bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Diversity and characteristics of the oral microbiome influenced by race and ethnicity

2 Qingguo Wang[1]*, Bing-Yan Wang [2], She'Neka Williams[3], Hua Xie[3]*

3 [1] School of Medicine, Meharry Medical College, Nashville, TN, USA

4 [2]School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX, USA

5 [3]School of Dentistry, Meharry Medical College, Nashville, TN, USA

6

7 *Corresponding authors: hxie@mmc.edu and qiwang@mmc.edu

8

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9 **Abstract**

10 Periodontitis disproportionately affects racial/ethnic populations. Besides social determinants contributing to
11 disparities in periodontal health, variations of oral microbial communities may also be a key factor influencing
12 oral immune responses. To characterize the oral microbiome from different racial/ethnic populations, we
13 collected 161 dental plaque samples from African Americans (AAs), Caucasian Americans (CAs), and Hispanic
14 Americans (HAs) with clinical gingival health or biofilm-induced gingivitis on an intact periodontium. Using
15 metagenomic sequencing, we found significant difference in diversity and abundance of microbial taxa in the
16 dental plaque samples from AA, CA, and HA groups and unique microbial species that can only be detected in
17 a particular racial/ethnic group. Moreover, we revealed racial/ethnic associated variations in functional potential
18 of the oral microbiome, showing that diversity and abundance of antibiotic resistant genes were greater in the
19 oral microbiome of the AAs than those in CAs or HAs, and that the AAs exhibited higher levels of genes
20 involving in modification of glycoconjugates, oligo- and polysaccharides. These findings indicate more
21 complex and higher virulence potential oral microbiome in AA and HA populations, which likely contributes to
22 higher prevalence of periodontitis in AAs and HAs.

23

24 **Importance**

25 Recognizing the variations in the oral microbiome among racial/ethnic populations offers insight into the
26 microbial determinants contributing to oral health disparities. In the study presented here, we found a higher
27 level of bleeding on probing (BOP), an indicator of tissue inflammatory response, in the AA group, which is
28 correspondence with a more complex oral microbiome detected in this group. Our observations suggest that the
29 variations of the oral microbiome associated with racial/ethnic backgrounds may directly relate to their
30 virulence potential including their abilities to induce host immune responses and to resist antibiotic treatment.
31 Therefore, these finding can be a stepping stone for developing precision medicine and personalized periodontal
32 prevention/treatment and for reducing oral health disparities.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34 **INTRODUCTION**

35 Periodontitis is a prevalent human disease affecting approximately 42% of adults aged 30 and older in the US.
36 The National Health and Nutrition Examination Survey (2009-2014) highlighted significant racial and ethnic
37 disparities in periodontitis, with higher rates observed in African Americans (AAs, 56.6%) and Hispanic
38 Americans (HAs, 59.7%) compared to Caucasian Americans (CAs, 37.0%)(1). These disparities are influenced
39 by multiple factors, including health care access and socioeconomic status (SES) (2, 3). A growing body of
40 evidence shows that human microbiome are dynamic microbial communities that shift with states of health (4).
41 Microbiome dysbiosis has been identified in a long list of diseases using meta-omic techniques, including
42 increased *Staphylococcus aureus* in the skin of atopic dermatitis (5) and decreased diversities and increased
43 Fusobacteria in the gut microbiome in patients with colorectal cancer (6). Interindividual differences of
44 microbiome are also influenced by genetics, environment, social determinants, and lifestyles (7, 8), which leads
45 to distinctive susceptible of population with different racial/ethnic backgrounds to diseases. A recent study
46 demonstrated a critical developmental window of gut microbiome variation at or shortly after 3 months of age,
47 which is driven by social status and environments and may lead to health disparities in adult (9). While oral
48 microbial dysbiosis is a recognized factor in oral diseases (10, 11), there remains a gap in understanding how
49 variations in the oral microbiome contribute to the disparities. Recent studies reported a differential oral
50 microbiota in the populations with oral malignant disorders and oral cancer when compared to that found in the
51 healthy counterparts (12-14). Although signatures of the oral microbiome associated with oral cancer has not
52 yet established, some bacterial taxa, including *Fusobacterium*, *Prevotella*, and *Porphyromonas*, were more
53 abundant in the oral cavities of patients with oral cancer. Population-level analysis of oral microbiome variation
54 was also investigated in racial and ethnic groups. A study by Yang et al. found a significant racial difference in
55 oral microbiome between African ancestry and European ancestry (15). The study of deep sequencing of 16S
56 rRNA genes revealed significantly increased abundances of four periodontitis-associated bacteria, including
57 *Porphyromonas gingivalis*, *Prevotella intermedia*, *Treponema denticola*, and *Filifactor alocis*, in mouth rinse

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58 samples of African ancestry, compared to those in the samples from European ancestry, although periodontal
59 status of the cohort in this study is not known.
60

61 We previously investigated microbiologic risk factors associated with periodontal health disparities using
62 qPCR. we determined and compared several key members of oral bacteria that play distinct roles in periodontal
63 health in AAs, CAs, and HAs. We detected much higher levels of *P. gingivalis* in the samples from the HA and
64 AA periodontitis patients than in samples from the CA patients, which appears to link to higher index of
65 bleeding on probing observed in the HAs and AAs with periodontitis (16). In addition, we examined *P.*
66 *gingivalis* level in the oral cavities of intact periodontium individuals with different racial/ethnic backgrounds.
67 Although significant difference of *P. gingivalis* distributions in these groups was not found using qPCR in our
68 previous study, the levels of *P. gingivalis* were higher in the HAs than that in the CAs (17). To further identify
69 population associated differential microbial profiles, we selected a total of 161 dental plaque samples from
70 AAs, CAs and HAs with intact periodontium for whole metagenome shotgun sequencing. We identified
71 significant difference in numbers of non-redundant bacterial genes and diversity and abundance of microbial
72 species among AA, CA, and HA groups. Additionally, several bacterial species were unique in a particular
73 racial/ethnic group. Moreover, functional potentials of the oral microbiome, such as antibiotic resistance and
74 LPS production, were higher in AA group than CA and HA groups. These results suggest that racial/ethnic
75 specific oral microbiome attributes to oral health disparities.

76

77 **RESULTS**

78

79 **Baseline characteristics of cohort**. In the previous studies, we determined detection rates of *P. gingivalis*,
80 using qPCR, among 340 individuals with intact periodontium with different racial/ethnic backgrounds (17).
81 Detection rates of *P. gingivalis* were 21.2% of the AAs, 18.2% of the CAs, and 30.6% of the HAs, respectively.
82 To further investigate variations of the oral microbiome associated with periodontal healthy disparities, we

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

83 selected dental plaque samples with *P. gingivalis* from the cohort and paired them with the *P. gingivalis*
84 negative samples with matched age, gender, and racial/ethnic background, resulting in 44 AAs (GpgA), 42 CAs
85 (GpgC) and 75 HAs (GpgH) (Table 1). GpgA, GpgC, and GpgH were further divided based on absence
86 (Gpg1A, Gpg1C, and Gpg1H) or presence (Group2A, Gpg2C, and Gpg2H) of *P. gingivalis* determined
87 previously by qPCR (17). Statistically significant differences were found in age and gender between AA and
88 HA and between CA and HA groups, but not between AA and CA (Table 1). Higher levels of BOP were found
89 in AAs, compared to CAs and HAs (Negative Binomial Regression, $p = 0.003$ after adjusted for covariates
90 including age and gender.
91
92 **Diversity and similarity of the oral microbiota among AA, CA, and HA groups**. A total of 2,025,809 non-
93 redundant genes were predicted from 161 dental plaque samples using MetaGeneMark. As shown in Fig. 1 and
94 Table 2, the median number of non-redundant genes per sample in AA groups were significantly higher than
95 those found in their counterparts CAs and HAs (Negative Binomial Regression, $p < 0.001$) after adjusting for
96 age and gender. There was no significant difference in number of non-redundant genes between CAs and HAs,
97 regardless of *P. gingivalis* detection with qPCR. In addition, unique genes were detected in each racial/ethnic
98 group. We found the highest number of unique genes (90,477) in the AA group followed by those (48,215 and
99 27,074) in HAs and CAs respectively (Fig. 1B). These finding suggest a more complex microbial composition
100 in AA population compared to HAs and CAs, which was also confirmed by taxonomic diversities among these
101 groups.
102
103 A total of 11,398 microbial species was identified in the 161 dental plaque samples using sequence or
104 phylogenetic similarity to microNR. An average of more than 3,000 species were predicted in each dental
105 plaque sample from AAs, which was significantly higher than those found in the samples from CAs and HAs
106 (Table 3). In addition, distinctive microbial species were identified in AAs, CAs, and HAs, which is consistent
107 with findings of unique genes among the groups. There were 501 microbial species found only in the AA group

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

108 compared to 153 and 310 detected in its counterparts of CAs and HAs, respectively (Fig. 2). Table 4 presents
109 the race/ethnic specific-species that were detected in 5 or more samples of each particular racial/ethnic group.
110 Interestingly, one bacterial species, *Pedobacter petrophilus*, was only detected in the samples from AAs with
111 higher level of *P. gingivalis* (Gpg2A). These results imply that specific microbial profiles contribute to
112 periodontal health disparities.
113

114 Notably, *P. gingivalis* were identified in all 161 samples with whole metagenome shotgun sequencing even
115 though *P. gingivalis* was detected in only 50% samples in this cohort using qPCR. However, samples (Gpg 2) in
116 which *P. gingivalis* can be detected using PCR had ten-fold higher abundance of *P. gingivalis* than those (Gpg
117 1) with *P. gingivalis* not detected using PCR (Table 5). Therefore, we re-designated Gpg 1 as the group with
118 low level of *P. gingivalis* and Gpg 2 as the high-level *P. gingivalis* group. In addition, *Tannerella forsythia*,
119 *Fusobacterium nucleatum*, *Streptococcus cristatus*, and *Streptococcus gordonii* were also detected in all 161
120 samples. *Treponema denticola* and *Filifactor alocis* were identified in 160 and 158 samples out of 161 samples,
121 respectively. We further determined the abundance of several well-studied bacterial species. We observed
122 significantly more abundance of *P. gingivalis* and *T. denticola* in the dental samples retrieved from AA and HA
123 groups compared to CA group (Table 6). Whereas, *F. alocis* and *T. forsythia* were detected significantly more
124 in HAs than CAs. We did not observe any significant difference in abundance of *F. nucleatum* and *S. cristatus*
125 among three groups. we also found a significantly higher number of S. gordonii in AAs than that in HAs, but
126 not between AAs and CAs.
127

128 To identify potential patterns of oral microbial profiles in different racial/ethnic groups, we projected the dental
129 plaque samples from high-dimensional space into a two-dimensional space using NMDS, a non-supervised
130 machine learning technique. NMDS retains the original sample rank-order similarity by measuring it using
131 Euclidean distance between the points that represent the samples. The visualization of the NMDS result is
132 provided in Fig. 3, which shows that the AA groups, Gpg1A and Gpg2A, are both located in the leftmost

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

133 section of the figure and are much closer to one another than CAs and HAs (Fig. 3), indicating that the
134 microbial profiles are more similar among the samples in the AA groups than those of the other two groups.
135 Additionally, both Gpg1H and Gpg2H, i.e., the HA groups, were projected toward the bottom right corner of
136 the figure, further away from the AA groups than the EA groups. More interestingly, the Gpg1 groups
137 (including Gpg1A, Gpg1C, and Gpg1H) occupy the upper section of the figure more than the Gpg2 groups (i.e.,
138 Gpg2A, Gpg2C, and Gpg2H), showing distinction between the two groups which were determined based on the
139 abundance of *P. gingivalis*.
140
141 **Functional diversities of the oral microbiome associated with levels of *P. gingivalis* and racial/ethnic**
142 **backgrounds.** Functional annotation was conducted by assembling metagenomic sequencing and mapping
143 against functional protein databases. Mapping against the Comprehensive Antibiotic Resistance Database
144 (CARD) allowed us to identified a total of 68 antibiotic resistance genes (ARGs) in the 161 dental plaque
145 samples (Fig. 4A). The median numbers of ARGs per sample were 40 in the plaque samples from the AA
146 group, 34 ARGs from the CAs, and 28 ARGs from the HAs. Additionally, more abundant ARGs (approximate
147 medium copies of 12,600) were also detected in the samples from the AAs, compared to 10,600 and 9,500 in
148 the samples from the CAs and the HAs, respectively (Fig. 4B). These data demonstrate more complex and
149 abundant ARGs were present in the oral microbiome of AAs compared to those in CAs and HAs, suggesting
150 that the oral microbiomes of AAs may have a higher ability to resist antibiotics. Moreover, Fig. 5 presents the
151 functions of the ARGs. The majority of ARGs act on antibiotic efflux and inactivation, and others are involved
152 in antibiotic target alteration, protection, and replacement. Most of the ARGs were found in *Bacillota*, a phylum
153 of mostly Gram-positive bacteria, *Pseudomonadota* including Gram-positive, negative, and variable bacterial
154 species, and *Bacteroidota*, a phylum of Gram-negative bacteria (Fig. 5), which presumably serve as reservoirs
155 for ARGs in the oral microbiome.
156

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

157 To determine the physiology of racial/ethnic-associated oral microbiome, we also analyzed the assembled
158 metagenomic protein sequences against the Carbohydrate-Active enZYmes Database (CAZy) (18). A cluster
159 heatmap shows the quantitation of the top 35 carbohydrate-active enzymes involving in the breakdown,
160 biosynthesis, and modification of glycoconjugates, oligo- and polysaccharides (LPS). The genes encoding these
161 enzymes were much more abundant in the oral microbiome of the AAs, compared to those in the CAs and HAs
162 (Fig. 6). This implies that Gram-negative bacteria in the oral microbial communities may be relatively higher in
163 AA population than in CAs and HAs, which can induce immune-inflammatory responses and initiates the onset
164 of periodontitis.
165
166 **DISCUSSION**

167 In this paper, we investigated variations of the oral microbial communities across races and ethnicities using
168 whole metagenomic sequencing. Recent studies of the oral microbiome, which mainly used less powerful
169 qPCR and 16S rRNA gene sequencing, found significant difference in microbial compositions between/among
170 racial/ethnic populations, including higher abundance of *P. gingivalis*, *T. forsythia*, *T. denticola*, and *F. alocis* in
171 AAs (15, 19-21). Our results on the cohort with intact periodontium are consistent with the previous reports
172 that some well-known periodontal pathogens including *P. gingivalis* and *T. forsythia* were more abundant in
173 AAs and HAs than those in CAs. However, comparing to previous qPCR-based studies that only detected *P.*
174 *gingivalis* at approximately 25% of samples from individuals without periodontitis (17, 22), we found *P.*
175 *gingivalis* in all tested dental plaques in this study with whole metagenomic sequencing, even though we
176 previously did not identified the *P. gingivalis* in half of the samples using qPCR. Moreover, *T. forsythia* was
177 also found in all samples, while detection rates of *T. denticola* and *F. alocis* were 99% and 98%, respectively.
178 These results are not in agreement with some other studies that reported much lower detection rates of these
179 bacteria using qPCR. For example, *T. forsythia* was detected in 10.5% of periodontally healthy participants
180 (32) and 68.0% of periodontitis patients (33). Moreover, our findings indicate that there is no difference in the
181 detection rates of most well-known periodontal pathogens among racial/ethnic groups. These results indicate

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

182 whole metagenomic sequencing is more sensitive than qPCR and 16S rRNA Sequencing for microbial detection
183 at the species level of all taxa including viruses. Therefore, detection rates of periodontal pathogens may not be
184 closely associated with periodontal health disparities, rather it is the levels of these pathogens in the oral
185 microbiome that contribute to the initiation and development of periodontitis. In addition, AAs, CAs, and HAs
186 differ in microbial diversities. We identified higher numbers of microbial taxa in the oral microbiome of AAs
187 and HAs than CAs and greater diversities of microbial compositions in AAs and HAs than CAs.
188 Racial/ethnic associated microbial species has not been reported to be associated with any human oral disease
189 so far. Interestingly, we identified that some bacterial species were only found in the oral microbiota of AAs,
190 such as *P. petrophilus,* which was clearly associated with higher levels of *P. gingivalis*. This bacterial species is
191 currently not easily cultured in laboratory, and hence difficult to analyze its relationship with other well-known
192 periodontitis pathogens and its role in the initiating periodontitis. Nevertheless, the unique presentence of this
193 bacterial species in the oral cavities of AA population may lead to a new research pathway linking to
194 periodontal health disparities.
195
196 One advantage of whole metagenomic sequencing is its ability to reveal information on functional potentials of
197 the oral microbiome, enabling us to compare antibiotic resistance and carbohydrate enzyme activities of the oral
198 microbiome among AAs, CAs, and CAs. Using metagenomic sequencing, we demonstrated more diverse and
199 abundant ARGs in the oral microbiome of the AA group, which may be an important determinant of virulence
200 that influence the ability of microorganisms to adapt and survive in the oral cavities. It remains to be
201 determined how complex and abundant ARGs evolve in the oral microbiome of the AA population. One
202 possible mechanism is that inappropriate antibiotic use was more common among AA patients compared to CA
203 patients (23). Another unique characteristic of the oral microbiome of AAs is its higher level of enzymes
204 involving in modification of glycoconjugates, oligo- and polysaccharides, which may be linked to higher level
205 of LPS. This observation is consistent with the levels of several Gram-negative periodontal pathogens
206 including *P. gingivalis*, *T. forsythia*, *T. denticola*, and *F. alocis*.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

207

208 Besides identification of functional potentials, other strengths of our study include a comprehensive periodontal
209 examination for each participant and taxonomic annotation coverage at every taxonomic level including the
210 species level. A previous study, using 16s rRNA sequencing, reported that African American had higher
211 microbial diversity in oral washes than Caucasians (15). By contrast, another study indicated that African
212 American adults had the lowest bacterial diversity in subgingival plaques while Chinese and Caucasian adults
213 had the highest diversity (24). Both of these studies lack detailed oral health parameters and used 16s rRNA
214 sequencing. It should be pointed out that one limitation of this study is based on self-reported race and ethnicity
215 data. However, previous studies suggested self-identified race and ethnicity of AAs and CAs are generally
216 reliable (25). In the present study, we observed significant difference in the oral microbiome among AA and
217 CA populations, and relatively less difference found between CAs and HAs, which may be due to HA
218 population with more complex genetic makeup. More accurate information on ancestry proportions of the
219 participants will be included in HAs for future health disparity studies.

220

221 In summary, this study shows significant differences in oral microbial diversity and abundance among AA, CA,
222 and HA groups. We also identified racial/ethnic specific bacterial species, such as *P. petrophilus* that was only
223 found in AAs with relatively higher *P. gingivalis* levels. Moreover, higher functional potential including
224 antibiotic resistance and LPS production were observed in the oral microbiome of AAs. These differences may
225 shape virulence potential of the oral microbiome and prepare microbiome as a whole to adapt their
226 environmental niches. Understanding and comparing racial/ethnic associated oral microbiome provide a
227 steppingstone for improving oral health disparities.

228

229 **MATERIALS AND METHODS**

230 **Study cohorts**

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

231 The research protocol was approved by the Committee for the Protection of Human Subjects of the University
232 of Texas Health Science Center at Houston (IRB number: HSC-DB-17-0636). Candidates were screened during
233 routine dental visits at the clinic of the School of Dentistry, University of Texas Health Science Center at
234 Houston between 2017 and 2022. Individuals aged 21-75 years with self-reported ethnicity/race of AAs, CAs,
235 or HAs were enrolled after the initial periodontal examination that included determination of plaque index (PI),
236 bleeding on probing (BOP) level, probing depth, and clinical attachment level on all teeth **(26)**. Radiographs
237 were taken during this screening phase to assess bone loss. The clinical oral examinations were performed by
238 faculty members of the School of Dentistry, University of Texas Health Science Center at Houston. The
239 examiners are calibrated annually in the diagnosis of periodontitis. Based on the 2017 World Workshop
240 classification (27, 28), all study participants diagnosed as clinical gingival health or biofilm-induced gingivitis
241 on an intact periodontium met the following criteria: >24 teeth; no alveolar bone loss or clinical attachment
242 loss; pocket depth ≤3mm (excluding pseudo pocket); no antibiotic therapy in the previous six months; and not
243 pregnant.

244

245 **Dental plaque sample collection**

246 Dental plaque samples, including supra and subgingival dental plaques, were collected by board-certified
247 periodontists using sterile paper points prior to any dental treatment and labelled numerically according to the
248 sampling sequences. The paper points were placed in the sulci of all the first molar in different quadrants for 1
249 min and then immersed immediately in an Eppendorf tube with 0.5 ml of Tris-EDTA (TE) buffer (pH 7.5) (29).
250 Bacterial pellets were harvested by centrifugation and then resuspended in 100 µl TE buffer. Samples were
251 stored at -80ºC until use.

252

253 **Sequencing and quality control**

254 Samples were sent to Novogene Co. (Sacramento, CA, USA) with dry ice for metagenomic sequencing.
255 Briefly, DNA extracted from human dental plaques was randomly sheared into short fragments, and the

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

256 resulting fragments were end-repaired, A-tailed, and ligated using Illumina adapters. The fragments with
257 adapters were subsequently amplified using PCR, followed by size selected and purified. Quality control of the
258 library was conducted using Qubit (⩾20 ng, ⩾10ng/μl), and quantification and size distribution detection were
259 performed using real-time and a bioanalyzer, respectively. The quantified libraries were pooled and sequenced
260 using an Illumina NovaSeq high throughput sequencer by Novogene Corporation, Inc., with a paired-end
261 sequencing length of 150 bp and an output of ~6 GB of raw data per sample.

262

263 **Data preprocessing**

264 The average size of raw data generated per sample was 6.394 GB. To ensure accuracy and reliability of the
265 subsequent data analysis, all low-quality bases (Q-value ≤ 38) that exceed certain threshold (40 bp), as well as
266 reads containing N nucleotides over 10 bp and those overlapping with adapters > 15 bp were trimmed. After
267 quality control, the size of clean data was 6.393 GB per sample, with 96.36% and 91.26% of bases having
268 quality scores greater than 20 and 30, respectively. To minimize host DNA contamination, raw reads that were
269 mapped to the human reference genome were discarded using the Bowtie2 software (30).

270

271 **Gene prediction and abundance analysis**

272 The MetaGeneMark software was used to predict open reading frames (ORFs) from scaftigs (> = 500bp) (31).
273 ORFs less than 100 nt were discarded. To generate gene catalogues, the remaining ORFs were dereplicated
274 using CD-HIT (32, 33) in default settings (i.e., identity = 95 %, coverage = 90 %). To calculate the gene
275 quantity, the clean data were mapped to the gene catalogue using Bowtie2 (parameters: –end-to-end, –sensitive,
276 –I 200, –X 400). Gene abundance (G_k) was calculated using the following formula:

$$G_k = \frac{r_k}{L_k} \frac{1}{\sum_{i=1}^{n} \frac{r_i}{L_i}}$$

277 where r represents the number of mapped reads and L represents gene length. Downstream analyses were
278 performed based on the abundance of the gene catalogues.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

279

280 **Taxonomy annotation**

281 The software tool DIAMOND (version 0.9.9.110) (34) was used to align the sequences of the identified genes
282 to those of bacteria, fungi, archaea, and viruses extracted from NCBI's NR database (version 2018-01-02).
283 MEGAN software was used to taxonomically annotate each metagenomic homolog (35). The sum of abundance
284 of genes annotated as a species in a sample was used as the abundance estimate of that species in that sample.
285 Based on the abundance of each taxonomic level, various analyses were performed, including heatmap of
286 abundance, principal coordinate analysis (PCoA), principal component analysis (PCA), and non-metric multi-
287 dimensional scaling (NMDS) analysis, which is an indirect gradient analysis approach that produces ordination
288 based on a distance matrix. R package ade4 (version 3.2.1) was used to perform PCA analysis and R package
289 vegan (version 2.15.3) was used for PCoA and NMDS analyses.

290

291 **Functional analysis**

292 The identified gene sequences were aligned to those in functional databases utilizing the DIAMOND software
293 (version 0.9.9.110) (34), with parameter settings: blastp, -e 1e-5. The functional databases used in this study
294 include Comprehensive Antibiotic Resistance Database (CARD) (36), Carbohydrate-Active Enzymes Database
295 (CAZy) (18), Kyoto Encyclopedia of Genes and Genomes (KEGG) (37), and Genealogy of Genes: Non-
296 supervised Orthologous Groups (eggNOG) (38). Based on the sequence alignment results, the best Blast hits
297 were selected for subsequent analysis and the relative abundance were calculated at different functional levels.

298

299 **Statistical analysis**

300 Statistical analyses were conducted using statsmodels (version 0.10.1) and SciPy (version 1.4.1), open-source
301 Python libraries for statistical modelling and scientific computing. Chi-square tests were performed to compare
302 gender of different racial/ethnic groups. Kruskal-Wallis tests were applied to compare ages of the sample
303 groups. For the total number of unique genes and taxa characterized in each sample, Negative Binomial

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

304 Regression was performed to adjust for age and gender while comparing the three racial groups. In addition, to
305 compare the abundance of a particular species in the three racial groups, linear regression was firstly performed
306 to fit the data, adjusting for covariates (age and gender), Kruskal-Wallis test was then applied to the residuals
307 extracted from the regression model to calculate p values. A p-value of < 0.05 was considered statistically
308 significant.

309

310 ACKNOWLEDGMENTS

311 The authors are grateful to all study participants for their contributions to this research. The authors thank
312 Krishna Kookal for abstracting clinical parameters from the Electronic Health Record at the School of
313 Dentistry, University of Texas Health Science Center at Houston.

314

315 The study was supported in part by grant U54MD007586 from the National Institute on Minority Health and
316 Health Disparities (NIMHD), USA; grant R16GM149359 from the National Institute of General Medical
317 Sciences (NIGMS), USA; grant UG3HG013248 from the National Human Genome Research Institute
318 (NHGRI), USA; and grant 1OT2OD032581 from the National Institutes of Health (NIH), USA. The work is
319 solely the responsibility of the authors and should not be interpreted as representing the official policies, either
320 expressed or implied, of the NIH.

321

322 H.X. contributed conception and design of the study. Q.W. and H.X. performed data analysis and prepared
323 Tables and Figures. B.W. prepared Table 1 and collected dental plaque samples. All authors were involved in
324 manuscript writing and revision.

325

326 ETHICS APPROVAL

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

327 The research protocol was approved by the Committee for the Protection of Human Subjects of the University
328 of Texas Health Science Center at Houston (IRB number: HSC-DB-17-0636). Consent forms were signed by all
329 participants, before sample collection.
330
331 **DATA AVAILABILITY**

332 Our metagenomic sequencing data are deposited at Sequence Read Archive (SRA) with accession number
333 PRJNA1160290.
334
335 **PATIENT CONSENT FOR PUBLICATION**

336 We give our consent for the publication of the manuscript including all tables and figures.
337
338 **CONFLICT OF INTEREST**

339 All authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication
340 of this article.
341
342
343

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

344 **REFERENCES**

345 1. **Eke PI, Thornton-Evans GO, Wei L, Borgnakke WS, Dye BA, Genco RJ.** 2018. Periodontitis in US
346 Adults: National Health and Nutrition Examination Survey 2009-2014. Journal of the American Dental
347 Association **149:**576-588 e576.

348 2. **Albandar JM.** 2024. Disparities and social determinants of periodontal diseases. Periodontology 2000.

349 3. **Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L,**
350 **Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D,**
351 **Tunyasuvunakool K, Wu Z, Zemgulyte A, Arvaniti E, Beattie C, Bertolli O, Bridgland A,**
352 **Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H,**
353 **Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A,**
354 **Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Zidek A, Bapst V, Kohli P,**
355 **Jaderberg M, Hassabis D, Jumper JM.** 2024. Accurate structure prediction of biomolecular
356 interactions with AlphaFold 3. Nature.

357 4. **Pflughoeft KJ, Versalovic J.** 2012. Human microbiome in health and disease. Annu Rev Pathol **7:**99-
358 122.

359 5. **Bjerre RD, Holm JB, Palleja A, Solberg J, Skov L, Johansen JD.** 2021. Skin dysbiosis in the
360 microbiome in atopic dermatitis is site-specific and involves bacteria, fungus and virus. BMC
361 microbiology **21:**256.

362 6. **Song M, Chan AT, Sun J.** 2020. Influence of the Gut Microbiome, Diet, and Environment on Risk of
363 Colorectal Cancer. Gastroenterology **158:**322-340.

364 7. **Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ,**
365 **Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L,**
366 **Lima-Mendez G, D'Hoe K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaudt L, Fu**
367 **J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J.** 2016. Population-level analysis of gut
368 microbiome variation. Science **352:**560-564.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

369 8. **Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, Herrema H, Tremaroli V, Bakker GJ,**
370 **Attaye I, Pinto-Sietsma SJ, van Raalte DH, Snijder MB, Nicolaou M, Peters R, Zwinderman AH,**
371 **Backhed F, Nieuwdorp M.** 2018. Depicting the composition of gut microbiota in a population with
372 varied ethnic origins but shared geography. Nat Med **24:**1526-1531.
373 9. **Mallott EK, Sitarik AR, Leve LD, Cioffi C, Camargo CA, Jr., Hasegawa K, Bordenstein SR.** 2023.
374 Human microbiome variation associated with race and ethnicity emerges as early as 3 months of age.
375 PLoS Biol **21:**e3002230.
376 10. **Kilian M, Chapple IL, Hannig M, Marsh PD, Meuric V, Pedersen AM, Tonetti MS, Wade WG,**
377 **Zaura E.** 2016. The oral microbiome - an update for oral healthcare professionals. British dental journal
378 **221:**657-666.
379 11. **Sedghi L, DiMassa V, Harrington A, Lynch SV, Kapila YL.** 2021. The oral microbiome: Role of key
380 organisms and complex networks in oral health and disease. Periodontology 2000 **87:**107-131.
381 12. **Lee WH, Chen HM, Yang SF, Liang C, Peng CY, Lin FM, Tsai LL, Wu BC, Hsin CH, Chuang**
382 **CY, Yang T, Yang TL, Ho SY, Chen WL, Ueng KC, Huang HD, Huang CN, Jong YJ.** 2017.
383 Bacterial alterations in salivary microbiota and their association in oral cancer. Scientific reports
384 **7:**16540.
385 13. **Wang K, Lu W, Tu Q, Ge Y, He J, Zhou Y, Gou Y, Van Nostrand JD, Qin Y, Li J, Zhou J, Li Y,**
386 **Xiao L, Zhou X.** 2016. Preliminary analysis of salivary microbiome and their potential roles in oral
387 lichen planus. Scientific reports **6:**22943.
388 14. **Burcher KM, Burcher JT, Inscore L, Bloomer CH, Furdui CM, Porosnicu M.** 2022. A Review of
389 the Role of Oral Microbiome in the Development, Detection, and Management of Head and Neck
390 Squamous Cell Cancers. Cancers (Basel) **14**.
391 15. **Yang Y, Zheng W, Cai Q, Shrubsole MJ, Pei Z, Brucker R, Steinwandel M, Bordenstein SR, Li Z,**
392 **Blot WJ, Shu XO, Long J.** 2019. Racial Differences in the Oral Microbiome: Data from Low-Income
393 Populations of African Ancestry and European Ancestry. mSystems **4**.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

394 16. **Wang BY, Lu T, Cai Q, Ho MH, Sheng S, Meng HW, Arsto L, Hong J, Xie H.** 2021. Potential
395 Microbiological Risk Factors Associated With Periodontitis and Periodontal Health Disparities.
396 Frontiers in cellular and infection microbiology **11:**789919.
397 17. **Wang BY, Cao A, Ho MH, Wilus D, Sheng S, Meng HW, Guerra E, Hong J, Xie H.** 2023.
398 Identification of microbiological factors associated with periodontal health disparities. Frontiers in
399 cellular and infection microbiology **13:**1137067.
400 18. **Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B.** 2009. The
401 Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic acids
402 research **37:**D233-238.
403 19. **Sabbagh S, Adatorwovor R, Kirakodu S, Rojas-Ramirez MV, Al-Sabbagh M, Dawson D,**
404 **Fernandes JG, Miguel MMV, Villasante-Tezanos A, Shaddox L.** 2024. Periodontal inflammatory
405 and microbial profiles in healthy young African Americans and Caucasians. Journal of clinical
406 periodontology **51:**895-904.
407 20. **Wang BY, Burgardt G, Parthasarathy K, Ho DK, Weltman RL, Tribble GD, Hong J, Cron S, Xie**
408 **H.** 2023. Influences of race/ethnicity in periodontal treatment response and bacterial distribution, a
409 cohort pilot study. Front Oral Health **4:**1212728.
410 21. **Renson A, Jones HE, Beghini F, Segata N, Zolnik CP, Usyk M, Moody TU, Thorpe L, Burk R,**
411 **Waldron L, Dowd JB.** 2019. Sociodemographic variation in the oral microbiome. Ann Epidemiol
412 **35:**73-80 e72.
413 22. **Griffen AL, Becker MR, Lyons SR, Moeschberger ML, Leys EJ.** 1998. Prevalence of
414 Porphyromonas gingivalis and periodontal health status. Journal of clinical microbiology **36:**3239-3242.
415 23. **Young EH, Strey KA, Lee GC, Carlson TJ, Koeller JM, Mendoza VM, Reveles KR.** 2022. National
416 Disparities in Antibiotic Prescribing by Race, Ethnicity, Age Group, and Sex in United States
417 Ambulatory Care Visits, 2009 to 2016. Antibiotics (Basel) **12**.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

418 24. **Mason MR, Nagaraja HN, Camerlengo T, Joshi V, Kumar PS.** 2013. Deep sequencing identifies
419 ethnicity-specific bacterial signatures in the oral microbiome. PloS one **8:**e77287.
420 25. **Lee YL, Teitelbaum S, Wolff MS, Wetmur JG, Chen J.** 2010. Comparing genetic ancestry and self-
421 reported race/ethnicity in a multiethnic population in New York City. J Genet **89:**417-423.
422 26. **Newman M, Takei, H., Klokkevold, H., & Carranza, F.** 2018. Newman and Carranza's Clinical
423 Periodontology.
424 27. **Chapple ILC, Mealey BL, Van Dyke TE, Bartold PM, Dommisch H, Eickholz P, Geisinger ML,**
425 **Genco RJ, Glogauer M, Goldstein M, Griffin TJ, Holmstrup P, Johnson GK, Kapila Y, Lang NP,**
426 **Meyle J, Murakami S, Plemons J, Romito GA, Shapira L, Tatakis DN, Teughels W, Trombelli L,**
427 **Walter C, Wimmer G, Xenoudi P, Yoshie H.** 2018. Periodontal health and gingival diseases and
428 conditions on an intact and a reduced periodontium: Consensus report of workgroup 1 of the 2017
429 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. Journal
430 of clinical periodontology **45 Suppl 20:**S68-S77.
431 28. **Trombelli L, Farina R, Silva CO, Tatakis DN.** 2018. Plaque-induced gingivitis: Case definition and
432 diagnostic considerations. Journal of clinical periodontology **45 Suppl 20:**S44-S67.
433 29. **Wang BY, Wu J, Lamont RJ, Lin X, Xie H.** 2009. Negative correlation of distributions of
434 Streptococcus cristatus and Porphyromonas gingivalis in subgingival plaque. Journal of clinical
435 microbiology **47:**3902-3906.
436 30. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357-359.
437 31. **Zhu W, Lomsadze A, Borodovsky M.** 2010. Ab initio gene identification in metagenomic sequences.
438 Nucleic acids research **38:**e132.
439 32. **Li W, Godzik A.** 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or
440 nucleotide sequences. Bioinformatics **22:**1658-1659.
441 33. **Fu L, Niu B, Zhu Z, Wu S, Li W.** 2012. CD-HIT: accelerated for clustering the next-generation
442 sequencing data. Bioinformatics **28:**3150-3152.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

443 34. **Buchfink B, Xie C, Huson DH.** 2015. Fast and sensitive protein alignment using DIAMOND. Nat
444 Methods **12:**59-60.
445 35. **Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M,**
446 **Batto JM, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jorgensen T, Brandslund I,**
447 **Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims**
448 **S, Zoetendal EG, Brunak S, Clement K, Dore J, Kleerebezem M, Kristiansen K, Renault P,**
449 **Sicheritz-Ponten T, de Vos WM, Zucker JD, Raes J, Hansen T, Meta HITc, Bork P, Wang J,**
450 **Ehrlich SD, Pedersen O.** 2013. Richness of human gut microbiome correlates with metabolic markers.
451 Nature **500:**541-546.
452 36. **Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong**
453 **H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding C,**
454 **Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J.** 2015.
455 Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell host &
456 microbe **17:**690-703.
457 37. **Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki**
458 **M, Hirakawa M.** 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic
459 acids research **34:**D354-357.
460 38. **Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldon T, Rattei T,**
461 **Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P.** 2014. eggNOG v4.0: nested orthology
462 inference across 3686 organisms. Nucleic acids research **42:**D231-239.
463

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Characteristics of the study cohort.

Characteristics	AAs	CAs	HAs	*p*-value
Gender (Male/Female)	20/24	16/26	17/58	0.027[+]
Age (year; Mean ± SD)	44.84±2.84	48.63±2.85	41.08±12.58	0.029[#]
BOP (%, Mean ± SD) [a]	34.92±5.23	17.30±3.11	22.53±18.29	0.006*
PI (%, Mean ± SD) [b]	61.04±5.87	50.12±5.24	44.94±25.10	0.093
Tooth number (Mean ± SD) [c]	26.72±0.47	27.04±0.35	27.17±2.22	0.963

[a] BOP: Bleeding on probing

[b] PI: Modified O'Leary plaque index

[c] Tooth number is based on a total of 32 teeth.

[+] The gender distribution of AA is significantly different from that of HA and there is no significant difference between AA and CA or between CA and HA (Chi-Square Test).

[#] The age distribution of AA is significantly different from other two groups which do not have significant difference in ages (Kruskal-Wallis test).

[*] The BOP of AA is significantly different from those of CA and HA which do not differ significantly in BOP (Negative Binomial Regression, adjusted for covariates Age and Gender).

464

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

465 Table 2. Non-redundant genes identified per samples in different
466 groups.

Comparison of sample groups	Median number of genes per sample	*P* - value*
Gpg1A vs. Gpg2A	445,282 vs. 513,094	0.596
Gpg1C vs. Gpg2C	178,020 vs. 235,183	0.371
Gpg1H vs. Gpg2H	171,843 vs. 167,841	0.884
Gpg1A vs. Gpg1C	445,282 vs. 178,020	4.34×10^{-3}
Gpg1A vs. Gpg1H	445,282 vs. 171,843	1.86×10^{-3}
Gpg1C vs. Gpg1H	178,020 vs. 171,843	0.862
Gpg2A vs. Gpg2C	513,094 vs. 235,183	1.38×10^{-2}
Gpg2A vs. Gpg2H	513,094 vs. 167,841	6.92×10^{-4}
Gpg2C vs. Gpg2H	235,183 vs. 167,841	0.295

467 **p* values were calculated using the Negative Binomial Regression model after adjusting for covariance.

468

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. The mean number of taxonomies identified per samples in different groups.

Comparison of sample groups	Mean number of species	*P* - value*
Gpg1A vs. Gpg2A	3,184 vs. 3,893	0.299
Gpg1C vs. Gpg2C	2,028 vs. 2,268	0.165
Gpg1H vs. Gpg2H	1,836 vs. 2,030	0.349
Gpg1A vs. Gpg1C	3,184 vs. 2,028	2.00×10^{-3}
Gpg1A vs. Gpg1H	3,184 vs. 1,836	4.19×10^{-4}
Gpg1C vs. Gpg1H	2,028 vs. 1,836	0.860
Gpg2A vs. Gpg2C	3,893 vs. 2,268	7.17×10^{-3}
Gpg2A vs. Gpg2H	3,893 vs. 2,030	8.26×10^{-5}
Gpg2C vs. Gpg2H	2,268 vs. 2,030	0.172

469 **P* values were calculated using the Negative Binomial Regression model after adjusting for covariance.

470

471

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

472

Table 4. Unique bacterial species found in AAs and HAs

Numbers of samples with the unique species					
AAs			HAs		
Species	Gpg1A	Gpg2A*	Species	Gpg1H	Gpg2H
Streptomyces sp. WAC04114	2	3	*Micrococcales bacterium*	3	5
Vagococcus hydrophili	5	0	*Megasphaera sp. DISK 18*	2	4
Veillonella sp. OK1	4	2	*Ruminococcus sp. NSJ-71*	3	3
Pedobacter petrophilus	0	7	*Ideonella azotifigens*	1	4
Hyphomicrobium sp. CS1GBMeth3	1	5	*Pseudomonas caricapapayae*	4	2
Halomonas zhangzhouensis	3	2	*Siphovirus Jomon_CT89*	2	4

*Gpg1 includes samples with the median abundance (FPKM) of of *P. gingivalis* (<5,741 FPKM), and Gpg2 had samples with the median abundance (FPKM) of *P. gingivalis* (>59,862 FPKM).

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

473

474

Table 5. Detection rates of well-known oral bacteria in different groups.

Sample group	Gpg1A	Gpg1C	Gpg1H	Gpg2A	Gpg2C	Gpg2H	Gpg1	Gpg2
Samples count	22	21	38	22	21	37	81	80
P. gingivalis	22	21	38	22	21	37	81	80
F. alocis	22	20	37	22	21	36	79	79
T. forsythia	22	21	38	22	21	37	81	80
T. denticola	22	21	37	22	21	37	80	80
F. nucleatum	22	21	38	22	21	37	81	80
S. cristatus	22	21	38	22	21	37	81	80
S. gordonii	22	21	38	22	21	37	81	80

475

476

477

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

478
479
480
481

Table 6. The average abundance (FPKM) of species per sample in different groups.

Sample group	GpgA	GpgC	*p*-value	GpgA	GpgH	*p*-value	GpgC	GpgH	*p*-value*
Porphyromonas gingivalis	159,821	82,068	0.002	159,821	112,880	0.015	82,068	112,880	0.021
Filifactor alocis	48,664	28,642	0.086	48,664	22,435	0.102	28,642	22,435	0.007
Tannerella forsythia	102,822	73,717	0.133	102,822	104,979	0.441	73,717	104,979	0.039
Treponema denticola	64,395	34,671	0.001	64,395	83,108	0.115	34,671	83,108	0.0003
Fusobacterium nucleatum	323,709	233,939	0.392	323,709	189,829	0.254	233,939	189,829	0.860
Streptococcus cristatus	26,989	28,667	0.749	26,989	29,825	0.908	28,667	29,825	0.712
Streptococcus gordonii	38,290	29,472	0.110	38,290	23,824	0.005	29,472	23,824	0.199

482 *P values were calculated by fitting a linear model (adjusting for covariance) and then applying Kruskal-Wallis
483 test to the residuals extracted from the model for the racial groups under comparison.

484

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

485 Figure legends

486

487 Fig. 1. Comparison of the number of non-redundant genes in dental plaque samples from different racial/ethnic
488 groups. (A) The violins represent the richness of non-redundant genes in the samples with low levels of *P.*
489 *gingivalis* in AAs (Gpg1A), CAs (Gpg1C), and HAs (Gpg1H) or high levels of *P. gingivalis* in AAs (Gpg2A),
490 CAs (Gpg2C), and HAs (Gpg2H). (B) Venn diagram of the total number of non-redundant genes identified in
491 the dental plaques samples from AAs (GpgA), CAs (GpgC), and (GpgH).

492

493 Fig. 2. Venn diagram of the microbial taxa at the species level identified in the racial/ethnic groups AAs
494 (GpgA), CAs (GpgC), HAs (GpgH).

495

496 Fig. 3. Visualization of NMDS analysis. Dots in a two-dimensional space represent dental plaque samples and
497 the distance between each pair of dots represents the dissimilarity between the corresponding two samples. The
498 samples in the same groups were assigned the same colors.

499

500 Fig. 4. Diversities and abundances of antibiotic-resistant genes (ARG). (A) Total number of ARG classes, and
501 (B) The abundances of ARGs in each sample. Each sample is presented as a red point and the number in each
502 boxplot represents median number of ARG classes and abundances in a group.

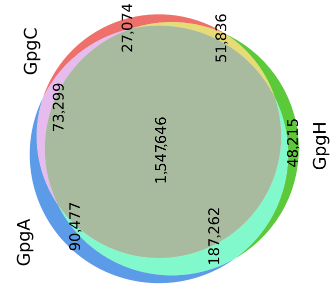
503

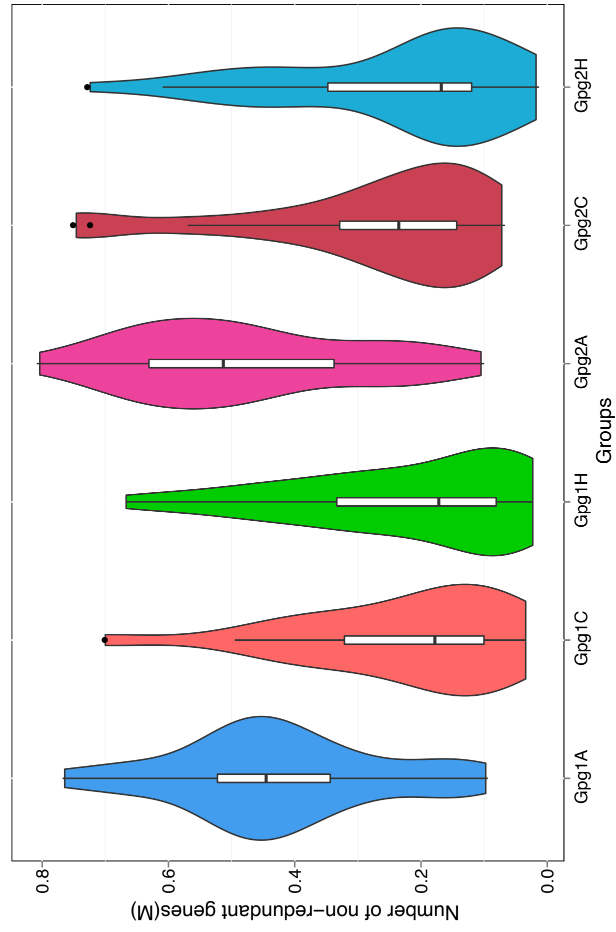
504 Fig. 5. Antibiotic resistance genes mechanism and the loop graph of species distribution. Circle chart is
505 divided into two parts, the right side shows the sample information, the left side shows the ARG tolerance of
506 antibiotic information. Inner circle different colors represent different samples and ARGs, scale for the relative
507 abundance (unit ppm). The left side is the sum of the relative abundance of the resistance genes in the sample,
508 the right side is the sum of the relative abundance of the resistance genes in each ARG. The left side of the outer

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

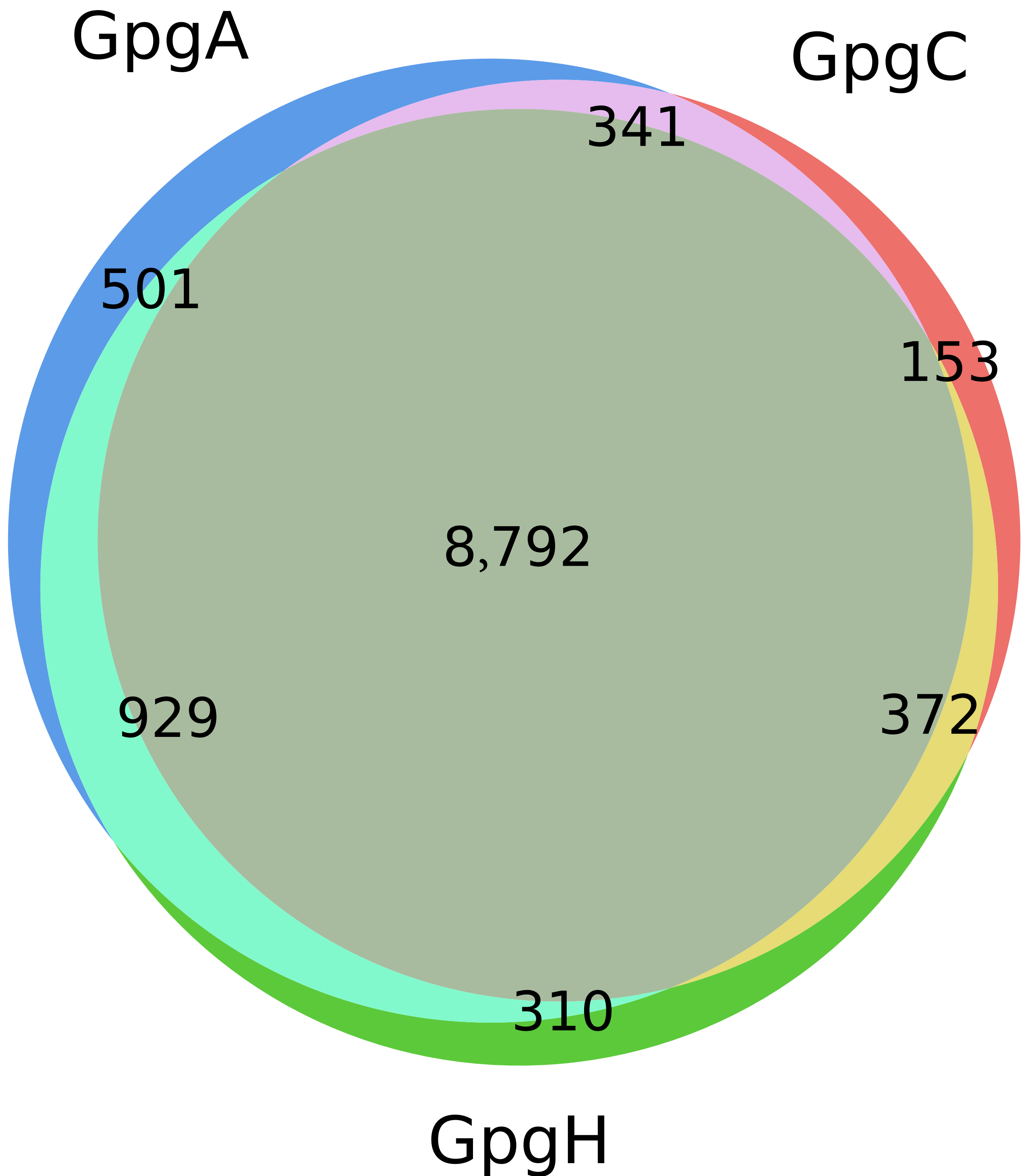
509 circle shows the relative percentage of the antibiotic to which the resistance gene belongs to, and the right side
510 of the outer ring shows the relative percentage of the sample in which the antibiotic resistance gene is located.
511
512 Fig. 6. Heatmap of carbohydrate-active enzyme genes identified in AA, CA, and HA groups. Columns
513 represent sample groups AA, CA, and HA and rows represent genes. The red color represents high gene
514 abundance to contrast with low abundance in blue color. The rows and columns are ordered based on the
515 correlations of z scores which were calculated based on gene abundance.
516

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

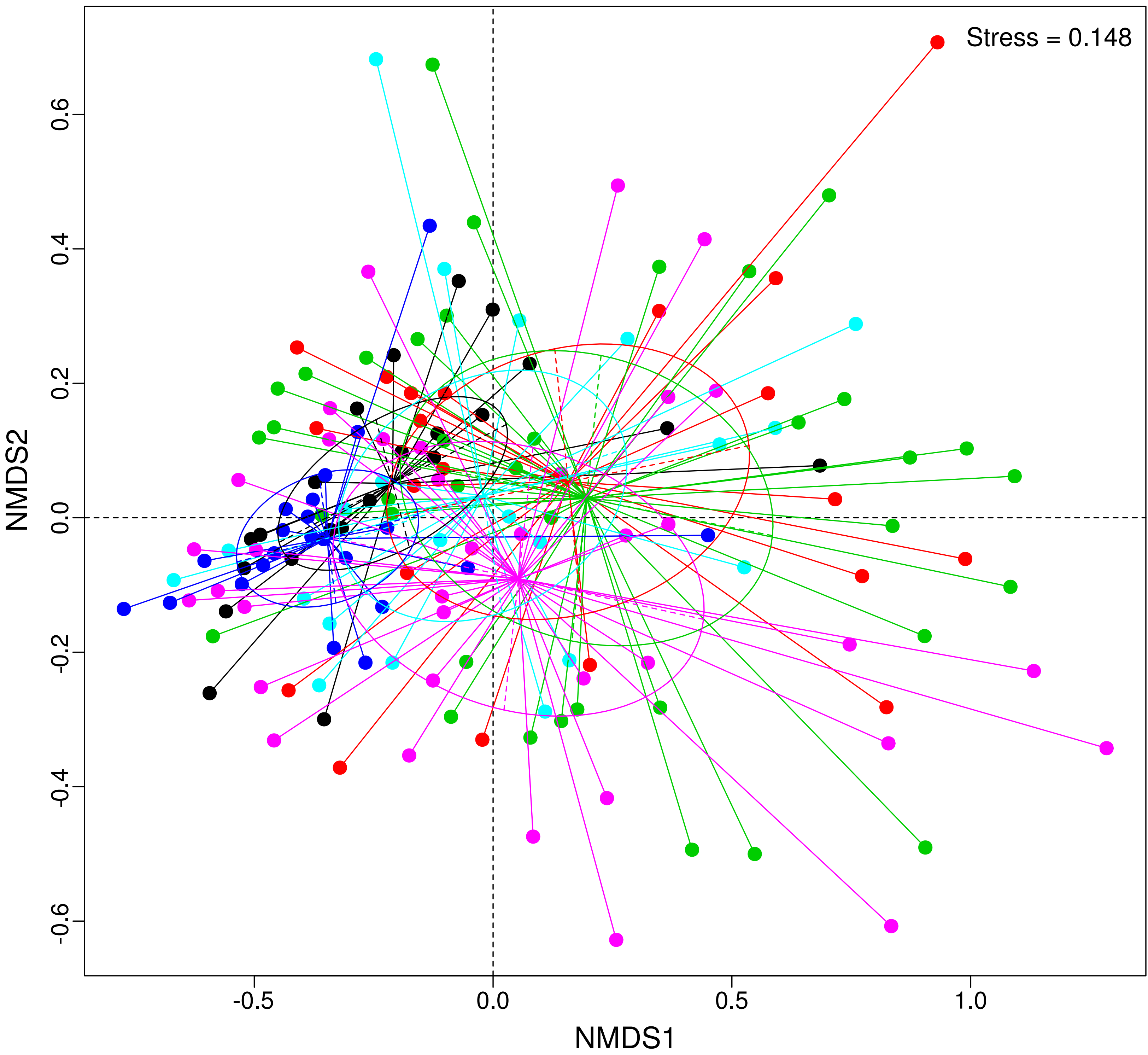




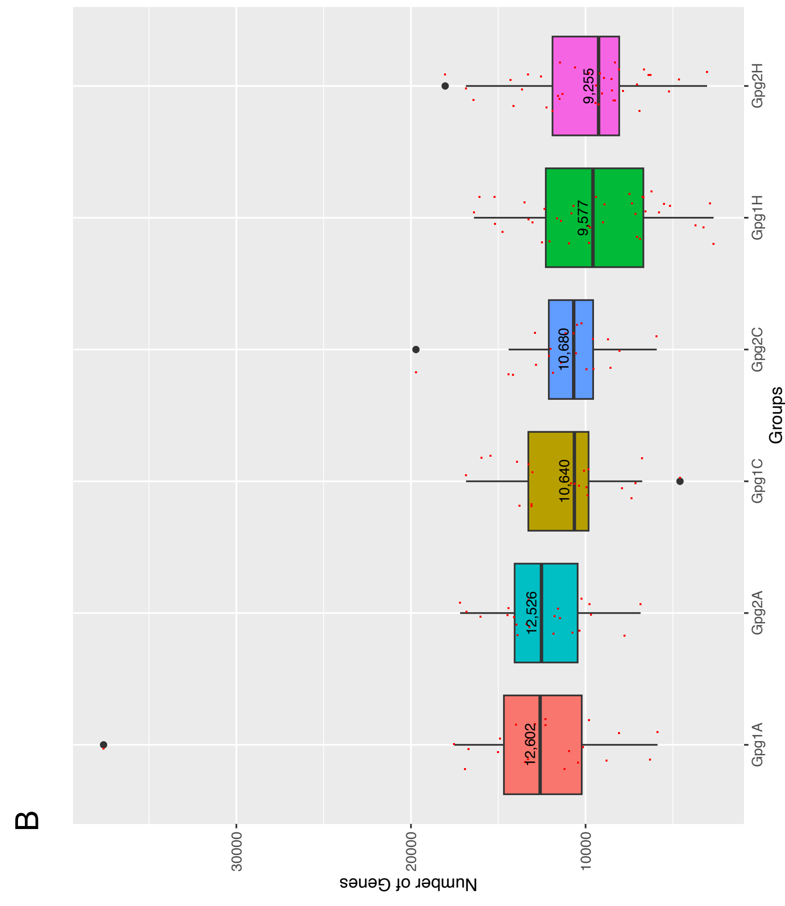
bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

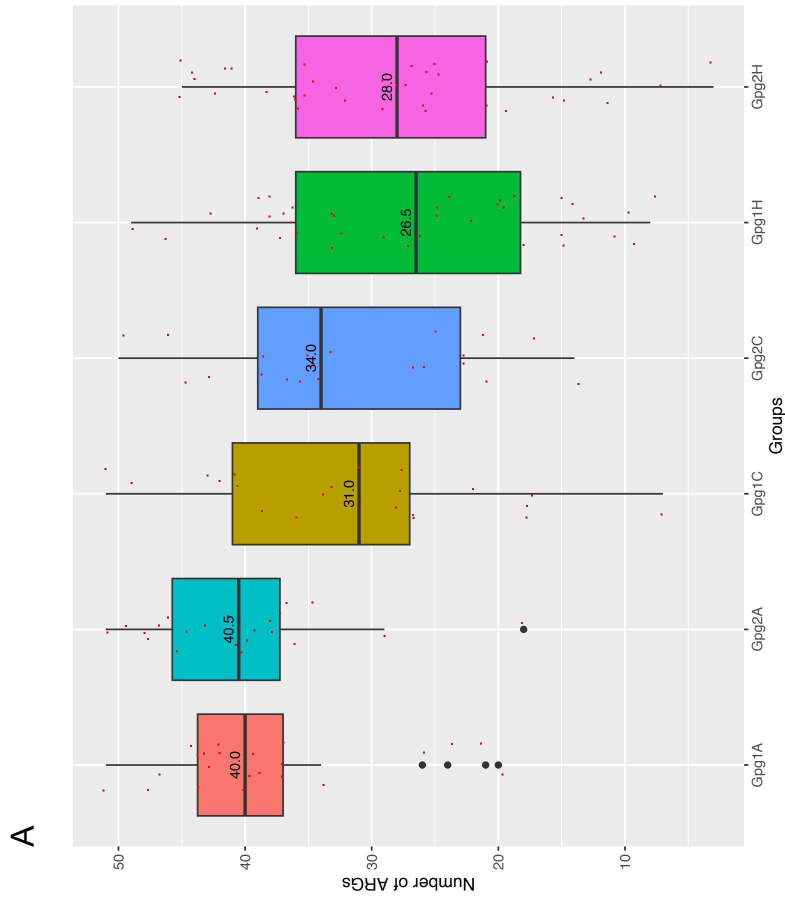


bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

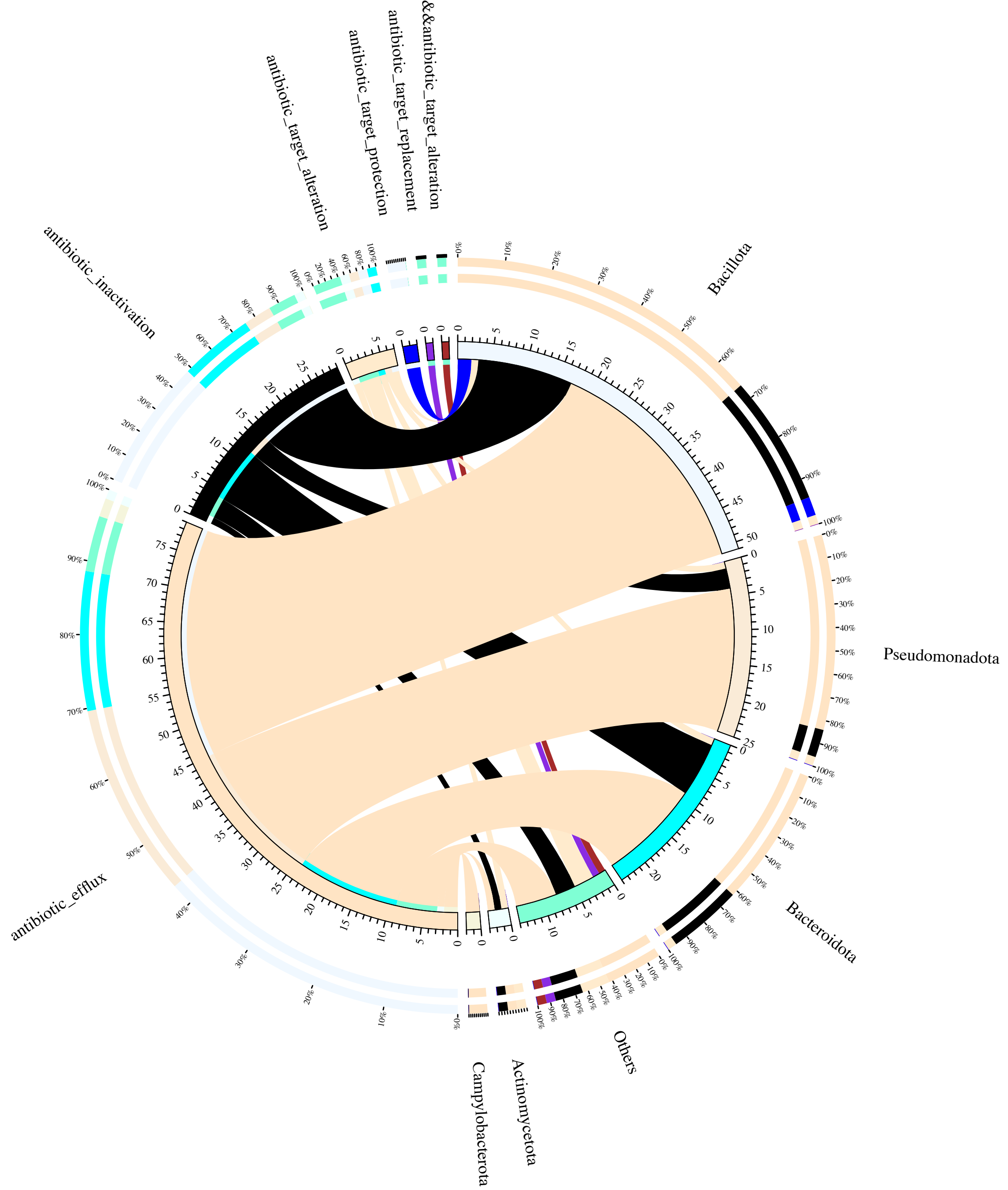


bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.





bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2024.10.07.617037; this version posted October 7, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.