**Multi-cohort fecal metagenomic analysis reveals the altered fungal signatures in colorectal cancer and the carcinogenic potential of *Aspergillus rambellii***

Yufeng Lin1, Yali Liu1, Nick Lung-Ngai Ting1, Xing Kang1, Thomas Ngai-Yeung Kwong1, Han Jing2, Yiwei Wang3, Harry Cheuk-Hay Lau1, Changan Liu1, Junjun She2, Sunny Hei Wong1,4, Joseph JY Sung1,4, Jun Yu1

1 Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China

2 Center for Gut Microbiome Research, Med-X Institute, The First Affiliated Hospital of Xi’an Jiao tong University, Xi’an, Shaanxi, China

3 School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

4 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

**Author contribution:**

**Yufeng LIN**, study design, dry lab analysis, wet lab validation, and written.

**Wet Lab:** Yali Liu, Xing Kang, Thomas Kwong.

**Dry Lab:** Yiwei Wang, Han Jing, Weixin Liu, SK, Yanqiang Ding.

**Clinical guidance:** Sunny Wong

**Manuscript revision:** Nick Ting, Thomas Kwong, Harry Lau.

**Study design and supervise:** Jun Yu

**Achnowleagment**

Weixin Liu, Siu-Kin Ng, Yanqiang Ding, Andrew Maltez Thomas, Nicola Segata

**Background** Gut microbiota alterations are associated with colorectal cancer (CRC) pathogenesis. However, the role of enteric fungi, an essential component of gut microbiota, in CRC remains largely elusive. We aim to characterize the contribution of enteric fungi to the development of CRC.

**Methods** We performed shotgun metagenomic analyses of 1325 fecal samples from seven public datasets and one new cohort (454 CRC patients, 350 Adenoma patients and 216 healthy controls). xx

**Results xxx** We identified 33 differentially abundant fungal species in CRC versus healthy individuals (false discovery rate (FDR) < 0.01), of which Aspergillus rambellii showed the most significant difference (FDR = 5.13E-18). In seven of the eight cohorts, the composite combined bacterial and fungal biomarkers classified CRC from healthy individuals with an AUC 1.44% - 10.60% relative change higher than the bacterial classifier. Among 14 biomarkers in the combined classifier, *A. rambellii* was the most important fungal species. Further abundance correlation analyses of the 64 differentially abundant species (33 fungi and 31 bacteria) showed that cross-kingdom interactions were associated with CRC. Particularly, strong differential correlations were shown between *A. rambellii* and two CRC-associated pathogens, *Fusobacterium* *nucleatum* and *Parvimonas* *micra*.

**Conclusions** This study revealed the mycobiome alterations in CRC particularly the enrichment of *A. rambellii* implying that the role of mycobiome in CRC is not negligeable.

Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related death globally1,2. Sporadic CRC, which arises without known contribution from germline mutations or significant family history, accounted for about 75% of CRC, implying the importance of environmental factors in CRC pathogenesis3. Recent studies have linked gut microbiota alteration to CRC occurence4–6. A meta-analysis with approximately 1,000 individuals from five cohorts has revealed the microbial signatures of genes specific for CRC7 and the association between the gut microbiome and choline degradation8. Even though >90% of the gut microbiome are composed of bacteria, a perturbed gut fungal composition has also been described to be associated with inflammatory bowel disease9, liver cirrhosis10 and CRC11. Fungi could influence the immunological responses of the host by dampening or promoting local inflammatory reactions9,10,12–14. The commensal fungi were shown to prevent hosts from colitis-associated colon cancer by prompting inflammasome activation and IL-18 maturation in murine model15. Therefore, it is apparent that fungi play a significant role in CRC development than our previous anticipation. Furthermore, previous research investigated how pathogenic fungi activated mannose-binding lectin to promote pancreatic ductal adenocarcinoma by activating the complement cascade1. However, apart from our previous study aiming to discover potential fungal biomarkers for CRC detection11, the compositional feature and the role of fungi in CRC pathogenesis remains unexplored.

In this study, we performed a multi-cohort fecal metagenomic analysis of eight available datasets from France16, Germany7,16, Australia17, America18,19, Canada19, Italy8, Chinese5, and Japan20. After rigorous and stringent data processing, a total of 1,329 samples from four continents, European, Occeanian, North American and Asian, including 525 healthy individuals, 350 adenoma patients, and 454 CRC patients, were included in this analysis. After consistent data reprocessing, we determined the compositional and ecologic alteraction of fungi in CRC. Our abundant differential analysis targeting the fungal genome revealed the significant enrichment of *Aspergillus rambellii* in CRC patients. Oncogenic function of *A. rambellii* in CRC was further validated in vitro and in vivo. We then evaluated and compared the differential correlations across the stages of colorectal carcinogenesis from healthy controls to adenoma, and CRC, respectively. *A. rambellii* and *Pichia kudriavzevii* exhibited the strong differential correlations with carcinogens (*F. nucleatum* and *P. micra*) and reported potential probiotics (*S. salivarius*, *A. hadrus* and *S. thermophilus*), respectively, in CRC compared with healthy controls. For clinical application, the trans-kingdoms CRC-screening models were 1.44%-10.60% relative change higher than pure bacterial models in seven of eight studies. And *A. rambellii* ranked fourth in the importance of trans-kingdoms model’s features, second only to the three famous carcinogens *P. micra*, *F. nucleatum*, and *G. morbillorum*. All these suggested that enteric fungi, especially*A. rambellii*might play a potential role in CRC carcinogenesis.

Methodology

Sample collection and data retrieva

Hong Kong cohort with CRC, adenoma patients and healthy controls

Recruitment criteria included presentations of digestive symptoms to the outpatient gastroenterology clinics and asymptomatic individuals 50 years or older receiving colonoscopy screening from the Chinese University of Hong Kong Jockey Club Bowel Cancer Education Centre. Stool samples were collected by participants and stored at –20°C within 4 hours. For long-term storage, all samples were stored at –80°C within 24 hours of stool collection. Total DNA was extracted from stool samples by using the QIAamp DNA Stool Mini Kit according to the manufacturer's instructions (Qiagen, Germany). All subjects had intact colonic at the time of stool collection. An independent Chinese cohort with 112 healthy individuals, 111 patients with CRC, and 197 patients with colorectal adenoma were recruited. A subset of these samples have been published in a previous research21.

Samples collected from NCBI dataset

Fecal shotgun metagenomic sequencing data of CRC-related studies from 2014 to 2020 with a minimum of 2 subject categories (CRC patients and healthy controls) were retrieved from the NCBI. Eight published cohorts and our recently completed but unpublished cohort were included in this meta-analysis; five of them also included adenoma patients8,16,17,19,20 (table 1 and supplementary table 1). We downloaded seven public fecal shotgun metagenomic CRC datasets from European Nucleotide Archive (ENA) using the following ENA identifiers: ERP005534 for Zeller et al.16, ERP008729 for Feng et al.17, PRJEB10878 for Yu et al.5, PRJEB12449 for Vogtmann et al.18, PRJNA389927 for Hanningan et al.19, PRJEB27928 for Wirbel et al.7, and SRP136711 for Thomas et al.8. The cohort from Yachida et al. was downloaded from the DNA Data Bank of Japan (DDBJ) with the Accession numbers: DRA006684 and DRA00815620. For our cohort, fecal metagenomic sequencing data were used from samples collected in Hong Kong from 2009 to 2012. A subset of samples in this cohort were published previously11. These nine studies were organized from eight countries and various sampling procedures, sample storage, and DNA extraction protocols.

Sample filtering

To ensure consistent and high-quality data, samples were subjected to filtering before analysis. Patients after surgery, or with ambiguous conditions (CRC, adenoma or healthy controls) were discarded. We only included the PCR-free cohort because the PCR-free kits could reduce bias and cell spike-in controls for a more accurate quantification22. Finally, we excluded the samples with low-alignment reads (less than 1,000,000), which might be due to low sequencing depth and host reads contamination. In the second part, we removed the outliers or suspected contaminated cases. These include samples with high-fungi composition (the fungi composition is more than 1% of total gut microbiota), low-Fungi composition (the fungi composition is less than 0.01% of the gut microbiota), and bacterial or fungal contaminated samples (a particular species constitutes more than 50% of the gut microbiota). Finally, the samples with low-fungal sequence depth (fungi aligned read counts less than 10,000 reads) would be discarded, which was consistent with a previous study revealing that fungi could not be detected in at least 30% of individuals23.

Sequence pre-processing and taxonomic profiling

We applied the KneadData default parameters for the quality control of all the metagenomic sequencing data. This separated microbial reads from the contaminated reads from the host or other user-defined sources using principled in silico methods. Next, taxonomic profiles were generated with the Kraken2 v2.0.9-beta across the custom database. Our custom library contained 9,543 bacterial and 909 fungal references from NCBI (https://www.ncbi.nlm.nih.gov/), FungiDB (https://fungidb.org/fungidb/), Ensemble (http://fungi.ensembl.org/index.html), and Broad Institute (<https://www.broadinstitute.org/>). The library was then established with the Jellyfish program by counting distinct 31-mer. We used the default parameters and discarded all reads with quality less than 20 and shorter than 50 nucleotides. Each query was classified to a taxon with the highest total hits of k-mer matched by pruning the general taxonomic trees affiliated with mapped genomes. The final metagenomic read counts were normalized by multiple methods, rarefaction, relative abundance (supplementary table 2 and supplementary table 8), and median normalization (supplementary table 3 and supplementary table 9) with the script (<https://github.com/ifanlyn95/multi-CRC-fungi>). To prevent the denominator from being zero, all zero values will be replaced by the normal distribution with a mean value of one-tenth of the non-zero minimum value and one-hundredth of the non-zero minimum value of the variance. The median normalization means dividing the relative abundance of each feature by the median of the control group:

: means the relative abundance of fungi or bacteria in sample , which belongs to cohort . In contrast, cohort has exactly sample to sample .

Overview of statistical analyses

To overcome the limitations of univariate statistics, relative abundance and median normalization were applied. The confounding factors, such as age, BMI, and tumor location, were not considered because these meta information were not complete in each cohort. The non-parametric test was used throughout for univariate association testing between the abundances of fungi and bacteria; all were two-sided Wilcoxon tests except were otherwise mentioned. Co-occurrence and co-exclusion relationships within fungi and between fungi and bacterial kingdoms were estimated using Differential Gene Correlation Analysis (DGCA)24. *P* value less than 0.05 after multiple comparisons correction using the false discovery rated method were considered significant except were otherwise mentioned. Areas under the receiver operating characteristic curves (AUCs) were determined by random forest model and the relative change values between trans-kingdom AUCs and pure bacteria AUCs present the models difference. Network parameters were estimated by using the network analyzer algorithm of Cytoscape, version 3.0.425, with default parameters and clustered by affinity propagation clusters methodology26 .

Differential abundance analysis

Three criteria were used to select the potential differentially abundant candidates between CRC and healthy individuals. First, we excluded the candidates with an average rarefied abundance less than 0.1% fungal composition. We selected the same trend features (SSTF), required more than 3/4 cohorts (not less than six cohorts) to perform the same trends. And the log2 of Multiple Median Fold Change (log2MultMedFC) was the evaluation index of SSTF. We define the log2MultMedFC as:

means the counts of CRC/CTRL samples in an individual cohort.

means the fungal names.

means the relative abundance of species in sample .

The second measure was based on the Wilcoxon rank-sum test. We identified differentially abundant features between two groups on a per species basis using Wilcoxon rank-sum test and with p-values being adjusted using the conservative Bonferroni correction. For the last criteria, we discarded features with an absolute value of log2 of features’ Fold Change less than 0.5. In addition, we ignored the unclassified strain of bacteria. The scripts were available on Github (https://github.com/ifanlyn95/multi-CRC-fungi).

Microbial association and network analysis

Inter-fungal and fungal-bacterial co-occurrence and co-exclusion relationships were estimated using the DGCA algorithm24. DGCA is an algorithm for systematically assessing the difference in feature-feature regulatory relationships under different conditions. In our case, DGCA was used to assess the difference of inter-fungal and fungal-bacterial correlations between different groups (CRC vs Adenoma vs Healthy controls). *P* values less than 0.05 were considered significant. When comparing the inter-fungal and fungal-bacterial correlation in different groups, DGCA leverages the permutation samples to calculate empirical p-values. The inclusion criterion for network plot features is correlation index less than -0.2 or more than 0.5. To stabilized the variance of sample correlation coefficients in different stages, the Fisher z-transformation is untilized:

where presents the sample correlation coefficient. Another important index used is the z-score, which represents the relative strength of differential correlation. The z-score is defined as:

wheres refers to the variance of the z-score in condition x. Using the difference in z-scores, a two-sided *p*-value can be calculated with the standard normal distribution. We considered the empirical p-values less than 0.05, and the absolute values of the z-score larger than 5 as a significantly different correlation between different groups. When the inter-fungal or fungal-bacterial correlation in CRC is weaker than that of the healthy control, the z-score would be positive. Whereas, if the correlation is stronger in CRC, the z-score would be negative. Based on a threshold of correlation significance (p-value less than 0.05) and the direction of correlation changes in different conditions (i.e. the correlation is stronger or weaker in CRC compared to healthy controls), species-species correlations in each condition could be categorized into three classes: significant positive correlation, no significant correlation, and significant negative correlation. As we have two conditions (CRC vs Healthy controls), there were nine classes for differential correlation analysis, namely '+/+', '+/0', '+/-', '0/+', '0/0', '0/-', '-/+', '-/0', and '-/-'. The interactions between these selected features were clustered with affinity propagation clusters methodology26.

Fungal Strain and Culture Condition

*Aspergillus rambellii* (CBS101887) was purchased from the Westerdijk Fungal Biodiversity Institute (Utrecht, NL), cultured on Sabouraud Dextrose Broth （SDB） (S3306; Sigma-Aldrich, St Louis, MO) agar plate for 3 days at 25°C in aerobic conditions. When the density of *A. rambellii* reached to XX, the fungal culture medium was screwed from the SDB-agar plate, centrifuged at 4500g for 15minutes and filtered through a 0.2-mm pore-size filter twice to obtain the *A. rambellii* conditional medium (*A.r* CM).

Cell Culture

Human normal colon epithelial cell line NCM460 was obtained from INCELL Corporation (San Antonio, TX). CRC cell lines HT29 and SW480 were obtained from the ATCC. All of the cell lines were grown in high-glucose Dulbecco’s Modified Eagle’s Medium (DMEM) (Gibco BRL, Grand Island, NY) supplemented with 10% (vol/vol) fetal bovine serum (FBS), and 1% penicillin/streptomycin in a humidified atmosphere containing 5% CO2.

Cell Viability Assay

Cell viability was determined by the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium (MTT) assay. For each well in the 96-well plate, 1000 cells were seeded and treated with the X% *A. r* CM or control medium SDB directly.

Colony Formation Assay

Colon cells (1,000 per well) were seeded on 6-well plates, followed by treatment with X% *A. r* CM in DMEM. SDB was used as the control. The treatment medium was changed every 3 days. After culturing for 14 to 18 days, cells were fixed with cold 100% Methanol and stained with 0.5% crystal violet solution. The colony with more than 50 cells was counted. All experiments were performed 3 times in triplicate.

Colorectal Cancer Patient-Derived Organoid Culture

Organoid 816 was originally from human tissue biopsy samples, obtained from a 46-year-old woman who was diagnosed with colorectal adenocarcinoma in the Prince of Wales Hospital, The Chinese University of Hong Kong. Organoid 828 was originally from human tissue biopsy samples, obtained from a 46-year-old woman who was diagnosed with colorectal adenocarcinoma in the Prince of Wales Hospital, The Chinese University of Hong Kong. The samples were processed with the patient’s consent. The pathologic specimens were embedded into Matrigel and placed in DMEM/F12 þ GlutaMAX (Invitrogen, Carlsbad, CA) containing N2 and B27 supplements (Invitrogen), 10 mmol/L HEPES, 1.25 mmol/L N-acetyl cysteine (Sigma-Aldrich, St Louis, MO), glutamine, 1% penicillin/streptomycin (Sigma-Aldrich), 10 mmol/L SB202190-monohydrochloride (Sigma-Aldrich), R-spondin-1 (RSPO-1), Noggin, WNT3A, and 50 ng/mL epithelial growth factor (Invitrogen). Treatment containing X% *A. r* CM was added into the culture medium directly. The treatment medium was changed every 3 days.

Results

Data pre-processing of metagenomic datasets for studying the associations between mycobiome and CRC

In the present study, a multi-cohort study was performed based on the metagenomics dataset from eight published studies (table 1 and supplementary table 1). On median, 68.28% and 0.1% of filtered reads were mapped to bacterial and fungal databases, respectively (supplementary figure 1a and supplementary table 1). Consistent with a previous study27 the fungus-derived reads observed in our study account for around 0.16% (10e-2.80) of the total enteric microbes. The rarefaction curve (supplementary figure 1b) showed that all cohort samples reached a plateau at 10,000 sequencing reads. We applied stringent filtering criteria to ensure rigorous outcomes and minimize the outlier effect (supplementary figure 1c). In total, 1,329 metagenomes (454 CRC patients, 350 adenoma and 525 healthy controls) were analyzed.

Alterations of enteric fungal composition in CRC

When we investigated at the overall fungal composition, we discovered that *Ascomycota* was the most prevalent fungal phylum across all cohorts, whereas other dominant fungal phyla exhibited considerable inter-cohort heterogeneity (supplementary figure 2a and 2b). Unlike all other cohorts, Yachida's Japanese cohort has Mucoromycota as the second most prevalent phylum, rather than Basidiomycota. Other instances include Asians having a lower prevalence of Microsporidia than non-Asians (supplementary figure 2b and 2c).

In agreement with previous researches showing distorted microbial diversity in the diseased group28, the alpha diversity of enteric fungi was reduced in CRC patients compared to healthy individuals when considering all the cohorts together and in three individual cohorts (chao1 index) (supplementary figure 2d). When comparing the CRC group to healthy controls, we found significant changes in fungal phylum composition and alpha diversity.

Identification of fungal species associated with CRC

We searched for the potential enteric fungal shifts in CRC patients as compared to healthy individuals. And we investigated that 296 species were obtained for further investigation after filtering low abundant (less than 0.1%) fungus from the 592 aligned species (figure 1a and supplementary table 2, 3). Using the Wilcoxon rank-sum test to compare data from all the cohorts together, 74 differentially abundant fungi were identified, which was named as the main set (FDR < 0.1). Among the 74 identified species, we further shortlisted 33 species that demonstrated significant alterations (FDR < 0.01) as the core set (figure 1a and supplementary table 4). We used the SSTF and Wilcoxon rank-sum test to see whether these 33 fungi (main set) were consistently altered across all eight cohorts. Except for the 2019\_ThomasAM and 2019\_Yachida cohorts, the enrichment and depletion status of the 33 species were constant in six cohorts (figure 1b). Interestingly, the most of the 33 species in the 2019\_ThomasAM cohorts exhibited considerable enrichment in CRC patients or no significant difference, with just a handful showing depletion in CRC patients (figure 1b). Whereas in the 2019\_Yachida group, most of the identified 33 fungi showed weak variance in CRC patients versus healthy individuals, unlike other cohorts (figure 1b). In addition, we noticed that three of the 33 species exhibited similar alterations across all cohorts, with *Aspergillus rambellii* and *Erysiphe pulchra* being enriched in CRC and *Trichophyton mentagrophytes* being reduced (figure 1b and supplementary table 5). We further identified 15 species that were consistently altered in 7 out of the eight cohorts. Ten of them were enriched in CRC patients, while the remaining five were depleted (supplementary table 5). For the 33 species in the core set, 10 were enriched, and the remaining 23 were depleted in CRC patients (figure 1c). The alterations of these 33 species in CRC patients versus healthy individuals were relatively consistent in most cohorts. Among them, *Aspergillus* *rambellii* showed the most remarkable difference between the CRC patients and the healthy control groups (-log10FDR = 17.29).

The fungal community of CRC and adenoma patients was also compared (supplementary figure 3, supplementary table 6 and supplementary table 7). In both CRC patients against adenoma patients and CRC patients versus healthy persons, seven fungus species varied substantially (FDR < 0.01). These species include *Aspergillus rambellii*, *Moniliophthora perniciosa*, *Erysiphe pulchra*, *Sphaerulina musiva*, *Phytophthora capsici*, *Aspergillus kawachii*, and *Cordyceps sp. RAO-2017* (supplementary table 4 and table 6). These species belong to the *Ascomycota* phylum except *Moniliophthora* *perniciosa* and *Phytophthora capsici* (supplementary table 12). We discovered universally differently numerous fungi in CRC patients as compared to adenoma patients, which was somewhat consistent with healthy controls.

*A. rambellii* is the most significant enriched fungus in CRC

We sought to idnetify the most significant fungi candidates associated with CRC using stringent criteria. As shown in figure 1c, enriched *A.* *rambellii* and depleted *A.* *kawachii* were the two significant altered fungi in CRC. *A. rambellii* was significantly enriched in seven cohorts (figure 1d); whilst, *A. kawachii* was significantly depleted in cohorts of 2014\_ZellerG, 2016\_VogtmannE, 2017\_JunY, and our unpublished dataset (figure 1d). Collectively, our multi-cohort analysis revealed that the enriched *A.* *rambellii* and depleted *A. kawachii*, were significantly associated with CRC in multiple cohorts.

*Aspergillus rambellii* promotes colon cancer cell growth

[TBA] figure 2

Ecological networks of CRC-enriched and depleted fungi increased with CRC progression

We evaluated the interactions among 33 core CRC-enriched and depleted fungi acorss steps of CRC progression using the correlation analysis with DGCA24. As shown in figure 3, we observed that both co-occurrence and co-excluding interactions among CRC-enriched and -depleted fungi were significantly different across the stages of healthy control, adenoma and CRC) – progressively stronger towards carcinogenesis. Four CRC-associated fungi including *Aspergillus* *rambellii*, *Erysiphe* *pulchra*, *Thielaviopsis* *punctulata*, and *Sphaerulina* *musiva,* showed significant co-occurrence centralities. These correlations weaken in adenoma and disappear in healthy individuals (figure 3 and supplementary figure 4) , indicating they are the most significant fungi in the CRC interaction network. In both healthy and adenoma conditions, *A. rambellii* only showed a strong correlation with *Moniliophthora permiciosa,* which was significantly distinct from CRC(figure 3).

Ecological interactions among differentially abundant fungi and bacteria with CRC progression

To identify the significant differentially abundant bacteria between CRC and healthy individuals, we performed Wilcoxon rank-sum test with stringent selection criteria (q-value < 0.01, , unclassified species removed) (supplementary table 8 and supplementary 9). Thirty-one differentially abundant bacteria were identified in CRC, which was more significant than fungi (supplementary table 10), including CRC-related enriched bateria *Fusobacterium nucleatum*, *Parvimonas micra*, and *Gemella morbillorum*29–40, and depleted beneficial bacteria *Roseburia* *intestinalis*, *Bifidobacterium* *bifidum*, and *Streptococcus* *thermophilus*41–46.

The possible interaction of differentially abundant fungi and bacteria in CRC progression was investigated using ecological network analysis. From healthy controls through ademona to CRC, we noticed that the fungal-bacterial link became stronger (figure 5, supplementary figure 5 and supplementary table 11). The number of significant trans-kingdoms interactions was increased during the CRC progression, from 143 (healthy control), to 156 (adenoma), and 184 (CRC) (supplementary table 11). And 99 fungal-bacterial correlations only appeared in CRC, which presented the important altered trans-kingdom community in CRC, such as *A. rambellii-F. nucleatum*. The 17 pairs of relationships represented by *A.rambellii-P.micra* were only significant in disease states (adenoma and CRC). This revealed that the fungal-bacterial interactions might be associated with CRC tumorigenesis.

Differential inter-fungal and fungal-bacterial correlation analysis in CRC versus healthy controls

We investigated whether the inter-fungal and fungal-bacterial correlation in CRC progression were substantially different between CRC and healthy control. In the density graph with z-score, which indicates the strength of the relationship difference, two peaks at -2 and +4 were detected in fungal-bacterial correlations (supplementary figure 6a). Inter-fungal, inter-bacterial, and fungal-bacterial interactions all showed substantial disparities in correlation changes, according to our differential correlation study (supplementary figure 6a).

We also defined the nine classes in the pair correlation comparison (supplementary figure 6b left panel). In cases pair correlation analysis, negative correlations were rare (supplementary table 11). Notably, only the intra-fungi had six (7.69%) '-/+' cases, which means the feature pair correlation in CRC was negative, while its association in healthy control was positive (supplementary figure 6b right panel). It might reveal some potential markers or changes in the stage alteration.

Sixty-four microbes (31 bacteria and 33 fungi) were separated into six clusters by affinity propagation clusters26 with z-score (figure 4). The majority of the candidates were found in two groups. Because 18 of the 22 microorganisms were fungi, we called the largest one the Fun\_cluster. Similarly, since 17 of the 21 microorganisms were bacteria, we dubbed the second largest cluster the Bac\_cluster. As the clustering results were based on the z-score, we can observe that the alteration of inter-bacteria and inter-fungal correlation have distinct differences. Notably, some bacteria were present in the Fun\_cluster while some fungi were present in the Bac\_cluster. This implies that these might be the special species that have more trans-kingdom interactions and might be important in CRC pathogenesis.

Fecal fungal-bacterial biomarkers to distingruish CRC patients from healthy subjects

In previous researches7,8, the classifier distinguishing CRC patients from healthy individuals was trained with only bacterial markers. We examined if the fungal markers identified in this study could improve the accuracy of the classifier and further increase the potential of using fecal metagenomic markers to early identify CRC patients from the population. To identify CRC from healthy people, we trained the model with single or multiple characteristics. Single feature refers to using only one fungus or bacteria in the core set as the predictor of the model. Whereas, multiple features refer to using a consortium of pure bacteria, pure fungi or a combination of fungi and bacteria in the core set as the predictor. Among the single-feature models, only six features’ average AUC value was greater than 60%. These include four bacteria (*Fusobacterium nucleatum*, *Parvimonas micra*, *Gemella morbillorum*, and *Porphyromonas asaccharolytica*) and two fungi (*Aspergillus rambellii* and *Aspergillus kawachii*) (table 2). *P. micra* had the highest performance, with an average AUC of 67.79%, but it had a poor performance in 2016\_VogtmannE (AUC: 56.15%), in which *A. rambellii* had the best performance (AUC: 67.57%). It was discovered that in certain cases, the predictive values of fungi might be higher the bacteria. Next, we sought whether employing a combination of fungi and bacteria as predictors would enhance the classifier model. We trained and compared the multi-features model with pure fungi, bacteria, and the fungal-bacterial combination, containing 17, 12, and 14 species respectively (figure 5a, supplementary figure 7). Remarkably, in 2016\_VogtmannE (fungi: 77.27% vs bacteria: 70.63%) and 2019\_WirbelJ, the fungal classifiers were more accurate than the bacterial one (fungi: 93.23% vs bacteria: 89.39%). The AUC of classifier with combined fungal and bacterial markers was 1.44% - 10.60% relative change higher than the bacterial classifier in seven of eight cohorts (figure 5b). Altogether, the classifier with combined fungal and bacterial markers was more accurate than the conventional pure fungal or bacterial classifiers.

Discussion

Previous research have mostly focused on the relationship between gut bacteria and host disease, while fungi are often disregarded due to their low proportion in the enteric microbiome47. We conducted the first comprehensive multi-cohort study of enteric fungal shotgun metagenomics in CRC of seven publicly accessible cohorts as well as our unpublished dataset. We were able to demonstrate the universal mycobiota alteration in CRC patients versus healthy individuals. Using robust statistical methods, we identified differentially abundant fungi and their ecological networks in stages of CRC progression. *A. rambelli* was shown to be the most different and enriched fungus in CRC, and its potential to promote CRC cell proliferation was confirmed in vivo and in vitro experiments. The analysis and comparison of intra-fungal and trans-kingdom ecological networks indicated that species interact differently at various stages of CRC progression, and *A. rambellii* and *P. kudriabzevii* may promote and inhibit colorectal carcinogenesis in collaboration with bacteria, respectively. Fungal-bacterial compound classifiers were also required, and they performed better than typical bacterial classifiers in terms of distinguishing CRC.

The multi-cohort-analysis approach has been used to evaluate and combine results from comparable studies50 with significant advantages of reducing the influence by cohort-specific bias and increasing statistical power. Using the rank-sum test and SSTF in our analysis, we identified 33 fungi that were associated with CRC across eight cohorts. Our results suggested that *A. rambellii* was the most significant CRC-enriched fungus, which showed universal associations with CRC in seven of eight cohorts. This finding was supported by previous studies showing the ability of *A. rambellii* in synthesizing carcinogenic products, aflatoxin and aflatoxin precursor sterigmatocystin51,52. It was the first time to report A. rambellii

XXX.

Furthermore, *A. kawachii*, a fungus from the genus Aspergillus, was the most important CRC-depleted fungus. Despite the fact that both fungi belong to the same genus, they have opposing functions in CRC. The crude enzyme extract derived from *A. kawachii* could enhance the antioxidative activities of Viscum album var. coloratum53, a promising agent for immunomodulation, treating colon cancer54 and hepatoma. The fermented silkworm produced by *A. kawachii* solid-state fermentation could inhibit the human hepatocellular carcinoma cells55. *R.* *irregularis* was the second most CRC-depleted fungus. *A. Officinalis*-*R. irregularis* symbiosis was reported to induce the production of salvianolic acid, which has anti-cancer effects56,57. *A. Officinalis*-*R. irregularis* symbiosis could produce rosmarinic acid, ferulic acid and caffeic acid, which related to beneficial properties of antioxidant, anti-inflammatory, and antimicrobial effects58,59. These findings support their roles of the enriched- or depleted fungi in the promotion or inhibition of colorectal carcinogenesis.

【here!!】

The intra-fungal interaction in CRC was considerably different with the healthy controls. Most previous studies have focused on the role of a single key microorganism or metabolite in CRC development60–62. However, microbes in our gut are interacting with one another. From our results, multiple strong positive or negative correlations in CRC disappeared or weakened in adenoma or healthy control groups. It maybe the interaction of multiple species that causes the carcinogenesis. Therefore, we cannot only focus on the abundance changes of fungi in two groups but also the alteration of inter-fungal correlation. The most inter-fungal interactions in these three groups were significantly different, but *Aspergillus* *rambellii*, *Rhizophagus* *irregularis*, *Rhizophagus* *clarus*, *Phytopythium* *vexans*, and *Edhazardia* *aedis* appeared in all groups. It suggested that they might play a vital role in the stability of the entire intestinal ecology.

From our differential correlation analysis, we obtained two main clusters the Bac\_Cluster and Fun\_Cluster. Our results showed that the inter-fungal correlations were weakened in CRC, while inter-bacterial correlations were enhanced. The disruption of the inter-fungal correlation may break the healthy intestinal environment and induce colorectal carcinogenesis. On the other hand, the increased bacterial correlations in CRC may potentially contribute to colorectal carcinogenesis. Interesting results were observed when comparing the of the fungal-bacterial interactions in two conditions (CRC vs Healthy controls). Our results showed that the fungal-bacterial correlations with smaller changes across two conditions (|z-score| < 2) contained low proportions. The CRC strengthen and weaken correlations performed the primary and secondary ratios in fungal-bacterial correlation comparisons, respectively. It revealed that the internal-kingdom associations and external-kingdom correlations were significantly different. This suggested that bacterial kingdom dysbiosis may cause the fungi to tremble rapidly, which was not similar to the warm alteration of internal-kingdom relationships as previously described11,63.

We discovered that most reported or potential probiotics were separated in the Bac\_Cluster, and *P. kudriavzevii* had multiple correlations with probiotics in this study. Supporting evidence from previous studies showed that *P. kudriavzevii* derived metabolites possess anticancer effects by inhibiting cell proliferation and inducing intrinsic and extrinsic apoptosis in colon cancer cells64. There were also strong correlations among *A. rambellii*, *F. nucleatum*29,65,66, and *P. micra*5, from which the latter two were the famous CRC-related pathogens. A previous study revealed that the altered trans-kingdom association between bacteria and virus are associated with CRC21. We proposed that the trans-kingdom interactions between bacteria and fungi are also important colorectal carcinogenesis. However, this discovery was explored only in metagenomic sequencing study. More experiments are needed to verify and prove this hypothesis.

for marker only,fecal fungi markers could be used together with bacterial markers to improve the accuracy of distinguishing CRC patients from tumor-free healthy individuals.

In conclusion,

References

Automatic citation updates are disabled. To see the bibliography, click Refresh in the Zotero tab.

**Figure legends**