The gut microbiota of rural and urban individuals is shaped by geography and lifestyle

Mubanga Hellen Kabwe[1], Surendra Vikram, Khodani Mulaudzi[1], Janet K. Jansson[2] and Thulani P. Makhalanyane[1*]

[1.] Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Lynwood Road, Hatfield, Pretoria, 0028, South Africa [2.] Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, Washington 99352, USA

Running title: mycobiome of urban and rural individuals

*Address correspondence to Thulani P. Makhalanyane, thulani.makhalanyane@up.ac.za

Abstract

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. 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 several factors including geographic location and lifestyle factors such as the smoking status were major drivers of gut mycobiota. Linear discriminant and the linear discriminant analysis effect size analysis revealed several distinct urban and rural biomarkers. Together, our analysis reveals distinct community structure in urban and rural South African individuals. Geography and lifestyle related factors were shown to be key drivers of rural and urban gut microbiota.

Importance

The past decade has revealed substantial insights regarding the ecological patterns of gut microbiomes. These studies have shown clear differences between the microbiomes of individuals living in urban and rural locations. Yet, in contrast to bacteria we know substantially less regarding the fungal gut microbiota (mycobiome). Here we provide the first insights regarding the mycobiome of individuals from urban and rural locations. We show that these communities are geographically structured. Further we show that lifestyle factors, such as diet and smoking, are strong drivers explaining community variability.

Introduction

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 drivers 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 beneficial 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 cohort 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 gene 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 cohorts with fewer than 20 individuals (25, 30, 31) with very few studies investigating larger cohorts (6, 32). This study represents the first analysis of the faecal mycobiota in a large cohort 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,

87 we show that geography and lifestyle structure the gut mycobiome of rural and urban South
88 African individuals.

89

90 **Results**

91 **Similarities and differences between urban and rural individuals**

92 We assessed the faecal mycobiota of South African adults living in rural (*n*=50) and urban
93 regions (*n*= 50) by assessing stool samples (see details regarding sample recruitment in
94 Methods). We recruited an equal number of male and female volunteers across two villages in
95 the Limpopo region of South Africa, which is located roughly 500km from the urban site
96 (Pretoria) (**Figure 1a**). To gain insights regarding predictive variables which may shape the gut
97 microbiome, detailed questionnaires were distributed to all volunteers. The volunteers from Ha-
98 Ravele and Tshikombani villages (population size of roughly 200,000, representing the rural
99 participants) were on average 24 years (mean ± 6.3). Volunteers from Pretoria (total population
100 of approximately 2.1 million) were on average 31 years (mean ± 9.1). The BMI of all
101 participants was above 26.45, resulting in a cohort of individuals classified as obese, less than
102 15.9% of participants were smokers.

103 Amplicon sequence data from 95 volunteers (samples from 5 rural volunteers were excluded due
104 to low quality reads) generated 5,936,454 raw reads. A total of 5,414,023 fungal reads were
105 retained after sequence filtering and chimera removal. Of these, 1,636,180 reads were assigned
106 to OTUs and these were further clustered into 1,911 OTUs using a 97% cut-off. A higher
107 proportion of fungal OTUs were unique to location with urban and rural samples accounting for
108 47.9% and 45.3% of reads, respectively (**Supplemental Figure 1a**).

After random subsampling to the lowest read count, 155 OTUs were excluded from further analyses. The resulting accumulation curves showed reasonable sequence saturation, at a regional level (**Supplemental Figure 1b**). Fungal species richness was significantly higher in the stool samples of rural volunteers (Observed; W = 615.5, *p*-Value = 1.471e4) compared to urban volunteers (**Figure 1b**). However, we did not find significant differences in richness based on gender and age group.

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 fungal mycobiome. In total, 13 distinct fungal classes were identified with *Saccharomycetes* constituting the majority of sequences (50.1%) followed by *Dothideomycetes* (20.3%), *Eurotiomycetes* (4.8%) and *Sordariomycetes* (3.9%). Unknown fungal genera dominated our cohort (18.4%), followed by *Pichia* (17.6%) *Candida* (17.1%) and *Cladosporium* (5.9%). However, no significant difference was found between taxa abundance at the class level for the gut mycobiota of rural and urban participants (W = 1054, *p*-value = 0.5992). The difference between the gut mycobiota of rural and urban individuals, across the two locations, was not statistically significant (Kruskal-Wallis chi-squared = 2.9875, df = 3, *p*-value = 0.3936).

To assess the degree of uniqueness of a given sample in relation to the overall community composition, we assessed the local contribution to beta diversity (LCTBD). We found that samples from urban volunteers contributed a greater fraction of the overall community diversity (*p*-Value < 0.05). Samples with high local contribution to beta diversity (LCTBD) had a high

abundance of *Basidiomycota* and other unknown taxa. In contrast, only two samples from the rural location contributed to overall community diversity (**Supplemental Figure 1c**).

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 NMDS plot (**Figure 2a**). Urban and rural samples formed distict clusters [PERMANOVA (R^2 = 0.070; *p*-Value = 0.0001), ANOSIM (R = 0.43, *p*-Value = 0.001) and ADONIS (R^2 = 0.07034 *p*-Value = 0.0001)]. However, male and female samples did not cluster separately. Pairwise analysis of permutational multivariate analysis of variance (PERMANOVA) showed that there was no significant difference between female vs male urban, and female vs male rural individuals (R^2 = 0.018; R^2 = 0.023; respectively and *p*-Value > 0.4 for both). However, there was a significant difference between the gut mycobiota of urban versus rural female and male participants (R^2 < 0.074 for all; *p*-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*-Value < 0.05) drivers of community composition and structure. Predictive variables 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 cohort: between *Mucor* and *Dipodascus*, *Mucor* and *Naganishia*, *Clavispora* and *Lentendrea*, and between *Udeniomyces and Lentendrea* (**Figure 4a**). 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 4b**). 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) (33) test for biomarkers was used to detect taxa that showed the strongest effect on group differentiation (**Figure 3a**). 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 ubundant* (**Figure 3b**). 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

175 OTU clustering. These results were further corroborated by two distinct clusters, consistent with
176 rural and urban locality.

177 Several studies have investigated the healthy human mycobiome (6, 7, 19, 25, 30, 34). In these
178 studies, geography was not considered as a potential factor structuring the gut mycobiota. For
179 instance, Nash et al. (2017) found no association between host phenotypic characteristics with
180 mycobiome profile. This study suggests that diet, the environment, diurnal cycles, and host
181 genetics may substantially influence the human gut mycobiome. However, the finding that the
182 majority of the variation could not be explained by their metadata does suggest that other
183 environmental factors, such as geography, may contribute to structuring the human microbiome
184 (6).

185 Our study provides the first results showing the importance of geography in African populations.
186 Geographic locality may be associated with different environmental factors, such as different
187 climatic regimes, which may effect structural changes in the mycobiota. For example, climate
188 significantly influences vegetation and farming practices and leads to region specific diets. These
189 region-specific diets may ultimately influence the gut mycobiota. This is a reasonable prediction
190 given previous findings showing that fungi have climate dependent biogeographic patterns (35,
191 36). These patterns are likely to determine the type of fungi individuals may be exposed to,
192 which may in turn impact the colonization of fungi in the human gut. The most abundant rural-
193 associated biomarker species found in this study, *Dipodascus geotrichum,* is ubiquitous in nature
194 (37) whereas, *Hypopichia burtonii* is commonly isolated from corn, wheat, and rice (38). The
195 urban-associated biomarkers were dominated by the species *Dekkera bruxellensis,* which are
196 commonly isolated from fermented food such as wine, beer, feta cheese and sour dough (39-41).
197 In contrast, *Hannaella sinensis* is commonly isolated from plants and soil (42, 43). The staple

198 diet of the rural South African population primarily consists of a corn-based porridge (called
199 'pap'). It is therefore not uncommon for a fungal species commonly isolated from corn to be a
200 dominant biomarker for the rural population. Conversely, the urban population diet was more
201 diverse and included fermented foods such as wine, sour dough bread and feta cheese, which are
202 commonly available in supermarkets. Thus, the species *Dekkera bruxellensis* was identified as a
203 dominant biomarker in the urban population.

204 In addition to geographic location, we found that smoking, mode of birth and breastfeeding
205 significantly influenced gut fungal communities. Several studies have previously reported that
206 these factors may significantly influence the initial colonization, subsequent composition and
207 structure of bacterial members of the human gut microbiome (28, 44-46). Suhr et al. (2016) and
208 Hallen-Adams et al. (2017) investigated the gut mycobiome of two cohorts that were exclusively
209 on a vegetarian or a western diet. These studies found that the distribution of fungi differed
210 considerably between the two cohorts (7, 47). Plant-associated *Fusarium, Malassezia,*
211 *Penicillium* and *Aspergillus* species were detected at higher abundances within the vegetarian
212 cohort, compared to the cohort on a conventional diet. The finding that smoking affected fungal
213 community composition and structure is supported by several studies (48-50). The
214 approximately 4000 chemical compounds produced by cigarettes have been shown to alter the
215 composition of the gut microbiome (48, 50-53). The reported increase of *Clostridia* induced by
216 smoking in murine models has also been indirectly confirmed in humans where an increased rate
217 of *C. difficile* infection was greater in former and current smokers compared to never smokers
218 (52). Moreover, the abundance of the fungus *Candida tropicalis* has also been reported to be
219 significantly higher in *C. difficile* infection patients compared to healthy individuals. (54). The
220 abundance of *C. tropicalis* has also been detected to be positively correlated with levels of anti-

Saccharomyces cerevisiae antibodies (ASCA) (54). 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).

Most studies have identified the genera *Candida, Saccharomyces*, *Malassezia* and *Aspergillus* as the three most abundant in the gut of healthy individuals (6, 25, 32). To the best of our knowledge, our study is the first to report *Pichia* as one of the top four 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 (34). 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*.

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 (34, 47, 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, 46). 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 taxa n 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, 34) 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. These results have significant health implications, particularly for immunocompromised individuals living in rural and urban locations. Such findings provide a valid basis for the development of novel therapeutics or preventative measures reliant on modulating the gut mycobiome.

Methods

Ethical clearance

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. All experimental methods and experiments were in accordance with the Helsinki Declaration.

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-Ravele and Tshikombani 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 (Figure 1a). 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)]. 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 in the sample collection sheets 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

311 approximately 1.2 mL of binding buffer was added. Next, 70% ethanol (500 µL) was added and
312 the contents precipitated by centrifugation at room temperature for 60 s at 10,000 x *g*. The DNA
313 was eluted with 100 µL filter-sterilised autoclaved Millipore water and quantified using the
314 NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific, Waltham, MA, USA). The
315 quality of isolated DNA was confirmed by agarose gel electrophoresis, on 1% (w/v) agarose gel
316 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
317 Volts for 45 min. in a BioRad Sub-Cell® GT gel electrophoresis system with gel red visualising
318 agent. The gel was visualised using the BioRad Gel Doc system and viewed with a UV Trans-
319 illuminator.

320 **ITS gene region amplification, sequencing and data processing**

321 The internal transcribed spacer (ITS) region was amplified using fungal-specific primers (60):
322 ITS1F (5‘ -CTTGGTCATTTAGAGGAAGTAA-3‘) and ITS4 (5‘-
323 TCCTCCGCTTATTGATATGC-3‘) and the HotStarTaq Plus Master Kit (Qiagen, Valencia,
324 CA). Amplicons from different samples were pooled to equimolar concentrations and purified of
325 short fragments using Agencourt Ampure beads (Agencourt Bioscience Corporation, USA).
326 Paired-end 2 x 250bp sequencing was performed on an Illumina MiSeq instrument (Illumina
327 Inc., San Diego, CA, USA) at Mr DNA (Shallowater, TX 79363).

328 The resultant data were analysed using the Quantitative Insights into Microbial Ecology
329 (QIIME2) software version 2018.8.0 (71). Demultiplexed sequences were assessed for quality
330 and those shorter than 200 bp, with quality scores below 25, containing more than two
331 ambiguous characters or more than one mismatch to the sample-specific barcode or the primer
332 sequences, were excluded from further downstream analyses. Chimeric sequence detection and
333 operation taxonomic unit (OTU) selection was performed at 97% sequence similarity using

USEARCH v11 (72). Taxonomies were assigned to each OTU using the UNITE (release 7_99) databases for fungi (73). 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 (74, 75). Alpha diversity indices (observed, Chao1, Shannon, and Simpson indices), together with rarefaction curves were calculated and visualized using the R packages "phyloseq" and "ggplot". First, the Shapiro-Wilk's test was applied to determine data distribution (76). Subsequently, the unpaired two-sample Wilcoxon rank sum test (77, 78) 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. (79, 80). In these analyses, the rural or urban location was specified as a random factor.

The R packages "phyloseq" (81) and "microbiomeseq" (82) 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 in the urban and rural samples. Whereas, the Kruskal-Wallis test (83) was applied to determine significant differences between taxa in the four sample types.

The local contribution to beta diversity (LCBD) was calculated according to (84). 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 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 (85), a permutational multivariate analysis of variance (PERMANOVA) (86) 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 (87, 88).

The effect of the different recorded environmental factors on fungal community composition and structure was determined through redundancy analysis (RDA). First, the OTU-count data were Hellinger-transformed. The contribution of highly correlating OTUs (p-Value < 0.05) with redundancy axes was identified using the envfit function from the R package vegan (85). To measure the relationship of abundant taxa with measured anthropometric factors (age, BMI, height, and weight), a Spearman correlation analysis was done and visualized in the R package microbiomeseq (82).

Fungal-fungal relationships were interrogated using SparCC (89). 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 (90) and the correlation matrix was visualized using the “corrplot” function (91) 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) (33). The LEfSe was calculated using the online Galaxy web application (92) 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.

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.

Acknowledgements

399 This research was supported by the South African Medical Research Foundation (TPM). MHK
400 and KM received scholarships from the National Research Foundation. We wish to thank the
401 Fulbright program for providing sabbatical support to TPM, JJ was funded by the Microbiomes
402 in Transition (MinT) initiative at the Pacific Northwest National Laboratory in Richland, WA,
403 USA. PNNL is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-
404 76RL01830. We would like to thank all the volunteers for donating their data, time and samples
405 to our research.

406

Figure captions

Figure 1 Geographic locations and diversity estimates **a**) The three sampling locations in Gauteng (Pretoria) and Limpopo (Ha-ravele 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 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-ravele 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

Figure 4 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.

Supplemental Figure legend

Figure S1 Comparison of mycobiota between urban and rural participants a) Venn diagram showing the unique and shared phylotypes for samples collected from urban and rural participants. b) Rarefaction plot showing sequencing coverage. c) 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.

References

1. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences 103:12115-12120.
2. Richard ML, Sokol H. 2019. The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. Nature Reviews Gastroenterology & Hepatology 16:331-345.
3. Underhill DM, Iliev ID. 2014. The mycobiota: interactions between commensal fungi and the host immune system. Nature Reviews Immunology 14:405.
4. Oever Jt, Netea MG. 2014. The bacteriome–mycobiome interaction and antifungal host defense. European journal of immunology 44:3182-3191.
5. Huffnagle GB, Noverr MC. 2013. The emerging world of the fungal microbiome. Trends in microbiology 21:334-341.
6. Nash AK, Auchtung TA, Wong MC, Smith DP, Gesell JR, Ross MC, Stewart CJ, Metcalf GA, Muzny DM, Gibbs RA. 2017. The gut mycobiome of the Human Microbiome Project healthy cohort. Microbiome 5:153.
7. Hallen-Adams HE, Suhr MJ. 2017. Fungi in the healthy human gastrointestinal tract. Virulence 8:352-358.
8. Rizzetto L, De Filippo C, Cavalieri D. 2014. Richness and diversity of mammalian fungal communities shape innate and adaptive immunity in health and disease. European journal of immunology 44:3166-3181.
9. Romani L. 2011. Immunity to fungal infections. Nature Reviews Immunology 11:275.
10. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosh D, Panzer AR, LaMere B, Rackaityte E, Lukacs NW. 2016. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nature medicine 22:1187.
11. Ott SJ, Kühbacher T, Musfeldt M, Rosenstiel P, Hellmig S, Rehman A, Drews O, Weichert W, Timmis KN, Schreiber S. 2008. Fungi and inflammatory bowel diseases: alterations of composition and diversity. Scandinavian journal of gastroenterology 43:831-841.
12. Sokol H, Leducq V, Aschard H, Pham H-P, Jegou S, Landman C, Cohen D, Liguori G, Bourrier A, Nion-Larmurier I. 2017. Fungal microbiota dysbiosis in IBD. Gut 66:1039-1048.
13. Iliev ID, Funari VA, Taylor KD, Nguyen Q, Reyes CN, Strom SP, Brown J, Becker CA, Fleshner PR, Dubinsky M. 2012. Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. Science 336:1314-1317.
14. Luan C, Xie L, Yang X, Miao H, Lv N, Zhang R, Xiao X, Hu Y, Liu Y, Wu N. 2015. Dysbiosis of fungal microbiota in the intestinal mucosa of patients with colorectal adenomas. Scientific reports 5:7980.
15. Rodríguez MM, Pérez D, Chaves FJ, Esteve E, Marin-Garcia P, Xifra G, Vendrell J, Jové M, Pamplona R, Ricart W. 2015. Obesity changes the human gut mycobiome. Scientific reports 5:14600.
16. Gewolb IH, Schwalbe RS, Taciak VL, Harrison TS, Panigrahi P. 1999. Stool microflora in extremely low birthweight infants. Archives of Disease in Childhood-Fetal and Neonatal Edition 80:F167-F173.
17. Strati F, Di Paola M, Stefanini I, Albanese D, Rizzetto L, Lionetti P, Calabrò A, Jousson O, Donati C, Cavalieri D. 2016. Age and gender affect the composition of fungal population of the human gastrointestinal tract. Frontiers in microbiology 7:1227.

493 18. Heisel T, Podgorski H, Staley CM, Knights D, Sadowsky MJ, Gale CA. 2015. Complementary
494 amplicon-based genomic approaches for the study of fungal communities in humans. PLoS One
495 10:e0116705.
496 19. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD. 2013. Archaea
497 and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One
498 8:e66019.
499 20. Gouba N, Raoult D, Drancourt M. 2013. Plant and fungal diversity in gut microbiota as revealed
500 by molecular and culture investigations. PloS one 8:e59474.
501 21. Kowalewska B, Zorena K, Szmigiero-Kawko M, Wąż P, Myśliwiec M. 2016. Higher diversity in
502 fungal species discriminates children with type 1 diabetes mellitus from healthy control. Patient
503 preference and adherence 10:591.
504 22. Gouba N, Raoult D, Drancourt M. 2014. Gut microeukaryotes during anorexia nervosa: a case
505 report. BMC research notes 7:33.
506 23. Zhang E, Tanaka T, Tajima M, Tsuboi R, Nishikawa A, Sugita T. 2011. Characterization of the skin
507 fungal microbiota in patients with atopic dermatitis and in healthy subjects. Microbiology and
508 immunology 55:625-632.
509 24. Nguyen LD, Viscogliosi E, Delhaes L. 2015. The lung mycobiome: an emerging field of the human
510 respiratory microbiome. Frontiers in microbiology 6:89.
511 25. Hamad I, Sokhna C, Raoult D, Bittar F. 2012. Molecular detection of eukaryotes in a single
512 human stool sample from Senegal. PloS one 7:e40888.
513 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.
514 2008. Bacterial, archaeal, and eukaryal diversity in the intestines of Korean people. The Journal
515 of Microbiology 46:491-501.
516 27. Hamad I, Raoult D, Bittar F. 2016. Repertory of eukaryotes (eukaryome) in the human
517 gastrointestinal tract: taxonomy and detection methods. Parasite immunology 38:12-36.
518 28. Xu Z, Knight R. 2015. Dietary effects on human gut microbiome diversity. Br J Nutr 113 Suppl:S1-
519 5.
520 29. Wang WL, Xu SY, Ren ZG, Tao L, Jiang JW, Zheng SS. 2015. Application of metagenomics in the
521 human gut microbiome. World J Gastroenterol 21:803-14.
522 30. Suhr MJ, Banjara N, Hallen-Adams HE. 2016. Sequence-based methods for detecting and
523 evaluating the human gut mycobiome. Letters in applied microbiology 62:209-215.
524 31. Raimondi S, Amaretti A, Gozzoli C, Simone M, Righini L, Candeliere F, Brun P, Ardizzoni A,
525 Colombari B, Paulone S. 2019. Longitudinal survey of fungi in the human gut: ITS profiling,
526 phenotyping and colonization. Frontiers in microbiology 10:1575.
527 32. Hallen-Adams HE. 2017. Fungi in the healthy human gastrointestinal tract. Virulence 8:352.
528 33. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011.
529 Metagenomic biomarker discovery and explanation. Genome biology 12:R60.
530 34. Hallen-Adams HE. 2015. Fungi inhabiting the healthy human gastrointestinal tract: a diverse and
531 dynamic community. Fungal Ecology 15:9.
532 35. Wang D, Rui Y, Ding K, Cui X, Hao Y, Tang L, Pang Z, Zhang B, Zhou S, Wang K. 2018. Precipitation
533 drives the biogeographic distribution of soil fungal community in Inner Mongolian temperate
534 grasslands. Journal of soils and sediments 18:222-228.
535 36. Summerell BA, Laurence MH, Liew EC, Leslie JF. 2010. Biogeography and phylogeography of
536 Fusarium: a review. Fungal Diversity 44:3-13.
537 37. Kolecka A, Khayhan K, Groenewald M, Theelen B, Arabatzis M, Velegraki A, Kostrzewa M, Mares
538 M, Taj-Aldeen SJ, Boekhout T. 2013. Identification of medically relevant species of arthroconidial
539 yeasts by use of matrix-assisted laser desorption ionization–time of flight mass spectrometry.
540 Journal of clinical microbiology 51:2491-2500.

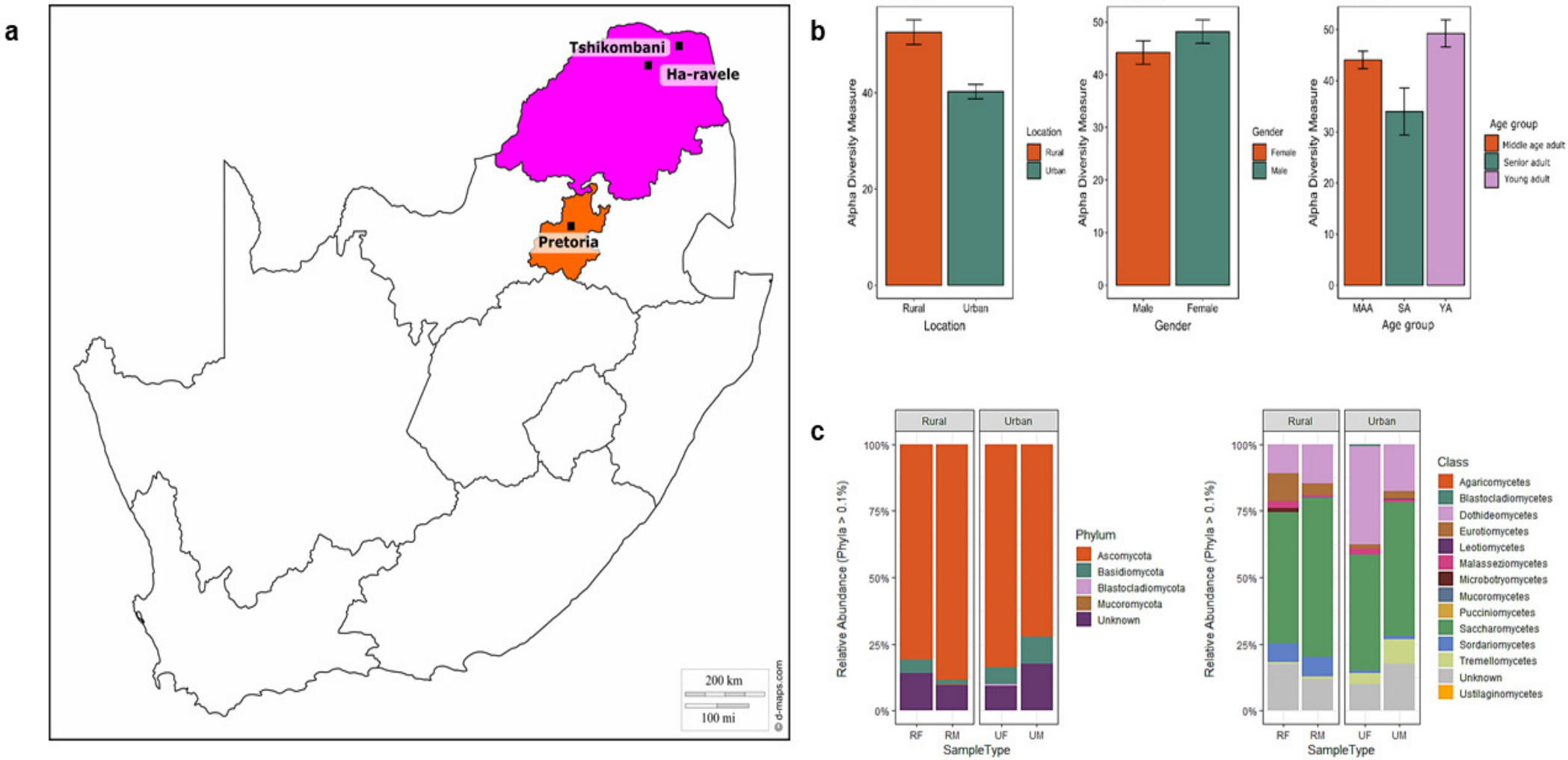
38. Kurtzman CP. 2011. Hyphopichia von Arx & van der Walt (1976), p 435-438, The Yeasts. Elsevier.
39. Schifferdecker AJ, Dashko S, Ishchuk OP, Piškur J. 2014. The wine and beer yeast Dekkera bruxellensis. Yeast 31:323-332.
40. Fadda ME, Cosentino S, Deplano M, Palmas F. 2001. Yeast populations in Sardinian feta cheese. International Journal of Food Microbiology 69:153-156.
41. Meroth CB, Hammes WP, Hertel C. 2003. Identification and population dynamics of yeasts in sourdough fermentation processes by PCR-denaturing gradient gel electrophoresis. Appl Environ Microbiol 69:7453-7461.
42. Into P, Pontes A, Sampaio JP, Limtong S. 2020. Yeast Diversity Associated with the Phylloplane of Corn Plants Cultivated in Thailand. Microorganisms 8:80.
43. Landell M, Brandão L, Barbosa A, Ramos J, Safar S, Gomes F, Sousa F, Morais P, Broetto L, Leoncini O, Ribeiro J, Fungsin B, Takashima M, Nakase T, Vainstein M, Fell J, Scorzetti G, Vishniac H, Valente P. 2014. Hannaella pagnoccae sp. nov., a tremellaceous yeast species isolated from plants and soil. International journal of systematic and evolutionary microbiology 64.
44. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP. 2012. Human gut microbiome viewed across age and geography. nature 486:222.
45. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proceedings of the National Academy of Sciences 107:11971-11975.
46. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105-108.
47. Suhr MJ. 2016. Sequencebased methods for detecting and evaluating the human gut mycobiome. Letters in Applied Microbiology 62:209.
48. Capurso G, Lahner E. 2017. The interaction between smoking, alcohol and the gut microbiome. Best practice & research Clinical gastroenterology 31:579-588.
49. Biedermann L, Brülisauer K, Zeitz J, Frei P, Scharl M, Vavricka SR, Fried M, Loessner MJ, Rogler G, Schuppler M. 2014. Smoking cessation alters intestinal microbiota: insights from quantitative investigations on human fecal samples using FISH. Inflammatory bowel diseases 20:1496-1501.
50. Allais L, Kerckhof FM, Verschuere S, Bracke KR, De Smet R, Laukens D, Van den Abbeele P, De Vos M, Boon N, Brusselle GG. 2016. Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. Environmental microbiology 18:1352-1363.
51. Biedermann L, Zeitz J, Mwinyi J, Sutter-Minder E, Rehman A, Ott SJ, Steurer-Stey C, Frei A, Frei P, Scharl M. 2013. Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. PloS one 8:e59260.
52. Rogers MA, Greene MT, Saint S, Chenoweth CE, Malani PN, Trivedi I, Aronoff DM. 2012. Higher rates of Clostridium difficile infection among smokers. Plos one 7:e42091.
53. Tomoda K, Kubo K, Asahara T, Andoh A, Nomoto K, Nishii Y, Yamamoto Y, Yoshikawa M, Kimura H. 2011. Cigarette smoke decreases organic acids levels and population of bifidobacterium in the caecum of rats. The Journal of toxicological sciences 36:261-266.
54. Hoarau G, Mukherjee P, Gower-Rousseau C, Hager C, Chandra J, Retuerto M, Neut C, Vermeire S, Clemente J, Colombel J-F. 2016. Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. MBio 7:e01250-16.
55. Dupuy AK, David MS, Li L, Heider TN, Peterson JD, Montano EA, Dongari-Bagtzoglou A, Diaz PI, Strausbaugh LD. 2014. Redefining the human oral mycobiome with improved practices in

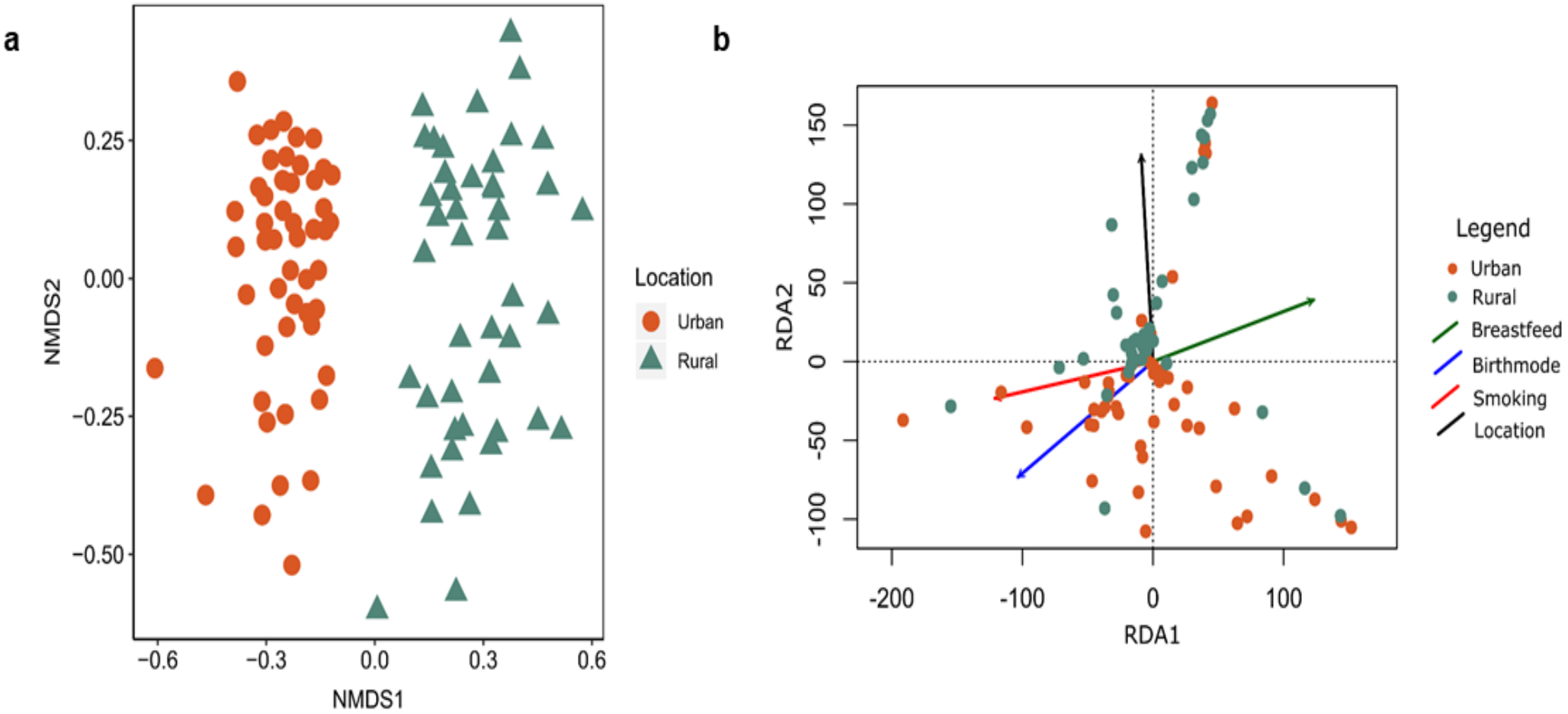
amplicon-based taxonomy: discovery of Malassezia as a prominent commensal. PLoS One 9:e90899.

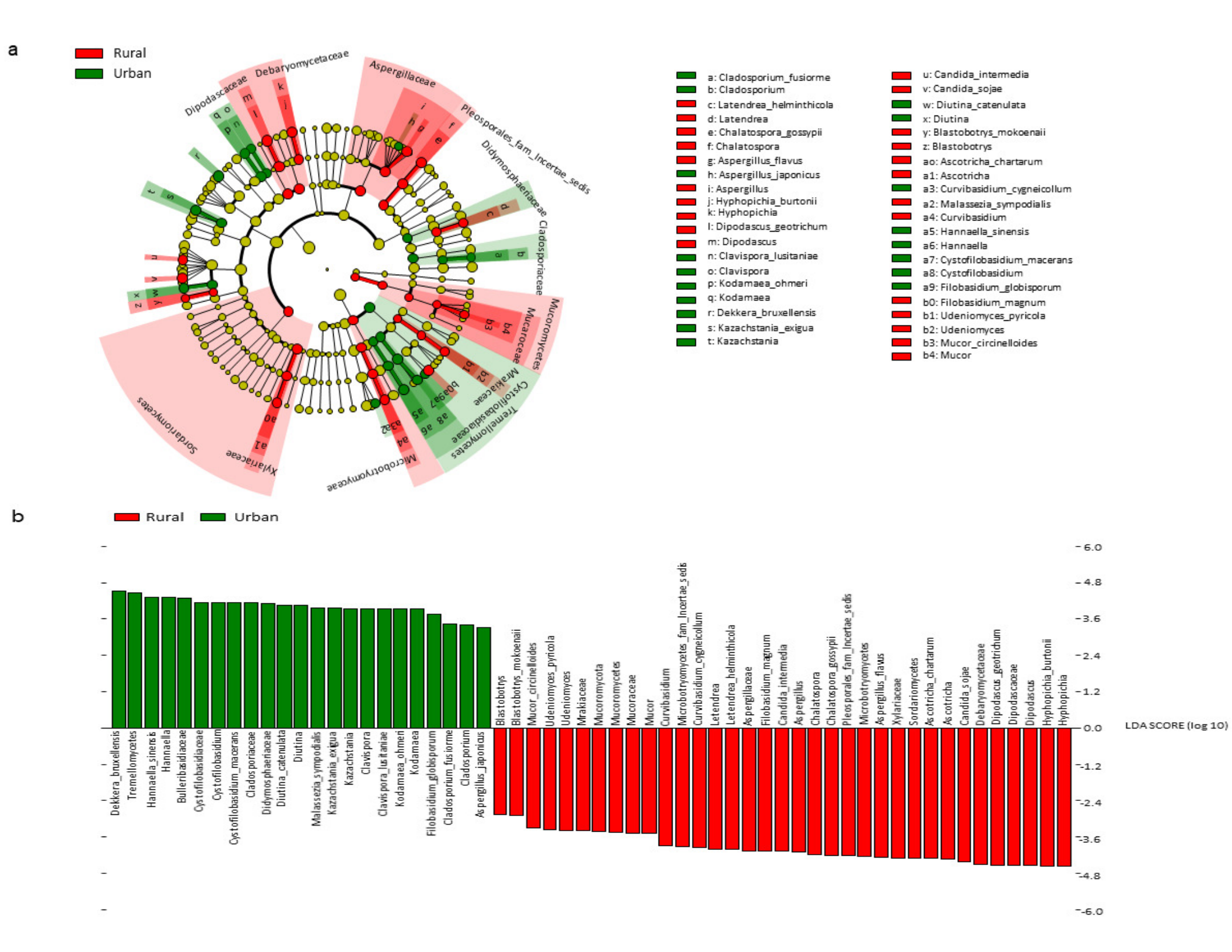
56. Mukherjee PK, Chandra J, Retuerto M, Sikaroodi M, Brown RE, Jurevic R, Salata RA, Lederman MM, Gillevet PM, Ghannoum MA. 2014. Oral mycobiome analysis of HIV-infected patients: identification of Pichia as an antagonist of opportunistic fungi. PLoS pathogens 10:e1003996.
57. Rodríguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, Avershina E, Rudi K, Narbad A, Jenmalm MC. 2015. The composition of the gut microbiota throughout life, with an emphasis on early life. Microbial ecology in health and disease 26:26050.
58. Strati F, Cavalieri D, Albanese D, De Felice C, Donati C, Hayek J, Jousson O, Leoncini S, Pindo M, Renzi D. 2016. Altered gut microbiota in Rett syndrome. Microbiome 4:41.
59. Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. Molecular ecology 2:113-118.
60. White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications 18:315-322.
61. Raimondi S, Amaretti A, Gozzoli C, Simone M, Righini L, Candeliere F, Brun P, Ardizzoni A, Colombari B, Paulone S, Castagliuolo I, Cavalieri D, Blasi E, Rossi M, Peppoloni S. 2019. Longitudinal Survey of Fungi in the Human Gut: ITS Profiling, Phenotyping, and Colonization. Frontiers in Microbiology 10.
62. Calderone RA, Clancy CJ. 2011. Candida and candidiasis. American Society for Microbiology Press.
63. Silva S, Negri M, Henriques M, Oliveira R, Williams DW, Azeredo J. 2012. Candida glabrata, Candida parapsilosis and Candida tropicalis: biology, epidemiology, pathogenicity and antifungal resistance. FEMS microbiology reviews 36:288-305.
64. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM. 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. PLoS pathogens 6:e1000713.
65. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspõllu A, Väin E, Saarma I, Salumets A, Donders GG, Metsis M. 2013. Characterization of the vaginal micro-and mycobiome in asymptomatic reproductive-age Estonian women. PloS one 8:e54379.
66. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, Sequencing NISCC. 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature 498:367.
67. Leung MH, Chan KC, Lee PK. 2016. Skin fungal community and its correlation with bacterial community of urban Chinese individuals. Microbiome 4:46.
68. Moran G, Coleman D, Sullivan D. 2012. An introduction to the medically important Candida species, p 11-25, Candida and Candidiasis, Second Edition. American Society of Microbiology.
69. Saleh HA, Moawad AA, El-Hariri M, Refai MK. 2011. Prevalence of yeasts in human, animals and soil sample at El-Fayoum Governorate in Egypt. Int J Microbiol Res 2:233-239.
70. Vogel C, Rogerson A, Schatz S, Laubach H, Tallman A, Fell J. 2007. Prevalence of yeasts in beach sand at three bathing beaches in South Florida. Water research 41:1915-1920.
71. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F. 2018. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints,
72. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.
73. Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytologist 166:1063-1068.

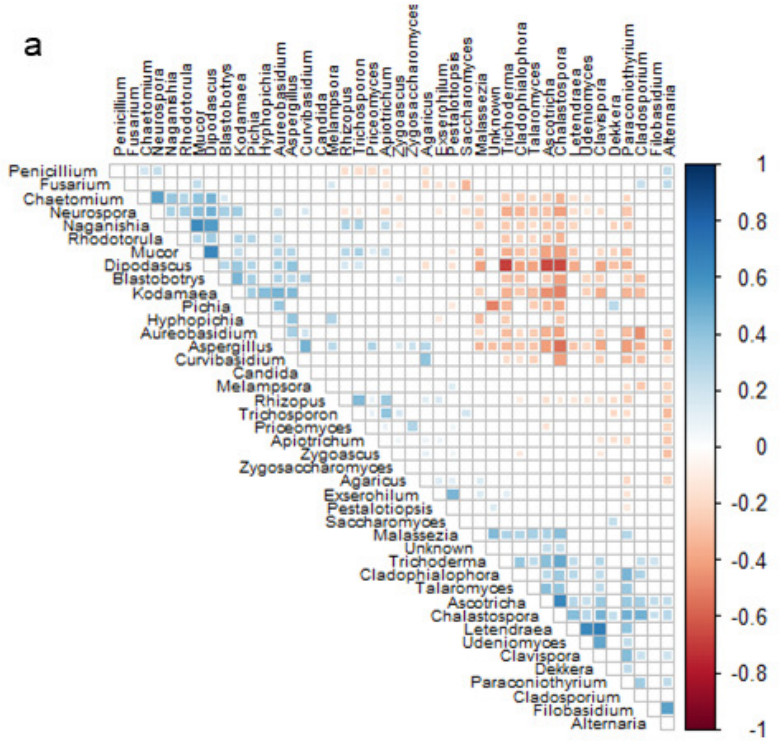
636 74. RDevelopmentCoreTeam. 2018. R: A Language and Environment for Statistical Computing: A
637 Graduate Course in Probability. R Foundation for Statistical Computing, Vienna, Austria.
638 75. Wickham H, Wickham MH. 2007. The ggplot package.
639 76. Villasenor Alva JA, Estrada EG. 2009. A generalization of Shapiro–Wilk's test for multivariate
640 normality. Communications in Statistics—Theory and Methods 38:1870-1883.
641 77. Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically
642 larger than the other. The annals of mathematical statistics:50-60.
643 78. Wilcoxon F. 1945. Individual comparisons by ranking methods. Biometrics bulletin 1:80-83.
644 79. Wickham H, Francois R, Henry L, Müller K. 2015. dplyr: A Grammar of Data Manipulation. R
645 package version 0.4. 3. R Found Stat Comput, Vienna https://CRAN R-project org/package=
646 dplyr.
647 80. Kassambara A. 2018. ggpubr:'ggplot2'Based Publication Ready Plots. R package version 0.1. 8.
648 81. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and
649 graphics of microbiome census data. PloS one 8:e61217.
650 82. Ssekagiri A, Sloan W, Ijaz U. microbiomeSeq: an R package for microbial community analysis in
651 an environmental context, p. *In* (ed),
652 83. Breslow N. 1970. A generalized Kruskal-Wallis test for comparing K samples subject to unequal
653 patterns of censorship. Biometrika 57:579-594.
654 84. Legendre P, De Cáceres M. 2013. Beta diversity as the variance of community data: dissimilarity
655 coefficients and partitioning. Ecology letters 16:951-963.
656 85. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'hara R, Simpson GL, Solymos P, Stevens MHH,
657 Wagner H. 2010. Vegan: community ecology package. R package version 1.17-4. http://cran r-
658 project org> Acesso em 23:2010.
659 86. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral
660 ecology 26:32-46.
661 87. Gray J, Aschan M, Carr M, Clarke K, Green R, Pearson T, Rosenberg R, Warwick R. 1988. Analysis
662 of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a
663 mesocosm experiment. Marine Ecology Progress Series 46:151-165.
664 88. Anderson MJ, Walsh DC. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of
665 heterogeneous dispersions: what null hypothesis are you testing? Ecological monographs
666 83:557-574.
667 89. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS
668 computational biology 8:e1002687.
669 90. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful
670 approach to multiple testing. Journal of the Royal statistical society: series B (Methodological)
671 57:289-300.
672 91. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. 2017. Package 'corrplot'. Statistician 56:316-324.
673 92. Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J.
674 2010. Galaxy: a web-based genome analysis tool for experimentalists. Current protocols in
675 molecular biology 89:19.10. 1-19.10. 21.

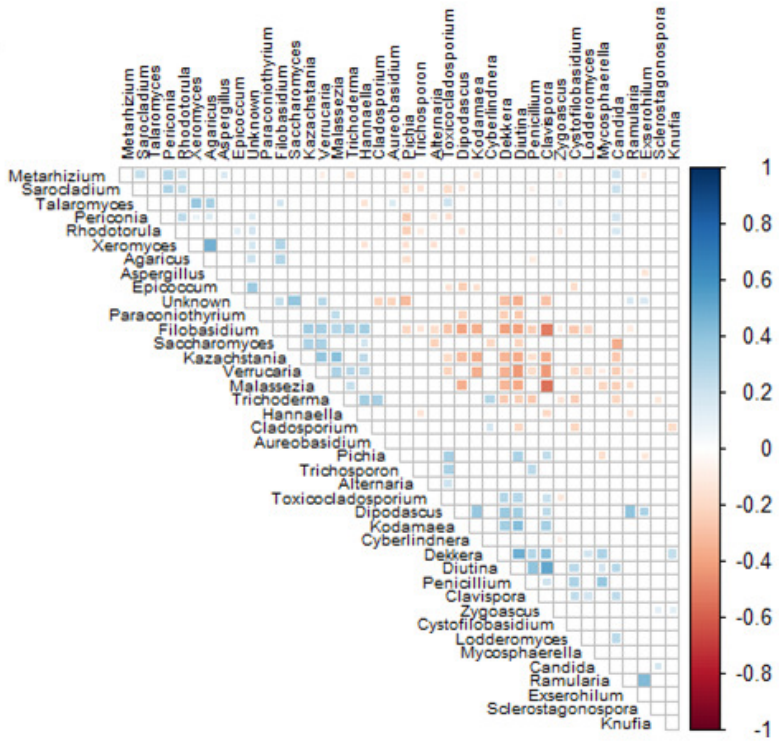
676


a
Tshikombani
Ha-ravele
Pretoria
200 km
100 mi
© d-maps.com
b
Observed
Alpha Diversity Measure
Location
Rural
Urban
Gender
Female
Male
Age group
Middle age adult
Senior adult
Young adult
MAA
SA
YA
c
Rural
Urban
Relative Abundance (Phyla > 0.1%)
SampleType
RF
RM
UF
UM
Phylum
Ascomycota
Basidiomycota
Blastocladiomycota
Mucoromycota
Unknown
Class
Agaricomycetes
Blastocladiomycetes
Dothideomycetes
Eurotiomycetes
Leotiomycetes
Malasseziomycetes
Microbotryomycetes
Mucoromycetes
Pucciniomycetes
Saccharomycetes
Sordariomycetes
Tremellomycetes
Unknown
Ustilaginomycetes


a
0.25
0.00
−0.25
−0.50
NMDS2
−0.6
−0.3
0.0
0.3
0.6
NMDS1
Location
Urban
Rural
b
150
100
50
0
-50
-100
RDA2
-200
-100
0
100
RDA1
Legend
Urban
Rural
Breastfeed
Birthmode
Smoking
Location


a
Rural
Urban
Dipodascaceae
Debaryomycetaceae
Aspergillaceae
Pleosporales_fam_Incertae_sedis
Didymosphaeriaceae
Cladosporiaceae
Mucoromycetes
Mucoraceae
Mrakiaceae
Cystofiliobasidiaceae
Tremellomycetes
Microbotryomycetes
Xylariaceae
Sordariomycetes
a: Cladosporium_fusiorme
b: Cladosporium
c: Latendrea_helminthicola
d: Latendrea
e: Chalatospora_gossypii
f: Chalatospora
g: Aspergillus_flavus
h: Aspergillus_japonicus
i: Aspergillus
j: Hyphopichia_burtonii
k: Hyphopichia
l: Dipodascus_geotrichum
m: Dipodascus
n: Clavispora_lusitaniae
o: Clavispora
p: Kodamaea_ohmeri
q: Kodamaea
r: Dekkera_bruxellensis
s: Kazachstania_exigua
t: Kazachstania
u: Candida_intermedia
v: Candida_sojae
w: Diutina_catenulata
x: Diutina
y: Blastobotrys_mokoenaii
z: Blastobotrys
ao: Ascotricha_chartarum
a1: Ascotricha
a3: Curvibasidium_cygneicollum
a2: Malassezia_sympodialis
a4: Curvibasidium
a5: Hannaella_sinensis
a6: Hannaella
a7: Cystofilobasidium_macerans
a8: Cystofilobasidium
a9: Filobasidium_globisporum
b0: Filobasidium_magnum
b1: Udeniomyces_pyricola
b2: Udeniomyces
b3: Mucor_circinelloides
b4: Mucor
b
Rural
Urban
Dekkera_bruxellensis
Tremellomycetes
Hannaella_sinensis
Hannaella
Bulleribasidiaceae
Cystofilobasidiaceae
Cystofilobasidium
Cystofilobasidium_macerans
Cladosporiaceae
Didymosphaeriaceae
Diutina_catenulata
Diutina
Malassezia_sympodialis
Kazachstania_exigua
Kazachstania
Clavispora
Clavispora_lusitaniae
Kodamaea_ohmeri
Kodamaea
Filobasidium_globisporum
Cladosporium_fusiorme
Cladosporium
Aspergillus_japonicus
Blastobotrys
Blastobotrys_mokoenaii
Mucor_circinelloides
Udeniomyces_pyricola
Udeniomyces
Mrakiaceae
Mucoromycota
Mucoromycetes
Mucoraceae
Mucor
Curvibasidium
Microbotryomycetes_fam_Incertae_sedis
Curvibasidium_cygneicollum
Letendrea
Letendrea_helminthicola
Aspergillaceae
Filobasidium_magnum
Candida_intermedia
Aspergillus
Chalatospora
Chalatospora_gossypii
Pleosporales_fam_Incertae_sedis
Microbotryomycetes
Aspergillus_flavus
Xylariaceae
Sordariomycetes
Ascotricha_chartarum
Ascotricha
Candida_sojae
Debaryomycetaceae
Dipodascus_geotrichum
Dipodascaceae
Dipodascus
Hyphopichia_burtonii
Hyphopichia
6.0
4.8
3.6
2.4
1.2
0.0
-1.2
-2.4
-3.6
-4.8
-6.0
LDA SCORE (log 10)


a
Penicillium
Fusarium
Chaetomium
Neurospora
Naganishia
Rhodotorula
Mucor
Dipodascus
Blastobotrys
Kodamaea
Pichia
Hyphopichia
Aureobasidium
Aspergillus
Curvibasidium
Candida
Melampsora
Rhizopus
Trichosporon
Priceomyces
Apiotrichum
Zygoascus
Zygosaccharomyces
Agaricus
Exserohilum
Pestalotiopsis
Saccharomyces
Malassezia
Unknown
Trichoderma
Cladophialophora
Talaromyces
Ascotricha
Chalastospora
Letendraea
Udeniomyces
Clavispora
Dekkera
Paraconiothyrium
Cladosporium
Filobasidium
Alternaria
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


b
Metarhizium
Sarocladium
Talaromyces
Periconia
Rhodotorula
Xeromyces
Agaricus
Aspergillus
Epicoccum
Unknown
Paraconiothyrium
Filobasidium
Saccharomyces
Kazachstania
Verrucaria
Malassezia
Trichoderma
Hannaella
Cladosporium
Aureobasidium
Pichia
Trichosporon
Alternaria
Toxicocladosporium
Dipodascus
Kodamaea
Cyberlindnera
Dekkera
Diutina
Penicillium
Clavispora
Zygoascus
Cystofilobasidium
Lodderomyces
Mycosphaerella
Candida
Ramularia
Exserohilum
Sclerostagonospora
Knufia
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1