

1 **Supplemental Materials**

2 **Landscape of the gut mycobiome dynamics during pregnancy and its**
3 **relationship with host metabolism and pregnancy health**

4

5 This file includes:

6 **Methods**

7 (1) Study population

8 (2) Sample collection

9 (3) Questionnaires and clinical data collection

10 (4) Bioinformatic analyses

11 (5) Statistical analysis

12

13 **Supplemental Figure S1 to S2**

14 (1) Supplemental Figure S1. Supplementary Table S1: Characteristics of
15 demographic, clinical and dietary factors stratified by early-pregnancy fungal
16 enterotypes.

17 (2) Supplemental Figure S2. The association between changes in consumption of food
18 groups during pregnancy and alterations in gut fungal richness.

19

20 **Methods**

21 **Study population**

22 The current study was based on the Tongji-Huaxi-Shuangliu Birth Cohort (THSBC),
23 in which pregnant women aged 18 to 41 years were recruited during the early
24 pregnancy when they presented to antenatal care clinics in in a local maternal and
25 child health hospital during their early pregnancy. Exclusion criteria were 1) receiving
26 infertility treatment (e.g., in vitro fertilization or intrauterine insemination); 2)
27 reporting severe chronic or infectious diseases (e.g., cancer, HIV infection, or
28 tuberculosis); or 3) were unable to or refused to sign the informed consent. The
29 THSBC study was approved by the Ethics Committee of Tongji Medical College,
30 Huazhong University of Science and Technology (No. [2017](S225)-1), and informed
31 consent was obtained from all participants.

32

33 In the present analysis, we included 4800 participants who had available ITS2
34 sequencing data, dietary information and clinical records during their pregnancy. This
35 dataset enables us to comprehensively profile the gut mycobiome among pregnant
36 women and investigate potential determinants contributing to the variations of gut
37 mycobiome. To examine how pregnancy impacts the gut mycobiome over time and
38 investigate their potential associations with host metabolism, we established a sub-
39 cohort of 1059 participants, which included 514 women who gave birth to preterm
40 (n=240), low birthweight (n=137), or macrosomia (n=216) infants, as well as 545
41 randomly selected participants who did not experience the above three adverse
42 pregnancy outcomes.

43

44 ITS2 sequencing was performed for all the 4800 participants, while the shotgun
45 metagenomics sequencing was performed for T1 samples within the established sub-
46 cohort (n=1059). Additionally, within the sub-cohort, 750 and 748 participants had
47 ITS2 and 16S sequencing data available, respectively, for all trimesters. We also
48 repeatedly measured serum metabolome throughout each trimester of pregnancy for
49 participants in this selected sub-cohort using an LC-ESI-MS/MS system.

50 Sample collection

51 Stool samples were collected and stored in ice boxes at the hospital or home by the
52 participants under instructions and then transferred to the hospital to store at $-40\text{ }^{\circ}\text{C}$
53 within 24 hours. A detailed standard operating procedure was given to the pregnant
54 women for instructions on fecal sample collection, temporary storage, and
55 transportation to the hospital. Stocks of frozen fecal samples were transported by dry
56 ice every 2 to 3 months to the laboratory in Huazhong University of Science and
57 Technology and stored at $-80\text{ }^{\circ}\text{C}$ before further processing.

58

59 Questionnaires and clinical data collection

60 All participants completed a set of structured questionnaires on sociodemographic
61 information, lifestyle and behaviors (such as cigarette smoking and alcohol drinking),
62 diet (including recent daily intakes of major food groups), history of pregnancy and
63 births, history of diseases and medications, and family history of diseases.

64 Anthropometric and blood pressure measurements were collected on site using
65 devices according to standard protocols. Body mass index (BMI) was calculated by
66 dividing the weight in kilograms by the square of height in meters. Underweight,
67 overweight and obesity was defined as $\text{BMI} < 18.5$, $\text{BMI} \geq 24\text{ kg/m}^2$ and $\text{BMI} \geq 28$
68 kg/m^2 , respectively. Pregnancy complications (e.g., gestational diabetes mellitus) and
69 birth outcomes were extracted from the electronic clinical records.

70

71 Bioinformatic analyses**72 Gut mycobiome analysis using ITS2 rRNA gene sequencing data**

73 Microbial DNA was extracted using the E.Z.N.A.[®] soil DNA Kit (Omega Bio-tek,
74 Norcross, GA, U.S.) according to manufacturer's protocols. The final DNA
75 concentration and purification were determined by NanoDrop 2000 UV-vis
76 spectrophotometer (Thermo Scientific, Wilmington, USA), and DNA quality was
77 checked by 1% agarose gel electrophoresis. The ITS2 hypervariable regions of the
78 fungal ITS rRNA gene were amplified with primers ITS3F:

79 GCATCGATGAAGAACGCAGC and ITS4R: TCCTCCGCTTATTGATATGC by

80 thermocycler PCR system (GeneAmp 9700, ABI, USA). Purified amplicons were
81 pooled in equimolar and paired-end sequenced (2×250) on an Illumina NovaSeq
82 platform (Illumina, San Diego, USA) according to the standard protocols by Majorbio
83 Bio-Pharm Technology Co. Ltd. (Shanghai, China).

84

85 The mean sequencing depth and its standard deviation for all samples were 118,177
86 and 7955, respectively. It is worth noting that within the sub-cohort analyzed
87 longitudinally, the number of sequencing reads remained relatively consistent across
88 trimesters. Specifically, the mean (SD) sequencing depths were 117,906 (8350),
89 118,882 (6125), and 116,688 (9198) for samples collected during the first, second,
90 and third trimester of pregnancy, respectively. The demultiplexed ITS2 sequences
91 were denoised and grouped into amplicon sequence variants (ASVs; i.e., 100% exact
92 sequence match) using DADA2.[1] During the process, marker gene Illumina
93 sequence data and low-quality regions of the sequences were detected and filtered.
94 We trimmed 28 bases (primer and barcode) from the beginning of the sequences. We
95 also truncated the sequences at the 245 bases as the quality dropped around position
96 245 (median of quality score <30). The ASV features that were presented in only one
97 sample were excluded as suggested by the Qiime2 tutorial, based on the suspicion that
98 these may not represent real biological diversity but rather PCR or sequencing errors.
99 The individual ASVs were taxonomically classified based on the UNITE (version 8.2,
100 99%) database using the VSEARCH tool wrapped in QIIME2 (version 2021.2).[2] α -
101 diversity analysis was conducted through the q2-diversity plugin at the sampling
102 depth of 10000. α -diversity was estimated by Shannon's diversity index (or Shannon;
103 a quantitative measure of community richness and evenness), Observed Features (or
104 Richness; a qualitative measure of community richness), and Faith's PD (or Faith's
105 Phylogenetic Diversity; a qualitative measure of community richness that incorporates
106 phylogenetic relationships between the observed features).

107

108 **Gut bacteria analysis using 16S rRNA gene sequencing data**

109 For the 16S analysis, raw sequencing reads were merge-paired, quality filtered and

110 analyzed using QIIME2 (version 2021.2). As described above, we used DADA2
111 denoised-paired plugin in QIIME2 to process the fastq files. We filtered the features
112 that were present in only a single sample. The taxonomies of ASVs were subsequently
113 determined using the Naive Bayes classifier trained on the Sliva_138 99% reference
114 database. α -diversity analysis was conducted at the sampling depth of 10000. A
115 diversity of the gut bacteria was estimated by the indices the same as ITS2 data.

116

117 **Microbial functional profiling using metagenome data**

118 Microbial DNA extractions were carried out by a standardized CTAB procedure.
119 DNA concentration was measured using Qubit dsDNA Assay Kit in Qubit 2.0
120 Fluorometer (Life Technologies, CA, USA). For DNA library preparation, a total
121 amount of 1 μ g DNA per sample was used. In addition, the NEBNext Ultra DNA
122 Library Prep Kit (NEB, USA) was used following manufacturer's recommendations
123 and index codes were added to attribute sequences to each sample. The DNA samples
124 were fragmented by sonication to a size of approximately 350 bp. Then, the DNA
125 fragments were end-polished, A-tailed, and ligated with the full-length adaptor for
126 Illumina sequencing with further PCR amplification. Thereafter, PCR products were
127 purified (AMPure XP system) and libraries were analyzed for size distribution by
128 Agilent2100 Bioanalyzer and quantified using real-time PCR. The clustering of the
129 index-coded samples was performed on a cBot Cluster Generation System according
130 to the manufacturer's instructions. Lastly, sequencing was performed using the
131 Illumina NovaSeq platform at Shanghai Personal Biotechnology Co. Ltd. (Shanghai,
132 China) and 150 bp paired-end reads were generated.

133

134 Next, raw sequencing reads were first quality-controlled with KneadData toolkit
135 (v0.10.0): 1) to trim the reads by quality score from the 5' end and 3' end with a
136 quality threshold of 20; 2) removed read pairs when either read was < 50 bp,
137 contained "N" bases or quality score mean below 30; and 3) deduplicate the reads.
138 Reads aligning to the human genome (H. sapiens, UCSC hg38) were removed via
139 KneadData integrated with Bowtie2 tool (v2.4.5).

140 Functional profiling was performed with HUMAnN3 (v3.0.1), which maps sample
141 reads against the sample-specific reference database to quantify gene presence and
142 abundance in a species-stratified manner, with unmapped reads further used in a
143 translated search against Uniref90 to include taxonomically unclassified but
144 functionally distinct gene family abundances. We extracted the Uniref90 gene
145 families of gut bacteria for downstream analyses. The Uniref90 gene families were
146 then converted into relative abundances of unstratified pathway.

147

148 **Serum metabolomics profiling**

149 The sample extracts were analyzed using an LC-ESI-MS/MS system (UPLC,
150 ExionLC AD, <https://sciex.com.cn/>; MS, QTRAP® System, <https://sciex.com/>) at
151 Wuhan Metware Biotechnology Co., Ltd. (Wuhan, China). LIT and triple quadrupole
152 (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer
153 (QTRAP), QTRAP® LC-MS/MS System, equipped with an ESI Turbo Ion-Spray
154 interface, operating in positive and negative ion mode and controlled by Analyst 1.6.3
155 software (Sciex). Instrument tuning and mass calibration were performed with 10 and
156 100 µmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively.
157 QQQ scans were acquired as MRM experiments with collision gas (nitrogen) set to 5
158 psi. Declustering potential (DP) and collision energy (CE) for individual MRM
159 transitions was done with further DP and CE optimization. A specific set of MRM
160 transitions were monitored for each period according to the metabolites eluted within
161 this period. The mass spectrum data were processed using Software Analyst 1.6.3.
162 The metabolite identification was conducted by referencing standards in self-built
163 metware database and public databases. The identified metabolites were matched with
164 the parent ion mass-to-charge ratio, the fragment ion mass-to-charge ratio as well as
165 retention time of their corresponding standards. The accuracy of metabolite
166 characterization was classified into three levels, depending on the presence of isotope
167 internal standards or matching score with the secondary mass spectrometry. The
168 matching score > 0.7 indicates the level 1 accuracy of metabolite characterization,
169 while 0.5-0.7 and < 0.5 indicate the level 2 and level 3 accuracy, respectively.

170 **Statistical analysis**

171 **Determinants of gut fungal composition among pregnant women**

172 The gut mycobiome data were analyzed at the genus level. We investigated the
173 determinants of the gut fungal composition using the data collected from 4800
174 participants during the first trimester, with each participant contributing only 1 stool
175 sample. We used the *vegdist* function in the R package *vegan* to calculate the gut
176 fungal Bray-Curtis dissimilarity matrix. The contribution of 20 environmental
177 variables (including demographics, physiologic traits, diseases, and habitual dietary
178 intakes) to fungal community variation was determined by PERMANOVA analysis
179 using the function *adonis2* in *vegan*.^[3] We applied a complete data analysis strategy
180 which excluded 19 participants with missing values for at least one of the
181 environmental variables. The 4800 samples included in this analysis were sequenced
182 in two separate batches. Therefore, we included the batch information as a covariate
183 in the model, to adjust for potential batch effects. The p value was determined through
184 999 permutations.

185

186 **Gut fungal and bacterial enterotype clustering**

187 The fecal samples (T1, n=4800; T2, n=890; T3, n=850) of ITS2 amplification were
188 clustered into fungal enterotypes by using a partitioning around medoid (PAM)
189 clustering method as those previously described.^[4] Briefly, the samples were grouped
190 into clusters with partitioning around medoid (PAM) based on the between-sample
191 Bray-Curtis distance calculated at genus-level. The optimal number of clusters was
192 determined by the silhouette index. The driver genus of each enterotype was
193 determined as the genus with the highest relative abundance in the enterotype. The
194 fecal samples of 16S amplification were clustered into bacterial enterotypes by using
195 the method as that for bacterial enterotype.

196

197 **Dynamics of within-sample α diversity throughout each trimester of pregnancy**

198 This analysis was conducted in the established sub-cohort of 750 participants who had
199 available gut fungi α diversity data for each trimester of pregnancy. We utilized

200 paired t-tests to assess the statistical significance of the changes in fungal α diversity
201 between T1 and T2, as well as between T2 and T3 independently.

202

203 Among the 750 participants included in the study to profile changes in gut fungal
204 richness from T1 to T3, data on the estimation of consumption changes in food groups
205 from T1 to T3 based on the FFQ were available for 639 participants. These food
206 groups consisted of rice, steamed bread, noodles, vegetables, meat, eggs, milk, and
207 fruit. The weight of each food group consumed per day was quantified based on the
208 FFQ. We applied a linear regression model to estimate the association between the
209 changes in richness from T1 to T3 and the respective food group. As covariates, we
210 incorporated age, pre-pregnancy BMI, interval time between sample collections,
211 parity, and gravidity. An FDR<0.05 was considered statistically significant.

212

213 **Loss rate calculations and discriminative genera identification**

214 Utilizing the available repeated measurements of gut fungi throughout pregnancy
215 within the designated sub-cohort (n=750), we elucidated the loss rate for each fungal
216 genus as the host underwent progression from T1 to T3. The loss rate for each fungal
217 genus was determined by quantifying the proportion of the decline in frequency
218 observed between T1 and T3.

219

220 To assess the gut fungi enriched or depleted during early or late pregnancy, we
221 conducted an analysis using paired t test analysis. This analysis was based on the
222 examination of 465 genera that were detected at either T1 or T3. We transformed the
223 taxa data using the centered log-ratio (CLR) method to address the compositional
224 nature of the mycobiome data before we perform the paired t test analysis. To
225 determine statistical significance, a false discovery rate (FDR)-adjusted p-value
226 threshold of less than 0.05 was used.

227 Additionally, we used these 465 fungal genera to construct a machine learning
228 framework of LightGBM for predicting the trimester that the samples belong to.[5]

229 The construction of prediction model was based on Scikit-learn (v0.15.2), and ten-

230 fold cross validation (CV) was applied. To evaluate binary classification performance,
231 receiver operating characteristic (ROC) curve analyses were conducted using the R
232 package pROC. We used the SHAP (Shapley Additive exPlanations) to interpret
233 predictions and the importance of each fungal genus to the prediction model is
234 represented using Shapley values.[6]

235

236 **Quantification of intra-individual gut fungi compositional alterations**

237 We applied the *vegdist* function from the R package *vegan* to calculate the Bray-
238 Curtis distance based on gut fungal genus-level composition to assess intra-individual
239 and inter-individual dissimilarities, respectively.[3] The sub-cohort, consisting of 750
240 participants, was included in this analysis, as they had relevant data throughout each
241 trimester of pregnancy. To determine the intra-individual distance, paired data was
242 used, with the gut mycobiome data at T1 serving as the reference. Subsequently, a
243 Bray-Curtis distance value was calculated for each participant in the sub-cohort,
244 reflecting the extent of gut fungal compositional alterations.

245 We fitted a multivariate regression model to examine the associations of pre-
246 pregnancy overweight status (category variable) or pre-pregnancy BMI (continuous
247 variable, z-score transformed) with the extent of gut fungal compositional alteration
248 within the established sub-cohort. The model was adjusted for potential confounders
249 including age, time interval between sample collection, parity, antibiotics use and
250 pregnancy complications. To examine the potential influence of gut fungal
251 compositional alteration on adverse birth outcomes, multivariate regression models
252 were constructed for preterm delivery, low birthweight, and macrosomia. The
253 aforementioned potential confounders, along with the extent of gut fungal
254 compositional alteration, were considered as exposure variables in these models.

255

256 With regard to the divergence between individuals, the inter-individual distance was
257 evaluated at various time points (i.e., T1, T2 or T3), separately. At each time point
258 (e.g., T1 or T3), we calculate the average Bray-Curtis distance for each participant
259 compared to all other participants. Thus, at each time point, each participant

260 possessed a distance value to reflect the similarity of her gut fungal composition with
261 others.

262

263 **Dynamic trajectory of the core gut fungal genera and their relationship with host** 264 **health status**

265 We conducted a longitudinal trajectory analysis for each core fungal genus in the
266 established sub-cohort comprising 750 participants. For each core fungi, every
267 participant had three measurements recorded at different time points, namely T1, T2,
268 and T3. Therefore, the trajectory of a genus within an individual participant could be
269 characterized by a vector consisting of three component values corresponding to these
270 time points. Thereafter, we applied the *vegdist* function from the R package *vegan* to
271 calculate the Canberra distance metric.[3] We performed PERMANOVA analysis
272 using the function *adonis2* in *vegan*, to assess the association of between-individual
273 variation in the trajectory of each fungal genus with pre-pregnancy overweight status
274 or adverse birth outcomes. The p value was determined through 999 permutations and
275 an FDR-adjusted p value of less than 0.05 was considered indicative of statistical
276 significance.

277

278 **Network analysis among gut fungal enterotype, functional pathways and host** 279 **serum metabolites**

280 The network analysis was conducted among the participants who had available
281 metagenome data during the first trimester of pregnancy. After excluding 35
282 participants who had antibiotics exposure within 2 weeks before stool sample
283 collection, this analysis included a total of 1001 women. We firstly performed
284 Kruskal-Wallis test to identify pathways whose distribution varied across fungal
285 enterotypes. Thereafter, we performed post-hoc pair-wise comparison to defined
286 which enterotype was enrich with the identified pathways. We fitted multivariate
287 regression models to examine the associations of identified pathways with host serum
288 metabolites. Here, we adjusted for potential confounders including age, gestation
289 week, parity and pre-pregnancy BMI. Both the fungi and pathway data were

290 standardized using z-score before the regression analysis. An FDR-adjusted p value of
291 less than 0.05 was considered indicative of statistical significance.

292

293 **Covarying relationship between gut fungi alterations and host metabolic changes**

294 We investigated the relationship between alterations in gut fungi and changes in host
295 metabolism among 709 participants. These participants had available gut mycobiome
296 sequencing data as well as serum metabolome data during each trimester of
297 pregnancy. To address the compositional nature of the mycobiome data, we first
298 applied the centered log-ratio (CLR) method to transform the taxa data. Next, we
299 calculated the changes in each core fungus from T1 to T3 for each participant.

300 Additionally, we calculated the changes in signal intensity of each serum metabolite
301 from T1 to T3 for each participant. This enabled us to construct a matrix of gut fungi
302 alterations and a matrix of host metabolic changes. To investigate the overall
303 relationship between gut fungi alterations and host metabolic changes, we conducted
304 Procrustes analysis in R using the ‘vegan’ R package. Procrustes. The p-value was
305 generated based on 999 permutations.[3]

306

307 Furthermore, we explored the covarying relationship between individual fungal
308 genera and individual serum metabolites. For this analysis, we applied pairwise
309 Spearman correlation analysis to each genus-serum metabolite pair in the
310 aforementioned dataset. We considered a false discovery rate (FDR)-adjusted p-value
311 of less than 0.05 as indicative of statistical significance.

312

313 **Pre-pregnancy overweight status impacts the metabolic changes during** 314 **pregnancy**

315 For those participants with available serum metabolomics data throughout each
316 trimester of pregnancy, we had constructed a matrix of host metabolic changes. Based
317 on this matrix, the *vegdist* function from the R package *vegan* was utilized to calculate
318 the Canberra distance metric.[3] To assess the contribution of pre-pregnancy
319 overweight to the variation in metabolic changes between individuals, a

320 PERMANOVA analysis was conducted using the *adonis2* function from the vegan
321 package.[3] The significance of the results was determined using 999 permutations,
322 and a false discovery rate (FDR)-adjusted p-value of less than 0.05 was considered
323 statistically significant.

324

325 To identify the distinct metabolic alterations in underweight and overweight pregnant
326 women, we classified the participants into three groups based on their pre-pregnancy
327 weight status: underweight, normal weight, and overweight/obese. Paired t-tests were
328 conducted for each serum metabolite measured at T1 and T3 within each group. A
329 FDR-adjusted p-value of less than 0.05 was considered statistically significant.

330 Metabolic changes that were observed as significant solely among the underweight
331 group, but not in the other groups, were defined as unique metabolic changes in
332 pregnant women who were underweight prior to pregnancy. Similarly, significant
333 metabolic changes observed solely among the overweight/obese group were defined
334 as unique metabolic changes in pregnant women who were overweight/obese prior to
335 pregnancy.

336

337 **Clinical impact of the gut mycobiome during pregnancy**

338 Logistic regression was used to evaluate the association between each core fungal
339 genus, specifically measured during the first trimester of pregnancy, and the
340 occurrence of pregnancy complications (GDM and PIH), as well as adverse birth
341 outcomes (e.g., preterm delivery, macrosomia, and low birthweight). The model
342 included age, pre-pregnancy BMI, parity as covariates. Specifically, when examining
343 GDM or PIH as exposures, the model was adjusted for these variables. Additionally,
344 when fungal genera were included as exposures, the fungal genera data were z-score
345 standardized and the model was further adjusted for the gestational week
346 corresponding to stool sample collection and the batch of sequencing.

347

348 The analysis pertaining to pregnancy complications was conducted among 4606
349 participants who possessed gut fungi sequencing data at T1 and information regarding

350 pregnancy complications. Likewise, the analysis concerning birth outcomes was
351 performed among 4656 participants who had gut fungi sequencing data at T1 and
352 information on birth outcomes. In both analyses, a complete data analysis strategy
353 was implemented for variables with an exceptionally low occurrence of missing
354 values, which consequently excluded 3 participants. For the variable "gestational
355 week at stool sample collection," which was missing for 278 participants, a multiple
356 imputation strategy was employed to impute these missing values. As a result, the
357 logistic regression model for pregnancy complications consisted of 4603 participants,
358 while the logistic regression model for adverse birth outcomes encompassed 4653
359 participants.

360

361 For pregnancy complications that exhibited significant associations with both gut
362 fungi and adverse birth outcomes, we conducted mediation analysis to investigate the
363 potential mediation effect of pregnancy complications on the link between the
364 mycobiome and adverse birth outcomes. All statistical analyses were performed using
365 Stata version 15 or R version 4.0.2.

366

367 **References**

- 368 1 Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample
369 inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
- 370 2 Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and
371 extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37:852–7.
- 372 3. Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B.,
373 Simpson, G.L., Sólymos, P., Stevens, M.H.H. & Wagner, H., *vegan: Community*
374 *Ecology Package*, 2012, Software. <http://CRAN.R-project.org/package=vegan>.
- 375 4. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome.
376 *Nature*. 2011;473:174–80.
- 377 5. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu T.Y. (2017)
378 LightGBM: a highly efficient gradient boosting decision tree. NIPS'17: Proceedings
379 of the 31st International Conference on Neural Information Processing Systems.

380 3149–3157.

381 6. Lundberg, S.M., Lee, S.I. (2017) A unified approach to interpreting model

382 predictions. NIPS'17: Proceedings of the 31st International Conference on Neural

383 Information Processing Systems. 4768–4777.

384

385

Table S1: Characteristics of demographic, clinical and dietary factors stratified by early-pregnancy fungal enterotypes

	<i>Candida</i> -dominated enterotype (n=907)	<i>Aspergillus</i> -dominated enterotype (n=2723)	<i>Saccharomyces</i> -dominated enterotype (n=1170)	<i>p</i> *
Age (year)	26.4 (3.6)	26.3 (3.6)	26.5 (3.5)	0.246
Pre-pregnancy BMI	20.8 (2.9)	20.9 (2.9)	21.0 (3.1)	0.343
Gestation week (wk)	10.2 (2.1)	10.2 (2.0)	10.1 (2.0)	0.374
Time to delivery (day)	198.2 (16.1)	198.4 (15.5)	198.5 (15.6)	0.940
Parity				0.013
Primiparous	54.0%	57.3%	60.5%	
Overweight status				0.249
Overweight or obese	21.8	19.2	19.7	
Normal weight	65.1	67.6	65.3	
Underweight	13.1	13.3	15	
Steam bread consumption				<0.001
Ever (during the past 1 year)	46.9%	47.3%	54.2%	
Drinking				0.070
Ever (during the past 1 year)	22.4%	22.0%	18.9%	
Tea consumption				0.024
Ever (during the past 1 year)	34.0%	33.7%	29.5%	
Coffee consumption				0.057
Ever (during the past 1 year)	24.6%	26.6%	23.1%	
Milk consumption				0.564
never	19.2%	21.0%	19.7%	
<1 / day	42.9%	42.9%	42.1%	
≥1/day	38.0%	36.1%	38.2%	

* One-way ANOVA was applied to examine the significance of difference between groups for continuous variables, while chi-square test was used for category variables.

Table S2. The association between changes in consumption of food groups during pregnancy and alterations in gut fungal richness*

Exposure (changes in consumption of food groups)	Beta coefficient	95% CI	p	FDR
Rice (per 50g/day)	-3.20	(-5.75, -0.64)	0.014	0.11
Steamed bread (per 1 bread/day)	-6.38	(-12.9, 0.14)	0.055	0.15
Noodle (per 50g/day)	-5.06	(-10.63, 0.50)	0.074	0.15
Vegetables (per 50g/day)	1.20	(-9.82, 3.22)	0.243	0.39
Meat (per 50g/day)	1.45	(-3.48, 6.37)	0.564	0.64
Egg (per 1 egg/day)	7.33	(0.26, 14.41)	0.042	0.15
Milk (per 50g/day)	0.01	(-0.01, 0.03)	0.51	0.64
Fruit (per 50g/day)	-0.30	(-1.89, 1.29)	0.712	0.71

* Covariates included in the regression model: age, pre-pregnancy BMI, interval time between sample collections, parity, and gravidity. 639 out of the 750 with available data on the estimation of consumption changes in food groups from T1 to T3 based on the FFQ were included.