Alterations in Gastric Microbial Communities are Associated with Risk of Gastric Cancer in a Korean Population: A Case-Control Study

Madhawa Neranjan Gunathilake
National Cancer Center Graduate School of Cancer Science and Policy
Jeonghee Lee
National Cancer Center
Il Ju Choi
National Cancer Center Graduate School of Cancer Science and Policy
Young-Il Kim
National Cancer Center Graduate School of Cancer Science and Policy
Jihyun Kim
Yonsei University
Jae Kyung Yoon
Yonsei University
Woo Jun Sul
Chung-Ang University
Jeongseon Kim (jskim@ncc.re.kr)
National Cancer Center Research Institute https://orcid.org/0000-0002-0889-2686

Research

Keywords: Gastric cancer, gastric microbiome, microbial dysbiosis index, pathways, case-control

Posted Date: July 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-45990/v1

Version of Record: A version of this preprint was published at Cancers on September 14th, 2020. See the published version at https://doi.org/10.3390/cancers12092619.

Abstract

Background: Although recent studies have reported the potential role of microbiome in GC, little is known about the microbial dysbiosis and their functions in GC. This study aimed to observe the associations between the alterations in gastric microbial communities and GC risk.

Results: The study participants included 268 GC patients and 288 controls. DNA was extracted from gastric biopsies, and 16S rRNA gene analysis was performed to assign into bacterial taxa. Linear discriminant analysis (LDA) effect size (LEfSe) was used to identify differentially abundant taxa and pathways. Compositionality corrected by renormalization and permutation (CCREPE) was applied to derive microbial dysbiosis index (MDI). Microbiota metabolic pathways were analyzed using Phylogenetic Investigation of Communities by Reconstruction of Observed States (PICRUSt) based on the Kyoto Encyclopedia of Genes and Genomes (KEGG). Based on LEfSe, *Streptococcus_NCVM* and *Prevotella melaninogenica* species were highly enriched in GC cases and controls, respectively. Those who are in the third tertile of *Prevotella melaninogenica* showed a significantly decreased risk of GC in total (odds ratio (OR): 0.91, 95% confidence interval (CI): 0.38-0.96, p-trend=0.071). In the cladogram, class *Bacilli* was phylogenetically enriched in GC cases while phylum *Actinobacteria*, class *Actinobacteria* were phylogenetically related in the controls. MDI was significantly higher for the GC cases compared to the healthy controls in the female population (p=0.002). Females, those who are in the third tertile of the MDI showed a significantly increased risk of GC (OR: 2.66, 95% CI: 1.19-5.99, p-trend=0.017; model II). There was a significant inverse correlation between MDI and Shannon index for the total population (R=-0.83, p<0.001). Secondary bile acid synthesis and biosynthesis of ansamycins pathways were highly abundant in cases and controls, respectively.

Conclusion: There is a significantly higher MDI in GC cases than controls and further, MDI is significantly positively associated with GC risk in the female population. Biosynthesis of ansamycins is a critical function which is highly enriched in healthy controls. Further microbiome studies are needed to confirm the findings of the current study.

Introduction

According to the GLOBOCAN estimates in 2018, gastric cancer (GC) has been identified as the fifth most common cancer type and it is one of the main causes of cancer related death in the world [1]. The GC incidence in eastern Asia, including Korea, is the highest worldwide, which is over 4 times higher than the rates in Western Europe [2]. In 2017, according to the Korea Central Cancer Registry (KCCR), the age-adjusted incidence rate of GC was 32.0 per 100,000 in the overall population, 46.4 per 100,000 in men, and 19.6 per 100,000 in women [3]. Various epidemiological studies have demonstrated that smoking [4–7], alcohol consumption [4, 5, 8], obesity and physical inactivity [4, 5, 9], family history of GC [10–14] and numerous dietary factors such as foods preserved with salts, pickled vegetables, low fruits and vegetables consumption, low dietary vitamin C and carotenoid intake, high salt consumption, high processed meat consumption, high salt intake [4, 5, 15, 16] are associated with the increased risk of GC.

In addition, some of recent epidemiological studies have suggested the involvement of gastric microbiome in GC occurrence by the induction of chronic inflammation or down regulation of host immunity [17–23].

In the gastric environment, innumerable bacteria form a complex and stable bacterial community that eventually plays important role in host mucosal immune response, energy and nutrient metabolism, pathogen elimination and several diseases development including GC [24–29]. *Helicobacter pylori* (HP) infection plays a crucial role in the initial steps of GC carcinogenesis by enhancing the inflammation [30]. It is widely implicated that GC carcinogenesis can be occurred because of imbalance of the gastric microbial community [31]. Such an imbalance of the gastric microbial community is known as dysbiosis that eventually leads to the pathological conditions such as GC [22]. When there is a dysbiosis condition with the imbalance of the microbial community in the stomach can lead to the formation of inflammatory biomarkers that can stimulate the carcinogenesis process [32, 33]. Moreover, bacteria prominent during gut dysbiosis can secrete toxins able to interfere with host cell growth, finally predisposing the host organism to cancer development [34].

The microbiome is involving with the metabolism and it communicates with the host to proceed with the physiological functions. They release several metabolites that are essential for our body functions [35]. It can be noted that there is a two-way communication between gut microbes and the human immune system [35]. Emphasizing the metabolic functions associated with gut microbiota is important because several possible microbial mechanisms are involved in carcinogenesis through microbiota. Dysbiosis helps to increase the pH and innate immune response that generates changes in the acquisition of nutrients, metabolism, and survival of members of the normal microbiota, and however expansion and development of other species such as *Clostridium*, *Lactobacillus*, *Peptostreptococcus*, *Bacteroides* that promote the carcinogenesis [36]. Thus, understanding how dysbiosis influences metabolic reactions and inflammatory responses is critical.

In this study, we hypothesized that the alterations in gastric microbial communities are associated with the risk of GC. Therefore we conducted a case-control study to apply compositional analysis of microbiome data to derive a microbial dysbiosis index (MDI) associated with GC risk and further to identify the microbial abundances and enriched metabolic functions associated with subsequent GC risk.

Results

General characteristics

The general characteristics of the study participants can be found in the previously published article [23] (**Additional file: Table S 1)**. In brief, GC cases had higher proportion of current smokers (29.1%) than the controls (17.7%). Higher proportion of GC cases had family history of GC ($p = 0.003$), had not engaged in regular exercise ($p < 0.001$), had lower education ($p < 0.001$), had lower employment rates ($p = 0.037$), had

low level of monthly income ($p < 0.001$), had HP positive status ($p < 0.001$) than the controls. Daily energy consumption was higher in GC cases than the controls ($p < 0.001$).

Microbial taxa distribution and microbial diversity

At the species level, HP was the most dominant species in GC cases and controls (91.57 ± 18.56 and 87.93 ± 27.18%), followed by *Haemophilus parahaemolyticus* (0.88 ± 3.61 and 0.88 ± 3.08%), *Streptococcus_NCVM* (0.78 ± 2.45 and 0.75 ± 2.87%), *Neisseria subflava* (0.49 ± 1.49 and 0.66 ± 2.56%), *Neisseria perflava* (0.47 ± 1.85 and 1.05 ± 4.61%), *Streptococcus_JUTE* (0.37 ± 0.84 and 0.51 ± 1.26%), *Prevotella melaninogenica* (0.32 ± 1.00 and 0.71 ± 2.21%), *Staphylococcus warneri* (0.23 ± 3.49% and 0.09 ± 0.001%), *Prevotella pallens* (0.20 ± 0.80% and 0.40 ± 1.45%), *Streptococcus_PAC001345* (0.20 ± 0.66% and 0.28 ± 1.06%), and *Veillonella atypica* (0.19 ± 0.80% and 0.31 ± 1.15%) (**Additional file: Fig. S1**).

Table 1 shows the alpha diversity indices between cases and controls. There was a significantly higher Shannon index in the controls than GC cases ($p = 0.030$). Significantly higher Shannon index was observed in controls than the GC cases in females ($p = 0.007$). There was a significantly higher richness in GC cases compared to controls in total population ($p = 0.009$) and male ($p < 0.001$). Pilou evenness based on the Shannon index was significantly higher in controls than GC cases in total population ($p = 0.004$) and in females ($p = 0.004$).

Table 1
Comparison of diversity measures between GC cases and controls

	All (n = 556)		
	Controls(n = 288)	Cases(n = 268)	p-value
Shannon index	2.06 ± 2.59	1.66 ± 1.06	0.030
Richness	32.25 ± 16.93	35.89 ± 16.00	0.009
Evenness	0.14 ± 0.06	0.15 ± 0.03	0.440
Pilou evenness	0.58 ± 0.61	0.46 ± 0.37	0.004
	Male (353)		
	Controls(n = 181)	Cases(n = 172)	
Shannon index	1.93 ± 2.37	1.79 ± 1.77	0.519
Richness	31.55 ± 16.44	38.26 ± 16.25	< 0.001
Evenness	0.14 ± 0.07	0.15 ± 0.03	0.560
Pilou evenness	0.55 ± 0.55	0.48 ± 0.40	0.190
	Female (n = 203)		
	Controls(n = 107)	Cases(n = 96)	
Shannon index	2.27 ± 2.93	1.43 ± 1.21	0.007
Richness	33.44 ± 17.76	31.66 ± 14.72	0.440
Evenness	0.14 ± 0.03	0.14 ± 0.03	0.578
Pilou evenness	0.63 ± 0.69	0.42 ± 0.28	0.004

Additional file: Fig. S2 represents a principal coordinate analysis (PCoA) plot of the Bray-Curtis distance based on the relative abundance table at the species level. The red dots represent the GC patients, while the blue triangles represent the controls. The red and blue ellipses represent cases in which 95% of the data belong to the GC patients and controls at a 5% significance level. The 2-D plot of the first two principal coordinates shows a marked divergence between the GC patients and the healthy controls. The total diversity captured by the first two principal coordinates was 32.4%. The microbiota composition of the patients with GC was significantly different from that of the healthy controls (Analysis of similarities (ANOSIM) R=-0.00015, p = 0.004) according to the Bray-Curtis dissimilarity measure.

Differentially abundant taxa

To identify the specific bacterial taxa associated with GC, different taxonomy levels were compared using Linear discriminant analysis (LDA) effect size (LEfSe) analysis based on the non-parametric factorial

Kruskal-Wallis (KW) sum rank test to detect bacterial taxa with significant differential abundance between GC cases and controls. Then, LEfSe uses LDA to estimate the effect size of each differentially abundant taxa with the criteria of LDA ≥ 2 ad $p < 0.05$. At the species level, 8 species were differentially abundant in GC cases and controls out of the total of 945 species. Of them, *Prevotella melaninogenica, Prevotella nigrescens, Prevotella intermedia, Streptococcus_CP003667, Gemella taiwanensis, Streptococcus vestibularis* were highly enriched in controls whereas *Campylobacter jejuni, Streptococcus_NCVM* species were highly abundant in GC cases (Fig. 1).

Association between candidate taxa and GC risk

Table 2 shows the association between the relative abundance of bacteria species and GC risk. Those who are in the third tertile of the *Prevotella melaninogenica* showed a significantly decreased risk of GC compared to the lowest tertile in total population (OR: 0.91, 95% CI: 0.38–0.96, p-trend = 0.071). Subjects who carry *Prevotella nigrescens* showed a significantly decreased risk of GC compared to those who did not carry (OR: 0.64, 95% CI: 0.43–0.94). In contrast those who carry *Streptococcus vestibularis* showed a significantly increased risk of GC (OR: 2.41, 95% CI: 1.63–3.56) in the total population.

Table 2
Association between relative abundance of bacterial species and GC risk

Candidate species	No. of controls (%)	No. of cases (%)	Model I OR (95% CI)	Model II OR (95% CI)
Campylobacter jejuni				
0 (Non-carriers)	288(100.0)	264(98.5)	1.00	1.00
> 0 (Carriers)	0(0.0)	4(1.5)	> 999.99(< 0.001->999.99)	> 999.99(< 0.001->999.99)
Streptococcus_CP003667				
0 (Non-carriers)	187(64.9)	203(75.8)	1.00	1.00
> 0 (Carriers)	101(35.1)	65(24.3)	0.59(0.41–0.86)	0.58(0.38–0.88)
Gemella taiwanensis				
0 (Non-carriers)	155(53.8)	118(44.0)	1.00	1.00
> 0 (Carriers)	133(46.2)	150(55.9)	1.48(1.06–2.07)	1.10(0.75–1.62)
Streptococcus_NCVM				
< 0.000458	95(33.0)	62(23.1)	1.00	1.00
0.000458–0.00204	96(33.3)	92(34.3)	1.47(0.96–2.26)	1.26(0.78–2.03)
>=0.00204	97(33.7)	114(42.5)	1.80(1.18–2.74)	1.37(0.85–2.21)
p-trend			0.022	0.308
Prevotella intermedia				
0 (Non-carriers)	133(46.2)	144(53.7)	1.00	1.00
> 0 (Carriers)	155(53.8)	124(46.3)	0.74(0.53–1.03)	0.69(0.47-1.00)
Prevotella melaninogenica				
< 0.000356	95(33.0)	113(42.2)	1.00	1.00
0.000356–0.00178	96(33.3)	77(28.7)	0.67(0.45–1.01)	0.68(0.43–1.07)
>=0.00178	97(33.7)	78(29.1)	0.68(0.45–1.01)	0.91(0.38–0.96)
p-trend			0.141	0.071
Prevotella nigrescens				

Model I: Crude

Model II: Adjusted for age, family history of GC, regular exercise, education, occupation, income, total energy intake

Candidate species	No. of controls (%)	No. of cases (%)	Model I OR (95% CI)	Model II OR (95% CI)
0 (Non-carriers)	157(54.50	179(66.8)	1.00	1.00
> 0 (Carriers)	1319(45.5)	89(33.2)	0.60(0.42–0.84)	0.64(0.43–0.94)
Streptococcus vestibularis				
0 (Non-carriers)	153(53.1)	81(30.2)	1.00	1.00
> 0 (Carriers)	135(46.9)	187(69.8)	2.62(1.85–3.71)	2.41(1.63–3.56)
Model I: Crude				
Model II: Adjusted for age, family history of GC, regular exercise, education, occupation, income, total energy intake				

Phylogenetic relationships

Figure 2-A represents a histogram of the LDA scores calculated for the taxa which showed the significant bacterial difference between GC cases and healthy controls and Fig. 2-B shows the cladogram based on the Ezbio database. LDA score at the $\log_{10}$ scale is indicated at the bottom. The greater the LDA score is, the more significant the microbial biomarker is in the comparison. According to the cladogram using the LEfSe method, it indicates the phylogenetic distribution of gastric microbes associated with patients with GC (red indicates phylotypes statistically over-represented in GC) and in healthy subjects (green indicates phylotypes over-represented in healthy subjects). Each filled circle in the cladogram represents one phylotype and phylum and class are indicated in their names on the cladogram and the order, family, or genera are given in the right panel. Taxa related to phylum *Actinobacteria*; class *Actinobacteria* was phylogenetically enriched in healthy controls whereas the taxa related to class *Bacilli* was phylogenetically enriched in GC cases.

Microbial dysbiosis index (MDI)

Microbial dysbiosis index (MDI) was derived for the current study population based on the compositional data analysis of the microbiome using Compositionality Corrected by Renormalization and Permutation (CCREPE) method. After selecting 64 out of 73 genera based on the CCREPE method, fold change values were calculated to identify the genera which were increased in GC and decrease in GC. Of them, 13 genera were identified as increased in GC (fold change > 1) (**Additional file: Table S 2**) while 51 genera were identified as decreased in GC (fold change < 1) (**Additional file: Table S 3**). Top 10 genera that were diverged from the segment line to the right side (> 1) and the top 10 genera which were diverged from the segment line to the left side (< 1) were represented as a diverging lollipop chart for simplicity (Fig. 3).

The MDI was higher for the GC cases compared to the healthy controls and the result was marginally significant ($p = 0.097$) for the total population. A significantly higher MDI was observed for the GC cases

than the controls in the female population (p = 0.002) (Table 3). Table 4 shows the association between MDI and GC risk for the total population, male and female. Even though the subjects who are in the third tertile of the MDI showed an increased risk of GC, the associations were not significant for total and male populations. However, there was a significantly increased risk of GC for those who have higher MDI in females (OR: 2.66, 95% CI: 1.19–5.99, p-trend = 0.017) in model II. **Additional file: Table S 4** shows the stratified analysis of the association between MDI and GC based on lifestyle factors. Among those who are never smokers, a significantly increased risk of GC was observed for the subjects in the third tertile of MDI in model II (OR: 2.37, 95% CI: 1.14–4.92, p-trend = 0.019).

Table 3
Comparison of microbial dysbiosis index (MDI) between cases and controls

Total	**Cases(268)**	**Controls(288)**	**p-value**
MDI	3.77 ± 1.94	3.45 ± 2.59	0.097
Male	Cases(172)	Controls(181)	
MDI	3.52 ± 2.04	3.58 ± 2.43	0.773
Female	Cases(96)	Controls(107)	
MDI	4.23 ± 1.65	3.22 ± 2.84	0.002
MDI: Microbial dysbiosis index			

Table 4
Association between microbial dysbiosis index (MDI) and GC risk

Microbial Dysbiosis Index (MDI)	No. of controls (%)	No. of cases (%)	Model I	Model II
Total				
T1(< 3.18)	96(33.3)	91(33.9)	1.00	1.00
T2(3.18–4.52)	97(33.7)	75(27.9)	0.82(0.54–1.24)	0.97(0.60–1.57)
T3(≥ 4.52)	95(33.0)	102(38.1)	1.13(0.76–1.69)	1.37(0.86–2.17)
p for trend			0.561	0.179
Male				
T1(< 3.25)	60(33.2)	74(43.0)	1.00	1.00
T2(3.25–4.48)	60(33.2)	42(24.4)	0.57(0.34–0.96)	0.80(0.43–1.52)
T3(≥ 4.48)	61(33.7)	56(32.6)	0.74(0.45–1.22)	1.15(0.63–2.11)
p for trend			0.225	0.657
Female				
T1(< 3.04)	36(33.6)	18(18.8)	1.00	1.00
T2(3.04–4.52)	36(33.6)	31(32.3)	1.72(0.82–3.62)	1.69(0.71–4.02)
T3(≥ 4.52)	35(32.7)	47(48.9)	2.69(1.31–5.49)	2.66(1.19–5.99)
p for trend			0.006	0.017
Model I: Crude				
Model II: Adjusted for age, family history of GC, regular exercise, education, occupation, income, total energy intake				

Nonmetric multidimensional scaling (NMDS)

Nonmetric multidimensional scaling (NMDS) was performed to relate the microbial composition with the metadata of the study population. The NMDS is one of the ordination methods to represent the multidimensional space into a two-dimensional scale. It regresses distances in the initial configuration against the observed distances (Bray-Curtis dissimilarity). The way it calculates how well the two-dimensional representation of multidimensional space is by using stress value. Stress is the

disagreement between 2-D configuration and predicted values from the regression. A rule of thumb: stress < 0.05 provides excellent representation in reduced dimensions. The multidimensional space of the microbial species level indicated that there is an excellent representation in reduced dimensions with the stress value 0.043. The nonmetric correlation coefficient between observed dissimilarity and ordination distance was (R^2 = 0.998). According to NMDS results, the microbial composition of the HP positive and negative groups was significantly different (Permutational Multivariate Analysis of Variance (PERMANOVA); p = 0.001) (**Additional file: Fig. S3**).

Microbial prediction functions

LEfSe analysis was performed to distinguish the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways between GC cases and controls. Fourteen pathways were presented significantly different between GC cases and control groups (Fig. 4). Seven pathways were differentially abundant in the healthy controls whereas eight pathways were differentially abundant in GC cases (LDA score > 2 at 5% significance). Biosynthesis of ansamycins pathway was identified as one of the critical bacterial metabolic functions that highly enriched in healthy controls.

The KEGG pathway of biosynthesis of ansamycins (ko01051, p < 0.001), one carbon pool by folate (ko00670, p < 0.001), peptidoglycan biosynthesis (ko00550, p < 0.001), lipopolysaccharide biosynthesis (ko00540, p = 0.028), RNA polymerase (ko03020, p < 0.001), nicotinate and nicotinamide metabolism (ko00760, p < 0.001), homologous recombination (ko03440, p < 0.001) were enriched in healthy controls whereas secondary bile acid biosynthesis (ko00121, p < 0.001), endocytosis (ko04144, p = 0.019), Vibriocholera pathogenic cycle (ko05111, p < 0.001), dioxin degradation (ko00621, p < 0.001), cyanoamino acid metabolism (ko00460, p < 0.001), beta-alanine metabolism (ko00410, p = 0.002), toluene degradation (ko00623, p = 0.012), and atrazine degradation (ko00791, p = 0.007) were abundant in GC cases (**Additional file: Table S 5**). Several metabolites that were involved in the KEGG pathways were identified based on the KEGG orthology (**Additional file: Table S 5**). Particularly, transketolase enzyme (K00615, p = 0.004)) which is involved in the biosynthesis of ansamycins (ko01051) pathway was significantly enriched in healthy controls than GC cases. Formamidase (K01455, p < 0.001) and L-asparaginase (K13051, p = 0.033) enzymes which are involved in the cyanoamino acid metabolism (ko00460) pathway were highly enriched in the GC cases than the controls. Secondary bile acid biosynthesis (ko00121) pathway had choloylglycine hydrolase (K01442) orthology which is highly enriched in GC cases than the controls.

Discussion

In this study, the gastric microbiota composition was characterized between GC cases and controls. Calculation of the MDI by applying compositional analysis of microbiome data is a novel approach to identify the gastric dysbiosis and further to observe the associations with GC risk. Moreover, the metagenomics functions were predicted to identify the pathways associated with GC risk. The MDI was higher for GC cases than healthy controls for the total population. A significant higher MDI was observed

for GC cases than controls in the female population (p = 0.002). In females, those who are in the third tertile of the MDI showed a significantly increased risk of GC (OR: 2.66, 95% CI: 1.19–5.99, p-trend = 0.017; model II). NMDS results indicated that the microbial composition of HP positive and negative groups was significantly different (PERMANOVA; p = 0.001). Regarding KEGG pathways, biosynthesis of ansamycins pathway was identified as a critical pathway which is differentially abundant in healthy controls.

In our study, *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Fusobacterium* were the most dominant phyla in case and control groups. Comparison with the previous research findings, *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, and *Fusobacteria* was identified as the most dominant phyla in the gastric environment which shows consistent results with the current study [19, 22, 37–39]. A study conducted using Chinese and Mexican populations revealed that HP was the predominant member of the microbiota in the gastric environment [40]. A study aimed to evaluate the microbial composition of the gastric mucosa found that the gastric microbial composition of the patients who had non-atrophic gastritis, intestinal metaplasia, and GC of the intestinal type was with *Firmicutes* and *Proteobacteria* phyla which account 70% of in each sample [41]. A similar result was observed in our study indicating that *Bacteroidetes*, *Firmicutes*, *Proteobacteria* are dominant phyla in the gastric environment in GC cases and controls. Another study evaluated the gastric microbiota in individuals based on different histopathological stages of GC reported that there were 19 genera with average relative abundance > 0.5% across 60 samples at the genus level where the high abundance genera were *Helicobacter*, *Flavobacterium*, *Haemophilus*, and *Serratia* [38]. The finding of this study related to the *Helicobacter*, *Haemophilus*, and *Neisseria* genera are consistent with our findings indicating those three genera are most dominant in both GC cases and control groups. A Korean study observed that the *Epsilonproteobacteria* class corresponding to HP species was the predominant, but the abundance of *Bacilli* class was relatively increased in the GC group which is consistent with our findings [42].

A study observed that there were a higher evenness and diversity of the gastric microbiota in the GC group in comparison with chronic gastritis and intestinal metaplasia groups [42]. We observed that the evenness was more or less similar in both GC cases and controls. Interestingly, richness was significantly higher in GC cases in the current study population. A similar finding has been reported in the study compared cancer tissues with non-cancer tissues where there is a high microbial richness in GC tissues [21]. Moreover, we observed that there was a higher Shannon index in the controls than the GC group. Another study that compared the chronic gastritis group with the GC group observed that there is a higher Shannon index in the chronic gastritis group compared to the GC group [22]. A study focused on the relationship between gastric dysbiosis and GC development found that there is an increased richness although the Shannon index is lower in the GC group compared to controls which is similar to our findings [43]. On the contrary, a study conducted on microbiota in gastric mucosa in the GC tissues compared with the non-cancer tissues revealed that there is a significantly higher Shannon index in the cancer group compared with the healthy controls [21]. A study reported that 75.86% was captured by the first two principal coordinates in the PCoA beta diversity plot according to the weighted UniFrac phylogenetic distance measure. They have observed that there was a significant divergence between non-

cancer and cancer samples since those samples were clustered separately [21]. A study conducted in Mexica performed an ordination analysis of the 44 taxa between non-atrophic gastritis and GC based on weighted-UniFrac distance measure and reported that there is a significant separation of the microbiota composition between two groups [41].

In species level, *Streptococcus_NCVM* and *Campylobacter jejuni* were differentially abundant in the GC cases while *Prevotella melaninogenica* was differentially abundant in the controls based on the LEfSe analysis. Although the evidence related to the effects of those bacterial species in GC occurrence is limited, a study has reported that there is a remarkable effect of the metabolic products produced by those *Prevotella* species such as lactic acid, acetic acid, butane diacid, isovaleric acid, isobutyric acid to the human gastric cell physiology [44]. Furthermore, *Prevotella melaninogenica* enrichment can create the gastric environment more likely to be acidic by lowering pH than non-atrophic gastritis where there is a restriction to colonize by other harmful bacterial species [44]. A study evaluated the microbiota composition in advanced gastric adenocarcinoma through the shotgun metagenomics approach and they have reported that the cladogram of the gastric microbiome phylogenetically associated with GC and superficial gastritis. Family *Porphyromonadaceae*, genus *Porphyromonas*, genus *Alloprevotella* were enriched in the GC group whereas genus *Actinomyces*, and genus *Atopobium* were enriched in the superficial gastritis group. However, when comparing to our study, this study does not have similar taxa which were highly enriched specifically in the GC case group [45]. Based on the cladogram, the *Bacilli* class was phylogenetically related in GC cases in the current study that is similar to the findings observed by Liu *et al.* [17].

Based on the compositional analysis of microbiome data in the genus level, there was a higher MDI in GC cases than controls and it was marginally significant for the total population ($p = 0.097$) and it was significant for female ($p = 0.002$). Furthermore, higher MDI showed a significantly increased risk of GC in females (OR: 2.66, 95% CI: 1.19–5.99, p-trend = 0.017; model II). A study conducted to observe the association between gastric dysbiosis of the gastric microbiome and GC risk concluded that there is a higher MDI in GC patients than those who had chronic gastritis ($p < 0.0001$) [22]. A study carried out to characterize the changes of the microbiome associated with histopathological stages of gastric tumorigenesis observed that there is a significant microbial dysbiosis of gastric mucosa in GC patients with a significant overrepresentation of 21 and reduction of 10 bacterial taxa in GC in comparison to superficial gastritis ($q < 0.05$) [46].

In one hand for deriving MDI based on the fold changes of selected genera, *Lactobacillus*, *Diaphorobacter*, *Acinetobacter*, *Atopobium*, *Actinobacillus*, and *Rhizobium* genera are top six genera out of 13 genera enriched in GC cases in the current study population. Particularly, *Lactobacillus* has a critical role in carcinogenesis because of N-nitroso compounds derived from the metabolism of nitrate/nitrite [28]. In fact, several previous microbiome studies have reported that there is an increment in the abundance of *Lactobacillus* in GC patients [41–43]. MDI was significantly positively associated with the risk of GC specifically in females in the current study. As a plausible biological mechanism, it has been reported that the gut microbiome is one of the principal regulators of circulating estrogen in females [47].

The gut microbiota secretes β-glucuronidase which is an enzyme that deconjugates estrogens into their active forms where there is a direct regulation of estrogens by gut microbiota. Once the dysbiosis of the microbiota is taken place that is characterized by lower microbial diversity, it can impair the above-mentioned deconjugation process where there is a reduction of the circulating estrogens. The alterations of circulating estrogens may affect to develop several pathological conditions particularly GC in females [47]. Three possible mechanisms have been proposed for the carcinogenesis due to the microbial dysbiosis [26].

The first mechanism is related to the bacterial-induced chronic inflammation. The inflammatory mediators produced due to chronic inflammation have harmful effects on epithelial, endothelial cells, and extracellular matrix compounds. During this inflammatory process, epithelial and immune cells trigger ROS and reactive nitrogen species (RNS) due to the direct influence of TNF-α, IL-6, and TGF-β [48]. Production of ROS and RNS occurs via induction of NADPH oxidase and nitric oxide synthase. NADPH oxidase catalyzes the superoxide anion leading to superoxide dismutase mediated hydrogen peroxide H_2O_2 production. Simultaneously, nitric oxide synthase generates nitric oxide (NO), which can be converted into nitrogen dioxide (NO_2), peroxinitrite, and dinitrogen trioxide (N_2O_3) to produce their ROS and RNS. Interestingly, increased expression of NADPH oxidase, nitric oxide synthase, and their ROS and RNS species have been identified in tumor microenvironment [49]. Additionally, cell proliferation, mutagenesis, oncogene activation, and angiogenesis can be facilitated by the inflammatory mediators produced by the above mechanism.

In the second mechanism, NF-κB can be activated and cellular apoptosis can be inhibited. Activation of NF-κB pathway that is related to oncogenic cell signaling in epithelial cells has been identified as a critical pathway for the TNF- α induced tumor growth. NF-κB signaling can be categorized into a “classical” pathway and “alternative” pathway. In classical pathway IκB kinase β (IKKβ) phosphorylates IκBα whereas in alternative pathway IKKα phosphorylates the p100 precursor of the NF-κB p52 subunit. There is an accumulation of the heterodimeric NF-κB transcription factors in the nucleus as a result of the above signaling events. Classical pathway regulates mainly p50/p65/ and p50/c-Rel dimers and the alternative pathway regulates the p52/relB dimers. Other kinases including the unconventional IKK family members, IKKε, and TBK1 can also activate the NF-κB pathway. Several signaling pathways converge on the NF-κB regulators provides significant evidence where cancers can aberrantly stimulate NF-κB [50]. Further, it has an effect to activate pro-inflammatory cytokines, oncogenes, and to induce cancer cell proliferation. In the third mechanism, bacterial substances can act as carcinogenic substances that induce the carcinogenesis process [26]. The integration of those mechanisms can potentially stimulate the carcinogenesis process with the involvement of microbial dysbiosis. Particularly, local concentration of various cytokines including interleukin-1β (IL-1β), IL-6, tumor necrosis factor-α (TNF-α) can be increased due to the microbial dysbiosis. Endothelial cells can be activated by IL-1β to produce vascular endothelial growth factor (VEGF). VEGF eventually generates an inflammatory microenvironment which is helpful for angiogenesis and tumor progression [51]. Also, TNF-α has an ability to produce ROS that can induce DNA damages. A study focused on the relationship between tumor-immune environment

associated with GC microbiota in patients who have GC identified that there is a correlation between regulatory T cells and plasmacytoid dendritic cells in the tumor microenvironment that is further associated with the dysbiosis of the gastric microbiota [24].

Regarding the metagenomics functional pathway results, biosynthesis of ansamycins pathway was highly enriched in controls. It has been reported that ansamycins is a groups of antibiotic produced by strains of several *Actinomycetes*. Ansamycins have proved to be very potent molecules displaying anticancer, antibacterial, and antiviral activities [52]. The one carbon pool by folate pathway was highly enriched in controls than GC cases. One carbon metabolism mediated by folate cofactor supports the multiple physiological processes including biosynthesis of purines and thymidine, amino acid homeostasis (glycine, serine, and methionine), epigenetic maintenance, and redox defense. While most gut bacteria can synthesize the folate, humans require the dietary folate intake to maintain the physiological processes [53]. It has an essential role in the nucleic acid synthesis. It has been noted that the adequate folate level over the long term may support genome integrity. Based on the evidence of experimental and epidemiological studies, there is a protective effect of the folate towards the colorectal cancer, breast cancer, and pancreatic cancer while there are inconsistent results for the gastric cancer [54–56]. Secondary bile acid biosynthesis pathway was also highly enriched in GC cases than controls. Secondary bile acids can induce reactive oxygen species production, genomic destabilization, apoptosis resistance, and cancer stem cells-like formation. There are diverse signals involved in the carcinogenesis mechanism of bile acids, with a major role of epidermal growth factor receptor, and its down-stream signaling, involving mitogen-activated protein kinase, phosphoinositide 3-kinase/Akt, and nuclear factor kappa-light-chain-enhancer of activated B cells. Bile acids regulate numerous genes including the human leukocyte antigen class I gene, p53, matrix metalloprotease, urokinase plasminogen activator receptor, Cyclin D1, cyclooxygenase-2, interleukin-8, and miRNAs of cancer cells [57].

There are several strengths of our study. To the best of our knowledge, this is the first study that is associated with the methodological approach that employed the compositional analysis of microbiome data using a novel statistical approach termed “CCREPE” to derive a MDI for a Korean population. The main strength of this approach is it abrogates the spurious correlations when determining the significance of a similarity measure. Second, the sample size of current microbiome-related study is comparatively large with 268 GC cases and 288 healthy subjects relative to previous microbiome studies and it improves the power of statistical analysis to observe the relevant associations between the microbiome and the risk of GC. Third, several potential covariates were considered in a multivariate analysis that are established risk factors for GC development. Those confounding variables are age, smoking, family history of GC, regular exercise, education, occupation, income, and total energy intake throughout the analysis.

However, there are potential limitations associated with the current study. Generally, selection bias and recall bias need to be raised since this study is a hospital-based case-control study. Selection bias might have occurred because healthy subjects were selected from the participants who attended the health screening. They may have a healthier lifestyle due to health concerns compared to those who do not

participate in screening. Therefore, healthy subjects might be less representative of the general population. Second, current study is not a follow-up study. Thus, the associations between the gastric microbiome and GC risk can occur without having a causal relation because patients with early GC have changed their microbial profile because of premalignant lesions that already have been progressed and due to their changes in the dietary habits. However, cases included only patients diagnosed with early GC in this study. Thus, the influence of the dietary changes on GC symptoms will be slight. Third, MDI has not been validated although it has been applied in an epidemiological case-control study. Furthermore, since a single sample was measured for the microbial measurements in the current study, the results related to microbial exposure may have less accuracy compared with the microbiome measurements in multiple time points [58]. However, it is important to note that repeating biopsies with those who have normal gastric histology has ethical issues.

Conclusions

In conclusion, evidence based on the current results found that specific bacterial pathogens may play a role in GC risk whereas other bacterial types may be associated with reduced risk. Moreover, it was observed that there is a significantly higher MDI in GC cases than controls and further, MDI is significantly positively associated with GC risk in the female population. Microbial composition was significantly different based on HP infection status. Particularly, it was identified that the microbial metabolic pathways of gastric microbiota associated with GC. Specifically, the biosynthesis of ansamycins pathway is highly enriched in the GC cases.. Evaluating the gastric microbial composition associated with GC will be beneficial to develop novel preventive guidelines to prevent GC risk based on the individual microbiome profile of the Koreans in the future.

Methods

Study population

Participants were recruited at the National Cancer Center Hospital in Korea between March 2011 and December 2014. Individuals who had been histologically confirmed as early GC patients within the preceding three months at the Center for Gastric cancer were included in the case group. Early GC was defined as an invasive carcinoma confined to the mucosa and/or submucosa, regardless of lymph node metastasis status. Patients diagnosed with diabetes mellitus, a history of cancer within the past five years, advanced GC, or severe systemic or mental disease, as well as women who were pregnant or breastfeeding, were excluded. The control group was selected from health-screening examinations at the Center for Cancer Prevention and Detection at the same hospital. Individuals with a history of cancer, diabetes mellitus, gastric ulcers, and HP treatment in the control group were excluded. The final sample of 556 participants was composed of 268 patients and 288 controls (men, 353; women, 203). This study was approved by the Institutional Review Board of the National Cancer Center [IRB Number: NCCNCS-11-438]. Written informed consent was obtained from all participants.

Data collection

Five gastric mucosa biopsy samples were collected from each study participant following the Sydney system after endoscopy and examination of the stomach. A biopsy sample in the greater curvature, at least 3 cm away from each tumor, was used for the metagenomics analysis. The HP infection status was determined by a rapid urease test, a serological test and histological evaluation. Regarding the rapid urease test, one biopsy sample was taken from the greater curvature of the corpus. Four biopsy samples were collected from the lesser curvature of the corpus and antrum for histological evaluation. The HP status was determined via Wright-Giemsa staining of the biopsy specimens by a pathologist who specialized in GC. A current infection was defined as at least one positive test result in the rapid urease test or histological evaluation of four biopsy sites [59]. Participants were asked to complete a self-administered questionnaire. Demographic, lifestyle, physical activity, and medical history data were collected from the participants. Total energy intake was obtained from the semi-quantitative food frequency questionnaire (SQFFQ), which has been previously reported as a reliable and valid questionnaire [60]. HP infection was assessed by a rapid urease test and histological evaluation.

DNA extraction

DNA was extracted from the biopsy samples using the MagAttract DNA Blood M48 kit (Qiagen, Hilden, Germany) and BioRobot M48 automatic extraction equipment (Qiagen), according to the manufacturers' instructions.

Metagenomics 16S rRNA gene sequencing

Input gDNA (12.5 ng) was amplified with 16S rRNA gene V3-V4 primers, and a subsequent limited cycle amplification step was performed to add multiplexing indices and Illumina sequencing adapters. The final products were normalized and pooled using PicoGreen, and the library sizes were verified using the LabChip GX HT DNA High Sensitivity Kit (PerkinElmer, Massachusetts, USA). Then, we sequenced using the MiSeq™ platform (Illumina, San Diego, USA). Each sequenced sample was prepared according to the Illumina 16S rRNA gene Metagenomic Sequencing Library protocols. DNA quantification and quality were measured by PicoGreen and Nanodrop analyses, respectively. The 16S rRNA genes were amplified using 16S rRNA gene V3-V4 primers for the 288 control samples and the 268 GC patient samples. The primer sequences are as follows: 16S rRNA gene V3-V4 primer.

16S rRNA gene Amplicon PCR Forward Primer.

5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG.

16S rRNA gene Amplicon PCR Reverse Primer.

5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC.

The paired-end FASTQ files that were already been demultiplexed were imported to make QIIME2 artifact files. After removing the barcodes/adaptors using Cutadapt, DADA2 pipeline was applied to remove noisy

reads, dereplicate sequences, cluster sequences, and remove chimera using QIIME v2.2019.7 [61]. Amplicon sequence variants (ASVs) table was obtained as the end product. Taxonomic abundance was counted with the Ezbio database [62]. Host mitochondrial and chloroplast, archaea, eukaryotes, and unassigned reads were filtered before calculating relative abundance. The microbial composition was normalized using the values calculated from the taxonomic abundance count divided by the number of preprocessed reads for each sample to obtain the relative abundance.

Statistical analysis

Descriptive statistics

To compare the demographic and lifestyle characteristics between the cases and controls, the chi-square test and Student's *t*-test were performed for categorical variables and continuous variables, respectively.

Linear discriminant analysis of effect size (LEfSe) and cladogram

LEfSe analysis [63] was used to estimate which microbiome attributes differ significantly by cancer status for six taxonomy levels namely phylum, class, order, family, genus and species. LEfSe couples a univariate nonparametric test for statistically significant phenotypic segregation with post hoc prioritization by the size of the effect as determined by LDA. Galaxy implementation of LEfSe [64] with default options was used. Differences were evaluated via a threshold for the logarithmic LDA score for discriminate features for 2.0. Ezbio feature table was prepared using the Ezbio database to plot the cladogram.

Deriving microbial dysbiosis index (MDI)

Compositional analysis of microbiome data was performed using CCREPE. This is a novel statistical methodology for co-variation analysis in compositional data [65]. It allows the derivation of accurate significance values for arbitrary association measures (correlation or other similarity scores) when applied to compositional data. It consists of an R package (publicly available through R/Bioconductor – (http://huttenhower.org/ccrepe) and N-dimensional checkerboard score (NC-score), a novel similarity measure specifically designed to detect association patterns in the human microbiome and other microbial communities. This NC-score is an extension to arbitrary nominal categories of the classical checkerboard score for ecological species co-occurrence. For each pair of microbe m_1 and m_2, the NC-score counts the normalized number of co-variation and co-exclusion over all pair of samples s_1 and s_2. CCREPE method was applied with the relative abundance data set containing 73 genera. CCREPE results were obtained with four matrices (P-values, Z-stat values, NC score, and False Discovery Rate (FDR) corrected Q values). The sub-correlation matrix of NC score was extracted by two following criteria; FDR corrected Q-values < 0.05 and pairs of genera NC score |>0.30|. Finally, 64 genera were selected for further analysis. Fold change of selected genera was calculated by dividing the mean abundance in the cases by that of the controls to identify the genera increased in GC (fold change > 1) and decreased in GC (fold change < 1) and presented in a diverging lollipop chart using the R package "ggplot2". The MDI was

calculated as the log of [total abundance in genera increased in GC] over [total abundance in genera decreased in GC].

Association between gastric microbiome and GC risk

The relative abundance of the candidate taxa was categorized into tertiles based on the relative abundance in the control group. Exceptionally, if more than one-third of the subjects have a relative abundance of zero, those bacterial species were categorized into two groups (non-carriers and carriers) based on the median distribution of the controls. Non-carriers were defined as subjects who had a relative abundance of zero. The group with the lowest relative abundance was used as the reference group. The ORs and 95% CIs were estimated using unconditional logistic regression models. The median values of relative abundance in each tertile category were used as continuous variables to test for trends. The OR estimates were calculated for the crude model (model I) and model II. Model II was adjusted for age, smoking, first-degree family history of GC, regular exercise, education, occupation, monthly income, and total energy intake.

Nonmetric multidimensional scaling (NMDS)

NMDS was performed to relate the microbial composition with the metadata of the study population. Ordination distances were calculated based on Bray-Curtis distance measure with 10000 iterations by using metaMDS function of “vegan” package in R. The Shepard plot was drawn to observe the correlation between observed dissimilarity and the ordination distance. Corresponding stress value was obtained to indicate the goodness of fit for two-dimensional representations of multidimensional space of microbial species. NMDS plots were drawn to observe the difference of microbial composition based on metadata of the study population. PERMANOVA; adonis function, vegan package in R was used to check the significance of the microbial composition difference based on selected covariates. All the statistical analysis was carried out using SAS version 9.4 software (SAS Inc., Cary, NC, USA) and the R platform (version 3.5.1) (The R Foundation for Statistical Computing, Vienna, Austria).

Metagenomics functional analysis

The gastric microbial functional gene contents were predicted using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt v2) [66, 67]. In brief, the paired-end FASTQ files that were already been demultiplexed were imported to make QIIME2 artifact files. After removing the barcodes/adaptors using Cutadapt, DADA2 pipeline was applied to remove noisy reads, dereplicate sequences, cluster sequences, and remove chimera using QIIME v2.2019.7 [61]. Amplicon sequence variants (ASVs) table was obtained as the end product. It was normalized and then the metagenome functional predictions were predicted via PICRUSt v2 [68].

List Of Abbreviations

ANOSIM: Analysis of similarities; CCREPE: Compositionality corrected by renormalization and permutation; CI: Confidence intervals; FDR: False discovery rate; GC: Gastric cancer; HP: *Helicobacter*

pylori; KEGG: Kyoto Encyclopedia of Genes and Genomes; LAB: Lactic acid bacteria; LDA: Linear discriminant analysis; LEfSe: Linear discriminant analysis of effect size; MDI: Microbial dysbiosis index; NC-score: N-dimensional Checkerboard scores; NMDS: Nonmetric multidimensional scaling; OTUs: Operational taxonomic units; OR: odds ratio; PCoA: Principal coordinate analysis; PERMANOVA: Permutational multivariate analysis of variance; PICRUSt; Phylogenetic Investigation of Communities by Reconstruction of Unobserved States; QIIME: Quantitative insights into microbial ecology; RNS: Reactive nitrogen species; ROS: Reactive oxygen species; SQFFQ: Semi-quantitative food frequency questionnaire; VEGF: Vascular endothelial growth factor.

Declarations

Ethic approval and consent to participate

This study was approved by the Institutional Review Board of the National Cancer Center [IRB Number: NCCNCS-11-438]. Written informed consent was obtained from all participants.

Consent for publication

Not applicable

Availability of data and materials

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by grants from the National Cancer Center in South Korea (no. 1410260, 1810980, and 1910330)

Authors' contribution

JK had the full access to all of the data in the study and take all responsibility for the integrity of the data and the accuracy of the data. JK and MNG conceived of the study concept and design. IJC, YIK, JK and JFK carried out the acquisition and curation of data. MNG, JL and JKY carried out the formal analysis. MNG wrote the original draft of the manuscript. IJC, YIK, JFK, WJS, and JK carried out the critical revision

of the manuscript for important intellectual content. JK acquired the funding. JK supervised the project. All authors read and approved the final manuscript.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
2. Ferlay JS, Bray F, Forman D, Mather C, Parkin D. GLOBOCAN, Cancer incidence and mortality worldwide: IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from http://globocan.iarc.fr. Cited June 2020.
3. Hong S, Won Y-J, Park YR, Jung K-W, Kong H-J, Lee ES, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2017. Cancer Res Treat. 2020;52(2):335-50.
4. World Cancer Research Fund International CUPD, Nutrition Physical Activity, http://www.Wcrf.Org/int/continuous-updateproject/cup-findings-reports/stomach-cancer. Cited June 2020.
5. Poorolajal J, Moradi L, Mohammadi Y, Cheraghi Z, Gohari-Ensaf F. Risk factors for stomach cancer: a systematic review and meta-analysis. Epidemiol Health. 2020:e2020004.
6. Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R, et al. Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. Cancer Causes Control. 2008;19(7):689-701.
7. Ramos MFKP, Ribeiro Júnior U, Viscondi JKY, Zilberstein B, Cecconello I, Eluf-Neto J. Risk factors associated with the development of gastric cancer—case-control study. AMB Rev Assoc Med Bras. 2018;64(7):611-9.
8. Kim MH, Kim SA, Park CH, Eun CS, Han DS, Kim YS, et al. Alcohol consumption and gastric cancer risk in Korea: a case-control study. Nutr Res Prac. 2019;13(5):425-33.
9. Gunathilake M, Lee J, Jang A, Choi I, Kim Y-I, Kim J. Physical activity and gastric cancer risk in patients with and without Helicobacter pylori infection in a Korean population: a hospital-based case-control study. Cancers. 2018;10(10):369.
10. Yaghoobi M, Bijarchi R, Narod S. Family history and the risk of gastric cancer. Br J Cancer. 2010;102(2):237-42.
11. Oliveira C, Pinheiro H, Figueiredo J, Seruca R, Carneiro F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol. 2015;16(2):e60-e70.
12. Lee HJ, Yang HK, Ahn YO. Gastric cancer in Korea. Gastric Cancer. 2002;5(3):177.
13. McLean MH, El-Omar EM. Genetics of gastric cancer. Nat Rev Gastroenterol Hepatol. 2014;11(11):664.
14. Choi YJ, Kim N. Gastric cancer and family history. Korean J Intern Med. 2016;31(6):1042.

15. Hoang BV, Lee J, Choi IJ, Kim Y-W, Ryu KW, Kim J. Effect of dietary vitamin C on gastric cancer risk in the Korean population. World J Gastroenterol. 2016;22(27):6257.
16. Kim JH, Lee J, Choi IJ, Kim YI, Kwon O, Kim H, et al. Dietary carotenoids intake and the risk of gastric cancer: a case-control study in Korea. Nutrients. 2018;10(8).
17. Liu X, Shao L, Liu X, Ji F, Mei Y, Cheng Y, et al. Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer. EBioMedicine. 2019;40:336-48.
18. Dong Z, Chen B, Pan H, Wang D, Liu M, Yang Y, et al. Detection of microbial 16S rRNA gene in the serum of patients with gastric cancer. Front Oncol. 2019;9:608.
19. Shao D, Vogtmann E, Liu A, Qin J, Chen W, Abnet CC, et al. Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China. Cancer. 2019;125(22):3993-4002.
20. Park CH, Lee JG, Lee A-R, Eun CS, Han DS. Network construction of gastric microbiome and organization of microbial modules associated with gastric carcinogenesis. Sci Rep. 2019;9(1):1-9.
21. Chen X-H, Wang A, Chu A-N, Gong Y-H, Yuan Y. Mucosa-associated microbiome in gastric cancer compared with non-cancer tissues. Front Microbiol. 2019;10:1261.
22. Ferreira RM, Pereira-Marques J, Pinto-Ribeiro I, Costa JL, Carneiro F, Machado JC, et al. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. Gut. 2018;67(2):226-36.
23. Gunathilake MN, Lee J, Choi IJ, Kim YI, Ahn Y, Park C, et al. Association between the relative abundance of gastric microbiota and the risk of gastric cancer: a case-control study. Sci Rep. 2019;9(1):13589.
24. Schulz C, Schütte K, Mayerle J, Malfertheiner P. The role of the gastric bacterial microbiome in gastric cancer: Helicobacter pylori and beyond. Therapeutic Advances in Gastroenterol. 2019;12:1756284819894062.
25. Olsen I, Yilmaz Ö. Possible role of Porphyromonas gingivalis in orodigestive cancers. J Oral Microbiol. 2019;11(1):1563410.
26. Karpiński TM. Role of oral microbiota in cancer development. Microorganisms. 2019;7(1):20.
27. Weng M-T, Chiu Y-T, Wei P-Y, Chiang C-W, Fang H-L, Wei S-C. Microbiota and gastrointestinal cancer. J Formos Med Assoc. 2019;118:S32-S41.
28. Vinasco K, Mitchell HM, Kaakoush NO, Castaño-Rodríguez N. Microbial carcinogenesis: lactic acid bacteria in gastric cancer. Biochim Biophys Acta -Rev Cancer. 2019.
29. Martha AC, A.; Hashimoto, D.; Bogunovic, M.; Spencer, S.P.; Belkaid, Y.; et al. Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. Science 2014, 343.
30. Noto JM, Peek Jr RM. The gastric microbiome, its interaction with Helicobacter pylori, and its potential role in the progression to stomach cancer. PLoS Pathogens. 2017;13(10):e1006573.
31. Thursby E, Juge N. Introduction to the human gut microbiota. Biochem J. 2017;474(11):1823-36.

32. Espinoza JL, Matsumoto A, Tanaka H, Matsumura I. Gastric microbiota: An emerging player in Helicobacter pylori-induced gastric malignancies. Cancer Letters. 2018;414:147-52.
33. Tao J, Li S, Gan R-Y, Zhao C-N, Meng X, Li H-B. Targeting gut microbiota with dietary components on cancer: Effects and potential mechanisms of action. Crit Rev Food Sci Nutr. 2019:1-13.
34. Vivarelli S, Salemi R, Candido S, Falzone L, Santagati M, Stefani S, et al. Gut microbiota and cancer: From pathogenesis to therapy. Cancers. 2019;11(1):38.
35. Martinez-Guryn K, Leone V, Chang EB. Regional Diversity of the Gastrointestinal Microbiome. Cell Host Microbe. 2019;26(3):314-24.
36. Dias-Jácome E, Libânio D, Borges-Canha M, Galaghar A, Pimentel-Nunes P. Gastric microbiota and carcinogenesis: the role of non-Helicobacter pylori bacteria: a systematic review. Rev Esp Enferm Dig. 2016;108(9):530-40.
37. Sung J, Kim N, Kim J, Jo HJ, Park JH, Nam RH, et al. Comparison of Gastric Microbiota Between Gastric Juice and Mucosa by Next Generation Sequencing Method. J Cancer Preve. 2016;21(1):60-5.
38. Li TH, Qin Y, Sham PC, Lau KS, Chu KM, Leung WK. Alterations in Gastric Microbiota After H. Pylori Eradication and in Different Histological Stages of Gastric Carcinogenesis. Sci Rep. 2017;7:44935.
39. Nardone G, Compare D. The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases? United European Gastroenterol J. 2015;3(3):255-60.
40. Yu G, Torres J, Hu N, Medrano-Guzman R, Herrera-Goepfert R, Humphrys MS, et al. Molecular characterization of the human stomach microbiota in gastric cancer patients. Front Cellular Infect Microbiol. 2017;7:302.
41. Aviles-Jimenez F, Vazquez-Jimenez F, Medrano-Guzman R, Mantilla A, Torres J. Stomach microbiota composition varies between patients with non-atrophic gastritis and patients with intestinal type of gastric cancer. Sci Rep. 2014;4:4202.
42. Eun CS, Kim BK, Han DS, Kim SY, Kim KM, Choi BY, et al. Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and gastric cancer using pyrosequencing methods. Helicobacter. 2014;19(6):407-16.
43. Castaño-Rodríguez N, Goh K-L, Fock KM, Mitchell HM, Kaakoush NO. Dysbiosis of the microbiome in gastric carcinogenesis. Sci Rep. 2017;7(1):15957.
44. Dong T, Feng Q, Liu F, Chang LK, Zhou X, Han M, et al. Alteration of stomach microbiota compositions in the progression of gastritis induces nitric oxide in gastric cell. Exp Ther Med. 2017;13(6):2793-800.
45. Hu Y-L, Pang W, Huang Y, Zhang Y, Zhang C-J. The Gastric Microbiome Is Perturbed in Advanced Gastric Adenocarcinoma Identified Through Shotgun Metagenomics. Front Cell Infect Microbiol. 2018;8(433).
46. Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut. 2018;67(6):1024-32.

47. Baker JM, Al-Nakkash L, Herbst-Kralovetz MM. Estrogen–gut microbiome axis: physiological and clinical implications. Maturitas. 2017;103:45-53.
48. Landskron G, De la Fuente M, Thuwajit P, Thuwajit C, Hermoso MA. Chronic inflammation and cytokines in the tumor microenvironment. J Immunol Res. 2014;2014.
49. Piao JY, Lee HG, Kim SJ, Kim DH, Han Hj, Ngo HKC, et al. Helicobacter pylori Activates IL-6-STAT3 Signaling in Human Gastric Cancer Cells: Potential Roles for Reactive Oxygen Species. Helicobacter. 2016;21(5):405-16.
50. Staudt LM. Oncogenic activation of NF-κB. CSH Perspect Biol. 2010;2(6):a000109.
51. Carmi Y, Dotan S, Rider P, Kaplanov I, White MR, Baron R, et al. The role of IL-1β in the early tumor Cell–Induced angiogenic response. J Immunol. 2013;190(7):3500-9.
52. https://www.sciencedirect.com/topics/chemistry/ansamycin. Cited July 2020.
53. Ducker GS, Rabinowitz JD. One-carbon metabolism in health and disease. Cell Metab. 2017;25(1):27-42.
54. Giovannucci E. Epidemiologic studies of folate and colorectal neoplasia: a review. J Nutr. 2002;132(8):2350S-5S.
55. Ericson U, Sonestedt E, Gullberg B, Olsson H, Wirfält E. High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmö Diet and Cancer cohort. AM J Clin Nutr. 2007;86(2):434-43.
56. Larsson SC, Håkansson N, Giovannucci E, Wolk A. Folate intake and pancreatic cancer incidence: a prospective study of Swedish women and men. J Natl Cancer I. 2006;98(6):407-13.
57. Nguyen TT, Ung TT, Kim NH, Jung YD. Role of bile acids in colon carcinogenesis. World J Clin Cases. 2018;6(13):577-88.
58. Hayes RB, Ahn J, Fan X, Peters BA, Ma Y, Yang L, et al. Association of Oral Microbiome With Risk for Incident Head and Neck Squamous Cell Cancer. JAMA Oncol. 2018;4(3):358-65.
59. Woo HD, Fernandez-Jimenez N, Ghantous A, Degli Esposti D, Cuenin C, Cahais V, et al. Genome-wide profiling of normal gastric mucosa identifies Helicobacter pylori-and cancer-associated DNA methylome changes. Int J Cancer. 2018;143(3):597-609.
60. Ahn Y, Kwon E, Shim J, Park M, Joo Y, Kimm K, et al. Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. Eur J Clin Nutr. 2007;61(12):1435.
61. http://qiime.org/. Cited June 2020.
62. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67(5):1613-7.
63. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60.
64. http://huttenhower.org/galaxy. Cited June 2020.

65. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell host Microbe. 2014;15(3):382-92.
66. http://picrust.github.io/picrust/. Cited June 2020.
67. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnol. 2013;31(9):814.
68. https://www.genome.jp/kegg/. Cited June 2020.

Figures

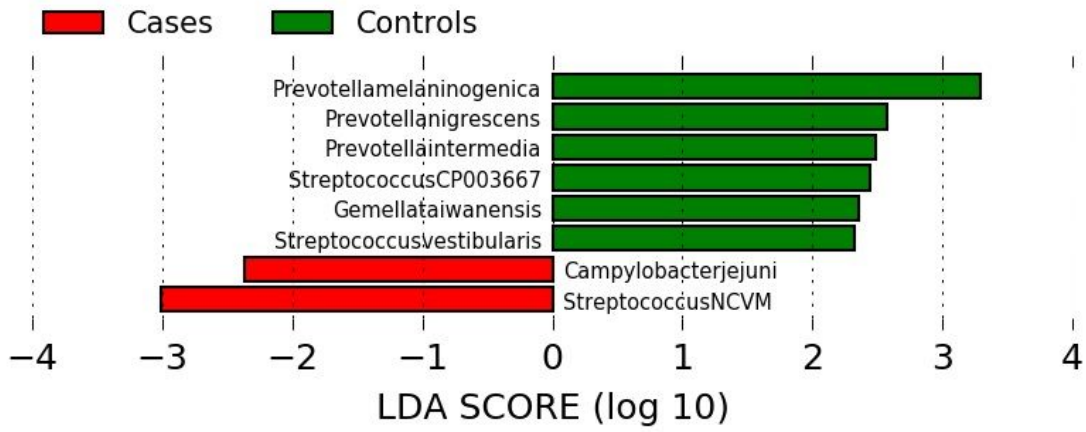

Figure 1

Linear discriminant analysis of effect size (LEfSe) analysis plot for the taxonomy species.

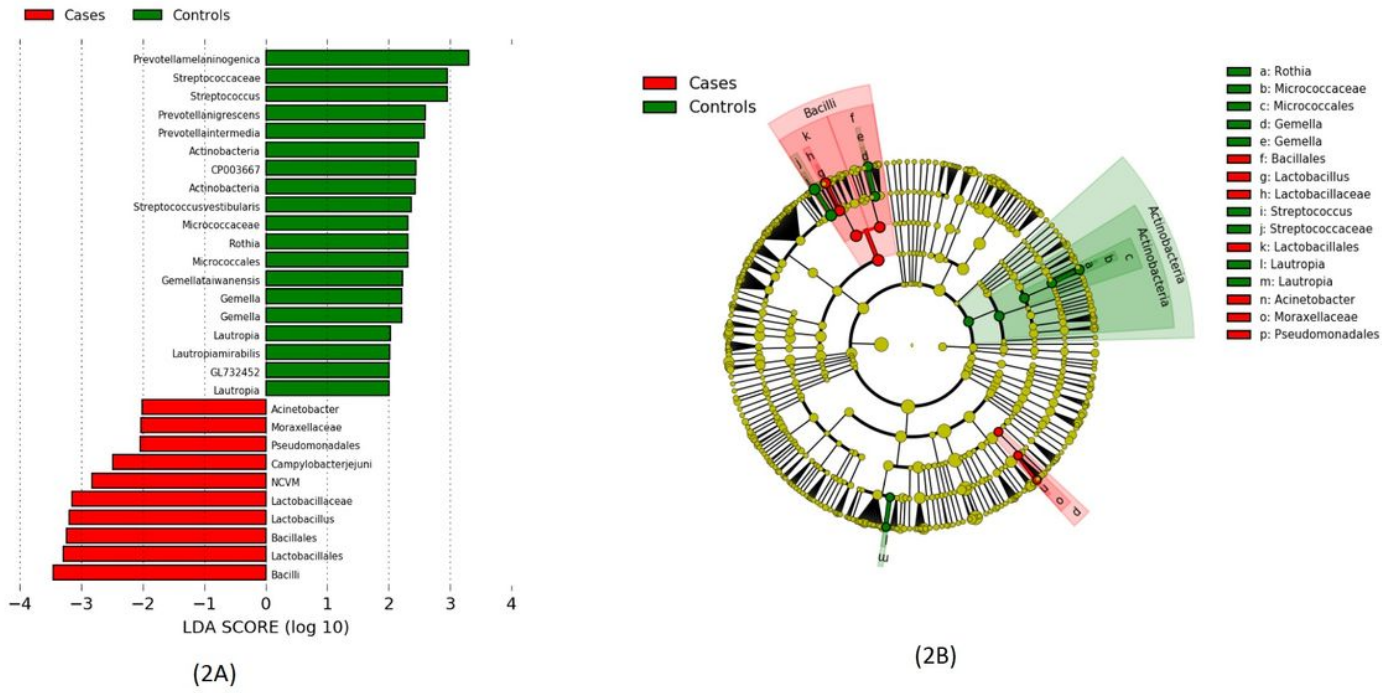

Figure 2

A. Linear discriminant analysis of effect size (LEfSe) analysis plot of the gastric microbial taxa differentially abundant between GC and control groups. B. Cladogram representation of gastric microbiome taxa associated with GC based on Ezbio database.

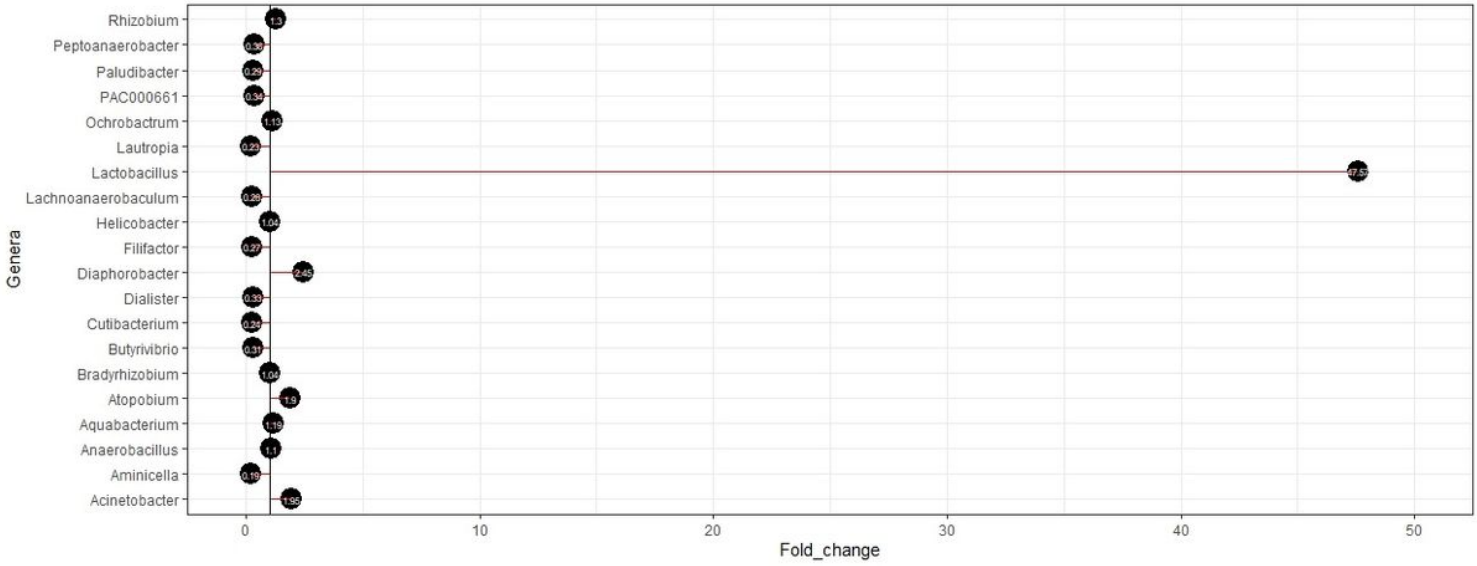

Figure 3

Diverging Lollipop chart for differences in abundances are shown for genera that were detected using compositionality corrected by renormalization and permutation (CCREPE). Fold change for each genus was calculated by dividing the mean abundance in the cases by that of the controls.

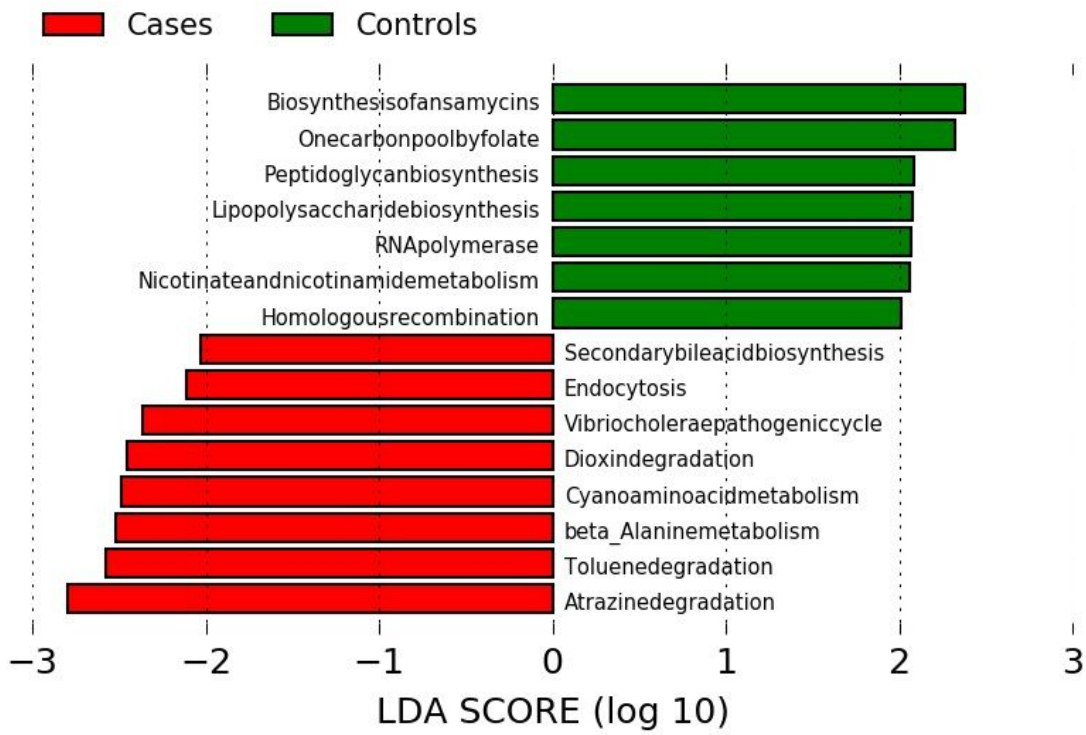

Figure 4

Comparison of the microbial functional pathways between cases and controls.

Supplementary Files

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

- Supplementaryfigures.zip
- Additionalfiles.docx