.7	268, 40.9%
Cardiometabolic disease N, % N=563	3, 1.5%*	15, 8.5%	-	28, 14.7	46, 8.2%

Cardiometabolic risk is defined by three out of five risk factors: waist circumference >102 cm in males and >88 cm in females; elevated blood pressure (≥ 130/85 mm Hg), or receiving treatment; hypertriglyceridemia (≥ 150 mg/dL), or receiving treatment; low HDL (<40 mg/dL in males and <50 mg/dL in females), or receiving treatment; and elevated fasting plasma glucose (≥ 110 mg/dL) or receiving treatment. *p<0.05, **p<0.01 compared to US.

Figure 1 | Gut bacterial alpha diversity associations with cardiometabolic risk factors

(adjusted for age, sex and BMI, only significant associated were shown here). **a)** Alpha diversity (Shannon index) in the four study sites (USA, RSA, Ghana and Jamaica); **b)** association with hypertension in RSA; **c)** association with elevated fasting blood glucose in Jamaica; **d)** association with hypertension in Ghana. * $p < 0.05$; ** $p < 0.01$.

Figure 2 | Gut bacterial structure and function correlates with cardiometabolic (CM) risk phenotype.

(a), Alpha diversity analysis (Shannon Index) from 16S rRNA gene sequence data of gut microbiota against CM risk; **(b-c)**, Principal coordinate analyses (PCoA) of weighted **(a)** and unweighted **(b)** UniFrac distance of gut microbiota composition against CM risk (fdr-corrected $p < 0.01$); **(d)**, Specific gut bacterial taxa are differentially abundant (in relative terms) between study sites with and without elevated CM risk (only significantly differential exact sequence variants (ESVs) with relative abundance $\geq 1\%$ in at least one group shown. Data shown are means \pm S.E.M.; $p(\text{fdr-corrected}) < 0.05$); **(e)**, Relative abundance of the genes involved in the four different pathways for butyrate synthesis against CM risk of gut microbiota across all study sites. (Data shown are means of percentages \pm S.E.M. $p(\text{fdr-corrected}) < 0.05$). CMD means CM risk defined as at least 3 CM risk factors from five: waist circumference, elevated blood pressure, elevated blood fasting glucose, elevated triglyceride and low HDL concentration in USA, RSA, and Ghana *** $p < 0.001$. false discovery rate

Figure 3 | Oral-associated bacterial structure correlates with CM risk phenotype. **(a-b)**, Principal coordinate analyses (PCoA) of weighted **(a)** and unweighted **(b)** UniFrac distance of oral microbiota composition against CM risk (fdr-corrected $p < 0.01$). **(c)**, Specific oral bacterial taxa are differentially abundant (in relative terms) between study sites with and without CM risk (only significantly differential exact sequence variants (ESVs) with relative abundance $\geq 1\%$ in at least one group shown. Data shown are mean \pm S.E.M. $p(\text{fdr-corrected}) < 0.05$). CMD means CM risk, including at least 3 CM risk factors of waist circumference, elevated blood pressure, elevated blood fasting glucose, high triglyceridemia and low HDL-cholesterolemia in USA, RSA, and Ghana. fdr, false discovery rate.

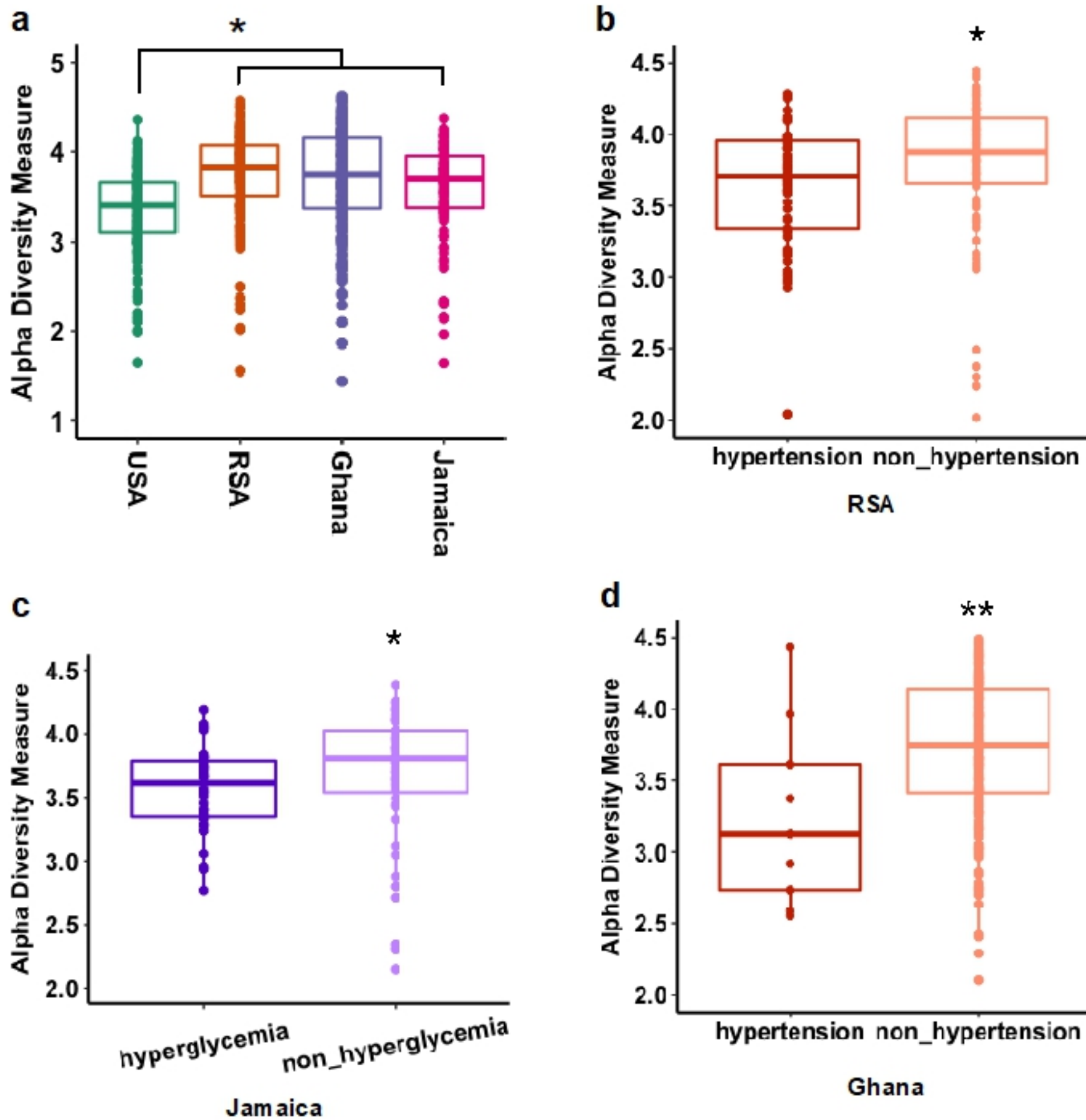


Figure 1

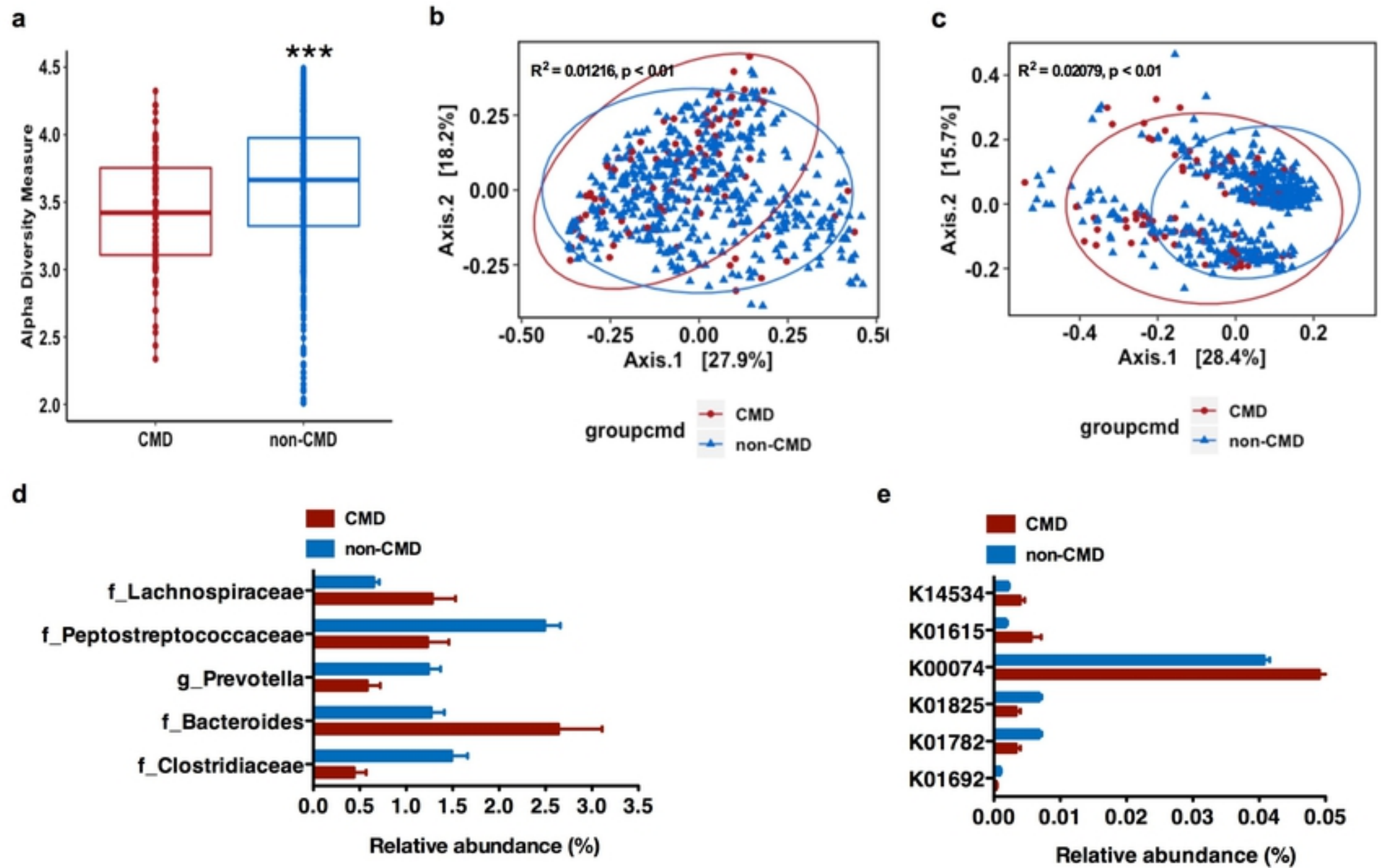


Figure 2

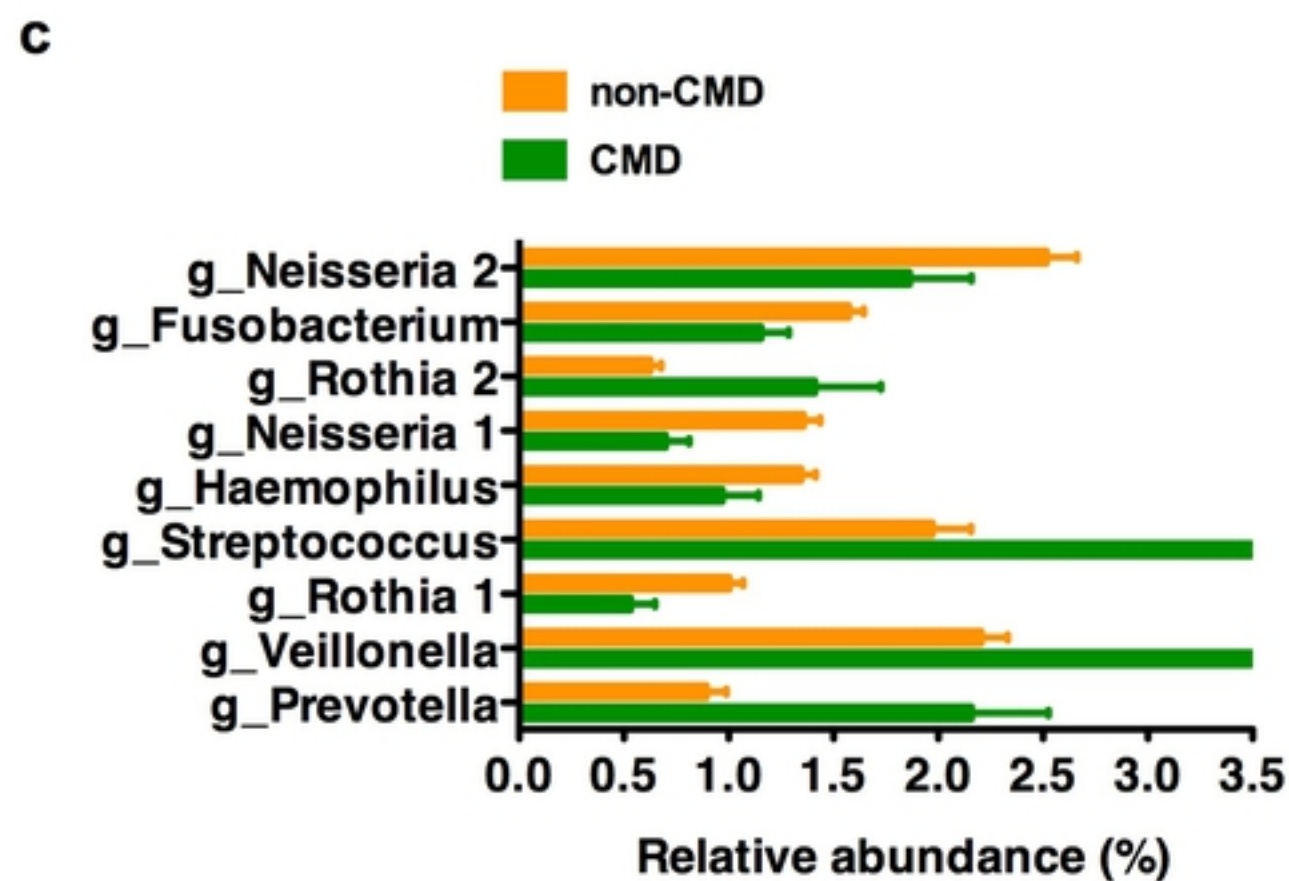
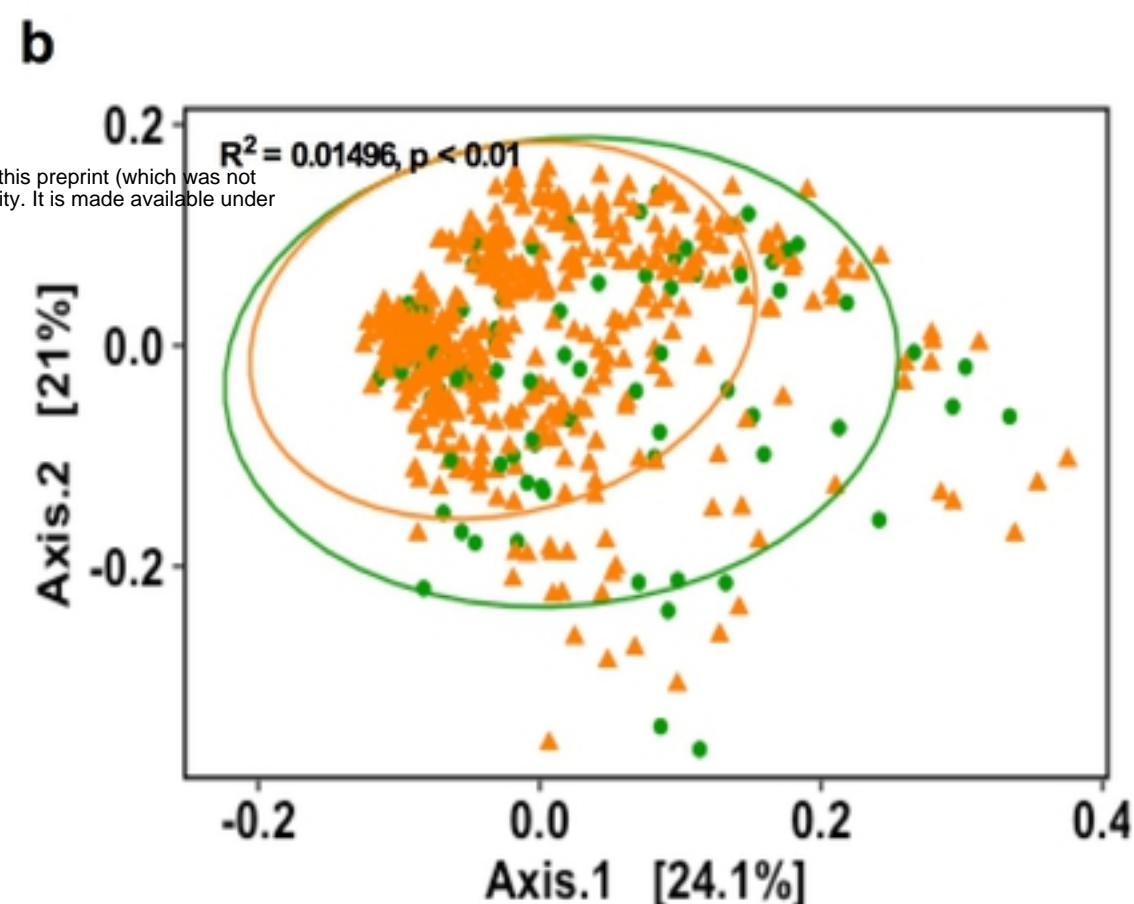
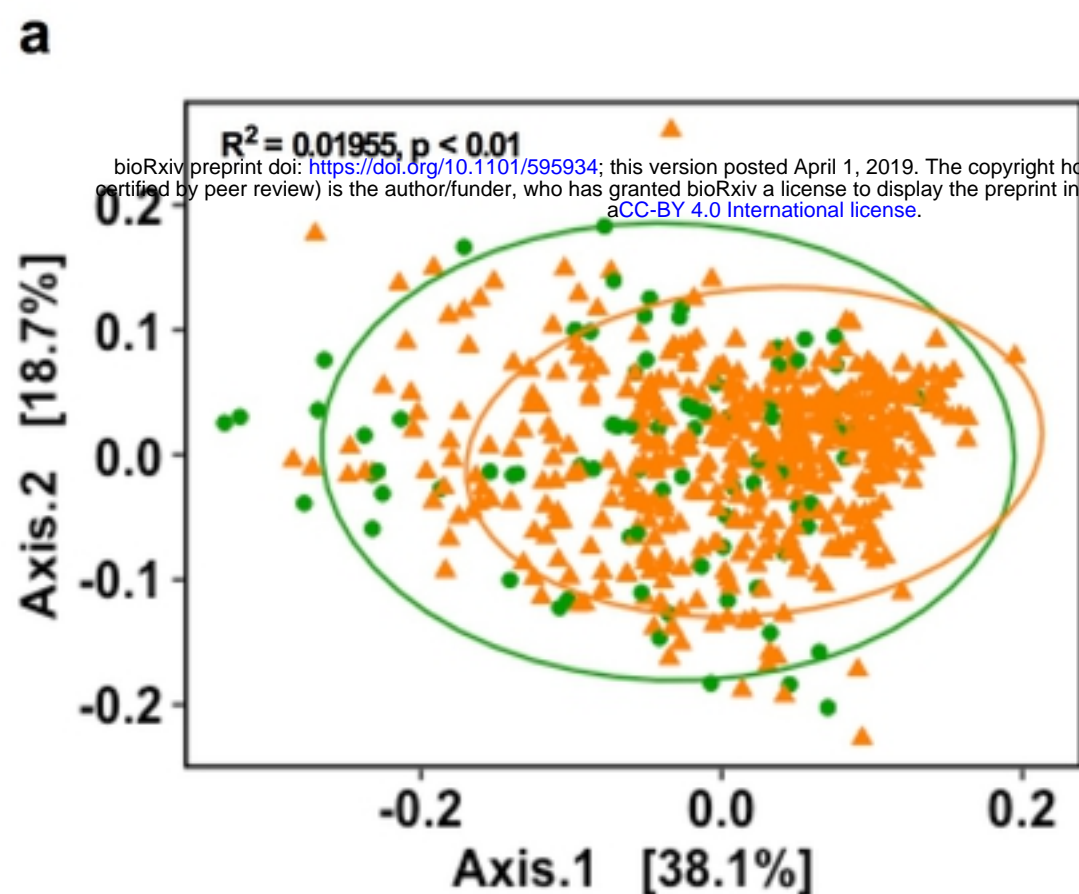


Figure 3