

The gut mycobiota of rural and urban individuals is shaped by geography

Mubanga Kabwe

University of Pretoria

Surendra Vikram

University of Pretoria

Khodani Mulaudzi

University of Pretoria

Janet Jansson

Pacific Northwest National Laboratory

Thulani Makhalyane (✉ Thulani.makhalanyane@up.ac.za)

University of Pretoria <https://orcid.org/0000-0002-8173-1678>

Research article

Keywords: Gut microbiome, ethnicity, mycobiota, diet, Africa, rural, urban

Posted Date: July 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-19519/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on August 17th, 2020. See the published version at <https://doi.org/10.1186/s12866-020-01907-3>.

Abstract

Background Understanding the structure and drivers of gut microbiota remains a major ecological endeavour. Recent studies have shown that several factors including diet, lifestyle and geography may substantially shape the human gut microbiota. However, most of these studies have focused on the more abundant bacterial component and comparatively less is known regarding fungi in the human gut. This knowledge deficit is especially true for rural and urban African populations. Therefore, we assessed the structure and drivers of rural and urban gut mycobiota.

Results Our participants (n=100) were balanced by geography and sex. The mycobiota of these geographically separated cohorts was characterized using amplicon analysis of the Internal Transcribed Spacer (ITS) gene. We further assessed biomarker species specific to rural and urban cohorts. In addition to phyla which have been shown to be ubiquitous constituents of gut microbiota, Pichia were key constituents of the mycobiota. We found that geographic location was a major driver of gut mycobiota. Other factors such as smoking were also determined gut mycobiota albeit to a lower extent, as explained by the small proportion of total variation. Linear discriminant and the linear discriminant analysis effect size analysis revealed several distinct urban and rural biomarkers.

Conclusions Together, our analysis reveals distinct community structure in urban and rural South African individuals. Geography was shown to be a key driver of rural and urban gut mycobiota.

Background

By comparison to prokaryotes (bacteria and archaea), eukaryotes are considered part of the rare “biosphere” of the gut [1, 2]. Despite their low abundances, fungi play significant roles in host physiology [2-5]. Recent studies have shown that the gut fungal community composition is less stable over time, compared to bacterial communities [4, 6, 7]. These studies suggest that the gut mycobiota is more variable than bacterial communities, and may be influenced substantially by environmental factors [3, 7]. Despite evidence confirming the gut microbiota is diverse and interacts with the host immune system [3, 8, 9], knowledge regarding the community structure of healthy human gut mycobiota remains scant.

Most studies have focused on the potential roles played by the mycobiota in the aetiology of gut diseases [10-12]. These studies have provided crucial insights on the role of the mycobiota as a potential driver of immunological disorders and as opportunistic pathogens in immunocompromised hosts [13]. Further, dysbiosis of gut mycobiota has been linked to obesity, colorectal cancer and Inflammatory Bowel Diseases (IBDs) [12, 14, 15]. Decreased abundances of *Saccharomyces cerevisiae* and higher proportions of *Candida albicans* were found in IBD patients compared to healthy controls. A recent study showed that Crohn’s disease-specific gut environments may select for fungi to the detriment of bacteria suggesting disease-specific inter-kingdom network alterations in IBD [12]. Yet, despite these effects, there remains a clear deficit in knowledge regarding the precise role played by the gut mycobiota in disease prevention. Relatedly, the factors which drive the diversity and community structure of gut mycobiome remain

underexplored. Assessing the influence of environmental factors on the gut mycobiome across a wider group of participants is crucial for determining the effects on host-microbiota dynamics and health.

Several studies have evaluated the effects of age [16-18], gender [17], diet [19], diabetes and obesity [15, 20, 21], anorexia nervosa [22], differences across body sites [23, 24] and geographical locations [6, 25, 26] on mycobiome composition and diversity. Yet, these studies are mostly disease centric or focussed on Asian [26] and/or Western populations [6, 7, 19]. To our knowledge, only one study has investigated the gut eukaryotic diversity of African individuals [27]. Although these studies improved our understanding of the mycobiome, there may be several confounding factors such as genetic differences. These differences make it difficult to assess, for instance, the effects of living in urban or rural areas on the microbiome. The effects of diet, geographic locality and lifestyle, on the gut microbiome are often assumed but rarely examined. Where these relationships are assessed, the majority of studies have primarily focused on the ecologically abundant bacteria [28, 29] with assertions that their patterns will likely hold for other taxa, including mycobiomes.

Here, we applied amplicon sequencing of the fungal internal transcribed spacer (ITS) of the rRNA genes on samples collected from individuals living in urban and rural areas in Africa. We provide the first insights regarding the drivers of mycobiome community structure and potential biomarkers specific to individuals from urban and rural locations. Previous studies of the gut mycobiome have primarily investigated small groups with fewer than 20 individuals [25, 30, 31] with very few studies investigating larger groups [6, 7]. This study represents the first analysis of the faecal mycobiota in a large group of healthy sub-Saharan individuals (100 volunteers). Furthermore, this is the first study which compares the composition and diversity of the gut mycobiome of geographically separated non-western individuals with the same ethnicity. We further explored potential biomarker taxa in urban and rural individuals and explore how these taxa vary between the two areas. Using extensive predictor variables collected from participants, we show that geography and lifestyle structure the gut mycobiome of rural and urban South African individuals.

Results

Similarities and differences between urban and rural individuals

We assessed the faecal mycobiota of South African adults living in rural ($n=50$) and urban ($n= 50$) locations by assessing stool samples (see details regarding sample recruitment in Methods). We recruited an equal number of male and female volunteers. The volunteers from the rural areas were from two villages in the Limpopo province of South Africa. These villages are located roughly 500km from the urban site in Pretoria (**Figure 1a**). To gain insights regarding predictive variables, which may shape the gut microbiome, detailed questionnaires were distributed to all volunteers (**Additional File 1 Questionnaire**). The volunteers from Ha-Ravale and Tshikombani villages (population size of roughly 200,000, representing the rural participants) were on average 24 years (mean \pm 6.3). Volunteers from Pretoria (total population of approximately 2.1 million) were on average 31 years (mean \pm 9.1). The mean

age of the participants was 27 years (mean \pm 7.9) across all samples. The average height and weight of the participants was 1.64 m (mean \pm 0.1) and 69.8 kg (mean \pm 17.6), respectively. The average body mass index (BMI) of all participants was 26.02 kg/m² (mean \pm 6.4), resulting in a group of participants classified as overweight and obese, less than 15.9% of participants were smokers.

Amplicon sequence data from 95 volunteers (samples from 5 rural volunteers were excluded due to low quality reads) generated 5,936,454 raw reads. Of these, 1,636,180 reads were clustered into OTUs at 3% divergence (97% similarity) and 1,911 OTUs were taxonomically classified. The resulting accumulation curves showed reasonable sequence saturation at a regional level (**Additional File 3 Figure S1**).

A higher proportion of fungal OTUs were unique to location with urban and rural samples accounting for 47.9% and 45.3% of reads, respectively (**Additional File 4 Figure S2**). Fungal species richness was higher in the stool samples of urban volunteers compared to rural volunteers (**Figure 1b**). However, there were no significant differences in species richness based on location ($W = 915$, p-value = 0.118) and sample type (Kruskal-Wallis chi-squared = 5.103, df = 3, p-value = 0.164). A significant difference was detected in species richness between the three age groups (Kruskal-Wallis chi-squared = 12.215, df = 2, p-value = 0.002).

Two ubiquitous fungal phyla in urban and rural locations

Overall, four distinct fungal phyla were detected in urban and rural gut mycobiomes, based on sequences with relative abundances above 0.1% (**Figure 1c**). The majority of sequences were assigned to members of the phyla *Ascomycota* and *Basidiomycota*, that constituted 80.7% and 6.1% of the total relative abundance, respectively. Unknown sequences constituted 12.9% of the total relative abundance of the mycobiome. The relative abundance of *Ascomycota* was almost equal between the urban (40.1%) and rural (39.8%) location. Whereas, the relative abundance of *Basidiomycota* was higher in the urban (4.4%) compared to the rural (1.7%) location. Nonetheless, no significant differences were detected in taxa relative abundance between location ($W = 8720$, p-value = 0.3826), sample type (Kruskal-Wallis chi-squared = 0.9454, df = 3, p-value = 0.8145) and age group (Kruskal-Wallis chi-squared = 0.29477, df = 2, p-value = 0.863) at phylum level. In total, 8 distinct fungal classes were identified with *Saccharomycetes* constituting the majority of sequences (52.6%) followed by *Dothideomycetes* (20.9%), *Eurotiomycetes* (4.3%), *Sordariomycetes* (4%) and *Tremellomycetes* (3.98%). The relative abundance of *Saccharomycetes* was higher in the rural (26.8%) compared to the urban (25.8%) location. Whereas, *Dothideomycetes* relative abundance was higher in the urban (14.8%) compared to the rural (6.1%) location. However, no significant differences were detected in taxa relative abundance between location ($W = 3480$, p-value = 1), sample type (Kruskal-Wallis chi-squared = 2.7506, df = 3, p-value = 0.4317) and age group (Kruskal-Wallis chi-squared = 0.55361, df = 2, p-value = 0.7582). *Pichia* dominated our participants (18.9% of total relative abundance), followed by *Candida* (18.2%) and *Cladosporium* (6.1%). Whereas, unknown fungal genera accounted for 18.1% of the relative abundance. The relative abundance of *Pichia* was higher in the rural (12.7%) compared to the urban (6.2%) location. Whereas, the relative abundance of *Candida* and *Cladosporium* was higher in the urban (10.9% and 4.2%, respectively) compared to the rural (7.3% and 2%,

respectively) location. However, no significant difference was found between taxa abundance at the genus level for the gut mycobiota of rural and urban participants based on location ($W = 4417$, $p\text{-value} = 0.9936$), sample type (Kruskal-Wallis chi-squared = 1.1117, df = 3, $p\text{-value} = 0.7742$) and age group (Kruskal-Wallis chi-squared = 0.51268, df = 2, $p\text{-value} = 0.7739$).

To assess the distribution and contribution of taxa in a given sample to the overall community composition, we assessed the local contribution to beta diversity (LCTBD). In line with findings from alpha diversity (observed species richness) analyses, we found that samples from urban volunteers (greater species richness) had a more significant contribution to the overall community diversity ($p\text{-Value} < 0.05$). Samples with high LCTBD had a high abundance of *Basidiomycota* and other unknown taxa. In contrast, only two samples from the rural location contributed more significantly to overall community diversity (**Additional File 5 Figure S3**).

Distinct mycobiota among urban and rural volunteers unrelated to gender

Differences in the fungal community structure between the rural and urban localities were visualized in an non-metric multi-dimensional scaling (NMDS) plot (**Figure 2a**). Urban and rural samples formed distinct clusters [permutational multivariate analysis of variance (PERMANOVA) ($R^2 = 0.070$; $p\text{-Value} = 0.0001$), ANOSIM ($R = 0.43$, $p\text{-Value} = 0.001$) and ADONIS ($R^2 = 0.07034$ $p\text{-Value} = 0.0001$)]. However, male and female samples did not cluster separately. Pairwise analysis using PERMANOVA showed that there was no significant difference between gender within each location ($R^2 = 0.018$; $R^2 = 0.023$; respectively and $p\text{-Value} > 0.4$ for both). Nevertheless, there was a significant difference between the gut mycobiota of female and male participants between the two locations ($R^2 < 0.074$ for all; $p\text{-Value} = 0.001$ for all).

Ecological drivers of gut mycobiota

Redundancy analysis (RDA) was performed to determine which predictor variables significantly explained the variation in fungal community composition (**Figure 2b**). Four predictive variables were significant ($r^2 > 0.2$; $p\text{-Value} < 0.05$) drivers of community composition and structure. Variation partitioning analysis demonstrated that only 3% community variations were explained by these four predictive variables (**Additional File 2 Partition of variance in RDA**). Predictive variables which included; breastfeeding, smoking, mode of birth and location; all of which significantly influenced the fungal community composition.

We conducted correlation analyses to explore the relationships among dominant gut species. Our results showed a few highly positive correlations in the rural participants: between *Mucor* and *Dipodascus*, *Mucor* and *Naganishia*, *Clavispora* and *Lentendrea*, and between *Udeniomyces* and *Lentendrea* (**Figure 3**). Whereas, the strongest negative correlation was found between *Dipodascus* with *Trichoderma*, *Dipodascus* with *Ascotricha* and *Dipodascus* with *Chalastospora*. Within the urban cohort, *Xeromyces* and *Agaricus*, *Diutina* and *Clavispora*, and *Dekkera* and *Diutina* exhibited the strongest positive

correlations (**Figure 3**). The strongest negative correlations were detected between *Clavispora* with *Filobasidium*, and with *Verrucaria* and *Malassezia*.

Biomarker taxa

Linear discriminant analysis (LDA) and the linear discriminant analysis effect size (LEfSe) [32] test for biomarkers was used to detect taxa that showed the strongest effect on group differentiation (**Figure 4a**). OTU level analysis uncovered 14 urban-associated species from 10 genera. Whereas, 17 rural-associated species from 11 genera, were detected as possible biomarkers. The most abundant rural-associated biomarker genera were *Hypopichia* and *Dipodascus*, with species *Hypopichia burtonii* and *Dipodascus geotrichum* being the most abundant (**Figure 4b**). Whereas, the urban-associated biomarkers were dominated by the class *Tremellomycetes* and genera *Dekkera* and *Hannaella*. Whereas, species *Dekkera bruxellensis* and *Hannaella sinensis* dominated the urban-associated biomarkers.

Discussion

The results from this study suggest that the gut mycobiome of the South African population is structured by geography and lifestyle. This finding is supported by the clustering of a large proportion of the fungal OTU's into discrete rural and urban groups within the Venn diagram. Only a small percentage of OTUs were shared between the two populations, which may suggest that factors such as the environment, age and diet may play a role in shaping the differences in OTU clustering. These results were further corroborated by NMDS and Permanova analyses, which not only show that both populations cluster distinctly and share just a few taxa, but that they have diverse fungal community composition consistent with rural and urban locality. Redundancy analysis (RDA) also predicted that the measured variables account for only a small proportion of the variability ($R^2 = 0.2$) in fungal community composition. Indicating that other undetermined factors may be driving differences fungal community composition between the two groups.

Several studies have investigated the healthy human mycobiome [6, 7, 19, 25, 30, 33]. In these studies, geography was not considered as a potential factor structuring the gut mycobiota. For instance, previous studies found no association between host phenotypic characteristics with mycobiome profile [6]. Nash et al. (2017) also suggests that diet, the environment, diurnal cycles, and host genetics may substantially influence the human gut mycobiome. However, the finding that the majority of the variation could not be explained by their metadata does suggest that other environmental factors, such as geography, may contribute to structuring the human mycobiome [6].

Our study provides the first results showing the importance of geography in African populations. Geographic locality may be associated with different environmental factors, such as different climatic regimes, which may effect structural changes in the mycobiota. For example, climate significantly influences vegetation and farming practices and leads to region specific diets. These region-specific diets may ultimately influence the gut mycobiota. This is a reasonable prediction given previous findings showing that fungi have climate dependent biogeographic patterns [34, 35]. These patterns are likely to

determine the type of fungi individuals may be exposed to, which may in turn impact the colonization of fungi in the human gut. The most abundant rural-associated biomarker species found in this study, *Dipodascus geotrichum*, is ubiquitous in nature [36] whereas, *Hypopichia burtonii* is commonly isolated from corn, wheat, and rice [37]. The urban-associated biomarkers were dominated by the species *Dekkera bruxellensis*, which are commonly isolated from fermented food such as wine, beer, feta cheese and sour dough [38-40]. In contrast, *Hannaella sinensis* is commonly isolated from plants and soil [41, 42]. The staple diet of the rural South African population primarily consists of a corn-based porridge (called 'pap'). It is therefore not uncommon for a fungal species commonly isolated from corn to be a dominant biomarker for the rural population. Conversely, the urban population diet was more diverse and included fermented foods such as wine, sour dough bread and feta cheese, which are commonly available in supermarkets. Thus, the species *Dekkera bruxellensis* was identified as a dominant biomarker in the urban population.

In addition to geographic location, we found that smoking, mode of birth and breastfeeding significantly influenced gut fungal communities. Several studies have previously reported that these factors may significantly influence the initial colonization, subsequent composition and structure of bacterial members of the human gut microbiome [28, 43-45]. Suhr et al. (2016) and Hallen-Adams et al. (2017) investigated the gut mycobiome of two cohorts that were exclusively on a vegetarian or a western diet. These studies found that the distribution of fungi differed considerably between the two cohorts [7, 46]. Plant-associated *Fusarium*, *Malassezia*, *Penicillium* and *Aspergillus* species were detected at higher abundances within the vegetarian cohort, compared to the cohort on a conventional diet. The finding that smoking affected fungal community composition and structure is supported by several studies [47-49]. The approximately 4000 chemical compounds produced by cigarettes have been shown to alter the composition of the gut microbiome [47, 49-52]. The reported increase of *Clostridia* induced by smoking in murine models has also been indirectly confirmed in humans where an increased rate of *C. difficile* infection was greater in former and current smokers compared to never smokers [51]. Moreover, the abundance of the fungus *Candida tropicalis* has also been reported to be significantly higher in *C. difficile* infection patients compared to healthy individuals [53]. The abundance of *C. tropicalis* has also been detected to be positively correlated with levels of anti-*Saccharomyces cerevisiae* antibodies (ASCA) [53]. In our study *C. tropicalis* was detected to be higher in individuals who smoke compared to non-smokers whereas the inverse was true for *S. cerevisiae*. These findings may confirm the antagonistic association between the species *C. tropicalis* and *S. cerevisiae*, as previously reported by Hoarau et al. (2016) [53].

Most studies have identified the genera *Candida*, *Saccharomyces*, *Malassezia* and *Aspergillus* as the three most abundant in the gut of healthy individuals [6, 25, 54]. To the best of our knowledge, our study is the first to report *Pichia* as one of the top four (*Pichia*, *Candida*, *Cladosporium* and *Paraconiothyrium*) most abundant genera in the human gut mycobiome. This may be due to several factors including differences in cohort characteristics (e.g., geographical location, diet, genetic predisposition and climate). *Pichia* have been identified as both constituent members of the human oral [55, 56] and gut microbiome [33]. Mukherjee detected a 1:1 abundance ratio in the oral mycobiome of individuals when *Candida* and

Pichia were present together [56]. *Pichia* was also observed to have an antagonistic effect against *Candida*, *Fusarium* and *Aspergillus* [56].

The yeast genera, *Pichia*, *Candida* and *Cladosporium*, dominated the South African gut mycobiome. Our findings agree with previous studies which show that members of the *Aspergillus*, *Candida*, *Debaryomyces*, *Malassezia*, *Penicillium*, *Pichia*, and *Saccharomyces* genera were the most recurrent and/or dominant fungal genera [33, 46, 57]. In contrast to previous findings, our data indicate higher relative abundances of *Cladosporium*, detection of *Mucor* and the absence or low abundance of genera such as *Cyberlindnera*, and *Galactomyces* [6, 19, 58]. Previous studies found that the gut mycobiome of a cohort from Houston, Texas, was dominated by *Saccharomyces*, *Malassezia* and *Candida* [6]. By contrast, the genus *Malassezia* was not detected in the gut mycobiome of a Pennsylvania cohort, which was instead dominated by the genera *Saccharomyces* and *Candida* [19]. Differences in study methodologies may be a source of these conflicting findings [6]. One study amplified the Internal Transcribed Spacer 2 (ITS2) region of the fungal rRNA gene [6], and the second amplified the ITS1 region [19]. Studies similar to the work by Gardes et al. (1993) and White et al. (1990), where ITS1F and ITS2 primer sets were used to amplify the ITS2 region, did not detect *Malassezia* [59, 60]. The second reason for the observed differences has been attributed to differences in cohort characteristics, such as diet and/or geographical location. Strati (2016) and Raimondi's (2019) investigating cohorts in Italy, detected same dominant fungal genera [58, 61], and the investigation of cohorts in two different states in the USA observed different results [6, 45]. We used ITS1 and ITS4 in this study and found that the genera *Pichia*, *Candida* and *Cladosporium* dominated the urban cohort, whereas genera *Pichia*, *Candida* and *Aspergillus* dominated the rural cohort. The dominant taxa identified in urban and rural locations further support our assertion that geographic location plays a major role in the observed differences.

Candida albicans was the most dominant taxon in our cohort and is frequently reported as the most abundant *Candida* species in both diseased [62] and healthy individuals [63]. *Candida* spp. not only colonize the gut [19, 33] but several other body sites, including the oral cavity [55, 64], vagina [65], and skin [66, 67]. However, *Candida* are autochthonous to the mammalian digestive tract and species including *Candida albicans*, *C. tropicalis*, *C. parapsilosis*, and *C. glabrata* may grow and colonize at 37°C [7]. A review of the literature suggest that *C. albicans* carriage in healthy individuals ranges from 30–60% [68] and that living mammals are considered a niche for these species as they are not found in significant concentrations in soil, food or air [69, 70]. Raimondi et al., (2019) reported that *C. albicans* was frequently detected and dominated the cultivable mycobiota of different faecal samples [61].

Conclusions

This study provides the first insight into the importance of geography and lifestyle factors on the gut mycobiome in rural and urban locations in Africa. We found that fungi in the gut display distinct patterns consistent with geographic locality. Redundancy analysis showed that several lifestyle factors were major drivers explaining the distinct community structure. The results of biomarker analysis revealed several ecologically important fungal taxa, which were unique to individuals from urban and rural areas.

The finding that certain taxa may be biomarker species have potential consequences for certain groups including immunocompromised individuals living in rural and urban locations. Increases in the abundances of these taxon may lead to deleterious effects on the health of these groups. Such findings provide a valid basis for the development of novel therapeutics or preventative measures reliant on modulating the gut mycobiome.

Methods

Participant enrolment criteria for urban and rural areas

Volunteers were recruited from two rural locations and one urban location. For rural volunteers, we recruited individuals following traditional diets, with generally low levels of processed foods. Urban cohorts reported mixed diets and increased consumption of processed foods. Volunteers from the Ha-Ravale (females; n = 1 and males; n = 15) and Tshikombani (females; n = 24 and males; n = 10) villages located in the Vhembe District of the Limpopo Province comprised the rural cohort. Both villages are approximately 391 km and 439 km, respectively, from the closest city (Pretoria). This city, in the Gauteng province of South Africa, served as the urban sampling area. In total, 100 stool samples were collected from healthy volunteers. These samples were equally divided between gender and locality [i.e. rural (25 males and 25 females) and urban (25 males and 25 females)]. Participants were categorized by age into young adults (ages 18-27 years; n = 61), middle-aged adults (ages 28-37 years, n = 28), and older adults (aged older than 37 years, n = 11). The height of the participants was measured in meters, weight in kilograms and the BMI was calculated using the formula $BMI = \text{kg}/\text{m}^2$ [71] where kg is a person's weight in kilograms and m^2 is their height in meters squared. Participants were categorized by BMI into Underweight = $BMI < 18.5$ (n=10), Normal weight $BMI = 18.5-24.9$ (n=46), Overweight $BMI = 25-29.9$ (n=22) and Obese = BMI of 30 or greater (n=22) [71]. Self-stool collection kits were provided to all volunteers (Easy Sampler® Stool collection Kit, Hounisen Lab Equipment A/S, Skanderborg, Denmark).

Inclusion and exclusion criteria

The participants were all healthy adults age 18 – 50 years. Volunteers reporting antibiotic use/other treatments within six months prior to participating in the study and sample collection were excluded from the study. Similarly, individuals who had been diagnosed with any inflammatory-related bowel diseases or gastrointestinal diseases within six months prior to sample collection were excluded from the study.

DNA extraction

DNA was isolated using the PowerSoil DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA) following the manufacturer's specifications with minor modifications. Briefly, approximately 0.25g of stool sample was transferred into the Power-Bead tubes using a sterile disposable wooden spatula (Lasec Laboratories, RSA). The sample was homogenized by gently vortexing the tubes for 10 s before adding 60 μL of the lysis buffer. This was then incubated for 30 min. at 55°C prior to centrifugation at room temperature for 30 s at 10,000 x g. The supernatant from this step was transferred to sterile 2 mL

tubes and 250 µL of inhibitor removal reagent was added to this. The samples were incubated on ice for 5 min., thereafter approximately 1.2 mL of binding buffer was added. Next, 70% ethanol (500 µL) was added and the contents precipitated by centrifugation at room temperature for 60 s at 10,000 x g. The DNA was eluted with 100 µL filter-sterilised autoclaved Millipore water and quantified using the NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific, Waltham, MA, USA). The quality of isolated DNA was confirmed by agarose gel electrophoresis, on 1% (w/v) agarose gel in 1 X TAE buffer (0.2% [w/v] Tris, 0.5% [v/v] acetic acid, 1% [v/v] 5 M EDTA [pH 8]) at 90 Volts for 45 min. in a BioRad SubCell® GT gel electrophoresis system with gel red visualising agent. The gel was visualised using the BioRad Gel Doc system and viewed with a UV Trans-illuminator.

ITS gene region amplification, sequencing and data processing

The internal transcribed spacer (ITS) region (420 to 825 bp) was amplified using fungal-specific primers [72]: ITS1F (5'-CTTGGTCATTAGAGGAAGTAA-3') [59] and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') [60] with barcode inserted on the forward primer. Briefly, the HotStarTaq Plus Master Mix Kit (Qiagen, USA) was used for the PCR amplification reaction (94°C for 3 minutes, followed by 30 cycles of 94°C for 30 seconds, 53°C for 40 seconds, 72°C for 1 minute and final elongation step at 72°C for 5 minutes). The PCR products were checked in 2% agarose gel to determine the success of amplification and the relative intensity of bands. Amplicons from different samples were pooled to equal proportions based on their molecular weight and DNA concentrations. The pooled DNA was purified of short fragments using Agencourt Ampure beads (Agencourt Bioscience Corporation, USA). Then the pooled and purified PCR product was used to prepare Illumina library. Paired end 2 x 250bp sequencing was performed on an Illumina MiSeq instrument (Illumina Inc., San Diego, CA, USA) at Mr DNA (Shallowater, TX, USA).

The resultant data were analysed using the Quantitative Insights into Microbial Ecology (QIIME2) software version 2018.8.0 [73]. Demultiplexed sequences were merged and assessed for quality. Sequences shorter than 200 bp, with quality scores below 25, containing more than two ambiguous characters or more than one mismatch to the sample-specific barcode or the primer sequences, were excluded from further downstream analyses. Sequences were denoised, chimeric sequence removed and operation taxonomic units (OTUs) were defined by clustering at 3% divergence (97% similarity) using USEARCH v11 [74]. Taxonomies were assigned to each OTU using the UNITE (release 7_99) databases for fungi [75]. Singletons were excluded, and each sample was randomly subsampled (rarefied) to the same number of sequences per sample (17 980).

Statistical analyses

All statistical analyses were performed in R version 3.5.1 using R studio [76, 77]. Alpha diversity (observed richness), together with rarefaction curves were calculated and visualized using the R packages "phyloseq" and "ggplot". First, the Shapiro-Wilk's test was used to determine whether the data had a normal distribution [78]. Subsequently, the unpaired two-sample Wilcoxon rank sum test [79, 80] was applied to determine significant differences between the alpha diversity indices using the R packages

“dplyr” version 0.4.3 and the “ggpubr” version 0.1.8. [81, 82]. In these analyses, the rural or urban location was specified as a random factor.

The R packages “phyloseq” [83] and “microbiomeseq” [84] were used to calculate and visualize relative taxa abundance at phylum and class level. OTU abundance was transformed to relative abundance and taxa with relative abundance less than 0.1% were removed. The Wilcoxon rank sum test was applied to determine significant differences between taxa relative abundance in the urban and rural samples. Whereas, the Kruskal-Wallis test [83][85] was applied to determine significant differences in taxa relative abundance between the four sample types (rural female, rural male, urban female and urban male).

The LCBD was calculated according to [86]. The LCBD describes the degree of uniqueness of a given sample in relation to the overall community composition. The taxa abundance was normalized to obtain the proportion of most abundant taxa per sample. Location was used as the grouping variable and the Hellinger method [87] was used for the dissimilarity coefficients calculation.

Pairwise similarities among samples were calculated using the Bray–Curtis index of similarity. The resulting matrix was represented visually in a nonmetric multidimensional scaling (NMDS) plot to observe community structure. Using the vegan package [88], a permutational multivariate analysis of variance (PERMANOVA) [89] based on 9999 permutations of the data, was performed to test whether differences between sample groupings in the NMDS ordinations were statistically significant. Microbial community similarities and the homogeneity of dispersion between the rural and urban sample groups were tested using the ANOSIM and ADONIS tests, respectively [90, 91].

The effect of the different recorded environmental factors on fungal community composition and structure was determined through redundancy analysis (RDA). The contribution of highly correlating OTUs (p -Value < 0.05) with redundancy axes was identified using the envfit function from the R package vegan [88].

Fungal-fungal relationships were interrogated using SparCC [92]. Correlation was based on measuring the linear relationship between log transformed abundances. First, data were filtered to remove OTUs that had less than 2 reads on average. SparCC was used to generate true correlation coefficients from which pseudo p-values were calculated. The calculate pseudo p-values were false discovery rate (FDR) adjusted [93] and the correlation matrix was visualized using the “corrplot” function [94] in R.

Potential biomarker taxa which differed in abundance and occurrence between the two geographic groups were detected by linear discriminant analysis (LDA) effect size (LEfSe) [32]. The LEfSe was calculated using the online Galaxy web application [95] with the Huttenhower lab’s tool (<https://galaxyproject.org/learn/visualization/custom/lefse/>). First the nonparametric factorial Kruskal–Wallis sum rank tests (alpha = 0.01) was used to detect differential abundant features (at genera, family, class and phylum level) within the two geographic locations (rural and urban). The phylogenetic consistency was then tested using the pairwise Wilcoxon rank-sum tests (alpha = 0.01). Finally, the effect size of each differentially abundant feature was estimated using the LDA. The all-against-all classes were

compared (most stringent) and a linear discriminant analysis score value of 2.0 was chosen as threshold for discriminative features.

Abbreviations

IBD: irritable bowel disease; ITS: internal transcribed spacer; HMO: human milk oligosaccharides; OTU: operational taxonomic unit; LCTBD: local contribution to beta diversity; NMDS: nonmetric multidimensional scaling;

Declarations

Ethical approval and consent to participate

All experiments were approved by the Ethics Approval Committee of the Faculty of Health Sciences at the University of Pretoria (EC 160630-051). Participants approved and provided informed consent prior to enrolment in this study. Participant provided written consent. All experimental methods and experiments were in accordance with the Helsinki Declaration.

Consent for publication

All participants provided consent for publication of study results of the collected biomaterials paired with anonymized information on age, sex, location, diet and other data.

Availability of data and materials

The sequence data generated in this study are available on the NCBI (<https://www.ncbi.nlm.nih.gov/>) under the following accession number: PRJNA589500.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the South African Medical Research Foundation (TPM). MHK and KM received scholarships from the National Research Foundation. We wish to thank the Fulbright program for providing sabbatical support to TPM, JJ was funded by the Microbiomes in Transition (MinT) initiative at the Pacific Northwest National Laboratory in Richland, WA, USA. PNNL is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

Authors' contributions

T.P.M designed the research; Faecal samples were collected by K.M with support from M.H.K and TPM. M.H.K and K.M performed the analyses; M.H.K, S.V and K.M analysed the data; M.H.K and T.P.M. wrote

the paper. T.P.M. and J.J. edited the paper. All authors read and approved the manuscript.

Acknowledgements

We would like to thank all the volunteers for donating their data, time and samples to our research. We also thank the Centre for High Performance Computing (Cape Town, South Africa) and the University of Pretoria's Centre for Bioinformatics and Computational Biology for providing computational resources.

Authors' information

¹Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa

2. Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, Washington 99352, USA

Corresponding author

Correspondence to Thulani P. Makhalyane

References

1. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ: **Microbial diversity in the deep sea and the underexplored “rare biosphere”**. *Proceedings of the National Academy of Sciences* 2006, **103**(32):12115-12120.
2. Richard ML, Sokol H: **The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases**. *Nature Reviews Gastroenterology & Hepatology* 2019, **16**(6):331-345.
3. Underhill DM, Iliev ID: **The mycobiota: interactions between commensal fungi and the host immune system**. *Nature Reviews Immunology* 2014, **14**(6):405.
4. Oever Jt, Netea MG: **The bacteriome–mycobiome interaction and antifungal host defense**. *European journal of immunology* 2014, **44**(11):3182-3191.
5. Huffnagle GB, Noverr MC: **The emerging world of the fungal microbiome**. *Trends in microbiology* 2013, **21**(7):334-341.
6. Nash AK, Auchtung TA, Wong MC, Smith DP, Gesell JR, Ross MC, Stewart CJ, Metcalf GA, Muzny DM, Gibbs RA: **The gut mycobiome of the Human Microbiome Project healthy cohort**. *Microbiome* 2017, **5**(1):153.
7. Hallen-Adams HE, Suhr MJ: **Fungi in the healthy human gastrointestinal tract**. *Virulence* 2017, **8**(3):352-358.
8. Rizzetto L, De Filippo C, Cavalieri D: **Richness and diversity of mammalian fungal communities shape innate and adaptive immunity in health and disease**. *European journal of immunology* 2014, **44**(11):3166-3181.

9. Romani L: **Immunity to fungal infections.** *Nature Reviews Immunology* 2011, **11**(4):275.
10. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosh D, Panzer AR, LaMere B, Rackaityte E, Lukacs NW: **Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation.** *Nature medicine* 2016, **22**(10):1187.
11. Ott SJ, Kühbacher T, Musfeldt M, Rosenstiel P, Hellwig S, Rehman A, Drews O, Weichert W, Timmis KN, Schreiber S: **Fungi and inflammatory bowel diseases: alterations of composition and diversity.** *Scandinavian journal of gastroenterology* 2008, **43**(7):831-841.
12. Sokol H, Leducq V, Aschard H, Pham H-P, Jegou S, Landman C, Cohen D, Liguori G, Bourrier A, Nion-Larmurier I: **Fungal microbiota dysbiosis in IBD.** *Gut* 2017, **66**(6):1039-1048.
13. Iliev ID, Funari VA, Taylor KD, Nguyen Q, Reyes CN, Strom SP, Brown J, Becker CA, Fleshner PR, Dubinsky M: **Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis.** *Science* 2012, **336**(6086):1314-1317.
14. Luan C, Xie L, Yang X, Miao H, Lv N, Zhang R, Xiao X, Hu Y, Liu Y, Wu N: **Dysbiosis of fungal microbiota in the intestinal mucosa of patients with colorectal adenomas.** *Scientific reports* 2015, **5**:7980.
15. Rodríguez MM, Pérez D, Chaves FJ, Esteve E, Marin-Garcia P, Xifra G, Vendrell J, Jové M, Pamplona R, Ricart W: **Obesity changes the human gut mycobiome.** *Scientific reports* 2015, **5**:14600.
16. Gewolb IH, Schwalbe RS, Taciak VL, Harrison TS, Panigrahi P: **Stool microflora in extremely low birthweight infants.** *Archives of Disease in Childhood-Fetal and Neonatal Edition* 1999, **80**(3):F167-F173.
17. Strati F, Di Paola M, Stefanini I, Albanese D, Rizzetto L, Lionetti P, Calabò A, Jousson O, Donati C, Cavalieri D: **Age and gender affect the composition of fungal population of the human gastrointestinal tract.** *Front Microbiol* 2016, **7**:1227.
18. Heisel T, Podgorski H, Staley CM, Knights D, Sadowsky MJ, Gale CA: **Complementary amplicon-based genomic approaches for the study of fungal communities in humans.** *PLoS One* 2015, **10**(2):e0116705.
19. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD: **Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents.** *PLoS One* 2013, **8**(6):e66019.
20. Gouba N, Raoult D, Drancourt M: **Plant and fungal diversity in gut microbiota as revealed by molecular and culture investigations.** *PloS one* 2013, **8**(3):e59474.
21. Kowalewska B, Zorena K, Szmigiero-Kawko M, Wąż P, Myśliwiec M: **Higher diversity in fungal species discriminates children with type 1 diabetes mellitus from healthy control.** *Patient preference and adherence* 2016, **10**:591.
22. Gouba N, Raoult D, Drancourt M: **Gut microeukaryotes during anorexia nervosa: a case report.** *BMC research notes* 2014, **7**(1):33.
23. Zhang E, Tanaka T, Tajima M, Tsuboi R, Nishikawa A, Sugita T: **Characterization of the skin fungal microbiota in patients with atopic dermatitis and in healthy subjects.** *Microbiology and immunology*

- 2011, **55**(9):625-632.
24. Nguyen LD, Viscogliosi E, Delhaes L: **The lung mycobiome: an emerging field of the human respiratory microbiome.** *Front Microbiol* 2015, **6**:89.
25. Hamad I, Sokhna C, Raoult D, Bittar F: **Molecular detection of eukaryotes in a single human stool sample from Senegal.** *PloS one* 2012, **7**(7):e40888.
26. Nam Y-D, Chang H-W, Kim K-H, Roh SW, Kim M-S, Jung M-J, Lee S-W, Kim J-Y, Yoon J-H, Bae J-W: **Bacterial, archaeal, and eukaryal diversity in the intestines of Korean people.** *The Journal of Microbiology* 2008, **46**(5):491-501.
27. Hamad I, Raoult D, Bittar F: **Repertory of eukaryotes (eukaryome) in the human gastrointestinal tract: taxonomy and detection methods.** *Parasite immunology* 2016, **38**(1):12-36.
28. Xu Z, Knight R: **Dietary effects on human gut microbiome diversity.** *Br J Nutr* 2015, **113 Suppl**:S1-5.
29. Wang WL, Xu SY, Ren ZG, Tao L, Jiang JW, Zheng SS: **Application of metagenomics in the human gut microbiome.** *World J Gastroenterol* 2015, **21**(3):803-814.
30. Suhr MJ, Banjara N, Hallen-Adams HE: **Sequence-based methods for detecting and evaluating the human gut mycobiome.** *Letters in applied microbiology* 2016, **62**(3):209-215.
31. Raimondi S, Amaretti A, Gozzoli C, Simone M, Righini L, Candelier F, Brun P, Ardizzone A, Colombari B, Paulone S: **Longitudinal survey of fungi in the human gut: ITS profiling, phenotyping and colonization.** *Front Microbiol* 2019, **10**:1575.
32. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C: **Metagenomic biomarker discovery and explanation.** *Genome biology* 2011, **12**(6):R60.
33. Hallen-Adams HE: **Fungi inhabiting the healthy human gastrointestinal tract: a diverse and dynamic community.** *Fungal Ecology* 2015, **15**:9.
34. Wang D, Rui Y, Ding K, Cui X, Hao Y, Tang L, Pang Z, Zhang B, Zhou S, Wang K: **Precipitation drives the biogeographic distribution of soil fungal community in Inner Mongolian temperate grasslands.** *Journal of soils and sediments* 2018, **18**(1):222-228.
35. Summerell BA, Laurence MH, Liew EC, Leslie JF: **Biogeography and phylogeography of Fusarium: a review.** *Fungal Diversity* 2010, **44**(1):3-13.
36. Kolecka A, Khayhan K, Groenewald M, Theelen B, Arabatzis M, Velegraki A, Kostrzewska M, Mares M, Taj-Aldeen SJ, Boekhout T: **Identification of medically relevant species of arthroconidial yeasts by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry.** *Journal of clinical microbiology* 2013, **51**(8):2491-2500.
37. Kurtzman CP: **Hyphopichia von Arx & van der Walt (1976).** In: *The Yeasts*. Elsevier; 2011: 435-438.
38. Schifferdecker AJ, Dashko S, Ishchuk OP, Piškur J: **The wine and beer yeast Dekkera bruxellensis.** *Yeast* 2014, **31**(9):323-332.
39. Fadda ME, Cosentino S, Deplano M, Palmas F: **Yeast populations in Sardinian feta cheese.** *International Journal of Food Microbiology* 2001, **69**(1):153-156.

40. Meroth CB, Hammes WP, Hertel C: **Identification and population dynamics of yeasts in sourdough fermentation processes by PCR-denaturing gradient gel electrophoresis.** *Appl Environ Microbiol* 2003, **69**(12):7453-7461.
41. Into P, Pontes A, Sampaio JP, Limtong S: **Yeast Diversity Associated with the Phylloplane of Corn Plants Cultivated in Thailand.** *Microorganisms* 2020, **8**(1):80.
42. Landell M, Brandão L, Barbosa A, Ramos J, Safar S, Gomes F, Sousa F, Morais P, Broetto L, Leoncini O *et al:* **Hannaella pagnoccae sp. nov., a tremellaceous yeast species isolated from plants and soil.** *International journal of systematic and evolutionary microbiology* 2014, **64**.
43. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP: **Human gut microbiome viewed across age and geography.** *nature* 2012, **486**(7402):222.
44. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R: **Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.** *Proceedings of the National Academy of Sciences* 2010, **107**(26):11971-11975.
45. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R: **Linking long-term dietary patterns with gut microbial enterotypes.** *Science* 2011, **334**(6052):105-108.
46. Suhr MJ: **Sequencebased methods for detecting and evaluating the human gut mycobiome.** *Letters in Applied Microbiology* 2016, **62**(3):209.
47. Capurso G, Lahner E: **The interaction between smoking, alcohol and the gut microbiome.** *Best practice & research Clinical gastroenterology* 2017, **31**(5):579-588.
48. Biedermann L, Brülisauer K, Zeitz J, Frei P, Scharl M, Vavricka SR, Fried M, Loessner MJ, Rogler G, Schuppler M: **Smoking cessation alters intestinal microbiota: insights from quantitative investigations on human fecal samples using FISH.** *Inflammatory bowel diseases* 2014, **20**(9):1496-1501.
49. Allais L, Kerckhof FM, Verschueren S, Bracke KR, De Smet R, Laukens D, Van den Abbeele P, De Vos M, Boon N, Brusselle GG: **Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut.** *Environmental microbiology* 2016, **18**(5):1352-1363.
50. Biedermann L, Zeitz J, Mwinyi J, Sutter-Minder E, Rehman A, Ott SJ, Steurer-Stey C, Frei A, Frei P, Scharl M: **Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans.** *PloS one* 2013, **8**(3):e59260.
51. Rogers MA, Greene MT, Saint S, Chenoweth CE, Malani PN, Trivedi I, Aronoff DM: **Higher rates of Clostridium difficile infection among smokers.** *Plos one* 2012, **7**(7):e42091.
52. Tomoda K, Kubo K, Asahara T, Andoh A, Nomoto K, Nishii Y, Yamamoto Y, Yoshikawa M, Kimura H: **Cigarette smoke decreases organic acids levels and population of bifidobacterium in the caecum of rats.** *The Journal of toxicological sciences* 2011, **36**(3):261-266.
53. Hoarau G, Mukherjee P, Gower-Rousseau C, Hager C, Chandra J, Retuerto M, Neut C, Vermeire S, Clemente J, Colombel J-F: **Bacteriome and mycobiome interactions underscore microbial dysbiosis**

- in familial Crohn's disease. *MBio* 2016, **7**(5):e01250-01216.
54. Hallen-Adams HE: **Fungi in the healthy human gastrointestinal tract.** *Virulence* 2017, **8**(3):352.
55. Dupuy AK, David MS, Li L, Heider TN, Peterson JD, Montano EA, Dongari-Bagtzoglou A, Diaz PI, Strausbaugh LD: **Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of Malassezia as a prominent commensal.** *PLoS One* 2014, **9**(3):e90899.
56. Mukherjee PK, Chandra J, Retuerto M, Sikaroodi M, Brown RE, Jurevic R, Salata RA, Lederman MM, Gillevet PM, Ghannoum MA: **Oral mycobiome analysis of HIV-infected patients: identification of Pichia as an antagonist of opportunistic fungi.** *PLoS pathogens* 2014, **10**(3):e1003996.
57. Rodríguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, Avershina E, Rudi K, Narbad A, Jenmalm MC: **The composition of the gut microbiota throughout life, with an emphasis on early life.** *Microbial ecology in health and disease* 2015, **26**(1):26050.
58. Strati F, Cavalieri D, Albanese D, De Felice C, Donati C, Hayek J, Jousson O, Leoncini S, Pindo M, Renzi D: **Altered gut microbiota in Rett syndrome.** *Microbiome* 2016, **4**(1):41.
59. Gardes M, Bruns TD: **ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts.** *Molecular ecology* 1993, **2**(2):113-118.
60. White TJ, Bruns T, Lee S, Taylor J: **Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics.** *PCR protocols: a guide to methods and applications* 1990, **18**(1):315-322.
61. Raimondi S, Amaretti A, Gozzoli C, Simone M, Righini L, Candeliere F, Brun P, Ardizzone A, Colombari B, Paulone S *et al*: **Longitudinal Survey of Fungi in the Human Gut: ITS Profiling, Phenotyping, and Colonization.** *Front Microbiol* 2019, **10**(1575).
62. Calderone RA, Clancy CJ: **Candida and candidiasis:** American Society for Microbiology Press; 2011.
63. Silva S, Negri M, Henriques M, Oliveira R, Williams DW, Azeredo J: **Candida glabrata, Candida parapsilosis and Candida tropicalis: biology, epidemiology, pathogenicity and antifungal resistance.** *FEMS microbiology reviews* 2012, **36**(2):288-305.
64. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM: **Characterization of the oral fungal microbiome (mycobiome) in healthy individuals.** *PLoS pathogens* 2010, **6**(1):e1000713.
65. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspöllu A, Väin E, Saarma I, Salumets A, Donders GG, Metsis M: **Characterization of the vaginal micro-and mycobiome in asymptomatic reproductive-age Estonian women.** *PloS one* 2013, **8**(1):e54379.
66. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, Sequencing NISCC: **Topographic diversity of fungal and bacterial communities in human skin.** *Nature* 2013, **498**(7454):367.
67. Leung MH, Chan KC, Lee PK: **Skin fungal community and its correlation with bacterial community of urban Chinese individuals.** *Microbiome* 2016, **4**(1):46.
68. Moran G, Coleman D, Sullivan D: **An introduction to the medically important Candida species.** In: *Candida and Candidiasis, Second Edition.* American Society of Microbiology; 2012: 11-25.

69. Saleh HA, Moawad AA, El-Hariri M, Refai MK: **Prevalence of yeasts in human, animals and soil sample at El-Fayoum Governorate in Egypt.** *Int J Microbiol Res* 2011, **2**(3):233-239.
70. Vogel C, Rogerson A, Schatz S, Laubach H, Tallman A, Fell J: **Prevalence of yeasts in beach sand at three bathing beaches in South Florida.** *Water research* 2007, **41**(9):1915-1920.
71. Lim JU, Lee JH, Kim JS, Hwang YI, Kim T-H, Lim SY, Yoo KH, Jung K-S, Kim YK, Rhee CK: **Comparison of World Health Organization and Asia-Pacific body mass index classifications in COPD patients.** *Int J Chron Obstruct Pulmon Dis* 2017, **12**:2465-2475.
72. Manter DK, Vivanco JM: **Use of the ITS primers, ITS1F and ITS4, to characterize fungal abundance and diversity in mixed-template samples by qPCR and length heterogeneity analysis.** *Journal of Microbiological Methods* 2007, **71**(1):7-14.
73. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F: **QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science.** In.: PeerJ Preprints; 2018.
74. Edgar RC: **Search and clustering orders of magnitude faster than BLAST.** *Bioinformatics* 2010, **26**(19):2460-2461.
75. Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E: **UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi.** *New Phytologist* 2005, **166**(3):1063-1068.
76. RDevelopmentCoreTeam: **R: A Language and Environment for Statistical Computing: A Graduate Course in Probability.** R Foundation for Statistical Computing, Vienna, Austria. In.; 2018.
77. Wickham H, Wickham MH: **The ggplot package.** In.; 2007.
78. Villasenor Alva JA, Estrada EG: **A generalization of Shapiro–Wilk's test for multivariate normality.** *Communications in Statistics—Theory and Methods* 2009, **38**(11):1870-1883.
79. Mann HB, Whitney DR: **On a test of whether one of two random variables is stochastically larger than the other.** *The annals of mathematical statistics* 1947:50-60.
80. Wilcoxon F: **Individual comparisons by ranking methods.** *Biometrics bulletin* 1945, **1**(6):80-83.
81. Wickham H, Francois R, Henry L, Müller K: **dplyr: A Grammar of Data Manipulation.** R package version 0.4. 3. R Found Stat Comput, Vienna <https://CRAN.R-project.org/package=dplyr> 2015.
82. Kassambara A: **ggpubr: 'ggplot2'Based Publication Ready Plots.** R package version 0.1. 8. In.; 2018.
83. McMurdie PJ, Holmes S: **phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data.** *PloS one* 2013, **8**(4):e61217.
84. Ssekagiri A, Sloan W, Ijaz U: **microbiomeSeq: an R package for microbial community analysis in an environmental context.** In: *ISCB Africa ASBCB conference: 2017*.
85. Breslow N: **A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship.** *Biometrika* 1970, **57**(3):579-594.
86. Legendre P, De Cáceres M: **Beta diversity as the variance of community data: dissimilarity coefficients and partitioning.** *Ecology letters* 2013, **16**(8):951-963.

87. Rao CR: **A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance.** *Qüestiió: quaderns d'estadística i investigació operativa* 1995, **19**(1).
88. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H: **Vegan: community ecology package.** R package version 1.17-4. <http://cran.r-project.org>> Acesso em 2010, **23**:2010.
89. Anderson MJ: **A new method for non-parametric multivariate analysis of variance.** *Austral ecology* 2001, **26**(1):32-46.
90. Gray J, Aschan M, Carr M, Clarke K, Green R, Pearson T, Rosenberg R, Warwick R: **Analysis of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment.** *Marine Ecology Progress Series* 1988, **46**:151-165.
91. Anderson MJ, Walsh DC: **PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing?** *Ecological monographs* 2013, **83**(4):557-574.
92. Friedman J, Alm EJ: **Inferring correlation networks from genomic survey data.** *PLoS computational biology* 2012, **8**(9):e1002687.
93. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *Journal of the Royal statistical society: series B (Methodological)* 1995, **57**(1):289-300.
94. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J: **Package 'corrplot'.** *Statistician* 2017, **56**:316-324.
95. Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J: **Galaxy: a web-based genome analysis tool for experimentalists.** *Current protocols in molecular biology* 2010, **89**(1):19.10. 11-19.10. 21.

Supplementary Figure Caption

Additional File 1 Questionnaire Details of the questionnaire provided to participants prior to enrolment in the study. The questionnaire details essential required information, clinical information, voluntary dietary information and questions regarding data sharing.

Additional File 2 Results from the partition of variance analysis in RDA

Additional File 3 Figure S1 Rarefaction plot showing sequencing coverage. The estimated average sequence coverage of high-quality paired end reads after quality control assessed using Nonpareil (in alignment mode).

Additional File 4 Figure S2 Venn diagram showing the unique and shared phylotypes for samples collected from urban and rural participants.

Additional File 5 Figure S3 Taxa abundance data was normalised to obtain the proportion of most abundant taxa per sample. The diameter of the points at the bottom of the plot corresponds to the

magnitude of the LCBD value for a particular sample. The bars correspond to taxa that are most abundant with the top taxa sharing a bigger portion of the bar for each sample.

Figures

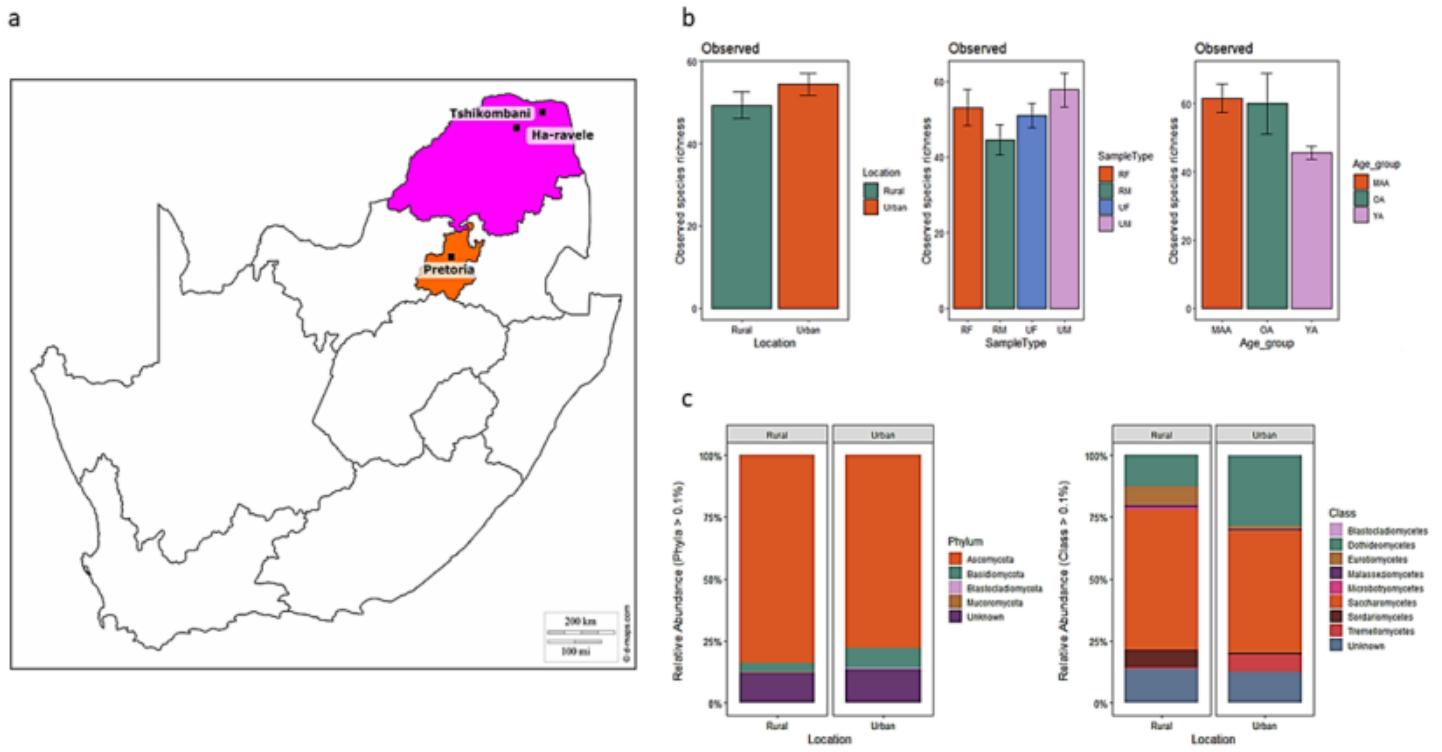


Figure 1

Geographic locations and diversity estimates a) The three sampling locations in Gauteng (Pretoria) and Limpopo (Ha-ravale and Tshikombani) provinces of South Africa b) The differences in mycobiota species richness between the two locations, gender and age group and, c) The relative abundances of taxa at phylum and class levels within each location. The abundance of each taxon was calculated as the percentage of sequences per gender (RF = Rural female, RM = Rural male, UF = urban female and UM = Urban male) from each location for a given microbial group. The group designated as 'Unknown' encompasses unclassified sequences together with classes representing >0.1% of the total sequences. The bar size represents the relative abundance of specific taxa in the particular group, with colours referring to taxa according to the legend.

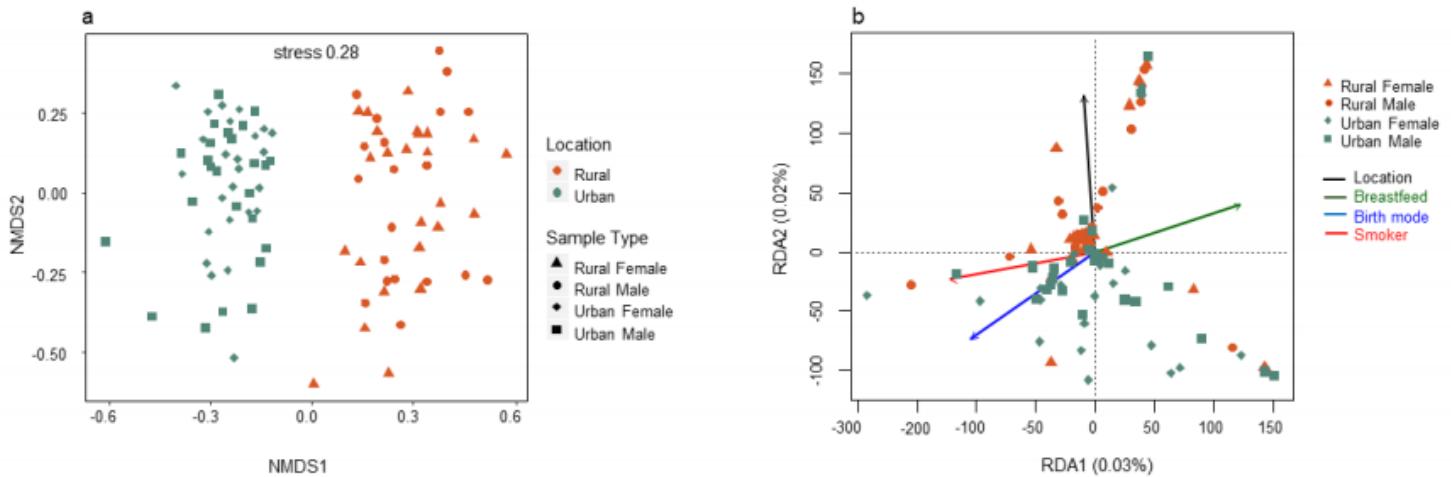


Figure 2

Overview of mycobiota structure and significant environmental drivers a) The non-metric multidimensional scaling (NMDS) plot based on Bray–Curtis dissimilarity and, b) Redundancy analysis (RDA) showing community structure in response to four selective variables. The filled shapes reflect fungal community composition in the different locations, with colours referring to location and the different explanatory variables according to the legend.

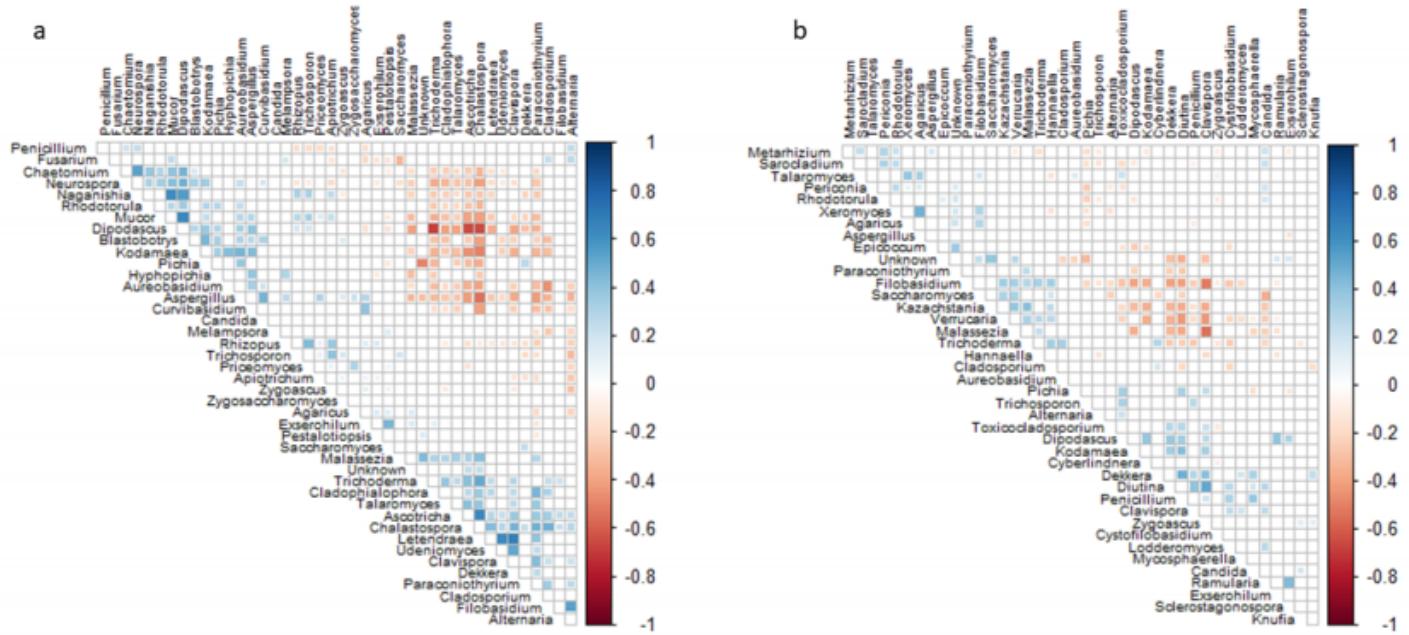


Figure 3

Correlations occurring between fungal taxa in a) rural and b) urban fungal mycobiota with $P < 0.05$ after FDR adjustment. Red squares represent significant negative correlations and blue squares represent significant positive correlations. The darker colours represent stronger correlations and non-significant correlations have been excluded from the plot.

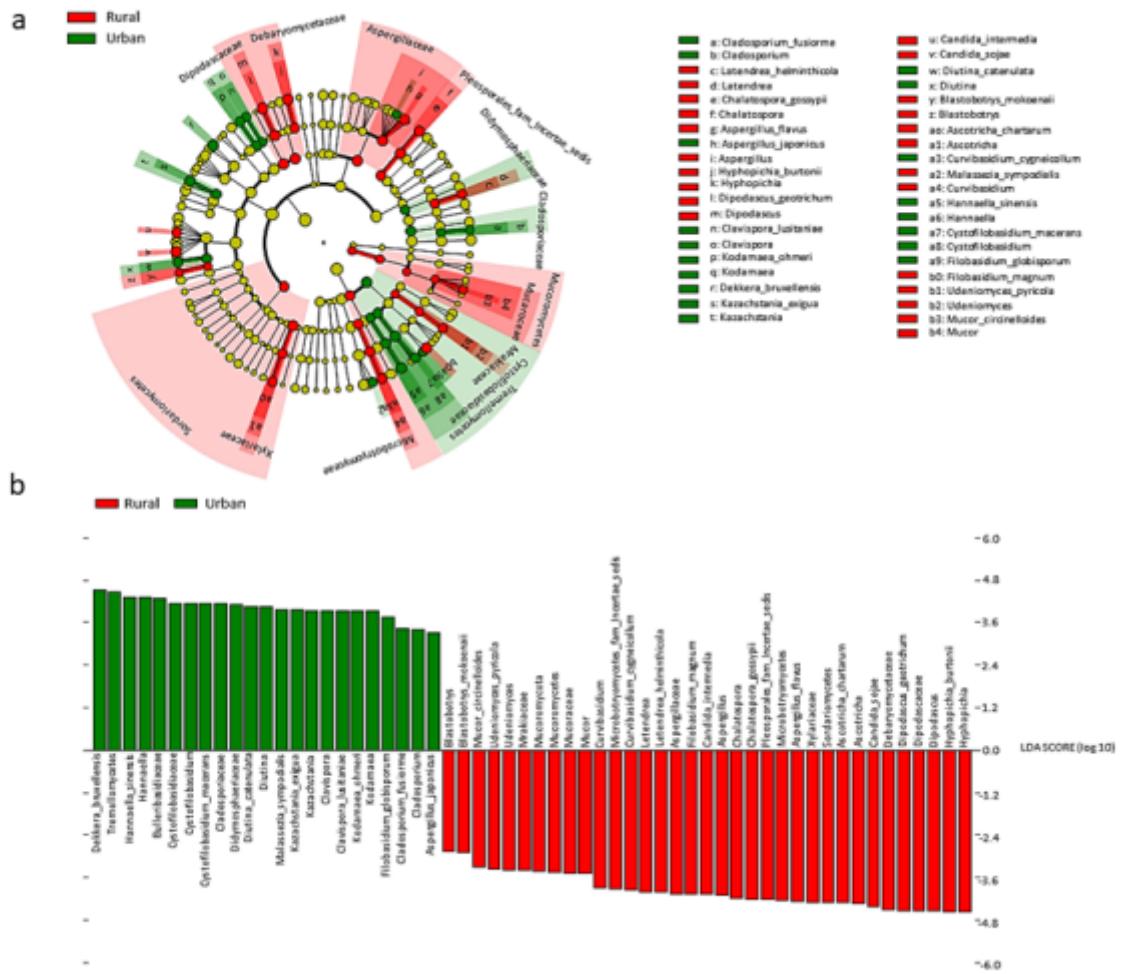


Figure 4

The results of Linear discriminant analysis (LDA) effect size (LefSe) analysis of rural and urban gut mycobiota a) The cladogram shows the output of the LEfSe algorithm, which identifies taxonomically consistent differences between rural (Ha-ravale and Tshikombani villages) and urban (Pretoria) fungal community members, respectively. Taxa with nonsignificant differences are represented as yellow circles and the diameter of the circle is proportional to relative abundances b) The histogram of the LDA scores was computed for differentially abundant taxa between the rural and urban gut mycobiota. The bar size represents the effect of the size of specific taxa in the particular group at species level

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryinformationforKabweetal.docx](#)