

ORIGINAL RESEARCH ARTICLE

Impact of dietary emulsifiers on the presence of adherent-invasive *Escherichia coli* in Crohn's disease

Supplementary File

1. Supplementary methods

1.1. The identification of Adherent-invasive *Escherichia coli* (AIEC)

Two terminal ileum samples were obtained during colonoscopy from cases and controls and stored in Minimum Essential Media (MEM, Gibco, USA) with 15% glycerol (Invitrogen, UK). AIEC was isolated according to the description from a previous study.¹ Briefly, *E. coli* adherent to intestinal mucosa, as well as invading the intestinal epithelial cells, were isolated from the preserved liquid and lysate of the biopsy of each subject, respectively. Drigalski agar was used to determine the colony-forming units per milligram (CFU/mg) of *E. coli* that was adherent to or invaded epithelial cells. Forty-five *E. coli* strains were isolated from each sample and co-cultured with Intestine-407 (ATCC® CCL-6™) cells according to the gentamicin protection assay to verify bacteria with invasive ability. Two strains, including LF82 (the type strain of AIEC) and K12C600 (an *E. coli* strain without invasive ability), were used as references to identify the invasive ability of isolated *E. coli* strains. Patients were considered AIEC-positive if isolated *E. coli* strains had an invasion rate equal to or superior to 10% of the reference.

1.2. The estimation of food additives consumption

The food additives questionnaire was described previously and was validated in a study to identify the food additives exposure of CD patients and their family members.^{2,3} Briefly, two dietitians drafted the questionnaire on food additives and constructed inflammatory bowel disease (IBD)-specific food additives databases. Initially, a systematic review was performed to search the literature on IBD studies to confirm the food additives related to IBD development. Ten food additives were considered potential pathogenic factors to IBD, including maltodextrin, carrageenan, carboxymethylcellulose, polysorbate-80, aspartame, saccharin, sucralose, titanium dioxide, sodium aluminosilicate, and sulfite preservatives. The highly processed food included in the food additives questionnaire covered a wide range of items belonging to fast food, soft drinks, snacks, and five categories of packaged foods

(dairy, meat, grains, fruits, and vegetables). The maximum permissible level of each food additive allowed in the Codex Alimentarius CXS 192e International Food Standards (CODEX) was used as the concentration for food additive consumption estimation. For a food additive not recorded in the CODEX, the concentration was determined based on published studies or technical protocols with a range from 20 mg/kg to 10,000 mg/kg. The amount of food consumed was obtained from participants through items inquiring about whether the foods were consumed (Yes/No) and the frequency of food consumption (how many times every day/week/month/year). The food additives questionnaire exhibited a high reliability, with over 75% of questions answered in a consistent manner on each attempt. We calculated the maximum consumption of each food additive on a yearly basis using the formula below:

- Annual exposure (mg/year) = Food additive concentration (mg/g) × Food amount per day (g) × 365 days

1.3. Sample DNA extraction and 16S amplicon sequencing

Microbial DNA was extracted from mucosa biopsy samples of the rural cohort. Mucosal biopsy with lysis buffer and zirconia beads was mixed, and the tissue lyser was applied to homogenize the samples. After centrifuging the biopsy lysate, the supernatant was collected and then added to the Maxwell Tissue DNA Purification kit (Promega, USA). The Maxwell RSC instrument was used for the purification of mucosal DNA. The extracted DNA was sequenced by the Novogene company, targeting the V3-V4 region of the 16S rRNA gene using forward primer 341F: 5'-CCTAYGGGRBGCASCAG-3' and reverse primer 806R: 5'-GGACTACNNGGGTATCTAAT-3'. The Illumina Novaseq 6000 platform (Illumina, USA) was used for sequencing to generate PE 250 bp raw reads. The DNA extraction and sequencing of the urban cohort were described in the previous study.⁴

1.4. The taxonomy annotation and functional prediction of mucosal microbiome data

Sequencing data from Yunnan and Hong Kong were preprocessed using the QIIME2 pipeline to remove

amplification primers and low-quality bases. Raw reads were clustered by VSEARCH, and those reads with 97% similarity were used to generate the operational taxonomic units (OTUs). The taxonomy of OTUs was assigned by the Silva 138 classifier that was trained using the Naïve Bayes algorithm.⁵ As the samples from the urban and rural areas were sequenced at different time points, we collapsed all the OTUs to the genus level to minimize the batch effect in sequencing.⁶ Low-frequency genera with a total abundance

of <0.01 of the abundance of all taxa were filtered from the OTU table. For the functional annotation, the functional pathways were predicted by the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) based on the marker gene sequenced from the mucosal samples.⁷ The functional pathway was annotated according to the MetaCyc database (<https://metacyc.org/>), and each functional pathway was summarized into Superclass levels.

2. Supplementary tables and figures

Table S1. The food additive intake between Crohn's disease patients and healthy controls in the rural cohort

Characteristic	Crohn's disease (mg/year)	Healthy control (mg/year)	p-value [†]
Polysorbate-80 (P80)	563954 (562112)	600511 (690442)	0.928
Carboxymethylcellulose (CMC)	1514634 (1695110)	1703765 (2279790)	0.677
Carrageenan (CRN)	610047 (643839)	687820 (986041)	0.491
Aluminum silicate (AlSiO)	8382 (30250)	2092 (8455)	0.038*
Sulfite (SO ₃) ₂	20543 (31777)	20626 (24562)	0.624
Titanium dioxide (TiO ₂)	127151 (457355)	29664 (118610)	0.035*
Aspartame (ASP)	103636 (179812)	114359 (215758)	1.000
Sucralose (SUC)	49771 (91345)	44951 (58059)	0.739
Saccharine (SAC)	35724 (67244)	25417 (36580)	0.453

Notes: The number in each cell denotes the median (interquartile range, IQR). [†]p-value was calculated according to the Wilcoxon rank sum test.

*indicates statistically significant p-values.

Table S2. The standardized mean difference and p-value of food additives between AIEC-positive and AIEC-negative groups in rural CD patients

Food additive	Overall CD patients		College/University		Middle school		Primary school		No formal schooling	
	SMD	p-value	SMD	p-value	SMD	p-value	SMD	p-value	SMD	p-value
Polysorbate-80 (P80)	-0.285	0.308	-0.727	0.080	-0.978	0.070	1.542	0.310	0.471	0.570
Carboxymethylcellulose (CMC)	-0.584	0.025	-0.463	0.264	-1.225	<0.001	0.385	0.760	-0.750	0.515
Carrageenan (CRN)	0.329	0.294	0.237	0.628	-0.652	0.140	3.810	0.010	1.174	0.172
Aluminum silicate (AlSiO)	-0.261	0.232	-0.316	0.424	-0.366	0.340	-0.492	0.300	-0.290	0.766
Sulfite (SO ₃) ₂	-0.553	0.023	-0.701	0.076	-0.900	0.010	0.064	0.960	-0.900	0.451
Titanium dioxide (TiO ₂)	-0.239	0.273	-0.323	0.410	-0.383	0.320	-0.466	0.320	0.179	0.826
Aspartame (ASP)	0.211	0.565	-0.881	0.035	0.648	0.630	0.598	0.530	0.597	0.447
Sucralose (SUC)	-0.227	0.375	-0.603	0.122	0.074	0.910	0.211	0.850	0.254	0.745
Saccharine (SAC)	-0.164	0.502	-0.621	0.110	-0.823	0.150	1.878	0.280	0.228	0.794

Abbreviations: AIEC: Adherent-invasive *Escherichia coli*; CD: Crohn's disease; SMD: Standardized mean difference.

Table S3. The association between AIEC presence and CD risk identified by multivariate analysis in the rural cohort

Items	OR (CI)	p-value	Category
AIEC presence	7.5 (1.04, 54.23)	0.046*	AIEC
Saccharine (SAC)	3.22 (0.89, 11.65)	0.075	Food additives
Sucralose (SUC)	0.33 (0.06, 1.97)	0.224	Food additives
Aspartame (ASP)	1.46 (0.36, 5.85)	0.595	Food additives
Titanium dioxide (TiO ₂)	1.03 (0.93, 1.13)	0.559	Food additives
Sulfite (SO ₃) ₂	1.47 (0.54, 4.03)	0.454	Food additives
Aluminum silicate (AlSiO)	1.03 (0.93, 1.15)	0.551	Food additives
Carrageenan (CRN)	1.35 (0.58, 3.16)	0.492	Food additives
Carboxymethylcellulose (CMC)	0.82 (0.62, 1.08)	0.162	Food additives
Polysorbate-80 (P80)	0.67 (0.16, 2.91)	0.595	Food additives
Education (College/University) [ref]	1 (1, 1)	NA	Demographics
Education (Middle school)	8.2 (1.35, 49.71)	0.022*	Demographics
Education (Primary school)	4.8 (0.66, 35.12)	0.122	Demographics
Education (No formal schooling)	3.47 (0.43, 28.06)	0.243	Demographics
BMI	0.77 (0.65, 0.9)	0.001*	Demographics
Age	0.95 (0.9, 1.01)	0.093	Demographics

Note: *indicates statistically significant p-values.

Abbreviations: AIEC: Adherent-invasive *Escherichia coli*; BMI: Body mass index; CD: Crohn's disease; CI: Confidence interval; OR: Odds ratio.

Table S4. The association between food additives and AIEC presence identified by multivariate analysis in the rural CD patients

Items	OR (CI)	p-value	Category
Saccharine (SAC)	0.86 (0.31, 2.4)	0.773	Food additives
Sucralose (SUC)	2.24 (0.33, 15.22)	0.410	Food additives
Aspartame (ASP)	0.51 (0.08, 3.16)	0.495	Food additives
Titanium dioxide (TiO ₂)	1.05 (0.91, 1.21)	0.852	Food additives
Sulfite (SO ₃) ₂	1.04 (0.68, 1.6)	0.467	Food additives
Aluminum silicate (AlSiO)	0.94 (0.82, 1.09)	0.412	Food additives
Carrageenan (CRN)	4.49 (1.28, 15.75)	0.019*	Food additives
Carboxymethylcellulose (CMC)	0.94 (0.72, 1.23)	0.646	Food additives
Polysorbate-80 (P80)	0.15 (0.02, 1.03)	0.054	Food additives
Education (College/University) [ref]	1 (1, 1)	NA	Demographics
Education (Middle school)	0.19 (0.02, 1.5)	0.114	Demographics
Education (Primary school)	0.18 (0.01, 3.12)	0.237	Demographics
Education (No formal schooling)	8.73 (0.42, 183.54)	0.163	Demographics
BMI	0.81 (0.57, 1.14)	0.224	Demographics
Age	1.08 (0.98, 1.18)	0.117	Demographics

Note: *indicates statistically significant p-values.

Abbreviations: AIEC: Adherent-invasive *Escherichia coli*; BMI: Body mass index; CD: Crohn's disease; CI: Confidence interval; OR: Odds ratio.

Table S5. The association between urbanization and AIEC presence identified by multivariate analysis combining urban and rural CD patients

Items	OR (CI)	p-value
Age	1.34 (0.39, 4.31)	0.628922
Region (Rural) (ref)	1 (1, 1)	NA
Region (Urban)	2.56 (1.03, 6.52)	0.043953*
Surgery history (No) (ref)	1 (1, 1)	NA
Surgery history (Yes)	1.06 (0.41, 2.73)	0.903215
CD Location (L1) [ref]	1 (1, 1)	NA
CD Location (L2)	3.56 (0.67, 27.90)	0.164243
CD Location (L3)	1.99 (0.46, 13.96)	0.407879
CD Location (Other)	0.005 (0.00045, 32.6)	0.993400

Note: *indicates statistically significant p-values.

Abbreviations: AIEC: Adherent-invasive *Escherichia coli*; CD: Crohn's disease; CI: Confidence interval; OR: Odds ratio.

Table S6. The mean log₁₀ abundance and p-values of differentially abundant functional pathways for AIEC presence in CD patients

Pathways	Median log ₁₀ abundance (AIEC-positive)	Median log ₁₀ abundance (AIEC-negative)	p-value
Gallate degradation I	2.031576	1.987025	0.035313
Acetylene degradation	4.33813	4.539657	0.016986
Mandelate degradation I	1.159324	1.419401	0.025132
Biotin biosynthesis II	2.459143	2.702125	0.047919
Superpathway of sulfur oxidation (Acidianus ambivalens)	3.534779	3.792717	0.025686
L-glutamate and L-glutamine biosynthesis	4.279106	4.512997	0.009919
Syringate degradation	2.328232	2.287731	0.048733
Chondroitin sulfate degradation I (bacterial)	2.842423	3.165462	0.031271
Mandelate degradation to acetyl-CoA	1.69211	1.903616	0.023412

Abbreviations: AIEC: Adherent-invasive *Escherichia coli*; CD: Crohn's disease

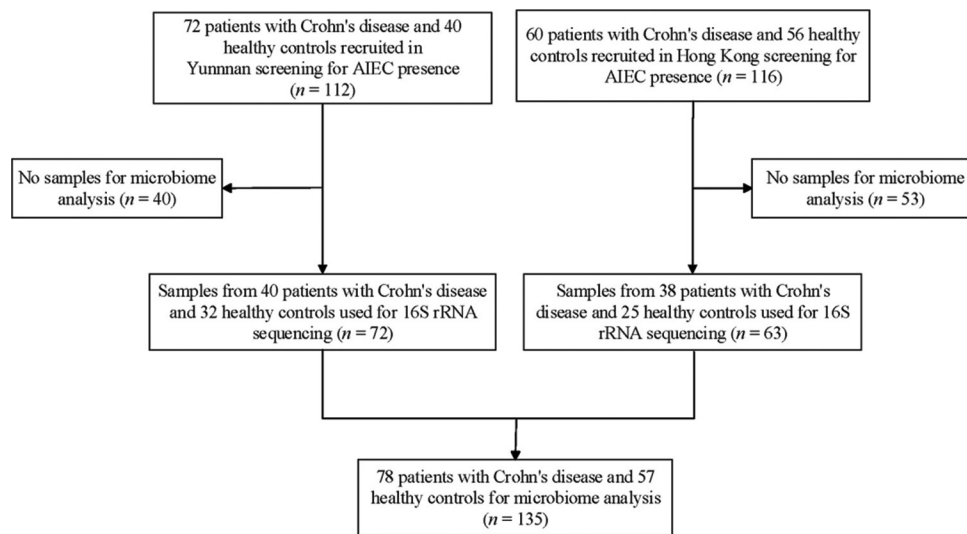


Figure S1. Overview of the study cohorts from Hong Kong (urban) and Yunnan (rural)

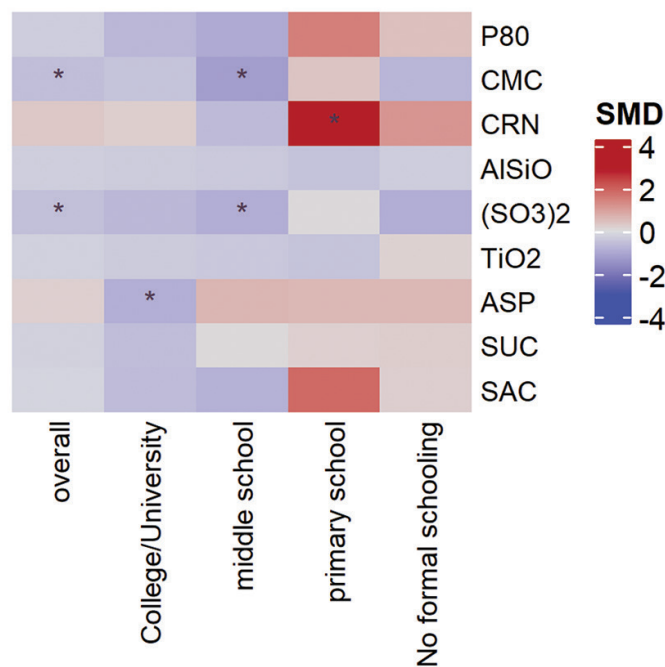


Figure S2. Comparison of food additive intakes in CD patients at different education levels in the rural cohort. The color in each cell represents the standardized mean difference (SMD) between CD patients with and without AIEC presence. The red color indicates increased food additive consumption in CD patients with AIEC. The blue color indicates decreased food additive consumption in CD patients with AIEC. The asterisk (*) indicates a significant difference with $p < 0.05$. The p -value was calculated using the Wilcoxon rank sum test.

Abbreviations: AIEC: Adherent-invasive *Escherichia coli*; AlSiO: Aluminum silicate; ASP: Aspartame; CD: Crohn's disease; CMC: Carboxymethylcellulose; CRN: Carrageenan; P80: Polysorbate-80; SAC: Saccharine; (SO₃)₂: Sulfite; SUC: Sucralose; TiO₂: Titanium dioxide.

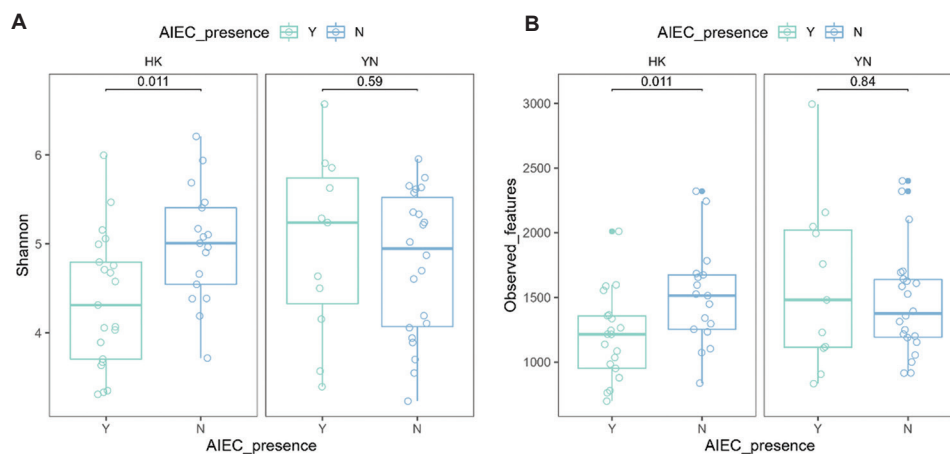


Figure S3. The Bacteroides/Prevotella ratio (B/P ratio) and alpha diversity in the urban and rural cohorts. (A) The Shannon diversity in CD patients with and without AIEC presence in the urban and rural cohorts. (B) The observed features in CD patients with and without AIEC presence in the urban and rural cohorts.

Abbreviations: AIEC: Adherent-invasive *Escherichia coli*; CD: Crohn's disease; HK: Hong Kong; YN: Kunming, Yunnan.

References

- O'Brien CL, Bringer MA, Holt KE, *et al.* Comparative genomics of Crohn's disease-associated adherent-invasive *Escherichia coli*. *Gut*. 2017;66:1382-1389.
doi: 10.1136/gutjnl-2015-311059
- Trakman GL, Lin W, Wilson-O'Brien AL, *et al.* Development and Validation of surveys to estimate food additive intake. *Nutrients*. 2020;12(3):812.
doi: 10.3390/nu12030812
- Trakman GL, Lin WYY, Hamilton AL, *et al.* Processed food as a risk factor for the development and perpetuation of crohn's disease-The ENIGMA study. *Nutrients*. 2022;14(17):3627.
doi: 10.3390/nu14173627
- Zhilu X, Xiangqian D, Keli Y, *et al.* Association of Adherent-invasive *Escherichia coli* with severe Gut mucosal dysbiosis in Hong Kong Chinese population with Crohn's disease. *Gut Microbes*. 2021;13(1):1994833.
doi: 10.1080/19490976.2021.1994833
- Robeson MS 2nd, O'Rourke DR, Kaehler BD, *et al.* RESCRIPt: Reproducible sequence taxonomy reference database management. *PLoS Comput Biol*. 2021;17(11):e1009581.
doi: 10.1371/journal.pcbi.1009581
- Gibbons SM, Duvallet C, Alm EJ. Correcting for batch effects in case-control microbiome studies. *PLoS Comput Biol*. 2018;14(4):e1006102.
doi: 10.1371/journal.pcbi.1006102
- Douglas GM, Maffei VJ, Zaneveld JR, *et al.* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38(6):685-688.
doi: 10.1038/s41587-020-0548-6