

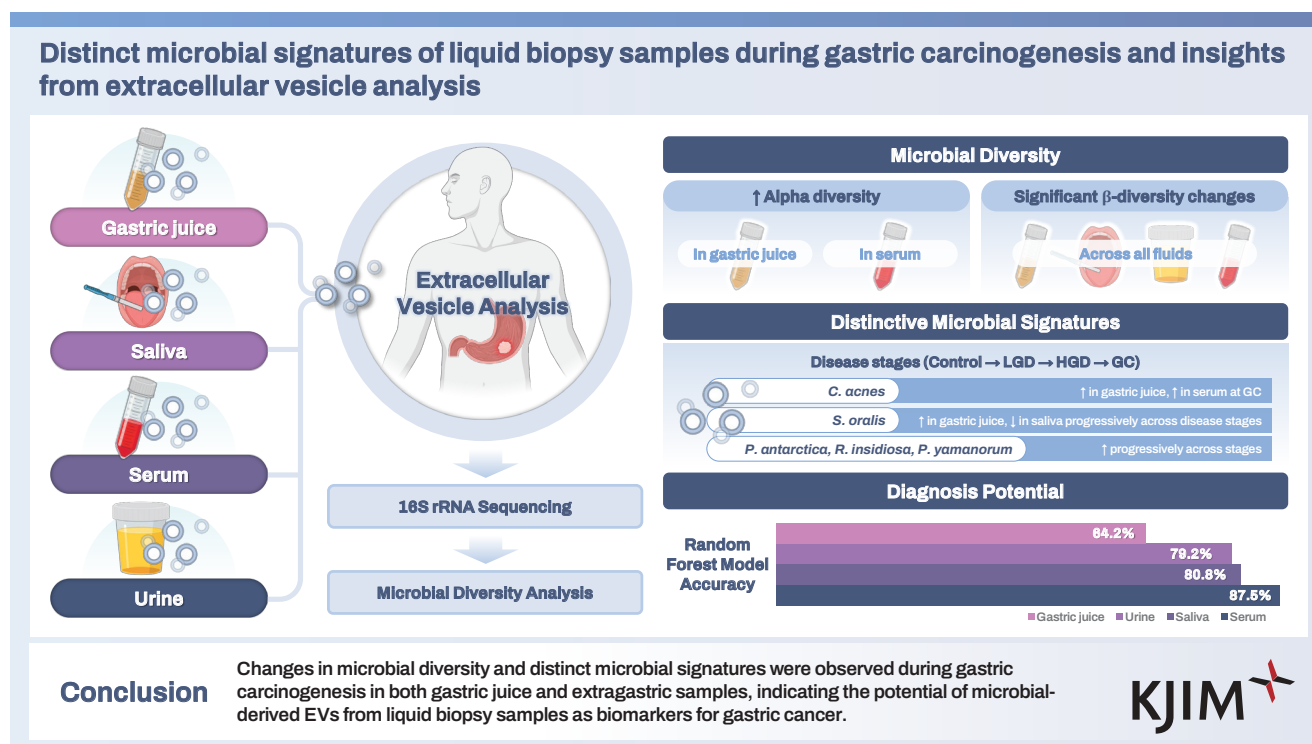


Distinct microbial signatures of liquid biopsy samples during gastric carcinogenesis and insights from extracellular vesicle analysis

Hee Sang You^{1,2,*}, Jae Yong Park^{1,*}, Hochan Seo^{1,2}, Beom Jin Kim¹, and Jae Gyu Kim¹

¹Department of Internal Medicine, Chung-Ang University College of Medicine, Seoul; ²Laboratory of Gastrointestinal Mucosal Immunology, Chung-Ang University College of Medicine, Seoul, Korea

*These authors contributed equally to this manuscript.



Background/Aims: The early detection of gastric cancer is crucial for improving patient outcomes. However, its pathogenesis is not fully understood. The microbiome and extracellular vesicles (EVs) might play a role in gastric carcinogenesis. We aimed to identify gastric-carcinogenesis-associated microbial signatures and evaluate whether these features vary across disease stages.

Methods: We enrolled 141 participants (132 patients with gastric cancer or dysplasia and 9 healthy controls). Microbial-derived EVs were isolated from gastric juice, saliva, serum, and urine. Next-generation sequencing of EV-derived bacterial DNA was performed.

Results: This sequencing revealed the alpha and beta diversities and microbial composition across different disease stages. The alpha diversity was significantly increased in the gastric juice and serum of disease groups. The beta diversity showed significant differences among patient groups. Distinct microbial signatures were observed across different disease stages in all four sample types. Specific bacterial species—*Cutibacterium acnes*, *Streptococcus oralis*, *Pseudomonas antarctica*, *Ralstonia insidiosa*, and *Pseudomonas yamanorum*—exhibited unique abundance patterns associated with disease progression, suggesting their potential as noninvasive biomarkers.

Conclusions: Changes in microbial diversity and distinct microbial signatures were observed during gastric carcinogenesis in both gastric juice and extragastric samples, indicating the potential of microbial-derived EVs from liquid biopsy samples as biomarkers for gastric cancer.

Keywords: Microbiome; Extracellular vesicles; Liquid biopsy; Stomach neoplasms; Biomarkers

INTRODUCTION

Gastric cancer (GC) remains one of the most prevalent and lethal malignancies worldwide and the fifth leading cause of cancer-related deaths [1]. Despite advancements in treatment, the prognosis of patients with GC remains poor, particularly for those diagnosed at advanced stages [2]. Early detection of GC, ideally during its premalignant stages, is critical for improving the prognosis and reducing the global burden of the disease. Although endoscopic inspection remains the mainstay of screening and diagnosis for GC, this method is invasive, causes patient discomfort, and increases healthcare costs [3]. However, current non-invasive screening and diagnostic methods for GC lack the sensitivity and specificity necessary for early stage detection, underscoring the need for novel biomarkers and innovative approaches [4].

Emerging evidence suggests that the microbiome plays a crucial role in the pathogenesis of GC [5]. *Helicobacter pylori* is a well-established and unmatched risk factor implicated in the development and progression of GC. Recent research suggests that alterations in the gastric microbiome other than *H. pylori* may play a role in disease progression from chronic gastritis to atrophic gastritis, intestinal metaplasia, dysplasia, and eventually carcinoma [6]. These findings highlight the potential of microbial signatures as biomarkers for early detection of GC.

Microbial-derived extracellular vesicles (EVs), which are nanosized particles secreted by various microorganisms, represent a promising frontier for exploring the role of the microbiome in cancer. These vesicles carry proteins, lipids, nucleic acids, and other molecules that reflect the biological

state of their parent cells, and play a significant role in cell signaling by acting as carriers of molecular information between cells [7]. Importantly, microbial-derived EVs are present in various biological fluids, such as saliva, serum, urine, and gastric juice, making them accessible targets for noninvasive diagnostic testing. EVs have attracted significant interest as potential biomarkers for various diseases, including cancer, owing to their ability to protect and transport microbial components in a stable form [8].

In this study, we analyzed microbially derived EVs from various liquid biopsy samples obtained from patients at different stages of gastric carcinogenesis. We aimed to explore the microbial features associated with GC using microbial-derived EVs and to assess whether these features vary across disease stages. Additionally, this study aimed to deepen our understanding of the role of microbes in GC progression and potentially inform the development of future biomarkers for disease detection and staging.

METHODS

Study participants and sample collection

Participants were prospectively recruited at the Chung-Ang University Hospital between December 2016 and October 2022. The inclusion criteria for the neoplasm group were newly diagnosed patients with histologically confirmed GC or adenoma. The control group included patients with only gastritis and no evidence of gastric neoplasms on endoscopic examination. The exclusion criteria included individuals with a history of malignancy (including GC), those who had undergone gastric surgery, those who had taken

antibiotics or probiotics within the past three months, or those under 20 years of age. The final cohort comprised 141 participants, including 132 patients and nine controls (100 male and 41 female). This study was approved by the Institutional Review Board (IRB) of the Chung-Ang University Hospital (IRB No. C2016047(1790)) and adhered to the Declaration of Helsinki. The informed consent was provided in written form by all participants. Gastric neoplasms were diagnosed and categorized using the Vienna Classification System [9]. Specifically, category 3 (non-invasive low-grade adenoma/dysplasia) was classified as low-grade dysplasia (LG), category 4.1 (high-grade adenoma/dysplasia) and 4.2 (non-invasive carcinoma [carcinoma in situ]) were classified as high-grade dysplasia (HG), and categories 4.3 (suspicion of invasive carcinoma) and 5 (invasive neoplasia) were classified as GC.

The following protocols were used for sample collection. For saliva samples, participants were instructed to swish 10 mL of saliva in their mouths and then expectorate into a specimen tube, which was stored at -80°C until use. Gastric juice samples were collected following a specific protocol. Patients fasted for > 8 hours prior to collection. For those undergoing endoscopic examination or resection, a trap tube was attached between the endoscope and the suction tube. Gastric juice (7–30 mL) was suctioned through the endoscope at the beginning of the procedure. Simultaneously, tissue samples from the antrum and body were collected and stored at -80°C . For patients undergoing surgical gastrectomy, gastric juice and tissue samples were obtained either endoscopically before surgery or during surgery, immediately after making an incision. The endoscopes were washed and disinfected by immersion in a detergent solution containing 7% proteolytic enzyme and 2% glutaraldehyde to prevent contamination. Gastric juice samples were immediately stored at -20°C and transported to the laboratory without preservative reagents. The gastric juice was then diluted to 40 mM with 1 M Tris base and centrifuged at $600 \times g$ for 10 minutes at 4°C to obtain a supernatant, which was further centrifuged at $1,500 \times g$ for 10 minutes at 4°C . Finally, the supernatant was collected and stored at -80°C .

DNA extraction and sequencing

Bacterial EVs were isolated from the gastric juice, saliva, serum, and urine of the individuals following a previously described procedure [10,11]. Each gastric juice, saliva, and

urine sample was centrifuged at $10,000 \times g$ for 10 minutes at 4°C . The supernatant was taken and passed through a $0.22 \mu\text{m}$ membrane filter to eliminate foreign particles and then quantified based on protein concentration [12,13]. For serum, after mixing 100 μL of serum and 900 μL PBS, it was centrifuged at $10,000 \times g$ for 10 minutes at 4°C to eliminate other components. The supernatant was taken and passed through a $0.22 \mu\text{m}$ membrane filter to eliminate foreign particles, and it was quantified based on protein concentration.

Bacterial DNA was extracted from the prepared EVs as previously described [11,14]. Isolated EVs (1 μg by protein, each sample) were boiled at 100°C for 40 minutes and centrifuged at $13,000 \times g$ for 30 minutes, and the supernatants were collected. The collected samples were then subjected to bacterial DNA extraction using a DNA extraction kit (PowerSoil DNA Isolation Kit; MO BIO, Carlsbad, CA, USA) following the manufacturer's instructions. The isolated DNA was quantified using the QIAxpert system (QIAGEN, Hilden, Germany). Microbial genomic DNA was extracted using a DNeasy PowerSoil kit (QIAGEN) according to the standard protocol provided in the manufacturer's instructions. The isolated genomic DNA was amplified by targeting the V3–V4 hypervariable regions of the 16S rRNA gene. This amplification was carried out using specific primers (16S_V3_F: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGG-NGGCWGCAG-3' and 16S_V4_R: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'). Amplicon libraries were prepared from the amplified DNA fragments. The MiSeq Reagent Kit v3 (600-cycle) (Illumina, San Diego, CA, USA) was used to sequence the library reagents. The Nextera XT Index Kit v2 Set A (96 indices, 384 samples) (Illumina) was used for barcodes and adapters. Library preparation for sequencing followed the 16S Metagenomic Sequencing Library Preparation (Part # 15,044,223 Rev. B). All amplicons were sequenced using MiSeq (Illumina) according to the manufacturer's instructions. The bacterial DNA in each sample was quantified using a QIAxpert system.

Taxonomic classification and microbial profiling

Paired-end 16S rRNA gene sequences were processed using the Quantitative Insights into Microbial Ecology (QIIME2 v2021.4) [15]. Adapter sequences were trimmed using Cutadapt, and the reads were subsequently filtered for quality and the presence of chimeras using DADA2, with specific

parameters set as follows: trim-left-f 0, trim-left-r 0, trunc-len-f 260, trunc-len-r 200, trunc-q 2, max-ee-f 3, and max-ee-r 3 [16]. A naïve Bayes classifier, trained on the V3–V4 region sequences from the SILVA 138 database, was used for taxonomic classification. Sequences identified as chloroplasts or mitochondria were excluded from the analysis.

Statistical analysis

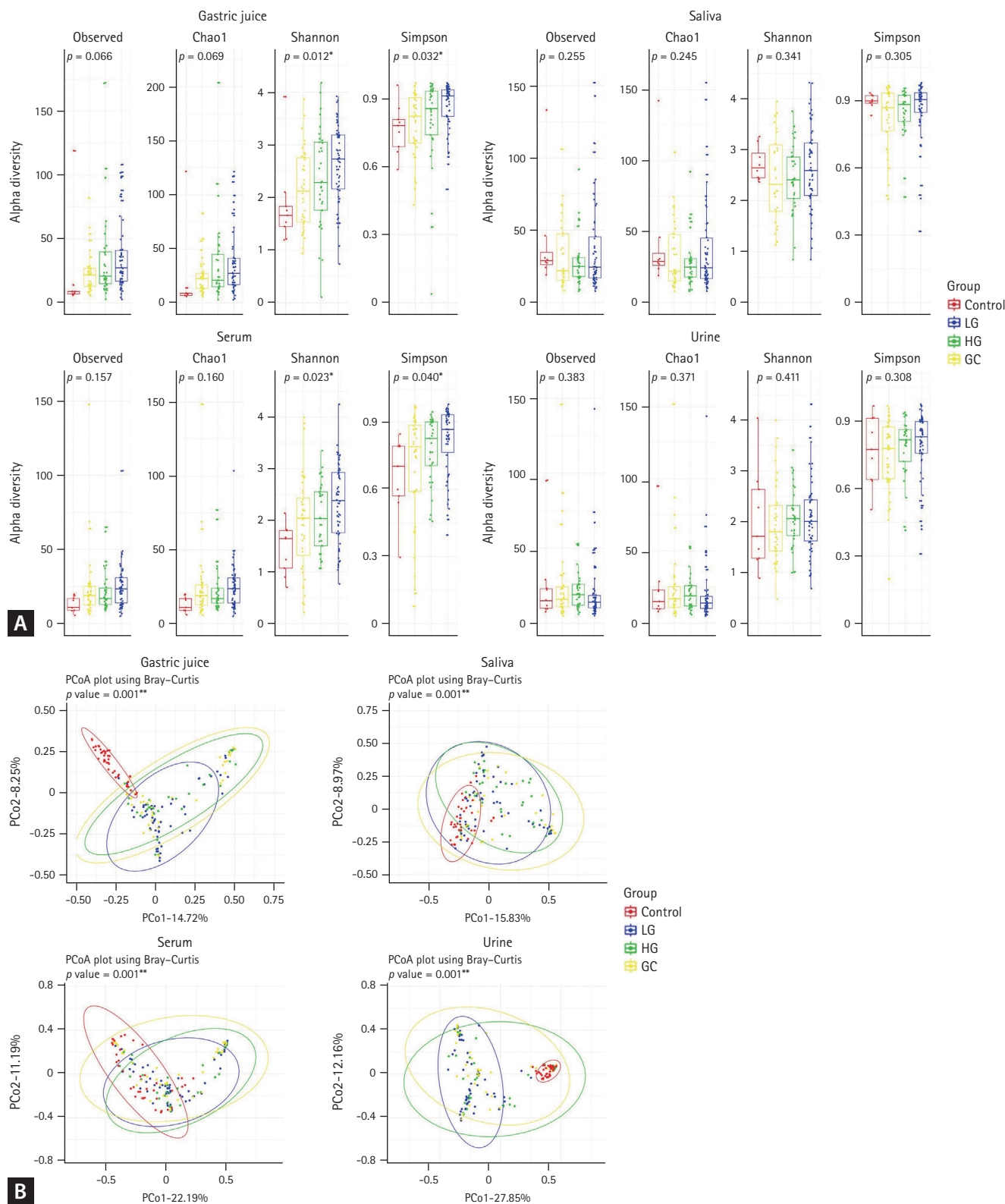
For alpha diversity analysis, the samples were rarefied to a minimum read count of 2,386 to normalize the data. The *p* values for alpha diversity were calculated using analysis of variance adjusted for age and sex and were further adjusted using the bootstrap method [17]. Principal coordinate analysis (PCoA) was performed to evaluate the taxon-level clustering of groups based on Bray–Curtis dissimilarity distances. The synthetic minority oversampling technique (SMOTE) algorithm was applied to increase the normal sample size for PCoA [17]. The *p* value for the PCoA was determined through permutational multivariate analysis of variance using dissimilarity matrices adjusted for age and sex. For the analysis of relative abundances and Multivariable Associations with Linear Models2 (MaAsLin2), operational taxonomic units (OTUs) that exhibited a low count across the dataset were filtered out; specifically, OTUs with a minimum count of less than or equal to four were removed [18]. Furthermore, samples were retained only if they contained OTUs that surpassed the minimum count threshold in more than 20% of the dataset, thereby ensuring the exclusion of

samples with sparse OTU representation. To address the low variance, which often indicates a lack of differential abundance across conditions, we implemented a filtering process based on the interquartile range. Features within the lowest 10% of the interquartile range were discarded. Following the filtering process, data normalization was performed to mitigate the effects of varying sequencing depths across the samples. Total sum scaling was used as the normalization technique, where the read count of each sample was divided by the total read count of the sample to ensure comparability across the dataset. The MaAsLin2 method was used to evaluate differences in the OTU counts of bacteria-derived EVs among groups. This analysis was performed using a negative binomial model for group comparisons. Analysis was conducted using MaAsLin2 on MicrobiomeAnalyst (<https://www.microbiomeanalyst.ca/>) and the MaAsLin2 package in R adjusting for age and sex as covariates. Statistical significance was set at a *q*-value of less than 0.05. MaAsLin2 was executed using the default parameters in conjunction with the Benjamini–Hochberg false discovery rate method [19]. To validate and complement the results of MaAsLin2, a random forest classifier analysis was performed. To balance the sample sizes across groups, SMOTE oversampling was applied, increasing the sample size of each group (control, LG, HG, and GC) to 100 samples. A random forest classifier was subsequently employed to distinguish between the disease groups. The dataset was split into a training set (70%) and a test set (30%) to rigorously assess the model's predictive ac-

Table 1. Baseline characteristics of participants in each sample group

Sample type	Variable	Total	Control	Low grade dysplasia	High grade dysplasia	Gastric cancer
Gastric juice	Number	141	9	58	33	41
	Age	64.4 ± 11.9	33.2 ± 11.4	66.2 ± 8.3	66.4 ± 8.0	67.0 ± 9.1
	Sex, male	96 (68.1)	3 (33.3)	42 (72.4)	24 (72.7)	27 (65.9)
Saliva	Number	129	8	58	33	30
	Age	64.2 ± 11.3	34.9 ± 11.0	66.2 ± 8.3	66.5 ± 7.9	65.5 ± 8.4
	Sex, male	96 (74.4)	3 (37.5)	42 (72.4)	24 (72.7)	27 (90.0)
Serum	Number	140	9	58	33	40
	Age	64.3 ± 11.9	33.2 ± 11.4	66.2 ± 8.3	66.4 ± 8.0	66.7 ± 9.2
	Sex, male	99 (70.7)	3 (33.3)	42 (72.4)	24 (72.7)	30 (75.0)
Urine	Number	140	9	58	33	40
	Age	64.3 ± 11.9	33.2 ± 11.4	66.2 ± 8.3	66.4 ± 8.0	66.7 ± 9.2
	Sex, male	99 (70.7)	3 (33.3)	42 (72.4)	24 (72.7)	30 (75.0)

Values are presented as mean ± standard deviation or number (%).



curacy. For feature selection, microbial biomarkers identified using MaAsLin2 were incorporated into the random forest model. These biomarkers were selected based on their statistically significant associations with the disease groups, serving as key input variables to improve the classifier's in-

terpretability and overall performance. All statistical analyses were performed in R Software (version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria).

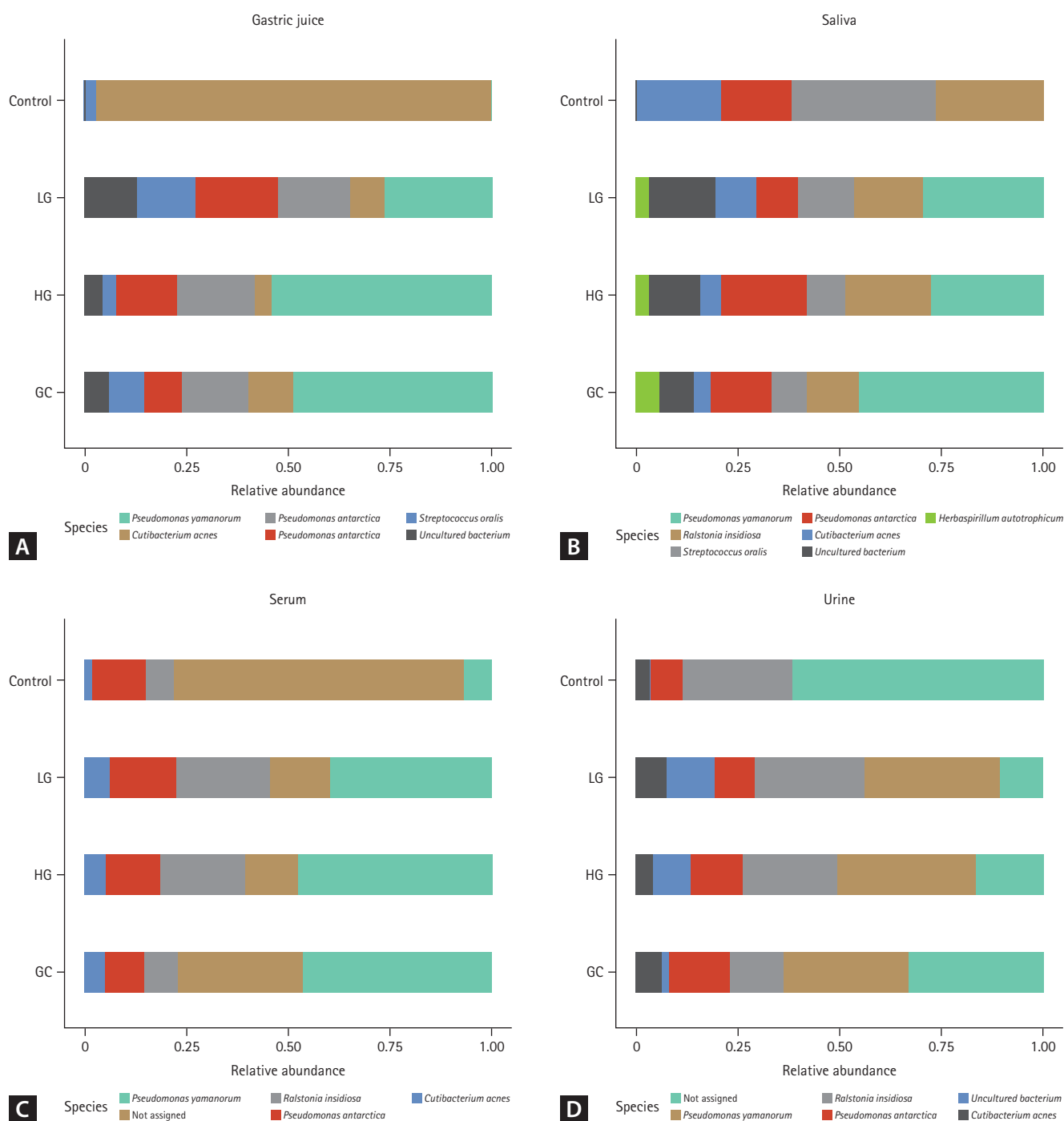


Figure 2. Microbial composition analysis across disease stages for gastric juice, saliva, serum and urine samples. Changes in microbial composition at the species level during gastric carcinogenesis in gastric juice (A), saliva (B), serum (C), and urine (D). LG, low-grade dysplasia; HG, high-grade dysplasia; GC, gastric cancer.

RESULTS

Clinical characteristics of the study population

A total of 141 participants were included in the study, including nine controls, 58 with LG, 33 with HG, and 41 with GC. We evaluated samples from four sources: gastric juice, saliva, serum, and urine. A total of 550 samples of all types and conditions were analyzed. The mean age of the study participants was 64.4 years, with males accounting for 70.9% (100 out of 141) of the study population. Table 1

summarizes the clinical characteristics of the enrolled patients, categorized by sample type.

Diversity of microbial communities in EV samples

The alpha diversity was assessed in the gastric juice, saliva, serum, and urine samples using various indices (Fig. 1A). In gastric juice, both the Shannon ($p = 0.012$) and Simpson ($p = 0.032$) indices showed significant changes in microbial diversity and species dominance with the progression of the

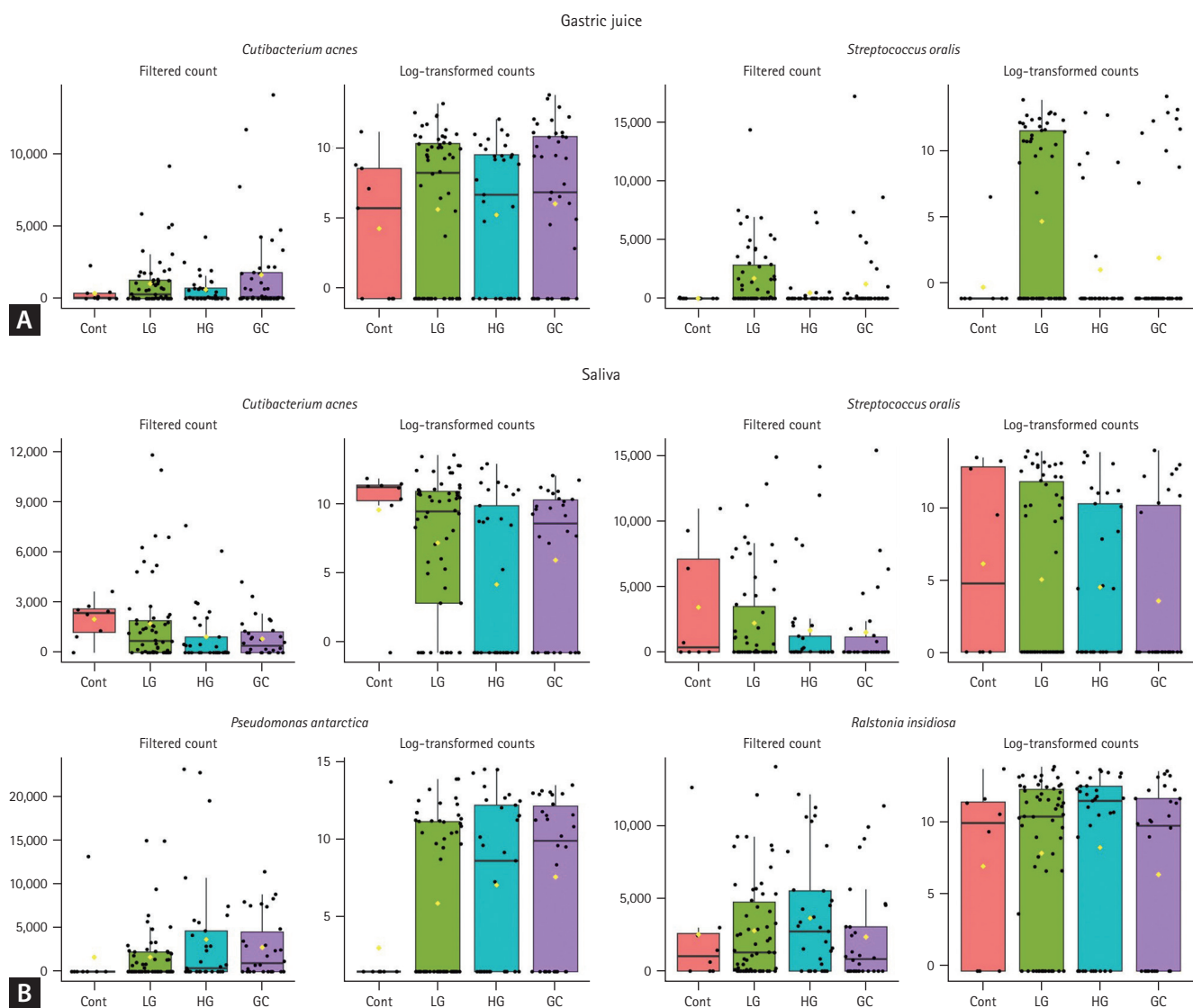


Figure 3. Multivariate Association with Linear Models 2 (MaAsLin2) analysis across GC stages for gastric juice, saliva, serum and urine samples. MaAsLin2 was used to identify the specific bacterial strains at the species level in gastric juice (A), saliva (B), serum (C), and urine (D) across various stages of GC development. Strains were identified through MaAsLin2 analysis of filtered counts and assessed using the non-parametric Kruskal–Wallis test with a significance threshold of a false discovery rate-adjusted p value (q -value) below 0.05. The control group is marked in orange, the LG group in green, the HG group in blue, and the GC groups in purple. Strains with a q -value of less than 0.05 are indicated. LG, low-grade dysplasia; HG, high-grade dysplasia; GC, gastric cancer.

pathological condition. Similarly, in the serum, significant changes in microbial diversity were observed in Shannon ($p = 0.023$) and Simpson ($p = 0.040$) indices as the disease progressed. In contrast, the saliva and urine samples did not show statistically significant differences in microbial diversity across the different stages of gastric carcinogenesis.

Beta diversity was assessed across different stages of gastric carcinogenesis in all sample types. The results showed significant differences in microbial composition across all sample types during disease progression ($p < 0.05$, Fig. 1B), with notable differences in gastric juice and urine ($p < 0.001$, Fig. 1B).

Alterations in microbiota across different disease stages in various samples

The relative abundances of bacterial communities in various samples (gastric juice, saliva, serum, and urine) were analyzed proportionally, as shown in Figure 2. The bar chart represents the relative abundances of bacterial species identified in the four groups: control, LG, HG, and GC.

In gastric juice samples (Fig. 2A), *Cutibacterium acnes* dominated in the control group, whereas in the LG, HG, and GC groups, there was a decrease in the abundance of *C. acnes* and an increase in the abundances of *Pseudomonas yamanorum*, *Pseudomonas antarctica*, *Ralstonia insidiosa*, and *Streptococcus oralis*. For saliva samples (Fig. 2B),

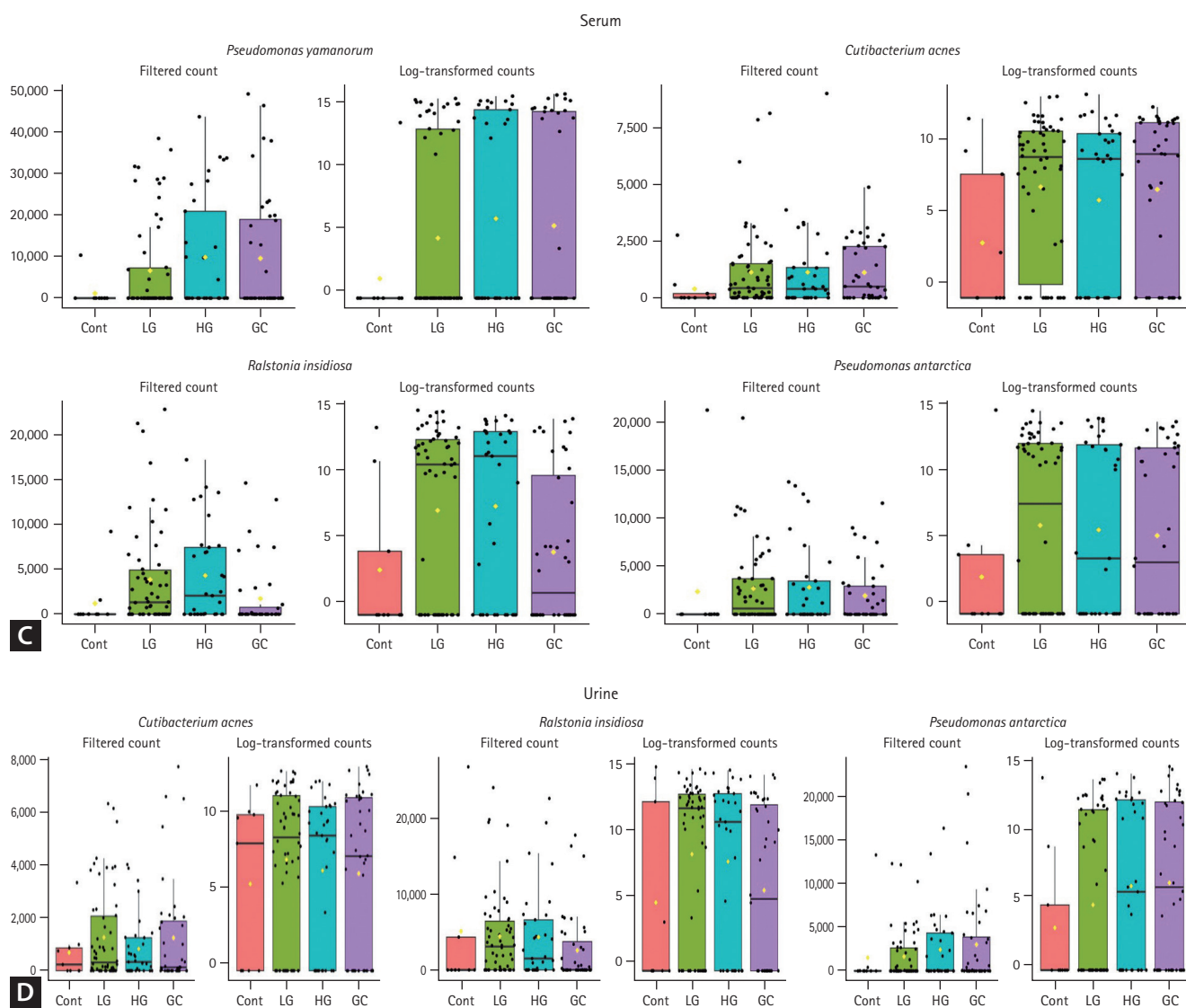


Figure 3. Continued

compared with the controls, the LG, HG, and GC groups showed a decrease in *C. acnes* and *S. oralis* abundances, along with an increase in *P. yamanorum* abundance. Notably, *Herbaspirillum autotrophicum* was detected exclusively in the disease groups (LG, HG, and GC). In the serum samples (Fig. 2C), compared to the controls, the LG, HG, and GC groups showed an increase in *C. acnes* and *R. insidiosa* abundances. Urine samples (Fig. 2D) revealed a heterogeneous bacterial composition. However, *P. yamanorum* was consistently prominent across the disease groups (LG, HG, and GC), indicating a noticeable increase compared to that in the control group.

Differential bacterial abundance in biological samples across disease states

Specific bacterial strains with significantly different filtered OTU counts in the disease groups (LG, HG, and GC) compared to those in the control group were identified. Figure 3 shows a comparative analysis of the abundances of these bacterial strains across different patient groups for various sample types. Bacterial abundance was represented by filtered counts to assess the correlations between groups. In the gastric juice, the abundance of *C. acnes* and *S. oralis* displayed a notable increase with the progression of disease severity ($q < 0.01$ each, Fig. 3A). In saliva, the levels of *C. acnes* and *S. oralis* significantly decreased as the disease progressed, whereas *P. antarctica* and *R. insidiosa* increased ($q < 0.01$ all, Fig. 3B). In the serum, the abundances of *P. yamanorum*, *C. acnes*, *R. insidiosa*, and *P. antarctica* significantly increased with advancing disease stage ($q < 0.01$ all, Fig. 3C). Similarly, in urine, *C. acnes*, *R. insidiosa*, and *P. antarctica* showed significantly increasing trends during disease progression ($q < 0.01$ each, Fig. 3D). To complement the MaAsLin2 results and further assess the discriminative power of these identified bacterial biomarkers across disease stages, the random forest classifier was employed. This analysis evaluated the distribution and classification potential of specific microorganisms associated with different stages of gastric carcinogenesis in each sample type (Supplementary Tables 1-4). The overall classification accuracy was 64.2% for gastric juice and 80.8% for saliva, while serum and urine samples demonstrated classification accuracies of 87.5% and 79.2%, respectively. Meanwhile, it is well known that differences in microbiome diversity and composition are greatly influenced by *H. pylori* infection status. To assess the impact of *H. pylori*, all analyses were re-evaluated

after excluding samples from *H. pylori*-positive participants. The overall trends and results remained consistent (data not shown).

DISCUSSION

This study aimed to investigate shifts in microbial communities across different stages of gastric carcinogenesis by analyzing microbial-derived EVs in various biological samples, including gastric juice, saliva, serum, and urine. We observed significant changes in microbial diversity and composition during gastric carcinogenesis, which were effectively captured through metagenomic analysis of samples obtained from liquid biopsy.

Our results demonstrated significant changes in the microbial diversity in gastric juice and serum samples as the disease progressed. The increased alpha diversity observed in gastric juice suggests greater microbial variation and shifts in species dominance associated with GC development. These findings are consistent with those of previous studies that highlighted the dynamic nature of the gastric microbiome in response to carcinogenic changes in the gastric environment [5]. Similar patterns of increased microbial diversity were observed in the serum samples, which may reflect systemic changes in the microbiome associated with advanced cancer stages. Conversely, saliva and urine samples did not show statistically significant changes in alpha diversity, suggesting that microbial alterations associated with GC may be more localized in the gastric environment and systemic circulation than in other bodily fluids. Although the alpha diversity in saliva remained relatively stable, significant changes were observed in gastric juice, which might reflect the selective pressure of the gastric environment during gastric carcinogenesis. These changes were particularly pronounced in the beta diversity analysis, which showed significant differences in the microbial community composition across all sample types during gastric carcinogenesis. These results support the hypothesis that GC progression is associated with specific shifts in microbial populations [20]. This finding aligns with emerging evidence suggesting that changes in the microbial community structure could serve as potential indicators of disease status and progression [21].

We further analyzed the microbial composition and abundance and identified distinct microbial signatures across different liquid biopsy samples during carcinogenesis. In gastric

juice, the abundance of *C. acnes* and *S. oralis* significantly increased with disease progression, suggesting that these bacteria might be involved in the pathogenesis or progression of GC or that they may thrive in the altered gastric environment associated with the disease. In contrast, the abundances of *C. acnes* and *S. oralis* decreased in the saliva as disease severity increased. This opposing pattern may indicate different microbial dynamics in the oral and gastric environments during GC progression. *C. acnes*, *P. antarctica*, and *R. insidiosa* also increased in both the serum and urine, with the latter two species displaying a similar increasing trend in the saliva. Meanwhile, *P. yamanorum* exhibited a notable increase, specifically in the serum. These common microbial signatures across different samples may indicate the importance of these species in the process of carcinogenesis and their association with changes in the microbial milieu of the stomach. This consistency in microbial presence across various liquid biopsy samples suggests the possible systemic circulation of microbial-derived EVs, as proposed in previous studies [22]. These variations in microbial populations align with a growing body of evidence suggesting a complex interaction between the microbiota and cancer cells [23]. Furthermore, these findings provide a basis for developing diagnostic approaches using extragastric liquid biopsy samples such as serum and urine to detect microbial changes associated with GC. In addition, we conducted additional analysis using a random forest classifier to evaluate GC stage-specific microorganisms in liquid biopsy samples (gastric juice, saliva, serum, and urine). The results showed relatively high classification accuracies for serum (87.5%), saliva (80.8%), and urine (79.2%) samples, whereas the classification accuracy for gastric juice samples was lower at 64.2%. These findings further support the potential of specific microorganisms as non-invasive biomarkers for predicting GC stages, particularly highlighting the utility of EVs from liquid biopsy samples such as serum and saliva as diagnostic tools. However, additional data supplementation and improvements in analytical algorithms are necessary to achieve clearer stage differentiation across all sample types.

This study utilized EV-based microbiota analysis to uncover novel microbial dynamics associated with gastric carcinogenesis. Unlike traditional approaches analyzing total microbial DNA from gastric juice or saliva, EV-based analysis focuses on vesicle-encapsulated genetic material, ensuring higher specificity and functional relevance [22,24]. EVs, bacterial-derived structures containing genetic mate-

rial, proteins, and metabolites, protect microbial information from degradation and facilitate intercellular communication, even in hostile conditions [25,26]. This makes EVs ideal for non-invasive biomarker analysis in liquid biopsies [26]. Serum EVs reflect systemic microbial shifts in advanced disease states, while urine EVs offer a stable medium for profiling localized and systemic changes. By encapsulating active microbial components, EVs highlight functional roles in host-pathogen interactions, adding mechanistic insights to microbiota profiling [24,26]. EV-based methods surpass direct 16S rRNA sequencing by capturing disease-relevant signals with greater precision, particularly from pathogenic microorganisms [22,24,26]. These features underscore the potential of EVs as biomarkers and mediators of gastric carcinogenesis, paving the way for future research and clinical applications [22,26].

Although not well established, we explored the potential roles of these bacteria in gastric pathology through a literature review. Mannosylerythritol lipids secreted by *P. antarctica* inhibit inflammatory mediators and produce anti-inflammatory effects, which could either contribute to or result from the pathophysiological changes in cancer [27]. Although *C. acnes* is traditionally linked to skin conditions, its increased abundance in GC raises questions regarding its systemic role, potentially as an agent of inflammation or as a response to the tumor microenvironment [28-30]. Similarly, *S. oralis* may have implications in GC progression, given its oncogenic potential in the oral cavity and its increased presence in gastric juice with disease progression [31,32]. The systemic increase in *P. antarctica*, *R. insidiosa*, and *P. yamanorum* highlights the potential of these microbes as biomarkers for GC and its precursors, including dysplasia, although their roles in the disease process warrant further exploration [33-36].

Efforts to understand the pathogenesis of GC and to enhance diagnostic methods have greatly benefited from new perspectives and insights gained through extensive research on various biofluids from patients at different stages of the disease [4,37]. Notably, the liquid biopsy approach, which utilizes next-generation sequencing, is a noninvasive alternative that may complement conventional endoscopic examination [38]. Our study extends the well-known association between *H. pylori* infection and GC by identifying additional bacterial species that could be implicated in the broader dysbiotic shift that accompanies or potentially contributes to cancer development. We highlight the potential

of microbial-derived EVs and specific microbial signatures, such as *C. acnes*, *S. oralis*, *P. antarctica*, *R. insidiosa*, and *P. yamanorum*, as non-invasive biomarkers for the early detection and monitoring of GC. By analyzing these microbial signatures across various liquid biopsy samples, including gastric juice and extragastric fluids, our study highlights a novel and impactful approach for developing diagnostic tools for GC.

Despite these strengths, some limitations must be acknowledged. Firstly, the control group was relatively small and comprised younger individuals compared to the disease group because efforts to include healthy controls with normal gastric conditions, matched to the age range of the patient groups, were challenging. Although we attempted to mitigate these differences through statistical adjustments for age and sex (Bootstrap, SMOTE, MaAsLin2), smaller and less diverse control groups may have affected the generalizability of our findings. Secondly, while this study demonstrates significant associations between microbial diversity and composition and disease progression in gastric carcinogenesis, it does not establish a direct causal relationship between these factors. The observed differences in microbial diversity and composition may either influence disease progression or result from it, and establishing causality requires further investigation. Experimental models and multi-omics approaches are essential for understanding the role of microbial shifts in gastric carcinogenesis, potentially guiding diagnostic and therapeutic advancements to improve patient outcomes. Thirdly, one critical factor influencing microbial diversity and composition is *H. pylori* infection, which is well-known to cause dysbiosis and impact gastric disease progression, including cancer [39]. In this study, the low prevalence of *H. pylori* infection among participants suggests a limited impact on the overall results. To validate our findings, we performed additional analyses by excluding samples from *H. pylori*-infected individuals. Minor changes were observed, but the primary outcomes remained consistent, suggesting that our findings are not substantially influenced by *H. pylori* infection status. Lastly, although we identified distinct microbial signatures associated with gastric carcinogenesis, this study did not include functional analyses or an assessment of diagnostic capabilities. Future research is required to explore the functional roles of these microbial signatures in EV-mediated communication pathways and evaluate their diagnostic potential. To strengthen the validity and applicability of the identified

biomarkers, future studies should involve larger and more diverse cohorts to establish the robustness and generalizability of the findings.

In conclusion, our study revealed significant changes in the microbial diversity and distinct microbial signatures during gastric carcinogenesis, which were consistently observed in both gastric juice and extragastric liquid biopsy samples. These findings underscore the potential of microbial derived EVs as early diagnostic tools for GC and its precursors. Our study provides a valuable foundation for future studies aimed at utilizing these microbial signatures to enhance early detection and improve diagnostic accuracy. Although the development of non-invasive biomarkers for early GC detection presents challenges, our study provides important insights and highlights the need for continued research to establish reliable and broadly applicable diagnostic tools.

KEY MESSAGE

1. Microbial diversity, particularly alpha and beta diversities, varies across different stages of gastric carcinogenesis, with notable changes in gastric juice and serum.
2. Distinct microbial signatures, including specific bacterial species like *C. acnes*, *S. oralis*, and *P. antarctica*, show unique abundance patterns associated with disease progression.
3. Microbial-derived EVs from liquid biopsy samples, such as gastric juice and extragastric fluids, have the potential to serve as noninvasive biomarkers for the early detection of GC.

REFERENCES

1. Ferlay J, Ervik M, Lam F, et al. Global cancer observatory: cancer today (version 1.0) [Internet]. Lyon (FR): International Agency for Research on Cancer, c2024 [cited 2024 Aug 20]. Available from: <https://gco.iarc.who.int/today/en/fact-sheets-cancers>.
2. Park SH, Kang MJ, Yun EH, Jung KW. Epidemiology of gastric cancer in Korea: trends in incidence and survival based on Korea Central Cancer Registry Data (1999-2019). *J Gastric Cancer* 2022;22:160-168.
3. Haque E, Esmail A, Muhsen I, Salah H, Abdelrahim M. Recent

- trends and advancements in the diagnosis and management of gastric cancer. *Cancers (Basel)* 2022;14:5615.
4. Conti CB, Agnesi S, Scaravaglio M, et al. Early gastric cancer: update on prevention, diagnosis and treatment. *Int J Environ Res Public Health* 2023;20:2149.
5. Yang J, Zhou X, Liu X, Ling Z, Ji F. Role of the gastric microbiome in gastric cancer: from carcinogenesis to treatment. *Front Microbiol* 2021;12:641322.
6. Park JY, Seo H, Kang CS, et al. Dysbiotic change in gastric microbiome and its functional implication in gastric carcinogenesis. *Sci Rep* 2022;12:4285.
7. Yang J, Shin TS, Kim JS, Jee YK, Kim YK. A new horizon of precision medicine: combination of the microbiome and extracellular vesicles. *Exp Mol Med* 2022;54:466-482.
8. Shin H, Choi BH, Shim O, et al. Single test-based diagnosis of multiple cancer types using Exosome-SERS-AI for early stage cancers. *Nat Commun* 2023;14:1644.
9. Schlemper RJ, Riddell RH, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000;47:251-255.
10. Yoo JY, Rho M, You YA, et al. 16S rRNA gene-based metagenomic analysis reveals differences in bacteria-derived extracellular vesicles in the urine of pregnant and non-pregnant women. *Exp Mol Med* 2016;48:e208.
11. Yang J, McDowell A, Seo H, et al. Diagnostic models for atopic dermatitis based on serum microbial extracellular vesicle metagenomic analysis: a pilot study. *Allergy Asthma Immunol Res* 2020;12:792-805.
12. Witwer KW, Buzás EI, Bemis LT, et al. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *J Extracell Vesicles* 2013;2:20360.
13. Skryabin GO, Vinokurova SV, Galetsky SA, et al. Isolation and characterization of extracellular vesicles from gastric juice. *Cancers (Basel)* 2022;14:3314.
14. Yang J, McDowell A, Kim EK, et al. Consumption of a *Leuconostoc holzapfelii*-enriched synbiotic beverage alters the composition of the microbiota and microbial extracellular vesicles. *Exp Mol Med* 2019;51:1-11.
15. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37:852-857.
16. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581-583.
17. Aborujilah A, Mohammad Nassr R, Al-Hadhrani T, et al. Comparative study of SMOTE and bootstrapping performance based on predication methods. In: Saeed F, Mohammed F, Al-Nahari A, eds. *Innovative systems for intelligent health informatics. Proceedings of the 5th International Conference of Reliable Information and Communication Technology (IRICT 2020)*; 2020 Dec 21-22; Langkawi. Cham: Springer, 2021: 3-9.
18. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* 2021;17:e1009442.
19. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990;9:811-818.
20. Zeng R, Gou H, Lau HCH, Yu J. Stomach microbiota in gastric cancer development and clinical implications. *Gut* 2024;73: 2062-2073.
21. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. *Nat Med* 2018;24:392-400.
22. Park JY, Kang CS, Seo HC, et al. Bacteria-derived extracellular vesicles in urine as a novel biomarker for gastric cancer: integration of liquid biopsy and metagenome analysis. *Cancers (Basel)* 2021;13:4687.
23. Jaye K, Li CG, Bhuyan DJ. The complex interplay of gut microbiota with the five most common cancer types: from carcinogenesis to therapeutics to prognoses. *Crit Rev Oncol Hematol* 2021;165:103429.
24. Heo M, Park YS, Yoon H, et al. Potential of gut microbe-derived extracellular vesicles to differentiate inflammatory bowel disease patients from healthy controls. *Gut Liver* 2023;17:108-118.
25. Li CC, Hsu WF, Chiang PC, Kuo MC, Wo AM, Tseng YJ. Characterization of markers, functional properties, and microbiome composition in human gut-derived bacterial extracellular vesicles. *Gut Microbes* 2023;15:2288200.
26. Ricci V, Carcione D, Messina S, Colombo GI, D'Alessandra Y. Circulating 16S RNA in biofluids: extracellular vesicles as mirrors of human microbiome? *Int J Mol Sci* 2020;21:8959.
27. Morita Y, Tadokoro S, Sasai M, Kitamoto D, Hirashima N. Biosurfactant mannosyl-erythritol lipid inhibits secretion of inflammatory mediators from RBL-2H3 cells. *Biochim Biophys Acta* 2011;1810:1302-1308.
28. Fischer K, Tschisnarov R, Pilz A, et al. *Cutibacterium acnes* infection induces type I interferon synthesis through the cGAS-STING pathway. *Front Immunol* 2020;11:571334.
29. Sá S, Fernandes R, Gestoso Á, et al. *Cutibacterium acnes* dysbiosis: alternative therapeutics for clinical application. *Applied Sciences* 2023;13:12086.

30. Cros MP, Mir-Pedrol J, Toloza L, et al. New insights into the role of *Cutibacterium acnes*-derived extracellular vesicles in inflammatory skin disorders. *Sci Rep* 2023;13:16058.
31. Zhou S, Li C, Liu L, et al. Gastric microbiota: an emerging player in gastric cancer. *Front Microbiol* 2023;14:1130001.
32. Karpiński TM. Role of oral microbiota in cancer development. *Microorganisms* 2019;7:20.
33. Dong Z, Chen B, Pan H, et al. Detection of microbial 16S rRNA gene in the serum of patients with gastric cancer. *Front Oncol* 2019;9:608.
34. Hao Y, Karaoz U, Yang L, et al. Progressive dysbiosis of human orodigestive microbiota along the sequence of gastroesophageal reflux, Barrett's esophagus and esophageal adenocarcinoma. *Int J Cancer* 2022;151:1703-1716.
35. Sun QH, Zhang J, Shi YY, Zhang J, Fu WW, Ding SG. Microbiome changes in the gastric mucosa and gastric juice in different histological stages of *Helicobacter pylori*-negative gastric cancers. *World J Gastroenterol* 2022;28:365-380.
36. You HS, Park JY, Seo H, Kim BJ, Kim JG. Increasing correlation between oral and gastric microbiota during gastric carcinogenesis. *Korean J Intern Med* 2024;39:590-602.
37. Felipez N, Montori S, Mendizuri N, et al. The human gastric juice: a promising source for gastric cancer biomarkers. *Int J Mol Sci* 2023;24:9131.
38. Chen M, Zhao H. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum Genomics* 2019;13:34.
39. Zhang L, Zhao M, Fu X. Gastric microbiota dysbiosis and *Helicobacter pylori* infection. *Front Microbiol* 2023;14:1153269.

Received : September 26, 2024

Revised : December 19, 2024

Accepted : December 27, 2024

Correspondence to

Jae Gyu Kim, M.D., Ph.D.

Department of Internal Medicine, Chung-Ang University College of Medicine, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea

Tel: +82-2-6299-3147, Fax: +82-2-749-9150

E-mail: jgkimd@cau.ac.kr

<https://orcid.org/0000-0002-4841-9404>

CRedit authorship contributions

Hee Sang You: methodology, resources, investigation, data curation, formal analysis, validation, software, writing - original draft, writing - review & editing; Jae Yong Park: methodology, resources, investigation, formal analysis, validation, writing - original draft, writing - review & editing; Hoochan Seo: methodology, resources, data curation, formal analysis, validation, software, writing - review & editing; Beom Jin Kim: resources, investigation, writing - review & editing, funding acquisition; Jae Gyu Kim: conceptualization, data curation, formal analysis, writing - review & editing, supervision, project administration, funding acquisition

Conflicts of interest

The authors disclose no conflicts.

Funding

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under grant number NRF-2022R1F1A1069321 and by the Chung-Ang University Research Grants in 2023.

Supplementary Table 1. Random forest classifier analysis of gastric juice samples by groups

Act. group	Pred. group (train set)				Correctly classified	Pred. group (test set)				Correctly classified
	Control	LG	HG	GC		Control	LG	HG	GC	
Control	52	0	20	0	72.2%	18	0	10	0	64.3%
LG	13	32	15	9	46.4%	2	19	9	1	61.3%
HG	20	8	39	2	56.5%	0	3	24	4	77.4%
GC	11	8	18	33	47.1%	2	3	9	16	53.3%
Overall correct class. Rate					55.7%	64.2%				

The random forest classifier was performed using *Cutibacterium acnes* and *Streptococcus oralis*, identified as group-discriminative by MaAsLin2 analysis of gastric juice samples. To address sample imbalance, SMOTE was applied, generating a balanced dataset of 100 samples.

Pred., predicted group; Act., actual group; LG, low-grade dysplasia; HG, high-grade dysplasia; GC, gastric cancer.

Supplementary Table 2. Random forest classifier analysis of saliva samples by groups

Act. group	Pred. group (train set)				Correctly classified	Pred. group (test set)				Correctly classified
	Control	LG	HG	GC		Control	LG	HG	GC	
Control	68	0	0	0	100.0%	32	0	0	0	100.0%
LG	2	57	0	13	79.2%	1	20	3	4	71.4%
HG	0	2	59	8	85.5%	3	1	20	7	64.5%
GC	2	1	1	67	94.4%	1	3	0	25	86.2%
Overall correct class. Rate					89.6%	80.8%				

The random forest classifier was performed using *Cutibacterium acnes*, *Ralstonia insidiosa*, *Streptococcus oralis*, and *Pseudomonas antarctica*, identified as group-discriminative by MaAsLin2 analysis of saliva samples. To address sample imbalance, SMOTE was applied, generating a balanced dataset of 100 samples.

Pred., predicted group; Act., actual group; LG, low-grade dysplasia; HG, high-grade dysplasia; GC, gastric cancer.

Supplementary Table 3. Random forest classifier analysis of serum samples by groups

Act. group	Pred. group (train set)				Correctly classified	Pred. group (test set)				Correctly classified
	Control	LG	HG	GC		Control	LG	HG	GC	
Control	72	0	0	0	100.0%	28	0	0	0	100.0%
LG	6	45	9	9	65.2%	3	23	1	4	74.2%
HG	7	7	50	5	72.5%	0	0	28	3	90.3%
GC	9	4	4	53	75.7%	4	0	0	26	86.7%
Overall correct class. Rate					78.6%	87.5%				

The random forest classifier was performed using *Pseudomonas yamanorum*, *Cutibacterium acnes*, *Ralstonia insidiosa*, and *Pseudomonas antarctica*, identified as group-discriminative by MaAsLin2 analysis of serum samples. To address sample imbalance, SMOTE was applied, generating a balanced dataset of 100 samples.

Pred., predicted group; Act., actual group; LG, low-grade dysplasia; HG, high-grade dysplasia; GC, gastric cancer.

Supplementary Table 4. Random forest classifier analysis of urine samples by groups

Act. group	Pred. group (train set)				Correctly classified	Pred. group (test set)				Correctly classified	
	Control	LG	HG	GC		Control	LG	HG	GC		
Control	72	0	0	0	100.0%	28	0	0	0	100.0%	
LG	6	43	11	9	62.3%	2	20	2	7	64.5%	
HG	3	5	56	4	82.4%	4	1	25	2	78.1%	
GC	9	5	5	52	73.2%	3	1	3	22	75.9%	
Overall correct class. Rate					79.6%						79.2%

The random forest classifier was performed using *Cutibacterium acnes*, *Ralstonia insidiosa*, and *Pseudomonas antarctica*, identified as group-discriminative by MaAsLin2 analysis of urine samples. To address sample imbalance, SMOTE was applied, generating a balanced dataset of 100 samples.

Pred., predicted group; Act., actual group; LG, low-grade dysplasia; HG, high-grade dysplasia; GC, gastric cancer.