Molecular profiling of microbiota in human oral cavity, stomach and intestine

Qianqian Liu [1], **Feizhou Zhu** [Corresp., 1], **Liyu Chen** [2], **Meihua Xu** [3], **Jianwei Chen** [4], **Colin K. He** [5,6], **Lingzhi Zou** [3], **Na Jiang** [3], **Xianghong Liu** [3], **Yang Meng** [1], **Shijing Liao** [2], **Yong Peng** [7], **Lili Wang** [2]

[1] Department of Biochemistry and Molecular Biology, School of Life Sciences, Central South University, Changsha, People's Republic of China

[2] Department of Medical Microbiology, School of Basic Medical Sciences, Central South University, Changsha, Hunan, People's Republic of China.

[3] Department of Gastroenterology, Xiangya Hospital, Central South University, Changsha, Hunan, People's Republic of China

[4] BGI-Qingdao, BGI, Qingdao, People's Republic of China

[5] Stegotech LLC, Audubon, PA, USA

[6] The Third Xiangya Hospital, Central South University, Changsha, Hunan, People's Republic of China

[7] Affiliated First Hospital, Hunan Traditional Chinese Medical College, Zhuzhou, Hunan, People's Republic of China

Corresponding Author: Feizhou Zhu
Email address: zhufeizhou@csu.edu.cn

The microbiota in the human gut is not only a complicated microecological system but also plays important roles in both health and disease. In order to understand the roles of these gut bacteria, we determined the distribution of microbiota in different regions of the gut by sequencing the 16S rRNA gene V4 region of the bacteria in the saliva, gastric juice, and stool of healthy individuals. The 16S rRNA gene V3-V5 region sequences of saliva and stool microbiota were obtained from Human Microbiome Project (HMP) and the V4 sequence was obtained from the V3-V5 sequences by a program designed by Perl language. We found that the microbiota of the gastric juice is more similar to those in the saliva rather than that in the stool. The frequency of some taxa was significantly different among the three groups with the *Streptococcus, Veillonella, Oribacterium, Selenomonas, Actinomyces,* and *Granulicatella* most abundant in the saliva; the *Prevotella, Neisseria, Actinobacillus, Treponema,* and *Helicobacter* most abundant in the gastric juice; and the *Bacteroides, Parabacteroides, Faecalibacterium, Sutterella, Ruminococcus, Oscillospira* and *Phascolarctobacterium* most abundant in the stool. In addition, results from PICRUSt analyses suggest that the functions of microbiota in the gastric juice are more similar as those in the saliva than in the stool. Moreover, we also found that the membrane transport of the microbiota in the saliva is higher than that in the stool and gastric juice. To our knowledge, this is the first comprehensive comparison of microbiota in the human oral cavity, stomach, and intestine.

Molecular profiling of microbiota in human oral cavity, stomach, and intestine

Qianqian Liu[1], Feizhou Zhu[1]*, Liyu Chen[2], Meihua Xu[3], Jianwei Chen[4], Colin K. He[5, 6], Lingzhi Zou[3], Na Jiang[3], Xianghong Liu[3], Yang Meng[1], Shijing Liao[2], Yong Peng[7] and Lili Wang[2]

[1] Department of Biochemistry and Molecular Biology, School of Life Sciences, Central South University, Changsha, Hunan, 410013, People's Republic of China; [2] Department of Medical Microbiology, School of Basic Medical Sciences, Central South University, Changsha, Hunan, 410013, People's Republic of China; [3] Department of Gastroenterology, Xiangya Hospital, Central South University, Changsha, Hunan, 410008, People's Republic of China; [4] BGI-Qingdao, BGI-Shenzhen, Qingdao, 266555, People's Republic of China; [5] Stegotech LLC, Audubon, PA 19403, USA; [6] The Third Xiangya Hospital, Central South University, Changsha, Hunan, 410013, People's Republic of China; and [7]Affliated First Hospital of Hunan Traditional Chinese Medical College, Zhuzhou, China, 412000.

*Correspondence: Feizhou Zhu, e-mail: zhufeizhou@csu.edu.cn

Abstract

The microbiota in the human gut is not only a complicated microecological system but also plays important roles in both health and disease. In order to understand the roles of these gut bacteria, we determined the distribution of microbiota in different regions of the gut by sequencing the 16S rRNA gene V4 region of the bacteria in the saliva, gastric juice, and stool of healthy individuals**.** The 16S rRNA gene V3-V5 region sequences of saliva and stool microbiota were obtained from Human Microbiome Project (HMP) and the V4 sequence was obtained from the V3-V5 sequences by a program designed by Perl language. We found that the microbiota of the gastric juice is more similar to those in the saliva rather than that in the stool. The frequency of some taxa was significantly different among the three groups with the *Streptococcus, Veillonella, Oribacterium, Selenomonas, Actinomyces,* and *Granulicatella* most abundant in saliva; the *Prevotella, Neisseria, Actinobacillus, Treponema,* and *Helicobacter* most abundant in the gastric juice; and the *Bacteroides, Parabacteroides, Faecalibacterium, Sutterella, Ruminococcus, Oscillospira* and *Phascolarctobacterium* most abundant in the stool. In addition, results from PICRUSt analyses suggest that the functions of microbiota in the gastric juice are more similar as those in the saliva than in the stool. Moreover, we also found that the membrane transport of the microbiota in the saliva is higher than that in the stool and gastric juice. To our knowledge, this is the first comprehensive comparison of microbiota in the human oral cavity, stomach, and intestine.

Keywords Oral Microbiota; Gastric Microbiota; Intestinal Microbiota; Metagenomics; 16S rRNA gene

Introduction

The microbiota in the human gut is not only a complicated microecological system, but also plays critical roles in both health and disease. The fundamental functions of gut microflora include salvaging energy and absorbing nutrients, exerting important trophic effects on intestinal epithelia

and immune structure and function to protect the host from alien microbes' invasion. Imbalanced gut flora are highly related to certain diseases such as colon cancer, inflammatory bowel diseases (Guarner & Malagelada, 2003), rheumatoid arthritis (Zhang et al., 2015) and type 2 diabetes (Qin et al., 2012).

Before the discovery of Helicobacter pylori (Hp), human stomach was considered sterile because the acidic environment was assumed unfavorable to the survival of microorganisms from the oral cavity. After the finding of the Hp in the stomach, the microbial ecosystem of the stomach became the focus of the research. Myriad bacteria such as *Enterococcus, Pseudomonas, Streptococcus, Staphylococcus* genera*, Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes* and *Fusobacteria* have been identified in the stomach by analyzing the16S rRNA gene sequences (Bik et al., 2006; Monstein et al., 2000). In the past few decades, molecular biology techniques have become more popular in identifying microorganisms. However, these techniques are not only complicated but also expensive. Therefore, metagenomic analysis based on high-throughput sequencing, bioinformatics and large-data statistics have gained popularity. We were interested in developing a rapid, high-sensitive, cheap and culture-free method for analyzing human microbiota to dissect the relationship between microecology and digestive diseases (Wu et al., 2016; Yu et al., 2017; Sun et al., 2016).

Gut microbiota start to develop immediately after birth and evolve throughout the lifespan. Mainly due to food consumption, the sterile digestive tract is populated with bacteria from the oral cavity to the stomach and eventually to the intestine. It is reasonable to assume that microbiota of oral cavity, stomach, and intestine share some similarity. Indeed, it has been reported that the stomach microbiota were dissimilar to those in the oral cavity and the intestine (Yu et al., 2017; Stearns et al., 2011). In this study, the gastric juice samples were collected from 28 healthy individuals and the 16S rRNA gene V4 region of their microbiota was sequenced by Miseq platform. The 16S rRNA gene V4 region sequences of saliva and stool microbiota of healthy individuals were obtained from the Human Microbiome Project (HMP). Since the 16S rRNA gene V4 region sequences are unavailable in HMP, the V4 regions were abstracted from the V3-V5

sequences in HMP by a program designed by Perl language. Therefore, the V4 region sequences of 248 saliva samples and 271 stool samples from healthy individuals were used for the analyses. To our knowledge, this is the first comprehensive comparison of microbiota in human oral cavity, stomach, and intestine.

Materials and methods

Gastric juice samples collection and bacterial DNA isolation

Twenty-eight healthy individuals were recruited in the Department of Gastrointestinal Endoscope of Xiangya Hospital, Changsha, Hunan, China, from October 2015 to November 2016 (Supplementary S1 Table). The study was approved by the independent Ethics Committee of Xiangya Hospital of Central South University in accordance with the ethical guidelines of the Declaration of Helsinki (No. 201512548). Participation was voluntary and written informed consent was obtained from all participants. The following criteria were used for exclusion: (1) age under 18 years; (2) the presence of a serious illness such as severe cardiopulmonary, renal or metabolic diseases; (3) prior medication history of antibiotics, acid drugs (proton pump inhibitors and H2 receptor antagonists), (4) probiotics, or anti-inflammatory drugs (aspirin, nonsteroidal and steroids) for past one month; (5) a large amount of alcohol consumption and smoking for past one month. Participants fasted for more than 12 h before the endoscopic examination. Approximately 10 mL of gastric juice was collected from the stomach during gastroscopy using a sterile syringe, filtered by double sterile gauze to remove food debris, stored in sterile 10-15 mL tubes and maintained at 0 °C no more than 12 h before DNA isolation. The bacterial sediments were collected by centrifugation of the gastric juice at 12000 rpm for 10 min at 4 °C. DNA isolation was followed by using the QIAamp® FAST DNA Stool Mini Kit (QIAGEN) according to the manufacturer's protocol.

Sequencing 16S rRNA gene V4 region of gastric juice microbiota

Bacterial 16S rRNA gene V4 region of gastric juice was amplified by PCR with the forward 515F (**5'-**gtgccagcmgccgcggtaa**-3'**) and reverse 806R (**5'-**ggactachvgggtwtctaat**-3'**) primers. The PCR product was purified by a QIAquick PCR Purification Kit (QIAGEN). The jagged ends of the amplicons were converted into blunt ends using T4 DNA polymerase, Klenow Fragment and T4 Polynucleotide Kinase. Then, an 'A' base was added to each 3' end to make it easier to add adapters, and special sequencing adapters were added to each end of amplicons to construct libraries. However, too short fragments were removed by Ampure beads. The libraries were screened and only the qualified library was used for constructing cluster and high-throughput sequencing. The high-throughput sequencing was conducted at the Beijing Genomics Institute (BGI, Wuhan, China) on the Illumina Miseq PE250 sequencer platform (Supplementary S2 Table).

The 16S rRNA gene V4 region sequences of saliva and stool microbiota

The 16S rRNA gene V3-V5 sequences from saliva and stool microbiota in healthy individuals were downloaded from HMP and the V4 sequences were derived from the V3-V5 sequences by a program designed by Perl language. Briefly, the PCR primers for the V4 region, the forward 515F (**5'-**gtgccagcmgccgcggtaa**-3'**) and reverse 806R (**5'-**ggactachvgggtwtctaat**-3'**), were mapped to tag two sides of the V3-V5 region. If 4 consecutive bases at the 3'-end of the primers could completely match the tags and the mismatching bases of the remaining primer were less than 2, the tags were retained and the sequences between 515F-806R were cut from the V3-V5 sequences. Otherwise, the sequence was discarded. In addition, samples with V4 tags less than 3000 were discarded. Therefore, the V4 region sequences of 248 saliva samples and 271 stool samples from healthy individuals were used for the following analyses (Supplementary S2 Table).

Bioinformatical analyses

After sequencing, the reads were de-multiplexed according to the barcodes using the Quantitative Insights into Microbial Ecology (QIIME) pipeline (Denver, CO, USA) with the default parameters

(Caporaso et al., 2010). The raw data were filtered to eliminate adapter pollution and low-quality reads to obtain clean reads, then paired-end clean reads with overlap were merged to tags. The QIIME pipeline was used to cluster and annotate the tags to the Operational Taxonomic Units (OTUs), map the taxonomic profiles of microbiota, compare the relative abundance of taxa as well as calculate the beta diversity. Briefly, the tags with at least 97% sequence identity were clustered and annotated into species-level OTUs according to the Greengenes database version 13_8 (McDonald et al., 2011). Rank curves were plotted by the script of plot_rank_abundance_graph.py. Area maps of the taxa were plotted by the script of plot_taxa_summary.py. The comparisons of the taxonomic relative frequencies were performed by the script of group_significance.py with Kruskal-Wallis nonparametric Analysis of Variance (ANOVA). The beta diversity and principal component analysis (PCA) was detected with a rarefaction depth of 3000 by the script of core_diversity_analyses.py. The comparison of the beta diversity was not only performed by the scripts of compare_categories.py with the Permutational Multivariate Analysis of Variance (PERMANOVA) using permutations of 999, but also was carried out by the script of make_distance_boxplots.py with two-sample t-tests. Multiple comparison was corrected by Bonferroni and Bonferroni_P < 0.05 was considered significant.

The biomarkers were discovered by the Linear Discriminant Analysis (LDA) Effect Size (LEfSe)**.** LEfSe is an algorithm for high-dimensional biomarker discovery and an explanation that identifies genomic features characterizing the differences between two or more biological conditions (http://huttenhower.sph.harvard.edu/galaxy/) (Segata et al., 2011). The threshold on the logarithmic LDA score for discriminative features was set to 3.0. The metagenomic functions of saliva, gastric juice and stool microflora were predicted and annotated by the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Langille et al., 2013). In addition, the Venn diagrams were plotted by Venn online software (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Results

Rank of species in the gut

In this study, it was the inclusion criteria that the V4 tags of an individual were not less than 3000 after being cut from the V3-V5 sequences. As a result, V4 tags of the saliva were 6355±3432 (mean ± SD) and V4 tags of the stool were 8343±6645 (mean ± SD) in this study (Supplementary S2 Table). However, the V4 tags of gastric juice were as much as 34976±3814 (mean ± SD), because they were obtained by the high-throughput sequencing (Supplementary S2 Table). Therefore, we raised the question whether the different amount of tags in the 3 groups would affect the analyses of this study.

The Rank curves of the species-level OTUs were plotted to answer the question (Fig. 1). According to the Rank diagram, all sample tags covered the major species well whose relative abundance was higher than 0.1%. Nevertheless, minor species, whose relative abundance was less than 0.1%, were covered differently regarding the number of sample tags. The more the sample tags were present, the more minor species were identified (Fig. 1A, 1B and 1C). As a result, though the number of sample tags was different in the 3 groups, the major species of 3 groups were profiled well. However, some minor species were profiled poorly due to insufficient tags. This fact did not affect most analyses much in this study, because the abundance of minor species was too low and people's interests focused on the major species. However, a few analyses might have been affected, such as the Venn diagram and alpha diversity.

Taxonomic profiles of gut microbiota

As the bacterial taxa are quite a lot in the gut, only the major taxa (mean of relative frequency is more than 1% in one group) are shown in the area diagram (Fig. 2A and 2B) and were analyzed by Kruskal-Wallis nonparametric ANOVA (Table 1 and 2) and LEfSe (Fig. 3).

In the gut, the major bacterial phyla are *Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria* and *Spirochaetes* (Fig. 2A). At the phylum level, the frequency of

some taxa was significantly different among the three groups with the *Firmicutes* and *Actinobacteria* most abundant in the saliva; the *Proteobacteria* and *Spirochaetes* most abundant in the gastric juice; and the *Bacteroidetes* most abundant in the stool (Table 1). For the frequency of *Fusobacteria*, the saliva and the gastric juice had no significant differences, but both were significantly higher than the stool (Table 1). At the genus level, the frequency of some taxa was significantly different among the three groups with the *Streptococcus, Veillonella, Oribacterium, Selenomonas, Actinomyces,* and *Granulicatella* most abundant in the saliva; the *Prevotella, Neisseria, Actinobacillus, Treponema,* and *Helicobacter* most abundant in the gastric juice; and the *Bacteroides, Parabacteroides, Faecalibacterium, Sutterella, Ruminococcus, Oscillospira,* and *Phascolarctobacterium* most abundant in the stool (Table 2). Some genera had no significant differences between the saliva and the gastric juice, but were least abundant in the stool, such as *Porphyromonas, Capnocytophaga, Fusobacterium, Leptotrichia, Lautropia, Campylobacter, Aggregatibacter,* and *Haemophilus* (Table 2). It is indicated that the microbiota of the gastric juice is more similar to that of the saliva than that of the stool (Fig. 2A and 2B, Table 2). In addition, the results of LEfSe (Fig. 3) were consistent with the results of the Kruskal-Wallis nonparametric ANOVA (Table 2).

The Venn diagram reflects the partial situations of the genera in the 3 groups for the reason above. In brief, there were 151 common genera in the saliva and gastric juice, 108 common genera in the saliva and stool, 89 common genera in the gastric juice and stool, and 76 common genera in the three groups (Supplementary S1 Fig).

Diversity of gut microbiota

Beta diversity was figured out to show the dissimilarities of microbial profiles in the saliva, gastric juice and stool. Principal Component Analysis (PCA) suggested that the microbial profile of the saliva was more similar to the gastric juice than that of the stool (Fig. 4A), consistent with the results of the Kruskal-Wallis nonparametric ANOVA (Table 1 and 2). However, according to the PERMANOVA, the beta diversity between each pair of the 3 groups all reached a highly

significant level (Bonferroni_P = 0.003, Fig. 4B).

Furthermore, the beta diversity between each pair of the 3 groups was significant different with the least dissimilarities between the gastric juice and the saliva; the most dissimilarities between the stool and the gastric juice; and the medium dissimilarities between the stool and the saliva (Fig. 4C). In addition, alpha diversity was not measured for the mentioned reasons.

Functional profiles of gut microbiota

PICRUSt provides a cost-saving way to study the metagenomic functions using 16S rRNA gene sequences because the metagenomic sequencing is very expensive. After the annotation by KEGG, it is showed that the cellular pathways were different in the microbiota of the saliva, gastric juice and stool. The carbohydrate metabolism and amino acid metabolism in the stool microbiota were higher than those of the saliva and those of gastric juice, but the replication, repair and translation in the stool microbiota were lower than those of the saliva and those of gastric juice. Specially, the membrane transport in the saliva microbiota was higher than that of the stool and that of gastric juice (Fig. 5A). It is indicated by the PCA that the cellular functions of the gastric juice microbiota were similar to those of saliva microbiota rather than those of stool microbiota (Fig. 5B), consistent with the PCA of β-diversity (Fig. 4A).

Discussion

We showed in this study that molecular profile of gastric juice microbiota is more similar to saliva microbiota than stool microbiota; presumably that many gastric bacteria were originated from the oral cavity carried by foods and drinks. This finding is consistent with the previously published data (Yu et al., 2017; Stearns et al., 2011). In addition, Stearns JC et al. found that the highest OTU richness and phylogenetic diversity of microflora were oral bacteria. In contrast, the bacteria in the gut showed lowest OTU richness (Stearns et al., 2011). In our study, the α-diversity analysis was not performed due to insufficient tags in the saliva and the stool.

Most of our results are generally consistent with findings of the others (Yu et al., 2017; Stearns et al., 2011). Of note, specific analytical methods used in our research: (1) the major bacteria in the saliva, gastric juice and stool were compared comprehensively by Kruskal-Wallis nonparametric ANOVA; (2) LEfSe was applied to find the biomarkers of the saliva, gastric juice and stool; and (3) PICRUSt was used to predict the differential cellular pathways of the saliva, gastric juice and stool. This enabled us to define specific bacteria in the oral cavity, stomach and intestine, such as *Oribacterium*in in the oral cavity, *Helicobacter* in the stomach, and *Parabacteroides*, *Faecalibacterium,* and *Sutterella* in the intestine (Table 2). These bacteria might have unique functions for the organs which they inhabited. For example, *Helicobacter pylori* is related to gastric ulcers and gastric cancer (Hwang et al., 2015), and *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium in Crohn's Disease (Sokol et al., 2008). These findings will be help in better understanding the distributions of microbiota and the importance of maintaining a balanced microbiota in the gastro-enteral system.

Author Contributions

F.Z. developed the research hypothesis and designed the experiment. Q.L., L.C., M.X and Y.P. performed the main experiments and wrote the main manuscript text. J.C. and C.K.H analyzed the data. L.Z. et al. collected samples. The final manuscript is an end product of joint writing efforts of all the authors.

Conflicts of interest

Colin K. He is a chief science officer (CSO) employed by Stegotech LLC (Audubon, USA). Jianwei Chen is a bioinformatic engineer employed by BGI-Qingdao (Qingdao, People's Republic of China).

Acknowledgments

This work was supported by the Natural Science Foundation of Hunan Province, PR. China (No.2015jj2162), Science and Technology Plan Project of Hunan Province, China (No.2017SK2092), the Research Foundation for Independent Innovation of Graduate Students in Central South University (No.2017zzts359), the Institute Research Project of Hunan Province, China (No. 14C0879).

Data availability

All 16S rRNA gene V4 sequences of gastric juice bacterial flora had been submitted to Sequence Read Archive (SRA) with accession number SRP139893 (the SRA submission will be released on 2019-04-02 or upon publication, whichever is first). All 16S rRNA gene sequences of saliva and stool bacterial flora in this study are available in the HMP.

Supplementary data

Additional supporting data may be found in the online version of this article:

Fig. S1. A Venn diagram of genera in the gut microbiota.

Table S1. Demographics of healthy people donating gastric juices.

Table S2. Statistics of sample sequences

References

Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, Perez-Perez G, Blaser MJ, Relman DA. 2006. Molecular analysis of the bacterial microbiota in the human stomach.

Proceedings of the National Academy of Sciences 103(3):732-737 DOI 10.1073/pnas.0506655103

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5):335-336 DOI 10.1038/nmeth.f.303

Guarner F, Malagelada J. 2003. Gut flora in health and disease. *Lancet* 361(9356):512-519 DOI 10.1016/S0140-6736(03)12489-0

Hwang JJ, Lee DH, Lee AR, Yoon H, Shin CM, Park YS, Kim N. 2015. Characteristics of gastric cancer in peptic ulcer patients with Helicobacter pylori infection. *World J Gastroenterol* 21(16):4954-4960 DOI 10.3748/wjg.v21.i16.4954

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* 31(9):814-821 DOI 10.1038/nbt.2676

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2011. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal* 6(3):610-618 DOI 10.1038/ismej.2011.139

293 Monstein HJ, Tiveljung A, Kraft CH, Borch K, Jonasson J. 2000. Profiling of bacterial flora in
294 gastric biopsies from patients with Helicobacter pylori-associated gastritis and histologically
295 normal control individuals by temperature gradient gel electrophoresis and 16S rDNA
296 sequence analysis. *Journal of Medical Microbiology* 49(9):817-822 DOI 10.1099/0022-1317-
297 49-9-817
298 Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D,
299 Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S,
300 Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T,
301 Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto
302 J, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R,
303 Pedersen O, Kristiansen K, Wang J. 2012. A metagenome-wide association study of gut
304 microbiota in type 2 diabetes. *Nature* 490(7418):55-60 DOI 10.1038/nature11450
305 Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011.
306 Metagenomic biomarker discovery and explanation. *Genome Biology* 12(6):R60 DOI
307 10.1186/gb-2011-12-6-r60
308 Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, Blugeon S,
309 Bridonneau C, Furet JP, Corthier G, Grangette C, Vasquez N, Pochart P, Trugnan G, Thomas
310 G, Blottiere HM, Dore J, Marteau P, Seksik P, Langella P. 2008. Faecalibacterium prausnitzii
311 is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn
312 disease patients. *Proceedings of the National Academy of Sciences* 105(43):16731-16736 DOI
313 10.1073/pnas.0804812105

Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Moreno-Hagelsieb G, Neufeld JD. 2011. Bacterial biogeography of the human digestive tract. *Scientific Reports* 1(1):170 DOI 10.1038/srep00170

Sun Y, Ma Y, Lin P, Tang Y, Yang L, Shen Y, Zhang R, Liu L, Cheng J, Shao J, Qi T, Tang Y, Cai R, Guan L, Luo B, Sun M, Li B, Pei Z, Lu H. 2016. Fecal bacterial microbiome diversity in chronic HIV-infected patients in China. *Emerging Microbes & Infections* 5(4):e31 DOI 10.1038/emi.2016.25

Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, Ma Y, Purdue MP, Jacobs EJ, Gapstur SM, Li H, Alekseyenko AV, Hayes RB, Ahn J. 2016. Cigarette smoking and the oral microbiome in a large study of American adults. *The ISME Journal* 10(10):2435-2446 DOI 10.1038/ismej.2016.37

Yu G, Torres J, Hu N, Medrano-Guzman R, Herrera-Goepfert R, Humphrys MS, Wang L, Wang C, Ding T, Ravel J, Taylor PR, Abnet CC, Goldstein AM. 2017. Molecular Characterization of the Human Stomach Microbiota in Gastric Cancer Patients. *Frontiers in Cellular and Infection Microbiology* 7:302 DOI 10.3389/fcimb.2017.00302

Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, Wu X, Li J, Tang L, Li Y, Lan Z, Chen B, Li Y, Zhong H, Xie H, Jie Z, Chen W, Tang S, Xu X, Wang X, Cai X, Liu S, Xia Y, Li J, Qiao X, Al-Aama JY, Chen H, Wang L, Wu Q, Zhang F, Zheng W, Li Y, Zhang M, Luo G, Xue W, Xiao L, Li J, Chen W, Xu X, Yin Y, Yang H, Wang J, Kristiansen K, Liu L, Li T, Huang Q, Li Y, Wang J. 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Medicine* 21(8):895-905 DOI 10.1038/nm.3914

Figure 1

Rank of bacterial species-level OTUs in the gut.

(A) The rank of bacterial species-level OTUs in the gut. (B) The rank of bacterial species-level OTUs in the gastric juice. (C) The rank of bacterial species-level OTUs in the saliva. (D) The rank of bacterial species-level OTUs in the stool.

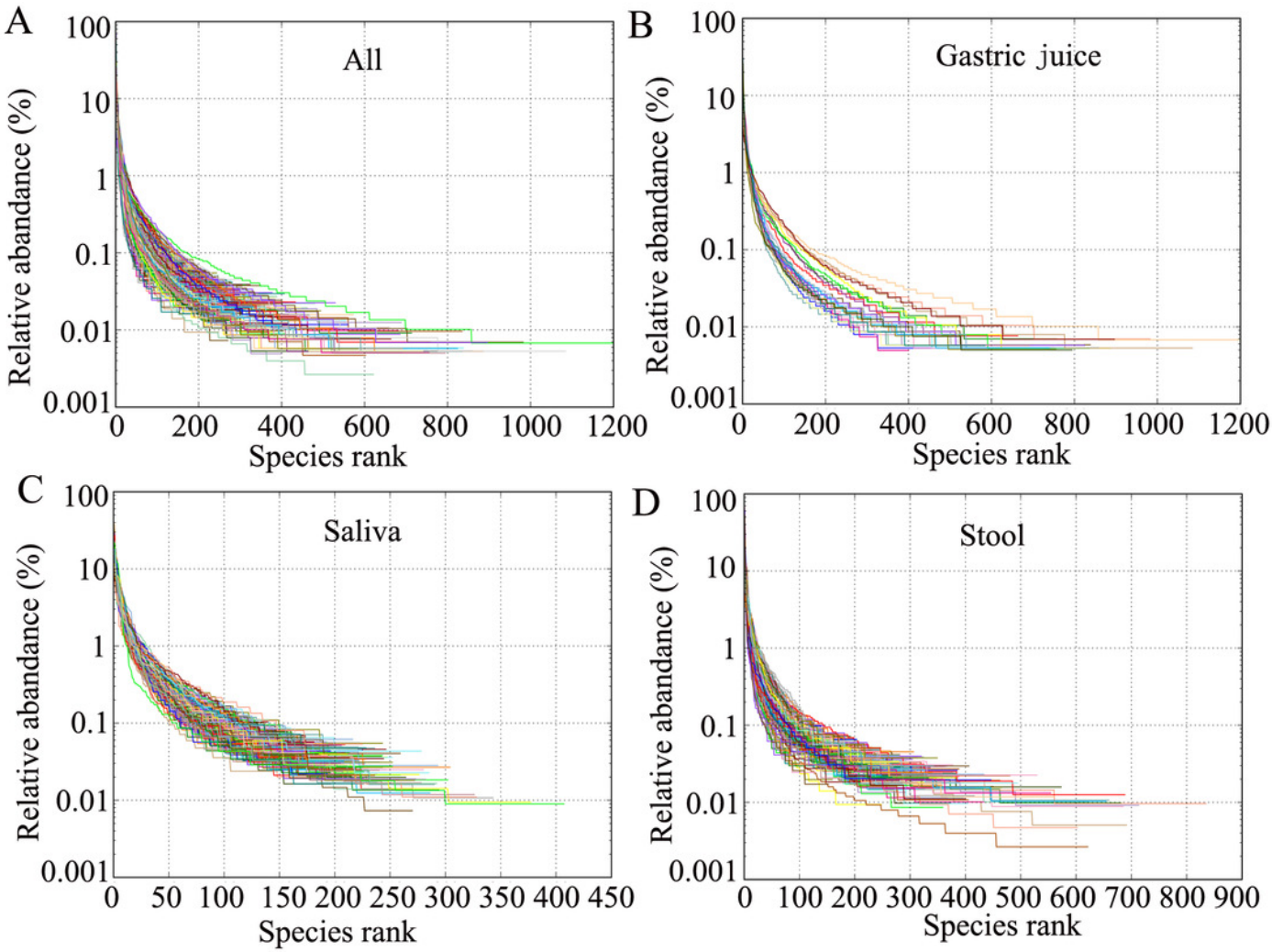


Figure 2

Profiles of bacterial flora in the gut at the phylum (A) and genus (B) level.

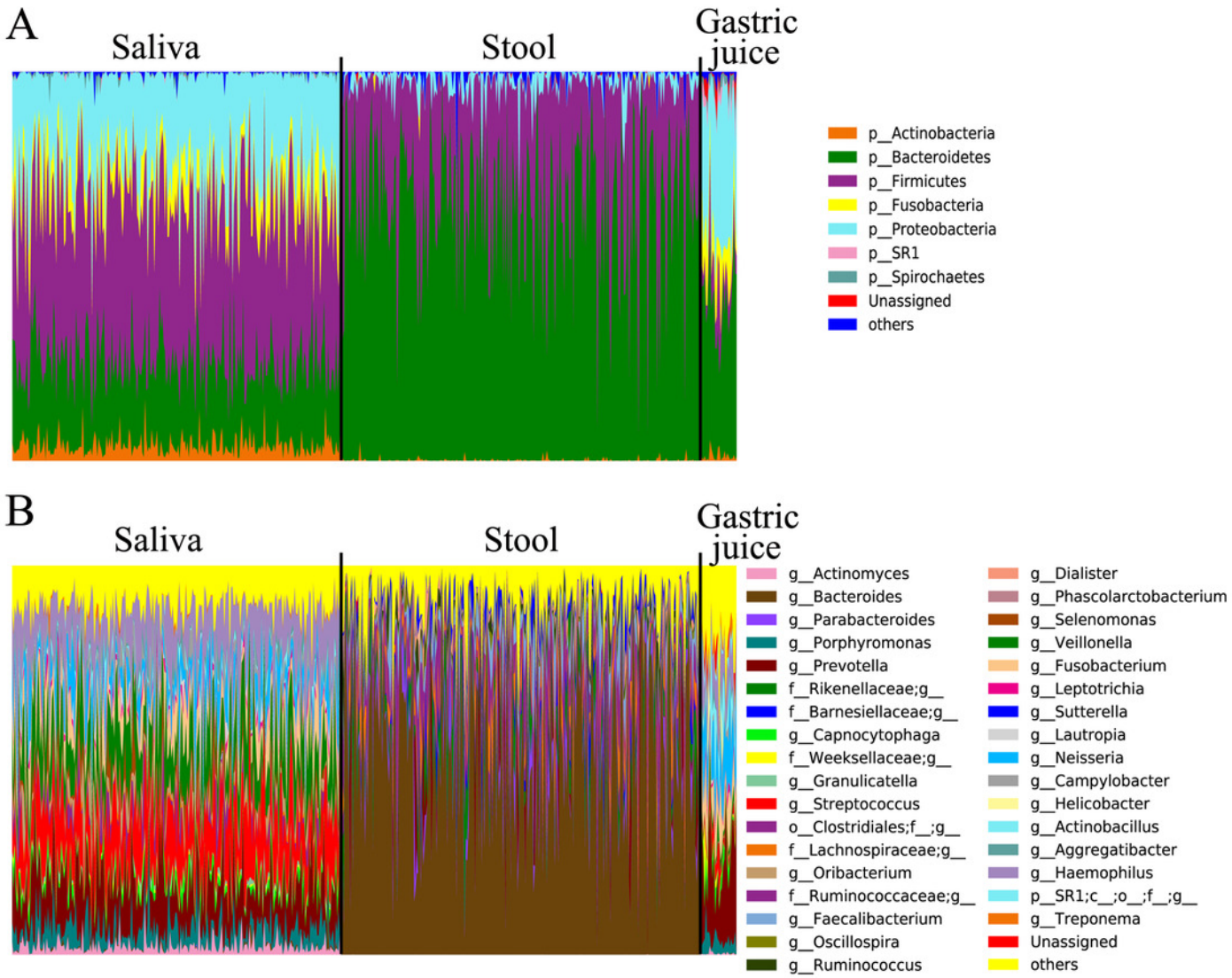
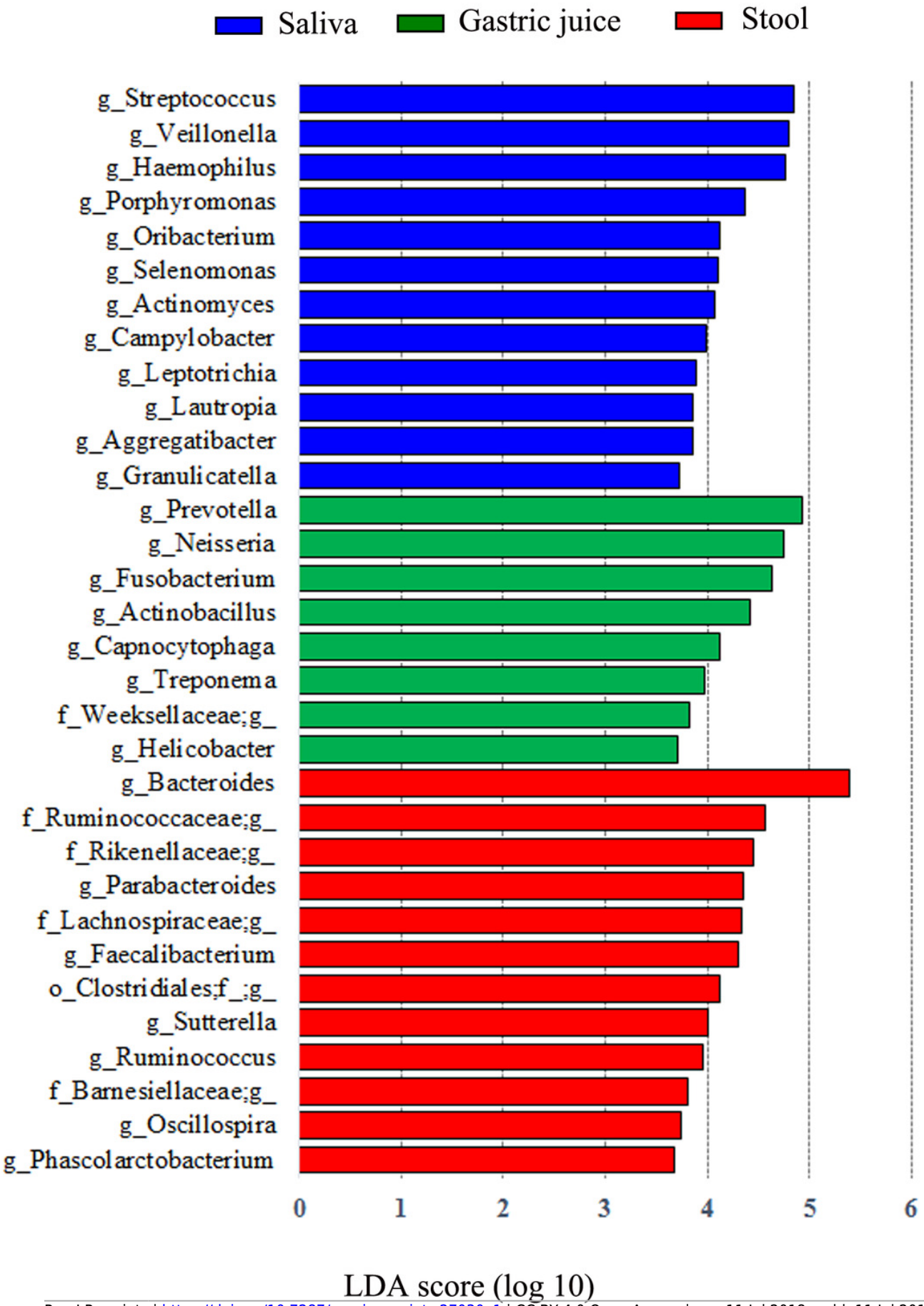


Figure 3

Biomarkers discovery by LEfSe.

The threshold on the logarithmic LDA score for discriminative features was set to 3.0.

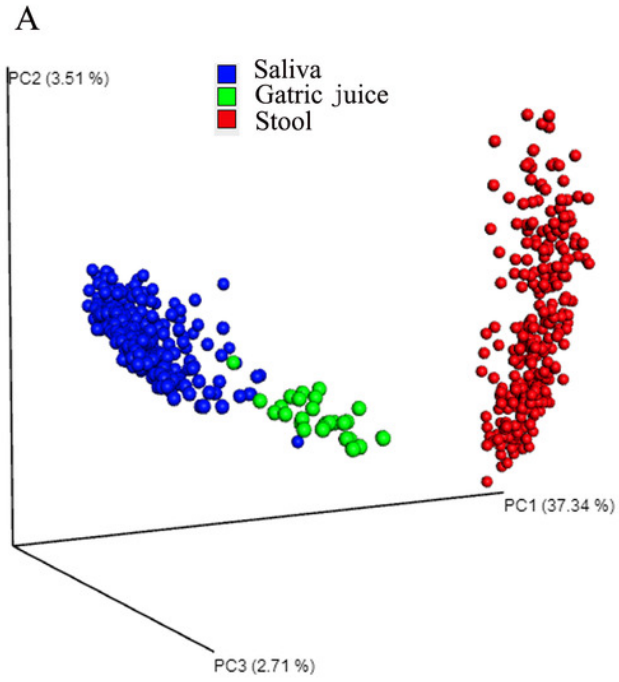


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27030v1 | CC BY 4.0 Open Access | rec: 11 Jul 2018, publ: 11 Jul 2018

Figure 4

Beta-diversity of microbiota in the gut.

(A) The PCA of microbiota in the human oral cavity, stomach, and intestine. (B) The PERMANOVA of microbiota in the human oral cavity, stomach, and intestine. (C) The unweighted Unifrac distances were compared to evaluate the dissimilarities of microbiota profiles in the human oral cavity, stomach, and intestine. Multiple comparison was corrected by Bonferroni and Bonferroni_P < 0.05 was considered significant. ***: Bonferroni_P < 0.001.



B PERMANOVA

Distance matrix	P	Bonferroni_P
Saliva VS Gastric juice	0.001	0.003
Saliva VS Stool	0.001	0.003
Stool VS Gastric juice	0.001	0.003

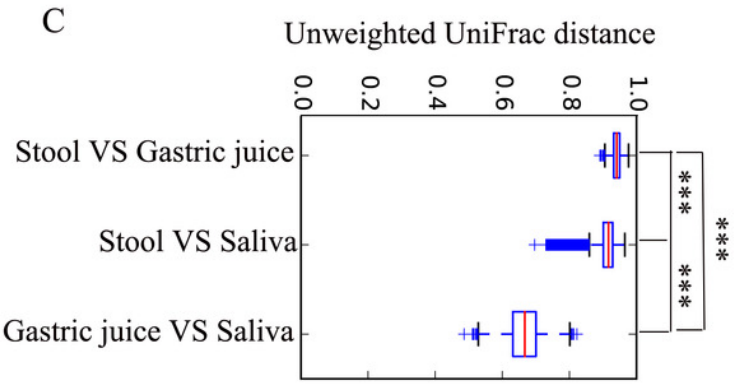


Figure 5

Functional analysis of the microbiota in the gut using PICRUSt.

A heatmap (A) and PCA (B) of the bacterial cellular pathway in microbiota of saliva, gastric juice and stool were analyzed by PICRUSt using 16S rRNA gene V4 sequences.

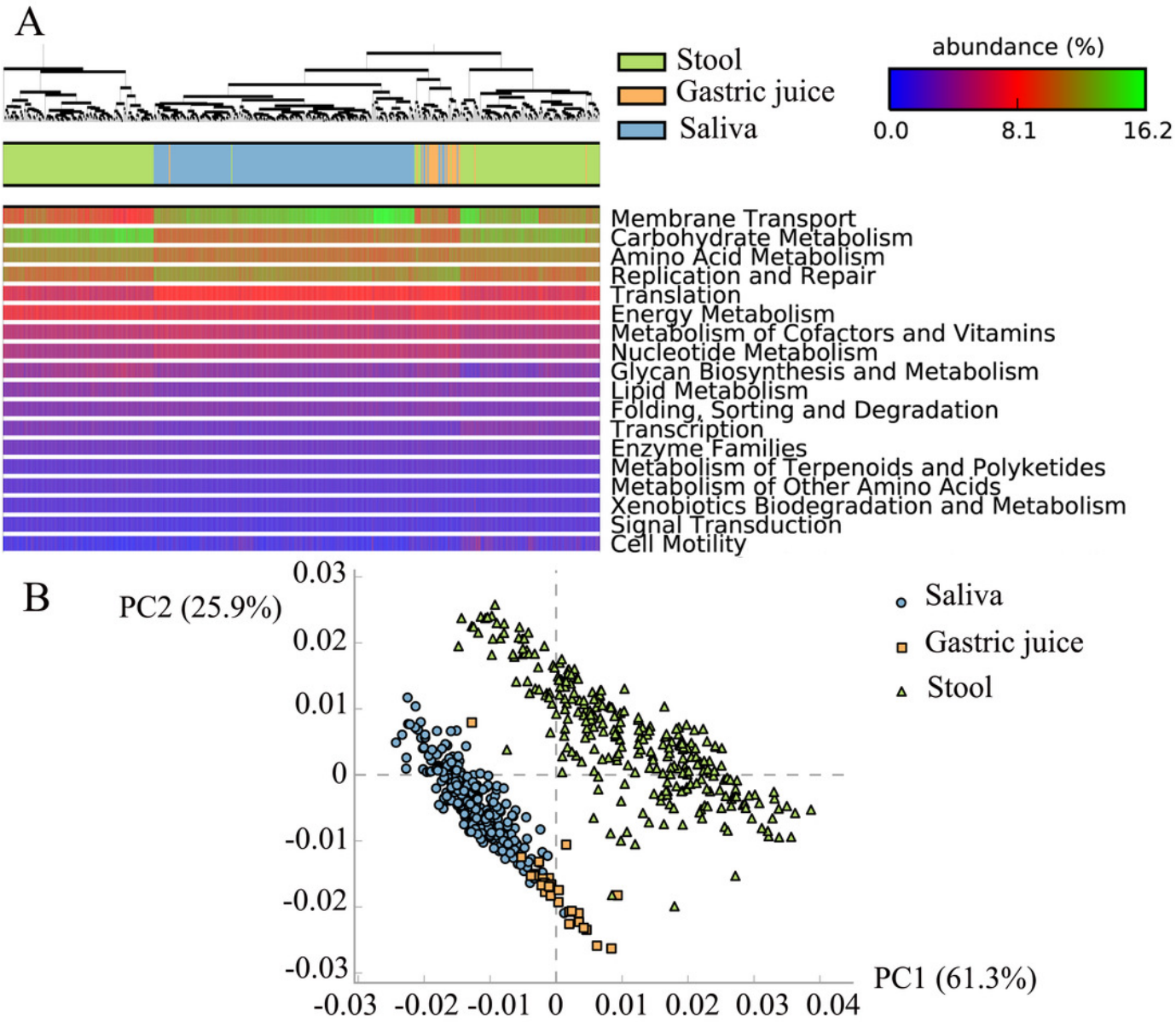


Table 1(on next page)

Comparison of bacterial phyla across the saliva, gastric juice and stool.

Comparisons were conducted by Kruskal-Wallis nonparametric ANOVA, and $P < 0.05$ after Bonferroni corrections (Bonferroni_P) were considered significant.

1

Phylum	Saliva mean	Gastric juice mean	Stool mean	Saliva vs. Gastric juice		Saliva vs. Stool		Stool vs. Gastric juice	
				P	Bonferroni_P	P	Bonferroni_P	P	Bonferroni_P
p__Actinobacteria	0.0404	0.0127	0.0021	2.24E-10	1.56E-09	1.44E-82	1.01E-81	1.12E-12	7.85E-12
p__Bacteroidetes	0.2187	0.4025	0.6732	9.46E-14	6.62E-13	2.63E-76	1.84E-75	1.52E-11	1.07E-10
p__Firmicutes	0.4099	0.0813	0.2843	1.18E-17	8.27E-17	4.70E-22	3.29E-21	2.41E-11	1.68E-10
p__Fusobacteria	0.0814	0.0992	0.0007	0.1567	1.0000	3.93E-94	2.75E-93	6.47E-35	4.53E-34
p__Proteobacteria	0.2405	0.3379	0.0283	3.56E-06	2.49E-05	1.90E-80	1.33E-79	8.40E-18	5.88E-17
p__Spirochaetes	0.0041	0.0191	4.07E-06	4.17E-11	2.92E-10	8.73E-67	6.11E-66	1.62E-64	1.14E-63
p__SR1	0.0006	0.0194	0.00E+00	3.28E-20	2.30E-19	7.56E-17	5.29E-16	1.58E-61	1.10E-60

2

3

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27030v1 | CC BY 4.0 Open Access | rec: 11 Jul 2018, publ: 11 Jul 2018

Table 2(on next page)

Comparison of bacterial genera across the saliva, gastric juice and stool.

Comparisons were conducted by Kruskal-Wallis nonparametric ANOVA, and $P < 0.05$ after Bonferroni corrections (Bonferroni_P) were considered significant.

Genus	Saliva mean	Gastric juice Mean	Stool mean	Saliva vs. Gastric juice		Saliva vs. Stool		Stool vs. Gastric juice	
				P	Bonferroni_P	P	Bonferroni_P	P	Bonferroni_P
g__Actinomyces	0.0231	0.0034	1.62E-05	4.95E-14	1.68E-12	5.95E-98	2.02E-96	2.05E-46	6.95E-45
f__Barnesiellaceae;g__	7.24E-05	0.0000	0.0125	0.1818	1	1.33E-28	4.51E-27	3.91E-06	0.000133
g__Bacteroides	0.0027	0.0003	0.4993	0.4540	1	2.35E-87	7.98E-86	3.01E-18	1.02E-16
g__Parabacteroides	0.0004	0.0000	0.0457	0.0058	0.1983	1.51E-71	5.13E-70	3.62E-15	1.23E-13
g__Porphyromonas	0.0465	0.0352	3.33E-05	0.2374	1	9.23E-97	3.14E-95	1.29E-41	4.38E-40
g__Prevotella	0.1422	0.3235	0.0302	4.03E-12	1.37E-10	6.74E-62	2.29E-60	1.46E-19	4.97E-18
f__Rikenellaceae;g__	0.0003	4.73E-06	0.0570	0.2055	1	5.83E-77	1.98E-75	2.08E-15	7.08E-14
f__Weeksellaceae;g__	0.0044	0.0126	9.57E-07	0.9821	1	1.75E-96	5.94E-95	1.51E-64	5.12E-63

g__Capnocytophaga	0.0190	0.0253	6.12E-06	0.1626	1	3.29E-97	1.12E-95	6.91E-55	2.35E-53
g__Granulicatella	0.0107	0.0028	3.34E-05	2.17E-11	7.39E-10	1.80E-97	6.11E-96	4.03E-49	1.37E-47
g__Streptococcus	0.1415	0.0299	0.0002	1.01E-14	3.43E-13	2.55E-90	8.66E-89	2.03E-23	6.90E-22
o__Clostridiales;f__;g__	0.0239	0.0024	0.0290	5.13E-08	1.74E-06	1.97E-06	6.71E-05	1.17E-12	3.99E-11
f__Lachnospiraceae;g__	0.0077	0.0015	0.0442	1.22E-12	4.13E-11	1.04E-49	3.55E-48	2.82E-17	9.59E-16
g__Oribacterium	0.0265	0.0011	6.39E-07	4.87E-17	1.65E-15	9.84E-99	3.34E-97	1.33E-62	4.51E-61
f__Ruminococcaceae;g__	0.0002	6.32E-05	0.0742	0.0002	0.0068	4.27E-91	1.45E-89	6.69E-18	2.28E-16
g__Faecalibacterium	0.0002	4.73E-06	0.0409	0.5383	1	1.01E-85	3.42E-84	7.58E-17	2.58E-15
g__Oscillospira	3.86E-05	0	0.0108	0.2801	1	4.72E-93	1.60E-91	6.74E-18	2.29E-16
g__Ruminococcus	7.78E-05	3.66E-06	0.0183	0.7821	1	1.23E-76	4.19E-75	4.65E-15	1.58E-13
g__Dialister	0.0026	0.0010	0.0100	0.0113	0.3835	0.0734	1	0.7902	1

g__Phascolarctobacterium	2.14E-05	3.80E-06	0.0099	0.3234	1	9.66E-42	3.28E-40	3.20E-07	1.09E-05
g__Selenomonas	0.0254	0.0025	0.0000	6.33E-10	2.15E-08	4.31E-96	1.46E-94	1.58E-61	5.36E-60
g__Veillonella	0.1262	0.0203	0.0008	1.55E-16	5.27E-15	1.75E-91	5.95E-90	1.00E-25	3.41E-24
g__Fusobacterium	0.0654	0.0869	0.0007	0.0573	1	1.23E-93	4.18E-92	3.91E-36	1.33E-34
g__Leptotrichia	0.0155	0.0117	2.27E-06	0.0839	1	1.28E-99	4.36E-98	2.92E-59	9.92E-58
g__Sutterella	9.17E-05	0.0000	0.0199	0.1197	1	2.79E-66	9.50E-65	3.94E-13	1.34E-11
g__Lautropia	0.0149	0.0052	6.48E-06	0.0136	0.4640	2.70E-83	9.19E-82	2.94E-52	9.99E-51
g__Neisseria	0.0581	0.1159	5.40E-06	1.77E-06	6.01E-05	5.22E-98	1.78E-96	2.94E-56	9.99E-55
g__Campylobacter	0.0198	0.0189	1.53E-05	0.1806	1	2.78E-73	9.45E-72	1.36E-53	4.63E-52
g__Helicobacter	4.99E-06	0.0097	0	1.09E-25	3.70E-24	0.1389	1	1.21E-32	4.10E-31
g__Actinobacillus	0.0096	0.0534	3.24E-06	3.22E-06	0.0001	7.78E-89	2.65E-87	8.74E-61	2.97E-59

g__Aggregatibacter	0.0143	0.0132	2.93E-05	0.6748	1	8.09E-97	2.75E-95	4.66E-56	1.59E-54
g__Haemophilus	0.1152	0.0826	0.0007	0.0289	0.9809	3.40E-91	1.16E-89	2.46E-25	8.37E-24
g__Treponema	0.0041	0.0190	0	4.17E-11	1.42E-09	2.14E-67	7.26E-66	1.45E-66	4.94E-65
p__SR1;c__;o__;f__;g__	0.0006	0.0194	0	3.28E-20	1.12E-18	7.56E-17	2.57E-15	1.58E-61	5.36E-60

1

2