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

Yufeng Lin1, Yali Liu1, Harry Cheuk-Hay Lau1, Xing Kang1, Nick Lung-Ngai Ting1, Thomas Ngai-Yeung Kwong1, Yiwei Wang3, Han Jing2, Siu-Kin Ng1, 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

**Correspondence:** Professor Jun Yu, Institute of Digestive Disease, Department of Medicine and Therapeutics, Prince of Wales Hospital, the Chinese University of Hong Kong, Shatin, NT, Hong Kong. Phone: (852) 37636099, Fax: (852) 21445330, Email: [junyu@cuhk.edu.hk](mailto:junyu@cuhk.edu.hk)

**Author Contributions:** YL performed data collection, study design, bioinformatic analyses and drafted the manuscript. YL, XK and TNYK performed the experiments. NLNT, HCHL and SKN revised the manuscript. HJ, YW and CL commented on the bioinformatic analyses. JS and JJYS commented on the study. SHW collected clinical data and revised the manuscript. JY designed and supervised the study and revised the manuscript. All authors approved the final version of manuscript.

**Acknowledgement**

We would like to thank Weixin Liu, Yanqiang Ding, Andrew Maltez Thomas and Nicola Segata for their comments on bioinformatic analyses.

**Fundings**

ABSTRACT

**Background & Aims:** Enteric fungi is a major component of human gut microbiota, but its role in colorectal cancer (CRC) remains largely elusive. We aimed to conduct a meta-analysis to uncover the contribution of fungal mycobiota to CRC progression.

**Methods:** We retrieved fecal metagenomic datasets from 7 previous publications and established an additional metagenomic cohort, totaling 1,329 metagenomes (454 CRC, 350 adenoma and 525 healthy controls). Analyses on mycobiota composition, fungal interactions, and trans-kingdom interactions between fungi and bacteria were conducted. Performance of fungal and bacterial biomarkers in CRC diagnosis was assessed.

**Results:** Our multi-cohort analysis revealed that alteration in enteric mycobiota was occurred in CRC. Abundances of 33 fungal species (10 enriched, 23 depleted) were significantly changed in CRC patients compared to healthy controls (FDR < 0.01). *Aspergillus rambellii* was the top enriched fungi in CRC patients. *Aspergillus rambellii* promoted colon cancer cell growth *in vivo* and *in vitro* in nude mice. Whereas co-occurrence interactions among *A. rambellii* and other CRC-enriched fungi were stronger in CRC. Our correlation analysis also demonstrated trans-kingdom interactions between enteric fungi and bacteriain CRC progression, of which *A. rambelli* was closely associated with well-established CRC-enriched bacteria including *Fusobacterium nucleatum*. Moreover, we found that a diagnostic panel with combined fungal and bacterial biomarkers was more accurate than panels with pure bacteria to discriminate CRC patients from healthy individuals (AUC relative change increased by 1.44%-10.60%).

**Conclusions:** This study revealed the involvement of enteric fungi and their trans-kingdom interactions with bacteria in CRC, implying the importance of fungal mycobiota in colorectal tumorigenesis.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer worldwide with over 500 thousands associated death every year1,2. The contribution of gut microbiota to CRC progression has been widely acknowledged, of which its composition constantly shifts across stages of colorectal tumorigenesis with enrichment of pathogenic bacteria3–5. In comparison, the role of microbial components other than gut bacteria, such as fungi, is largely unexplored in CRC. Our previous study reported the alteration in enteric mycobiota in CRC6. Whereas increasing evidence has revealed the mechanistic association of commensal fungi with pathogenesis of different diseases, as exemplified by the contribution of fungal mycobiota to pancreatic tumorigenesis through modulating the host immunity7. These studies therefore imply the importance of enteric fungi in disease progression.

Metagenomic sequencing is a major tool for depicting the complete profile of microbes from different taxonomic kingdoms including bacteria, fungi and viruses that are present in human. However, fungal mycobiota study has been heavily obstructed due to the low abundance of fungi in human gut (≤ 1% of total microbes) and the lack of well-characterized reference fungal genomes for aligning sequencing reads8. As consequence, enteric fungi have been frequently neglected in the analyses of previous metagenomic studies. Moreover, metagenomic sequencing is readily influenced by sample quality, sequencing platform and the bioinformatic pipelines used for analysis, and difference in these factors could lead to inconsistency among metagenomic studies9. Whilst such inconsistency could be worsened by inter-study variations of gut microbiota which is attributed to numerous environmental factors including diet, geography, and ethnicity10. Hence, given by its increased statistical power, meta-analysis could facilitate identification of key CRC-associated fungi that are consistent in multiple metagenomic studies.

In this study, we performed a meta-analysis on 7 published fecal metagenomic datasets and an additional cohort to examine the correlation between enteric fungi and CRC, totaling 1,329 metagenomes with 454 CRC, 350 adenoma and 525 healthy controls. We confirmed that alteration in enteric mycobiota is occurred in CRC progression. In particular, enrichment of *Aspergillus rambellii* and depletion of *Aspergillus kawachii* were significantly associated with CRC. The interactions among different fungal species and trans-kingdom interactions between fungi and bacteria were also revealed, of which *A. rambelli* are closely linked with pathogenic bacteria such as *Fusobacterium nucleatum* in CRC progression. We further demonstrated that a diagnostic panel with mixture of fungal and bacterial biomarkers is more accurate than panels with pure bacteria to identify CRC patients from healthy individuals.

METHODS

Hong Kong cohort of patients with CRC, adenomas, and controls

of and112 healthy individuals in Hong Kongsstudy11 Subjects were selected when receiving colonoscopy screening from Jockey Club Bowel Cancer Education Centre, the Chinese University of Hong Kong. Individuals with presentations of digestive symptoms were recruited as symptomatic patients, and asymptomatic individuals with age ≥ 50 years old were recruited as controls. Individuals were excluded if they had history of surgery. Fecal samples were collected per each subject and stored at -80°C within 24 hours after collection. Total DNA was extracted from fecal samples by QIAamp DNA Stool Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. All subjects had intact colonic at the time of stool collection.

Published metagenomic cohorts of patients with CRC, adenomas, and controls

CRC studies from January 2014 to January 2021 with shotgun metagenomic sequencing on human fecal samples were searched on National Center for Biotechnology Information (NCBI). studies withtoavoidamplification ensure 12Eight studies with published metagenomic datasets of CRC patients and healthy controls were included, of which five of them also involved patients with adenoma13–17 (**Table 1** and **Supplementary Table 1**). Seven metagenomic datasets were downloaded from European Nucleotide Archive (ENA) using the following ENA identifier: ERP005534 for Zeller et al.13, ERP008729 for Feng et al.14, PRJEB10878 for Yu et al.4, PRJEB12449 for Vogtmann et al.18, PRJNA389927 for Hanningan et al.15 (excluded because of non-PCR-free preprocessing and low sequencing depth), PRJEB27928 for Wirbel et al.19, and SRP136711 for Thomas et al.16. The metagenomic dataset of Yachida et al.17 was downloaded from the DNA Data Bank of Japan (accession number: DRA006684 and DRA008156). These eight cohorts plus our additional cohort covered samples from 8 different countries, totaling 454 patients with CRC, 350 adenoma, and 525 controls.

Sample filtering

To ensure consistent and data quality, samples were subjected to strict filtering before analysis. The abnormal patients, such as IBD and history sugary patients, and confused stage samples would be first discarded. And only the cohorts with PCR-free processing were left. Samples with low alignment reads (≤ 1,000,000) which may be attributed to low sequencing depth and host reads contamination, were excluded. Outliers and suspected contaminated cases were also removed, including samples with high-fungi content (fungal reads were ≥ 1% of total metagenomic reads), low-fungi content (fungal reads were ≤ 0.01% of total metagenomic reads), and high-bacteria content (bacterial reads were ≥ 50% of total metagenomic reads). Moreover, samples with low-fungi sequencing depth (fungi-aligned read counts ≤ 10,000) were discarded, which was consistent with a previous study revealing that fungi could not be detected in at least 30% of individuals20.

Sequence pre-processing and taxonomic profiling

The KneadData default parameters were applied for quality control of all metagenomic sequencing data which could separate microbial reads from contaminated reads from the host or other user-defined sources of *in silico* methods. Taxonomic profiles were then generated by Kraken2 (version 2.0.9-beta) across our custom database, containing 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 using Jellyfish program by counting distinct 31-mer. Default parameters were used and all reads with quality less than 20 and shorter than 50 nucleotides were discarded. 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 rarefaction, relative abundance (**Supplementary Table 2** and **Supplementary Table 3**), and median normalization (**Supplementary Table 4** and **Supplementary Table 5**) using a GitHub script (<https://github.com/ifanlyn95/multi-CRC-fungi>). To prevent the denominator from being zero, all zero values were replaced by normal distribution with a mean value of 1/10 of the non-zero minimum value and 1/100 of the non-zero minimum value of variance. The median normalization was calculated by dividing the relative abundance of each feature by the median of control group:

where is the relative abundance of fungi or bacteria in sample whichbelongs to cohort . In contrast, cohort has exactly sample to sample .

Statistical analyses

To overcome limitations of univariate statistics, relative abundance and median normalization were applied. While confounding factors such as age, BMI, and tumor location were not considered because this meta information was not complete in each cohort. Non-parametric test was used during univariate association testing between abundances of fungi and bacteria; two-sided Wilcoxon tests were used unless otherwise mentioned. Co-occurrence and co-exclusive interactions among fungi and between fungi and bacteria were estimated using Differential Gene Correlation Analysis (DGCA)21. *P* value less than 0.05 after multiple comparisons correction using the false discovery rated method was considered significant unless otherwise mentioned. Areas under the receiver operating characteristic curves (AUCs) were determined by random forest model and the relative change values between AUCs of trans-kingdom panel and AUCs of pure bacteria panel represented the difference in models. Network parameters were estimated using the network analyzer algorithm of Cytoscape (version 3.0.4)22 with default parameters and clustered by the plugin, affinity propagation clusters23 .

Differential abundance analysis

Three criteria were used to select potential candidates with differentially abundance between CRC and healthy controls. First, candidates with an average rarefied fungal abundance less than 0.1% were excluded. Candidates with same trend features (SSTF; same trends in ≥ 6 cohorts) were selected. The log2 of Multiple Median Fold Change (log2MultMedFC) was used as the evaluation index of SSTF and calculated by:

represents the counts of CRC or control samples in an individual cohort;

represents the names of fungi;

represents the relative abundance of species in sample .

The second criteria was based on Wilcoxon rank-sum test. We identified differentially abundant features between two groups on a per species basis using Wilcoxon rank-sum test with *p*-values being adjusted by the conservative Bonferroni correction. For the last criteria, we discarded features with an absolute value of log2 fold change in features less than 0.5 and 1 in fungi and bacteria, respectively. In addition, all unclassified strains were not included in analysis. The scripts were available on GitHub (same above).

Microbial association and network analysis

Co-occurrence and co-exclusive interactions among fungi and between fungi and bacteria were estimated using the DGCA algorithm for systematically assessing the difference in feature-feature regulatory relationships under different conditions21. Here, DGCA was used to assess the difference of fungi-fungi and fungi-bacteria correlations among distinct groups of subjects (CRC, adenoma or healthy controls). *P* value less than 0.05 was considered significant. When comparing correlations in different groups, DGCA leverages the permutation samples to calculate empirical *p*-values (**Ref**). A correlation was plotted in the network only if its correlation index was less than -0.2 or greater than 0.5. To stabilize variance of sample correlation coefficients in different stages, Fisher z-transformation was used:

where represents the sample correlation coefficient, and *z* represents the z-score which is the relative strength of differential correlation. The z-score was defined as:

where refers to the variance of z-score in condition x, and *dz* represents the difference in z-score. Obtaining the difference in z-score *(dz*) allowed calculation of two-sided *p*-value with standard normal distribution. Empirical *p*-value less than 0.05 was considered significant, and z-score with absolute value larger than 5 was considered as a significantly different correlation between distinct groups. When a fungi-fungi or fungi-bacteria correlation in CRC was weaker than that in healthy controls, the z-score was positive and vice versa. Based on the cutoff of correlation significance (*p*-value ≤ 0.05) and direction of correlation changes in different conditions (i.e. the correlation was stronger or weaker in CRC compared to healthy controls), species-species correlations could be categorized into three classes: significantly positive, insignificant, and significantly negative. As we compared correlations between CRC and healthy controls, there were a total of nine classes of differential correlation, namely “+/+”, “+/0”, “+/-“, “0/+”, “0/0”, “0/-“, “-/+”, “-/0”, and “-/-“. Interactions between these features were clustered by affinity propagation clusters methodology23.

Random forest-based machine learning

Our machine learning analyses exploited median normalized relative abundance at taxonomic species level using Kraken2 and its plugin Bracken. To reduce bias, leave-one-set-out (LOSO) approach was used to perform nested cross-validation. Feature selection and model training were performed by the R package “random Forest”. To choose the best model, we utilized the maximal average AUC and the best AUC in multi-features and single feature as the selection criteria, respectively. Only species ranked as the top three features in at least one cohort were included in selection of multi-features model characters.The code and figures generated during analyses were available on GitHub (same above).

Fungal strain and culture condition

*Aspergillus rambellii* (CBS101887) was purchased from Westerdijk Fungal Biodiversity Institute (Utrecht, Netherlands) and cultured on Sabouraud Dextrose Broth (SDB) (SDBS3306; Sigma-Aldrich, St Louis, MO) agar plate for 4 days at 25°C in aerobic conditions. The fungal culture medium was screwed from the SDB-agar plate, centrifuged at 4,500 g for 15 minutes, and filtered through 0.2-mm pore-size filter twice to obtain the *A. rambellii* conditional medium (ARCM).

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 American Type Culture Collection (Manassas, VA). All cell lines were grown in high-glucose Dulbecco’s Modified Eagle’s Medium (DMEM) (Gibco BRL, Grand Island, NY) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin in a humidified atmosphere containing 5% carbon dioxide.

Cell viability assay

Cell viability was determined by the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium (MTT) assay (Thermo Fisher, M6494). Cells were seeded onto 96-well plates (1000 cells per well) and directly treated with DMEM supplemented with 10% ARCM or control medium SDB.

Colony formation assay

Colon cells (1,000 cells per well) were seeded onto 6-well plates, followed by treatment with DMEM supplemented with 10% ARCM or control medium SDB . The treatment medium was changed every 3 days. After culturing for 10-14 days, cells were fixed in cold 100% methanol and stained with 0.5% crystal violet solution. Colony with more than 50 cells was counted. All experiments were performed in triplicate.

CRC patient-derived organoid culture

Organoid 816 was originally extracted from human tissue biopsy samples, obtained from a 54-year-old woman who was diagnosed with metastatic colorectal adenocarcinoma (Princess Margaret Living Biobank, Canada). Organoid 828 was also originally extracted from human tissue biopsy samples, obtained from a 46-year-old woman who was diagnosed with colorectal adenocarcinoma (Princess Margaret Living Biobank). Pathologic specimens were embedded into Matrigel and placed in DMEM/F12 þ GlutaMAX containing N2 and B27 supplements, 10 mmol/L HEPES, 1.25 mmol/L N-acetyl cysteine (Sigma-Aldrich), glutamine, 1% penicillin/streptomycin (Sigma-Aldrich), 10 mmol/L SB202190-monohydrochloride (Sigma-Aldrich), R-spondin-1, Noggin, WNT3A, and 50 ng/mL epithelial growth factor. Treatment containing 10% ARCM was added into the culture medium directly. The treatment medium was changed every 3 days. Unless specified, all reagents used for organoid culture were purchased from Invitrogen, Carlsbad, CA.

RESULTS

Meta-analysis of metagenomic datasets to study the association of enteric mycobiota with CRC

In this meta-analysis, we included seven published studies that used fecal shotgun metagenomics to characterize CRC patients compared to healthy controls (**Table 1** and **Supplementary Table 1**). For an additional eighth study population, we generated new fecal metagenomic data from samples collected in Hong Kong (**Table 1** and **Supplementary Table 1**). Among these metagenomes, median 0.16% (10-2.80) of total metagenomic reads were mapped to fungal genomes (**Supplementary Figure 1A** and **Supplementary Table 1**), which is consistent with previous studies at 0.1%24. The rarefaction curve showed that all cohort samples reached a plateau at 10,000 fungal sequencing reads (**Supplementary Figure 1B**). To ensure the quality of acquired metagenomic datasets, we applied a series of filtering with stringent criteria to remove samples with contamination or low-fungi content (**Supplementary Figure 1C**). Upon filtering, a total of 1,329 metagenomes were retained including 454 CRC patients, 350 patients with adenoma, and 525 healthy controls.

Alteration of enteric mycobiota in CRC

The fungal composition at phylum level was first investigated. *Ascomycota* was the most prevalent fungal phylum in both CRC patients and healthy controls across all cohorts (**Supplementary Figure 2A and B**). We then compared abundances of the top five prevalent phyla between CRC patients and healthy controls, and found that none of these phyla had consistent change in abundance across all cohorts (**Supplementary Figure 2C**). For fungal richness, the alpha diversity (Chao1 index) was reduced in CRC patients compared to healthy controls when combining all cohorts together, in line with previous findings25 (**Supplementary figure 2D**). Whereas significant difference in alpha diversity was observed only in three cohorts, implicating the presence of inter-cohort variation which could be attributed to ethnic and geographical disparities among cohorts.

Identification of CRC-associated fungal species

To depict CRC-associated fungi, abundance of individual fungus with relative abundance > 0.1% (*n* = 296) in all CRC patients was compared to healthy controls (**Supplementary Table 2** and **4**). We identified 74 fungal species with differential abundance (False discovery rate (FDR) < 0.1, Wilcoxon rank-sum test), and we then trimmed down these 74 fungi to shortlist 33 species at FDR < 0.01 (**Figure 1A** and **Supplementary Table 6**). To identify fungi that were consistently changed across cohorts, SSTF and Wilcoxon rank-sum test were conducted on these 33 fungi (**Figure 1B**) and we identified 15 species (10 CRC-enriched, 5 CRC-depleted) that were constantly altered in 7 out of 8 cohorts (**Supplementary Table 7**). Among these 33 fungi (termed as “core set” hereafter), 10 of them were enriched in CRC patients and the remaining 23 were depleted (**Figure 1C**). In the core set, *Aspergillus* *rambellii* was the most significantly enriched fungal species in CRC patients compared to healthy controls (*p* < 0.05 in 7 out of 8 cohorts) (**Figure 1D** and **Supplementary Table 8**). Whilst *Aspergillus kawachii* was the most fungus CRC patients (*p* < 0.05 in 4 out of 8

Fungal communities in CRC and adenoma patients were also compared and fungi with significantly differential abundance were displayed (**Supplementary Figure 3A**, **Supplementary Table 9**). At FDR < 0.01, 24 fungal species with differential abundance were identified (16 CRC-enriched, 8 CRC-depleted) and 7 of them were also identified (1 CRC-enriched, 6 CRC-depleted) in CRC and normal comparison, including *A. rambellii* and *A. kawachii* (**Supplementary Figure 3B, Supplementary Table 9**). Collectively, our multi-cohort metagenomic analysis revealed that enrichment of *A.* *rambellii* and depletion of *A. kawachii* are significantly associated with CRC.

*Aspergillus rambellii* promotes colon cancer cell growth

To validate the in-silico findings, the direct tumorigenic effect of *A.* *rambellii* was determined by coculturing colon normal epithelial cell NCM460, colon cancer cell lines SW480 and HT29 and CRC organoids with *A.r* CMor control medium SDB. As shown in figure 2a and c, cocuture with *A.r* CM coused a significant increase of cell viability of normal epithelial cell, colon cancer cell lines and CRC organoids. The growth-promoting effect was further confirmed by colony formation assay (figure 2b). These results indicated that *A.* *rambellii* promotes colon cancer cell growth in vitro. [TBA]

Fungi-fungi and fungi-bacteria interactions in CRC progression

Interactions among microbes are known to be involved in CRC progression (**Ref**). We therefore evaluated interactions among the core set of 33 fungal species in CRC, adenoma and healthy controls using the correlation analysis with DGCA21. Both co-occurrence and co-exclusive interactions among fungi were significantly different among distinct groups of subjects, and such fungi-fungi interactions were progressively stronger towards CRC progression (**Figure 3**). In CRC, 4 CRC-enriched fungi including *A.* *rambellii*, *Erysiphe* *pulchra*, *Thielaviopsis* *punctulata*, and *Sphaerulina* *musiva,* showed significant co-occurrence centralities. While these co-occurrence correlations were weaker in adenoma patients and disappeared in healthy controls (**Figure 3** and **Supplementary Figure 4**). Moreover, compared to CRC, *A. rambellii* had completely distinct interactions with other fungi in adenoma and healthy controls which it only showed a strong correlation with *Moniliophthora perniciosa*. Our results identified the key fungi especially *A. rambellii* that are crucial in the enteric mycobiota of CRC patients.

Human gut microbiota is dominated by bacteria of which our analysis revealed that median 68.28% of total metagenomic reads were mapped to bacterial genomes (Supplementary Figure 1A). Given by the predominance of bacteria, we therefore evaluated interactions between fungi and bacteria. Thirty-one bacteria with differential abundance were identified in CRC (**Supplementary Table 10**), including enrichments of pathogenic *Fusobacterium nucleatum*, *Parvimonas micra*, and *Gemella morbillorum*26,27, as well as depletions of probiotic *Roseburia* *intestinalis*, *Bifidobacterium* *bifidum*, and *Streptococcus* *thermophilus*28–30. We then assessed trans-kingdom interactions between fungi and bacteria, and observed an increase in strength of fungi-bacteria interactions along CRC progression (**Figure 3** and **Supplementary Figure 5**). The amount of significant fungi-bacteria interactions was increased from healthy controls (*n* = 143), adenoma (*n* = 156), to CRC (*n* = 184) (**Supplementary Table 11**). In particular, 99 fungi-bacteria interactions were only associated with CRC, involving the significant correlation between *A. rambellii* and *F. nucleatum*. The 17 pairs of relationships represented by *A.rambellii-P.micra* were only significant in disease states (adenoma and CRC) (**Supplementary Table 11**). Collectively, these findings implicated that trans-kingdom interactions between enteric fungi and bacteria could be associated with CRC progression.

Differential correlation of trans-kingdom microbial interactions in CRC progression

We further investigated whether fungi-fungi and fungi-bacteria interactions were distinct in CRC progression. By comparing the correlation strength between CRC and healthy controls, we showed that the fungi-fungi interaction network (z-score = +4) was different from the bacteria-bacteria network (z-score = -2) (**Supplementary Figure 6A**). Such disparity was confirmed by the trans-kingdom fungi-bacteria interaction network in which two separated peaks were observed (z-score = -2 and +4). Nine classes of differential paired correlation between CRC and controls were then defined (**Supplementary Figure 6B**). More positive fungi-fungi correlations were observed in healthy controls (class “0/+”, 30.77% of total fungi-fungi correlations; the left and right symbol in each class represents the status of correlation in CRC and controls, respectively), while the amount of positive correlations decreased along CRC progression. The majority of fungi-bacteria correlations was classified to the class "+/0" (56.72% of total fungi-bacteria correlations), indicating that these trans-kingdom interactions were positive in CRC but insignificant in healthy controls.

To visualize the differential interaction network, DGCA was used and microbes with similar differential correlation strength (z-score) were grouped together by affinity propagation clusters23. In general, the core set of 33 fungal species plus 31 bacteria with differential abundance (**Supplementary Table 10**) were separated into 6 clusters (**Figure 4**). Two of these 6 clusters showed dominance, thus they were termed as Fun\_cluster (81.8% microbes involved were fungi) and Bac\_cluster (81.0% microbes involved were bacteria). Whilst these two large clusters also implied the difference between fungi-fungi and bacteria-bacteria interactions in CRC progression. In particular, three well-established CRC-enriched bacteria, *F. nucleatum*, *F. periodonticum*, and *P. micra,* were also included in the Fun\_cluster and displayed strong correlations with other fungi. For instance, the fungal *A. rambellii* had a z-score of -5.95 and -5.07 in correlation with *F. nucleatum* and *P. micra*, respectively (both correlations belonged to the class “+/0”; **Supplementary Table 11**), implicating pathogenic bacteria could have closer interactions with fungi than other bacteria in CRC. For the Bac\_cluster, it involved several probiotic species including *Streptococcus thermophilus*, *Streptococcus salivarius*, and *Eubacterium eligens*. Altogether, our differential correlation analysis revealed trans-kingdom interactions among species of fungi and bacteria in CRC progression.

Improved diagnostic performance of CRC using biomarker panel with combined fungi and bacteria

Previous studies have shown the potential of using bacterial biomarkers for CRC diagnosis16,19. We therefore evaluated whether enteric fungi could also be used as diagnostic markers of CRC. Among the core set of 33 fungi and 31 bacteria, only 6 of them had an average AUC greater than 60% in discriminating CRC patients from healthy controls across all eight cohorts, including 2 fungi (*A. rambellii* and *A. kawachii*) and 4 bacteria (*F. nucleatum*, *P. micra*, *G. morbillorum*, and *Porphyromonas asaccharolytica*) (**Table 2**). *P. micra* had the best diagnostic accuracy with average AUC of 67.79%, yet such performance was considered to be clinically unsatisfactory31. To enhance the accuracy, we next employed a biomarker panel with combinations of fungi and bacteria. We trained the panel for CRC diagnosis with pure fungi (*n* = 17), bacteria (*n* = 12), or mixture of both (5 fungi and 9 bacteria) (**Figure 5A** and **Supplementary Figure 7**). In two cohorts (2016\_VogtmannE and 2019\_WirbelJ), the accuracy of panel with pure fungi in CRC diagnosis was higher than the panel with pure bacteria. When comparing to the panel with pure bacteria, the accuracy of panel with combined fungus and bacteria was greatly improved with AUC reaching over 80% in five out of eight cohorts (AUC relative change increased by 1.44%-10.60%) (**Figure 5B**). Hence, our results demonstrated that a diagnostic panel with mixture of fungal and bacterial markers is more accurate than conventional panels with pure bacteria to identify CRC patients from healthy individuals.

DISCUSSION

In this study, we retrieved metagenomic datasets from 7 previous studies of CRC microbiota and established an additional cohort to uncover the landscape of fungal mycobiota in CRC. To our knowledge, this study is the first meta-analysis specifically focusing on enteric fungi in CRC progression. The association of bacteria with CRC has been well-acknowledged, while fungi are often disregarded due to their relatively low abundance in the gut microbiota32. Such low abundance has greatly affected investigations on these fungi as a single metagenomic dataset may not have sufficient sequencing reads and sample size to obtain a comprehensive profile of enteric mycobiota. Meanwhile, by combining results from comparable studies, multi-cohort analysis could improve statistical power and allow reduction of cohort-specific bias33. Here, we revealed the alteration of enteric mycobiota in CRC patients from multiple independent cohorts. Using rank-sum test and SSTF, 33 fungi that are significantly associated with CRC across 8 cohorts were identified, including 10 CRC-enriched and 23 CRC-depleted fungal species.

In particular, *A. rambellii* was found to be the most significant CRC-enriched fungal species.

This finding was supported by previous studies showing the ability of *A. rambellii* in synthesizing carcinogenic products, aflatoxin and aflatoxin precursor sterigmatocystin34,35. Aflatoxin G, aflatoxin B, aflatoxin M, and sterigmatocystin could all be produced by *A.rambellii*36, and all of them were the most toxic and carcinogenic mycotoxins, due to their extreme hepatocarcinogenesis37. XXX. Even though the genus *Aspergillus* was frequently reported with cancer, especially *A. flavus*38, it was the first time to report the promoting cancer cell proliferation function of *A. rambellii* and detected in human fecal.

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, *A. kawachii* has been shown to be atoxigenic and acceptable for use in food and beverage fermentation39. In addition, the crude enzyme extract derived from *A. kawachii* could enhance the antioxidative activities of Viscum album var. coloratum40, a promising agent for immunomodulation, treating colon cancer41. The fermented silkworm produced by *A. kawachii* solid-state fermentation could inhibit the human hepatocellular carcinoma cells42. These previous studies indicate that *A. kawachii* and its metabolites might be used to cure cancer or inhibit cancer cells proliferation. It could also explain why the genus *Aspergillus* cannot drive pancreatic oncogenesis7. *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 effects43,44. *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 effects45,46. These findings support their roles of the enriched- or depleted-fungi in the promotion or inhibition of colorectal carcinogenesis.

Microbes are known to coexist in human gut microbiota, whereas interactions among them could be linked with health and disease. For example, our previous study reported that co-exclusive bacterial correlations become more common in CRC25. Here, we evaluated the interactions between enteric fungi, and revealed the significant differences in fungal correlation network among CRC patients, patients with adenoma, and healthy controls. Four CRC-enriched fungi including *A.* *rambellii*, *E.* *pulchra*, *T.* *punctulata*, and *S.* *musiva* displayed co-occurrence centralities in the correlation network of CRC, of which the correlation between *A. rambelli* and other fungi is greatly altered in CRC (**Figure 3**). Our correlation analyses and results of *in vitro* studies altogether suggest that *A. rambelli* could be the major contributor in mediating fungal dysbiosis along CRC progression. Further studies are therefore recommended to unravel the functional relationship between *A. rambelli* and other enteric fungi in colorectal tumorigenesis.

Given that majority of microbes in the human gut are bacteria, we, therefore, assessed intra-kingdom microbial interactions among fungi and bacteria and clustered microbes with similar differential correlation strength. Our results demonstrated that fungi-fungi and bacteria-bacteria differential correlations are significantly different which could be grouped into two separated clusters (Fun\_Cluster and Bac\_Cluster, respectively; **Figure 4**). Of note, fungi-fungi correlations were found to be diminished in CRC, in contrast to bacteria-bacteria correlations which showed enhanced strength across CRC progression. These findings suggested that enteric fungi and bacteria could have antagonistic interplays in CRC: fungal interactions in healthy individuals could be disrupted by altered bacterial composition in colorectal tumorigenesis. An increase in bacteria-bacteria correlations may impact the community of enteric fungi, thereby providing a favorable condition for the growth of pathogenic bacteria.

To further assess how microbial interactions are involved in CRC progression, we evaluated the trans-kingdom correlations between enteric fungi and bacteria. Our results revealed that majority of fungi-bacteria correlations are positive in CRC but insignificant in healthy controls. As fungi-fungi interactions were shown to be negatively associated with CRC, these results implied that the marked increase in bacterial interactions could dominate the fungal mycobiota to contribute CRC. On the other hand, some fungal species are closely linked with pathogenic bacteria, as exemplified by the significant positive association of withthe well-established CRC-associated bacteria 4,47,48*.* Consistently, previous studies have reported the co-occurrence correlations between several fungal and bacterial species in CRC progression6,49. Human gut microbiota comprises of microbes from different taxonomic kingdoms including bacteria, fungi and viruses, thus it is reasonable to expect that trans-kingdom microbial interactions are frequently occurred in both health and disease conditions. A study in 2018 reported the trans-kingdom interactions between enteric bacteria and virus in CRC11, whereas here our differential correlation analysis revealed that interactions among fungi and bacteria are involved and contribute to CRC progression. It is noteworthy that these findings are solely based on metagenomic sequencing analysis, hence mechanistic investigation is required for validation of such trans-kingdom microbial interactions in colorectal tumorigenesis.

Numerous studies have shown the potential of using bacterial biomarkers for CRC diagnosis16,19,28. Here, we demonstrated that fungal species in fecal samples could also be used as biomarkers to identify CRC patients from 8 cohorts with an average AUC of 73%. Whereas a biomarker panel with mixture of fungi and bacteria had great improvement in diagnostic performance with an average AUC of 83%. Increase in AUC relative change by 1.44%-10.60% across seven of eight cohorts was observed when compared to the panel with pure bacterial biomarkers. Measuring fecal abundances of microbes has been increasingly acknowledged worldwide due to its lower invasiveness than the conventional endoscopic screening for malignancy. However, identification of signature microbes at population level has been challenging, as the gut microbiota in different populations is highly heterogeneous10. Previous meta-analyses by Wirbel *et al.* and Thomas *et al.* have reported several CRC-enriched bacterial species and their capabilities as biomarkers of CRC diagnosis16,19. In this study, our multi-cohort analysis demonstrated that combined use of fungal and bacterial biomarkers could further enhance diagnostic performance, thus highlighting the potential use of enteric fungi in clinical application.

In summary, our meta-analysis revealed the alteration in enteric fungal mycobiota in colorectal tumorigenesis. We identified 33 fungal species that are consistently associated with CRC in multiple cohorts, particularly *A. rambellii* which could XXX (*in vitro* results). We unraveled the trans-kingdom interactions between enteric fungi and bacteria of which bacteria could override the fungal community in CRC progression. While some fungi including *A. rambelli* are positively linked with pathogenic bacteria to contribute CRC. Moreover, we highlighted the potential of fungi as biomarkers for CRC diagnosis. Overall, this study illustrated the crucial involvement of enteric fungi and their trans-kingdom interactions with bacteria in CRC. Further work to depict the functional consequences of dysbiosis fungal mycobiota is warranted for deeper understanding of the mechanistic roles of fungi in colorectal tumorigenesis.

**REFERENCES**

1. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **n/a**,.

2. Lin, Y., Wang, G., Yu, J. & Sung, J. J. Y. Artificial intelligence and metagenomics in intestinal diseases. *Journal of Gastroenterology and Hepatology* **36**, 841–847 (2021).

3. Wong, S. H. & Yu, J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nature Reviews Gastroenterology & Hepatology* **16**, 690–704 (2019).

4. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).

5. Dai, Z. *et al.* Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* **6**, 70 (2018).

6. Coker, O. O. *et al.* Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. *Gut* **68**, 654–662 (2019).

7. Aykut, B. *et al.* The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature* **574**, 264–267 (2019).

8. Underhill, D. M. & Iliev, I. D. The mycobiota: interactions between commensal fungi and the host immune system. *Nat Rev Immunol* **14**, 405–416 (2014).

9. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome Biology* **14**, R51 (2013).

10. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).

11. Nakatsu, G. *et al.* Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. *Gastroenterology* **155**, 529-541.e5 (2018).

12. Jones, M. B. *et al.* Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci USA* **112**, 14024–14029 (2015).

13. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* **10**, 766 (2014).

14. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nature Communications* **6**, 6528 (2015).

15. Hannigan, G. D., Duhaime, M. B., Ruffin, M. T., Koumpouras, C. C. & Schloss, P. D. Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome. *mBio* **9**, (2018).

16. Thomas, A. M. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nature Medicine* **25**, 27 (2019).

17. Yachida, S. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine* **25**, 27 (2019).

18. Vogtmann, E. *et al.* Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PLoS ONE* **11**, e0155362 (2016).

19. Wirbel, J. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine* **25**, 27 (2019).

20. Schulze, J. & Sonnenborn, U. Yeasts in the Gut: From Commensals to Infectious Agents. *Dtsch Arztebl Int* **106**, 837–842 (2009).

21. McKenzie, A. T., Katsyv, I., Song, W.-M., Wang, M. & Zhang, B. DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC Systems Biology* **10**, 106 (2016).

22. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504 (2003).

23. Leone, M., Sumedha & Weigt, M. Clustering by soft-constraint affinity propagation: applications to gene-expression data. *Bioinformatics* **23**, 2708–2715 (2007).

24. Chin, V. K. *et al.* Mycobiome in the Gut: A Multiperspective Review. *Mediators of Inflammation* **2020**, e9560684 (2020).

25. Nakatsu, G. *et al.* Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nature Communications* **6**, 8727 (2015).

26. Stott, K. J., Phillips, B., Parry, L. & May, S. Recent advancements in the exploitation of the gut microbiome in the diagnosis and treatment of colorectal cancer. *Biosci Rep* **41**, BSR20204113 (2021).

27. Loftus, M., Hassouneh, S. A.-D. & Yooseph, S. Bacterial community structure alterations within the colorectal cancer gut microbiome. *BMC Microbiol* **21**, 98 (2021).

28. Liang, Q. *et al.* Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. *Clin Cancer Res* **23**, 2061–2070 (2017).

29. Yu, X. *et al.* A Comparative Characterization of Different Host-sourced Lactobacillus ruminis Strains and Their Adhesive, Inhibitory, and Immunomodulating Functions. *Front. Microbiol.* **8**, (2017).

30. Department of Microbiology, Islamic Azad University School of Science, Fars, Iran *et al.* Anti-colon cancer activity of Bifidobacterium metabolites on colon cancer cell line SW742. *Turk J Gastroenterol* **30**, 835–842 (2019).

31. Mandrekar, J. N. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology* **5**, 1315–1316 (2010).

32. Pérez, J. C. & Johnson, A. D. Regulatory Circuits That Enable Proliferation of the Fungus Candida albicans in a Mammalian Host. *PLOS Pathogens* **9**, e1003780 (2013).

33. Pereira, M. B., Wallroth, M., Jonsson, V. & Kristiansson, E. Comparison of normalization methods for the analysis of metagenomic gene abundance data. *BMC Genomics* **19**, 274 (2018).

34. Cary, J. W., Ehrlich, K. C., Beltz, S. B., Harris-Coward, P. & Klich, M. A. Characterization of the Aspergillus ochraceoroseus aflatoxin/sterigmatocystin biosynthetic gene cluster. *Mycologia* **101**, 352–362 (2009).

35. Frisvad, J. C., Skouboe, P. & Samson, R. A. Taxonomic comparison of three different groups of aflatoxin producers and a new efficient producer of aflatoxin B1, sterigmatocystin and 3-O-methylsterigmatocystin, Aspergillus rambellii sp. nov. *Systematic and Applied Microbiology* **28**, 442–453 (2005).

36. Navale, V., Vamkudoth, K. R., Ajmera, S. & Dhuri, V. Aspergillus derived mycotoxins in food and the environment: Prevalence, detection, and toxicity. *Toxicology Reports* **8**, 1008–1030 (2021).

37. Perrone, G. & Gallo, A. Aspergillus Species and Their Associated Mycotoxins. in *Mycotoxigenic Fungi* (eds. Moretti, A. & Susca, A.) vol. 1542 33–49 (Springer New York, 2017).

38. Uka, V. *et al.* Chemical repertoire and biosynthetic machinery of the Aspergillus flavus secondary metabolome: A review. *Compr Rev Food Sci Food Saf* **19**, 2797–2842 (2020).

39. Hong, S.-B. *et al.* ASPERGILLUS LUCHUENSIS, AN INDUSTRIALLY IMPORTANT BLACK ASPERGILLUS IN EAST ASIA. *PLOS ONE* **8**, e63769 (2013).

40. Kim, S.-Y., Yang, E.-J., Son, Y. K., Yeo, J.-H. & Song, K.-S. Enhanced anti-oxidative effect of fermented Korean mistletoe is originated from an increase in the contents of caffeic acid and lyoniresinol. *Food Funct.* **7**, 2270–2277 (2016).

41. Khil, L.-Y. *et al.* Mechanisms involved in Korean mistletoe lectin-induced apoptosis of cancer cells. *WJG* **13**, 2811 (2007).

42. Cho, H.-D. *et al.* Solid state fermentation process with Aspergillus kawachii enhances the cancer-suppressive potential of silkworm larva in hepatocellular carcinoma cells. *BMC Complement Altern Med* **19**, 241 (2019).

43. Ma, L., Tang, L. & Yi, Q. Salvianolic Acids: Potential Source of Natural Drugs for the Treatment of Fibrosis Disease and Cancer. *Frontiers in Pharmacology* **10**, 97 (2019).

44. Das, G. *et al.* Cordyceps spp.: A Review on Its Immune-Stimulatory and Other Biological Potentials. *Frontiers in Pharmacology* **11**, 2250 (2021).

45. Boskovic, I., Đukić, D. A., Maskovic, P., Mandić, L. & Perovic, S. Phytochemical composition and antimicrobial, antioxidant and cytotoxic activities of Anchusa officinalis L. extracts. *Biologia* **73**, 1035–1041 (2018).

46. Luo, C. *et al.* A Review of the Anti-Inflammatory Effects of Rosmarinic Acid on Inflammatory Diseases. *Frontiers in Pharmacology* **11**, 153 (2020).

47. Yu, T. *et al.* Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* **170**, 548-563.e16 (2017).

48. Guo, S. *et al.* Exosomes derived from *Fusobacterium nucleatum* -infected colorectal cancer cells facilitate tumour metastasis by selectively carrying miR-1246/92b-3p/27a-3p and CXCL16. *Gut* **70**, 1507–1519 (2021).

49. Sovran, B. *et al.* Enterobacteriaceae are essential for the modulation of colitis severity by fungi. *Microbiome* **6**, 152 (2018).

FIGURE LEGENDS

**Figure 1.** **Multi-cohorts analysis identified a set of intestinal fungi strongly associated with CRC. (a)** Among 296 non-rare abundance candidates, the meta-analysis significant of fungi (n = 33, FDR < 0.01; n = 74, FDR < 0.1) derived from Mann-Whitney U test and adjusted the p-value with the conservative Bonferroni correction. **(b)** The code set (n = 33) performance across cohorts. The heatmaps revealed the two-sided Wilcoxon test and generalized fold change within individual studies. **(c)** The core set (n = 33) with pair fold change among all cohorts. The blue and red bar represent the enriched- and depleted-CRC fungi, respectively. And highlight the candidates whose fold change is larger than two times or less than a half. **(d)** Violin graph for the two outstanding performance fungi, *A. rambellii* and *A. kawachii*, in different studies. The healthy control and CRC groups are shown by green and red colors, respectively.

**Figure 2. *Aspergillus rambellii* culture medium exhibit colorectal cancer-promoting effects in vitro and vivo. (a)** *A.r* CM (10.0%) promoted both norma epithelial cell (NCM460) and CRC (SW480, HT29) cell viability. SDB was utilized as a control. **(b)** *A.r* CM supported the colony formation of CRC cells. **(c)** Human colon cancer organoid growth enhanced after *A.r* CM treatment. Representative images of *A.r* CM and SDB treated organoids (upper panel, scale bar= ), Surface area (down panel, 104 µm2) of organoids in each observation field were measured. Data are expressed as mean ± SD. Statistical significance was determined by Student’s T test.

**Figure 3. Global correlations among altered fungal and bacterial in CRC compared with the healthy control and adenoma. (a)** Left panel: correlation between the 33 selected fungi and 31 selected bacterial candidates in CRC samples. Right panel: the top intra-fungal relationship in CRC. **(b)** and **(c)** The correlations in adenoma and healthy control using the same method, respectively. The pink diamond, blue circle, and green circle represented a substantially different paired correlation in CRC against healthy control, CRC versus adenoma, and adenoma versus healthy control, respectively.

**Figure 4. Differential relationships between CRC and healthy control.** Network for the differential correlation between CRC and control. The differential correlation index (z-score) was used to evaluate the differences in species-species regulatory interactions under various conditions (CRC vs healthy control). Six clusters were automatically separated through the methodology affinity propagation cluster. The two primary clusters were termed Fun\_cluster and Bac\_cluster, respectively, because they mostly included fungi and bacteria.

**Figure 5. Feature ranking and performance comparison of fungi-bacteria combination CRC-diagnosis models across 8 cohorts. (a)** Top panel: The importance of each species for the cross-validation prediction performance in each cohort estimated using the internal random forest scores. Only species appearing in the three top-ranking features in at least one dataset were reported. Bottom panel: AUC of pure fungal, bacterial, and fungal-bacterial CRC-diagnosis models across 8 cohorts, and the relative changes between bacterial and combined CRC-diagnosis models. **(b)** Diagnostic performance of bacterial and combined classification of CRC from control by random forest with LOSO.