Fungal mycoflora dysbiosis in gastric cancer

Mengya Zhong
Xiamen University
Yubo Xiong
Xiamen University
Jiabao Zhao
Department of Gastrointestinal Surgery, Zhongshan Hospital of Xiamen University, Xiamen, Fujian, China.
Zhi Gao
Guangxi Medical University
Jingsong Ma
Xiamen University
Zhengxin Wu
Guangxi University
Xuehui Hong (hongxu@xmu.edu.cn)
Zhongshan Hospital Xiamen University https://orcid.org/0000-0002-8853-7384
Yongxi Song
The First Affiliated Hospital of China Medical University

Research

Keywords: gastric cancer, fungal dysbiosis, Candida albicans, mycobiome, biomarker

Posted Date: August 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-58045/v1

Abstract

Background: Bacterial infection is associated with gastric carcinogenesis. However, the relationship between nonbacterial components and gastric cancer (GC) has not been fully explored. We aimed to characterize the fungal mycobiome in GC.

Results: We observed significant gastric fungal dysbiosis in GC. Principal component analysis revealed separate clusters for the GC and control groups, and Venn diagram analysis indicated that the GC group showed a lower OTU abundance than the control. At the genus level, the abundances of 15 fungal biomarkers distinguished the GC group from the control, of which *Candida* ($p = 0.000246$) and *Alternaria* ($p = 0.00341$) were enriched in GC, while *Saitozyma* ($p = 0.002324$) and *Thermomyces* ($p = 0.009158$) were decreased. Combining the results of Welch's t test and Wilcoxon rank sum test, *C. albicans* was significantly elevated in GC. The species richness Krona pie chart further revealed that *C. albicans* occupied 22% and classified GC from the control with an area under the receiver operating curve (AUC) of 0.743. Random forest analysis also confirmed that *C. albicans* could serve as a biomarker with a certain degree of accuracy. Moreover, compared with that of the control, the alpha diversity index was significantly reduced in the GC group. The Jaccard distance index and the Bray abundance index of the PCoA clarified separate clusters between the GC and control groups at the species level ($p = 0.00051$). Adonis (PERMANOVA) analysis and ANOVA showed that there were significant differences in fungal structure among groups ($p = 0.001$). Finally, FUNGuild functional classification predicted that saprotrophs were the most abundant taxa in the GC group.

Conclusions: This study revealed GC-associated mycobiome dysbiosis characterized by an altered fungal composition and ecology and demonstrated that *C. albicans* can be a fungal biomarker for GC. In addition, *C. albicans* may mediate GC by reducing the diversity and richness of fungi in the stomach, contributing to the pathogenesis of GC.

Background

Gastric cancer (GC) is the fourth most common malignancy and one of the main causes of cancer-related deaths worldwide [1]. The majority of GC cases are the intestinal type of noncardia gastric cancer, which undergoes a predictable histological progression from atrophic gastritis (AG) to intestinal metaplasia (IM) and eventually to GC[2]. Initially, *Helicobacter pylori* infection causes inflammation of the gastric mucosa and destruction of related hydrochloric acid secreting glands, leading to atrophic gastritis [3]. Atrophic gastritis is a chronic inflamed, hypochloremic state, which may cause GC. Although it is known that *H. pylori* infection contributes to this cascade, only approximately 1–3% of infected individuals subsequently develop GC [4, 5]. Some host-related factors mentioned in the current research, including age, smoking status, genetic susceptibility and environmental factors, such as consumption of a high-salt diet and smoked foods containing nitrates, as well as microbial infections, have been shown to contribute to gastric carcinogenesis[6]. However, the relationship between gastric microbial components (such as fungi) other than *H. pylori* and GC has not been fully explored.

Over the past decade, due to the difficulty in culturing the commensal microorganisms that reside in the stomach, compared with intestinal flora, gastric flora studies are few, with only recent increases in studied on this topic [7]. In recent years, combined with advances in PCR techniques and metagenomics, the robust microbiome of the stomach has attracted extensive attention [8]. Most of the research efforts on the microbiome have focused on characterizing bacteria in healthy and diseased states, while the relatively low abundance of nonbacterial components has been neglected because of various technical challenges ranging from sample preparation to inadequate reference databases. Studies have provided evidence that bacteria, mainly the phyla *Proteobacteria, Firmicutes, Actinobacteria* and *Fusobacteria* [9, 10], can be regularly detected in gastric biopsies with gastric microbial dysbiosis associated with GC. Although *H. pylori* is still the main risk factor for histological changes, the chance of evolving GC after infection is not high, indicating that the presence of other components plays a key role in the development of GC.

With the advancement of high-throughput sequencing technology, sequencing methods provide access to the fungi of the gastric mycoflora. Genomic equivalence estimates that the fungal composition of the mammalian microbiota comprises less than 1% of all commensal microbial species, but fungi are significantly larger than bacteria in cell size and possess specialized metabolic gene clusters in response to specific ecological needs. Emerging research has revealed that fungi play a stable role in the development and maintenance of the host immune system and can be altered in various diseases [11, 12]. The latest Nature journal reports that fungi, like bacteria, can also be transferred from the intestine to the pancreas, and related changes in the fungal mycobiome promote pancreatic oncogenesis[13]. Thus, the dynamic exploration of the changes in the composition of gastric fungi in the progression from health to GC not only provides direction for future high-throughput fungal sequencing research on tumors but is also essential for further investigating the mechanisms of gastric carcinogenesis other than *Helicobacter pylori*.

In this study, we characterized fungal compositional and ecological changes by analyzing metagenomic sequences in cancer lesions and adjacent noncancerous tissues of 45 patients with GC. *Candida albicans* was also discovered as a fungal indicator for GC. For the first time, we used ITS sequencing to demonstrate the importance of fungi in the pathogenesis of GC, providing a theoretical scientific basis for the development of potential prevention and treatment strategies.

Methods

Sample collection and PCR amplification

A total of 90 samples were obtained from 45 pairs of patients diagnosed with GC at the First Affiliated Hospital of China Medical University, Shenyang, China. Surgical biopsies were obtained from sites of cancer lesions and adjacent noncancerous tissues in each patient. All specimens were stored at -80 °C until DNA extraction. In addition, subjects provided informed consent for obtaining study specimens, and

the study was approved by the Clinical Research Ethics Committees of the First Affiliated Hospital of China Medical University.

Microbial DNA was extracted using HiPure DNA Kits (Magen, Guangzhou, China) according to the manufacturer's protocols. The internal transcribed spacer (ITS) of the ITS2 region between the 18S and 28S genes of the ribosomal DNA gene was amplified by PCR (94 °C for 2 min, 30 cycles at 98 °C for 10 s, 62 °C for 30 s, and 68 °C for 30 s, and a final extension at 68 °C for 5 min) using the fungal-specific primers ITS3_KYO2: GATGAAGAACGYAGYRAA and ITS4: TCCTCCGCTTATTGATATGC [14]. PCRs were performed in triplicate in a 50-μL mixture containing 5 μL of 10 × KOD buffer, 5 μL of 2 mM dNTPs, 3 μL of 25 mM MgSO4, 1.5 μL of each primer (10 μM), 1 μL of KOD polymerase, and 100 ng of template DNA. The related PCR reagents used in the experiment were from TOYOBO, Japan.

Metagenomics sequencing

Amplicons were extracted from 2% agarose gels, purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) according to the manufacturer's instructions and quantified using the ABI StepOnePlus Real-Time PCR System (Life Technologies, Foster City, USA). The purified amplicons were pooled in equimolar amounts and paired-end sequenced (PE250) on an Illumina platform according to standard protocols. The raw reads were deposited into the NCBI Sequence Read Archive (SRA) database.

Quality control and read assembly

Raw data containing adapters or low-quality reads affect subsequent assembly and analyses. Thus, to obtain high-quality clean reads, the raw reads were further filtered according to the following rules using FASTP [15] (version 0.18.0): reads containing more than 10% of unknown nucleotides-(N) and reads with less than 50% of bases with a quality value (Q-value) > 20 were removed. Paired-end clean reads were merged as raw tags using FLASH [16] (version 1.2.11) with a minimum overlap of 10 bp and a mismatch error rate of 2%.

The noisy sequences of raw tags were filtered using the QIIME [17] (version 1.9.1) pipeline based on specific filtering conditions [18] to obtain high-quality clean tags. The filtering conditions were as follows: briefly, raw tags from the first low-quality base site where the number of bases in the continuous low-quality value (the default quality threshold is < = 3) reached the set length (the default length is 3) were broken. Then, tags whose continuous high-quality base length was less than 75% of the tag length were filtered.

OTU and community composition analyses

The effective tags were clustered into operational taxonomic units (OTUs) with at least 97% similarity using the UPARSE [19] (version 9.2.64) pipeline. The tag sequence with the highest abundance was selected as the representative sequence within each cluster. For the analyses between groups, Venn diagram-based analyses were performed in the R project VennDiagram package [20] (version 1.6.16), and an upset plot was developed in the R project UpSetR package [21] (version 1.3.3) to identify unique and common OTUs.

The representative sequences were classified into organisms by a naive Bayesian model using the RDP classifier [22] (version 2.2) based on the ITS2[23] database (version update_2015), with a confidence threshold value of 0.8. The abundance statistics of each taxa were visualized using Krona [24] (version 2.6). The stacked bar plot of the community composition was visualized in the R project ggplot2 package [25] (version 2.2.1). Circular layout representations of species abundance were graphed using Circos [26] (version 0.69-3). A heatmap of species abundance was plotted using the pheatmap package (version 1.0.12) [27] in the R project.

Statistical analysis

The random forest package [28] (version 4.6.12), pROC package [29] (version 1.10.0) and labdsv package [30] (version 2.0–1) were used in the R project. A ternary plot of species abundance was plotted using the R ggtern package [31] (version 3.1.0). Chao1, Simpson and all other alpha diversity indexes were calculated in QIIME [17] (version 1.9.1). Comparisons of the alpha indexes between groups were performed with Welch's t-test and Wilcoxon rank test using the R project [32] (version 2.5.3).

The R project [32] (version 2.5.3) was also used to analyze the data based on multivariate statistical techniques, including Jaccard and Bray-Curtis distance matrixes, principal component analysis (PCA), principal coordinate analysis (PCoA) and nonmetric multidimensional scaling (NMDS) of weighted UniFrac distances, and the results were plotted in the R project ggplot2 package [25] (version 2.2.1). Welch's t-test, Wilcoxon rank test, Adonis (also called PERMANOVA) and ANOSIM test were performed using the R project, and the functional groups (guilds) of the fungi were inferred using FUNGuild[33] (version 1.0).

Results

Gastric fungal dysbiosis is associated with GC

We evaluated 90 samples from 45 pairs of patients and divided them into a GC group and a control group (adjacent noncancerous tissue) for comparison. We also analyzed the clinical characteristics closely related to GC and found no significant differences. The detailed characteristics of the patients are shown in Table S1. We first assessed and compared the fungal composition in the specimens. The PCA showed that the GC and control groups aggregated separately, revealing that the gastric mucosal fungal community discriminated GC and the control into two significantly distinct groups. The GC group

exhibited more unique fungal profiles than the control group (Fig. 1a, Table S2). To clarify the OTU crossover between different groups, we used a Venn diagram to indicate the differences among the groups according to OTU abundance. We found that both groups shared a total OTU abundance of 207. Simultaneously, the GC group showed a lower OTU abundance than the control group (Fig. 1b). Based on these OTU clustering results, it is suggested that alterations in stomach fungal composition may lead to gastric carcinogenesis.

Taxonomic coverage and alterations of fungi in GC

For the distribution of fungal taxa, in both the GC and control groups, the phylum *Ascomycota* was the dominant mycoflora, and *Basidiomycota* was considered to be the second most abundant phylum (Fig. 2a). The corresponding species abundance heat map is shown in Fig. 2b. We further analyzed the differences at the lower taxonomic level of class, finding a significant depletion of *Eurotiomycetes*, *Agaricomycetes*, *Tremellomycetes*, *Microbotryomycetes* and *Mortierellomycetes* and enrichment of *Saccharomycetes* and *Dothideomycetes* in the GC group compared with the control group (Fig. 2c). At the family level, we found 17 fungi with significant differences (Table S3), so we only showed data with a P value less than 0.01. *Pseudeurotiaceae, Trimorphomycetaceae, Chaetomiaceae* and *Aspergillaceae* were significantly decreased in the GC group, while *Saccharomycetales_fam_Incertae_sedis* and *Pleosporaceae* were increased, compared to the control (Fig. 2d). Furthermore, at the genus level, there were 15 different fungi between the two groups (Table S4); 2 fungal genera were enriched in the GC group, including *Candida* ($p = 0.000246$) and *Alternaria* ($p = 0.00341$), while *Saitozyma* ($p = 0.002324$) and *Thermomyces* ($p = 0.009158$) were decreased, compared to the control (Fig. 2e).

Candida albicans as a fungal indicator species for GC

To better identify fungal taxa with value as potential GC indicators, we evaluated fungal alterations at the species level. We initially used Welch's t test and found that there were 13 species with significant differences in the mean abundance when comparing the two groups (Table S5). Then, the Wilcoxon rank sum test was applied to determine whether the median species abundance was statistically significant, and we confirmed that 59 species had significant differences between the two groups (Table S6). The species with higher contents and greater than two-fold changes in abundance were selected for the next analysis.

With the Welch's t test, *Candida albicans* ($p = 0.000015$) and *Fusicolla acetilerea* ($p = 0.01691$) were increased, while *Aspergillus montevidensis* ($p = 0.001437$), *Saitozyma podzolica* ($p = 0.002324$) and *Penicillium arenicola* ($p = 0.00722$) were obviously decreased in the GC group (Fig. 3a). With the Wilcoxon rank sum test, the abundance of *C. albicans* ($p = 0.000072$), *Arcopilus aureus* ($p = 0.040759$) and *Fusicolla aquaeductuum* ($p = 0.026626$) was higher in the GC group, while *Candida glabrata* ($p = 0.014443$) and *Aspergillus montevidensis* ($p = 0.000586$) were less abundant, compared to the control (Fig. 3b). These results demonstrated that *Candida albicans* was significantly elevated in the GC group (p

< 0.0001). Next, we dynamically displayed the composition of species at different classification levels through the species composition pie chart and found that the abundance of *C. albicans* at the species level accounted for 22% (Fig. 3c, Table S7). We evaluated the accuracy based on the ROC curve and observed an AUC value of 0.743 (Fig. 3d). Random forest analysis was used to screen potential indicator species, and the values of the Gini index (Fig. 4a) and the mean decrease in accuracy (Fig. 4b) were the largest for *C. albicans*. Combined with the indicator analysis, we comprehensively considered the strong indicator ability of *C. albicans* among the groups (Fig. 4c). These results all indicated that *C. albicans* had an obvious effect in distinguishing GC and non-GC tissues and can be used as a biomarker with a certain degree of accuracy.

Altered fungal microbiota diversity in GC

Next, we conducted a diversity analysis to further understand the species richness and flora structure among the groups. Alpha diversity indexes (Chao1, ACE, Sobs, Shannon, Simpson and Good's Coverage) were significantly reduced in the GC group compared with those of the control (Fig.S1, Table S8). Briefly, we measured fungal alpha diversities and determined whether, through a t test (Fig. 5a-e) or rank sum test, five indexes, namely, the Chao1, ACE, Sobs, Shannon and Simpson indexes, were significantly different between the GC and control groups ($P < 0.05$) (Table 1).

Table 1
Difference of alpha diversity index between the GC and control groups.

Group	Testing method	Index	P value
GC-VS-Control	**T-test**	Sobs	0.001267
		Shannon	0.000003
		Simpson	0.000156
		Chao	0.001763
		Ace	0.006286
	Wilcoxon	Sobs	0.0118
		Shannon	0.000001
		Simpson	0.000001
		Chao	0.025431
		Ace	0.044121

We used PCoA to analyze two classic beta diversity indexes, the Jaccard distance index (Fig. 5f) and the Bray abundance index (Fig. 5g), and confirmed separate clusters for the GC and control groups at the species level. To overcome the shortcomings of linear models (PCA, PCoA) and better reflect the nonlinear

structure, we evaluated the accuracy of the model through NMDS stress values. We ensured the reliability of the model, confirming that the stress values of the Jaccard and Bray indexes were less than 0.1 (Fig. 5h). The significant difference of the two indexes between groups was shown by the Wilcoxon rank sum test at the genus level (P = 0.00051, Fig. 5i-j). We then evaluated and verified the fungal composition in our groups. Both Adonis (PERMANOVA) analysis (Table S9) and the ANOSIM test (Fig. 5k) revealed that there were significant differences in fungal structure between the GC and control groups. Combining the two diversity index results, our analysis suggested that with gastric carcinogenesis, the richness of the related fungal composition decreases, and the structure of the fungal community is quite different.

Ecological guilds of sampled taxa

Based on the OTU abundance, we used FUNGuild to perform functional classification prediction. The fungal taxa were grouped into 83 ecological guilds, and the most diverse guild was undefined saprotrophs (Fig.S2a). In addition, trophic mode divided fungal taxa into 9 types, of which the most diverse type was saprotrophs (Fig. S2b). In particular, heatmaps were drawn to describe the functional predictions under the two analytical methods, as shown in Fig. 6a and Fig. 6b, respectively. Thus, our analyses showed that a symbiotic ecological relationship may be important for the homeostasis of gastric fungi, while malnutrition might provide a suitable environment for the fungal dysbiosis of GC and ultimately contribute to the negative effects of gastric carcinogenesis.

Discussion

Gastric cancer causes one of the major types of digestive tract tumor worldwide [1]. After the continuous development of high-throughput sequencing technology, research on the correlation between gastric flora (other than *Helicobacter pylori*) and GC has gradually emerged. In this study, we tried to describe the fungal spectrum associated with GC, which has not been explained to date; the focus was on gastric fungal dysbiosis caused by GC. Compared with fecal samples, the colonization performance of tissue samples can better demonstrate the dynamic changes in the surrounding environment for gastric carcinogenesis. Therefore, we analyzed the ITS metagenome sequences of cancer lesions and adjacent noncancerous tissues to investigate the composition and ecological alterations of fungi associated with GC and identify fungal indicators. To ensure that the most effective data were clustered into OTUs, we filtered low-quality reads, and assembled and refiltered the data. After obtaining the OTUs, under the condition that the GC and control groups were effectively grouped, we carried out species identifications and alpha and beta diversity analysis, and compared differences between groups. *Candida albicans* was identified for the first time as a key fungus that can be used to distinguish between GC and control groups. We also combined FUNGuild functional annotation to study fungal functions from other ecological perspectives. For the first time, we showed the characteristics of the fungal flora in the stomach tissues of GC patients, demonstrating fungal malnutrition in the GC ecosystem and proving that *C. albicans* can be used as a biomarker with a certain degree of accuracy.

We clarified specific fungal composition changes in GC. Overall, the GC group showed a lower OTU abundance. At the phylum level, *Ascomycota* was the most enriched in the GC group compared with the control group, while *Basidiomycota* was less enriched. We further analyzed the differences at lower taxonomic levels and finally, at the species level, confirmed that *C. albicans, Fusicolla acetilerea, Arcopilus aureus* and *Fusicolla aquaeductuum* were excessively colonized in the GC tissue. At present, *C. albicans* is the most researched of these organisms with regard to its role in various diseases. This species normally exists in the body and does not cause damage. However, when the host's defense capacity is weakened, *C. albicans* will cause disease. Therefore, *C. albicans* is recognized as an opportunistic pathogen. Since immunosuppression caused by cancer chemotherapy promotes *C. albicans* infection, the relationship between *C. albicans* and cancer development or progression has been widely reported. For example, for hematological malignancies or solid tumors, up to 35% of patients with underlying disease have candidiasis, and the most common underlying disease among patients with candidiasis is also solid tumor [34]. *Candida albicans* can produce carcinogenic nitrosamines, which can cause abnormal proliferative changes in oral epithelial cancer [35]. The risk of malignant transformation of oral leukoplakia is higher than that of oral lichenoid lesions, and *C. albicans* strains isolated from patients can produce more carcinogenic acetaldehyde in ethanol [36]. The role of *C. albicans* in tumor adhesion and metastasis has been associated with TNF-α and IL-18 [37–39]. Recently, Bertolini et al. confirmed that *C. albicans* induced mucosal bacterial malnutrition and promoted invasive infection [40].

Notably, we first confirmed the indicative role of *C. albicans* in GC. In our study, compared with the control, the species richness of *C. albicans* occupied 22% in the GC group. Both the Welch's t test and Wilcoxon rank sum test confirmed that *C. albicans* was significantly more abundant in the GC group than the control group. In addition, the ROC curve showed that the AUC value of *C. albicans* was 0.743. Combined with the results of the Gini index and the mean decrease in accuracy, all results indicated that *C. albicans* could be used as a biomarker with a certain degree of accuracy. By diversity analysis, compared with the control group, the GC group had a decrease in species richness, diversity and uniformity. The structure of the species flora between the groups also showed a significant change, suggesting that *C. albicans* has an adverse effect on the diversity and richness of the stomach microbiome. Aykut et al. stated that identifying the species most associated with cancer may guide future attempts to use targeted antifungal drugs to slow tumor growth and avoid side effects and reported *Malassezia* as a pathogenic fungus associated with pancreatic cancer that promotes pancreatic oncogenesis via activation of MBL[13]. Our discovery that *C. albicans* may have contributed to the pathogenesis of GC not only lays a scientific foundation for the exploration of innovative therapies for GC but also provides a new idea for treating specific patients by adjusting their intestinal microbial flora as an adjuvant therapy or developing immunotherapies for targeted control of fungal infections, which is worthy of further study.

Due to the current lack of fungal genomic data, we integrated published article data and used FUNGuild to predict fungal functions from other ecological perspectives based on OTU abundance. The guild classification revealed that the most diverse guilds were undefined saprotrophs. Simultaneously, the trophic mode implied that the most diverse fungal type was the saprotrophs. Our analysis clarified the

importance of fungal homeostasis in the stomach and showed that fungal dysbiosis eventually promotes the occurrence of GC.

Conclusions

In conclusion, compared with most studies focusing on the bacterial spectrum associated with GC, our study described the gastric fungal dysbiosis in gastric carcinogenesis for the first time and showed that *C. albicans* can be used as a fungal marker for GC. In addition, *C. albicans* may mediate GC by reducing the diversity and richness of fungi in the stomach, contributing to the pathogenesis of GC. We also revealed the importance of homeostasis for gastric fungi. Additional analysis investigating the potential role of *C. albicans* in gastric carcinogenesis is warranted to delineate its use as a noninvasive biomarker for GC diagnosis.

Abbreviations

GC
gastric cancer
AUC
area under the receiver operating curve
C. albicans
Candida albicans
AG
atrophic gastritis
IM
intestinal metaplasia
ITS
internal transcribed spacer
SRA
Sequence Read Archive
OUTs
operational taxonomic units
PCA
principal component analysis
PCoA
principal coordinate analysis
NMDS
nonmetric multidimensional scaling

Declarations

Ethics approval and consent to participate

The study was approved by the Clinical Research Ethics Committees of the First Affiliated Hospital of China Medical University.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Outstanding Youth Fund Project of Fujian Province (2018D0016).

Authors' contributions

MYZ, XHH and YXS designed the study. MYZ, YBX and JBZ conducted the experiments. YBX, ZG, JSM and ZXW analyzed the results. XHH and YXS collected the clinical samples. MYZ, YBX and XHH collectively conceptualized the manuscript. YXS edited the manuscript and provided critical comments. All the authors reviewed and approved the final version of the manuscript.

Acknowledgments

We would like to thank the First Affiliated Hospital of China Medical University for providing clinical samples and the entire research team for their contributions and support.

Availability of supporting data

The raw reads were deposited into the NCBI SRA database (Accession Number: SRA: SRP276371 and Bioproject PRJNA650666).

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69:7–34.
2. Park YH, Kim N. Review of atrophic gastritis and intestinal metaplasia as a premalignant lesion of gastric cancer. J Cancer Prev. 2015;20:25–40.
3. Amieva M, Peek RM Jr. Pathobiology of Helicobacter pylori-Induced Gastric Cancer. Gastroenterology. 2016;150:64–78.
4. Stewart OA, Wu F, Chen Y. The role of gastric microbiota in gastric cancer. Gut Microbes. 2020;11:1220–30.
5. Peek RM Jr, Crabtree JE. Helicobacter infection and gastric neoplasia. J Pathol. 2006;208:233–48.
6. Compare D, Rocco A, Nardone G. Risk factors in gastric cancer. Eur Rev Med Pharmacol Sci. 2010;14:302–8.
7. Sheh A, Fox JG. The role of the gastrointestinal microbiome in Helicobacter pylori pathogenesis. Gut Microbes. 2013;4:505–31.
8. Monstein HJ, Tiveljung A, Kraft CH, Borch K, Jonasson J. Profiling of bacterial flora in gastric biopsies from patients with Helicobacter pylori-associated gastritis and histologically normal control individuals by temperature gradient gel electrophoresis and 16S rDNA sequence analysis. J Med Microbiol. 2000;49:817–22.
9. Nardone G, Compare D. The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases? United European Gastroenterol J. 2015;3:255–60.
10. Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, Wu WK, Wong SH, Chen Z, Sung JJY, Yu J. Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut. 2018;67:1024–32.
11. Coker OO, Nakatsu G, Dai RZ, Wu WKK, Wong SH, Ng SC, Chan FKL, Sung JJY, Yu J. Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. Gut. 2019;68:654–62.
12. Lai GC, Tan TG, Pavelka N. The mammalian mycobiome: A complex system in a dynamic relationship with the host. Wiley Interdiscip Rev Syst Biol Med. 2019;11:e1438.
13. Aykut B, Pushalkar S, Chen R, Li Q, Abengozar R, Kim JI, Shadaloey SA, Wu D, Preiss P, Verma N, et al. The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. Nature. 2019;574:264–7.
14. Toju H, Tanabe AS, Yamamoto S, Sato H. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PLoS One. 2012;7:e40863.
15. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
16. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27:2957–63.
17. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data.

Nat Methods. 2010;7:335–6.
18. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10:57–9.
19. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10:996–8.
20. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.
21. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33:2938–40.
22. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7.
23. Ankenbrand MJ, Keller A, Wolf M, Schultz J, Forster F. ITS2 Database V: Twice as Much. Mol Biol Evol. 2015;32:3030–2.
24. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 2011;12:385.
25. ggplot2. An implementation of the Grammar of Graphics [URL: http://CRAN. R-project. org/package = ggplot2].
26. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.
27. Kolde RKMR. Kolde M R. Package 'pheatmap'[J]. Package 'pheatmap'[J] 2015, 1(7).
28. Liaw AWM. Classification and regression by randomForest[J]. R news. 2002;2(3):18–22.
29. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
30. Roberts DW RMDW: Package 'labdsv'[J]. Ordination and Multivariate 2016.
31. Hamilton NE. FM: ggtern: Ternary diagrams using ggplot2[J]. Journal of Statistical Software 2018:87(81): 81–17.
32. Oksanen JBFG, Kindt R, et al.: Vegan: community ecology package.. 2010, 23: 2010.
33. Nguyen NH, et al. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology. 2016;20:241–8.
34. Zirkel J, Klinker H, Kuhn A, Abele-Horn M, Tappe D, Turnwald D, Einsele H, Heinz WJ. Epidemiology of Candida blood stream infections in patients with hematological malignancies or solid tumors. Med Mycol. 2012;50:50–5.
35. Sanjaya PR, Gokul S, Gururaj Patil B, Raju R. Candida in oral pre-cancer and oral cancer. Med Hypotheses. 2011;77:1125–8.
36. Gainza-Cirauqui ML, Nieminen MT, Novak Frazer L, Aguirre-Urizar JM, Moragues MD, Rautemaa R. Production of carcinogenic acetaldehyde by Candida albicans from patients with potentially

malignant oral mucosal disorders. J Oral Pathol Med. 2013;42:243–9.

37. Ramirez-Garcia A, Arteta B, Abad-Diaz-de-Cerio A, Pellon A, Antoran A, Marquez J, Rementeria A, Hernando FL. Candida albicans increases tumor cell adhesion to endothelial cells in vitro: intraspecific differences and importance of the mannose receptor. PLoS One. 2013;8:e53584.
38. Rodriguez-Cuesta J, Hernando FL, Mendoza L, Gallot N, de Cerio AA, Martinez-de-Tejada G, Vidal-Vanaclocha F. Candida albicans enhances experimental hepatic melanoma metastasis. Clin Exp Metastasis. 2010;27:35–42.
39. Ramirez-Garcia A, Rementeria A, Aguirre-Urizar JM, Moragues MD, Antoran A, Pellon A, Abad-Diaz-de-Cerio A, Hernando FL. Candida albicans and cancer: Can this yeast induce cancer development or progression? Crit Rev Microbiol. 2016;42:181–93.
40. Bertolini M, Ranjan A, Thompson A, Diaz PI, Sobue T, Maas K, Dongari-Bagtzoglou A. Candida albicans induces mucosal bacterial dysbiosis that promotes invasive infection. PLoS Pathog. 2019;15:e1007717.

Figures

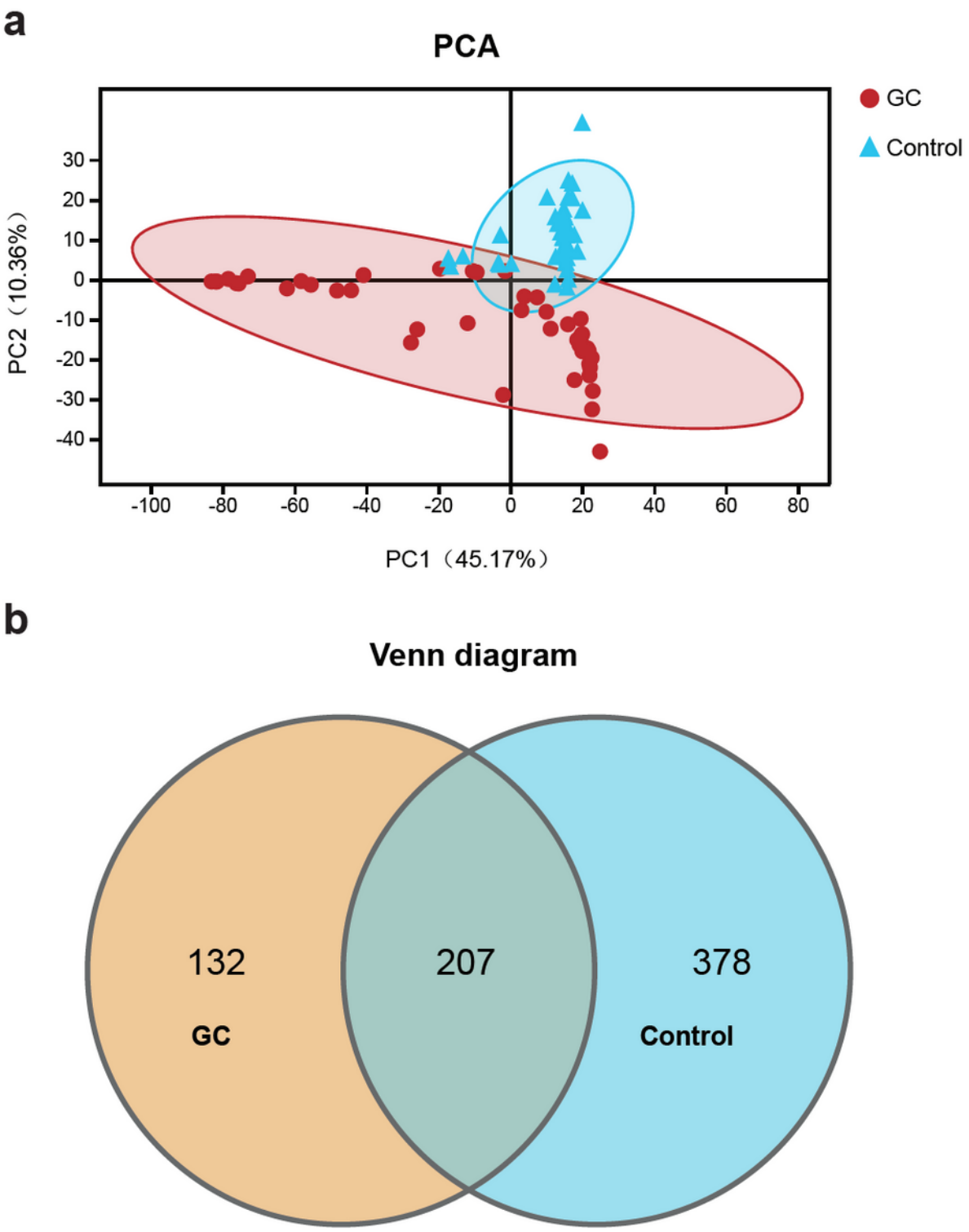

Figure 1

Classification and distribution of fungi in the stomachs of gastric cancer (GC) patients. (a) Through the principal component analysis (PCA) dynamic display, GC (n=45) and control (n=45) samples showed clustering distributions. PC1 and PC2 represent the first two main components, and they reflect the contribution to the sample difference, expressed as a percentage. (b) Based on the OTU abundance, Venn

diagram analysis was performed. Unique OTUs between the GC (yellow) and control (blue) groups was found as well as common OTUs (green) between the two groups.

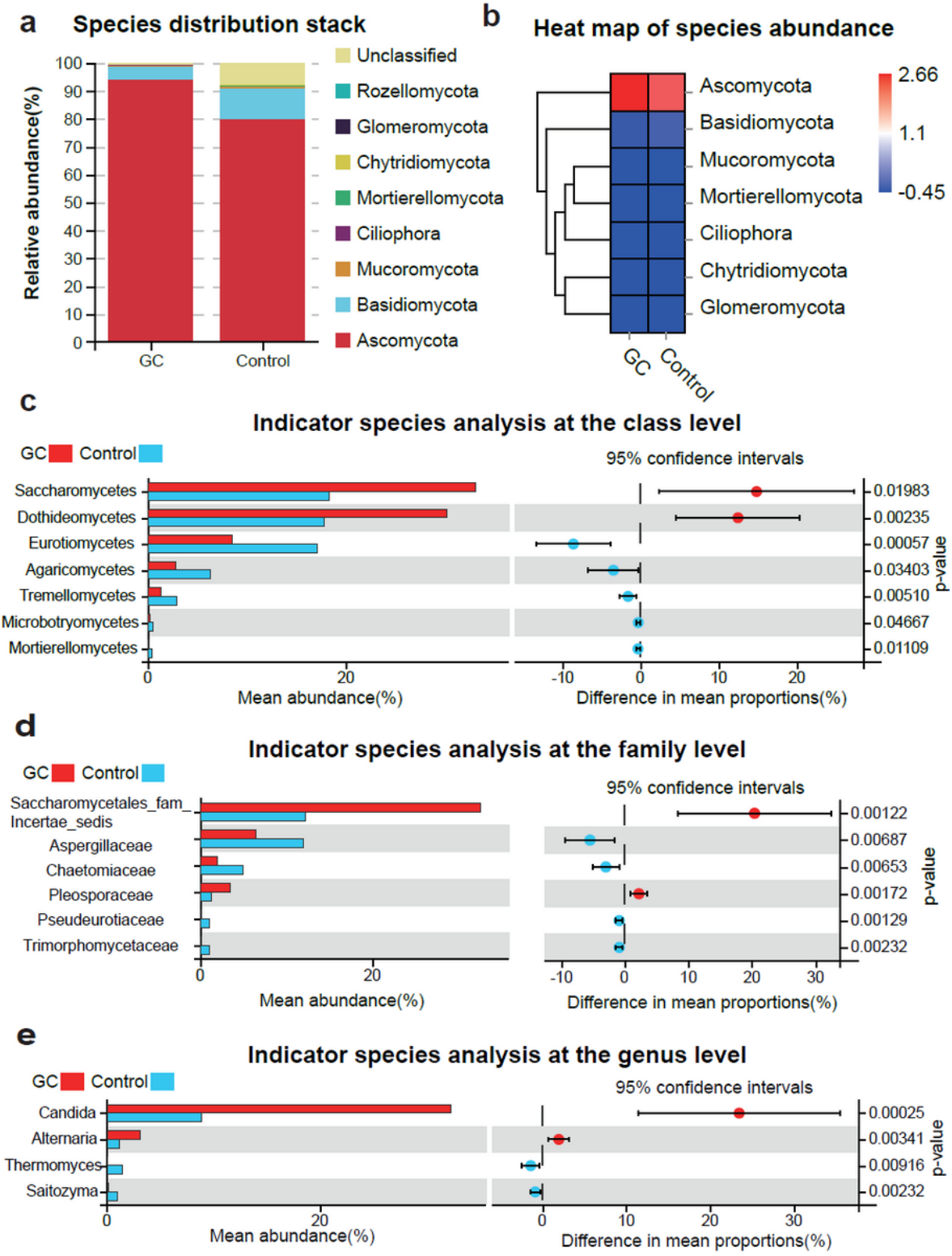

Figure 2

Changes in the fungal composition in the stomachs of gastric cancer (GC) patients. (a) Relative abundance of dominant gastric fungal phyla in the GC and control groups. The dominant phyla were Ascomycota and Basidiomycota in both groups. (b) The corresponding heatmap also shows changes in

the fungal phyla in the GC and control groups. Differences in fungal composition and abundance between GC (n=45) and the control (n=45) were detected using Welch's t test. The variation in the relative abundance of species represented in different groups was demonstrated graphically. Differences in OTUs appear in the left rows, and the corresponding P values are shown in the right rows. (c) Differentially abundant fungal classes between the GC and control groups. OTUs and taxa differences are shown with p-values less than 0.05. Differentially abundant fungal families (d) or genera (e) between the GC and control groups. OTUs and taxa differences are shown with p-values less than 0.01.

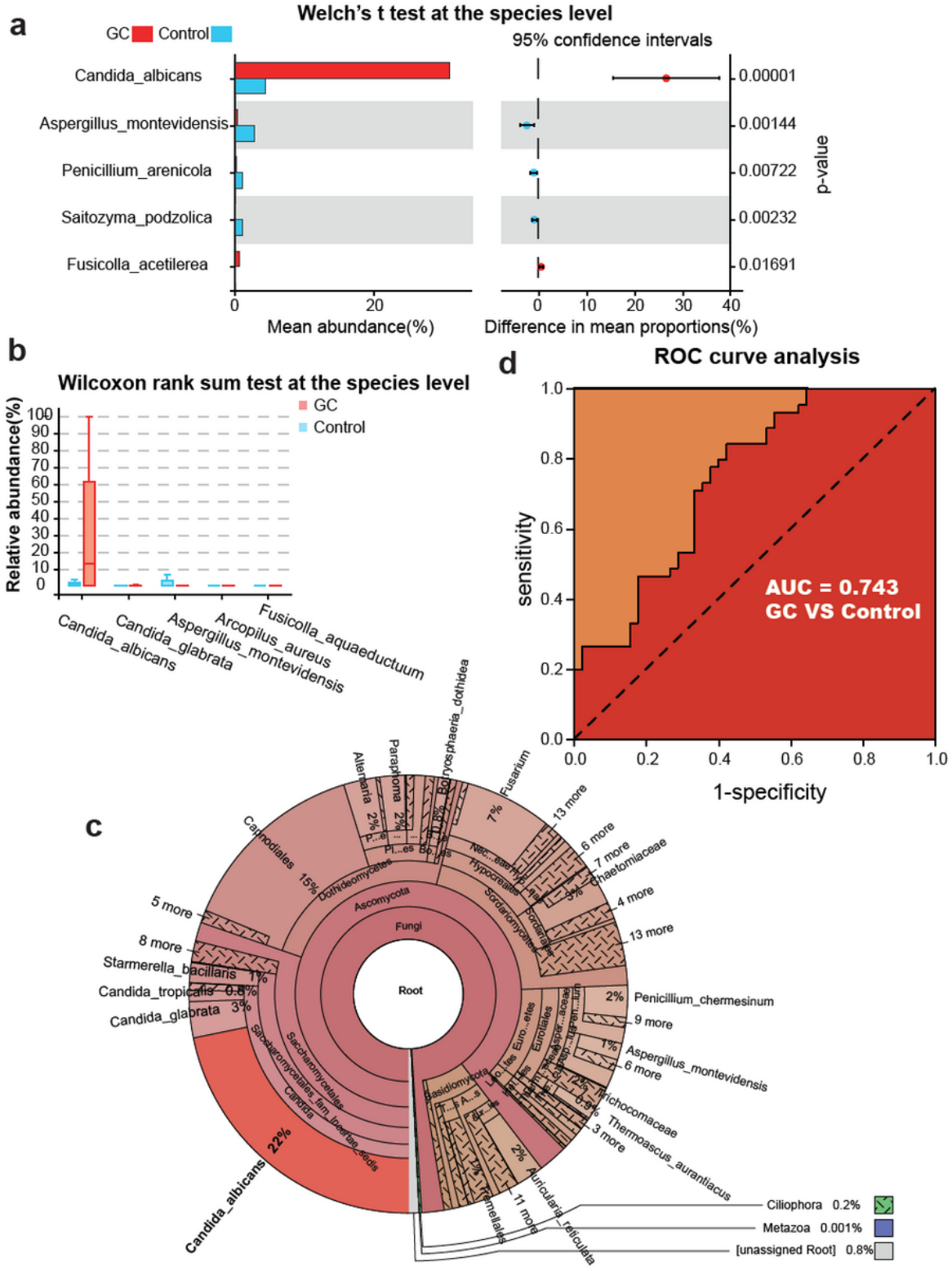

Figure 3

Changes in the fungal composition in the stomachs of gastric cancer (GC) patients. (a) Relative abundance of dominant gastric fungal phyla in the GC and control groups. The dominant phyla were Ascomycota and Basidiomycota in both groups. (b) The corresponding heatmap also shows changes in the fungal phyla in the GC and control groups. Differences in fungal composition and abundance between GC (n=45) and the control (n=45) were detected using Welch's t test. The variation in the relative abundance of species represented in different groups was demonstrated graphically. Differences in OTUs appear in the left rows, and the corresponding P values are shown in the right rows. (c) Differentially abundant fungal classes between the GC and control groups. OTUs and taxa differences are shown with p-values less than 0.05. Differentially abundant fungal families (d) or genera (e) between the GC and control groups. OTUs and taxa differences are shown with p-values less than 0.01.

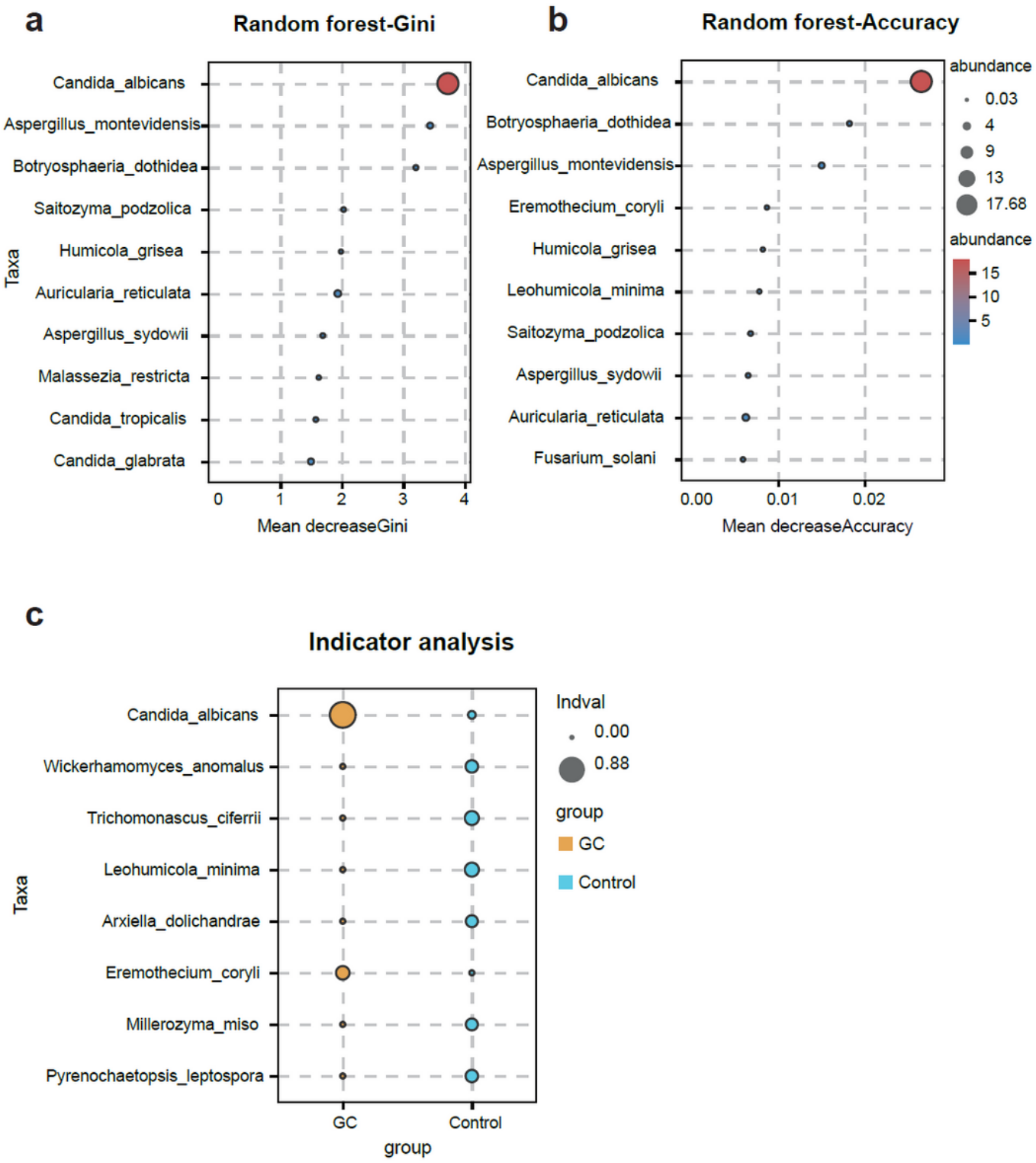

Figure 4

Candida albicans has a strong indication ability. Using the random forest algorithm to calculate the contribution of C. albicans to the grouping difference at the species level, it is found that the Gini index (a) and average accuracy (b) values were both largest for C. albicans. (c) The indicator analysis considers the frequency and abundance of C. albicans between groups.

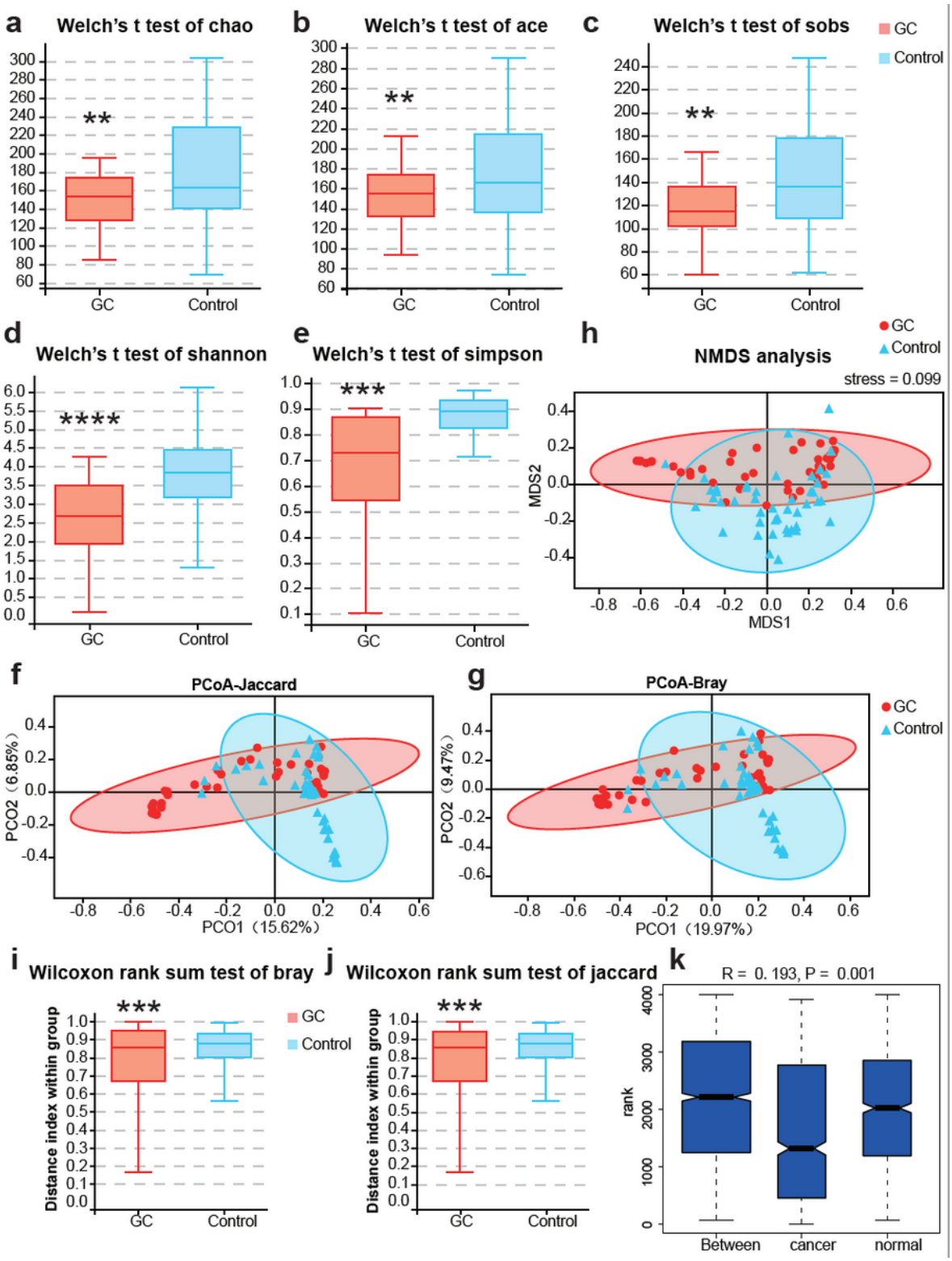

Figure 5

Changes in fungal flora diversity in GC. Hypothesis tests of the alpha diversity index through Welch's t test, Chao1 (a), ACE (b), Sobs (c), Shannon (d) and Simpson (e) diversity indexes between the GC (n=45) and control (n=45) groups confirmed that there were significant differences in species diversity between groups. Principal coordinate analysis (PCoA) of Jaccard distances (f) or Bray-Curtis distances (g) showed the stratification of GC (n=45) from control (n=45) samples by their fungal compositional profiles. (h)

Nonmetric multidimensional scaling (NMDS) analysis of the fungal compositional profiles stratified GC (n=45) from control (n=45) samples. A stress value less than 0.1 indicates that the model grouping is reliable. At the genus level, the Wilcoxon rank sum test was used to judge the significant difference between the Bray-Curtis distance (i) and Jaccard distance (j), and the degree of difference in fungal flora structure within the groups was compared. (k) Based on the distance index ranking, ANOSIM (analysis of similarities) confirmed that the distance between groups was significantly greater than the distance within groups, indicating that the flora structure of different groups was significantly different. **P<0.01, ***P<0.001, ****P<0.0001.

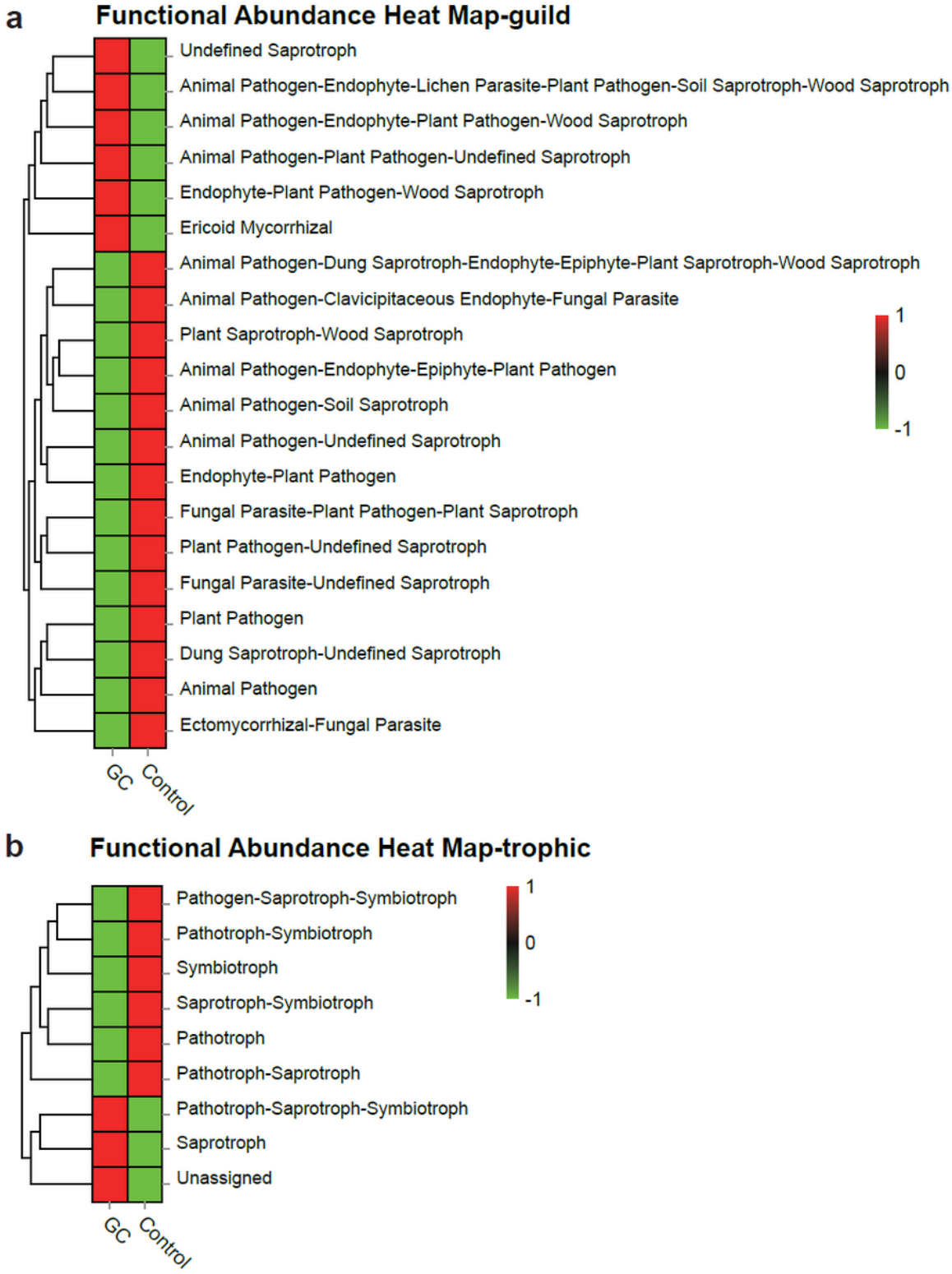

Figure 6

Saprotrophs are the most common functional category associated with GC. Based on the OTU abundance, fungal functional annotation was carried out using FUNGuild. Using functional groups (guilds), fungi were divided into categories based on their absorption and utilization of environmental resources. The three major categories and twelve subcategories of fungi distinguished the GC (n=45) and control groups (n=45) at the guild (a) and trophic (b) levels.

Supplementary Files

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

- TableS9.docx
- TableS8.docx
- TableS7.xls
- TableS6.xls
- TableS5.docx
- TableS4.docx
- TableS3.docx
- TableS2.xls
- TableS1.docx
- FigureS2.pdf
- FigureS1.pdf