



# Species-Level Analysis of the Human Gut Microbiome Shows Antibiotic Resistance Genes Associated With Colorectal Cancer

Chuanfa Liu<sup>1,2,3†</sup>, Zhiming Li<sup>2,3†</sup>, Jiahong Ding<sup>2,3†</sup>, Hefu Zhen<sup>2,3</sup>, Mingyan Fang<sup>2,3\*</sup> and Chao Nie<sup>2,3\*</sup>

<sup>1</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, <sup>2</sup> BGI-Shenzhen, Shenzhen, China, <sup>3</sup> China National GeneBank, BGI-Shenzhen, Shenzhen, China

## OPEN ACCESS

### Edited by:

Ravi Kant,  
University of Helsinki, Finland

### Reviewed by:

Shenghui Li,  
China Agricultural University, China  
Seungha Kang,  
University of Queensland, Australia

### \*Correspondence:

Mingyan Fang  
fangmingyan@genomics.cn  
Chao Nie  
niechao@genomics.cn

† These authors share first authorship

### Specialty section:

This article was submitted to  
Antimicrobials, Resistance  
and Chemotherapy,  
a section of the journal  
Frontiers in Microbiology

Received: 26 August 2021

Accepted: 11 November 2021

Published: 15 December 2021

### Citation:

Liu C, Li Z, Ding J, Zhen H,  
Fang M and Nie C (2021)  
Species-Level Analysis of the Human  
Gut Microbiome Shows Antibiotic  
Resistance Genes Associated With  
Colorectal Cancer.  
Front. Microbiol. 12:765291.  
doi: 10.3389/fmicb.2021.765291

Colorectal cancer (CRC) is the second leading cause of cancer deaths and continuously increases new cancer cases globally. Accumulating evidence links risks of CRC to antibiotic use. Long-term use and abuse of antibiotics increase the resistance of the gut microbiota; however, whether CRC is associated with antibiotic resistance in gut microbiota is still unclear. In this study, we performed a *de novo* assembly to metagenomic sequences in 382 CRC patients and 387 healthy controls to obtain representative species-level genome bins (rSGBs) and plasmids and analyzed the abundance variation of species and antibiotic resistance genes (ARGs). Twenty-five species and 65 ARGs were significantly enriched in the CRC patients, and among these ARGs, 12 were multidrug-resistant genes (MRGs), which mainly included *acrB*, *ToIC*, *marA*, *H-NS*, *Escherichia coli acrR* mutation, and *AcrS*. These MRGs could confer resistance to fluoroquinolones, tetracyclines, cephalosporins, and rifamycin antibiotics by antibiotic efflux and inactivation. A classification model was built using the abundance of species and ARGs and achieved areas under the curve of 0.831 and 0.715, respectively. Our investigation has identified the antibiotic resistance types of ARGs and suggested that *E. coli* is the primary antibiotic resistance reservoir of ARGs in CRC patients, providing valuable evidence for selecting appropriate antibiotics in the CRC treatment.

**Keywords:** antibiotic resistance gene (ARG), colorectal cancer (CRC), human gut metagenome, species-level genome bins, *Escherichia coli*

## INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers worldwide and has led to nearly 1 million deaths in 2020 only (Ferlay et al., 2021). Many factors are associated with an elevated risk of CRC, including genetic predisposition, colorectal polyps, inflammatory bowel disease, smoking, and alcohol intake (Wang et al., 2014). Accumulating evidence suggests that long-term, frequent, and/or combined antibiotic use could also be risk factors for CRC (Wang et al., 2014;

Dik et al., 2016; Cao et al., 2018; Crockett and Nagtegaal, 2019; Zhang et al., 2019). Antibiotics, such as metronidazole, ciprofloxacin, and rifaximin, are frequently used to treat colitis and intestinal lesions (Bernstein et al., 2016; Nitzan et al., 2016). During the long-term development and progression of CRC, the detrimental effect of antibiotics may be present even at the early stage of colitis, adenomatous polyps, or other precursors of the CRC. It is worth noting that antibiotic use could increase the richness of antibiotic-resistance bacterial species and the abundance of antibiotic resistance genes (ARGs) in the gut microbiota (Casals-Pascual et al., 2018; Dubinsky et al., 2020). Subsequent to antibiotic use and increased resistance, bowel dysbacteria may occur, and concomitantly, colonization resistance, and mucus production of the colon mucosal may be impaired (Becattini et al., 2016; Schwartz et al., 2020). Existing literature indicates that gut microbiota dysbiosis and colon mucosal surface changes are associated with the occurrence and progression of CRC (Yachida et al., 2019; Cheng et al., 2020; Xing et al., 2021). Therefore, research on drug-resistant microbiota and resistance genes may help to understand the progression of CRC.

The compositional patterns of antibiotic-resistant species and ARGs in the gut microbiota of CRC patients were scantily studied. To examine their potential effects exerted upon CRC patients and healthy people, we have downloaded published human gut metagenomic data of CRC patients and healthy controls to study the antibiotic resistance species and ARG distribution in the gut microbiota. We performed the metagenomic assembly to obtain representative species-level genome bins (rSGBs) to investigate ARG abundance in each species. Based on the Genome Taxonomy Database (GTDB) and Comprehensive Antibiotic Resistance Database (CARD), we annotated species and ARGs in rSGBs and analyzed their abundance. These analyses have revealed how the burden of antibiotic resistance changes in the intestine of CRC patients, stressing significant associations between these changes and microbiota composition. Our study has characterized the resistance of the gut microbiota in CRC patients and may shed new light on the proper antibiotic use for avoiding drug resistance.

## RESULTS

### Reconstruction and Annotation of Microbial Genomes and Plasmids

In this study, we downloaded metagenomic data of 382 CRC patients (the CRC group) and 387 healthy controls (CTR group) from eight studies (Table 1 and Supplementary Table 1). The genome reconstruction was performed using a pipeline reported by Pasolli et al. (2019) and carried out a *de novo* single-sample metagenomes assembly and binning. More than 23.5 million contigs (mean  $\pm$  SD, 30,688.6  $\pm$  13,972.6) were assembled from these samples; 5,880 high-quality metagenome-assembled genomes (MAGs) and 5,390 medium-quality MAGs were obtained (Supplementary Table 2). After clustering and filtering the rSGBs for the high-quality MAGs, we obtained 696 rSGBs with genome sizes ranging from 0.95 to 6.41 Mb (2.39  $\pm$  0.43 Mb) (Supplementary Tables 2–4). We then aligned

the high-quality sequencing reads to the 696 rSGBs. The read mapping rate in our results (76.6%  $\pm$  7.8%, Supplementary Table 2) was similar to that of a large-scale gut microbiota study (range, 67.76–87.51%) (Pasolli et al., 2019). Based on the quality of the mapping rate, it is acceptable to use our data for subsequent species and ARG annotations.

Thereafter, we classified rSGBs using the GTDB Toolkit (GTDB-Tk, see *Methods* for details on the taxonomy nomenclature used) (Chaumeil et al., 2019). However, 60 rSGBs (8.62%) could not be assigned to an existing species, and 402 (57.75%) rSGBs belonged to uncultured species (Figure 1 and Supplementary Table 5). All 696 rSGBs were classified into 13 phyla and 306 genera. In line with previous studies (Yeoh et al., 2020), Firmicutes (including Firmicutes A), Bacteroidota, Actinobacteriota, and Proteobacteria were predominant phyla, and the total relative abundance accounted for more than 90% of the gut microbiota (mean  $\pm$  SD, CRC group: 93.69%  $\pm$  8.76%; CTR group: 96.67%  $\pm$  5.75%) (Supplementary Figure 1A). *Bacteroides*, *Phocaeicola*, *Faecalibacterium*, *Prevotella*, *Alistipes*, and *Blautia* A were the dominant genera in the gut (Figure 2A). It indicated that our rSGBs covered the dominate species in the gut microbiota.

In addition, we assembled the plasmids of gut microbiota using metaplasmidSPAdes (Antipov et al., 2019). We obtained 24,692 plasmid-sourced contigs (N50 = 42,448 bp; max = 473,623 bp; min = 2,564 bp) with a mean of 32 contigs in each sample. The plasmid-sourced genes were predicted and clustered with MetaGeneMark and cd-hit, respectively (Li and Godzik, 2006; Zhu et al., 2010). A non-redundant plasmid-sourced gene catalog (159,890 genes, N50 = 701 bp) was obtained. Next, we applied reference-based taxonomy annotation of the gene catalog using the NCBI-NT database. Finally, 157,504 (98.5%) of the genes in the gene catalog could be uniquely and reliably assigned to a species. We found those genes mainly from *Escherichia coli*, *Faecalibacterium prausnitzii*, *Bacteroides dorei*, *Bacteroides fragilis*, *Bacteroides uniformis*, and *Klebsiella pneumoniae*.

### Alterations of Gut Microbial Composition in Colorectal Cancer and CTR Groups

We analyzed the effect size of cohorts and host characteristics on the variance of the gut microbiome by permutational multivariate analysis of variance (PERMANOVA) test, where results revealed the factor “cohort” to have a predominant impact on the species and ARG composition of the subjects (Supplementary Figure 2). To test the accuracy of results in the analysis, we select two cohorts randomly to confirm our species and ARG results [PRJEB7774 ( $n = 109$ ) and PRJEB12449 ( $n = 104$ )]. The analysis of the  $\alpha$  and  $\beta$  diversity of microbial composition revealed that the CRC group had a slightly lower species diversity than the CTR group [Shannon–Wiener index ( $H'$ ),  $p = 0.24$ , Figure 2B], which was consistent with the literature (Gao et al., 2015). The dimensionality-reduction analysis [principal coordinate analysis (PCoA) and non-metric multidimensional scaling (NMDS) analysis] of the rSGBs relative abundance of the CRC and CTR groups showed that the CRC and CTR groups

**TABLE 1** | Characteristics of the data sets included in this study.

Accession number <sup>a</sup>	Group (n <sup>b</sup> )	Age (years) (mean ± SD <sup>c</sup> )	Gender (F/M,% <sup>d</sup> )	BMI (kg/m <sup>2</sup> ) (mean ± SD <sup>c</sup> )	Read counts (mean ± SD <sup>c</sup> )
PRJEB7774 Feng et al. (2015)	CRC (46)	67.07 ± 10.91	39.13/60.87	26.46 ± 3.54	49,936,552 ± 7,270,051
	CTR (63)	67.06 ± 6.37	41.27/58.73	27.57 ± 3.74	46,090,975 ± 7,068,141
PRJNA389927 Hannigan et al. (2018)	CRC (28)	58.86 ± 11.02	28.57/71.43	28.57/71.43	4,851,235 ± 2,209,656
	CTR (28)	55.46 ± 9.52	60.71/39.29	60.71/39.29	5,848,176 ± 3,578,299
PRJEB10878 Yu et al. (2017)	CRC (74)	66.04 ± 10.60	35.14/64.86	23.98 ± 3.16	49,225,941 ± 10,554,038
	CTR (54)	61.76 ± 5.67	38.89/61.11	23.46 ± 2.96	53,622,545 ± 8,128,850
PRJEB6070 Zeller et al. (2014)	CRC (91)	64.66 ± 12.23	40.66/59.34	26.04 ± 4.49	43,899,076 ± 19,556,571
	CTR (66)	58.61 ± 12.79	50.00/50.00	24.68 ± 3.17	48,152,424 ± 23,181,408
PRJEB27928 Wirbel et al. (2019)	CRC (22)	66.55 ± 10.6	50.00/50.00	25.33 ± 4.93	48,769,786 ± 18,597,297
	CTR (60)	57.57 ± 11.08	46.67/53.33	24.88 ± 3.2	28,494,527 ± 7,144,313
PRJNA447983 Thomas et al. (2019)	CRC (29)	71.45 ± 8.23	20.69/79.31	25.71 ± 4.14	46,797,011 ± 22,256,056
	CTR (24)	67.92 ± 7.01	45.83/54.17	25.32 ± 3.51	58,117,998 ± 40,294,533
PRJDB4176 Yachida et al. (2019)	CRC (40)	59.05 ± 12.83	47.50/52.50	22.36 ± 2.72	40,249,523 ± 12,291,484
	CTR (40)	63.63 ± 12.36	42.50/57.50	22.90 ± 2.44	46,232,480 ± 14,228,285
PRJEB12449 Vogtmann et al. (2016)	CRC (52)	61.85 ± 13.58	28.85/71.15	24.89 ± 4.25	52,664,424 ± 16,619,909
	CTR (52)	61.23 ± 11.03	28.85/71.15	25.34 ± 4.28	52,629,971 ± 12,022,921

<sup>a</sup>References of the study.

<sup>b</sup>Counts of samples.

<sup>c</sup>Standard deviation.

<sup>d</sup>Ratio of the percentage of female and male.

were separated (PERMANOVA analysis  $p = 0.01$ ,  $R = 0.0731$ ) (Figure 2C and Supplementary Figure 1B).

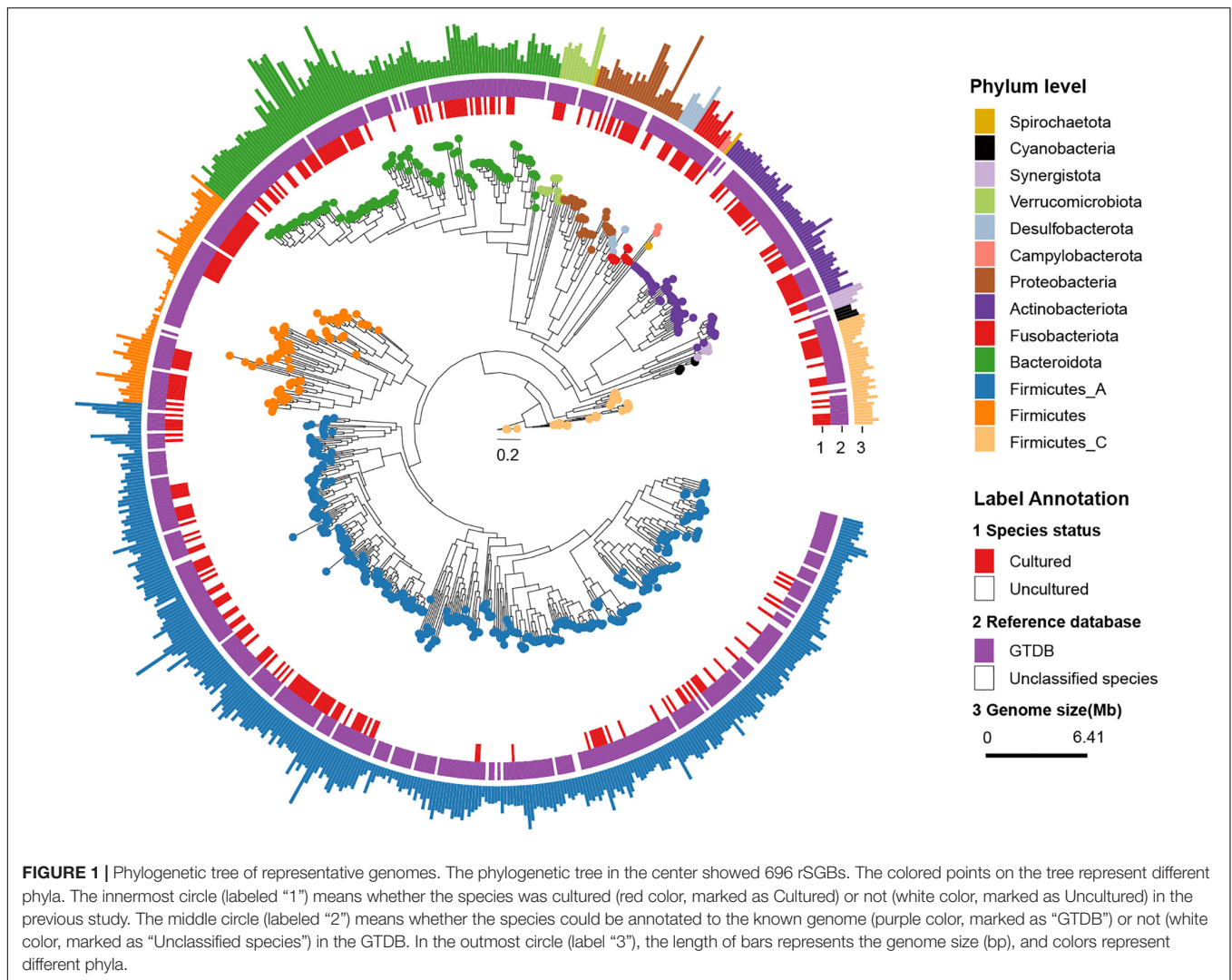
Then, the composition of microbiota between CRC patients and healthy controls was compared at phylum and genus levels. On the phylum levels, Bacteroidota, Desulfobacterota, and Fusobacteriota phyla were enriched in the CRC group, and Firmicutes A phylum was enriched in the CTR group (Wilcoxon test, adjusted  $p < 0.05$ ; Supplementary Figure 1C). On the genus level, *Anaerotignum*, *Bilophila*, *Bulleidia*, *Flavonifractor*, *Gemella*, *Intestinimonas*, *Parvimonas*, *Peptostreptococcus*, *Porphyromonas*, *Prevotella*, and *Ruthenibacterium* genera were enriched in the CRC group; meanwhile, *Agathobacter*, *Anaerostipes*, *Butyrivibrio* A, *Butyrivibrio* A, *CAG-41*, *Eubacterium* G, *Eubacterium* R, *Faecalibacterium*, *GCA-900066135*, *Lachnospira*, *TF01-11*, and *UBA11524* genera were enriched in the CTR group significantly (Wilcoxon test, adjusted  $p < 0.05$ ; Supplementary Figure 1C).

Next, we compared the microbiota composition between CRC patients and healthy controls at species levels using the linear discriminant analysis effect size (LefSe) algorithm (Segata et al., 2011). Within the 25 species enriched in the CRC group (Figure 2D and Supplementary Table 6), nine species had been reported to be increased in the CRC group before, that is, *E. coli* (GTDB classification: *E. coli* D), *Parabacteroides distasonis*, *B. fragilis*, *Porphyromonas* species, *Alistipes fingoldii*, *Alistipes onderdonkii*, *Akkermansia muciniphila*, *Bacteroides thetaiotaomicron*, *Mediterraneibacter torques* (previously named *Ruminococcus torques*), and *Ruminococcus B gnavus* (Zhang et al., 2018; Ai et al., 2019; Dai et al., 2019; Sahankumari et al., 2019; Wong and Yu, 2019; Yang et al., 2019). Moreover, two species were the first discovered species that were enriched in the CRC group, that is, *CAG-180 sp000432435* and *CAG-177 sp003514385*. Among the 35 species enriched in the CTR group (Figure 2D and

Supplementary Table 6), *Anaerostipes hadrus*, *Bifidobacterium catenulatum*, *Fusicatenibacter saccharivorans*, and butyrate-producing species *F. prausnitzii*, *Agathobacter rectalis*, and *Agathobacter faecis* had been reported in the literature as enriched in the healthy controls and possibly beneficial to the gut health (Ai et al., 2019; Kim et al., 2020; Ma et al., 2021). Focusing on the 60 microbiota that exhibited significantly different abundances between the CRC and CTR groups in all samples, we further compared the relative abundances of these microbiota between the CRC and CTR groups in PRJEB7774 and PRJEB12449 cohorts using the Wilcoxon rank-sum tests. We found that the enrichment of 85 and 83.3% of the significantly different species (CRC and CTR) in the PRJEB7774 and PRJEB12449 cohorts were congruent with those in the composite cohort of all samples (Supplementary Figure 3).

## Antibiotic Resistance Genes Conferred to Multiple Antibiotics

To analyze the antibiotic resistance information in microbiota, we characterized the ARGs and analyzed their abundance in the rSGBs and plasmids by annotating them to the CARD (Alcock et al., 2020), obtaining 164 ARGs in 189 rSGBs (Supplementary Tables 7, 8). The top 11 ARGs accounted for 51.16% of all abundance, mainly including *adeF*, *TolC*, *E. coli soxS* mutation, *AcrS*, *E. coli soxR* mutation, and *marA* (Figure 3A and Supplementary Figure 4A). The *adeF* had the highest relative abundance, which could encode the membrane fusion proteins of the multidrug efflux complex AdeFGH (Coyne et al., 2010). Within the plasmid-sourced genes, we obtained 43 ARGs in 49 genes (0.03% of all plasmid genes) that conferred resistance to antibiotics (Supplementary Table 8), and the ARG *tetQ* had the highest abundance.



We analyzed the resistance mechanisms and resistance drug types of identified ARGs and found that these ARGs in the rSGBs and plasmids could confer resistance to 33 and 18 types of antibiotics, respectively. Notably, 53.05% of rSGB-sourced ARGs (87 out of 164) could confer resistance to more than one antibiotic (**Supplementary Table 7**). ARGs could affect antibiotic resistance through the mechanisms of antibiotic efflux, inactivation, target alteration, target protection, target replacement, and reduced permeability to antibiotics.

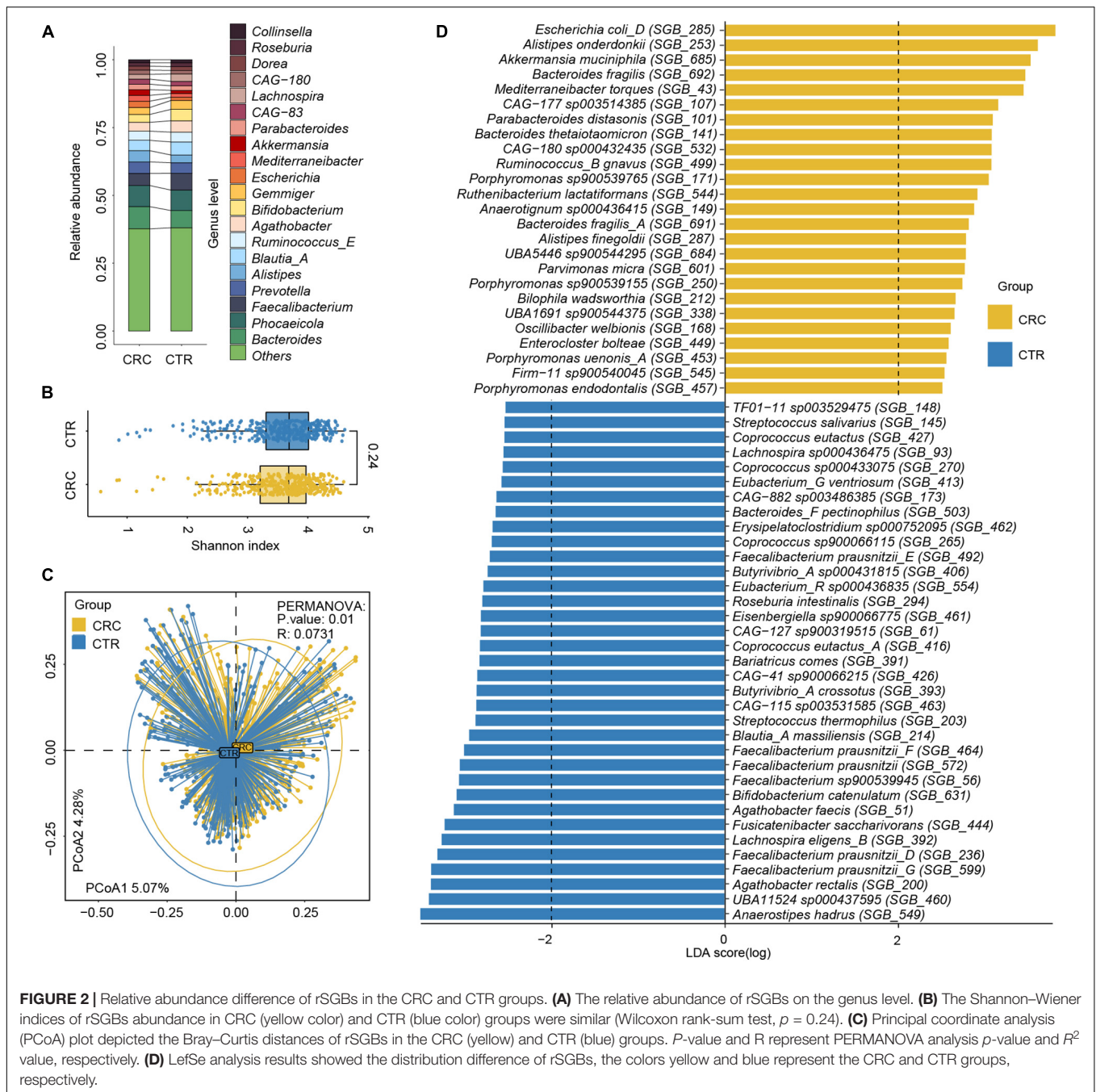
Regarding the percentage of abundance, the antibiotic efflux accounted for 53.36% of rSGB-sourced ARGs resistance mechanism, and antibiotic inactivation and antibiotic targets alteration accounted for 21.21 and 14.49%, respectively (**Figure 3B** and **Supplementary Figure 4B**). Eighty percent of the total rSGB-sourced ARG abundance could be ascribed to the top 10 resistance drug types, including nucleoside antibiotics (17.3%), cephalosporins (16.77%), macrolides (11.64%), phenicol antibiotics (8.06%), fluoroquinolones (5.97%), penams (5.2%), rifamycins (2.96%), and other antibiotics (**Figure 3C** and **Supplementary Figure 4C**).

We found that those with top resistance types were also the most consumed antibiotics globally (Van Boeckel et al., 2014). For example, ARGs in the gut of the CRC group could confer six types of global high-consumption antibiotics, including penicillins, cephalosporins, macrolides, fluoroquinolones, tetracyclines, and rifamycins. The antibiotic resistance types in our findings were consistent with the global antibiotic consumption. Among these antibiotics, 19.23% belong to the Access Class, 30.77% belong to the Watch Class, and 23.08% are part of the Reserve Class (World Health Organization AWaRe classification, version 2019; **Supplementary Table 9** and **Supplementary Figure 4D**; World Health Organization, 2019).

## The Divergence and Heterogeneity of Antibiotic Resistance Gene in the Colorectal Cancer and CTR Groups

We performed a group comparison in the abundance of ARGs, antibiotic mechanisms, and their resistance drug types to analyze ARG variations. First, the component of ARGs in the CRC and



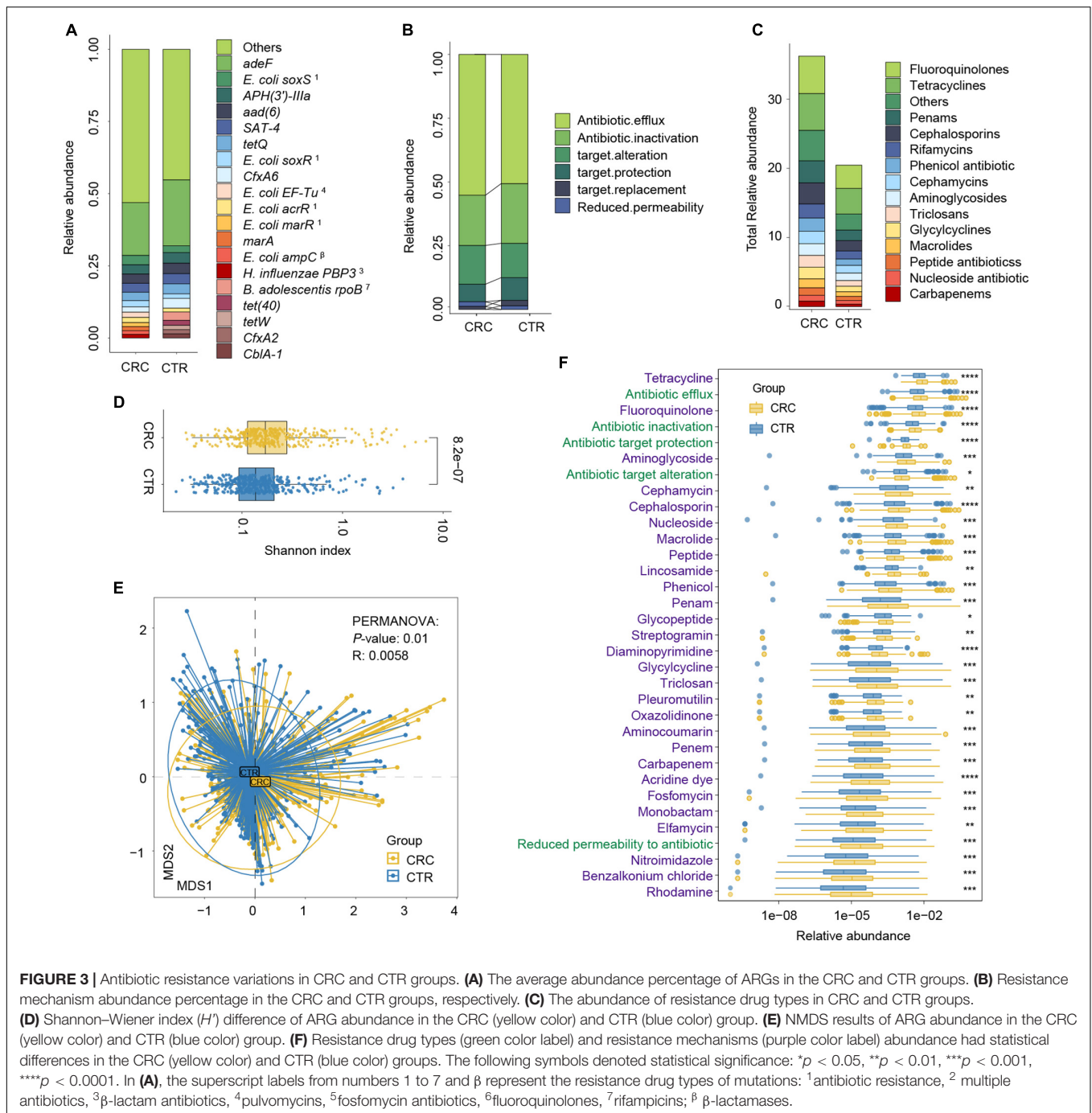


CTR groups were analyzed. Compared with the CTR group, the CRC group had a higher rSGB-sourced ARG diversity [Shannon–Wiener index ( $H'$ ),  $p < 0.001$ , **Figure 3D**], which indicated that gut microbiota in the CRC group had more complexity of ARGs. The  $\beta$  diversity of ARG abundance showed that the ARGs in the CRC group differed from those in the CTR group (NMDS, PERMANOVA:  $p = 0.01$ ,  $R = 0.0058$ ) (**Figure 3E**).

We then analyzed the rSGB-sourced ARG abundance grouped by resistance mechanisms and resistance drug types in the CRC and CTR groups. Eighty percent (33 of 39) resistance mechanisms and resistance drugs were significantly

enriched in the CRC group, mainly including antibiotic efflux and inactivation mechanism, tetracyclines, fluoroquinolones, penams, carbapenem, and cephalosporins (**Figure 3F**). The abundance of ARGs in the rSGBs was significantly enriched in the CRC group on the mechanisms and resistance drug type scales.

To further demonstrate the variability in the resistance burden between the CRC and CTR groups, we analyzed the abundance difference of every ARG. Fifty ARGs from rSGBs were significantly enriched in the CRC group (Wilcoxon test, adjusted  $p < 0.05$ ) (**Figure 4**). Correspondingly, the total abundance of ARGs from plasmids was higher in the CRC group than that in

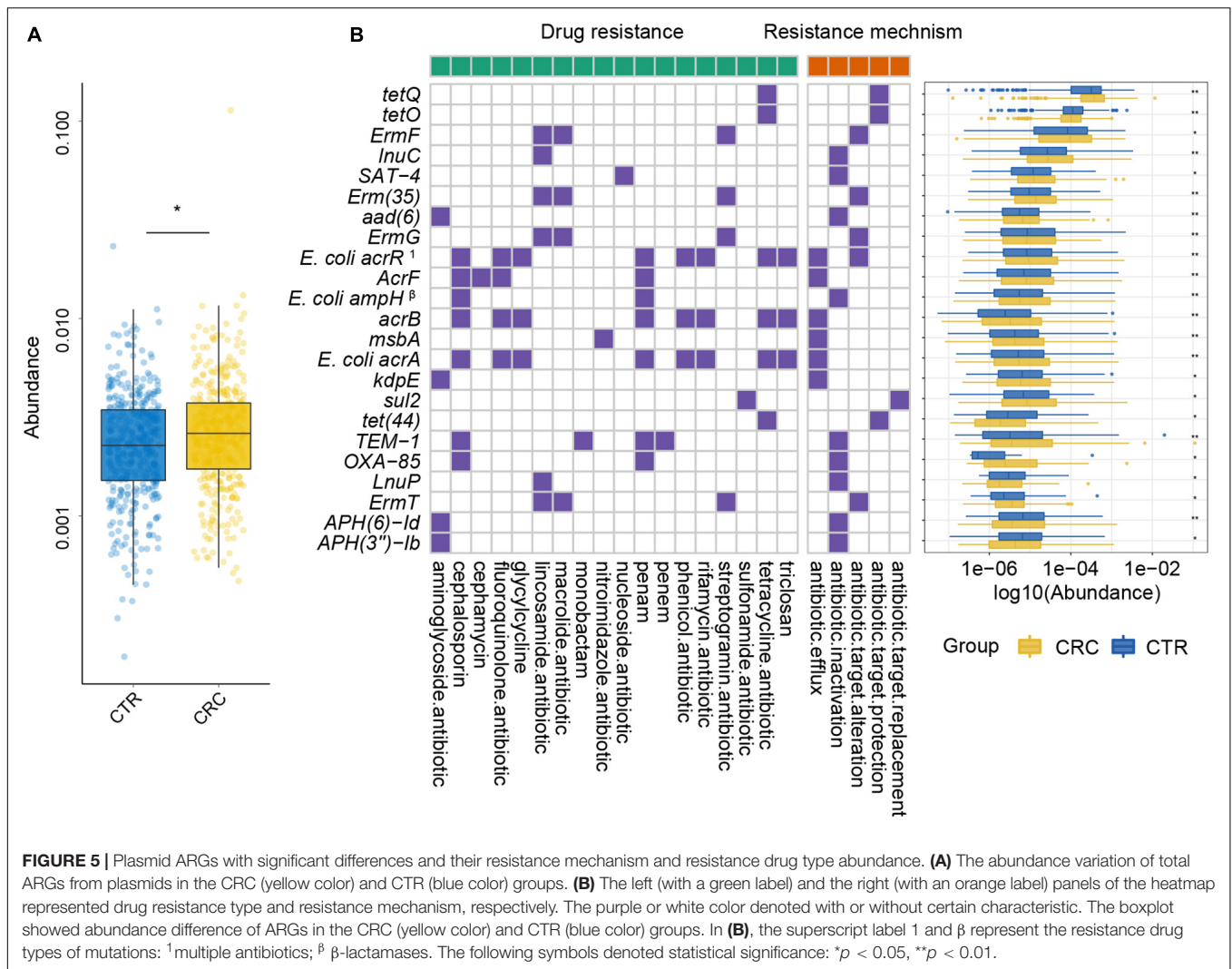


the CTR group (Figure 5A), and 70% of ARGs (16 of 23) were enriched in the CRC group significantly (Wilcoxon test, adjusted  $p < 0.05$ ) (Figure 5B). In our two validation cohorts, 100 and 96% rSGB-sourced ARGs as well as 60.9 and 73.9% plasmid-sourced ARGs in PRJEB7774 and PRJEB12449 cohorts were enriched congruently with all the samples (Supplementary Figures 5, 6). The ARGs from both the rSGBs and plasmids in the validation cohorts showed a notable consistency with those in all samples.

Among these rSGB-sourced ARGs, 90% (45 of 50) were found in the *E. coli*, mainly composing of *TolC*, *E. coli soxS* mutation,

*marA*, *AcrS*, *acrB*, and *mdtM*. The remaining five ARGs, including *adeF*, *CcrA*, *cepA*, *tet(W/N/W)*, and *tetQ*, were found in *P. distasonis*, *B. fragilis* A, *B. fragilis*, *UMGS693 sp900544555*, and *A. onderdonkii*, respectively. Although the ARGs existing in *E. coli*, *Bacteroides* species, and *Alistipes* species had been reported, their association with CRC was not robust (Dubinsky et al., 2020; Parker et al., 2020). In terms of resistance mechanisms and resistance drug types, these rSGB-sourced ARGs enriched in CRC could confer resistance to 33 types of antibiotics by five mechanisms, and 37 of 50 ARGs encoded antibiotic efflux





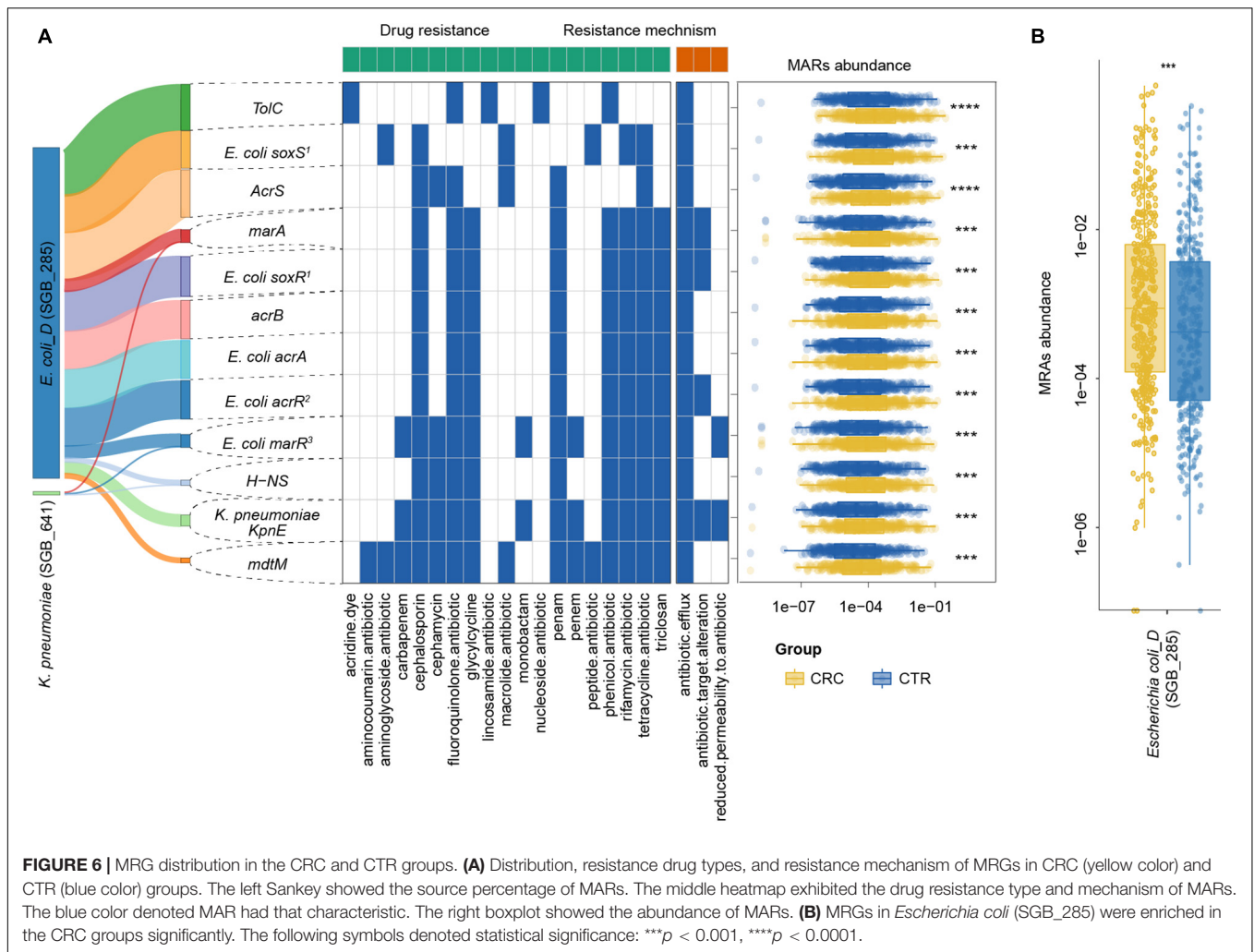
where the antibiotics included cephalosporins, fluoroquinolones, macrolides, tetracycline, and penams. Seven ARGs had resistance to penams and cephalosporins, six ARGs to tetracyclines and lincosamides, respectively.

Interestingly, penicillins, cephalosporins, and fluoroquinolones had been reported to be associated with the onset of colitis (Nitzan et al., 2016). Meanwhile, fluoroquinolones, cephalosporins, tetracyclines, and rifamycins were the most commonly used antibiotics in colitis treatment (Theochari et al., 2018). These results suggested that the increase in resistance burden might be due to antibiotic treatment of precancerous colitis and antibiotic use related to intestinal colitis and CRC by an unknown mechanism.

To further investigate the multidrug resistance of ARGs, we denoted ARGs that could confer resistance to five or more antibiotic drug types as multidrug-resistant genes (MRGs) (Supplementary Table 7). A total of 21 species from rSGBs were found that carried 41 MRGs, of which 12 MRGs were significantly enriched in the CRC group (Supplementary Figure 7). In the plasmid-sourced ARGs, three MRGs were found (*acrB*,

*E. coli acrR* mutation, and *E. coli acrA*) enriched in the CRC group. All these MRGs enriched in the CRC group were derived from *E. coli* and *K. pneumoniae* (Figures 5B, 6A). Meanwhile, compared with the CTR group, the rSGB-sourced MRG abundance in *E. coli* was significantly higher in the CRC group (Figure 6B). These MRGs could confer resistance to 19 types of antibiotic drugs, which included fluoroquinolones, cephalosporins, tetracyclines, penams, phenicol antibiotics, and rifamycins (Figures 5B, 6A). Among these MRGs, *mdtM* could confer resistance to 15 types of antibiotics, where both *E. coli marR* mutant and *K. pneumoniae KpnE* could confer resistance to 12 types of antibiotics, and *H-NS* conferred resistance to 9 types of antibiotics (Figure 6A). The intergroup difference of MAGs was completely confirmed in the PRJEB7774 and PRJEB12449 cohorts, respectively (Supplementary Figure 8). This finding showed that the abundance of MRGs and species with MARs was enriched in the CRC group, which suggested that the increase in MRG abundance was also associated with the CRC. Findings thus far indicated that the CRC group had a higher antibiotic resistance burden and had resistance to multiple antibiotics.





## Species Encoded by Antibiotic Resistance Genes Were Enriched in the Colorectal Cancer Group

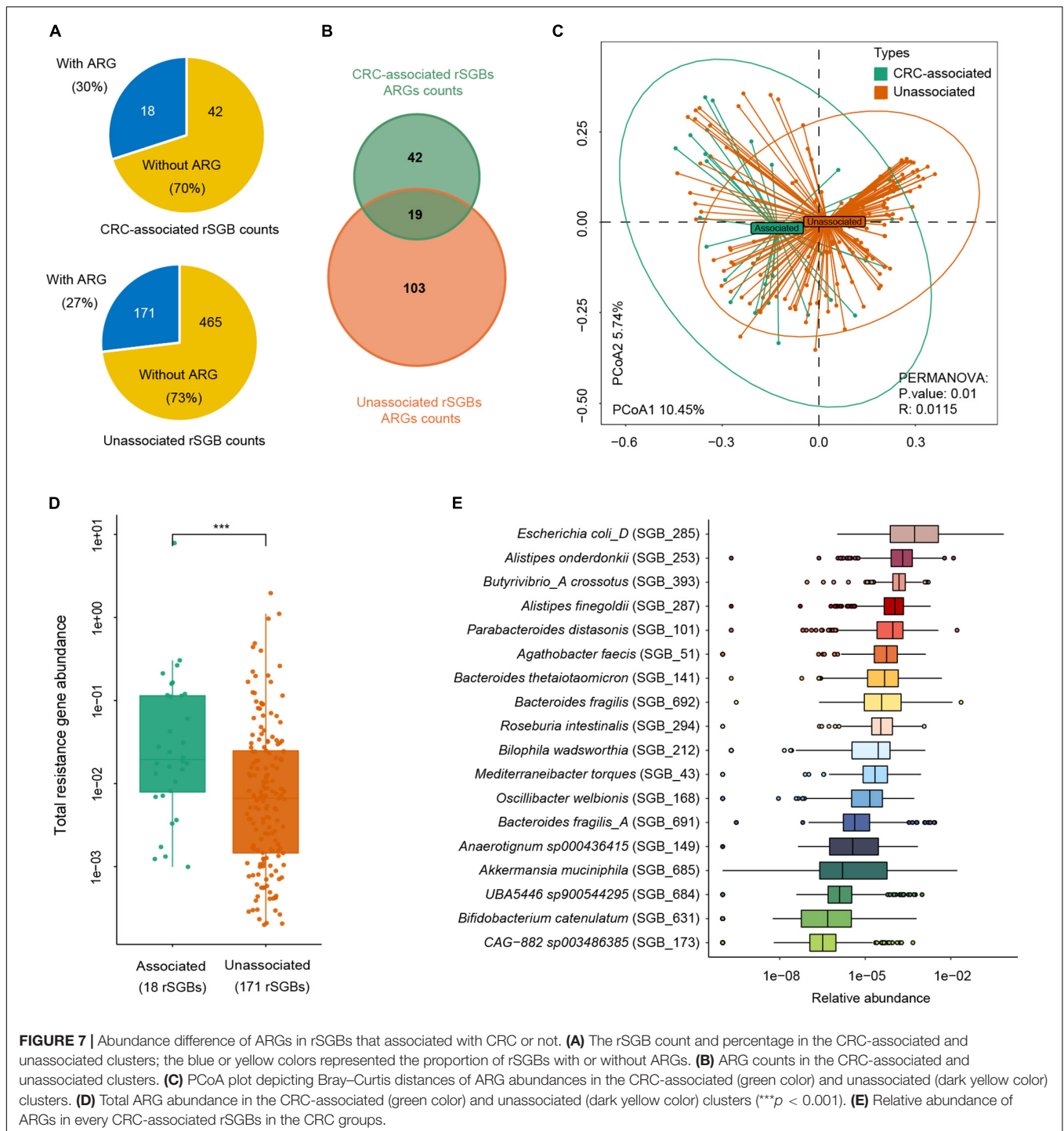
We further analyzed the sum of ARG abundance in the CRC-associated and unassociated species, where species enriched in the CRC and CTR groups were denoted as the “CRC-associated” cluster and others as the “unassociated” cluster. In the CRC-associated cluster, there were 30% of species (18 of 60 species) carrying ARGs; accordingly, there were 27% of species (171 of 636 rSGBs) with ARGs in the unassociated cluster (Figure 7A). Among all ARG types, 61 ARG types are found in CRC-associated clusters, and 122 ARG types are in unassociated clusters, whereas 19 are common to both clusters (Figure 7B). Moreover, the results showed that the CRC-associated cluster and unassociated cluster were divided in the PCoA analysis (PERMANOVA analysis,  $p = 0.01$ , Figure 7C), and the total abundance of ARGs in the CRC-associated cluster was significantly higher than that in the unassociated cluster (Wilcoxon test,  $p < 0.0001$ , Figure 7D).

The analysis of ARG abundance in the CRC group reported that *E. coli*, *A. onderdonkii*, *P. distasonis*, and *A. fingoldii* had a

relatively high abundance of ARGs (median value) (Figure 7E). Notably, these four species were also significantly enriched in the CRC groups, primarily *E. coli*, which was encoded by as many as 37 ARGs and 13 MRGs. Although the ARG type counts in the CRC-associated cluster were lower than that in the unassociated cluster, the former cluster carried a higher abundance of ARGs. Therefore, *E. coli* in the CRC-associated cluster may act as an antibiotic resistance reservoir in the gut microbiota because of its high abundance of drug resistance genes. Taken together, we reported here that more ARGs encoded CRC-associated species than unassociated species, and *E. coli* was a critical antibiotic reservoir in the gut.

## Predictive Effectiveness of Species and Antibiotic Resistance Genes

To demonstrate the plausible clinical prediction of ARGs and species, we next built a series of random-forest prediction models using all the species, rSGB-sourced ARGs, and these features enriched in two groups. First, we tested the classification effect of the abundance of species and ARGs. The results showed that the classification performance of species features (area under the

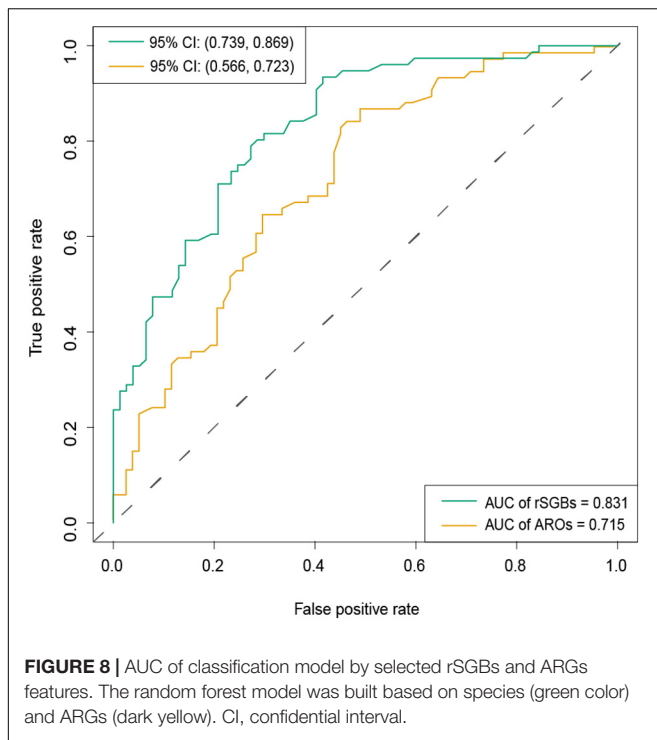


**FIGURE 7 |** Abundance difference of ARGs in rSGBs that associated with CRC or not. **(A)** The rSGB count and percentage in the CRC-associated and unassociated clusters; the blue or yellow colors represented the proportion of rSGBs with or without ARGs. **(B)** ARG counts in the CRC-associated and unassociated clusters. **(C)** PCoA plot depicting Bray–Curtis distances of ARG abundances in the CRC-associated (green color) and unassociated (dark yellow color) clusters. **(D)** Total ARG abundance in the CRC-associated (green color) and unassociated (dark yellow color) clusters (\*\* $p < 0.001$ ). **(E)** Relative abundance of ARGs in every CRC-associated rSGBs in the CRC groups.

curve [AUC] = 0.802) outperformed that of the ARG features (AUC = 0.663) (**Supplementary Figures 9A,B**).

To improve the precision, we rebuilt two models using 118 selected species (32 carried ARGs) features and 19 selected ARG features, using the random forest method (**Supplementary Table 10** and **Supplementary Figures 9C,D**). From **Figure 8**, it could be observed that the species model was more effective (AUC = 0.831) than the ARGs model (AUC = 0.715), when

classifying the CRC and CTR groups. Our species-based classification model performed better than the previous report, where the AUC  $\geq 0.8$  (Wirbel et al., 2019). The model accuracy based on ARG features in our study was approximate to that of a previously reported species model, although the accuracy of our model was unsatisfying (Thomas et al., 2019). In summary, species and ARGs in microbiota could predict CRC patients with modest precision.



## DISCUSSION

Antibiotic resistance is one of the most critical public health threats to human beings in recent years (Hernando-Amado et al., 2019). Although evidence has shown that antibiotic use may increase the risk of CRC, whether antibiotic resistance is also related is far from clear (Dik et al., 2016; Zhang et al., 2019; Armstrong et al., 2020; Wan et al., 2020). Here, we discovered 696 rSGBs, 24,692 plasmids, and 187 ARGs in the gut microbiota, where 25 species were enriched in the CRC group, and 13 species carried ARGs, such as *E. coli*, *A. onderdonkii*, *B. fragilis*, *A. muciniphila*, and *M. torques*. The abundance of ARGs was enriched in the CRC group, and *E. coli* was the essential ARG carrier.

To analyze the ARGs on the species level, we reconstructed the genomes *via* single-sample assembly from the metagenomes as reported in the literature (Pasolli et al., 2019; Zhu et al., 2019). However, plasmids also harbor many ARGs. It is reported that while the bacterial genome is assembled, its plasmids often remain unidentified because it is not clear which contigs in the genome assembly have arisen from plasmids (Antipov et al., 2016, 2019). We used metaplasmidSPAdes, which could identify novel plasmids and assemble plasmids from metagenomic data sets in order to annotate as many ARGs as possible.

With the species-level genomes reconstructed, representative genomes had a high genome quality of more than 90% completeness and less than 5% contamination. As a result, we found 25 species enriched in the CRC group, such as *E. coli*, *P. distasonis*, *A. muciniphila*, *B. thetaiotaomicron*, etc. These species had been reported to induce CRC by producing inflammatory polysaccharides, cell cycle inhibiting factors, and

cytotoxic distending toxins (Wassenaar, 2018; Sahankumari et al., 2019). For species enriched in the CTR group, mounting evidence showed the potential benefits of *F. prausnitzii* for improving intestinal healthy *via* producing butyrate (Ferreira-Halder et al., 2017; Kim et al., 2020). *B. catenulatum*, as a significant commensal bacterium of *F. prausnitzii*, could improve its growth, gut colonization, and butyrate production by producing short-chain fatty acids (Kim et al., 2020). Moreover, *Lachnospira eligens* (formerly *Eubacterium eligens*) could effectively suppress intestinal inflammation and prevent colitis and CRC (Feng et al., 2016). Although more than 90% of species we found had been reported or cultured before, we had also discovered unknown species enriched in the CRC group, such as *UBA5446 sp900544295* and *Anaerotrignum sp000436415*, with both of the species carrying ARGs. It indicated that reconstructing genomes was useful to find out more species carrying ARGs and increased the probability of illustrating the ARG distribution in gut microbiota.

In these species of rSGBs, we detected 164 ARGs, which was slightly higher than another large cohort study on antibiotics (149 ARG types) (Hu et al., 2013). In our results, 30.49% of ARGs were enriched in the CRC group significantly. These CRC-associated species had a high antibiotic resistance abundance. The *adeF* gene was the largest abundant ARG in the CRC group. In another cohort of a healthy population, the highest abundance of ARG was *TcR* (Hu et al., 2013). The ARG abundance difference in different populations may be due to the disease-specific variations in antibiotic resistance under different antibiotic exposures (Shamsaddini et al., 2021).

Antibiotic resistance could be acquired *via* gut microbiota through the use of broad-spectrum antibiotics, including cephalosporins, fluoroquinolones, penams, and rifaximin, to name a few. We noticed that the nucleoside antibiotic type had the highest abundance of antibiotics, followed by the cephalosporins, macrolides, and phenicol antibiotic types. Although a previous meta-analysis reported no significant associations between CRC and some antibiotics, for example, quinolones, tetracyclines, and macrolide antibiotics (Wan et al., 2020), a significantly higher abundance of these antibiotic resistance drugs was detected in the CRC group. In this study, penams were significantly higher in the CRC group, which may be caused by higher penicillin usage in CRC patients (Wan et al., 2020). Another interesting finding was that we found no evident difference in resistance to rifaximin between the CRC and CTR groups. Rifaximin, a rifamycin antibiotic, was popularly used in travelers' diarrhea and irritable bowel syndrome. It had been reported to neither affect ARGs nor increase the ARG burden because the use of rifaximin would rarely bring about the development of drug resistance compared with other antibiotics (Shamsaddini et al., 2021). The high ARG burden in CRC patients suggested that it is recommended to consider possible antibiotic resistance when selecting appropriate antibiotic treatments, such as short-term alternating antibiotics and microbiome-based interventions (Dubinsky et al., 2020).

We found that ARGs were resistant to antibiotics in Reserve Class. Reserve Class antibiotics were treated as the "last resort"

options and usually used for highly selected patients (life-threatening infections due to multidrug-resistant bacteria). They were closely monitored and prioritized as targets of stewardship programs to ensure their continued effectiveness (World Health Organization, 2019). However, drug-resistance species in the CRC group were resistant to the “last resort” antibiotics, such as penems, glycolcyclines, and streptogramins. Increasing resistance of intestinal bacteria to “last resort” antibiotics reduces the number of antibiotics that can be used to control intestinal infections. Therefore, even the use of reserve group antibiotics still calls for caution.

Our study was limited to illustrating the association between CRC and antibiotic use for lacking information on antibiotic administration in these studies. A more directed study was needed to establish their association in the future. As our metagenomic data were collected from eight cohorts, the abundance of rSGBs and ARGs was affected by the host properties. Two cohorts were selected to analyze the abundance variation of rSGBs and ARGs, and consistent results were obtained. In our study, *E. coli* carried many MRGs and was significantly enriched in the CRC group. Although phylotype D *E. coli* (*E. coli* D) was one of these species with cyclomodulin-encoding genes and could produce cytotoxic necrotizing factors, phylotype B2 *E. coli* was the main strain associated with CRC (Wassenaar, 2018). Therefore, whether the phylotype D *E. coli* is associated with CRC warrants further investigation.

## CONCLUSION

We analyzed the species and ARGs distribution in the gut microbiota of CRC and CTR groups and found that the CRC group's gut microbiota had higher ARGs and MRG abundance than that of the CTR group. And bacteria with ARGs were enriched in the CRC group, such as *E. coli*, *P. distasonis*, *B. thetaiotaomicron*, and *B. fragilis*. Meanwhile, CRC-associated species carried abundant ARGs. *E. coli* was the primary antibiotic resistance reservoir of species in the CRC patients. Using species and ARGs could classify CRC patients from healthy controls. It showed that the gut microbiota in CRC patients could confer resistance to fluoroquinolones, cephalosporins, penams, and tetracyclines. Our investigation proposes antibiotic resistance guidance to CRC patients, and this may help develop antibiotic use strategies to reduce the detrimental effects of antibiotic resistance.

## MATERIALS AND METHODS

### Datasets and Samples Details

We downloaded a total of 769 metagenomic paired-end data, including 382 CRC patients (CRC group, aged  $64 \pm 11$  years) and 387 healthy controls (CTR group, aged  $61 \pm 10$  years) (Table 1). Data were selected from eight published studies with the NCBI SRA database accession codes PRJEB10878 (Yu et al., 2017), PRJNA389927 (Hannigan et al., 2018), PRJEB12449 (Vogtmann et al., 2016), PRJEB27928 (Wirbel et al., 2019),

PRJEB6070 (Zeller et al., 2014), PRJNA447983 (Thomas et al., 2019), PRJEB7774 (Feng et al., 2015), and PRJDB4176 (Yachida et al., 2019). Participants who had a history of cancers, used antibiotics in the past period, or with gastrointestinal disease, including inflammatory bowel disease and intestinal infection, were excluded from the CTR groups (Zeller et al., 2014; Feng et al., 2015; Vogtmann et al., 2016; Yu et al., 2017; Hannigan et al., 2018; Thomas et al., 2019; Wirbel et al., 2019; Yachida et al., 2019). Basic information of participants, including gender, age, body mass index (BMI), vegetarian or not, smoking or not, health status (health or CRC), and the American Joint Committee on Cancer Staging (AJCC Staging) information for CRC participation, was also collected (Supplementary Table 1).

### Metagenomes *de novo* Assembly, Binning, and Quality Evaluation

All the 769 paired-end fastq data went through quality control by fastp (Chen et al., 2018); the host sequence (human reference genome version: hg38) in the data was removed using soap2 (Li et al., 2009). Next, data were applied *de novo* assembled using metaSPAdes genome assembler (Nurk et al., 2017). Metagenomic binning was performed by MetaBAT2, which generated 36,461 bins in total. Completeness and contamination rates of bins were calculated by checkm qa workflow (Parks et al., 2015). We filtered bins into high-quality bins (completeness > 90%, contamination < 5%), medium-quality bins (completeness > 50%, contamination < 5%), and low-quality bins (the residual bins). To obtain more high-quality MAGs, we rebinned contigs using the same parameters mentioned previously for these contigs tagged “bin.unbinned” in the MetaBAT2 results and low-quality bins.

### Species-Level Genome Bin Cluster and Representation Selection

The completeness and contamination rates and the quality filtering of new bins were assessed again to remove low-completeness and high-contamination bins. Finally, we obtained 5,880 high-quality MAGs and 5,390 medium-quality MAGs. The 5,880 high-quality MAGs were clustered into species-level genome bins (SGBs) by a two-step clustering strategy based on genetic distance calculation by Metapi (Zhu et al., 2019). Then, representative genomes were selected for each cluster by SGB properties, including completeness, contamination, genome size, and strain heterogeneity index. The sequence with the maximal rank value was selected as representative genomes (rSGBs). The maximal rank value was computed according to a formula:  $Rv = Cp - Ct + \log(Gs) - Th$ , where  $Rv$  means rank value,  $Cp$  and  $Ct$  represent completeness value and contamination value, and  $Gs$  and  $Th$  represent genome size and train heterogeneity.

### Taxonomy and Relative Abundance of Species

Taxonomy annotation for all the rSGBs was performed by GTDB-Tk (Chaumeil et al., 2019) based on the genome taxonomy database (GTDB, Release 95) (Parks et al., 2020). The high-quality reads were aligned to the rSGBs by bwa (default



parameters) (Li, 2013). Sequence-based contigs abundance profiling was performed by `jgi_summarize_bam_contig_depths` (default parameters) (Kang et al., 2019). Reads were mapped to the rSGBs, and the number of reads counted formed a mapping depth. Considering the different sequencing depths of different samples, we used the mapping depth matrix of normalization to estimate the abundances of contigs. For the rSGB profile, we used the species assignment of each contig from the rSGBs and took the median of the relative abundance of contigs from the same rSGBs to generate the abundance of certain rSGBs (**Supplementary Table 11**). Then the  $\alpha$  diversity (Shannon–Wiener index) of relative abundance of species was computed by the `vegan` (Oksanen et al., 2018). Next,  $\beta$  diversity was computed by PCoA and NMDS based on `ape` packages (Paradis and Schliep, 2018).

## Plasmid Assembly and Acquisition of Non-redundant Gene Catalog in the Plasmids

To avoid the possible ARG omission in the progress of MAGs assembly, we assembled the whole plasmid sequence in the gut microbiota from the host genome and removed high-quality reads using the `metaplasmidSPAdes` assembler (Antipov et al., 2019). Then, we called the genes in the contigs by `MetaGeneMark` (Zhu et al., 2010) and processed the gene cluster to the genes by `cd-hit` (Li and Godzik, 2006) to get the non-redundant gene catalog of the plasmids. After that, we computed the relative abundance of non-redundant genes in all samples. The genes were annotated to the NCBI-NT database (20191213) to get the species source of plasmids.

## Antibiotic Resistance Gene Identity, Resistance Mechanism, and Drug Type Analysis

To annotate ARGs in rSGBs, we predicted the open reading frame by `Prodigal` (Hyatt et al., 2010) and then identified ARGs using `Resistance Gene Identifier` based on `CARD` (version 3.0.7) for both rSGBs and plasmid genes (Alcock et al., 2020). Then, antibiotic resistance ontology (ARG types) was matched to the species by contigs id. Based on the best hit antibiotic resistance ontology results, relative abundances of ARGs, drug types, and resistance mechanism types were obtained. In our research, the ARGs were used to represent ARGs types. We also matched resistance antibiotics of rSGBs to `AWaRe` classification according to the WHO `AWaRe` classification of antibiotics (2019 version) (**Supplementary Table 9**). We computed the  $\alpha$  diversity (Shannon–Wiener index) of ARG relative abundance in the rSGBs by the `vegan` (Oksanen et al., 2018). Then, PCoA and NMDS were performed to compute  $\beta$  diversity by `ape` packages (Paradis and Schliep, 2018). Meanwhile, `PERMANOVA` was used to test the statistical significance of  $\beta$  diversity by the `vegan` package (Oksanen et al., 2018).

## Machine Learning Train Model

To assess the classification effect of species and rSGB-sourced ARGs, we built and trained a series of machine

learning models using selected elements from relative abundance profiles of species and ARGs. Data splitting, preprocessing, feature selection, model training, model tuning, and variable importance estimation were finished using the `caret` package (Kuhn, 2020). Before model training, near-zero variance and high correlation (absolute value of correlations coefficient  $> 0.75$ ), variables were removed. And then, data were centered and scaled. Next, the 10-fold cross-validation approach was used to select features, and random forest methods were applied to train models. Finally, we assessed the effect of models by the area under the receiver operating characteristic curve (AUC) value using the `ROCR` package (Sing et al., 2005).

## Statistical Analysis

During the analysis, we carried out Wilcoxon test and `LefSe` analysis for the relative abundance of all species (Segata et al., 2011), with cutoff  $p < 0.05$  and absolute values of the LDA score  $> 2.0$  (**Supplementary Table 6**). Then, we analyzed the ARG profile by calculating Wilcoxon test with a cutoff of adjusted  $p < 0.05$ . Finally, 50 ARGs were filtered from rSGBs. To access the effect of ARGs in the CRC-associated species, we marked the selected 60 species above as “CRC-associated” cluster and other species with ARGs as “unassociated” cluster. And then, we compared the sum of ARG abundance in these two clusters, and then a PCoA analysis was performed on the classes. During the analysis and figure visualization, `ggplot2` (Wickham, 2016), `ggtree` (Yu, 2020), `ggpubr` (Kassambara, 2020), and `networkD3` (Allaire et al., 2017) packages were used in our study.

## DATA AVAILABILITY STATEMENT

All the 5880 SGBs and 696 rSGBs data in this study have been deposited into CNGB Sequence Archive (CNSA) (Guo et al., 2020) of China National GeneBank DataBase (CNGBdb) (Chen et al., 2020) with accession number CNP0001862.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Clearance the Institutional Review Board of BGI. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

CL and ZL downloaded and analyzed the data. CL wrote the manuscript. JD modified the manuscript. HZ and CN gave beneficial advice. ZL and MF conceived the study and commented on the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported by the National Key Research and Development Program of China (No. 2020YFC2002902) and the National Natural Science Foundation of China (No. 31800765).

## REFERENCES

- Ai, D., Pan, H., Li, X., Gao, Y., Liu, G., and Xia, L. C. (2019). Identifying gut microbiota associated with colorectal cancer using a zero-inflated lognormal model. *Front. Microbiol.* 10:826. doi: 10.3389/fmicb.2019.00826
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935
- Allaire, J. J., Gandrud, C., Russell, K., and Yetman, C. (2017). *networkD3: D3 JavaScript Network Graphs from R*. R package version 0.4. Available online at: <https://CRAN.R-project.org/package=networkD3>
- Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P. A. (2016). plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32, 3380–3387. doi: 10.1093/bioinformatics/btw493
- Antipov, D., Raiko, M., Lapidus, A., and Pevzner, P. A. (2019). Plasmid detection and assembly in genomic and metagenomic data sets. *Genome Res.* 29, 961–968. doi: 10.1101/gr.241299.118
- Armstrong, D., Dregan, A., Ashworth, M., White, P., McGee, C., and de Lusignan, S. (2020). The association between colorectal cancer and prior antibiotic prescriptions: case control study. *Br. J. Cancer* 122, 912–917. doi: 10.1038/s41416-019-0701-705
- Becattini, S., Taur, Y., and Pamer, E. G. (2016). Antibiotic-Induced changes in the intestinal microbiota and disease. *Trends Mol. Med.* 22, 458–478. doi: 10.1016/j.molmed.2016.04.003
- Bernstein, C. N., Eliakim, A., Fedail, S., Fried, M., Gearry, R., Goh, K. L., et al. (2016). World gastroenterology organisation global guidelines inflammatory bowel disease: update August 2015. *J. Clin. Gastroenterol.* 50, 803–818. doi: 10.1097/MCG.0000000000000660
- Cao, Y., Wu, K., Mehta, R., Drew, D. A., Song, M., Lochhead, P., et al. (2018). Long-term use of antibiotics and risk of colorectal adenoma. *Gut* 67, 672–678. doi: 10.1136/gutjnl-2016-313413
- Casals-Pascual, C., Vergara, A., and Vila, J. (2018). Intestinal microbiota and antibiotic resistance: perspectives and solutions. *Hum. Microbiome J.* 9, 11–15. doi: 10.1016/j.humic.2018.05.002
- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848
- Chen, F. Z., You, L. J., Yang, F., Wang, L. N., Guo, X. Q., Gao, F., et al. (2020). CNGBdb: China national genebank database. *Yi Chuan* 42, 799–809. doi: 10.16288/j.yczz.20-080
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Cheng, Y., Ling, Z., and Li, L. (2020). The intestinal microbiota and Colorectal Cancer. *Front. Immunol.* 11:615056. doi: 10.3389/fimmu.2020.615056
- Coyne, S., Rosenfeld, N., Lambert, T., Courvalin, P., and Perichon, B. (2010). Overexpression of resistance-nodulation-cell division pump AdeFGH confers multidrug resistance in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54, 4389–4393. doi: 10.1128/AAC.00155-110
- Crockett, S. D., and Nagtegaal, I. D. (2019). Terminology, molecular features, epidemiology, and management of serrated colorectal neoplasia. *Gastroenterology* 157, 949–966.e4. doi: 10.1053/j.gastro.2019.06.041
- Dai, Z., Zhang, J., Wu, Q., Chen, J., Liu, J., Wang, L., et al. (2019). The role of microbiota in the development of colorectal cancer. *Int. J. Cancer* 145, 2032–2041. doi: 10.1002/ijc.32017
- Dik, V. K., van Oijen, M. G., Smeets, H. M., and Siersema, P. D. (2016). Frequent use of antibiotics is associated with colorectal cancer risk: results of a nested case-control study. *Dig. Dis. Sci.* 61, 255–264. doi: 10.1007/s10620-015-3828-3820
- Dubinsky, V., Reshef, L., Bar, N., Keizer, D., Golan, N., Rabinowitz, K., et al. (2020). Predominantly antibiotic-resistant intestinal microbiome persists in patients with pouchitis who respond to antibiotic therapy. *Gastroenterology* 158, doi: 10.1053/j.gastro.2019.10.001 610–624 e613.
- Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., et al. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun.* 6:6528. doi: 10.1038/ncomms7528
- Feng, Q., Zhang, D., Liu, C., Xiao, L., Tang, L., and Wang, J. (2016). *Use of Eubacterium in the Prevention and Treatment for Colorectal Cancer Related Diseases*. Geneva: World Intellectual Property Organization. PCT/CN2014/083689.
- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., et al. (2021). Cancer statistics for the year 2020: an overview. *Int. J. Cancer* doi: 10.1002/ijc.33588 [Online ahead of print].
- Ferreira-Halder, C. V., Faria, A. V. S., and Andrade, S. S. (2017). Action and function of *Faecalibacterium prausnitzii* in health and disease. *Best Pract. Res. Clin. Gastroenterol.* 31, 643–648. doi: 10.1016/j.bpg.2017.09.011
- Gao, Z., Guo, B., Gao, R., Zhu, Q., and Qin, H. (2015). Microbiota dysbiosis is associated with colorectal cancer. *Front. Microbiol.* 6:20. doi: 10.3389/fmicb.2015.00020
- Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., et al. (2020). CNSA: a data repository for archiving omics data. *Database (Oxford)* 2020:baaa055. doi: 10.1093/database/baaa055
- Hannigan, G. D., Duhaime, M. B., Ruffin, M. T. T., Koumpouras, C. C., and Schloss, P. D. (2018). Diagnostic potential and interactive dynamics of the colorectal cancer virome. *mBio* 9:e02248-18 doi: 10.1128/mBio.02248-2218.
- Hernando-Amado, S., Coque, T. M., Baquero, F., and Martinez, J. L. (2019). Defining and combating antibiotic resistance from one health and global health perspectives. *Nat. Microbiol.* 4, 1432–1442. doi: 10.1038/s41564-019-0503-509
- Hu, Y., Yang, X., Qin, J., Lu, N., Cheng, G., Wu, N., et al. (2013). Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat. Commun.* 4:2151. doi: 10.1038/ncomms3151
- Hyatt, D., Chen, G.-L., LoCasio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. doi: 10.7717/peerj.7359
- Kassambara, A. (2020). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R Package Version 0.4.0. Available online at: <https://CRAN.R-project.org/package=ggpubr>
- Kim, H., Jeong, Y., Kang, S., You, H. J., and Ji, G. E. (2020). Co-Culture with bifidobacterium catenulatum improves the growth, gut colonization, and butyrate production of *Faecalibacterium prausnitzii*: in vitro and in vivo studies. *Microorganisms* 8:788. doi: 10.3390/microorganisms8050788
- Kuhn, M. (2020). *caret: Classification and Regression Training*. R package Version 6.0-86. Available online at: <https://CRAN.R-project.org/package=caret>
- Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [Preprint]*. arXiv:1303.3997. Available online at: <https://arxiv.org/abs/1303.3997v2> (accessed May 26, 2013).
- Ma, C., Chen, K., Wang, Y., Cen, C., Zhai, Q., and Zhang, J. (2021). Establishing a novel colorectal cancer predictive model based on unique gut microbial single

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.765291/full#supplementary-material>

- nucleotide variant markers. *Gut Microbes* 13, 1–6. doi: 10.1080/19490976.2020.1869505
- Nitzan, O., Elias, M., Peretz, A., and Saliba, W. (2016). Role of antibiotics for treatment of inflammatory bowel disease. *World J. Gastroenterol.* 22, 1078–1087. doi: 10.3748/wjg.v22.i3.1078
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). *vegan: Community Ecology Package. Ordination Methods, Diversity Analysis and Other Functions for Community and Vegetation Ecologists. Version 2.5-1*. Available online at: <https://CRAN.R-project.org/package=vegan>
- Paradis, E., and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Parker, B. J., Wearsch, P. A., Veloo, A. C. M., and Rodriguez-Palacios, A. (2020). The genus *alisticipes*: gut bacteria with emerging implications to inflammation, cancer, and mental health. *Front. Immunol.* 11:906. doi: 10.3389/fimmu.2020.00906
- Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* 38, 1079–1086. doi: 10.1038/s41587-020-0501-508
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., et al. (2019). Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 176, 649–662.e20. doi: 10.1016/j.cell.2019.01.001
- Sahankumari, A., Gamage, B. D., and Malavige, G. N. (2019). Association of the gut microbiota with colorectal cancer in a South Asian cohort of patients. *bioRxiv* [Preprint]. doi: 10.1101/694125
- Schwartz, D. J., Langdon, A. E., and Dantas, G. (2020). Understanding the impact of antibiotic perturbation on the human microbiome. *Genome Med.* 12:82. doi: 10.1186/s13073-020-00782-x
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60
- Shamsaddini, A., Gillevet, P. M., Acharya, C., Fagan, A., Gavis, E., Sikaroodi, M., et al. (2021). Impact of antibiotic resistance genes in gut microbiome of patients with cirrhosis. *Gastroenterology* 161, 508–521.e7. doi: 10.1053/j.gastro.2021.04.013
- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCr: visualizing classifier performance in R. *Bioinformatics* 21, 3940–3941. doi: 10.1093/bioinformatics/bti623
- Theochari, N. A., Stefanopoulos, A., Mylonas, K. S., and Economopoulos, K. P. (2018). Antibiotics exposure and risk of inflammatory bowel disease: a systematic review. *Scand. J. Gastroenterol.* 53, 1–7. doi: 10.1080/00365521.2017.1386711
- Thomas, A. M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., et al. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* 25, 667–678. doi: 10.1038/s41591-019-0405-407
- Van Boeckel, T. P., Gandra, S., Ashok, A., Caudron, Q., Grenfell, B. T., Levin, S. A., et al. (2014). Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *Lancet Infect. Dis.* 14, 742–750. doi: 10.1016/s1473-3099(14)70780-70787
- Vogtmann, E., Hua, X., Zeller, G., Sunagawa, S., Voigt, A. Y., Herczeg, R., et al. (2016). Colorectal Cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS One* 11:e0155362. doi: 10.1371/journal.pone.0155362
- Wan, Q. Y., Zhao, R., Wang, Y., Wu, Y., and Wu, X. T. (2020). Antibiotic use and risk of colorectal cancer: a meta-analysis of 412 450 participants. *Gut* 69, 2059–2060. doi: 10.1136/gutjnl-2020-320826
- Wang, J. L., Chang, C. H., Lin, J. W., Wu, L. C., Chuang, L. M., and Lai, M. S. (2014). Infection, antibiotic therapy and risk of colorectal cancer: a nationwide nested case-control study in patients with Type 2 diabetes mellitus. *Int. J. Cancer* 135, 956–967. doi: 10.1002/ijc.28738
- Wassenaar, T. M. (2018). E. coli and colorectal cancer: a complex relationship that deserves a critical mindset. *Crit. Rev. Microbiol.* 44, 619–632. doi: 10.1080/1040841X.2018.1481013
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
- Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* 25, 679–689. doi: 10.1038/s41591-019-0406-406
- Wong, S. H., and Yu, J. (2019). Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* 16, 690–704. doi: 10.1038/s41575-019-0209-208
- World Health Organization (2019). *The 2019 WHO AWaRe Classification of Antibiotics for Evaluation and Monitoring of Use*. Geneva: World Health Organization.
- Xing, C., Wang, M., Ajibade, A. A., Tan, P., Fu, C., Chen, L., et al. (2021). Microbiota regulate innate immune signaling and protective immunity against cancer. *Cell Host Microbe* 29, 959–974.e7. doi: 10.1016/j.chom.2021.03.016
- Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., et al. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* 25, 968–976. doi: 10.1038/s41591-019-0458-457
- Yang, J., McDowell, A., Kim, E. K., Seo, H., Lee, W. H., Moon, C. M., et al. (2019). Development of a colorectal cancer diagnostic model and dietary risk assessment through gut microbiome analysis. *Exp. Mol. Med.* 51, 1–15. doi: 10.1038/s12276-019-0313-314
- Yeoh, Y. K., Chen, Z., Wong, M. C. S., Hui, M., Yu, J., Ng, S. C., et al. (2020). Southern Chinese populations harbour non-nucleatum *Fusobacteria* possessing homologues of the colorectal cancer-associated FadA virulence factor. *Gut* 69, 1998–2007. doi: 10.1136/gutjnl-2019-319635
- Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinformatics* 69, e96. doi: 10.1002/cpbi.96
- Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. Y., Qin, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 66, 70–78. doi: 10.1136/gutjnl-2015-309800
- Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst Biol.* 10:766. doi: 10.15252/msb.20145645
- Zhang, J., Haines, C., Watson, A. J. M., Hart, A. R., Platt, M. J., Pardoll, D. M., et al. (2019). Oral antibiotic use and risk of colorectal cancer in the United Kingdom, 1989–2012: a matched case-control study. *Gut* 68, 1971–1978. doi: 10.1136/gutjnl-2019-318593
- Zhang, Y., Yu, X., Yu, E., Wang, N., Cai, Q., Shuai, Q., et al. (2018). Changes in gut microbiota and plasma inflammatory factors across the stages of colorectal tumorigenesis: a case-control study. *BMC Microbiol.* 18:92. doi: 10.1186/s12866-018-1232-1236
- Zhu, J., Tian, L., Chen, P., Han, M., Song, L., Tong, X., et al. (2019). Over 50000 metagenomically assembled draft genomes for the human oral microbiome reveal new taxa. *bioRxiv* [Preprint]. doi: 10.1101/820365
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132. doi: 10.1093/nar/gkq275

**Conflict of Interest:** All authors were employed by company BGI-Shenzhen.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Liu, Li, Ding, Zhen, Fang and Nie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.