Supplementary Materials

Mapping the human gut mycobiome in middle-aged and elderly adults: multi-omics insights and implications for host metabolic health

This word file includes:

Methods

Study participants and stool sample collection

The current study was based on the Guangzhou Nutrition and Health Study (GNHS) cohort that involved 4048 participants aged 45-70 years at recruitment between 2008-2013. All participants had lived in Guangzhou city (China) for at least 5 years and were followed up approximately every 3 years (Fig. S1). Detailed information on the study design has been reported previously[1]. The study was registered at clinicaltrials.gov (NCT03179657). In the present study, 1244 participants with ITS2 sequencing data were included in the discovery cohort (Table S1). To validate the interaction analysis, we used the control arm of a case-control study of hip fracture with ITS2 and 16S rRNA sequencing data (n = 94) as an independent replication cohort (Table S15). The study design of the replication cohort has previously been described in detail[2]. Stool samples of 1244 participants were collected during on-site study visits between 2014-2018. Among these participants, 184 had stool samples collected twice at different follow-up visits with a median follow-up time of 3.2 years.

Stool samples were collected on-site at Sun Yat-sen University, temporarily stored in an ice box, manually stirred and then dispensed into centrifuge tubes and stored in a refrigerator at -80°C within 4 hours. After each sample was collected, we immediately checked whether the sample was sufficient and whether the collection date was recorded. The stool samples were stored at -80 °C before further processing. The study protocol was approved by the Ethics Committee of the School of Public Health

at Sun Yat-sen University and Ethics Committee of Westlake University. All participants provided written informed consent.

Phenotype data

Phenotypic data assessed in the present study included demographics (e.g., age, gender, BMI), 14 physiologic traits, 16 dietary intakes and 3 diseases (e.g., obesity, type 2 diabetes, metabolic syndrome). Information on demographics, lifestyle and physical activity was collected by questionnaires. Habitual dietary intakes at baseline were estimated from a validated food frequency questionnaire (FFQ), which recorded the frequencies of foods consumed in the past 12 months. Food intake was divided into 16 food groups according to the Guidelines for Measuring Household and Individual Dietary Diversity. Anthropometrics data were obtained by the measurements of trained staff on site. Fasting venous blood samples were collected at both baseline and follow-up visits. Insulin was measured using the electrochemiluminescence immunoassay (ECLIA, Roche cobas 8000 e602) method. The homeostatic model assessment of insulin resistance (HOMA-IR) was calculated as fasting blood glucose (FBG) (mmol/L) times fasting insulin (mIU/L) divided by 22.5. The measurements of the other blood tests (e.g., FBG, triglycerides, total cholesterol) had been described elsewhere[3].

Obesity was defined as a BMI ≥ 28 kg/m^2. Type 2 diabetes (T2D) was defined in adults as meeting one of the criteria: FBG ≥ 7.0 mmol/L, glycated hemoglobin

(HbA1c) ≥ 6.5% or self-reported medical treatment for diabetes. Metabolic syndrome was defined as meeting three of the five following items: (i) waist circumference > 90 cm (male) or waist circumference > 85 cm (female), (ii) FBG ≥ 6.1 mmol/L (110 mg/dl) or previously diagnosed with T2D, (iii) triglycerides ≥ 1.7 mmol/L (150 mg/dl), (iv) high-density lipoprotein (HDL) cholesterol < 1.04 mmol/L (40 mg/dl) and (v) systolic blood pressure (SBP) / diastolic blood pressure (DBP) ≥ 130/85 mm Hg or previously diagnosed with high blood pressure.

Bioinformatic analyses

ITS2 data to detect fungal identity, diversity and abundance

Microbial DNA was extracted using the E.Z.N.A.® soil DNA Kit (Omega Bio-tek, Norcross, GA, U.S.) according to manufacturer's protocols. The final DNA concentration and purification were determined by NanoDrop 2000 UV-vis spectrophotometer (Thermo Scientific, Wilmington, USA), and DNA quality was checked by 1% agarose gel electrophoresis. The ITS2 hypervariable regions of the fungal ITS rRNA gene were amplified with primers ITS3F: GCATCGATGAAGAACGCAGC and ITS4R: TCCTCCGCTTATTGATATGC by thermocycler PCR system (GeneAmp 9700, ABI, USA). The PCR reactions were conducted using the following program: 3 min of initial denaturation at 95 °C, 35 cycles of 30 s at 95 °C, 30s for annealing at 55 °C, and 45s for elongation at 72 °C, and a final extension at 72 °C for 10 min. PCR reactions were performed in triplicate in a 20 μL mixture containing 4 μL of 5 × FastPfu Buffer, 2 μL of 2.5 mM dNTPs, 0.8

μL of each primer (5 μM), 0.4 μL of FastPfu Polymerase and 10 ng of template DNA. The resultant PCR products were extracted from a 2% agarose gel and further purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) and quantified using QuantiFluor™-ST (Promega, USA) according to the manufacturer's protocol. Purified amplicons were pooled in equimolar and paired-end sequenced (2 × 300) on an Illumina MiSeq platform (Illumina, San Diego, USA) according to the standard protocols by Majorbio Bio-Pharm Technology Co. Ltd. (Shanghai, China).

Raw data was demultiplexed by the MiSeq Controller Software (Illumina Inc.). QIIME2 (version 2020.2) was used for the downstream analysis[4]. The demultiplexed ITS2 sequences were denoised and grouped into amplicon sequence variants (ASVs; i.e., 100% exact sequence match) using DADA2[5]. The ASV features that were presented in only one sample were excluded for all cohorts. The individual ASVs were taxonomically classified based on the UNITE (version 8.2, 99%) database using the VSEARCH tool wrapped in QIIME2. α-diversity analysis was conducted through the q2-diversity plugin at the sampling depth of 4500. α-diversity was estimated by Shannon's diversity index (or Shannon; a quantitative measure of community richness and evenness), Observed Features (or Richness; a qualitative measure of community richness), Evenness (or Pielou's Evenness; a measure of community evenness), and Faith's PD (or Faith's Phylogenetic Diversity; a qualitative measure of community richness that incorporates phylogenetic relationships between the observed features).

To compare the gut mycobiome composition with different ethnicity, we also analysed the ITS2 sequencing data from the Human Microbiome Project (HMP) cohort (PRJNA356769, n = 313) and Danish cohort (PRJEB34758, n = 99). The detailed information about both cohorts had been described previously[6 7].

16S rRNA data to detect bacterial identity, diversity and abundance

For the 16S analysis, raw data was obtained using the same method as previously described. Fastq files were demultiplexed by the MiSeq Controller Software (Illumina Inc.) using a mapping file as input. Sequences were merge-paired, quality filtered and analysed using QIIME2 (version 2019.10). As described above, we used DADA2 denoised-paired plugin in QIIME2 to process the fastq files. We filtered the features that were present in only a single sample. The taxonomies of ASVs were subsequently determined using the Naive Bayes classifier trained on the Sliva_132 99% reference database. α-diversity analysis was conducted at the sampling depth of 1000. α-diversity of the gut bacteria was estimated by the indices the same as ITS2 data.

Metagenome data

Faecal DNA extractions were carried out by a standardized CTAB procedure. DNA concentration was measured using Qubit dsDNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, CA, USA). For DNA library preparation, a total amount of 1μg DNA per sample was used as input material. In addition, the NEBNext Ultra DNA

Library Prep Kit (NEB, USA) was used following manufacturer's recommendations and index codes were added to attribute sequences to each sample. The DNA samples were fragmented by sonication to a size of approximately 350 bp. Then, the DNA fragments were end-polished, A-tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. After that, PCR products were purified (AMPure XP system) and libraries were analysed for size distribution by Agilent2100 Bioanalyzer and quantified using real-time PCR. The clustering of the index-coded samples was performed on a cBot Cluster Generation System according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform and 150 bp paired-end reads were generated. Finally, we obtained on average 42.4 million paired-end raw reads for each sample.

Next, raw sequencing reads were first quality-controlled with PRINSEQ (v0.20.4)[8]: 1) to trim the reads by quality score from the 5′ end and 3′ end with a quality threshold of 20; 2) removed read pairs when either read was < 60 bp, contained "N" bases or quality score mean below 30; and 3) deduplicate the reads. Reads that could be aligned to the human genome (*H. sapiens*, UCSC hg19) were removed (aligned with Bowtie2 v2.2.5 using --reorder --no-contain --dovetail).

Functional profiling was performed with HUMAnN2 v2.8.1[9], which maps sample reads against the sample-specific reference database to quantify gene presence and

abundance in a species-stratified manner, with unmapped reads further used in a translated search against Uniref90 to include taxonomically unclassified but functionally distinct gene family abundances. We extracted the Uniref90 gene families of gut bacteria for downstream analyses. The Uniref90 gene families were then converted into KEGG Orthologs (KOs). The abundances of KOs were normalized into relative abundance for each sample. In addition, the KOs presented in less than 10% of the 1009 samples were exclude from the downstream analysis.

Targeted faecal metabolome profiling

The absolute quantification of faecal samples (n = 913) was performed by an ultra-performance liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS) system. Detailed information about the measurements has been described previously[10]. The same list of metabolites was selected to capture the microbiota-related metabolites and some key host metabolites. The metabolites mainly included amino acids, bile acids, fatty acids, carbohydrates, organic acids, benzenoids, pyridines, and carnitines.

Statistical analysis

All statistical analyses were performed using Stata version 15 or R version 4.0.2. To compare the inter-individual variability of the human gut fungi and bacteria, we applied a Mann-Whitney U test on the Bray-Curtis dissimilarity. The dissimilarity matrix was calculated based on the composition at the genus level using the *vegdist()*

function from the *vegan* R package[11]. We used a linear mixed model to examine the association of sequencing depth and sequencing run (as a random effect) with z-score normalized α-diversity indices. To examine the association between α-diversity of gut fungi and bacteria, the residuals of the linear mixed model were taken as technique-adjusted α-diversity indices for the Spearman correlation analysis. Procrustes analysis was performed to investigate the overall relationship between gut fungal and bacterial taxa, and the *p* value was generated based on 999 permutations. We then used the *envfit* function in *vegan* to explore the contributors for the variation of the gut fungal community. In addition, we excluded the fungal genera that were present in less than 5% of the 1244 participants. The remaining genera were identified as a group of core taxa in our population. To explore the stability of the human gut mycobiome, the relative abundance of the core taxa between baseline and follow-up visits were compared using the paired two-sided Wilcoxon signed-rank test, among 184 participants with repeated measures of ITS2 data. To investigate the relationship between age and gut mycobiome, we further performed a Wilcoxon rank-sum test and a linear regression analysis to study the gut mycobial differences between middle-aged and elderly adults. The regression model was adjusted for gender and BMI.

To investigate the determinants of the gut fungal composition, 40 factors (including demographics, physiologic traits, diseases, and habitual dietary intakes) were correlated to the gut fungal distance matrix (Bray-Curtis) using permutational multivariate analysis of variance (PERMANOVA; permutations = 999). The linear

regression models were performed to examine the prospective associations between the core taxa and the habitual dietary intakes (e.g., dairy, fish, cereals), adjusted for age, gender, BMI, total energy intake, smoking status, drinking status, education attainment, income level, physical activity, and Bristol stool scale. In these models, the rank-based inverse normal transformation was applied for the relative abundance of the gut fungi. The associations between the core taxa of gut fungi and technique-adjusted bacterial α-diversity indices were investigated using Spearman correlation analysis.

To further explore the ecological connections between gut fungi and bacteria, a 5% prevalence filtering for the bacterial genera was performed. The SparCC algorithm, wrapped in the FastSpar application[12], was used to estimate the cross-kingdom taxonomic associations based on the raw read counts tables. FastSpar is a fast and parallelizable implementation of the SparCC algorithm with an unbiased *p*-value estimator, which was designed to provide rapid and robust correlation estimation for compositional data. The analysis was conducted with default parameters and 1000 bootstrap samples were used to infer pseudo *p* values. To estimate the robustness of our SparCC-inferred associations, we compared them to the associations from the Spearman correlation analysis. In addition, we examined the associations between the relative abundance of the gut mycobiome and bacterial KOs using the Spearman correlation analyses. The significant KOs were then mapped to the PATHWAY, MODULE and BRITE databases (https://www.genome.jp/kegg/ko.html) that stored

the cellular and organism-level functions.

We used the Spearman correlation analysis to examine the associations between gut mycobiome and faecal metabolites. Procrustes analysis was performed to investigate the overall relationship between gut mycobiome and faecal metabolome, and the *p* value was generated based on 999 permutations. To evaluate the link between the gut fungal composition and faecal metabolites, we conducted the PERMANOVA analysis based on the Bray-Curtis distance matrix of the gut mycobiome (permutations = 999). *p* values were controlled by Benjamini Hochberg method for multiple tests. FDR corrected (*q* values) or raw *p* values < 0.05 were considered significant.

We applied two linear regression models to examine the associations between the gut mycobiome and metabolic traits, including BMI, waist circumference, triglycerides, total cholesterol (TC), HDL cholesterol, LDL cholesterol, TC/HDL, fasting glucose, HbA1c, insulin and HOMA-IR: dependent variable (metabolic traits) ~ (intercept) + independent variable (fungi, continuous / binary) + age + gender. The 'continuous' model used the relative abundance of fungi as independent variable. As for the 'binary' model, we transformed the relative abundance of fungal into binary variables (presence or absence). The dependent variables with skewed distribution were log-transformed before analyses (fasting glucose, insulin, triglycerides, TC/HDL and HOMA-IR). In addition, we used a linear regression model to examine the associations of the changes in relative abundance of gut mycobiome between baseline

and follow-up with the parallel changes in metabolic traits, adjusted for age and gender. To test the potential effect modification of the gut mycobiome, we examined the interaction of gut fungi with bacterial α-diversity on the metabolic traits in the discovery cohort. We further validated the results from the interaction analyses in an independent replication cohort. To evaluate whether bacterial genes or faecal metabolites can mediate the mycobiomic impact on the host metabolic traits, we first checked whether the bacterial KOs/faecal metabolites were associated with the gut mycobiome. Then we applied the bi-directional mediation analysis using the *mediate* function from the R package *mediation* to infer the causal role of the gut mycobiome in contributing to the human phenotype through bacterial gene function or faecal metabolites[13]. Here, the host metabolic traits were the change between the baseline and the follow-up visit while the gut mycobiome and bacterial gene/faecal metabolites were from the baseline. We conducted the z-score normalization for the host phenotype before the mediation analysis. The mediation models were adjusted for age and gender. We estimated the total effect, direct effect, and indirect effect of the gut mycobiome or bacterial KOs/faecal metabolites on the metabolic traits from these models. The results were confirmed by the simulation exercises bootstrapped 1000 times. To estimate whether the association between the *Saccharomycetales spp*. and *Pichia* changes with increased age, we compared the association not only between the middle-aged and the elderly based on the cross-sectional data, but also between the baseline and follow-up visits based on the fungal repeated measures.

Fig. S1. Study design for the present analyses. Our study was based on the Guangzhou Nutrition and Health Study (GNHS) which is an ongoing community based cohort study in China. We included 1244 participants in the current study. We profiled the gut mycobiome using ITS2 sequencing. 184 individuals have collected stool sample twice at different time points. The dietary information was collected using the food frequency questionnaires (FFQs) every three years since 2008. To examine the associations between gut fungi and bacterial taxonomy and function, we applied qiime2 and HUMAnN2 to perform the taxonomic and functional profiling from 16S rRNA sequencing data and stool metagenomics data. In addition, we investigated the associations between gut mycobiome and fecal metabolome. We employed linear regression models to examine the interaction between gut fungi and bacterial diversity on the cardiometabolic health. Finally, the bi-directional mediation analyses were performed to check whether gut mycobiome can influence the human phenotypes through gut bacterial gene function. Figure created with BioRender.com.

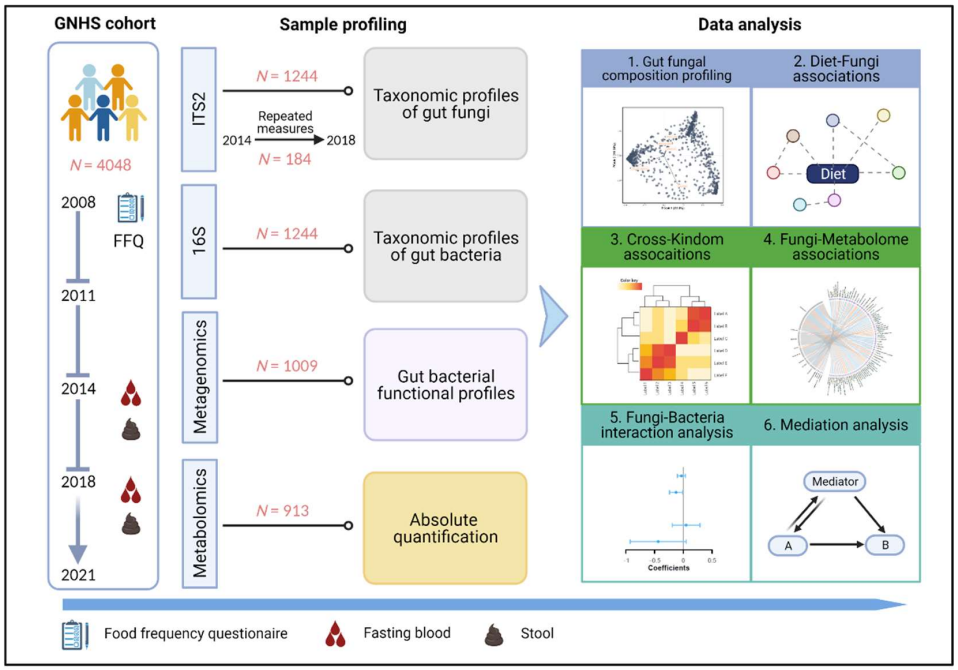


Fig. S2. The composition of gut mycobiome in different cohorts. (A) Age distribution of the GNHS cohort in the present study (N = 1244). (**B**) Bar plot shows the mean abundance ratio between the *Ascomycota* and *Basidiomycota*. (**C**) Venn plot shows the number of genera detected in three cohorts.

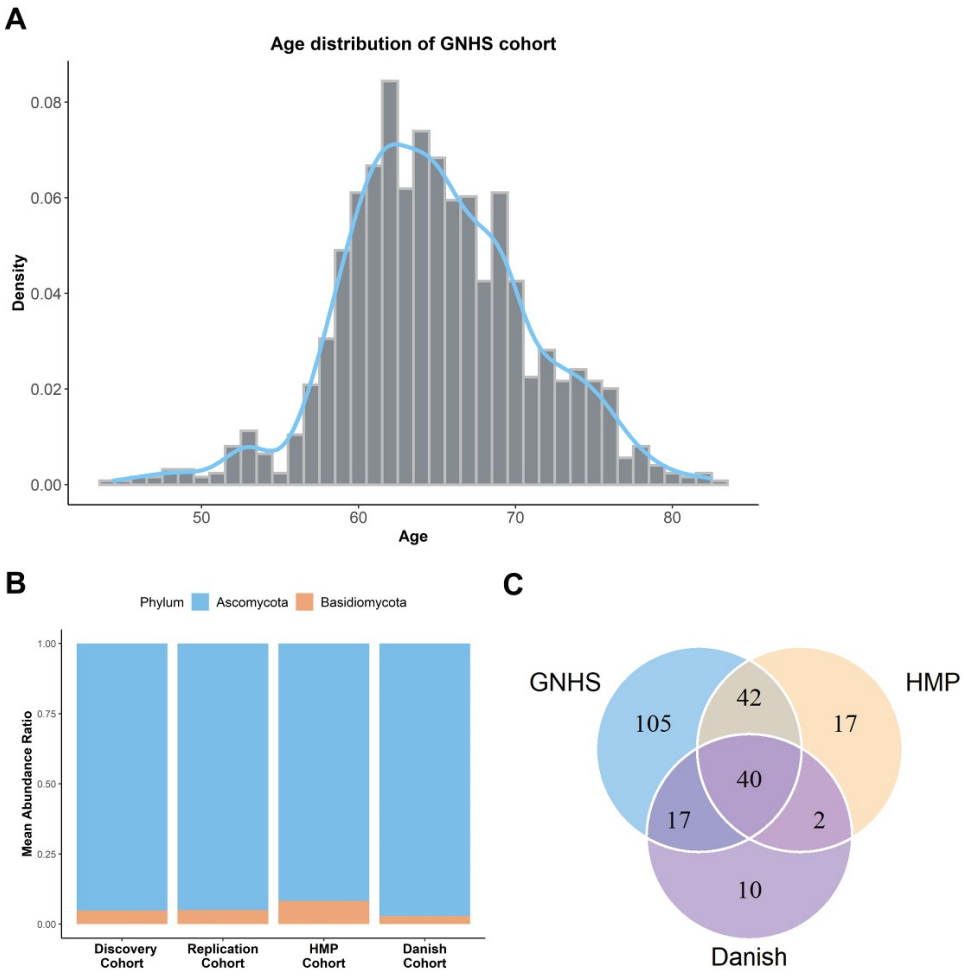


Fig. S3. Comparisons of the gut mycobiome between baseline and follow-up visit in the population with repeated measures (N = 184). The bar plots show the mean relative abundance (**A**) and prevalence (**B**) of the gut mycobiome at genus level. The median follow-up time was 3.2 years. **(C)** Bar plot shows the probability that the taxonomy (prob > 40%; 2 phyla and 7 genera) could be detected repeatedly at two time points.

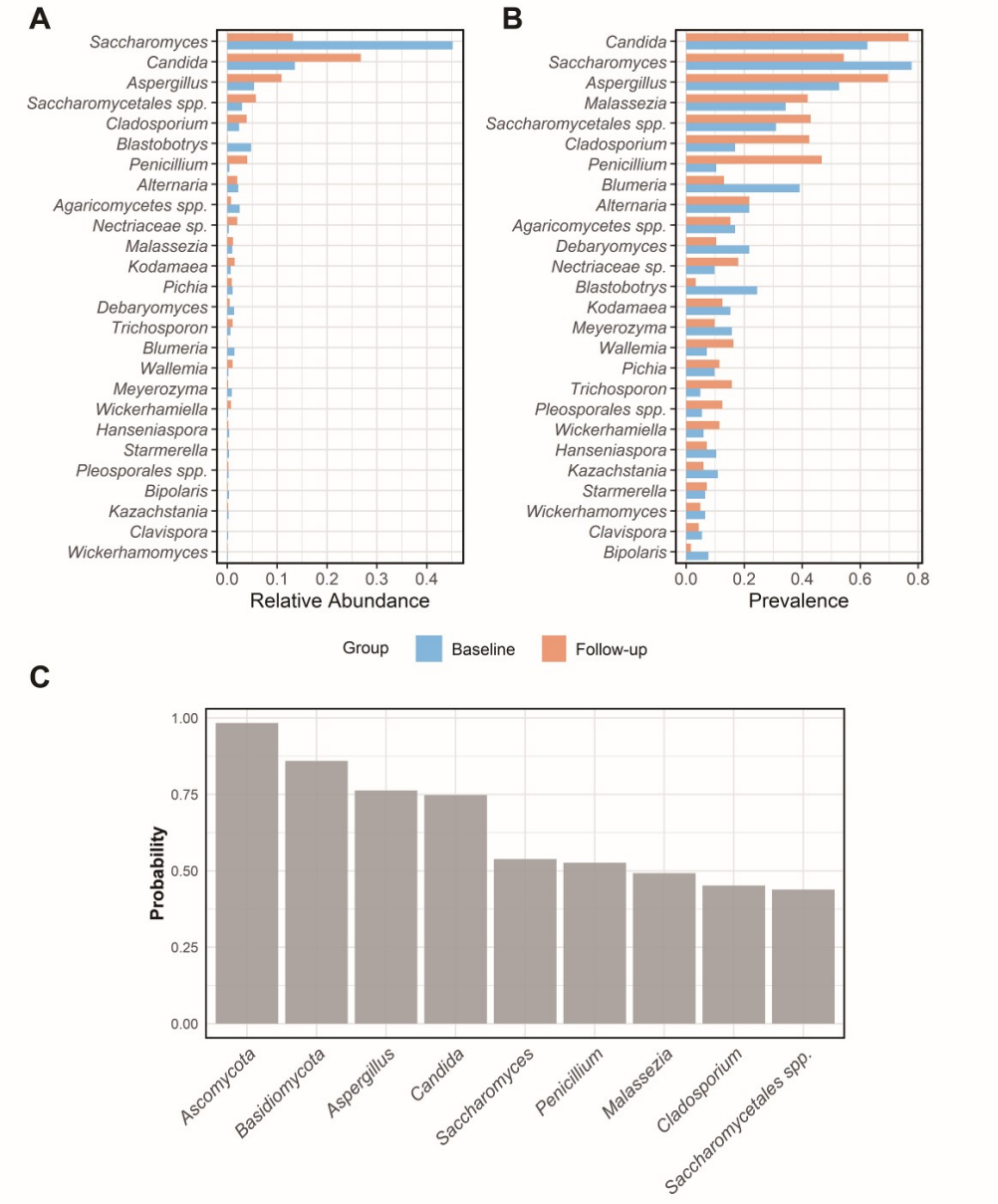


Fig. S4. Comparison of the composition between gut fungi and bacteria. A, The comparison of α-diversity indices between gut fungi and bacteria. **B,** The comparison of the inter-individual Bray-Curtis distance of gut fungi and bacteria. All box plots are the median with the interquartile range. *p* value was from Mann-Whitney U test.

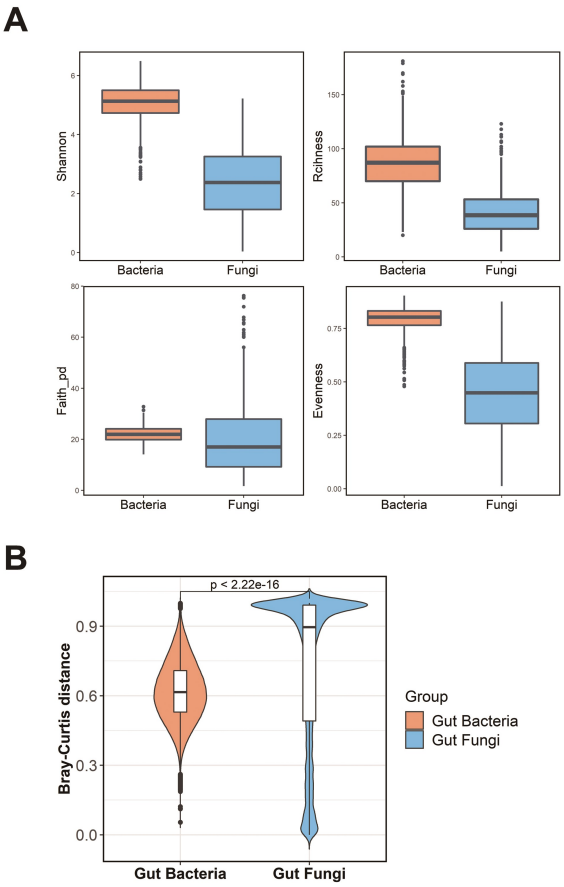


Fig. S5. Associations between gut fungal and bacterial composition. Procrustes analysis of gut fungi (ITS2) versus gut bacteria (16S) in 1244 participants. Fungi and bacteria are shown as orange and blue dots, respectively. Fungi and bacteria from the same individual are connected by grey lines.

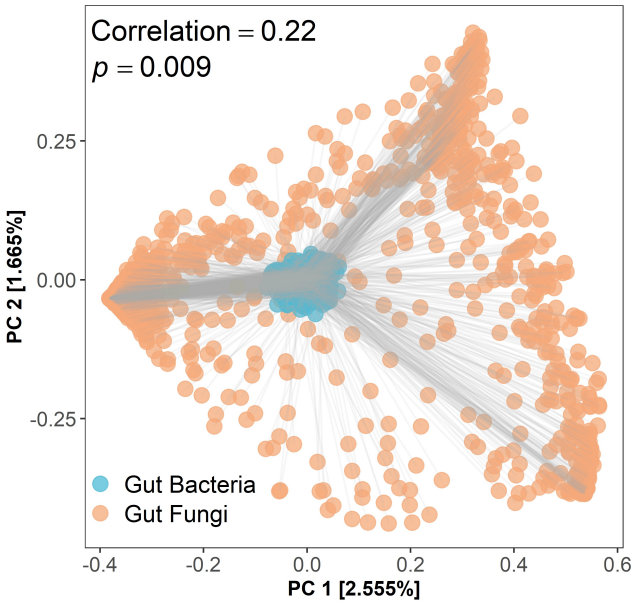


Fig. S6. Associations between the α-diversity indices of gut mycobiome and host phenotypes. Heatmap shows the Spearman correlation coefficients between the α-diversity indices and host phenotypes. Total number of participants in each analysis was 1235 for waist circumference and hip circumference, 1242 for neck circumference, 1210 for fasting glucose, 1019 for insulin, 1212 for HbA1c, 1211 for triglycerides, total cholesterol, HDL-C, LDL-CH and TC/HDL, 1239 for diet and 1244 for the other variables. bA1c, glycated hemoglobin; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC/HDL, total Cholesterol/HDL Cholesterol ratio; Middle-aged, age < 65 (N = 669); Elderly, age ≥ 65 (N = 575).

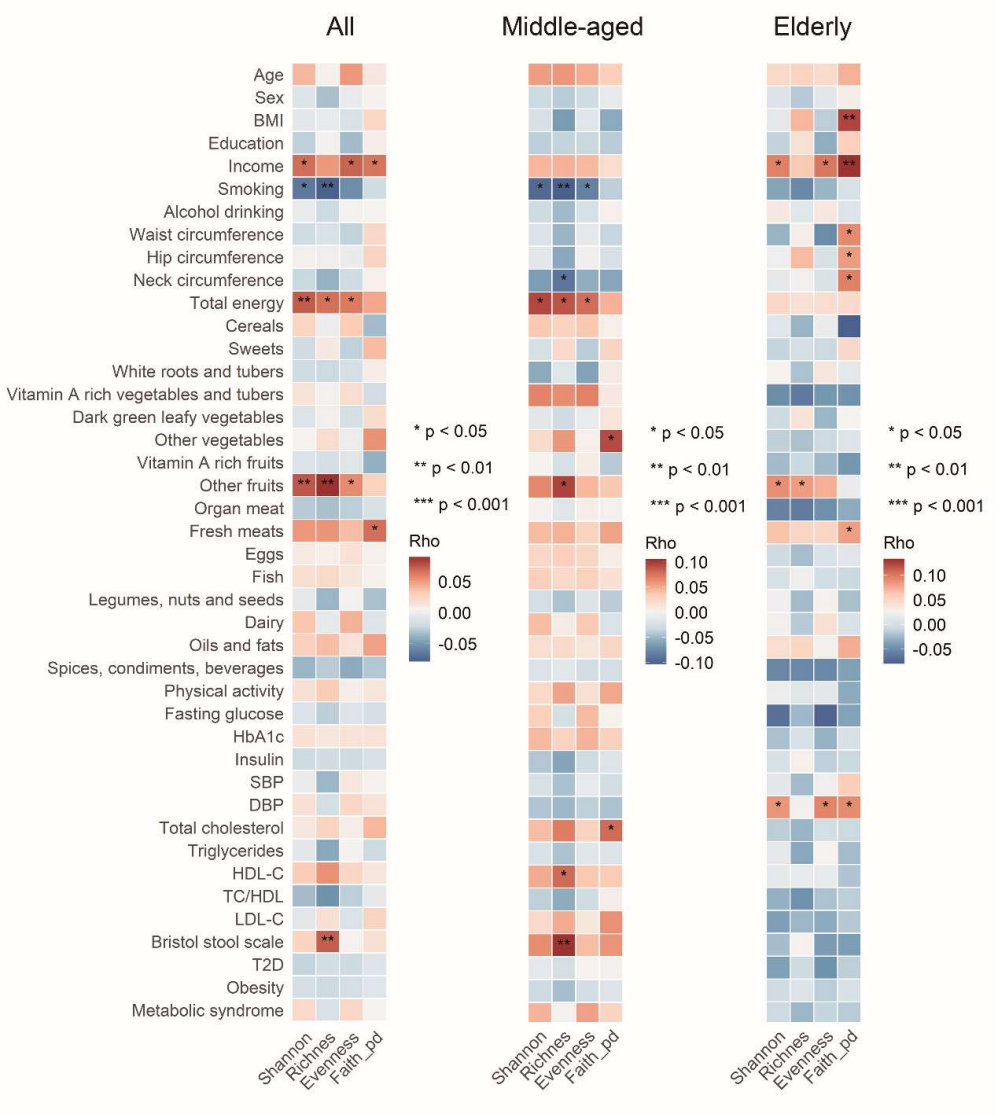


Fig. S7. The number of associations within the microbiome. A, Venn plot shows the number of associations between the gut fungi and bacteria assessed by SparCC and Spearman correlation analysis. **B,** Venn plot shows the number of inter- and intra-kingdom associations between gut bacteria and fungi.

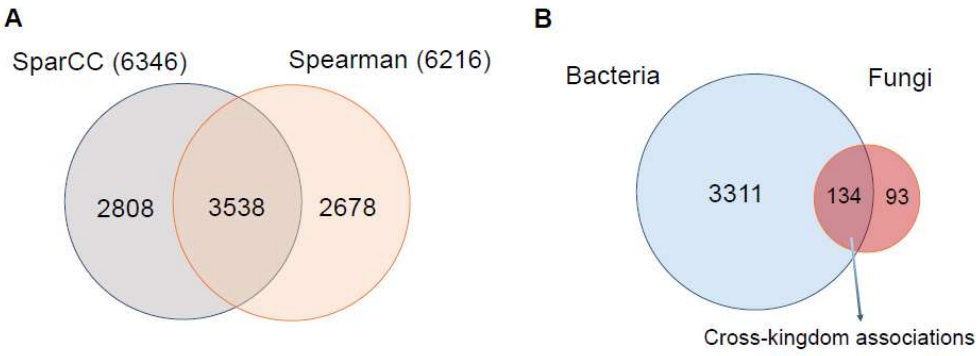


Fig. S8. Associations between the α-diversity indices of gut bacteria and gut mycobiota. Heatmap shows the Spearman correlation coefficients between specific gut fungi and the α-diversity indices of gut bacteria. *sp*. or *spp*. here represents unidentified fungal genus.

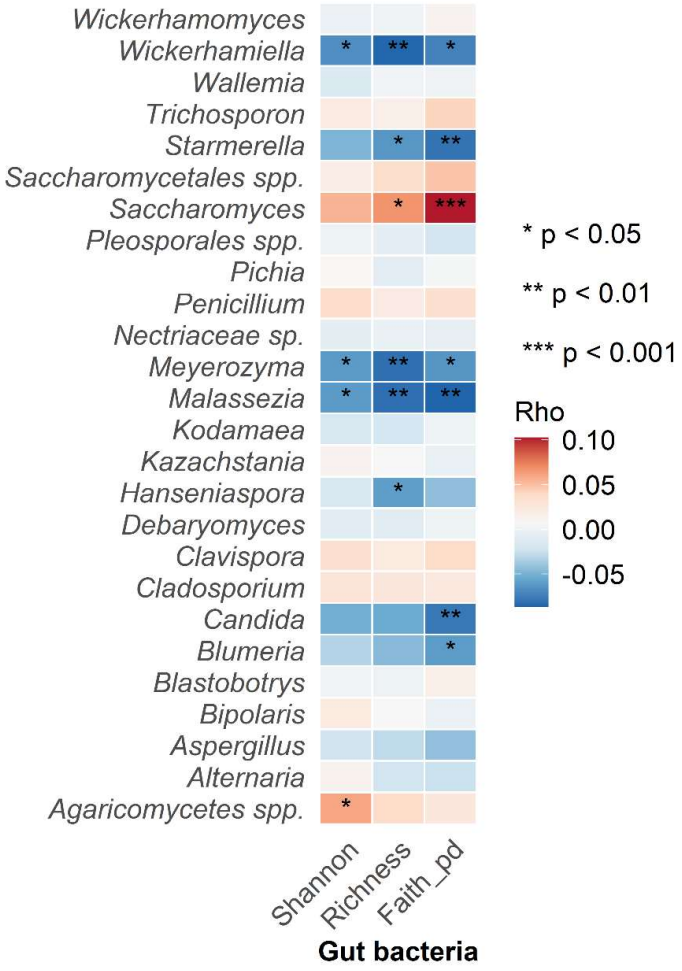


Fig. S9. Associations between gut mycobiome and fecal metabolome. A, Procrustes analysis of gut mycobiome versus fecal metabolome in 913 participants. Fungi and metabolites are shown as purple and green dots, respectively. Fungi and metabolites from the same individual are connected by grey lines. **B,** Venn plot shows the number of metabolites associated with gut fungal composition and specific fungi.

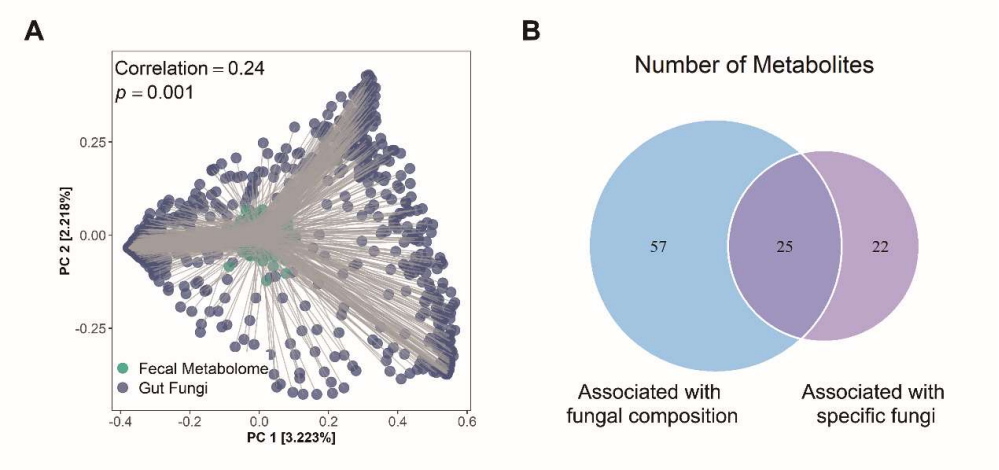


Fig. S10. Associations between gut mycobiome and metabolic traits in replication cohort. Heatmap of cross-sectional associations between gut fungi abundance and presence (left: continuous variable; right: binary variable) and metabolic traits. Linear regression models were adjusted by age and sex. Total number of participants in each analysis was 89 for fasting glucose, HDL cholesterol, HOMA-IR, LDL cholesterol, triglycerides, total cholesterol and TC/HDL, 90 for HbA1c and insulin, and 93 for BMI and waist circumference. Triglycerides, fasting glucose, TC/HDL and insulin were log-transformed.

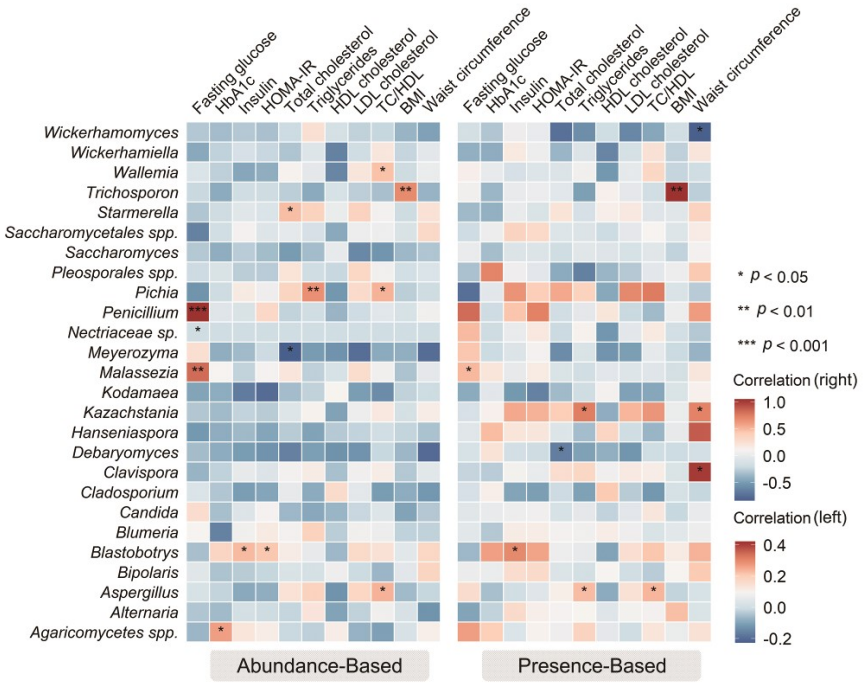


Fig. S11. Forest plot of the interaction of gut fungi with bacterial diversity on cardiometabolic traits. Associations were expressed as the difference in cardiometabolic traits (in SD unit) per 1 SD in each bacterial α-diversity index. Linear regression models were adjusted for age and sex. Associations were stratified by the presence of gut fungi. The error bars represent confidence intervals.

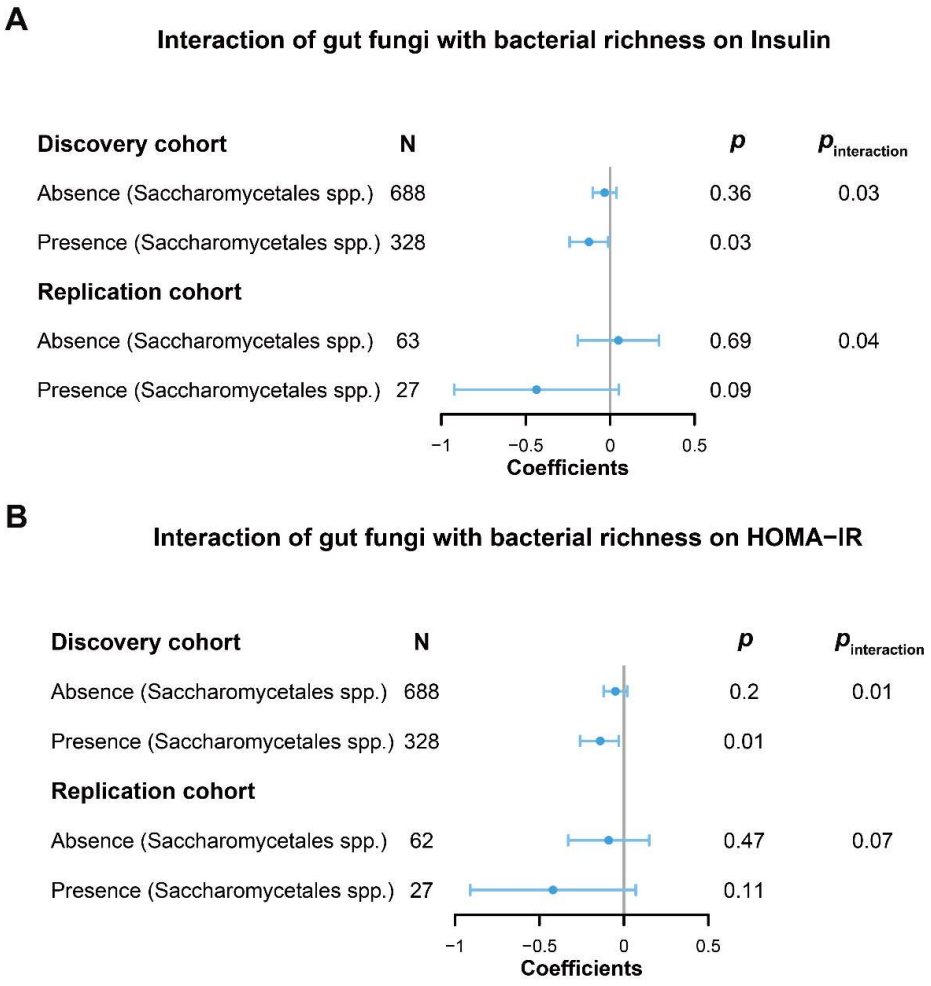


Fig. S12. Mediation linkages among the gut mycobiome, faecal metabolites and metabolic traits. p_{IDE} and $p_{inv\text{-}IDE}$ were estimated by the bi-directional mediation analysis. The metabolic traits were the longitudinal change between the baseline and follow-up visits while the gut mycobiome and faecal metabolites were from the baseline (N = 270). The red arrowed lines indicate the gut fungal effects on the changes of phenotype mediated by faecal metabolites with corresponding mediation p_{IDE}. Inverse mediation was performed to check whether faecal metabolites can influence the human phenotype through gut mycobiome. The percentage number in the center of each figure are effect ratios reflecting the proportion of the total effect of the independent variable on the dependent variable that is explained by the mediator, i.e., 29%. IDE, indirect effect; inv-IDE, inverse indirect effect. Figure created with BioRender.com.

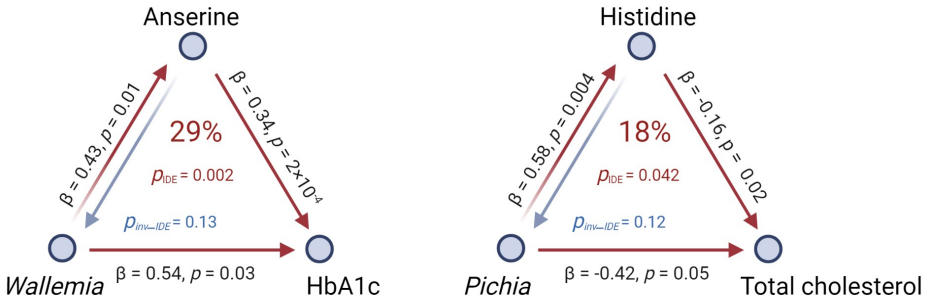
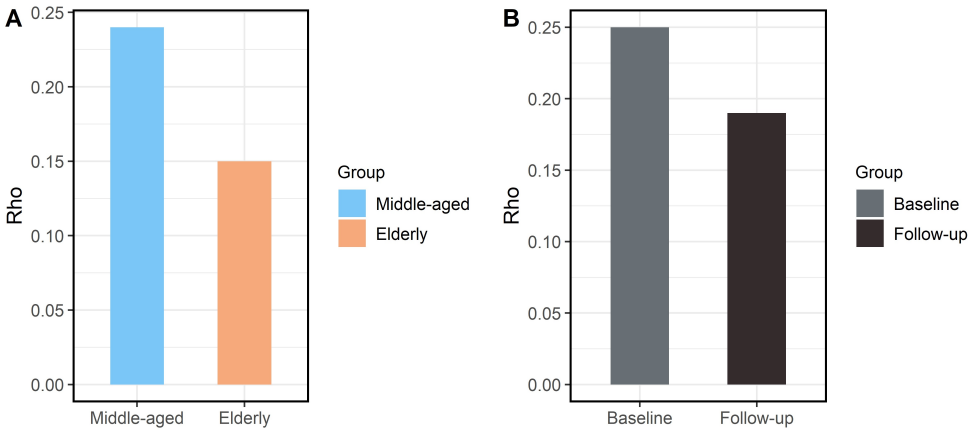


Fig. S13. Association between *Saccharomycetales spp.* and *Pichia*. A, Spearman correlation coefficients between the middle-aged (age < 65, N = 669, baseline) and the elderly (age ≥ 65, N = 575, baseline). **B,** Spearman correlation coefficients between the baseline and follow-up visits (N = 184). The median follow-up time was 3.2 years. All $p < 0.01$.



References

1. Zhang ZQ, He LP, Liu YH, et al. Association between dietary intake of flavonoid and bone mineral density in middle aged and elderly Chinese women and men. *Osteoporosis Int* 2014;25(10):2417-25. doi: 10.1007/s00198-014-2763-9

2. Fan F, Xue WQ, Wu BH, et al. Higher Fish Intake Is Associated with a Lower Risk of Hip Fractures in Chinese Men and Women: A Matched Case-Control Study. *Plos One* 2013;8(2) doi: ARTN e5684910.1371/journal.pone.0056849

3. Shuai ML, Zuo LSY, Miao ZL, et al. Multi-omics analyses reveal relationships among dairy consumption, gut microbiota and cardiometabolic health. *Ebiomedicine* 2021;66 doi: ARTN 10328410.1016/j.ebiom.2021.103284

4. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37(8):852-57. doi: 10.1038/s41587-019-0209-9

5. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13(7):581-+. doi: 10.1038/Nmeth.3869

[dataset] 6. Nash AK, Auchtung TA, Wong MC, et al. The Intestinal Mycobiome of the Human Microbiome Project Healthy Cohort. Sequence Read Archive, November 4, 2017. doi: 10.1186/s40168-017-0373-4

[dataset] 7. Ahmad HF, Mejia JLC, Krych L, et al. Gut Mycobiome Dysbiosis Is Linked to Hypertriglyceridemia among Home Dwelling Elderly Danes. European Nucleotide Archive, February 20, 2020. https://doi.org/10.1101/2020.04.16.044693

8. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011;27(6):863-4. doi: 10.1093/bioinformatics/btr026 [published Online First: 2011/02/01]

9. Franzosa EA, McIver LJ, Rahnavard G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 2018;15(11):962-68. doi: 10.1038/s41592-018-0176-y [published Online First: 2018/11/01]

10. Jiang Z, Sun TY, He Y, et al. Dietary fruit and vegetable intake, gut microbiota, and type 2 diabetes: results from two large human cohort studies. *BMC Med* 2020;18(1):371. doi: 10.1186/s12916-020-01842-0 [published Online First: 2020/12/04]

11. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* 2003;14(6):927-30. doi: Doi 10.1658/1100-9233(2003)014[0927:Vaporf]2.0.Co;2

12. Watts SC, Ritchie SC, Inouye M, et al. FastSpar: rapid and scalable correlation estimation for compositional data. *Bioinformatics* 2019;35(6):1064-66. doi: 10.1093/bioinformatics/bty734

13. Tingley D, Yamamoto T, Hirose K, et al. mediation: R Package for Causal Mediation Analysis. *J Stat Softw* 2014;59(5)