Title: Multiomic Insights into Human Health: Gut Microbiomes of Hunter-Gatherer, Agropastoral, and Western Urban Populations

Authors: Harinder Singh[1,*], Rosana Wiscovitch-Russo[1,2], Claire Kuelbs[2], Josh Espinoza[2], Amanda E. Appel[1, 2], Ruth J. Lyons[4], Sanjay Vashee[1,3], Hagen E.A. Förtsch[5], Jerome E. Foster[6], Dan Ramdath[6], Vanessa M. Hayes[4,7,8], Karen E. Nelson[1,2], Norberto Gonzalez-Juarbe[1,2]

Affiliations: [1]Infectious Diseases Group, J. Craig Venter Institute, Rockville, MD, USA. [2]Genomic Medicine Group, J. Craig Venter Institute, Rockville, MD, USA. [3]Synthetic Biology Group, J. Craig Venter Institute, Rockville, MD, USA. [4]Garvan Institute of Medical Research, New South Wales, Australia. [5]Windhoek Central Hospital, Windhoek Khomas, Namibia. [6]Faculty of Medical Sciences, University of the West Indies, Trinidad. [7]Ancestry and Health Genomics Laboratory, Charles Perkins Centre, University of Sydney, Camperdown, NSW, Australia. [8]School of Health Systems and Public Health, University of Pretoria, Pretoria, Gauteng, South Africa.

***Corresponding authors:**

Harinder Singh, hsingh@jcvi.org

Keywords: Gut microbiome, Ju/'hoansi, hunter-gatherers, multi-omics, metagenomics, metatranscriptomics, metabolomics, dietary patterns, agropastoral, Bantu, healthy, obese, diabetics, Afro-Trinidadian, Caribbean

Abstract:

Societies with exposure to preindustrial diets exhibit improved markers of health. Our study used a comprehensive multi-omic approach to reveal that the gut microbiome of the Ju/'hoansi hunter-gatherers, one of the most remote KhoeSan groups, exhibit a higher diversity and richness, with an abundance of microbial species lost in the western population. The Ju/'hoansi microbiome showed enhanced global transcription and enrichment of complex carbohydrate metabolic and energy generation pathways. The Ju/'hoansi also show high abundance of short-chain fatty acids that are associated with health and optimal immune function. In contrast, these pathways and their respective species were found in low abundance or completely absent in Western populations. Amino acid and fatty acid metabolism pathways were observed prevalent in the Western population, associated with biomarkers of chronic inflammation. Our study provides the first in-depth multi-omic characterization of the Ju/'hoansi microbiome, revealing uncharacterized species and functional pathways that are associated with health.

Introduction:

For the past 10,000 years, the human way of life and diet have drastically transformed due to the development of agriculture and even more after the Industrial Revolution[1]. The rapid shift from a Paleolithic to a Western diet has disrupted highly efficient metabolic pathways that evolved over millions of years of human and host-associated microbial dietary adaptation, leading to the development of Western diseases (WD)[2]. WDs are defined as non-infectious chronic degenerative diseases that are rare or absent in agropastoral (AP) and hunter-gatherer (HG) populations[3]. They are the leading causes of morbidity and mortality in Western populations[3]. WD includes chronic inflammation, obesity, type 2 diabetes (T2D), cardiovascular diseases (CVD), and cancer[4]. In contrast, infectious diseases are the primary cause of death in AP and HG populations[3]. In addition, several studies of HG populations have observed low circulating insulin levels and optimal insulin sensitivity if they maintain their traditional diet[5]. These observations underline the importance of considering the evolutionary history of human dietary patterns in the context of modern dietary practices to promote optimal health outcomes[2].

In the past decade, our understanding of the relationship between human health and the gut microbiome has expanded significantly. It is now well-established that the gut microbiome is crucial for overall health, influencing conditions such as nutrition deficiencies, obesity, chronic inflammation, diabetes, and even cancer[6]. It regulates energy harvest, signaling molecule release, and plays an invaluable role in maintaining the intestinal barrier and immune cell repertoire in the mucosa[7-9]. Understanding the gut microbiome's complexities, its relationship with human health, and the factors that shape its composition is critical for the development of effective therapeutic strategies that can mitigate chronic diseases. Due to recent significant studies from the National Institutes of Health-Human Microbiome Project and the European Metagenomics of the Human Intestinal Tract, our understanding of the genes, function, and complexity within the composition of the human gut bacteria has become more complete[10,11]. Notably, the microbiome's diversity plays a pivotal role in facilitating the production of nutrients from substrates that the host cannot metabolize[12]. Differences in gut microbiota between Western urban populations, HG, and AP are driven by dietary variations[13], with seasonal dietary dependencies causing significant microbiome shifts in groups such as the Hazda HG of Tanzania[14,15]. Western populations have less diverse gut microbiota, which lacks organisms that

promote health and metabolically efficient digestion. In contrast, HG and AP have more diverse and stable gut microbiota due to their high-fiber diets and low-fat and sugar consumption[15,16]. These differences have significant health implications, since alterations in gut microbiota composition and diversity are linked to chronic diseases such as obesity, type 2 diabetes, and inflammatory bowel disease[14,16-18].

The modern human originated in Africa over 200 thousand years ago and primarily relied on a hunter-gatherer's lifestyle when the everyday diet consisted mainly of hunted wild animals, fish, and unfarmed plant-based foods[19,20]. Of significant importance, a recent report has shown that anatomically modern humans originated from the Kalahari Desert region in Namibia[21]. The Ju/'hoansi community in the Kalahari Desert represents a unique and remote segment within the ethnolinguistic KhoeSan framework. Furthermore, the Ju/'hoansi are among the 'oldest' extant (or earliest diverged) contemporary human populations, pre-dating the Baka Pygmy divergence[22]. One of the oldest hunter-gatherer cultures that reside in the Kalahari Desert, the Ju/'hoansi provide a living example of a foraging and hunting lifestyle, with a plant-centric diet rich in ~105 edible plants consisting primarily of nuts, fruits, roots, gums, and leafy greens with meat being scarce[23,24]. However, due to the seasonal scarcity of natural resources, the Ju/'hoansi now depend on a mix of traditional hunter-gatherer methods and market-based strategies. Of note, a recent study has provided a low-depth V3-V4 16S rRNA and ITS1 sequencing for bacterial and fungal identification, respectively, from the Ju/'hoansi living at Nyae Nyae community[25]. While lacking species-level identification, this study documented the microbiota of Ju/'hoansi and established the unique role of diet and culture in core microbial composition compared to urban populations. Here, we discern the composition and function of the microbiome of geographically related HGs and APs and subsequently compare them to a Western diet cohort of genetically similar descendants from the West Indies.

For this research, we collected fecal samples from the Ju/'hoansi population residing in the Kalahari Desert of Northern Namibia, the Bantu AP group in the nearby vicinity, and a cohort comprising both healthy and diabetic/obese (defined as any participant with BMI $\geq$30 kg.m^2, or with an HbA1c value of $\geq$6.5%) individuals from Trinidad (West Indies) and then conducted extensive metagenomic, metatranscriptomic, and metabolomics sequencing to analyze their gut microbiome profiles. Our metatranscriptomics approach offers one of, if not, the most deeply

sequenced human gut microbiomes to date (~150 million paired reads on average). Moreover, we obtained a significant amount of high-quality metagenomic and metabolomic data, including 97 GB of metagenomic assemblies, 9,137 high-quality assembled genomes, 916 global metabolites, and 8 short-chain fatty acids abundance data. This data allowed us to identify microbes, microbial genes/pathways, and metabolites that are highly abundant or absent in the Ju/'hoansi population compared to the Western urban population. The data also revealed the presence/absence of microbes, microbial genes/pathways, and metabolites associated with WD that are not associated with the Ju/'hoansi population. All data we generated is freely available for future study by other scientists, including (1) metagenomic sequencing reads, (2) high-quality metagenome-assembled genomes (MAGs), (3) metatranscriptomics sequencing reads, (4) metabolomics data, and (5) organized metadata.

Results:

A multi-omics approach to study complex microbiomes.

In this study, a total of 219 fecal samples were collected for analysis from 55 Ju/'hoansi HG, 16 Bantu AP, 90 healthy and 58 obese/diabetic (BMI≥30kg.m^2 and HbA1c value of ≥6.5%) people of African descent from urban Trinidad, respectively (**Figure 1**). To perform ultra-high-depth metatranscriptomics (MetaT) and a high-depth sequencing approach for metagenomics (MetaG), we used the Illumina NovaSeq 6000 system. After the exclusion of human data, MetaT yielded approximately 30.4 billion microbial read pairs. Similarly, for MetaG, the removal of human data resulted in 10.5 billion microbial reads. MetaSPAdes was then used to generate 219 assemblies totaling 96.6 GB of assembled data. Using a combinatorial bioinformatic approach[26] (**Figure 1**), we identified 9,010 bacterial MAG, 127 archaeal genomes with 726 genomes not classified at the species level, 1025 viral genomes, and 2 eukaryotic genomes. After performing whole-genome average nucleotide identity-based clustering using FastANI, the final clusters delineated 1,184 bacterial species-level clusters (SLC), 7 archaeal SLC, 650 viral (bacteriophage) SLC, and 2 eukaryotic SLC (**Table S1**). The bar plot in Figure 1 displays the counts of genomes at the phylum level. Blue-colored MAGs represent organisms classified at the species level, while those not classified at the species level are depicted in orange using the GTDB-Tk[27]. The MetaT and Taxonomical analysis assessment involved utilizing the VEBA pipeline, a computational pipeline designed for metagenomic analysis [26]. Gene Set Enrichment

Analysis (GSEA) was also employed to identify enzymes, biological pathways, or processes overrepresented in the analyzed gene sets[28]. The set of 916 global metabolites offers a meticulously curated analysis of metabolon pathways, encompassing both super-pathways and sub-pathways. Super-pathways are composite biochemical pathways that elucidate the biosynthesis or metabolism of interconnected compounds, while subp-athways represent a segment of biochemical pathways. In summary, while the microbiome analysis of the combined samples showed mapping to a vast number of microbes present in the Genome Taxonomy Database[29], many of the recovered genomes were absent in such databases, detailing the novelty of high-depth-sequencing approaches for the discovery of unique microbial groups (**Figure 1, and Table S1**).

The gut microbiome of the Ju/'hoansi people.

First, we assessed the microbial abundance of the Ju/'hoansi gut microbiome. The circular representations of the taxonomic and phylogenetic tree of 1,184 bacterial species clusters present among all four groups are shown in **Figure 2A**. The size of clade markers inside the circular plot represents the relative abundance of the bacterial clusters in the Ju/'hoansi gut microbiome. The Ju/'hoansi gut microbiome is mainly composed of bacteria from the Phyla of Firmicutes (47.5%), Bacteroidota (39.8%), Proteobacteria (4.9%), Spirochaetota (2.8%), and Verrucomicrobiota (1.8%). Although Firmicutes is the most abundant and diverse phyla, genera from Bacteroidota (*Cryptobacteroides, Prevotella, Parabacteroides, Bacteroides*), Proteobacteria (*Succinivibrio*) and Spirochaetota (*Treponema_D)* are the most abundant genera. Among the most abundant classified SLC are *Succinivibrio sp000431835* (3.1%), *Cryptobacteroides sp000433355* (2.8), *Prevotella copri A* (1.9%), *Prevotella sp900313215* (1.8%), *Parabacteroides sp900549585* (1.7%), *Prevotella sp900546535* (1.3%), *Prevotella sp900556795* (1.8%), *Cryptobacteroides sp000432515* (1%), *Treponema D succinifaciens* (1%) and *Bacteroides uniformis* (1%) (**Figure 2A**). Additional prokaryotes found in the Ju/'hoansi gut microbiome were Archaea, comprising less than 1% of the microbiome. The most abundant Archaea species in the Ju/'hoansi were *Methanomassillicoccales intestinalis, Methanobrevibacter smithii, Methanosphaera stadtmanae, Methanomethylophilus alvus* (**Figure 2B**). In the case of eukaryotes, only two metagenome-assembled genomes (MAGs) belonging to the species subtype 4 of Blastocystis sp. and the genus

Blastocystis were identified. In the case of DNA/RNA viruses, we observed a major abundance of bacteriophages from the Caudovirales, Microviridae, and Martellivirales (**Figure 2C**).

Microbiome of Ju/'hoansi people is distinct from Western urban African descendants.

When comparing the microbiome of the rural population (Ju/'hoansi and Bantu) and the Western urban (WU, healthy and obese/diabetic), we observed 620 significant (adjusted p-value <= 0.05) differentially abundant SLC between the groups. Among these 620 SLC, 269 are more abundant in the rural population, with 132 SLC having an effect size more than 1.0, and 351 were more abundant in the WU population, with 230 SLC with an effect size greater than -1.0 (**Figure 3A- B, Table S1**). The Ju/'hoansi and Bantu showed higher alpha diversity than the WU healthy group, followed by the obese/diabetic group (**Figure 3C**). The principal component axis (PCA) plot shows a separate clustering of Ju/'hoansi, Bantu, WU healthy and obese/diabetic cohorts, with Bantu showing a 95% confidence interval ellipse overlapping healthy and obese/diabetic cohorts (**Figure 3D**). At the phylum level, Spirochaetota and Campylobacterota were the most abundant in the Ju/'hoansi (1.53, 0.96 effect size) population, followed by Bantu (0.25, 0.18), WU healthy (-0.55, -0.34) and WU obese/diabetic (-0.34, -0.35). The least abundant phyla in Ju/'hoansi were Actinobacteriota and Firmicutes A (-1.64, -1.18), with increasing abundance in Bantu (-1.24, -0.71), WU obese/diabetic (0.47, 0.45) and WU healthy (0.94, 0.67), respectively. Overall, six phyla were more abundant (1.53 to 0.43), and thirteen phyla were less abundant in the Ju/'hoansi as a whole (-0.03 to -1.64) when compared to the other populations (**Figure 3B, 3E, Table S2**).

When looking at the differences in the topmost abundant bacterial species, an increase in *UMGS1225 sp900549725, CAG-390 sp900753295, UBA4636 sp900770945, Fimimonas sp900769665, UBA1259 sp900760875, Treponema D berlinense* and several unclassified species from the genera *Faecousia, HGM11416, Phascolarctobacterium A, Eggerthellaceae, Anaerovibrio, Ellagibacter* (effect size from 1.56 to 1.3) was observed in Ju/'hoansi (**Figure 3A, Table S3**). Among other abundant genera with effect size >1.0 are: *Anaerovibrio, Aphodocola, Cryptobacteroides, ER4, Faecousia, Fimimonas, Gemmiger, Limivicinus, Onthousia, Onthovivens, Phascolarctobacterium A, Prevotella, Scatousia, Stercorousia,* and several unclassified/uncultured genus/species belong to class Acholeplasmatales, Acidaminococcales, Bacteroidales, Christensenellales, Coriobacteriales, Elusimicrobiales, Erysipelotrichales,

Gastranaerophilales, c_Bacilli;o_ML615J-28, Monoglobales, Oscillospirales, Peptostreptococcales, c_Alphaproteobacteria;o_RF32, c_Bacilli;o_RF39, c_Bacilli; RFN20, c_Kiritimatiellae;o_RFP12, Selenomonadales, c_Clostridia;o_TANB77. The majority of prevalent species in the Ju/'hoansi community remain either unclassified or uncultured. (**Figure 3B, Table S3**). Of interest, microbial groups with the highest abundance in the Ju/'hoansi sequentially decreased in the Bantu and the WU healthy and obese/diabetic populations, respectively (**Figure 3A, 3F, Table S3**).

Common probiotic species such as *Bifidobacterium catenulatum, Bifidobacterium angulatum, Bifidobacterium longum, Bifidobacterium bifidum, Bifidobacterium adolescentis, Bifidobacterium pseudocatenulatum* were significantly reduced and/or absent in the rural population when compared to the WU population (**Figure 3A, 3G, Table S3**). Other species with effect size greater than -1.3 in the Ju/'hoansi were *Eggerthella lenta, Faecalitalea cylindroides, Klebsiella pneumoniae, Paraprevotella clara, Alistipes putredinis, Bacteroides salyersiae, Anaerostipes hadrus A, Lachnospira rogosae A, Ruthenibacterium merdipullorum, Copromorpha excrementipullorum, Enorma massiliensis* and species from *Adlercreutzia, Alistipes, Anaerostipes, Avimicrobium, Bacteroides, Bifidobacterium, Collinsella, Copromorpha, Eggerthella, Enorma, Enterocloster, Faecalitalea, Lachnospira, Paraprevotella, Ruthenibacterium, f_Acutalibacteraceae;g_UMGS1071, f_Lachnospiraceae;g_AM51-8* genera (**Figure 3A, 3G, Table S3**). This data conclusively demonstrates the gut microbiome profile of the Ju/'hoansi is distinct from the AP and WU populations, highlighting the significance of this unique microbiome.

Microbial gene expression in the Ju/'hoansi compared to the Bantu and WU populations.

We also characterized the microbial gene expression of the gut microbiome by utilizing MetaT. Among the 620 differentially abundant SLCs, 72,055 orthologs were present in 70% of either the rural or the WU population. Using these 72,055 orthologs, we found 48,114 orthologs differentially abundant between the rural and WU groups (adjusted p-value <= 0.05). From these 48,114 orthologs, we calculated the NES (normalized enrichment score) for various KEGG pathways, modules, ortholog groups, and BRITE hierarchies. There were 141 KEGG orthologs with significant NES scores (adjusted P value < 0.05), 50 KEGG orthologs with NES below -1, and 91 with NES above +1 (**Figure 4A, Table S4)**. The KEGG orthologs with the highest

expression in the rural population (Ju/'hoansi and Bantu) as compared to the WU population (healthy and obese/diabetic) included pyruvate-ferredoxin, elongation factor G/Tu, membrane protein, pyruvate orthophosphate dikinase, DNA-directed RNA polymerase, phosphoenolpyruvate carboxykinase, peptide transport system permease, iron (III) transport system substrate-binding protein, DNA gyrase subunit A, DNA-directed RNA polymerase subunit, preprotein translocase subunit SecY, glyceraldehyde 3-phosphate dehydrogenase, topoisomerase IV subunit A, diphosphate reductase, aldehyde oxidoreductase, H+/Na+-transporting ATPase subunit A/B, oligoendopeptidase F, glucose-1-phosphate adenylyltransferase, CO dehydrogenase maturation factor, iron (III) transport system permease protein, maltose regulon positive regulatory protein, phosphoenolpyruvate carboxykinase, NAD+ oxidoreductase, chaperonin GroEL, methyl-accepting chemotaxis protein, fructose-bisphosphate aldolase, starch synthase, and both small and large ribosomal proteins (**Figure 4A, Table S4**).

At the KEGG BRITE hierarchy level, gene activity related to RNA polymerases, bacterial motility proteins, enzymes of 2-oxocarboxylic acid metabolism, DNA replication proteins, membrane trafficking, glycosyltransferases, messenger RNA biogenesis, prokaryotic defense system, DNA polymerases, and secretion system was more active in the rural populations. (**Figure 4B**). Of the 547 Pfam domains, 293 have NES scores higher than +1, whereas 254 Pfam domains showcase NES scores lower than -1 (**Table S5**). The Pfam domains more expressed in the rural population include multiple pyruvate ferredoxin/flavodoxin oxidoreductase domains, extracellular solute-binding proteins, inner membrane component, ABC transporter, multiple iron/4Fe-4S binding domain, S-layer homology domain, NADH dehydrogenase, Elongation Factor G/Tu, Oligopeptide transporter, Aldehyde oxidase and xanthine dehydrogenase, Bacterial SH3 domain, CoA dehydratase, amino acid transport system, Glyceraldehyde 3-phosphate dehydrogenase, various RNA polymerase domains, PEP-utilising enzyme, MalK OB fold domain, and ATP synthase alpha/beta family. The global metabolites profile comparison shows that metabolites from carbohydrate and nucleotide superpathways exhibit comparable abundance levels in both rural and urban populations. (**Figure 4D**).

The KEGG orthologs with the highest expression in the WU population included Fe^{2+} transmembrane sensor, transposase, macrolide efflux protein, multiple antibiotic resistance proteins, iron complex outer membrane receptor, periplasmic protein TonB, K+:H+ antiporter, RNA polymerase sigma-70 factor, acyl carrier protein, phosphoglycerate mutase, ferritin,

methylglyoxal synthase, LemA protein, REP-associated tyrosine transposase, beta-fructofuranosidase, sensor histidine kinase AgrC, biopolymer transport protein ExbD, regulator for asnA, asnC and gidA, cold shock protein, response regulator YesN, phosphoglycolate phosphatase, transcription termination/antitermination protein NusG, starch synthase, and large subunit ribosomal proteins (**Figure 4A, Table S4**). At the KEGG BRITE hierarchy level, genes involved in lipopolysaccharide biosynthesis proteins and antimicrobial resistance genes were more active in the WU population (**Figure 4B**). Pfam domains more expressed in the WU populations include: Two component regulator propeller, TonB-dependent Receptor, Carboxypeptidase regulatory-like, Bacterial regulatory helix-turn-helix proteins, FecR protein, Starch-binding, SusD family, multiple Tetratricopeptide repeat domain, many Outer membrane proteins domain, Aminoglycoside-2"-adenylyltransferase, ISXO2, Gluconolactonase, Transposase zinc-ribbon, Glycosyl hydrolase, NigD domain, Anaphase-promoting complex MarC family, Arm DNA-binding domain, HU domain, Winged helix-turn-helix, Phage integrase, and many signal peptides (**Figure 4C, Table S5**). The global metabolites profile comparison shows that metabolites in amino acid, lipid, cofactors and vitamins, partially characterized molecules, and peptides superpathway were more active in the WU population with log 2-fold changes ranging from -1.01 to -17.60. The average log 2-fold change of Energy and Xenobiotics superpathways metabolites demonstrates a tendency toward the WU population (**Figure 4D**).

The highly active and defensive Ju/'hoansi gut microbiome lacks the presence of major antimicrobial resistance genes and several typical food metabolites.

The Ju/'hoansi gut microbiome displayed significant metabolic activity and a heightened resilience to stress-inducing environments when contrasted with the Bantu and WU population gut microbiome, showing an effect size twice for respective microbial active genes (**Figure 5B**). Such active pathways and enzymes include DNA-directed RNA polymerase, Bacterial motility proteins, Trehalose biosynthesis, Quorum sensing, ABC transporters, Flagellar assembly, Membrane trafficking, RNA degradation, Mismatch repair, DNA replication, mRNA biogenesis, Bacterial secretion system, Protein export and Secretion system (**Figure 5B**). The Ju/'hoansi group has exhibited significantly higher Trehalose biosynthesis pathway activity than the WU population. Trehalose, a stable disaccharide molecule, is a source of energy and carbon metabolism when responding to abiotic stresses and has been shown to regulate blood glucose levels[30-34]. In addition, the rural population exhibited elevated RNA-based microbial activity, as

evidenced by observed alterations in nucleotide metabolite levels compared to the WU population (**Figure 5C**).

Among cofactors and vitamins subpathways, metabolites from ascorbate and aldarate metabolism, a crucial carbohydrate metabolic pathway that protects cells from oxidative damage[35], nicotinate, and nicotinamide metabolism, which plays a central role in cellular energy metabolism, DNA damage repair, gene expression, and stress response[36] were highly abundant (18.29 log fold change of NAD+) in the rural compared to WU population. Other metabolite subpathways, including hemoglobin and porphyrin metabolism, riboflavin metabolism, thiamine metabolism, and vitamin B6 metabolism, showed similar abundance in rural and WU populations (**Figure 5E**). The Ju/'hoansi have a high abundance of metabolites from the byproducts of the Bacterial Fungal subpathway, including 2-acetamidobutanoic acid, Urolithin B (improves muscle function)[37,38], and glutamyl-meso-diaminopimelate, diaminopimelate which are involved in peptidoglycan synthesis[39] (**Figure 5G**). The rural population surprisingly exhibits a high presence of two caffeine metabolites, namely 5-acetylamino-6-amino-3-methyluracil and 1,3-dimethyluric acid[40], along with metabolites of cannabis such as THC carboxylic acid and 11-hydroxytetrahydrocannabinol[41] (**Figure 5F, G)**. Among plant food metabolites, ferulate and 2,8-quinolinediol were also highly abundant in the rural groups. Three plant metabolites, equol, 4-hydroxycinnamate, and Feruloylputrescine, were abundant only in the Ju/'hoansi (**Figure 5G)**. They possess diverse bioactive properties, including antioxidant, antimicrobial, anticancer, anti-inflammatory, anti-diabetic, anti-melanogenic, and cardioprotective effects, and can inhibit aflatoxin production [42-48].

In the case of the WU population, a significant enrichment of various antimicrobial resistance (AMR) pathways was observed in the WU cohort. These include efflux pump MdtEF-TolC rhodamine 6G, which confers resistance against erythromycin, doxorubicin, and ethidium bromide[49-52]; the efflux pump MexAB-OprM, which confers resistance against carvacrol, quinolones, macrolides, novobiocin, chloramphenicol, tetracyclines, lincomycin, and β-lactam antibiotics[53,54], and the efflux pump AcrEF-TolC, which confers resistance against fluoroquinolone, quinolone, ciprofloxacin[55-58]. The WU was also enriched in aurachin A, a type II polyketide that inhibits the respiratory chain in prokaryotes and eukaryotes, as well as Monobactam, which inhibits peptidoglycan synthesis process biosynthesis pathways (**Figure 5A**)[59]. The WU population exhibits elevated expression levels of various metabolic pathways,

notably including cofactor and vitamin metabolism, tetrahydrofolate biosynthesis, riboflavin biosynthesis, pantothenate, CoA biosynthesis, and glycosaminoglycan metabolism, including heparan sulfate degradation. These pathways were found to be more prevalent in probiotic bacteria, which are in turn more abundant in the WU population, particularly within the healthy cohort (**Figure 5D**)[60].

Metabolite subpathways, including hemoglobin and porphyrin metabolism, riboflavin metabolism, thiamine metabolism, and vitamin B6 metabolism, showed similar abundance in rural and WU populations (**Figure 5E**). Xenobiotic pathways including chemical, drug analgesics, food metabolites and several metabolites including some drugs were abundant in the WU population, including ectoine, acetaminophen, trizma acetate, diglycerol, 3,4-dihydroxybenzoate, carboxyibuprofen, 1-methylxanthine, and diethyl phosphate (**Figure 5F, G)**. Diethyl phosphate is a neurotoxin[61], diglycerol is used as a fat restriction in the management of diabetics[62], it is a food stabilizer, and triethanolamine is found in cosmetic products, herbicides, and algicide[63]. Ectoine helps survival under extreme osmotic stress[64] (**Figure 5G**). The food metabolites prevalent in the WU population, especially in the obese/diabetic group, were advanced glycation end products (AGEs), which are biomarkers of aging, diabetes, diabetic ketoacidosis, and chronic kidney disease. These metabolites consist of pyrraline, nicotianamine, 3-dehydroshikimate, beta-guanidinopropanoate, calystegine A3, quinate, 4-hydroxycyclohexylcarboxylic acid, steviol, deoxymugineic acid, ginkgolic acid C15:1, vanillin sulfate, 5-hydroxymethylfurfural, dihydrocaffeate sulfate, 2-isopropylmalate, furaneol sulfate, 2-keto-3-deoxy-gluconate, pheophytin A, and 2-aminophenol. These metabolites are intermediates in amino acid biosynthesis pathways, aromatic bioproducts, and industrial chemicals. They are naturally present in foods such as potatoes, peppers, paprika, eggplants, strawberries, pineapples, tomatoes, buckwheat, nutritional supplements, and flavoring compounds used in foods, beverages, and pharmaceuticals. Some metabolites have antioxidant, potential allergenic properties, and cytotoxic and carcinogenic effects [65-76] (**Figure 5F, 5G**).

Low Amino acid, Fatty acid, and lipid metabolism pathway expression in the Ju/'hoansi.

The abundance and expression of amino acid pathways were significantly downregulated in rural populations (**Figure 6A-C**). Metabolites derived from the metabolism of leucine, isoleucine, and valine display heightened abundance within the rural population (**Figure 6B and**

Table S6). At the transcriptomics pathway expression level, branched-chain amino acids (BCAAs), including the leucine, isoleucine, and valine degradation pathway, were found more enriched in the Ju/'hoansi population, which supports the high abundance of leucine amino acids and peptides (**Figure 6C and Table S7**). The rural population exhibits abundance in the pathways of Phosphatidylcholine PC and Ceramides lipids (**Table S8)**. In rural areas, the most prevalent lipid was oleoyl-linoleoyl-glycerol, which was found in high abundance (**Table S8)**. Finally, a line plot delineates the prevalence of bacterial species, including novel ones, that have enhanced expression of Leucine, Valine, and Isoleucine pathways in the Ju/'hoansi and the Bantu populations and not the WU population (**Figure 6F)**.

The compounds in amino acids metabolism with the highest abundance (log 2-fold change) in the WU population were: (R)-salsolinol (-18.50), hydroxy-N6,N6,N6-trimethyllysine (-7.63), m-tyramine (-6.43), guanidine (-5.67), 4-imidazoleacetate (-4.86), N2-acetyllysine (-4.51), 3-methylglutaconate (-4.25), N-methylphenylalanine (-4.16) (**Figure 6B, Table S6**). The metabolism of tyrosine, histidine, arginine, proline, lysine, polyamine, methionine, cysteine, taurine, alanine, aspartate, guanidino, and acetamido were higher in the WU population (**Figure 6B, Table S6, S7**). Amino acid pathways enriched in the WU population include methionine biosynthesis, leucine biosynthesis, proline biosynthesis, proline metabolism, tryptophan biosynthesis, shikimate pathway, aromatic amino acid metabolism, isoleucine biosynthesis, histidine biosynthesis, histidine metabolism, arginine and proline metabolism, phenylalanine tyrosine and tryptophan biosynthesis, and valine, leucine and isoleucine biosynthesis (**Figure 6C**). Metabolites of lipids pathways were increased in abundance in the WU population, including androgenic steroids, corticosteroids, endocannabinoid, fatty acid dicarboxylate, fatty acid dihydroxy, fatty acid metabolism acyl, fatty acid monohydroxy, galactosyl glycerolipids, long chain fatty acid, lysophospholipid, lysoplasmalogen, medium chain fatty acid, monoacylglycerol, pregnenolone steroids, primary bile acid metabolism, secondary bile acid metabolism, and sterol (**Figure 6D, Table S8**). The WU population exhibits an abundance of various lipid metabolites, including deoxycholic acid 12-sulfate, chenodeoxycholic acid sulfate, 3-hydroxy sebacate, 1-palmitoyl glycerol (16:0), androsterone sulfate, pregnanediol sulfate, pregnenolone sulfate, 5alpha-androstan-3beta, 17beta-diol mono sulfate, pregnenolone sulfate, and stigmasterol. These lipid metabolites are prevalent in patients with ketoacidosis and are formed from fatty acids through a combination of omega-oxidation, abnormal fatty acid

oxidation, and incomplete beta-oxidation. Additionally, they are present in small quantities in commercial food products, plant sterol additions to food, and various components of olive oil and other vegetable oils. Several of these metabolites are natural steroids, belonging to sulfated steroids, neurosteroid, mono-sulfate, and disulfate fractions[77-84] (**Table S8**). The transcriptomic analysis also showed high fatty acid metabolism and biosynthesis expression, mainly in the WU populations (**Figure 6E**). The WU populations were strongly associated with bacteria involved in the amino acid biosynthesis/metabolism and lipid metabolism pathway (**Figure 6G**).

Highly diverse carbohydrate metabolism and short-chain fatty acid production in Ju/'hoansi gut microbiome flora

In the Ju/'hoansi gut microbiome, an extensive enrichment of microbial genes associated with various carbohydrate metabolism pathways has been identified (**Figure 7A**). The Ju/'hoansi gut microbiome is highly enriched with microbial genes responsible for carbon fixation, pyruvate oxidation, central carbohydrate metabolism, citrate cycle, the De Ley-Doudoroff pathway, gluconeogenesis, the Embden-Meyerhof pathway, incomplete reductive citrate cycle, the Wood-Ljungdahl pathway, the Arnon-Buchanan cycle, the Calvin cycle, the semi-phosphorylative Entner-Doudoroff pathway, beta-Oxidation, short-chain fatty acids metabolism, and reductive, oxidative and non-oxidative pentose phosphate pathways (**Figure 7A)**. The analysis of carbohydrate metabolites identified several compounds that were either involved in or produced during complex carbohydrate metabolism. Among these metabolites are glucose 6-phosphate, N-acetylmuramate, ribose, N-acetylglucosamine, xylose, mannose, N-acetylneuraminate, ribulose/xylulose, and arabinose, which were found in higher abundance in the rural populations (**Figure 7C**). Rural populations have abundant energy metabolites, including Malate, Fumarate, and Alpha-ketoglutarate (**Figure 7D**). The microbial genes found in the energy generation and carbohydrate metabolism pathways **(Figure 7A)**, prevalent among the Ju/'hoansi, consist of either novel species or species with low to negligible presence in WU populations. Out of 70 abundant species in HG Ju/'hoansi, the top five most abundant ones are: *g_Faecousia;s, CAG-390 sp900753295, UBA1259 sp900760875, UMGS1225 sp900549725, g_Anaerovibrio;s_* (**Figure 7E, Table S9**). The abundant taxa at higher taxonomic levels were the genera Faecousia and Limivicinus. The abundant family comprises Oscillospiraceae, CAG-272, and

Ruminococcaceae. At the order level, Oscillospirales, Christensenellales, and Clostridia were the most abundant, with Clostridia being the abundant class.

The pathways and enzymes enriched in the WU population include succinate dehydrogenase, central carbon metabolism in cancer, glyoxylate cycle, glucagon signaling, dTDP-L-rhamnose biosynthesis, galactose metabolism, and NAD biosynthesis (**Figure 7B**). Multiple metabolites, such as glycerate, galactonate, arabitol/xylitol, mannitol/sorbitol, lyxonate, N6-carboxymethyl lysine, arabonate/xylonate, play a role in advanced glycation end product formation, representing an intermediate step in sugar acid metabolism. These metabolites were notably prevalent in thermally processed foods, sugar alcohols used as sweeteners, milk sugars, and their derivatives, particularly in the WU population [85-87] (**Figure 7C**). WU populations had abundant succinate, 2-methylcitrate, isocitric lactone, and tricarballylate energy metabolites (**Figure 7D**). The pathways and enzymes enriched in the WU population are distributed across 48 species **(Figure 7B, Table S9)**. The most abundant species in the WU population are *Bacteroides salyersiae*, *Alistipes putredinis, Acidaminococcus intestini*, *Bacteroides stercoris, Paraprevotella clara, Alistipes putredinis,* and belong to Bifidobacterium, Collinsella genera (**Figure 7F**).

The data analysis reveals a significant abundance of all seven short-chain fatty acids in rural populations, with a calculated p-value of less than 0.05. Hexanoic acid is especially noteworthy, displaying the highest fold-change at 10.8, indicating a substantial increase. Its mean abundance is reported at 124.29 µg/g in the HG Ju/'hoansi population, 136.52 µg/g in the AP Bantu population, 10.38 µg/g in the WU Healthy group, and 13.66 µg/g in the WU Obese/Diabetic cohort. Isovaleric acid closely follows this trend, exhibiting a fold-change of 8.3 and a mean abundance of 226.72 µg/g in the HG Ju/'hoansi group, 283.59 µg/g in the AP Bantu population, 27.30 µg/g in the WU Healthy cohort, and 30.49 µg/g in the WU Obese/Diabetic cohort. Isobutyric acid also shows a significant fold-change of 7.3, with a mean abundance of 196.70 µg/g in the HG Ju/'hoansi group, 258.99 µg/g in the AP Bantu population, 27.42 µg/g in the WU Healthy group, and 30.93 µg/g in the WU Obese/Diabetic cohort. Valeric acid (reported abundances: HG-324.68, AP-379.64, WU-46.45/56.53) and 2-methylbutyric acid (HG-141.50, AP-177.50, WU-21.35/24.17) demonstrate similar fold-changes at 6.6. Additionally, the three most abundant SCFAs, propionic acid (abundances: HG-1407.98, AP-1566.00, WU-203.14/263.46), acetic acid (abundances: HG-2790.42, AP-3611.25, WU-416.50/564.58), and

butyric acid (abundances: HG-1436.47, AP-1309.79, WU-206.48/260.37); all show substantial levels in the rural populations **(Figure 7K)**.

Ju/'hoansi shows a low abundance of microbes associated with biomarkers and pathways of chronic inflammatory diseases.

We utilized Maaslin2 to identify potential microorganisms and pathways that are associated with lipid metabolism, HbA1c, and inflammation. There are 27 microbes that are associated with lipid metabolism, including cholesterol to HDL (high-density lipoprotein) ratio (CHOLD), direct high-density lipoprotein cholesterol (DHDLC), triglycerides, and very low-density lipoprotein cholesterol (VLDL) as shown in **Figure 7G**. Of the 27 identified microbes, only 5 exhibit higher prevalence among the rural population, all distinct microbial strains. Concerning inflammation markers, 20 distinct microbes were found to be linked with TNF Alpha, while 6 microbes were associated with C-reactive protein (CRP). Among the microbes associated with TNF Alpha, only 4 distinct species, 3 of which exhibited an effect size of less than 1.0, were prevalent in the rural population. In contrast, all CRP-associated microbes with effect sizes greater than 1.0 were prevalent in the WU population (**Figure 7H**). Among the 14 microbes associated with HbA1C, only 4 distinct microbes, with an average effect size of 0.9, were prevalent in the rural population despite a high prevalence of carbohydrate-metabolizing microbes in this population. The remaining 10 microbes, with an average effect size of -1.2, were abundant in the WU population (**Figure 7I**). A comparative analysis of effect sizes, depicted in a box plot, reveals that the mean effect size is notably lower in the Ju/'hoansi cohort, followed by the Bantu, WU healthy, and obese/diabetic subjects, in descending order. Significant differences were evident in the abundance of microbes associated with CRP, TNF Alpha, and DHDLC **(Figure 7J)**. These findings imply that most microbes linked to inflammation, lipid, and sugar metabolism were more abundant within the WU population than the rural population.

In order to minimize the possibility of false negatives, we followed a rigorous approach in which only those microbial genes exhibiting statistically significant enrichment in the gene set enrichment analysis (GSEA) pathway undergo thorough testing for their potential association with inflammation markers. By focusing specifically on these enriched genes, we aimed to capture a comprehensive understanding of the relationship between microbial genetic factors and inflammatory processes. The genes we have observed to be enriched in galactose metabolism,

glucagon signaling, and central carbon metabolism pathways in cancer have been identified to be associated with TNF Alpha and have significantly higher occurrence among the WU population (**Table S10**). Among the microbial genes associated with CRP abundance, only one gene, the bifunctional 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, present in the isoprenoid and terpenoid biosynthesis pathway, exhibits heightened expression in the rural population. (**Table S10**). Genes that were highly expressed in the WU population were enriched for the central carbon metabolism in cancer, multidrug/antimicrobial resistance, aromatic amino acid metabolism, glyoxylate cycle, lipopolysaccharide metabolism and biosynthesis, cofactor, and vitamin metabolism, pantothenate biosynthesis and shikimate pathways which links metabolism of carbohydrates to biosynthesis of aromatic compounds (**Table S10**). The expression levels of CRP-associated microbial genes were notably lower in the HG Ju/'hoansi population, with subsequent increased expression observed in the AP Bantu and WU populations.

Discussion

The Ju/'hoansi are among the few remaining groups of hunter-gatherers in the world. While some of their subgroups have transitioned away from a traditional hunter-gatherer lifestyle, their dietary patterns strongly resemble those of their ancestors. This study provides a multi-omics characterization of the gut microbiome of the Ju/'hoansi and compares it to geographically close AP and African descendants from the island of Trinidad in the Caribbean. The findings of this study offer a valuable representation of ancient microbiomes lost due to the impact of Western lifestyles, medicine, and dietary patterns. In addition, this study provides the scientific community with a vast amount of multi-omic data, including metagenomic sequencing reads, metatranscriptomics sequencing reads, global metabolomics data, and detailed metadata to enable secondary analysis by other groups. This research holds significant relevance in elucidating the composition of the primitive microbiome and the impact of industrialized food on the optimal functionality of the gut microbiome.

Our research findings indicate that the Ju/'hoansi population has a significantly greater abundance and diversity of gut microorganisms compared to the geographically proximate AP group, and this distinction is even more pronounced when contrasted with individuals from the Caribbean. Higher bacterial and fungal diversity via 16S and ITS sequencing was also observed in a recent Ju/'hoansi study in the Nyae Nyae Conservancy in northeastern Namibia[25]. The study

highlights a significant presence of diverse genera in the HG gut microbiota, demonstrating their crucial role in breaking down complex carbohydrates from plants and producing essential short-chain fatty acids. Notably, several predominant species inhabiting the HG are actively engaged in the metabolism of complex carbohydrates and the generation of energy through various metabolic pathways, encompassing carbon fixation, pyruvate oxidation, the pentose phosphate pathway, the citrate cycle, glycolysis, as well as central carbohydrate metabolic pathways and genes associated with the production of short-chain fatty acids. The high functionality of the HG gastrointestinal microbiome is punctuated by the abundance of specific metabolites such as urolithins, equols, 4-hydroxycinnamate, and feruloylputrescine. These metabolites have been demonstrated to enhance mitochondrial activity and mitophagy, regulate estrogenic activity, and exhibit anti-melanogenic, anti-inflammatory, and antioxidant properties, thereby mitigating the systemic effects of aging and age-related chronic conditions[42-48,88,89]. Numerous beneficial metabolites from bacterial-fungal subpathways were abundant among the Ju/'hoansi. One particular metabolite, 4-acetamidobutanoic acid, derived from the gamma-amino acid pathway, is recognized for its antioxidant and antibacterial properties and its capability to impede the proliferation of pathogens[90]. Our comprehensive multi-omic analysis revealed elevated gene expression of innate prokaryotic defense systems in both the Ju/'hoansi and AP groups, contrasting with increased gene expression of antibiotic-resistant genes in the western groups. This discrepancy may be attributed to the extensive use of antibiotics in the Western diet [91]. These findings collectively suggest that the Ju/'hoansi population possesses a more intricate metabolic repertoire, potentially contributing to their lower incidence of acute and chronic diseases.

The ingestion of animal products and saturated fats provokes significant alterations in androgen release within the body[92]. On the other hand, a dietary plan that includes higher dietary fiber, reduced intake of refined carbohydrates, decreased consumption of trans and saturated fats, and the addition of omega-3 and omega-9 fatty acids has demonstrated the potential to improve the normalization of the androgen profile. [93]. Steroid hormones, such as corticosteroids and lipid-based neurotransmitters, including endocannabinoids, play pivotal roles in modulating stress, immune responses, and metabolism. These compounds are integral in regulating energy balance and food intake functions[94,95]. The Ju/'hoansi and the AP groups benefit from a fiber-rich diet while minimizing intake of refined carbohydrates, trans fats, and saturated fats. The metabolic

pathways observed in the Trinidad samples predominantly contribute to producing medium- and long-chain fatty acids. Research indicates that these fatty acids can instigate alterations in the composition of gut microbiota, consequently influencing the physiological and immunological aspects of the host[96,97]. The gut microbiota controls the synthesis of these compounds within the intestinal tract. They have the potential to disrupt cellular membranes, induce lipid peroxidation, facilitate the progression of cancer, and modify cellular signaling pathways[98,99]. Many of these compounds are found in vegetable oils, animal fats, and dairy products and are commonly included in small quantities in various commercial Western food items.

In addition to metabolic complexity, the microbiomes of the Ju/'hoansi and AP groups exhibit a significantly increased production of short-chain fatty acids (SCFAs), compared to the obese and healthy cohorts in Trinidad. This heightened SCFA production is attributed to the enzymatic breakdown of carbohydrates by bacteria into simple sugars, which are then fermented to generate SCFAs. This research underscores the potential health implications and benefits associated with understanding and harnessing the metabolic activities of the microbiome. These fatty acids account for up to 10% of the caloric requirements of human cells and play a critical role in maintaining intestinal health. SCFAs are associated with improved gut barrier integrity and enhanced glucose and lipid metabolism. Furthermore, they regulate the immune system, inflammatory response, and blood pressure. In addition, SCFAs play a crucial role in providing energy to colonic cells, promoting the development of tight junctions, facilitating gut regeneration, and exerting direct or indirect influence on gut-brain communication and brain function. Finally, SCFAs enhance T_{reg} activity, improve mitochondrial metabolism, promote keratinocyte differentiation, reduce inflammatory factor expression, and inhibit the inflammatory response[100-103]. The data from our study collectively provides a direct mechanistic link to the lack of reported age-associated chronic inflammatory diseases in the Ju/'hoansi and the AP groups. This suggests a potential avenue for further in-depth exploration of the specific underlying biological mechanisms contributing to this phenomenon in these populations.

The microbiome of the Ju/'hoansi people exhibits a deficiency in microbes associated with fatty acid metabolism, as well as blood inflammation markers TNF Alpha and CRP. The enriched pathway associated with the inflammation marker TNF Alpha, glucagon, regulates glycemia, amino acid, and lipid metabolism, suggesting a higher abundance in urban populations. The CRP-associated enriched 81 pathway, another blood inflammation marker,

encompasses crucial elements such as multidrug/antimicrobial resistance, amino acid metabolism, and lipopolysaccharide metabolism/biosynthesis. While the association coefficient between microbes and microbial genes may be small, the results of differential abundance strongly reinforce the connection to lipid metabolism and blood inflammation biomarkers. The study is subject to certain limitations, primarily stemming from the small sample size of the agropastoral population. Despite this, notable differences were observed between the agropastoral and the other three groups. Analysis indicates that the microbiome of the agropastoral population falls between the Ju/'hoansi and the urban population. The observations indicate a potential progression in the microbiome, possibly originating from the rural Ju/'hoansi population and exhibiting an increase or commencing from the urban population and displaying a decrease as it approaches the Ju/'hoansi. Additionally, a limitation of the study involves merging two different metabolomics processed batches, potentially introducing biased analysis. However, the complementary nature of our metatranscriptomics and metabolomics analyses suggests an absence of biased analysis due to the use of different metabolomics batches. It is important to note that the findings from this study are preliminary and would necessitate validation in a case-controlled study using a model organisms or with additional sampling, however, the latter may be restricted by lifestyle, geography and environmental changes.

In summary, our study results provide compelling evidence supporting the unique functionality of the Ju/'hoansi microbiome, which appears to have diminished in the context of industrialized societies. Specifically, we observed distinct microbial profiles associated with the Ju/'hoansi and, to some extent, the AP group that hold potential as live therapeutics for addressing chronic inflammatory diseases prevalent in Western lifestyles. This suggests a potential avenue for targeted interventions. Moreover, this study shows that the differences between healthy and obese/diabetic populations are minimal when compared to the Bantu and the Ju/'hoansi, suggesting that there is a microbial bottleneck for the western microbiome. Our findings also offer valuable insights into the methodologies and datasets necessary for future studies and secondary analyses. These would play a pivotal role in fully elucidating the complex relationships between the microbiome, dietary habits, and environmental factors in the context of health and disease.

Indigenous community engagement

In order to ensure that scientific discoveries and understanding truly represent all populations rather than being limited to industrialized nations, it is imperative to involve indigenous communities in research endeavors. This approach is crucial to address potential biases and ensure that the benefits of scientific progress are shared equitably among diverse communities, thereby fostering a more inclusive and comprehensive representation of research. In order to maintain ethical standards and prevent exploitation, it is crucial to consider specific considerations when conducting this research. In this study, we performed deep metagenomics, metatranscriptomics sequencing, and metabolomics on anonymized fecal samples collected from Ju/'hoansi and Bantu agropastoralist populations from Namibia in 2019. Dr. Hayes has and continues to have, a close working relationship with the communities, with this study forming a part of a larger volume of work across the region. Specifically, Dr. Hayes has been working with the Ju/'hoansi community participants of this study for over a decade, beginning in 2008. The research presented in the outline of this project has also been communicated back to the community during engagement visits post-COVID-19. Born and educated in neighboring South Africa with a Namibian family, Dr. Hayes is not only well acquainted with the country and cultures, but she also shares a common language with the community, allowing for direct discussions and ensuring a deep level of community involvement and understanding. Dr. Hayes is further supported by a well-established local team, which includes key community members who facilitate site visits and ensure further study translation into the local click-language. As per Ju/'hoansi culture, decisions regarding participation are made at a community rather than at an individual level. The study was conducted without any biases that could have impacted its findings. All eligible participants residing within the focused communities who expressed willingness to participate were included in the study.

Declarations

Funding

This project was funded National Institute of Diabetes and Digestive, and Kidney Diseases grant R01 DK112381 01. The funding bodies did not have a role in study design, data collection and analysis, the decision to publish, and the manuscript's preparation.

Author Contributions

Managed study participant's consent and sample collection: VH, RH, HF, JF, DR. Performed computational analysis: HS, JE. Wrote the paper: HS, NG-J. Provided input to the manuscript: VH, SV, KEN. Laboratory work: RR, CK, AA. Conceived, designed, and supervised: HS, NG-J.

Conflict of Interest Statement

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

Ethics approval and consent to participate.

This study was reviewed and approved by the in-house Institutional Review Board at Garvan Institute of Medical Research (protocol number #280/2017), by the Campus Ethics Committee at the University of the West Indies at St Augustine (reference number: CEC262/08/16) and JCVI (protocol number 2019 277).

Consent for publication

Informed consent (and assent, where appropriate) and authorization to use, create, and disclose health information for research was obtained from and documented for each research participant enrolled to study at the Garvan Institute of Medical Research and the University of the West Indies at St Augustine, Trinidad.

Acknowledgments

We in dept to the study participants and their representative communities for their contribution and continued engagement. We are grateful to C.P. Bennett from Evolving Picture in Sydney (https://evolvingpicture.com/) for providing the community recording required to meet local ethical requirements and facilitating with community collections in Namibia, while we acknowledge the community members who have provided further logistical assistance and cultural insights, including B. G/aq'o, N. /kun, J. /kunta, A. Oosthuysen and E. Oosthuysen. We thank the hard working team members of the Department of PreClinical Sciences, University of the West Indies, for their dedicated support.

Data availability

Materials and Methods:

Description of sample collection and categorization into groups.

In a comprehensive study, we carried out a thorough investigation that involved enrolling and meticulously collecting stool samples from a specific cohort that comprised 61 Ju/'hoansi individuals, who are indigenous people residing in the Kalahari Desert of Namibia, located in southern Africa, as well as 16 agropastoral Bantu individuals living in the immediate vicinity. The collection of samples was undertaken in two phases, with 16 samples obtained in February 2019 and an additional 45 samples gathered in September 2019. These samples were collected for the purpose of the research. Out of the 61 samples, 33 were obtained from female participants, while 28 were obtained from male participants. Additionally, the participants were further categorized based on age, revealing that 30 individuals were over 45 years old and 31 individuals were under 45 years old.

The stool samples were collected on-site with meticulous attention to detail and subsequently preserved in liquid nitrogen tanks to ensure their integrity for subsequent analysis. The study protocol received ethical approval from the Garvan Institute of Medical Research in New South Wales, Australia, as well as from JCVI/NIH (2019-277). Dr. Hayes and Dr. Förtsch were granted the necessary permit to collect the samples from the Ministry of Healthy and Social Services (MoHSS Reference: 17/3/3 HEAF).

In Trinidad, a cross-sectional convenience sampling approach was used to recruit study participants characterized as individuals of African-descent (as defined by having at least 3 grandparents of pure African ancestry; *Miller et al. 1989*) who, for the obese/diabetic population, were diagnosed with T2D and/or being obese (BMI≥30kg.m$^{2;}$ n=75); and for the control healthy population (n=75). The study protocol was approved prior to start of recruitment by the following research ethics boards: The University of the West Indies; North Central Regional Health Authority (NCRHA); Southwest Regional Health Authority (SWRHA) and JCVI/NIH. The protocol was modified to adhere to public health guidelines during the Covid-19 pandemic.

Inclusion criteria: Individuals, aged 21- 60 years of age during consent process, of African-descent (as defined by having at least 3 grandparents of pure African ancestry; *Miller et al. 1989*)

who have been diagnosed with type 2 diabetes (T2D) and/or with a BMI≥30kg.m2; and healthy controls.

Exclusion criteria: Pregnant women. Individuals with recent antibiotic use; recent history (past 3 months) of gastrointestinal infection or diarrhea; individuals consuming large doses of probiotics. History of: cancer, neurological complications; significant weight loss (>5kg); major dietary changes; chronic alcohol consumption; recent major surgery or hospitalization for more than 24hrs. Positive for Covid-19, HIV, HBV or HCV. History of immunosuppression. Unwillingness to provide blood and stool sample.

For controls: in addition to the above: diabetes, markers of metabolic syndrome.

Recruitment

This was done on a phased basis whereby Controls were recruited first, whilst administrative approvals for visiting health centers were sought to recruit diabetic participants. This phase began during the COVID-19 pandemic, so an electronic flyer detailing the study, its benefits, criteria, and stipend was generated and distributed through the University of the West Indies (UWI) community via its weekly Marketing and Communication electronic newsletter. This flyer was also shared by project members through their social media. Subsequently in the second phase of recruitment a separate electronic flyer was designed specifically to recruit diabetic participants. This second flyer was also circulated by UWI, as above, and in collaboration with the Diabetes Association of Trinidad and Tobago (DATT), shared with its members utilizing the DATT Facebook page. Physical copies of the flyers designed for recruiting participants were placed at high visibility and high traffic areas throughout the UWI as well as in select areas of the North Central Regional Health Authority (NCRHA). During this period nondiabetic controls continued to be recruited through use of the electronic flyers. During Phase 3, we were able to conduct recruitment of diabetic patients through RHA databases. Further, with the end of the global COVID-19 pandemic, restrictions in public health facilities were eased and distribution of physical flyers for diabetic participants was then expanded to include the hospitals and clinics of both the North Central Regional Health Authority (NCRHA) and South West Regional Health Authority (SWRHA) catchments. These physical flyers were also distributed through the private practices of a select few medical practitioners that were willing to do so. The diabetic participants in this study were mainly recruited from health centers and clinics in the NCRHA.

Screening

Once potential participants agreed to study conditions, having been self-identified or identified by the research group as meeting the study requirements, screening was done to determine eligibility via a project specific screening questionnaire (Appendix A) which was administered either in person or over the phone. If eligibility criteria were met and participants agreed to take part in the study, they were scheduled to make their first visit to the research study site. Screening interviews during recruitment were conducted using standard social distancing protocol (recruiters wore PPE and maintained a distance of 2m). Interview questionnaires also ascertained whether potential participants had, or did have, COVID-19 or flu-like symptoms and/or came into contact with someone who had the disease, and required a history of travel to COVID-19 dense regions in the period between January 2020 and the date of interview.

Study visit

Once successfully screened, participants were invited. Upon arrival introductions were shared and formally explained the project to the participants. Initial questions were answered. Signed informed consent forms were obtained and stool collection kits were distributed along with the instructions on its use and handling. These kits were red insulated lunch bags, containing two Omnimet gut stool collection kits (DNA Genotek Inc. Ref #ME-200), and a reusable ice pack. Participants were asked to keep the sample cool/frozen once collected. The kits contained a collection swab/spatula, collection vials with stabilizers, and collection assistance paper. After an explanation of the kits and their use, participants were offered a chance to withdraw from the study should this be too difficult or uncomfortable for them.

For sample collection, the participants had an overnight (14 hours) fast then the following morning handed over the collected stool samples. At the study site a questionnaire was administered (Appendix C), along with biometric data collection and blood sampling by the project's phlebotomist. The participants were initially asked when their last meal was, to ensure they were in a fasting state. After which, the stool samples were checked to ensure both kits were used and were in good condition. Once these checks were done, the study questionnaire was administered to collect demographic information (e.g. age, gender, and occupation). Additionally, baseline medical and medication history were collected such as family medical history, obstetric history, use of medication/supplements and dietary habits. Next, a quick physical examination was completed and included a rapid surface examination and mucous membranes. Participant's height, weight, blood pressure and heart rate were then recorded using digital and analog scales.

Height was measured using a stadiometer. Blood pressure and heart rate were recorded using an Omron Automatic Blood Pressure Monitor.

Statistical analyses of sample collection

Data entry: participants were assigned unique identifiers and identifiers were removed from the data files. Participants were assigned to one of two study groups: Obese/Diabetic defined as any participant with BMI ≥30 kg.m^2, or with an HbA1c value of ≥6.5%; Healthy Controls (denoted as just "Healthy in the manuscript") were defined as participants with BMI <30kg.m^2 and not previously diagnosed as type 2 diabetic. All data were validated by a second team member using printouts of the data sheets. Biochemical variables were tested for normality prior to analyses, following which cross tabulation and independent sample t-test were performed to assess significant differences between the two study groups. An adjusted p. value (FDR) < 0.05 were deemed significant. A total of 116 healthy controls responded to advertisements and expressed interest in being a study participant. Of those, 103 responded to follow up and were screened; 78 were deemed eligible and 75 completed the study. A total of 43 obese or type 2 diabetics responded to the advertisements and social media posts; 35 who expressed interest were screened and 27 of those were deemed eligible to participate in the study. The remaining participants in this group were recruited through patient NCRHA data base of patients at the chronic disease clinics in close proximity to the UWI.

Workstation Decontamination for Isolation and Library Preparation

DNA/RNA isolation and library preparations, the procedures were carried out in a disinfected Class II Biosafety cabinet or Purifier Vertical Clean Bench (PCR Hood) treated with Eliminase (Cat# 1101, Decon Labs Inc., PA, USA) and 70% ethanol afterward exposed to UV light for 30 minutes.

DNA Extraction of Human Stool Samples

DNA was extracted according to protocol using QIAGEN DNeasy PowerSoil Pro Kit (Cat# 47016, QIAGEN, Hilden, Germany) with mechanically lysed using QIAGEN PowerLyzer 24 Homogenizer (110/220 V) (Cat# 13155, QIAGEN, Hilden, Germany) at 3000 rpm for 30 seconds. After extraction, DNA was eluted in 50uL of C6 buffer (10 mM Tris). Samples were

quantified using Qubit HS Assay (Cat# Q32856, ThermoFisher Scientific, Waltham, MA, USA) and DNA quality was measured using an Agilent Genomic DNA TapeStation (Cat# 5067-5366 and 5067-5365, Agilent Technologies, Santa Clara, CA, USA).

Metagenomic Library Preparation

Metagenomic libraries were generated using the NEBNext Ultra II FS DNA Library Prep Kit (Cat# E7805L, New England Biolab, Ipswich, MA, USA) with the NEBNext Multiplex Oligos for Illumina (Cat# E6448S, New England Biolab, Ipswich, MA, USA). Library prepping was carried out according to protocol for use with gDNA input <100ng. As instructed, all samples were diluted for an estimated 100ng in 26uL volume for starting reaction. Fragmentation was carried out at 37°C for 15 minutes. For adaptor ligation, the NEBNext Adaptor for Illumina were diluted 10-folds in 10 mM Tris-HCl, pH 7.5-8.0 with 10 mM NaCl (recommended for samples input gDNA <100ng). Samples were cleaned using Beckman Coulter Ampure XP beads (Cat# A63881, Beckman Coulter, Pasadena, CA, USA) at a 0.9X ratio. PCR enrichment was carried out by setting the thermal cycler at the following conditions: 98°C for 30 seconds, 98°C for 10 seconds, 65°C for 75 seconds for 4 cycles, 65°C for 5 minutes. Afterwards, libraries were cleaned once more using Ampure XP beads at a 0.9X ratio. Finally, libraries were quantified using Thermo Fisher Qubit 1X dsDNA HS Assay Kit (Cat# Q33231, ThermoFisher Scientific, Waltham, MA, USA) and average fragment size was estimated using Agilent Bioanalyzer High Sensitivity DNA Analysis (Cat# 5067-4626, Agilent Technologies, Santa Clara, CA, USA). Libraries were manually normalized based on the DNA concentration and average fragment size. The pooled library concentration was estimated by qPCR using KAPA Library Quantification kit for Illumina (Kapa Biosystems; Roche Diagnostics Corporation, Indianapolis, IN, USA). Finally, the library was loaded onto a NovaSeq 6000 S2, 300 cycles (2x150 bp) v1.5 as instructed by the manufacturer (Cat# 20028314, Illumina Inc., La Jolla, USA).

RNA Extraction of Human Stool Samples

RNA was extracted according to protocol using Zymo Quick-RNA Fecal/Soil Microbe Microprep Kit (Cat# R2040, Zymo Research, Irvine, CA, USA). Samples were lysed by bead beating using Vortex Genie-2 set at max speed for 10 minutes. After extraction, RNA was eluted in 15uL of DNase/RNase-Free Water. Samples were quantified using Thermo Fisher Qubit RNA

HS kit 100 assays (Cat# Q32852, ThermoFisher Scientific, Waltham, MA, USA). RNA integrity number (RIN) was measured using Agilent Bioanalyzer RNA 6000 nano kit (Cat# 5067-1511, Agilent Technologies, Santa Clara, CA, USA).

RNA Library Prepping

RNA libraries were generated using llumina's Stranded Total RNA Prep, Ligation with Ribo-Zero Plus (Cat# 20040529, Illumina Inc., La Jolla, USA) with IDT for Illumina RNA UD Indexes Set A and B (Cat# 20040553 and 20040554, Illumina Inc., La Jolla, USA). Library prepping was carried out according to protocol with some adjustments to fragmentation (fragmentation time was reduced to 30 seconds) and PCR amplification (set to 16 cycles) [RR1] steps because of samples low-quality RNA (the average RIN value was 2.5). Regarding the samples low-quality RNA, input volume was set to a total of 100ng in 11uL starting volume.

After library prepping, samples were quantified using Thermo Fisher Qubit 1X dsDNA HS Assay Kit (Cat# Q33231, ThermoFisher Scientific, Waltham, MA, USA) and average fragment size was estimated using Agilent Bioanalyzer High Sensitivity DNA Analysis (Cat# 5067-4626, Agilent Technologies, Santa Clara, CA, USA). Libraries were manually normalized based on the DNA concentration and average fragment size. The pooled library concentration was estimated by qPCR using KAPA Library Quantification kit for Illumina (Kapa Biosystems; Roche Diagnostics Corporation, Indianapolis, IN, USA). Finally, the library was loaded onto a NovaSeq 6000 S4, 300 cycles (2X150) v1.5 as instructed by the manufacturer (Cat# 20028312, Illumina Inc., La Jolla, USA).

Metabolomics using the Metabolon Platform:

Metabolon received 61 frozen feces and 147 OMNI-met feces samples. Global metabolic profiles comprise a total of 915 compounds of known identity (named biochemicals) were determined. The processing of the samples and the identification of metabolites were carried out by Metabolon in accordance with the specified procedures.

Sample Accessioning: Following receipt, samples were inventoried and immediately stored at -80^{o}C. Each sample received was accessed into the Metabolon LIMS system and assigned by the LIMS a unique identifier that was associated with the original source identifier only. This

identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80^{o}C until processed.

Sample Preparation: Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added before the extraction process's first step for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

QA/QC: Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS): All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer

interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

Bioinformatics: The informatics system consists of four major components: the Laboratory Information Management System (LIMS), the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by data analysts. The hardware and software foundations for these informatics components were the LAN backbone and a database server running Oracle 10.2.0.1 Enterprise Edition.

LIMS: The purpose of the Metabolon LIMS system was to enable fully auditable laboratory automation through a secure, easy to use, and highly specialized system. The scope of the Metabolon LIMS system encompasses sample accessioning, sample preparation and instrumental analysis and reporting and advanced data analysis. All of the subsequent software systems are grounded in the LIMS data structures. It has been modified to leverage and interface with the in-

house information extraction and data visualization systems, as well as third party instrumentation and data analysis software.

Data Extraction and Compound Identification: Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. These systems are built on a web-service platform utilizing Microsoft's .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (*m/z)*, and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

Curation: A variety of curation procedures were carried out to ensure that a high quality data set was made available for statistical analysis and data interpretation. The QC and curation processes were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, mis-assignments, and background noise. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

Metabolite Quantification and Data Normalization: Peaks were quantified using area-under-the-curve. For studies spanning multiple days, a data normalization step was performed to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected in run-day blocks by registering the medians to equal one (1.00) and normalizing each data point proportionately. In certain instances, biochemical data may have been normalized to an additional factor (e.g., cell counts, total protein as determined by Bradford assay, osmolality, etc.) to account for differences in metabolite levels due to differences in the amount of material present in each sample.

Batch normalization and Imputation of Data:

Normalizing the data is crucial to eliminate instrument batch effects, especially since we have merged data from two different collection samples. This step will ensure the accuracy and integrity of our analysis. For each metabolite, the raw values in the experimental samples are divided by the median of those samples in each instrument batch, giving each batch and, thus, the metabolite a median of one. To remove batch variability, for each metabolite, the values in the experimental samples are divided by the median of those samples in each instrument batch, giving each batch and, thus, the metabolite a median of one. For each metabolite, the minimum value across all batches in the median scaled data is imputed for the missing values. For each sample, the Batch-normalized data is divided by the value of the normalizer. Then, each metabolite is re-scaled to have a median = 1 (divide the new values by the overall median for each metabolite). Then, imputation is performed. The batch-norm-imputed data is transformed using the natural log. Metabolomic data typically displays a log-normal distribution; therefore, the log-transformed data is used for statistical analyses.

Assembly, classification, and dereplication for metagenome-assembled genomes

We used VEBA v1.2.0 to pursue a metagenomics approach by assembling and binning each metagenome separately, performing species-level clustering, and a within SLC orthology analysis [26,104]. Using VEBA's assembly.py module, each metagenome was assembled per sample using SPAdes v3.15.2 (metaSPAdes mode)[105]. Viruses were identified and classified using geNomad v1.3.3 [106] (VEBA's binning-viral.py and classify-viral.py modules). Prokaryotes were binned using several binning algorithms, including: 1) MaxBin2 v2.2.7[106], 2) Metabat2 v2.15[107],

and 3) CONCOCT[108]. The binning assignments were refined using DAS Tool v1.1.2[109], yielding 9,011 MAGs that were classified using GTDB-Tk v2.2.3 (reference database: R207[27]) and quality assessed using the lineage_wf pipeline of CheckM v1.1.3[110] (VEBA's binning-prokaryotic.py and classify-prokaryotic.py modules). Eukaryotic genomes were binned using Metabat2 v2.15, quality assessed using BUSCO v5.4.3[111], and classified against VEBA's microeukaryotic protein database (VEBA's binning-eukaryotic.py and classify-eukaryotic.py modules). Both prokaryotic and eukaryotic genomes were only used if they were at least medium quality according to the guidelines established by the Genomic Standard Consortium (>50% completeness and <10% contamination;[112]. The recommended cutoffs in CheckV for considering viral contigs are the following: 1) the number of viral genes > 0; 2) the number of viral genes ≥ 5 times the number of host genes; 3) completeness ≥ 50%; 4) checkv_quality and miuvig_quality are above medium quality[113]. Only non-proviruses were used in downstream analysis to prevent host bias.

Genome and protein-level clustering was performed using VEBA's cluster.py module. The dereplication of redundant species was clustered using FastANI v1.32[114] with a cutoff of ≥ 95% ANI; as the authors recommended, connected components were determined using NetworkX v2.5. FastANI clustering was separately prokaryotic, eukaryotic, and viral organisms to yield SLCs. For each SLC, the proteins within the pangenome were clustered into SSPCs using MMseqs2 with a query coverage of 80% and percent identity of 50%[115].

Gene models, functional annotation, and orthology analysis

VEBA's binning modules handle all the gene modeling in the backend. To be specific, gene models were created for prokaryotic genomes using Prodigal v2.6.3 (--meta mode;)[116], viral genomes using Prodigal-GV v2.10.0[106], and eukaryotic genomes with MetaEuk generating putative proteins and GFF files used for annotations and read mapping[117]. The putative proteins were annotated using both Diamond v2.1.7[118] against UniRef90[119] and KofamScan v1.3.0[120].

Read preprocessing and mapping to the genome catalog

Using VEBA's index.py and mapping.py, we aligned reads to the contigs for each genome to assess abundance and expression for metagenomics and metatranscriptomics datasets, respectively. Bowtie2 v2.5.2[121] was used for all read mapping to MAGs. Read counts tables were

constructed with featureCounts v2.0.6[122] using the bam files generated from Bowtie2, the assembly fasta, and the gene models. Counts for SLC or MAGs were computed by summing contig-level read counts using a SAF file generated from MAGs.

Reads were preprocessed using VEBA's preproces.py, which is a wrapper around Fastq Preprocessor. In the backed, reads were preprocessed using FastP v0.23.4[123] to quality trim and remove adapters. Quality trimmed reads were to the human genome (GRCh37.p4) via Bowtie2 v2.5.2 and removed aligned reads for assembly and mapping. Reads were quantified using SeqKit v2.3.1[124].

Statistical Analyses of Gut Microbiome Data

Differential (relative) abundance metrics fail to address compositionality. Extensive tool benchmarking has demonstrated the susceptibility of commonly employed differential (relative) abundance tools to sparsity, leading to unacceptably high rates of false positive identifications. It is imperative to consider high-throughput sequencing-derived microbiome study datasets as compositions throughout the analysis process. To circumvent the challenges associated with relative abundance analysis, we conducted all microbiome analyses using Compositional Data Analysis (CoDA) with ALDEx2. The analysis relied on the ALDEx2 (ANOVA-Like Differential Expression) package version 1.34 in R 4.3.2125 to identify differential abundances in the gut microbiota at a species level. For a rigorous effect size calculation in ALDEx2, we obtained analysis using 1,000 Dirichlet Monte-Carlo Instances (DMC) at 0.75 gamma. The identical parameters were employed to identify differentially abundant microbial genes/orthologs. In comparing rural and urban (WU) settings, merging HG and AP datasets results in the formation of rural datasets, while the combination of WU healthy and WU obese/diabetic datasets yields the WU datasets. However, it is important to note that all four datasets were categorized as separate groups to calculate each group's effect size.

The ALDEx2 tool is a powerful solution for conducting statistical tests on the centered log-ratio (clr) values obtained from a modeled probability distribution of the dataset. This tool effectively addresses the issue of false-positive identification by providing expected values of parametric and non-parametric statistical tests, along with effect-size estimates. It has demonstrated a remarkable ability to minimize this problem to near-zero levels in real and modeled microbiome datasets, proving its robustness even when dealing with dataset subsets[125,126]. Among the various

methods available for analyzing microbial differential abundance, Aldex2 stands out for its exceptional performance in controlling the false discovery rate (FDR)[127].

We utilized the gseKEGG function within the clusterProfile v4.10.0 R package to perform the KEGG pathway gene set enrichment analysis[28]. This involved identifying significantly enriched KEGG pathways associated with the differentially abundant microbes. For the enrichment analysis, we used all KEGG orthologs present in the differentially abundant microbes as the background genes, ensuring a comprehensive and informative pathway enrichment analysis. The KEGG pathway, modules, and BRITE information were obtained through the KEGGREST R package, an R interface to the KEGG database[128]. In the context of metabolomics data analysis, differentially abundant metabolites were identified using the t-test with a false discovery rate (FDR) threshold set at less than 0.05. The fold changes are determined by computing the ratios between the rural and WU populations. All of the analyses in the current study were performed at adjusted p.value < 0.05.

References:

1. Alt, K.W., Al-Ahmad, A., and Woelber, J.P. (2022). Nutrition and Health in Human Evolution-Past to Present. Nutrients *14*. 10.3390/nu14173594.

2. Kopp, W. (2019). How Western Diet And Lifestyle Drive The Pandemic Of Obesity And Civilization Diseases. Diabetes Metab Syndr Obes *12*, 2221-2236. 10.2147/DMSO.S216791.

3. Clatici, V.G., Voicu, C., Voaides, C., Roseanu, A., Icriverzi, M., and Jurcoane, S. (2018). Diseases of Civilization - Cancer, Diabetes, Obesity and Acne - the Implication of Milk, IGF-1 and mTORC1. Maedica (Bucur) *13*, 273-281. 10.26574/maedica.2018.13.4.273.

4. Gurven, M.D., Trumble, B.C., Stieglitz, J., Blackwell, A.D., Michalik, D.E., Finch, C.E., and Kaplan, H.S. (2016). Cardiovascular disease and type 2 diabetes in evolutionary

perspective: a critical role for helminths? Evol Med Public Health *2016*, 338-357. 10.1093/emph/eow028.

5. Janssen, J. (2023). The Impact of Westernization on the Insulin/IGF-I Signaling Pathway and the Metabolic Syndrome: It Is Time for Change. Int J Mol Sci *24*. 10.3390/ijms24054551.

6. Hills, R.D., Jr., Pontefract, B.A., Mishcon, H.R., Black, C.A., Sutton, S.C., and Theberge, C.R. (2019). Gut Microbiome: Profound Implications for Diet and Disease. Nutrients *11*. 10.3390/nu11071613.

7. Shreiner, A.B., Kao, J.Y., and Young, V.B. (2015). The gut microbiome in health and in disease. Curr Opin Gastroenterol *31*, 69-75. 10.1097/MOG.0000000000000139.

8. Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027-1031. 10.1038/nature05414.

9. Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., and Gordon, J.I. (2011). Human nutrition, the gut microbiome and the immune system. Nature *474*, 327-336. 10.1038/nature10213.

10. Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207-214. 10.1038/nature11234.

11. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59-65. 10.1038/nature08821.

12. Mosca, A., Leclerc, M., and Hugot, J.P. (2016). Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? Front Microbiol *7*, 455. 10.3389/fmicb.2016.00455.

13. Quercia, S., Candela, M., Giuliani, C., Turroni, S., Luiselli, D., Rampelli, S., Brigidi, P., Franceschi, C., Bacalini, M.G., Garagnani, P., and Pirazzini, C. (2014). From lifetime to evolution: timescales of human gut microbiota adaptation. Front Microbiol *5*, 587. 10.3389/fmicb.2014.00587.

14. Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., et al. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science *357*, 802-806. 10.1126/science.aan4834.

15. Carter, M.M., Olm, M.R., Merrill, B.D., Dahan, D., Tripathi, S., Spencer, S.P., Yu, F.B., Jain, S., Neff, N., Jha, A.R., et al. (2023). Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. Cell *186*, 3111-3124 e3113. 10.1016/j.cell.2023.05.046.

16. Cho, I., and Blaser, M.J. (2012). The human microbiome: at the interface of health and disease. Nat Rev Genet *13*, 260-270. 10.1038/nrg3182.

17. De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A *107*, 14691-14696. 10.1073/pnas.1005963107.

18. Manzel, A., Muller, D.N., Hafler, D.A., Erdman, S.E., Linker, R.A., and Kleinewietfeld, M. (2014). Role of "Western diet" in inflammatory autoimmune diseases. Curr Allergy Asthma Rep *14*, 404. 10.1007/s11882-013-0404-6.

19. Stringer, C., and Galway-Witham, J. (2017). Palaeoanthropology: On the origin of our species. Nature *546*, 212-214. 10.1038/546212a.

20. Rito, T., Richards, M.B., Fernandes, V., Alshamali, F., Cerny, V., Pereira, L., and Soares, P. (2013). The first modern human dispersals across Africa. PLoS One *8*, e80031. 10.1371/journal.pone.0080031.

21. Chan, E.K.F., Timmermann, A., Baldi, B.F., Moore, A.E., Lyons, R.J., Lee, S.S., Kalsbeek, A.M.F., Petersen, D.C., Rautenbach, H., Fortsch, H.E.A., et al. (2019). Human origins in a southern African palaeo-wetland and first migrations. Nature *575*, 185-189. 10.1038/s41586-019-1714-1.

22. Petersen, D.C., Libiger, O., Tindall, E.A., Hardie, R.A., Hannick, L.I., Glashoff, R.H., Mukerji, M., Indian Genome Variation, C., Fernandez, P., Haacke, W., et al. (2013). Complex patterns of genomic admixture within southern Africa. PLoS Genet *9*, e1003309. 10.1371/journal.pgen.1003309.

23. Koot, S., and van Beek, W. (2017). Ju/'hoansi Lodging in a Namibian Conservancy: CBNRM, Tourism and Increasing Domination. Conservation and Society *15*, 136-146. 10.4103/cs.cs_15_30.

24. Lee, R.B. (2013). The Dobe Ju/'hoansi, 4th Edition (Wadsworth Cengage Learning).

25. Truter, M., Koopman, J.E., Jordaan, K., Tsamkxao, L.O., Cowan, D.A., Underdown, S.J., Ramond, J.B., and Rifkin, R.F. (2024). Documenting the diversity of the Namibian Ju|'hoansi intestinal microbiome. Cell Rep *43*, 113690. 10.1016/j.celrep.2024.113690.

1110 26. Espinoza, J.L., and Dupont, C.L. (2022). VEBA: a modular end-to-end suite for in silico
1111 recovery, clustering, and analysis of prokaryotic, microeukaryotic, and viral genomes
1112 from metagenomes. BMC Bioinformatics *23*, 419. 10.1186/s12859-022-04973-8.
1113 27. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2:
1114 memory friendly classification with the genome taxonomy database. Bioinformatics *38*,
1115 5315-5316. 10.1093/bioinformatics/btac672.
1116 28. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for
1117 comparing biological themes among gene clusters. OMICS *16*, 284-287.
1118 10.1089/omi.2011.0118.
1119 29. Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.A., and Hugenholtz,
1120 P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a
1121 phylogenetically consistent, rank normalized and complete genome-based taxonomy.
1122 Nucleic Acids Res *50*, D785-D794. 10.1093/nar/gkab776.
1123 30. Luyckx, J., and Baudouin, C. (2011). Trehalose: an intriguing disaccharide with potential
1124 for medical application in ophthalmology. Clin Ophthalmol *5*, 577-581.
1125 10.2147/OPTH.S18827.
1126 31. Vanaporn, M., and Titball, R.W. (2020). Trehalose and bacterial virulence. Virulence *11*,
1127 1192-1202. 10.1080/21505594.2020.1809326.
1128 32. Iturriaga, G., Suarez, R., and Nova-Franco, B. (2009). Trehalose metabolism: from
1129 osmoprotection to signaling. Int J Mol Sci *10*, 3793-3810. 10.3390/ijms10093793.
1130 33. Marathe, S.J., Shah, N.N., and Singhal, R.S. (2020). Enzymatic synthesis of fatty acid
1131 esters of trehalose: Process optimization, characterization of the esters and evaluation of
1132 their bioactivities. Bioorg Chem *94*, 103460. 10.1016/j.bioorg.2019.103460.

34. Yoshizane, C., Mizote, A., Yamada, M., Arai, N., Arai, S., Maruta, K., Mitsuzumi, H., Ariyasu, T., Ushio, S., and Fukuda, S. (2017). Glycemic, insulinemic and incretin responses after oral trehalose ingestion in healthy subjects. Nutr J *16*, 9. 10.1186/s12937-017-0233-x.

35. Sun, L., Jia, H., Li, J., Yu, M., Yang, Y., Tian, D., Zhang, H., and Zou, Z. (2019). Cecal Gut Microbiota and Metabolites Might Contribute to the Severity of Acute Myocardial Ischemia by Impacting the Intestinal Permeability, Oxidative Stress, and Energy Metabolism. Front Microbiol *10*, 1745. 10.3389/fmicb.2019.01745.

36. Amjad, S., Nisar, S., Bhat, A.A., Shah, A.R., Frenneaux, M.P., Fakhro, K., Haris, M., Reddy, R., Patay, Z., Baur, J., and Bagga, P. (2021). Role of NAD(+) in regulating cellular and metabolic signaling pathways. Mol Metab *49*, 101195. 10.1016/j.molmet.2021.101195.

37. Selma, M.V., Beltran, D., Luna, M.C., Romo-Vaquero, M., Garcia-Villalba, R., Mira, A., Espin, J.C., and Tomas-Barberan, F.A. (2017). Isolation of Human Intestinal Bacteria Capable of Producing the Bioactive Metabolite Isourolithin A from Ellagic Acid. Front Microbiol *8*, 1521. 10.3389/fmicb.2017.01521.

38. Ryu, D., Mouchiroud, L., Andreux, P.A., Katsyuba, E., Moullan, N., Nicolet-Dit-Felix, A.A., Williams, E.G., Jha, P., Lo Sasso, G., Huzard, D., et al. (2016). Urolithin A induces mitophagy and prolongs lifespan in C. elegans and increases muscle function in rodents. Nat Med *22*, 879-888. 10.1038/nm.4132.

39. Wehrmann, A., Phillipp, B., Sahm, H., and Eggeling, L. (1998). Different modes of diaminopimelate synthesis and their role in cell wall integrity: a study with

Corynebacterium glutamicum. J Bacteriol *180*, 3159-3165. 10.1128/JB.180.12.3159-3165.1998.

40. Tang, B.K., Kadar, D., and Kalow, W. (1987). An alternative test for acetylator phenotyping with caffeine. Clin Pharmacol Ther *42*, 509-513. 10.1038/clpt.1987.189.

41. Huestis, M.A. (2005). Pharmacokinetics and metabolism of the plant cannabinoids, delta9-tetrahydrocannabinol, cannabidiol and cannabinol. Handb Exp Pharmacol, 657-690. 10.1007/3-540-26573-2_23.

42. Mayo, B., Vazquez, L., and Florez, A.B. (2019). Equol: A Bacterial Metabolite from The Daidzein Isoflavone and Its Presumed Beneficial Health Effects. Nutrients *11*. 10.3390/nu11092231.

43. Tehami, W., Nani, A., Khan, N.A., and Hichami, A. (2023). New Insights Into the Anticancer Effects of p-Coumaric Acid: Focus on Colorectal Cancer. Dose Response *21*, 15593258221150704. 10.1177/15593258221150704.

44. Pei, K., Ou, J., Huang, J., and Ou, S. (2016). p-Coumaric acid and its conjugates: dietary sources, pharmacokinetic properties and biological activities. J Sci Food Agric *96*, 2952-2962. 10.1002/jsfa.7578.

45. Lee, J.S., Bok, S.H., Park, Y.B., Lee, M.K., and Choi, M.S. (2003). 4-hydroxycinnamate lowers plasma and hepatic lipids without changing antioxidant enzyme activities. Ann Nutr Metab *47*, 144-151. 10.1159/000070037.

46. Choi, S.W., Lee, S.K., Kim, E.O., Oh, J.H., Yoon, K.S., Parris, N., Hicks, K.B., and Moreau, R.A. (2007). Antioxidant and antimelanogenic activities of polyamine conjugates from corn bran and related hydroxycinnamic acids. J Agric Food Chem *55*, 3920-3925. 10.1021/jf0635154.

1178 47. Kim, E.O., Min, K.J., Kwon, T.K., Um, B.H., Moreau, R.A., and Choi, S.W. (2012). Anti-
1179 inflammatory activity of hydroxycinnamic acid derivatives isolated from corn bran in
1180 lipopolysaccharide-stimulated Raw 264.7 macrophages. Food Chem Toxicol *50*, 1309-
1181 1316. 10.1016/j.fct.2012.02.011.
1182 48. Mellon, J.E., and Moreau, R.A. (2004). Inhibition of aflatoxin biosynthesis in Aspergillus
1183 flavus by diferuloylputrescine and p-coumaroylferuloylputrescine. J Agric Food Chem
1184 *52*, 6660-6663. 10.1021/jf040226b.
1185 49. Nishino, K., and Yamaguchi, A. (2004). Role of histone-like protein H-NS in multidrug
1186 resistance of Escherichia coli. J Bacteriol *186*, 1423-1429. 10.1128/JB.186.5.1423-
1187 1429.2004.
1188 50. Nishino, K., and Yamaguchi, A. (2001). Analysis of a complete library of putative drug
1189 transporter genes in Escherichia coli. J Bacteriol *183*, 5803-5812.
1190 10.1128/JB.183.20.5803-5812.2001.
1191 51. Nishino, K., and Yamaguchi, A. (2002). EvgA of the two-component signal transduction
1192 system modulates production of the yhiUV multidrug transporter in Escherichia coli. J
1193 Bacteriol *184*, 2319-2323. 10.1128/JB.184.8.2319-2323.2002.
1194 52. Nishino, K., Senda, Y., and Yamaguchi, A. (2008). CRP regulator modulates multidrug
1195 resistance of Escherichia coli by repressing the mdtEF multidrug efflux genes. J Antibiot
1196 (Tokyo) *61*, 120-127. 10.1038/ja.2008.120.
1197 53. Li, X.Z., Nikaido, H., and Poole, K. (1995). Role of mexA-mexB-oprM in antibiotic
1198 efflux in Pseudomonas aeruginosa. Antimicrob Agents Chemother *39*, 1948-1953.
1199 10.1128/AAC.39.9.1948.

1200 54. Li, X.Z., Plesiat, P., and Nikaido, H. (2015). The challenge of efflux-mediated antibiotic
1201 resistance in Gram-negative bacteria. Clin Microbiol Rev *28*, 337-418.
1202 10.1128/CMR.00117-14.
1203 55. Zheng, J., Cui, S., and Meng, J. (2009). Effect of transcriptional activators RamA and
1204 SoxS on expression of multidrug efflux pumps AcrAB and AcrEF in fluoroquinolone-
1205 resistant Salmonella Typhimurium. J Antimicrob Chemother *63*, 95-102.
1206 10.1093/jac/dkn448.
1207 56. Nishino, K., Yamada, J., Hirakawa, H., Hirata, T., and Yamaguchi, A. (2003). Roles of
1208 TolC-dependent multidrug transporters of Escherichia coli in resistance to beta-lactams.
1209 Antimicrob Agents Chemother *47*, 3030-3033. 10.1128/AAC.47.9.3030-3033.2003.
1210 57. Chetri, S., Dolley, A., Bhowmik, D., Chanda, D.D., Chakravarty, A., and Bhattacharjee,
1211 A. (2018). Transcriptional response of AcrEF-TolC against fluoroquinolone and
1212 carbapenem in Escherichia coli of clinical origin. Indian J Med Microbiol *36*, 537-540.
1213 10.4103/ijmm.IJMM_18_308.
1214 58. Lau, S.Y., and Zgurskaya, H.I. (2005). Cell division defects in Escherichia coli deficient
1215 in the multidrug efflux transporter AcrEF-TolC. J Bacteriol *187*, 7815-7825.
1216 10.1128/JB.187.22.7815-7825.2005.
1217 59. Brewer, N.S., and Hellinger, W.C. (1991). The monobactams. Mayo Clin Proc *66*, 1152-
1218 1157. 10.1016/s0025-6196(12)65797-8.
1219 60. Khan, S., and Chousalkar, K.K. (2021). Functional enrichment of gut microbiome by
1220 early supplementation of Bacillus based probiotic in cage free hens: a field study. Anim
1221 Microbiome *3*, 50. 10.1186/s42523-021-00112-5.

1222 61. Mai, J., and Ott, S. (2019). The Fascinating World of Phosphanylphosphonates: From
1223 Acetylenic Phosphaalkenes to Reductive Aldehyde Couplings. Synlett *30*, 1867-1885.
1224 10.1055/s-0039-1690129.
1225 62. Yamamoto, K., Asakawa, H., Tokunaga, K., Watanabe, H., Matsuo, N., Tokimitsu, I., and
1226 Yagi, N. (2001). Long-term ingestion of dietary diacylglycerol lowers serum
1227 triacylglycerol in type II diabetic patients with hypertriglyceridemia. J Nutr *131*, 3204-
1228 3207. 10.1093/jn/131.12.3204.
1229 63. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Some
1230 Industrial Chemicals. Lyon (FR): International Agency for Research on Cancer; 2000.
1231 (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 77.)
1232 Triethanolamine.
1233 64. Lessmann, H., Uter, W., Schnuch, A., and Geier, J. (2009). Skin sensitizing properties of
1234 the ethanolamines mono-, di-, and triethanolamine. Data analysis of a multicentre
1235 surveillance network (IVDK) and review of the literature. Contact Dermatitis *60*, 243-
1236 255. 10.1111/j.1600-0536.2009.01506.x.
1237 65. van Putten, R.J., van der Waal, J.C., de Jong, E., Rasrendra, C.B., Heeres, H.J., and de
1238 Vries, J.G. (2013). Hydroxymethylfurfural, a versatile platform chemical made from
1239 renewable resources. Chem Rev *113*, 1499-1597. 10.1021/cr300182k.
1240 66. Ruiz, H.H., Ramasamy, R., and Schmidt, A.M. (2020). Advanced Glycation End
1241 Products: Building on the Concept of the "Common Soil" in Metabolic Disease.
1242 Endocrinology *161*. 10.1210/endocr/bqz006.
1243 67. Gureev, A.P., Shaforostova, E.A., Starkov, A.A., and Popov, V.N. (2018). beta-
1244 Guanidinopropionic Acid Stimulates Brain Mitochondria Biogenesis and Alters Cognitive

Behavior in Nondiseased Mid-Age Mice. J Exp Neurosci *12*, 1179069518766524. 10.1177/1179069518766524.

68. Li, K., Mikola, M.R., Draths, K.M., Worden, R.M., and Frost, J.W. (1999). Fed-batch fermentor synthesis of 3-dehydroshikimic acid using recombinant Escherichia coli. Biotechnol Bioeng *64*, 61-73.

69. Lindner, H.A., Nadeau, G., Matte, A., Michel, G., Menard, R., and Cygler, M. (2005). Site-directed mutagenesis of the active site region in the quinate/shikimate 5-dehydrogenase YdiB of Escherichia coli. J Biol Chem *280*, 7162-7169. 10.1074/jbc.M412028200.

70. Suzuki, M., Tsukamoto, T., Inoue, H., Watanabe, S., Matsuhashi, S., Takahashi, M., Nakanishi, H., Mori, S., and Nishizawa, N.K. (2008). Deoxymugineic acid increases Zn translocation in Zn-deficient rice plants. Plant Mol Biol *66*, 609-617. 10.1007/s11103-008-9292-x.

71. Noor, E.T., Das, R., Lami, M.S., Chakraborty, A.J., Mitra, S., Tallei, T.E., Idroes, R., Mohamed, A.A., Hossain, M.J., Dhama, K., et al. (2022). Ginkgo biloba: A Treasure of Functional Phytochemicals with Multimedicinal Applications. Evid Based Complement Alternat Med *2022*, 8288818. 10.1155/2022/8288818.

72. Koch, E., Jaggy, H., and Chatterjee, S.S. (2000). Evidence for immunotoxic effects of crude Ginkgo biloba L. leaf extracts using the popliteal lymph node assay in the mouse. Int J Immunopharmacol *22*, 229-236. 10.1016/s0192-0561(99)00080-6.

73. Monien, B.H., Engst, W., Barknowitz, G., Seidel, A., and Glatt, H. (2012). Mutagenicity of 5-hydroxymethylfurfural in V79 cells expressing human SULT1A1: identification and

mass spectrometric quantification of DNA adducts formed. Chem Res Toxicol *25*, 1484-1492. 10.1021/tx300150n.

74. Verhaeghe, B.J., Van Bocxlaer, J.F., and De Leenheer, A.P. (1992). Gas chromatographic/mass spectrometric identification of 3-hydroxydicarboxylic acids in urine. Biol Mass Spectrom *21*, 27-32. 10.1002/bms.1200210107.

75. Matsubara, K., Kohling, R., Schonenberger, B., Kouril, T., Esser, D., Brasen, C., Siebers, B., and Wohlgemuth, R. (2014). One-step synthesis of 2-keto-3-deoxy-d-gluconate by biocatalytic dehydration of d-gluconate. J Biotechnol *191*, 69-77. 10.1016/j.jbiotec.2014.06.005.

76. Gallage, N.J., and Møller, B.L. (2018). Vanilla: The Most Popular Flavour. In Biotechnology of Natural Products, W. Schwab, B.M. Lange, and M. Wüst, eds. (Springer International Publishing), pp. 3-24. 10.1007/978-3-319-67903-7_1.

77. Tserng, K.Y., and Klein, P.D. (1979). Bile acid sulfates. III. Synthesis of 7- and 12-monosulfates of bile acids and their conjugates using a sulfur trioxide-triethylamine complex. Steroids *33*, 167-182. 10.1016/0039-128x(79)90024-2.

78. Greter, J., Lindstedt, S., Seeman, H., and Steen, G. (1980). 3-hydroxydecanedioic acid and related homologues: urinary metabolites in ketoacidosis. Clin Chem *26*, 261-265.

79. Tserng, K.Y., Griffin, R.L., and Kerr, D.S. (1996). Distinction of dicarboxylic aciduria due to medium-chain triglyceride feeding from that due to abnormal fatty acid oxidation and fasting in children. Metabolism *45*, 162-167. 10.1016/s0026-0495(96)90047-5.

80. Pettersen, J.E., Jellum, E., and Eldjarn, L. (1972). The occurrence of adipic and suberic acid in urine from ketotic patients. Clin Chim Acta *38*, 17-24. 10.1016/0009-8981(72)90202-1.

81. Zwicker, H., and Rittmaster, R.S. (1993). Androsterone sulfate: physiology and clinical significance in hirsute women. J Clin Endocrinol Metab *76*, 112-116. 10.1210/jcem.76.1.8380602.

82. Harteneck, C. (2013). Pregnenolone sulfate: from steroid metabolite to TRP channel ligand. Molecules *18*, 12012-12028. 10.3390/molecules181012012.

83. Laatikainen, T., and Vihko, R. (1970). Identification of C19O2 and C21O2 steroids in the mono- and disulphate fractions of human faeces. Eur J Biochem *13*, 534-538. 10.1111/j.1432-1033.1970.tb00957.x.

84. Cabral, C.E., and Klein, M. (2017). Phytosterols in the Treatment of Hypercholesterolemia and Prevention of Cardiovascular Diseases. Arq Bras Cardiol *109*, 475-482. 10.5935/abc.20170158.

85. Delgado-Andrade, C. (2016). Carboxymethyl-lysine: thirty years of investigation in the field of AGE formation. Food Funct *7*, 46-57. 10.1039/c5fo00918a.

86. Thiele, I., Swainston, N., Fleming, R.M., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013). A community-driven global reconstruction of human metabolism. Nat Biotechnol *31*, 419-425. 10.1038/nbt.2488.

87. Rice, T., Zannini, E., E, K.A., and Coffey, A. (2020). A review of polyols - biotechnological production, food applications, regulation, labeling and health effects. Crit Rev Food Sci Nutr *60*, 2034-2051. 10.1080/10408398.2019.1625859.

88. Garcia-Villalba, R., Gimenez-Bastida, J.A., Cortes-Martin, A., Avila-Galvez, M.A., Tomas-Barberan, F.A., Selma, M.V., Espin, J.C., and Gonzalez-Sarrias, A. (2022).

Urolithins: a Comprehensive Update on their Metabolism, Bioactivity, and Associated Gut Microbiota. Mol Nutr Food Res *66*, e2101019. 10.1002/mnfr.202101019.

89. Beltran, D., Romo-Vaquero, M., Espin, J.C., Tomas-Barberan, F.A., and Selma, M.V. (2018). Ellagibacter isourolithinifaciens gen. nov., sp. nov., a new member of the family Eggerthellaceae, isolated from human gut. Int J Syst Evol Microbiol *68*, 1707-1712. 10.1099/ijsem.0.002735.

90. Imai, T., Tanaka, K., Yonemitsu, T., Yakushiji, Y., and Ohura, K. (2017). Elucidation of the Intestinal Absorption of para-Aminobenzoic Acid, a Marker for Dietary Intake. J Pharm Sci *106*, 2881-2888. 10.1016/j.xphs.2017.04.070.

91. Browne, A.J., Chipeta, M.G., Haines-Woodhouse, G., Kumaran, E.P.A., Hamadani, B.H.K., Zaraa, S., Henry, N.J., Deshpande, A., Reiner, R.C., Jr., Day, N.P.J., et al. (2021). Global antibiotic consumption and usage in humans, 2000-18: a spatial modelling study. Lancet Planet Health *5*, e893-e904. 10.1016/S2542-5196(21)00280-1.

92. L, M.I.-A., Chavarro, J.E., Mendiola, J., Roca, M., Tanrikut, C., Vioque, J., Jorgensen, N., and Torres-Cantero, A.M. (2017). Fatty acid intake in relation to reproductive hormones and testicular volume among young healthy men. Asian J Androl *19*, 184-190. 10.4103/1008-682X.190323.

93. Liepa, G.U., Sengupta, A., and Karsies, D. (2008). Polycystic ovary syndrome (PCOS) and other androgen excess-related conditions: can changes in dietary intake make a difference? Nutr Clin Pract *23*, 63-71. 10.1177/011542650802300163.

94. Hannich, J.T., Umebayashi, K., and Riezman, H. (2011). Distribution and functions of sterols and sphingolipids. Cold Spring Harb Perspect Biol *3*. 10.1101/cshperspect.a004762.

95. Komorowski, J., and Stepien, H. (2007). [The role of the endocannabinoid system in the regulation of endocrine function and in the control of energy balance in humans]. Postepy Hig Med Dosw (Online) *61*, 99-105.

96. Jia, M., Zhang, Y., Gao, Y., and Ma, X. (2020). Effects of Medium Chain Fatty Acids on Intestinal Health of Monogastric Animals. Curr Protein Pept Sci *21*, 777-784. 10.2174/1389203721666191231145901.

97. Montoro-Huguet, M.A., Belloc, B., and Dominguez-Cajal, M. (2021). Small and Large Intestine (I): Malabsorption of Nutrients. Nutrients *13*. 10.3390/nu13041254.

98. D'Arrigo, P., and Servi, S. (2010). Synthesis of lysophospholipids. Molecules *15*, 1354-1377. 10.3390/molecules15031354.

99. Nomura, D.K., Long, J.Z., Niessen, S., Hoover, H.S., Ng, S.W., and Cravatt, B.F. (2010). Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. Cell *140*, 49-61. 10.1016/j.cell.2009.11.027.

100. Silva, Y.P., Bernardi, A., and Frozza, R.L. (2020). The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. Front Endocrinol (Lausanne) *11*, 25. 10.3389/fendo.2020.00025.

101. Scheppach, W. (1994). Effects of short chain fatty acids on gut morphology and function. Gut *35*, S35-38. 10.1136/gut.35.1_suppl.s35.

102. Nogal, A., Valdes, A.M., and Menni, C. (2021). The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. Gut Microbes *13*, 1-24. 10.1080/19490976.2021.1897212.

103. Frolova, M.S., Suvorova, I.A., Iablokov, S.N., Petrov, S.N., and Rodionov, D.A. (2022). Genomic reconstruction of short-chain fatty acid production by the human gut microbiota. Front Mol Biosci *9*, 949563. 10.3389/fmolb.2022.949563.

104. Espinoza, J.L., Phillips, A., Prentice, M.B., Tan, G.S., Kamath, P.L., Lloyd, K.G., and Dupont, C.L. (2024). Unveiling the microbial realm with VEBA 2.0: a modular bioinformatics suite for end-to-end genome-resolved prokaryotic, (micro)eukaryotic and viral multi-omics from either short- or long-read sequencing. Nucleic Acids Res *52*, e63. 10.1093/nar/gkae528.

105. Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res *27*, 824-834. 10.1101/gr.213959.116.

106. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P.S.G., Nayfach, S., and Kyrpides, N.C. (2024). Identification of mobile genetic elements with geNomad. Nat Biotechnol *42*, 1303-1312. 10.1038/s41587-023-01953-y.

107. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359. 10.7717/peerj.7359.

108. Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nat Methods *11*, 1144-1146. 10.1038/nmeth.3103.

109. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol *3*, 836-843. 10.1038/s41564-018-0171-1.

110. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res *25*, 1043-1055. 10.1101/gr.186072.114.

111. Manni, M., Berkeley, M.R., Seppey, M., Simao, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol *38*, 4647-4654. 10.1093/molbev/msab199.

112. Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol *35*, 725-731. 10.1038/nbt.3893.

113. https://bitbucket.org/berkeleylab/checkv/issues/38/recommended-cutoffs-for-analyzing-checkv.

114. Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun *9*, 5114. 10.1038/s41467-018-07641-9.

115. Steinegger, M., and Soding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol *35*, 1026-1028. 10.1038/nbt.3988.

116. Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119. 10.1186/1471-2105-11-119.

117. Levy Karin, E., Mirdita, M., and Soding, J. (2020). MetaEuk-sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. Microbiome *8*, 48. 10.1186/s40168-020-00808-x.

118. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat Methods *12*, 59-60. 10.1038/nmeth.3176.

119. Mirdita, M., von den Driesch, L., Galiez, C., Martin, M.J., Soding, J., and Steinegger, M. (2017). Uniclust databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Res *45*, D170-D176. 10.1093/nar/gkw1081.

120. Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics *36*, 2251-2252. 10.1093/bioinformatics/btz859.

121. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.

122. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930. 10.1093/bioinformatics/btt656.

123. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884-i890. 10.1093/bioinformatics/bty560.

124. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962. 10.1371/journal.pone.0163962.

125. Thorsen, J., Brejnrod, A., Mortensen, M., Rasmussen, M.A., Stokholm, J., Al-Soud, W.A., Sorensen, S., Bisgaard, H., and Waage, J. (2016). Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. Microbiome *4*, 62. 10.1186/s40168-016-0208-8.

126. Fernandes, A.D., Reid, J.N., Macklaim, J.M., McMurrough, T.A., Edgell, D.R., and Gloor, G.B. (2014). Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome *2*, 15. 10.1186/2049-2618-2-15.

127. Yang, L., and Chen, J. (2022). A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. Microbiome *10*, 130. 10.1186/s40168-022-01320-0.

128. D, T., and B, M. (2024). KEGGREST: Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG). R package version 1.44.1.

Figure 1: Resource for human gut microbiome evolution. The computational workflow used to generate high-quality metagenome-assembled genomes (MAGs) from Illumina metagenomic (MetaG) and metatranscriptomics (MetaT) sequencing. Blue MAGs represent classified organisms, while unclassified ones appear in orange using GTDB-Tk.

Figure 2: Gut microbiome diversity and abundance profile of Ju/'Hoansi peoples. A) The bacterial SLC relative abundance as circular representations of taxonomic and phylogenetic trees. Each color represents a different phylum, and the circle size represents the abundance of the respective taxonomic level. **B**) Relative abundance distribution across different taxonomic levels of Archaea SLC. **C**) The visualization illustrates the frequency (inner circle) and relative abundance of different viral/phage SLC (outer circle).

Figure 3: Microbiome comparison of Ju/'hoansi, Agropastoral, Urban Healthy and Urban obese/diabetic population. A) Volcano plot showing the differentially abundant species between rural and WU using ALDEx2. **B)** Multiple heat trees showing significant differentially abundant kingdom, phylum, class, order, family, genus, and species in the Ju/'hoansi population (Benjamini-Hochberg corrected p-value using Wilcoxon test < 0.05) with labels up to genus are displayed. The AP Bantu population is displayed in **(B.1)**, WU healthy in **(B.2)**, and WU obese/diabetic in **(B.3). C)** The Alpha Diversity plot uses the Shannon diversity index to compare the diversity of four cohorts. **D)** PCA plot generated using isometric log-ratio data based Euclidean distance plot between the four cohorts. **E)** The phylum-level relative abundance composition bar plot shows the distribution of various phylum in four cohorts. **F)** The topmost species with decreasing abundance from Ju/'hoansi to AP Bantu to WU cohort. **G)** The topmost species with increasing abundance from Ju/'hoansi to AP Bantu to WU cohort.

Figure 4: Metatranscriptomics comparison overview: A) NES (normalized enrichment score) for the top KEGG orthologs with corresponding effect size across four cohorts plotted as a bar plot. **B)** NES score for KEGG BRITE hierarchies as x-axis and adjusted p. Value as color code. **C)** Super pathway box plot displaying all metabolites' log 2-fold change (> 1.0). **D)** The Pfam domain with the highest NES scores and the effect size of four cohorts is displayed as a line graph.

Figure 5. Microbial resistance, activity, and food metabolism. A) Antibiotic resistance and prokaryotic defense system pathways. The four-bar color ratio represents the proportion of the total effect size, and the orange line denotes the NES score for the respective pathway. **B)** Various microbial activity pathways, each bar representing the effect size and the orange line denoting the NES score for the respective pathway. **C)** Nucleotide metabolites log 2-fold changes with negative values representing the log-fold change in the WU population. **D)** Microbial genes and **E)** metabolites involved in the cofactor and vitamin pathway based on metatranscriptomics and metabolomics analysis. **F)** The abundance of the Xenobiotic pathway and **G)** metabolites in the WU and rural populations. **G)** The Food Metabolites heatmap shows the abundance of significant food metabolites in four cohorts.

Figure 6. A) Box plot illustrating the log 2-fold changes for all amino acid metabolism pathway metabolites. **B)** Amino acid metabolism pathway metabolites with log 2-fold change greater than 2. **C)** Enriched amino acid biosynthesis and degradation KEGG pathway with respective effect size and NES score across four groups. **D)** Box plot illustrating the log 2-fold changes of all metabolites within the lipid subpathway. **E)** Bar plot illustrating the enrichment of microbial genes in various pathways related to fatty acid biosynthesis and metabolism with respective effect size and NES score. **F)** Microbes with microbial genes with a positive NES score from section C are presented in a plot alongside their corresponding effect sizes across four groups, arranged in sorted order from HG to WU. **I)** Microorganisms carrying microbial genes with a negative NES score from section C are depicted in a chart, displaying their respective effect sizes across four groups, organized in ascending order from HG to WU.

Figure 7. Energy Production Pathway and Blood Biomarker Association: A) Multiple energy production pathways highly enriched in the Ju/'hoansi population. **B)** Energy generation pathway enriched in the urban population. **C)** Carbohydrate metabolites and **D)** Energy pathway metabolites abundant in rural and WU populations. **E)** The line plot depicts the effect size of microorganisms carrying microbial genes from section A across four cohorts, sorted based on their abundance in the Ju/'hoansi population. **F)** The chart portrays microorganisms carrying microbial genes falling under section B, showcasing their varying effect sizes within four

cohorts. **G)** Microbes associated with lipid profile and **H)** Microbes associated with blood inflammation markers including TNF Alpha and C-Reactive Protein, and **I)** Microbes associated with HbA1c and effect size represent which group of the respective microbe is significantly more abundant in rural or WU. **J)** Box plot of respective microbes' effect sizes for four cohorts. **E)** Absolute abundance of eight short-chain fatty acids in the four cohorts. The Y-axis represents abundance in terms of µg/g.

55 HG Ju/'hoansi, 16 AP Bantu
90 WU Healthy, 58 WU Obese/Diabetic

Illumina 2x150 bp
NovaSeq 6000 S2

NovaSeq 6000 S4 flow

11.4 billion Read pairs (MetaG)

31.0 billion Read pairs (MetaT)

Read QC

BBtools

Fastp

Human DNA/RNA Removed

10.5 billion Read pairs (MetaG)

30.4 billion Read pairs (MetaT)

Assembly (metaSPAdes)

219 Assemblies
96.6 Gbp of assembled data

Reads Maped

Prokaryotic Binning
Viral Binning
Eukaryotic Binning

MetaBAT2, CONCOCT, MAXBIN 2
CheckM, VirFinder
CheckV

9,010 bacterial genomes
127 archaeal genomes
1025 bacteriophages genomes
2 Eukaryotic genomes

FastANI

Cluster

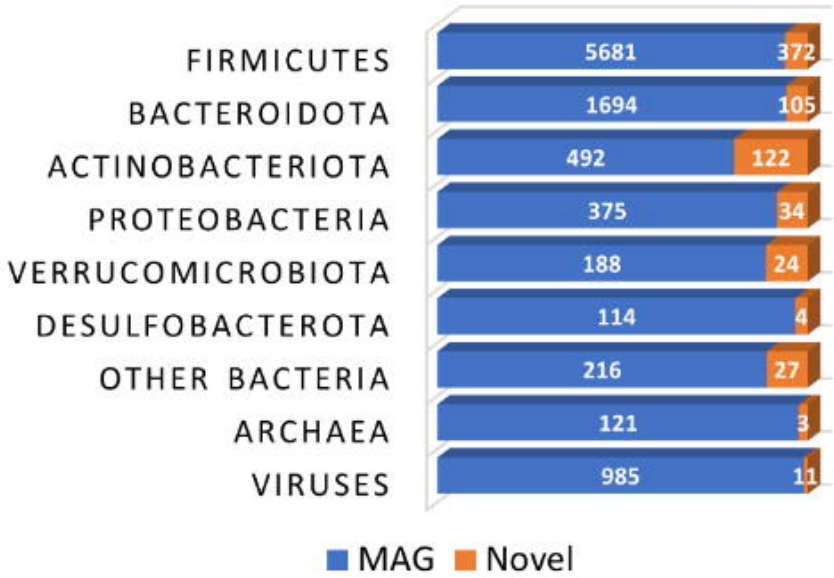
1184 bacterial SLC
7 archaeal SLC
650 bacteriophages SLC
2 Eukaryotic SLC

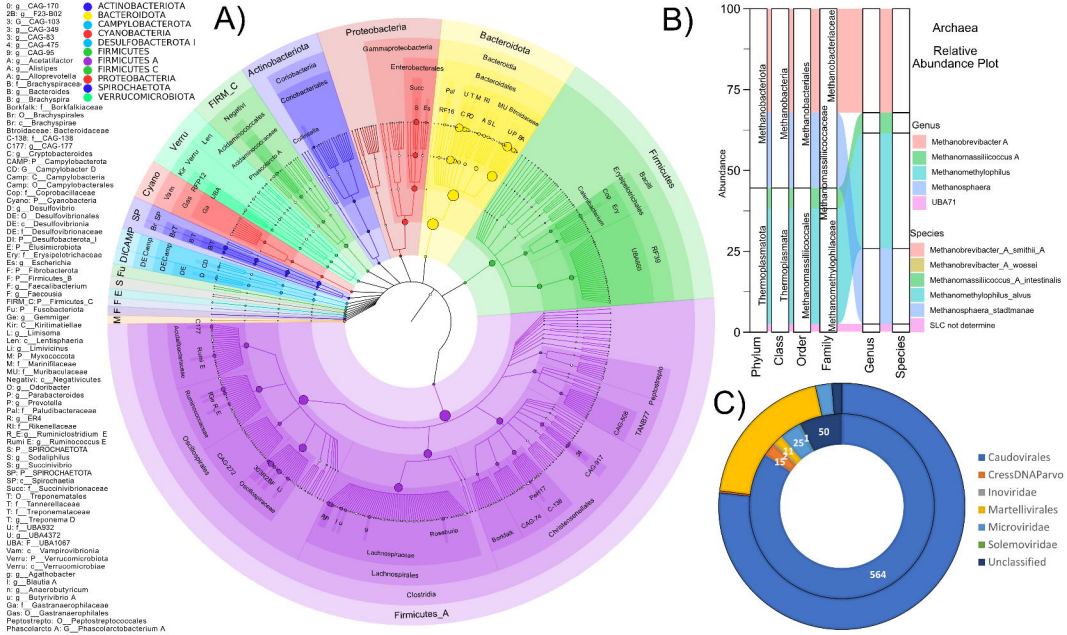
915 Global Metabolites
8 Short Chain Fatty Acids

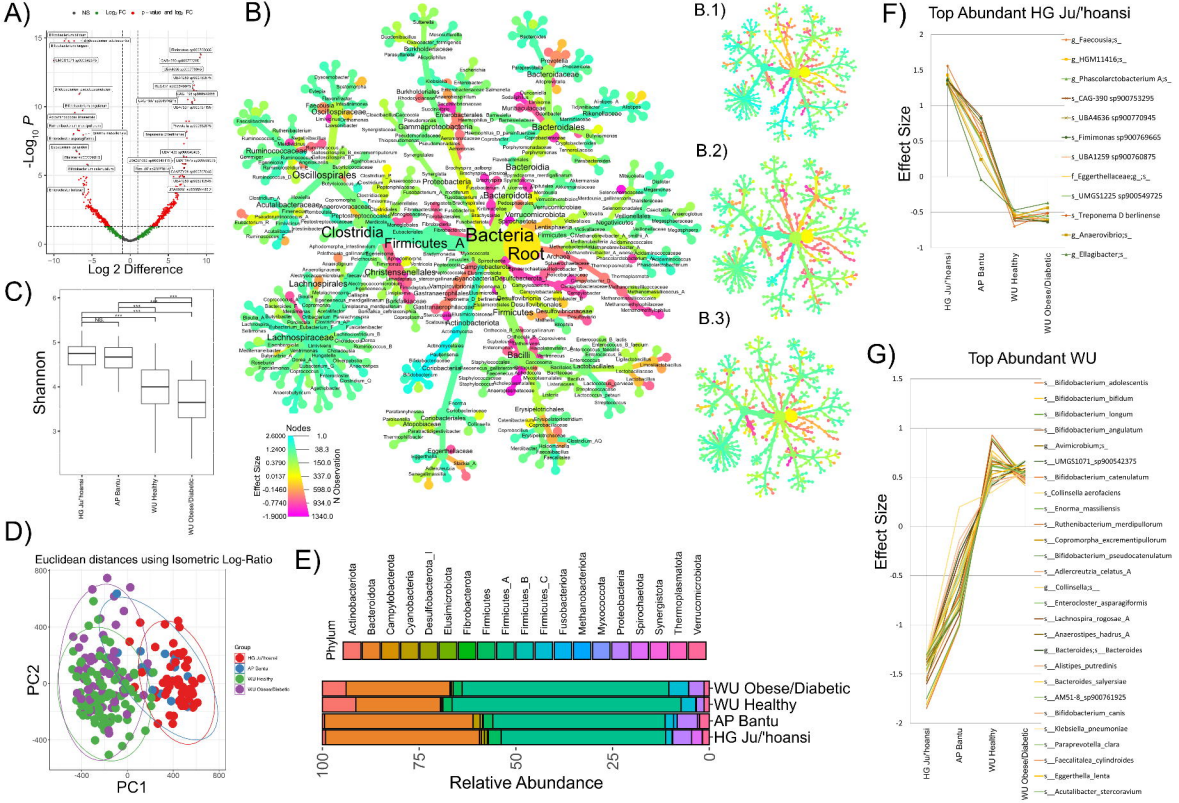
Transcriptomics Analysis

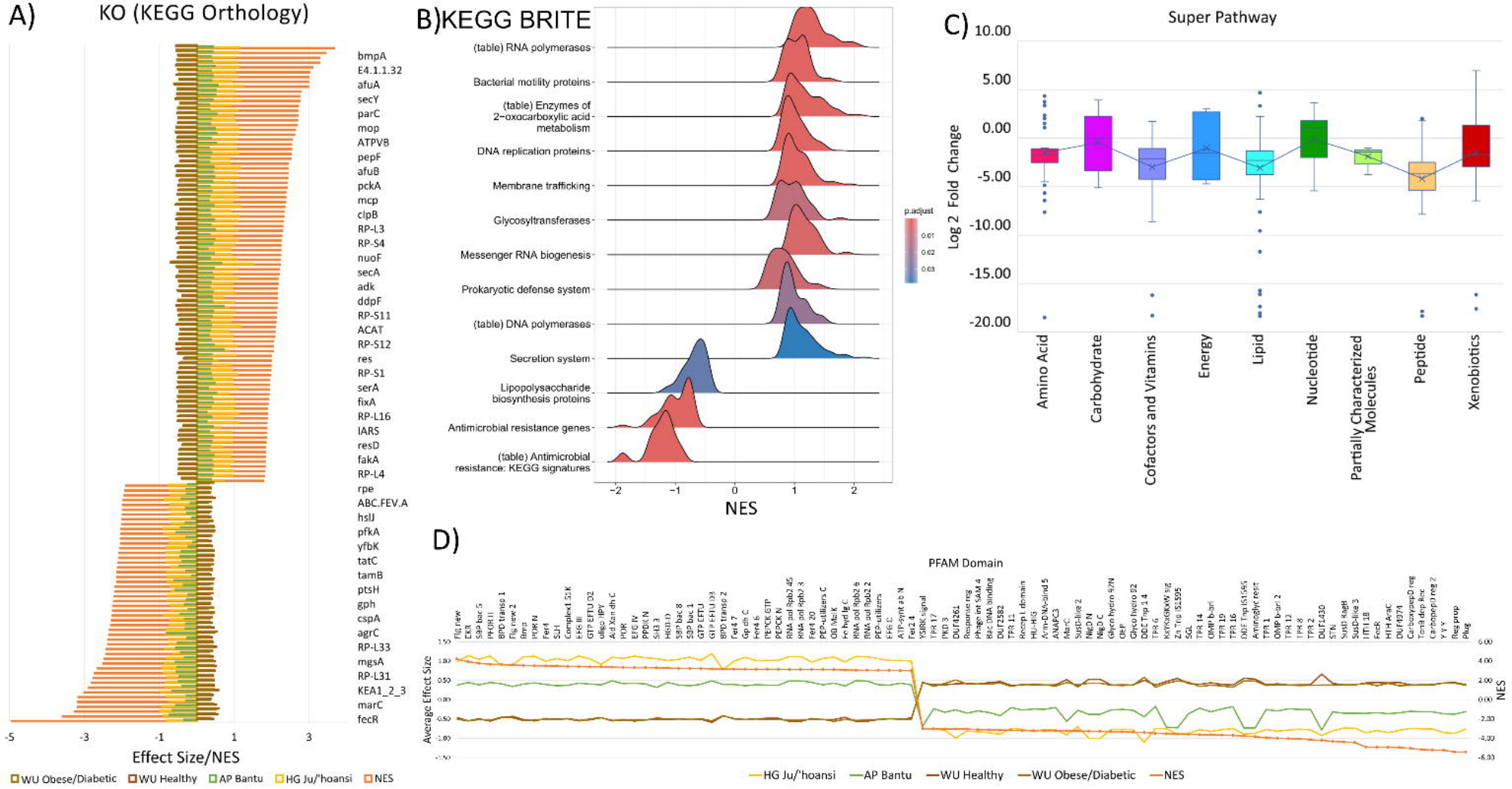
Taxonomy Analysis

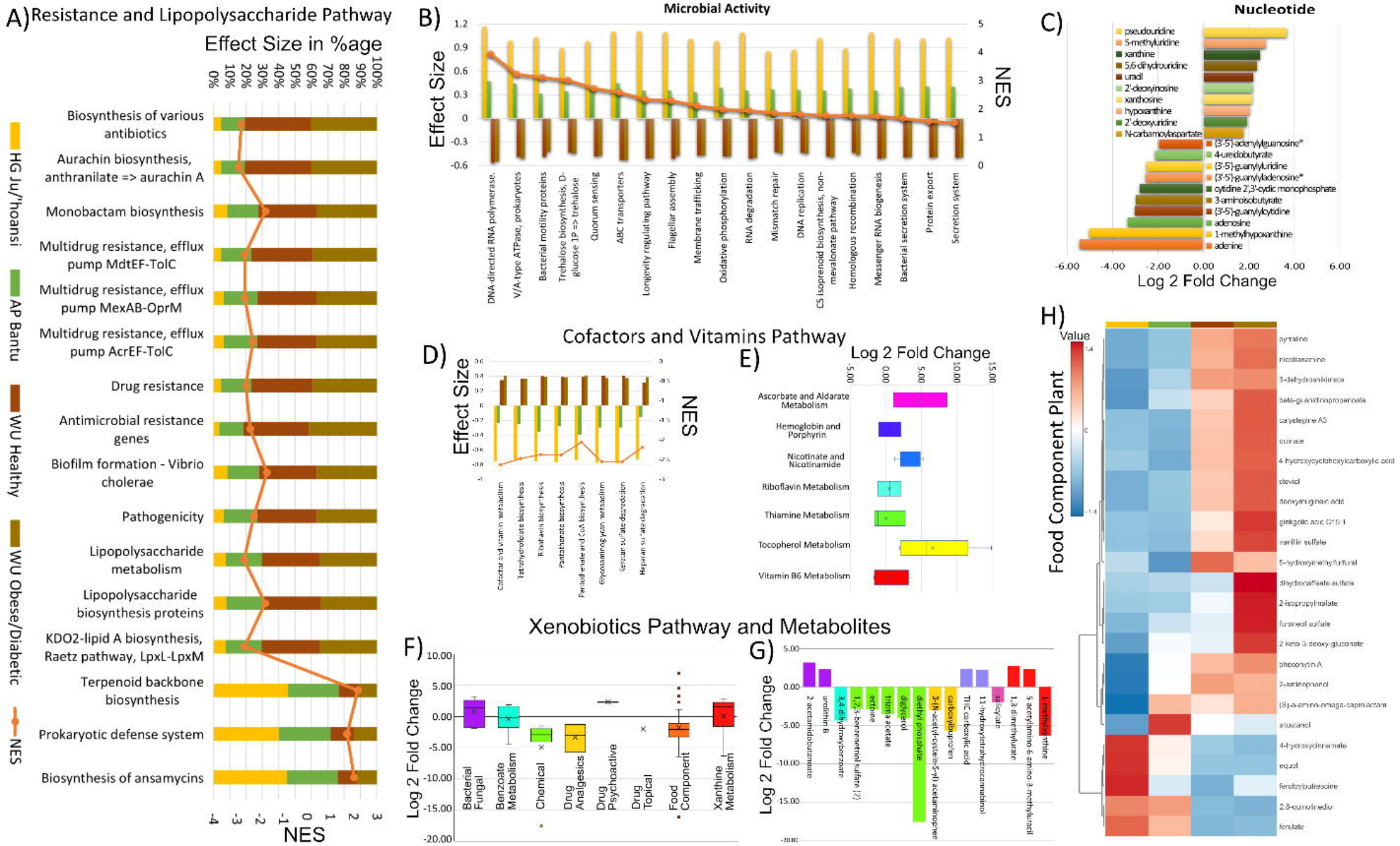
No. of Orthologs 4,096,206
No. of ORFs 21,339,640

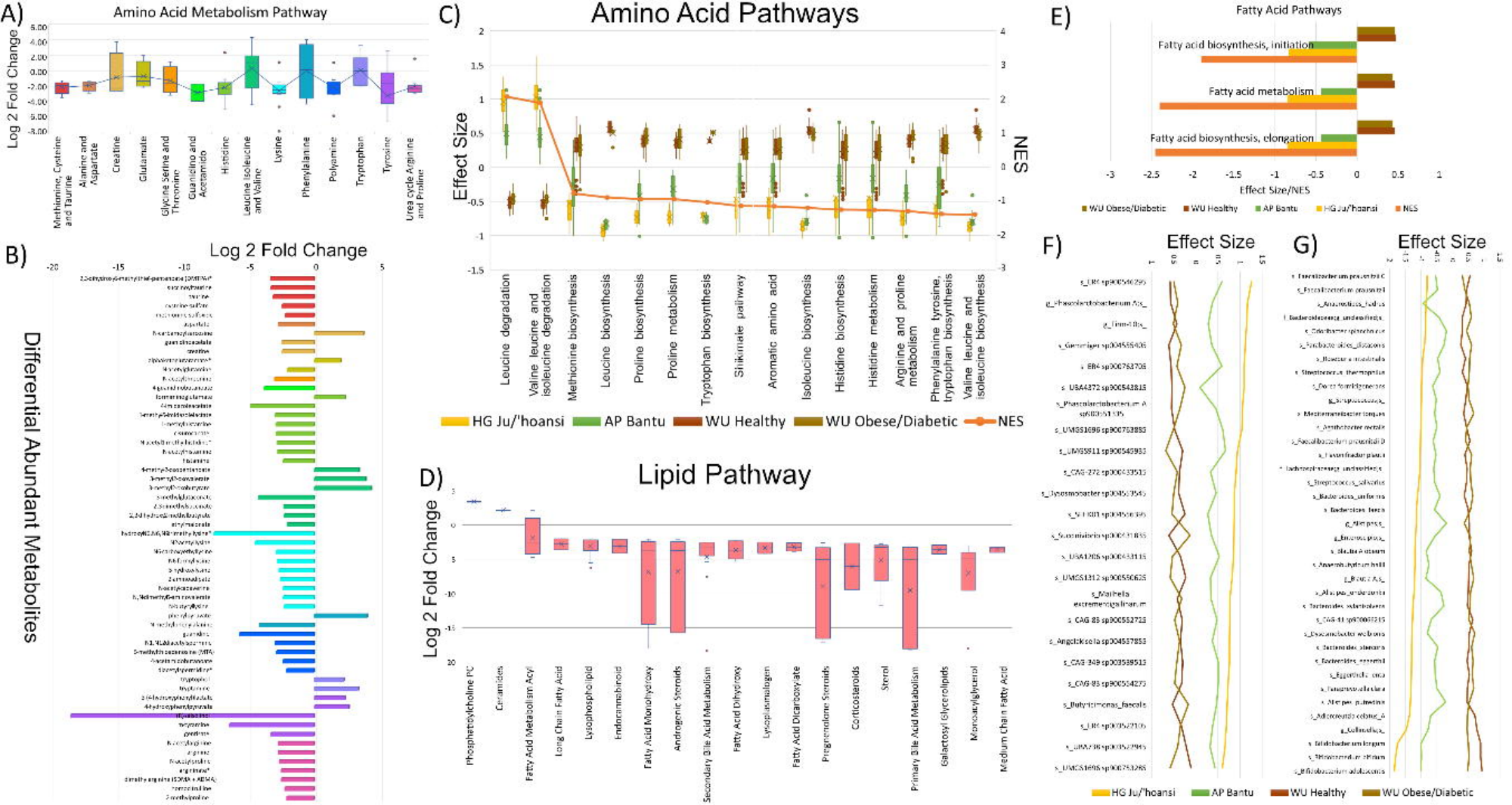


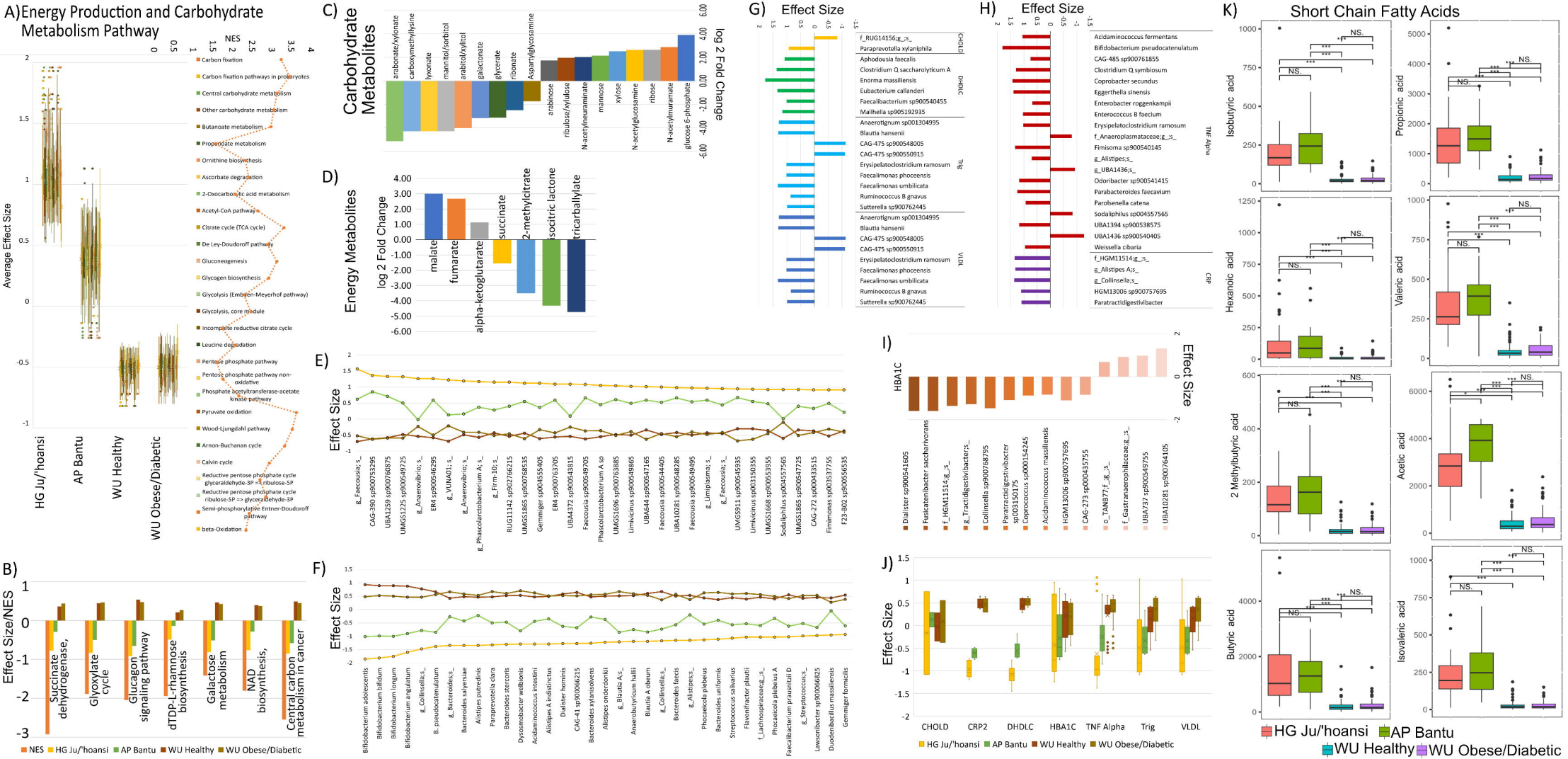

A)
B)
C)
ACTINOBACTERIOTA
BACTEROIDOTA
CAMPYLOBACTEROTA
CYANOBACTERIA
DESULFOBACTEROTA I
FIRMICUTES
FIRMICUTES A
FIRMICUTES C
PROTEOBACTERIA
SPIROCHAETOTA
VERRUCOMICROBIOTA
Archaea
Relative
Abundance Plot
Genus
Methanobrevibacter A
Methanomassiliicoccus A
Methanomethylophilus
Methanosphaera
UBA71
Species
Methanobrevibacter_A_smithii_A
Methanobrevibacter_A_woesei
Methanomassiliicoccus_A_intestinalis
Methanomethylophilus_alvus
Methanosphaera_stadtmanae
SLC not determine
Abundance
Phylum
Class
Order
Family
Genus
Species
Caudovirales
CressDNAParvo
Inoviridae
Martellivirales
Microviridae
Solemoviridae
Unclassified
564
50
25
21
19


A)
NS
Log2 FC
p-value and log2 FC
-Log10 P
Log 2 Difference
B)
B.1)
B.2)
B.3)
Root
Bacteria
Firmicutes_A
Clostridia
Bacteroidia
Bacteroidales
Bacteroidaceae
Oscillospiraceae
Ruminococcaceae
Oscillospirales
Christensenellales
Lachnospirales
Lachnospiraceae
Acidobacteriota
Actinobacteriota
Verrucomicrobiota
Nodes
Effect Size
No observation
C)
Shannon
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic
D)
Euclidean distances using Isometric Log-Ratio
PC1
PC2
Group
E)
Phylum
Actinobacteriota
Bacteroidota
Camysfobacterota
Cyanobacteria
Desulfobacteriota_I
Elusimicrobiota
Fibrobacterota
Firmicutes
Firmicutes_A
Firmicutes_B
Firmicutes_C
Fusobacteriota
Methanobacteriota
Myxococcota
Proteobacteria
Spirochaetota
Synergistota
Thermoplasmatota
Verrucomicrobiota
Relative Abundance
F)
Top Abundant HG Ju/'hoansi
Effect Size
g_Faecousia;s_
g_HGM11416;s_
g_Phascolarctobacterium A;s_
s_CAG-390 sp900753295
s_UBA4636 sp900770945
s_Fimimonas sp900769665
s_UBA1259 sp900760875
f_Eggerthellaceae;g_;s_
s_UMGS1225 sp900549725
s_Treponema D berlinense
g_Anaerovibrio;s_
g_Ellagibacter;s_
G)
Top Abundant WU
s__Bifidobacterium_adolescentis
s__Bifidobacterium_bifidum
s__Bifidobacterium_longum
s__Bifidobacterium_angulatum
g__Avimicrobium;s_
s__UMGS1071_sp900542375
s__Bifidobacterium_catenulatum
s_Collinsella aerofaciens
s_Enorma_massiliensis
s__Ruthenibacterium_merdipullorum
s__Copromorpha_excrementipullorum
s__Bifidobacterium_pseudocatenulatum
s__Adlercreutzia_celatus_A
g__Collinsella;s__
s_Enterocloster_asparagiformis
s_Lachnospira_rogosae_A
s__Anaerostipes_hadrus_A
g__Bacteroides;s__Bacteroides
s_Alistipes_putredinis
s_Bacteroides_salyersiae
s_AM51-8_sp900761925
s__Bifidobacterium_canis
s_Klebsiella_pneumoniae
s_Paraprevotella_clara
s__Faecalitalea_cylindroides
s__Eggerthella_lenta
s__Acutalibacter_stercoravium


A)
KO (KEGG Orthology)
bmpA
E4.1.1.32
afuA
secY
parC
mop
ATPVB
pepF
afuB
pckA
mcp
clpB
RP-L3
RP-S4
nuoF
secA
adk
ddpF
RP-S11
ACAT
RP-S12
res
RP-S1
serA
fixA
RP-L16
IARS
resD
fakA
RP-L4
rpe
ABC.FEV.A
hslJ
pfkA
yfbK
tatC
tamB
ptsH
gph
cspA
agrC
RP-L33
rngsA
RP-L31
KEA1_2_3
marC
fecR
-5
-3
-1
1
3
Effect Size/NES
WU Obese/Diabetic
WU Healthy
AP Bantu
HG Ju/'hoansi
NES
B)KEGG BRITE
(table) RNA polymerases
Bacterial motility proteins
(table) Enzymes of 2-oxocarboxylic acid metabolism
DNA replication proteins
Membrane trafficking
Glycosyltransferases
Messenger RNA biogenesis
Prokaryotic defense system
(table) DNA polymerases
Secretion system
Lipopolysaccharide biosynthesis proteins
Antimicrobial resistance genes
(table) Antimicrobial resistance: KEGG signatures
−2
−1
0
1
2
NES
p.adjust
0.01
0.02
0.03
C)
Super Pathway
10.00
5.00
0.00
-5.00
-10.00
-15.00
-20.00
Log 2 Fold Change
Amino Acid
Carbohydrate
Cofactors and Vitamins
Energy
Lipid
Nucleotide
Partially Characterized Molecules
Peptide
Xenobiotics
D)
PFAM Domain
Average Effect Size
NES
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic
NES


A)Resistance and Lipopolysaccharide Pathway
Effect Size in %age
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
Biosynthesis of various antibiotics
Aurachin biosynthesis, anthranilate => aurachin A
Monobactam biosynthesis
Multidrug resistance, efflux pump MdtEF-TolC
Multidrug resistance, efflux pump MexAB-OprM
Multidrug resistance, efflux pump AcrEF-TolC
Drug resistance
Antimicrobial resistance genes
Biofilm formation - Vibrio cholerae
Pathogenicity
Lipopolysaccharide metabolism
Lipopolysaccharide biosynthesis proteins
KDO2-lipid A biosynthesis, Raetz pathway, LpxL-LpxM
Terpenoid backbone biosynthesis
Prokaryotic defense system
Biosynthesis of ansamycins
-4 -3 -2 -1 0 1 2 3
NES
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic
NES
B)
Microbial Activity
Effect Size
NES
DNA directed RNA polymerase.
V/A-type ATPase, prokaryotes
Bacterial motility proteins
Trehalose biosynthesis, D-glucose 1P => trehalose
Quorum sensing
ABC transporters
Longevity regulating pathway
Flagellar assembly
Membrane trafficking
Oxidative phosphorylation
RNA degradation
Mismatch repair
DNA replication
CS isoprenoid biosynthesis, non-mevalonate pathway
Homologous recombination
Messenger RNA biogenesis
Bacterial secretion system
Protein export
Secretion system
C)
Nucleotide
pseudouridine
5-methyluridine
xanthine
5,6-dihydrouridine
uracil
2'-deoxyinosine
xanthosine
hypoxanthine
2'-deoxyuridine
N-carbamoylaspartate
(3'-5')-adenylylguanosine*
4-uredobutyrate
(3'-5')-guanylyluridine
(3'-5')-guanylyladenosine*
cytidine 2',3'-cyclic monophosphate
3-aminoisobutyrate
(3'-5')-guanylylcytidine
adenosine
1-methylhypoxanthine
adenine
-6.00 -4.00 -2.00 0.00 2.00 4.00 6.00
Log 2 Fold Change
Cofactors and Vitamins Pathway
D)
Effect Size
NES
Cofactor and vitamin metabolism
Tetrahydrofolate biosynthesis
Riboflavin biosynthesis
Pantothenate biosynthesis
Pantothenate and CoA biosynthesis
Glycosaminoglycan metabolism
Keratan sulfate degradation
Heparan sulfate degradation
E)
Log 2 Fold Change
-5.00 0.00 5.00 10.00 15.00
Ascorbate and Aldarate Metabolism
Hemoglobin and Porphyrin
Nicotinate and Nicotinamide
Riboflavin Metabolism
Thiamine Metabolism
Tocopherol Metabolism
Vitamin B6 Metabolism
Xenobiotics Pathway and Metabolites
F)
Log 2 Fold Change
10.00 5.00 0.00 -5.00 -10.00 -15.00 -20.00
Bacterial Fungal
Benzoate Metabolism
Chemical
Drug Analgesics
Drug Psychoactive
Drug Topical
Food Component
Xanthine Metabolism
G)
Log 2 Fold Change
5.00 0.00 -5.00 -10.00 -15.00 -20.00
2-acetamidobenzoate
urolithin B
3,4-dihydroxybenzoate
1,3-benzenediol sulfate (2)
ectoine
trans-a-acetate
diphenol
diethyl phosphate
3-(N-acetyl-L-cysteine-S-yl) acetaminophen
carboxyibuprofen
TMC carboxylic acid
11-hydroxytetrahydrocannabinol
salicylate
1,3-dimethylurate
5-acetylamino-6-amino-3-methyluracil
1-methylxanthine
H)
Value
Food Component Plant
pyrraline
nicotianamine
3-dehydroshikimate
beta-guanidinopropanoate
calystegine A3
quinate
4-hydroxycyclohexylcarboxylic acid
steviol
deoxymugineic acid
ginkgolic acid C15:1
vanillin sulfate
5-hydroxymethylfurfural
dihydrocaffeate sulfate
2-isopropylmalate
furaneol sulfate
2-keto-3-deoxy-gluconate
pheophytin A
2-aminopropanol
(S)-a-amino-omega-caprolactam
ectoine
4-hydroxycinnamate
equol
feruloylputrescine
2,8-quinolinediol
ferulate


A)
Amino Acid Metabolism Pathway
Log 2 Fold Change
Methionine, Cysteine and Taurine
Alanine and Aspartate
Creatine
Glutamate
Glycine Serine and Threonine
Guanidino and Acetamido
Histidine
Leucine Isoleucine and Valine
Lysine
Phenylalanine
Polyamine
Tryptophan
Tyrosine
Urea cycle Arginine and Proline
B)
Log 2 Fold Change
Differential Abundant Metabolites
C)
Amino Acid Pathways
Effect Size
NES
Leucine degradation
Valine leucine and isoleucine degradation
Methionine biosynthesis
Leucine biosynthesis
Proline biosynthesis
Proline metabolism
Tryptophan biosynthesis
Shikimate pathway
Aromatic amino acid
Isoleucine biosynthesis
Histidine biosynthesis
Histidine metabolism
Arginine and proline metabolism
Phenylalanine tyrosine, tryptophan biosynthesis
Valine leucine and isoleucine biosynthesis
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic
NES
D)
Lipid Pathway
Log 2 Fold Change
Phosphetidylcholine PC
Ceramides
Fatty Acid Metabolism Acyl
Long Chain Fatty Acid
Lysophospholipid
Endocannabinoid
Fatty Acid Monohydroxy
Androgenic Steroids
Secondary Bile Acid Metabolism
Fatty Acid Dihydroxy
Lysoplasmalogen
Fatty Acid Dicarboxylate
Progesterone Steroids
Corticosteroids
Sterol
Primary Bile Acid Metabolism
Galactosyl Glycerolipids
Monoacylglycerol
Medium Chain Fatty Acid
E)
Fatty Acid Pathways
Fatty acid biosynthesis, initiation
Fatty acid metabolism
Fatty acid biosynthesis, elongation
Effect Size/NES
WU Obese/Diabetic
WU Healthy
AP Bantu
HG Ju/'hoansi
NES
F)
Effect Size
G)
Effect Size
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic


A)Energy Production and Carbohydrate Metabolism Pathway
NES
Average Effect Size
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic
Carbon fixation
Carbon fixation pathways in prokaryotes
Central carbohydrate metabolism
Other carbohydrate metabolism
Butanoate metabolism
Propanoate metabolism
Ornithine biosynthesis
Ascorbate degradation
2-Oxocarboxylic acid metabolism
Acetyl-CoA pathway
Citrate cycle (TCA cycle)
De Ley-Doudoroff pathway
Gluconeogenesis
Glycogen biosynthesis
Glycolysis (Embden-Meyerhof pathway)
Glycolysis, core module
Incomplete reductive citrate cycle
Leucine degradation
Pentose phosphate pathway
Pentose phosphate pathway non-oxidative
Phosphate acetyltransferase-acetate kinase pathway
Pyruvate oxidation
Wood-Ljungdahl pathway
Arnon-Buchanan cycle
Calvin cycle
Reductive pentose phosphate cycle glyceraldehyde-3P => ribulose-5P
Reductive pentose phosphate cycle ribulose-5P => glyceraldehyde-3P
Semi-phosphorylative Entner-Doudoroff pathway
beta-Oxidation
C)
Carbohydrate Metabolites
log 2 Fold Change
arabionate/xylonate
carboxymethyllysine
lyxonate
mannitol/sorbitol
arabitol/xylitol
galactonate
glycerate
ribonate
Asparylglycosamine
arabinose
ribulose/xylulose
N-acetylneuraminate
mannose
xylose
N-acetylglucosamine
ribose
N-acetylmuramate
glucose 6-phosphate
D)
Energy Metabolites
log 2 Fold Change
malate
fumarate
alpha-ketoglutarate
succinate
2-methylcitrate
isocitric lactone
tricarballylate
E)
Effect Size
g__Faecousia;s__
CAG-390 sp900753295
UBA1259 sp900760875
UMGS1225 sp900549725
g__Anaerovibrio;s__
ER4 sp900546295
g__VUNA01;s__
g__Anaerovibrio;s__
g__Phascolarctobacterium A;s__
g__Firm-10;s__
RUG11142 sp902766215
UMGS1865 sp900768535
Gemmiger sp004555405
ER4 sp900763705
UBA4372 sp900543815
Faecousia sp900549705
Phascolarctobacterium A sp
UMGS1696 sp900763885
Limivicinus sp900549865
UBA644 sp900547165
Faecousia sp900544405
UBA10281 sp900548285
Faecousia sp900549495
g__Limiplasma;s__
g__Faecousia;s__
UMGS911 sp900545935
Limivicinus sp003150355
UMGS1668 sp900553955
Sodaliphilus sp900457565
UMGS1865 sp900547725
CAG-272 sp900433515
Finimonas sp003537755
F23-B02 sp900556535
F)
Effect Size
Bifidobacterium adolescentis
Bifidobacterium bifidum
Bifidobacterium longum
Bifidobacterium angulatum
g__Collinsella;s__
B. pseudocatenulatum
g__Bacteroides;s__
Bacteroides salyersiae
Alistipes putredinis
Paraprevotella clara
Bacteroides stercoris
Dysosmobacter welbionis
Acidaminococcus intestini
Alistipes A indistinctus
Dialister hominis
CAG-41 sp900066215
Bacteroides xylanisolvens
Alistipes onderdonkii
g__Blautia A;s__
Anaerobutyricum hallii
Blautia A obeum
g__Collinsella;s__
Bacteroides faecis
g__Alistipes;s__
Phocaeicola plebeius
Bacteroides uniformis
Streptococcus salivarius
Flavonifractor plautii
f__Lachnospiraceae;g__;s__
Phocaeicola plebeius A
Faecalibacterium prausnitzii D
g__Streptococcus;s__
Lawsonibacter sp900066825
Duodenibacillus massiliensis
Gemmiger formicilis
G)
Effect Size
CHOLD
DHDLC
TNF
VLDL
f_RUG14156;g__;s__
Paraprevotella xylaniphila
Aphodousia faecalis
Clostridium Q saccharolyticum A
Enorma massiliensis
Eubacterium callanderi
Faecalibacterium sp900540455
Mailhella sp905192935
Anaerotignum sp001304995
Blautia hansenii
CAG-475 sp900548005
CAG-475 sp900550915
Erysipelatoclostridium ramosum
Faecalimonas phoceensis
Faecalimonas umbilicata
Ruminococcus B gnavus
Sutterella sp900762445
Anaerotignum sp001304995
Blautia hansenii
CAG-475 sp900548005
CAG-475 sp900550915
Erysipelatoclostridium ramosum
Faecalimonas phoceensis
Faecalimonas umbilicata
Ruminococcus B gnavus
Sutterella sp900762445
H)
Effect Size
TNF Alpha
CRP
Acidaminococcus fermentans
Bifidobacterium pseudocatenulatum
CAG-485 sp900761855
Clostridium Q symbiosum
Coprobacter secundus
Eggerthella sinensis
Enterobacter roggenkampii
Enterococcus B faecium
Erysipelatoclostridium ramosum
f_Anaeroplasmataceae;g__;s__
Fimisoma sp900540145
g_Alistipes;s__
g_UBA1436;s__
Odoribacter sp900541415
Parabacteroides faecavium
Parolsenella catena
Sodaliphilus sp004557565
UBA1394 sp900538575
UBA1436 sp900540405
Weissella cibaria
f_HGM11514;g__;s__
g_Alistipes A;s__
g_Collinsella;s__
HGM13006 sp900757695
Paratractidigestivibacter
I)
HBA1C
Effect Size
Dialister sp900541605
Fusicatenibacter saccharivorans
f_HGM11514;g__;s__
g_Tractidigestivibacter;s__
Collinsella sp900768795
Paratractidigestivibacter sp003150175
Coprococcus sp000154245
Acidaminococcus massiliensis
HGM13006 sp900757695
CAG-273 sp000435755
o_TAN877;f__;g__;s__
f_Gastranaerophilaceae;g__;s__
UBA737 sp900549755
UBA10281 sp900764105
J)
Effect Size
CHOLD
CRP2
DHDLC
HBA1C
TNF Alpha
Trig
VLDL
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic
K)
Short Chain Fatty Acids
Isobutyric acid
Propionic acid
Hexanoic acid
Valeric acid
2 Methybutyric acid
Acetic acid
Butyric acid
Isovaleric acid
NS.
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic
B)
Effect Size/NES
Succinate dehydrogenase,
Glyoxylate cycle
Glucagon signaling pathway
dTDP-L-rhamnose biosynthesis
Galactose metabolism
NAD biosynthesis,
Central carbon metabolism in cancer
NES
HG Ju/'hoansi
AP Bantu
WU Healthy
WU Obese/Diabetic