bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

1 # Assessing the variation within the oral microbiome of healthy adults

2

3 Jacob T. Nearing[1], Vanessa DeClercq[2], Johan Van Limbergen[3,4], and Morgan G.I. Langille[1,5]

4 1. Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova
5 Scotia, Canada
6 2. Population Cancer Research Program, Dalhousie University, Halifax, Nova Scotia,
7 Canada
8 3. Division of Pediatric Gastroenterology & Nutrition, Emma Children's Hospital,
9 Amsterdam University Medical Centers, Amsterdam, The Netherlands
10 4. Dept of Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada.
11 5. Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia,

12 **Abstract:**

13 Over 1000 different species of microbes have been found to live within the human oral
14 cavity where they play important roles in maintaining both oral and systemic health. Several
15 studies have identified the core members of this microbial community, however, the factors that
16 determine oral microbiome composition are not well understood. In this study we exam the
17 salivary oral microbiome of 1049 Atlantic Canadians using 16S rRNA gene sequencing in order
18 to determine which dietary, lifestyle, and anthropometric features play a role in shaping
19 microbial community composition. Features that were identified as being significantly associated
20 with overall composition were then additionally examined for genera and amplicon sequence
21 variants that were associated with these features. Several associations were replicated in an
22 additional secondary validation dataset. Overall, we found that several anthropometric
23 measurements including waist hip ratio, height, and fat free mass, as well as age and sex, were

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

24 associated with oral microbiome composition in both our exploratory and validation cohorts. We
25 were unable to validate dietary impacts on the oral microbiome but did find evidence to suggest
26 potential contributions from factors such as the number of vegetable and refined grain servings
27 an individual consumes. Interestingly, each one of these factors on their own were associated
28 with only minor shifts in the oral microbiome suggesting that future biomarker identification for
29 several diseases associated with the oral microbiome may be undertaken without the worry of
30 confounding factors obscuring biological signal.

31 **Importance:**

32 The human oral cavity is inhabited by a diverse community of microbes known as the
33 human oral microbiome. These microbes play a role in maintaining both oral and systemic health
34 and as such have been proposed to be useful biomarkers of disease. However, to identify these
35 biomarkers, we first need to determine the composition and variation of the healthy oral
36 microbiome. Within this report we investigate the oral microbiome of 1049 healthy individuals
37 to determine which genera and amplicon sequence variants are commonly found between
38 individual oral microbiomes. We then further investigate how lifestyle, anthropometric, and
39 dietary choices impact overall microbiome composition. Interestingly, the results from this
40 investigation showed that while many features were significantly associated with oral
41 microbiome composition no single biological factor explained a variation larger than 2%. These
42 results indicate that future work on biomarker detection may be encourage by the lack of strong
43 confounding factors.

44

45 **Introduction:**

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

46 The human oral cavity is colonized by numerous bacteria, fungi, viruses and archaea that
47 make a rich microbial community known as the oral microbiome. This microbial community is
48 one of the most diverse sites of microbial growth within the human body being only secondary to
49 the colon (1). To date over 1000 different bacterial species have been found to colonize the oral
50 cavity (2) on various surfaces including the tongue, teeth, cheek, and gingivae (1). These
51 communities of microbes are responsible for various functions that can both maintain and
52 deplete oral health. For example, the presence of biofilms containing bacterial species such as
53 *Streptococcus mutans* and other aciduric bacteria can damage hard dental surfaces and lead to
54 dental caries (3, 4). Furthermore, the oral microbiome is known to play a role in a myriad of
55 other oral diseases including oral cancer (5), periodontitis (6, 7), and gingivitis (8, 9). In addition
56 to well-established associations between oral and cardiac health (10), recent work has also begun
57 to show that the oral microbiome may play a role in the health of other distal sites within the
58 human body. This includes diseases such as colorectal cancer (11, 12), pancreatic cancer (13),
59 prostate cancer (14), atherosclerosis (15) and inflammatory bowel disease (16).

60 Due to the associations between these diseases and the oral microbiome, its composition
61 has been proposed as a useful biomarker for human health and disease. With this in mind,
62 various studies have attempted to identify core members of the "healthy" oral microbiome (1,
63 17–20) to help aid in disease detection. These studies have uncovered that, at the genus level, the
64 oral microbiome remains relatively stable between individuals(1, 20) and across multiple
65 geographic locations(18, 21), but at deeper taxonomic resolutions, it can be variable. This
66 variability has indicated that dietary, anthropometric or sociodemographic factors may play a
67 role in shaping the oral microbiome(17, 19, 22–25). Various studies have focused on individual
68 factors that may cause shifts in the oral microbiome such as ethnicity(1, 25), alcohol

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

69 consumption(26), smoking(27), obesity(28, 29), and dietary patterns(30). However, to date only
70 a small number of studies have looked at the relative contributions of each of these factors to oral
71 microbiome variability in a single cohort. Takeshita et al., examined the oral microbiome of
72 2343 adults living in Japan using 16S rRNA gene sequencing and identified that higher
73 abundances of *Prevotella*, and *Veillnella* species were associated with old age, higher body mass
74 index (BMI), and poor overall oral health (19). Another study by Renson et al., in adults living in
75 New York city also found that variation in taxonomic abundances could be linked to marital
76 status, ethnicity, education and age (23). Further, work by Belstrøm et al., examined the oral
77 microbiome of 292 Danish individuals with low levels of dental caries and periodontitis using
78 microarrays and found that while socioeconomic status impacted oral microbiome profiles, diet,
79 BMI, age, and sex had no statistical impact on microbial abundances (22). This study, however,
80 was only able to identify the abundances of taxa that had a corresponding probe which, could
81 explain its disagreement with other work. Overall, these studies have indicated that both
82 biological differences such as sex and BMI as well as lifestyle and sociodemographic differences
83 can impact oral microbiome composition.

84 While these studies have shed light on the variation of the oral microbiome, it is currently
85 unclear to what extent these factors play a role in shaping the oral microbiome of an individual.
86 Without identifying the effect size of each of these factors relative to one another, it is difficult to
87 identify the correct variables that should be controlled for in case-control studies of the oral
88 microbiome. Furthermore, each of these studies have identified different taxa that are impacted
89 by various factors such as sex, BMI and age. This could be due to many factors, including
90 systemic bias introduced via the use of different protocols or differences in the studied cohorts.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

91 Therefore, the identification of microbes that are impacted by factors such as sex, BMI, or diet
92 could help identify potential interactions between the oral microbiome, health, and disease.
93 Herein, we report the variation within the healthy oral microbiome by examining 741
94 samples from non-smoking healthy individuals living within the Atlantic Provinces of Canada.
95 We then validated our results on a smaller subset of individuals (n=308) from the same cohort
96 (**Sup Fig 1**). The bacterial oral microbiome composition of these individuals was investigated
97 through 16S rRNA gene sequencing from saliva samples provided by each participant.
98 Compositions were then compared with 41 different variables including anthropometric, dietary
99 and sociodemographic factors. In this investigation, we determined which of these factors play a
100 role in shaping the oral microbiome and to what extent these factors can explain the overall oral
101 microbiome composition.
102
103
104
105 **Methods:**
106 Study design and population:
107 The current study includes the analysis of saliva samples from the Atlantic Partnership for
108 Tomorrow's Health (PATH) study. Atlantic PATH is part of the Canadian Partnership for
109 Tomorrow's Health (CanPath) project, a pan-Canadian prospective cohort study examining the
110 influence of environmental, genetic and lifestyle factors on the development of chronic disease
111 (31). The applicable provincial and regional ethics boards approved the study protocol and all
112 participants provided written informed consent prior to participation. The primary inclusion
113 criteria were that participants were aged 30-74 years at time of recruitment, a resident in one of

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

114 the Atlantic Canadian provinces (Nova Scotia, New Brunswick, Prince Edward Island, and
115 Newfoundland and Labrador). Recruitment and baseline data for all participating regions was
116 collected between 2000 and 2019. Details on participant recruitment and a descriptive cohort
117 profile have been published elsewhere (31). The questionnaire included sociodemographic
118 information, health information, behaviours, environmental factors, and self-reported
119 anthropometric information. Participants also had anthropometric measures (height, weight,
120 waist and hip circumferences, body composition, blood pressure, grip strength, and resting heart
121 rate) and biological samples (blood, urine, saliva, and toenails) collected. Approximately 9000
122 participants in the Atlantic PATH cohort provided a saliva sample. Participants were instructed
123 to refrain from eating, smoking, or chew gum for at least 30 minutes prior to oral specimen
124 collection. Oral samples (3 ml) were collected in sterile 50 ml conical tubes after rinsing with
125 water. Samples were stored at 4°C and batch shipped on ice to the central processing facility at
126 the QEII Health Sciences Centre in Halifax, Nova Scotia. Samples were processed within 24
127 hours of collection, aliquoted into cryovials and stored at -80°C until analysis.

128

129 The current analysis includes a total of 1214 saliva samples from healthy Atlantic Canadians
130 living within the provinces of Nova Scotia, New Brunswick, and Prince Edward Island. Based on
131 self-reported data, participants were defined as healthy if they had not been diagnosed with any
132 of the following conditions: hypertension, myocardial infarction, stroke, asthma, chronic
133 obstructive pulmonary disease, major depression, diabetes, inflammatory bowel disease, irritable
134 bowel syndrome, chronic bronchitis, emphysema, liver cirrhosis, chronic
135 hepatitis, dermatologic disease (psoriasis and eczema), multiple sclerosis,
136 arthritis, lupus, osteoporosis, and cancer. A total of 165 of these samples were removed due to

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

137 insufficient sequencing depth and of the remaining 1049, 308 were removed due to incomplete
138 answering of the 41 variables examined in this study. These 308 samples that were removed
139 were then used in validation analysis (details below) to confirm findings within the larger 741
140 participant cohort.

141

142 Socio-demographic, lifestyle and anthropometric variables:

143 Questionnaires were used to collect socio-demographic and lifestyle variables. Self-reported
144 variables included age, sex, education level, household income, rural/urban, province, dental
145 visits, sleep patterns, alcohol consumption, smoking status, and dietary variables such as food
146 avoidance, the use of specific types of fat/oil and artificial sweeteners, the frequency of dessert,
147 soda drinks, soy/fish sauce, seasoning with salt seasoning, and fast food, as well as servings of
148 vegetables, fruit, juice, whole grains, refined grains, dairy products, eggs, fish, tofu, beans, and
149 nuts/seeds. Anthropometric measures were collected by trained personnel in assessment centres.
150 Waist and hip circumferences were measured using Lufin steel tape. Height was measured by a
151 Seca stadiometer. Height and weight measures were used to calculate body mass index (BMI;
152 weight in kilograms divided by height in meters squared; kg/m^2). Body weight, fat mass, and
153 fat-free mass were measured using the Tanita bioelectrical impedance device (Tanita BC-418,
154 Tanita Corporation of America Inc., Arlington Heights, Illinois). Table 1 lists all variables that
155 were used for analysis.

156

157 Oral Microbiome 16S rRNA Sequencing:

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

158 Frozen saliva samples were thawed at room temperature and aliquoted into 96 well plates. DNA
159 from samples were then extracted using a QIAamp 96 PowerFecal QIAcube HT Kit following
160 the manufacturer's instructions using a TissueLyser II and the addition of Proteinase K.
161 Sequencing of the 16S rRNA gene was performed by the Integrated Microbiome Resource at
162 Dalhousie University. The V4-V5 region was amplified from extracted DNA in a PCR using 16S
163 rRNA gene V4-V5 fusion primers (515FB – 926R) (32) and high-fidelity Phusion polymerase.
164 Amplified DNA concentrations were then normalised and pooled together to be sequenced on an
165 Illumina MiSeq. Sequencing of samples was conducted over 6 Illumina MiSeq runs producing
166 300 base pair paired-end reads.
167

168 *16S rRNA Gene Sequence Processing:*

169 Primers were removed from paired-end 300 base pair sequences using cut adapt(33). Primer free
170 reads were then stitched together using the QIIME2 (v. QIIME2-2018.8)(34) VSEARCH(35)
171 join-pairs plugin. Stitched reads were then filtered using the QIIME2 plugin q-score-joined using
172 the default parameters. Quality filtered reads were then input into the QIIME2 plugin Deblur(36)
173 to produce amplicon sequence variants (ASV). A trim length of 360 base pairs and a minimum
174 number of reads required to pass filtering was set to 1. Amplicon sequence variants that were
175 found in an abundance of less than 0.1% of the mean sample depth (18) were then removed from
176 analysis. This is to keep inline with the approximate bleed-through rate on an Illumina MiSeq
177 sequencer. After filtering a total of 13248 ASVs were recovered. Representative sequences were
178 then placed into the Greengenes 13_8 99%(37) reference 16S rRNA tree using the QIIME2
179 (2019.7) fragment-insertion SEPP(38, 39) plugin . Rarefaction curves were then generated using
180 the QIIME2 alpha-rarefaction plugin and a suitable rarefaction depth of 5000 was chosen for

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

181 diversity analysis based on when the number of newly discovered ASVs came to a plateau (**Sup**
182 **Fig 2**). Representative sequences were then assigned taxonomy using a custom trained V4-V5
183 16S rRNA naive Bayesian QIIME2 classifier(40) trained on the 99% Silva V132 database(41).
184
185 Oral Microbiome Composition Analysis:
186 Taxonomic composition tables were generated using the QIIME2 taxa plugin and collapsed at
187 the genus level. All samples over 5000 reads in depth (1049) were subsampled to a depth of 5000
188 reads each and taxa that contributed less than a mean relative abundance of 1% were grouped
189 together under an "Other" category. The composition stacked bar chart was then generated in R
190 using ggplot2(42) and the x-axis was order based on the PC1 weighted Unifrac coordinates of
191 each sample.
192
193 Core Oral Microbiome Analysis:
194 Taxonomic tables subsampled previously at 5000 reads were collapsed at the genus and ASV
195 level using QIIME2. Genera/ASVs were removed at varying different sample presence cut-offs
196 and the remaining total mean relative abundance of non-filtered out genera/ASVs was then
197 calculated.
198
199 Oral Microbiome Alpha Diversity analysis:
200 Alpha diversity metrics were generated using QIIME2 (v2019.7) and the previously generated
201 tree containing both representative sequences and reference sequences. All samples were
202 subsampled to a depth of 5000 reads. Association between four different alpha diversity metrics
203 (Faith's Phylogenetic Diversity, Shannon, Evenness, Number of ASVs) were then tested using

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

204 general linear models while controlling for DNA extraction. A base model containing only DNA
205 extraction as a covariate and a testing modelling containing DNA extraction and the covariate of
206 interest were then compared using an ANOVA and p-values were recorded. Recorded p-values
207 were then corrected for false discovery (Benjamini and Hochberg(43)) with a chosen alpha of q
208 < 0.1.
209

210 Oral Microbiome Beta Diversity analysis:

211 Beta diversity metrics were generated using QIIME2 and the previously generated phylogeny.
212 All sequences were subsampled to a depth of 5000 reads based on the plateauing stage of
213 rarefaction plots (**Sup Fig 2**). Association between two different beta diversity metrics (weighted
214 UniFrac distance, Bray Curtis dissimilarity) were then tested using a PERMANOVA (adonis2
215 function in Vegan(44)) while controlling for DNA extraction. Marginal p values were then
216 corrected for false discovery (Benjamini and Hochberg) and an alpha value of q < 0.1 was
217 chosen. Significant features from univariate analysis were then included in a single multivariate
218 model that underwent backwards covariate selection, where each co-variation with the highest p-
219 value was removed from the model until all features were found to be significant. Additional
220 testing using adonis2 on fat free mass and height were done while controlling for both sex and
221 DNA extraction.
222

223 Differential abundance analysis:

224 Differential abundance analysis was conducted using the Corncob(45) (v 0.1.0) and
225 Phyloseq(46) R packages. A genus level taxonomic table was generated using QIIME2 (2019.7)
226 and genera that were not found in at least 10% of samples were removed. The fifteen covariates

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

227 that were found to be significantly associated to either weighted UniFrac or Bray Curtis
228 dissimilarities were chosen for testing. Testing of each covariate was done using the
229 "differentialtest" function in the Corncob package while controlling for differences in DNA
230 extraction and differential variability across DNA extraction and the covariate of interest.
231 Heatmaps were then constructed containing any genera/ASV that were significantly associated to
232 at least one of the covariates that were tested.
233

234 Validation analysis:

235 A total of 308 samples had not completely answered all 41 metadata variables of interest and
236 therefore were removed from the original analysis. This smaller cohort was used to test our
237 previous results by removing samples during testing of each covariate that had not answered that
238 question on the questionnaire. Both beta diversity analysis and differential abundance analysis
239 were carried out in the same manner as previously explained except for only testing features that
240 were previously identified as being significantly associated with that covariate/metric.
241 Furthermore, as there was previous evidence that these features were associated with that
242 covariate/metric, p-values were not corrected for false discovery but an alpha value of 0.05 was
243 chosen.
244

245 Results:

246 The Healthy Oral Microbiome is Stable at the Genus Level but Variable at Higher Resolutions:

247 We examined the oral microbiome composition of the overall cohort containing 1049
248 healthy individuals (**Sup Fig 1**) from Atlantic Canada to understand how anthropometric, socio-
249 demographic and dietary choices could alter oral microbiome composition. We found that 16

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

250 genera were found to have a mean relative abundance greater than 1% (**Fig 1A**) with *Veillonella*
251 having the largest mean contribution (21.49% +- 0.38%) followed by *Neisseria* (13.04% +-
252 0.40%), *Streptococcus* (11.86% +- 0.26%) and *Prevotella 7* (11.55% +- 0.24%).
253 To characterise the core relative abundance of core genera and ASVs within the oral
254 microbiome of these samples the mean relative abundance of genera/ASVs that were present in
255 greater than a specific percentage of samples was analysed. Interestingly, we found that at the
256 genus level the oral microbiome is relatively stable with 11 genera (**Sup Fig 3A)** present in
257 greater than 99% of all individuals making up on average a total relative abundance of 77.82%
258 (**Fig 1B**). However, this was not the case when we examined composition at a higher taxonomic
259 resolution. We then found that only 5.17% on average of the total relative abundance of the oral
260 microbiome was made up of 3 ASVs (**Sup Fig 3B**) shared between 99% of all participants in the
261 study (**Fig 1C**). These ASVs were classified as being in the *Granulicatella*, *Streptococcus*, and
262 *Gemelli* genera but could not confidently be assigned to a specific species.
263
264 Demographic, Anthropometric, and lifestyle choices have small but significant impacts on oral
265 microbiome composition
266 We examined the relationship of both alpha and beta diversity of the oral microbiome
267 between 41 different variables that described various demographic, lifestyle, and anthropometric
268 measures (Table 1). Samples were split into two different cohort based on whether they had
269 answered all 41 variables of interest. A total of 741 individuals answered all 41 variables and
270 were included in the exploratory cohort. From this cohort we did not find any significant
271 associations between any of the 41 variables tested and four different alpha diversity metrics
272 (Faith's PD, number of ASVs, Shannon, Evenness) after correction for multiple testing using

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

273 linear models that were adjusted for DNA extraction batch (**Sup file 1**)**.** We did, however, find
274 ten variables that were associated with differences in beta diversity as measured by both
275 weighted UniFrac (**Fig 2A**) and Bray Curtis dissimilarity (**Fig 2C**) (PERMANOVA, $q < 0.1$)
276 (**Sup file 2**). We found two additional variables that were only associated with weighted UniFrac
277 distances and three variables additional variables only associated with Bray Curtis dissimilarity
278 (PERMANOVA, $q < 0.1$). Principal component analysis of both the weighted UniFrac distances
279 and Bray Curtis dissimilarity of each sample revealed that anthropometric measures such as
280 height, weight, waist hip ratio, waist size, and fat free mass were all correlated in similar
281 directions along PC1, whereas features such as vegetable servings, age, and being female
282 correlated in opposite directions (**Fig 2C**). As sex plays an important role in determining the
283 height, fat free mass and waist hip ratio of an individual, we attempted to determine whether sex
284 was confounding our results from these variables. A separate analysis on weighted UniFrac
285 distances controlling for sex indicated that fat free mass (p=0.02, r2=0.0039) and waist hip ratio
286 (p=0.03, r2=0.0039), but not height (p=0.44, r2=0.0012) was significantly associated to
287 microbial composition despite differences in sex. Examining the amount of variation explained
288 by each variable by itself after controlling for DNA extraction showed small effect sizes for both
289 weighted UniFrac distances and Bray Curtis dissimilarities (R2 0.0030 - 0.009) (**Fig 2B, 2D**). Of
290 the features that were significant, sleeping light exposure explained the least amount of variation
291 in both weighted UniFrac distances (r2 = 0.0036) and Bray-Curtis dissimilarity (r2=0.0030). We
292 also found that fat free mass explained the largest amount of variation in both weighted UniFrac
293 (r2=0.009) and Bray Curtis dissimilarity (r2=0.006). In generally we found that the rankings of
294 effect sizes between these two different metrics agreed (Fig 2B, 2D). Also, the directionality of
295 each feature along PC1 and PC2 were similar between both weighted UniFrac and Bray Curtis

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

296 dissimilarity (Fig 2A, 2C). Examining each significant factor in our weighted UniFrac analysis
297 using a backward selected multivariate PERMANOVA, we found that 7.0% of total oral
298 microbiome variation could be explained by a total of 6 significant factors including DNA
299 extraction batch despite using the same protocol, equipment and personnel for each round (**Sup**
300 **Tab 1**). Interestingly, of these 6 factors DNA extraction number explained a considerable
301 amount of the variation alone (4.18%) (**Sup Table 1**). We found similar results examining beta
302 diversity variation using Bray Curtis dissimilarity with a slightly higher number of significant
303 features and lower total variation explained (5.87%) (**Sup Table 2**).
304
305 Various oral bacterial genera and ASVs are associated with anthropometric measurements, and
306 dietary choices in healthy individuals
307 We next decided to identify genera that were associated with the fifteen features previously
308 identified as being associated with beta diversity in either the weighted UniFrac or Bray Curtis
309 dissimilarity analysis. We found 42 genera (**Fig 3A**) and 42 ASVs (**Fig 3B**) that were
310 significantly associated with at least one of these features after controlling for DNA extraction.
311 We found that sex, height, and fat free mass shared similar genera and ASV associations. To
312 control for the possibility of sex confounding our height and fat free mass associations we
313 reanalysed the data controlling for sex. We found that no ASVs or genera were significantly
314 associated to fat free mass after controlling for sex and only 3 genera Chloroplast,
315 Burkholderiaceae unclassified and *Treponema 2* were significantly associated to height.
316 Interestingly two of these three genera were not previously associated to height in our initial
317 analysis. These results suggest that many of these features associated to height or fat free mass
318 may be driven by differences in sex. To test this, we also tested for differences in sex while

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

319 controlling for both fat free mass and height. Interestingly, we did not find any significantly
320 associated ASVs and only three significantly associated genera Defluvittaleaceae *UCG-011*,
321 *Leptotrichia*, and *Treponema 2*.
322 We did not find any other features that shared similar patterns of taxonomic associations
323 but there were multiple genera with multiple feature associations. The genus *Prevotella 7* had the
324 highest number of features (5) associated with its relative abundance including four
325 anthropometric measurements (height, fat free mass, waist size, waist hip ratio, and weight) and
326 sex. Interestingly, BMI did not have any genera or ASVs significantly associated despite many
327 other anthropometric measures showing strong taxonomic signals. We were unable to identify
328 any single ASVs associated to waist size and weight but were able to identify a small number of
329 genera including *Prevotella 7*, which was related to both and *Mogibacterium* with waist size. We
330 also found that for some phyla, all taxa with significant associations had the same effect size
331 direction. For example, genera in the Actinobacteria or Proteobacteria phyla tended to be
332 negatively associated with fat free mass, height and being male. We also found several genera in
333 the Proteobacteria phylum that were significantly associated with the amount of time since an
334 individuals last dental appointment.
335 In contrast, examining the ASVs associated with each feature we found that in a small
336 number of cases ASVs in the same genera had opposite directions of association to the same
337 features. For example, two ASVs classified as *Rothia* uncultured were both significantly
338 associated to age but in opposite directions suggesting that lower taxonomic resolution is
339 required to identify some associations. Furthermore, we also identified cases were ASVs that
340 were associated to a feature were classified in a genus that was found not to be related to that
341 feature. For example, ASV-4ca02 *Selenomonas* uncultured was strongly associated with being

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

342 male even though this entire collective genus was not (**Fig 3**). Further examples include ASV-
343 e2cc4 which was classified in the genus *Alysiella*, and significantly associated with reduced
344 refined grain servings. Examples of the opposite occurrence are also present with genera such as
345 *Mycoplasma* being associated with age but no single ASV for this associated could be identified.
346
347 Validation of diversity and differential abundance analysis:
348 To help validate our findings we analyzed an additional 308 samples from a smaller
349 subset of the Atlantic PATH cohort that had not completely answered all 41 variables of interest.
350 We found that associations between beta diversity and anthropometric features such as height,
351 weight, waist hip ratio, and fat free mass were recoverable within our smaller cohort (**Table 2,**
352 **Sup Fig 4**). Furthermore, we also found that the associations between age and sex with oral
353 microbiome composition were also recoverable, validating our previous analysis. We were
354 unable to recover any significant dietary associations within this smaller validation cohort. We
355 also were unable to recover associations between lifestyle variables such as sleeping light
356 exposure or the time since an individuals last dental visit. The inability to recover these
357 differences could have been due to the highly reduced sample size within this validation cohort.
358 We further validated our differential abundance analysis using this cohort and found 8/17
359 genera associated with sex, 8/16 genera associated with fat free mass, 5/15 genera associated
360 with height, and 3/11 genera associated with age were recoverable within this smaller cohort.
361 Additionally, the negative association between *Prevotella 2* and waist hip ratio was also verified
362 within this cohort. Furthermore, several associations between ASVs and features such as sex
363 (5/14), height (4/12), fat free mass (2/3) and sleeping light exposure (1/2) were also found within
364 this smaller validation cohort. All significant effect sizes that were recovered in the validation

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

365 cohort except for one, between sleeping light exposure and ASV-d4746 *Streptococcus*, remained
366 in same direction as the original cohort indicating relationships that were robust to sample
367 choice.

368 **Discussion:**

369 Our analysis of 1049 healthy (**Sup Fig 1**) individuals from Atlantic Canada revealed that
370 much of the oral microbiome of Atlantic Canadians was made up of eleven "core" genera that
371 belong to six different phyla (*Actinobacteria*, *Fusobacteria*, *Proteobacteria*, *Firmicutes*,
372 *Bacteroidetes*, and *Fusobacteria*). Interestingly some of these core genera found in 99% of all
373 samples were found in relatively low abundance (<2% mean abundance) indicating that bacteria
374 within the oral microbiome can be consistently observed with minor contributions. In contrast,
375 the composition at the ASV level had only 3 ASVs being present in 99% of samples and only
376 contributing 5.17% of the total oral microbiome composition on average. Overall, these results
377 indicate that individuals tend to share similar genera within the oral cavity, but the species/strains
378 shared between individuals is highly variable. These findings are inline with previous work from
379 the Human Microbiome project that found the oral microbiome to be relatively stable at the
380 genus level(1).

381 We found that various anthropometric and lifestyle features were significantly associated
382 with oral microbiome composition, however, they explained only a small amount of total oral
383 microbiome variance while controlling for DNA extraction batch (5.87-7.00%). We found that
384 fat free mass explained the highest amount of variance (0.6-0.9%) of all biological features.
385 While this feature had many differential abundant genera and ASVs associated with it, we were
386 unable to recover any of them after controlling for differences in sex. This could indicate that
387 these associations could be driven by sex and not underlying fat mass, however, we were also

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

388 unable to recover many relationships between sex and taxonomic abundance while controlling
389 for fat free mass indicating that both of these factors significant confound the other. However,
390 despite this issue there is previous evidence to suggest that some bacteria are related to
391 differences in body size. A study in children found reduced abundance of *Veillonella*, *Prevotella*,
392 *Selenomonas* and *Streptococcus* in obese children (47). Interestingly, in our adult population we
393 found similar trends with members of the *Veillonella* family being positively associated with
394 increasing fat free mass, and members of the *Provetella* genus also being linked with higher fat
395 free mass. Another publication on the Southern Community Cohort Study found that both
396 *Granulicatella* and *Gemella* were associated with obesity (29) which we also found within our
397 cohort at both the genus and ASV level. One interesting result from our study was our inability
398 to identify any genera or ASVs linked to body mass index, despite numerous relationships
399 between anthropometric measurements being identified. These results indicate that future studies
400 should be advised to include sex and other measurements of body composition, such as lean
401 body mass, when looking at relationships between the microbiome and obesity.
402 We found two genera *Defluviitaleaceae* UCG-011 and an uncultured genus from
403 *Veillonellaceae* which were strongly associated with being male. However, neither of these
404 associations were recovered in our validation cohort indicating that they could either be false
405 positives or require a larger sample size to recover. Despite this we were still able to recover
406 eight genus level associations in our validation cohort, however, only a few of these associations
407 match those that were previously reported. Renson et al. found two genera *Lactobacillus* and
408 *Actinobacillus* to be higher in males, which we did not find within our study (23). This could
409 have been due to multiple differences including sampling procedures or systemic protocol bias.
410 Raju et al. found that there was a high relative abundance of *Haemophilus* in females which we

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

411 also found in our study, however they also found *Oribacterium* to be increased in females which
412 was opposite from what was found in this study (47). Differences between these studies and ours
413 can in part be attributed to differences in samples collection and sequencing primers used
414 highlighting the importance of standardizing protocols within the field.

415 We were unable to recover any relationships between dietary features within our
416 validation cohort, however, refined grain servings per day had the largest impact on oral
417 microbiome composition in our initial analysis. During this initial analysis, we found that
418 bacteria from four genera *Bergeyella*, *Parvimonas, Veillonella,* and *Neisseria* decreased in
419 relative abundance with increasing refined grain intake. Interestingly, refined grain intake had a
420 very strong association with inflammatory bowel disease in a previous analysis of this
421 cohort(48), and alterations in the oral microbiome have been linked to inflammatory bowel
422 disease in the past (49). Previous work by Said et al., found multiple genera in differential
423 abundance between individuals with and without IBD including the increased presence of
424 *Veillonella* (16), which we found to be linked positively with refined grain intake.

425 Other dietary factors we found linked to oral microbiome composition in our original
426 analysis include both juice servings and vegetable servings. However, were only able to find a
427 small number of genera and taxa linked to vegetable serving intake and juice serving intake.
428 Furthermore, we were unable to recover these effects in our validation cohort indicating the
429 possibility for a false positive or the requirement of a large sample size to see these effects.
430 Previous work within the field has found conflicting evidence on the role of diet impacting oral
431 microbiome composition. This previous evidence along with the inability to recover these
432 relationships within our validation cohort indicates that diet may only have a small impact on

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

433 oral microbiome variation, and that these effects require large samples to recover them or that
434 different dietary capture methods have a strong influence on the observed results.

435 Looking at all features that were significantly associated to oral microbiome composition
436 together in a single model we were only able to explain a small portion of the total variance
437 between samples (6.75-6.92%). This indicates that while many of these features are significantly
438 related to microbial composition each one by themselves tends to only cause small shifts in
439 overall microbial composition. Furthermore, a majority of the variance accounted for was due to
440 differences in DNA extraction date. This shows that while slight technical variations such as the
441 time when DNA extraction was done can have larger impacts on sample composition
442 emphasizing the need to control for these technical variations during large population-based
443 studies.

444 One large limitation to our study was our lack of detailed dental history information from
445 participants. While we did record how recently each individual last visited the dentist, we were
446 unable to retrieve detailed information on dental health, which has been found to have dramatic
447 impacts on oral microbiome composition (19). This could explain some of the missing variation
448 that was not accounted for in our study, however, it is unlikely to explain all 93.25% indicating
449 we are still missing a suitable amount of information on what determines an individual's oral
450 microbiome composition.

451 In conclusion, our study indicates that the healthy oral microbiome is relatively stable
452 between individuals at the genus level and is impacted very little by any one factor. Future
453 studies that attempt to identify oral microbial biomarkers associated with disease may be
454 encouraged by the lack of major confounding variables and may be justified in controlling only
455 for sex, body composition, oral health, and basic dietary information.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

456

457 **Availability of data and materials:**

458 All sequencing data has been uploaded to the European Nucleotide Archive and is
459 available under the accession number PRJEB38175. Code used to analysis all data is available at
460 https://github.com/nearinj/Nearing_et_al_2020_Oral_Microbiome. De-identified metadata used
461 in this project can be accessed by contacting the Atlantic Partnership for Tomorrow's Health
462 project.

463

464 **Funding:**

465 JTN is supported by both a Research Nova Scotia, Scotia Scholars award (2019-2022) as
466 well as a Nova Scotia Graduate Scholarship (2019-2023). JVL was supported by a Canadian
467 Institutes of Health Research (CIHR)-Canadian Association of Gastroenterology-Crohn's Colitis Canada
468 New Investigator Award (2015–2019), a Canada Research Chair Tier 2 in Translational Microbiomics
469 (2018-2019) and a Canadian Foundation of Innovation John R. Evans Leadership fund (awards #35235
470 and #36764), a Nova Scotia Health Research Foundation (NSHRF) establishment award (2015–2019), an
471 IWK Health Centre Research Associateship, a Future Leaders in IBD project grant, a donation from the
472 MacLeod family and by a CIHR-SPOR-Chronic Diseases grant (Inflammation, Microbiome, and
473 Alimentation: Gastro-Intestinal and Neuropsychiatric Effects: the IMAGINE-SPOR chronic disease
474 network). The data used in this research were made available by the Atlantic Partnership for
475 Tomorrow's Health (Atlantic PATH) study, which is the Atlantic Canada regional component of
476 the Canadian Partnership for Tomorrow's Health Project funded by the Canadian Partnership
477 Against Cancer and Health Canada. The views expressed herein represent the views of the
478 authors and do not necessarily represent the views of Health Canada.

479

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

480 **Acknowledgements:**

481 This research has been conducted using Atlantic PATH data and biosamples, under application
482 #2018-103. We would like to thank the Atlantic PATH participants who donated their time,
483 personal health history and biological samples to this project. We would also like to thank the
484 Atlantic PATH team members for data collection and management.

485

486

487 Table 1: Cohort characteristic and variables compared against oral microbiome composition. NA
488 represents responses of prefer not to answer or missing data.

	Overall
Number of participants	1214
Rural/Urban (%)	
Urban	1050 (86.5)
Rural	126 (10.4)
NA	38 (3.1)
Province (%)	
New Brunswick	124 (10.2)
Nova Scotia	1070 (88.1)
Prince Edward Island	16 (1.3)
NA	Data repressed
Economic Region	
Annapolis Valley	52
Cape Breton	142
Edmundston – Woodstock	Data repressed
Fredericton – Oromocto	44
Halifax	773
Moncton – Richibucto	32
North Shore	41
Prince Edward Island	16
Saint John – St., Stephen	45
Southern Shore	28
Sex (%)	

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

Female	846 (69.7)
Male	368 (30.3)
Body mass index (mean (SD))	27.30 (4.55)
Waist size (mean (SD))	90.96 (12.79)
Hip size (mean (SD))	104.29 (9.45)
Waist hip ratio (mean (SD))	0.87 (0.08)
Height (mean (SD))	167.06 (8.90)
Weight (mean (SD))	76.39 (14.99)
Age (mean (SD))	55.39 (7.80)
Fat mass (mean (SD))	25.26 (9.55)
Fat free mass (mean (SD))	51.05 (10.87)
Body fat percentage (mean (SD))	32.68 (8.61)
Vegetable servings (mean (SD))	2.56 (1.98)
Fruit servings (mean (SD))	2.00 (1.45)
Juice servings (mean (SD))	0.69 (0.95)
Whole grain servings (mean (SD))	2.11 (1.43)
Refined grain servings (mean (SD))	0.67 (0.86)
Milk product servings (mean (SD))	2.04 (1.29)
Egg servings per week (mean (SD))	3.25 (2.68)
Meat/poultry servings (mean (SD))	1.53 (1.35)
Fish servings (mean (SD))	0.51 (0.67)
Tofu servings (mean (SD))	0.04 (0.18)
Bean servings (mean (SD))	0.36 (0.55)
Nut/seed servings (mean (SD))	0.69 (0.68)
Dessert Frequency (%)	
Never	109 (9.0)
Less than once a month	153 (12.6)
About once a month	228 (18.8)
2 to 3 times a month	173 (14.3)
Once a week	85 (7.0)
2 to 3 times a week	115 (9.5)
4 to 5 times a week	58 (4.8)
6 to 7 times a week	169 (13.9)
NA	124 (10.2)
Avoidance of particular foods (%)	
Never	853 (70.3)
Often	11 (0.9)
Prefer not to answer	15 (1.2)
Rarely	163 (13.4)
Sometimes	52 (4.3)
NA	120 (9.9)

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

Oil on bread (%)	
Butter	371 (30.6)
Low fat margarine	272 (22.4)
Full fat margarine	300 (24.7)
None	109 (9.0)
Olive oil	36 (3.0)
NA	126 (10.4)
Artificial sweeteners (%)	
Almost never	976 (80.4)
About 1/4 of the time	24 (2.0)
About 1/2 of the time	16 (1.3)
About 3/4 of the time	12 (1.0)
Almost always or always	53 (4.4)
NA	133 (11.0)
Non-diet soda frequency (%)	
Zero days a week	432 (35.6)
One to three days per month	459 (37.8)
One to five days a week	167 (13.8)
Six to seven days a week	27 (2.2)
NA	129 (10.6)
Diet sugar drink frequency (%)	
Zero days a week	513 (42.3)
One to three days per month	356 (29.3)
One to five days a week	156 (12.9)
Six to seven days a week	57 (4.7)
NA	132 (10.9)
Soy/fish usage (%)	
Never at the table	424 (34.9)
Rarely at the table	441 (36.3)
Sometimes at the table	217 (17.9)
At most meals of eating occasions	9 (0.7)
NA	123 (10.1)
Salt seasoning (%)	
Never	368 (30.3)
Rarely	347 (28.6)
Sometimes	219 (18.0)
Most meals	157 (12.9)
NA	123 (10.1)
Fast food frequency (%)	
Never	149 (12.3)

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

Less than once per month	384 (31.6)
One - three times per month	366 (30.1)
One - six per week	191 (15.7)
One or more times per day	Data Repressed
NA	122 (10.0)
Alcohol Frequency (%)	
Never	61 (5.0)
Less than once a month	192 (15.8)
About once a month	70 (5.8)
2 to 3 times a month	171 (14.1)
Once a week	170 (14.0)
2 to 3 times a week	259 (21.3)
4 to 5 times a week	127 (10.5)
6 to 7 times a week	112 (9.2)
NA	52 (4.3)
Education level (%)	
Highschool or below	208 (17.1)
Non-bachelors post secondary	425 (35.0)
Bachelors	334 (27.5)
Graduate	242 (19.9)
NA	Data Repressed
Income (%)	
Below $25 000 CAD	41 (3.4)
$25 000 - $49 999 CAD	157 (12.9)
$50 000 - $74 999 CAD	244 (20.1)
$75 000 - $99 999 CAD	244 (20.1)
$100 000 - $149 999 CAD	291 (24.0)
Greater than $150 000 CAD	179 (14.7)
NA	58 (4.8)
Sleeping trouble frequency (%)	
None	104 (8.6)
A little of the time	411 (33.9)
Some of the time	507 (41.8)
Most of the time	161 (13.3)
All the time	25 (2.1)
NA	Data Repressed
Last dental visit (%)	
Less than 6 months ago	851 (70.1)
6 months to less than 1 year ago	221 (18.2)
1 year to less than 2 years ago	56 (4.6)

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

2 years to less than 3 years ago	17 (1.4)
3 or more years ago	24 (2.0)
NA	45 (3.7)
Sleeping light exposure (%)	
Virtually no light	561 (46.2)
Some light	613 (50.5)
A lot of light	36 (3.0)
NA	Data Repressed
DNA extraction batch (%)	
Extraction.1	85 (7.0)
Extraction.10	66 (5.4)
Extraction.11	80 (6.6)
Extraction.12	78 (6.4)
Extraction.13	85 (7.0)
Extraction.14	57 (4.7)
Extraction.15	79 (6.5)
Extraction.16	0 (0.0)
Extraction.17	67 (5.5)
Extraction.2	85 (7.0)
Extraction.3	81 (6.7)
Extraction.4	68 (5.6)
Extraction.5	85 (7.0)
Extraction.6	92 (7.6)
Extraction.7	85 (7.0)
Extraction.8	60 (4.9)
Extraction.9	61 (5.0)

489

490

491

492

493

494

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

495

496 Table 2: Validation of Beta Diversity Results

Metric	Feature	P-value	R^2
Weighted UniFrac	Waist Hip Ratio	0.0190	0.0116
	Height	0.001	0.0117
	Weight	0.010	0.0102
	Fat Free Mass	0.002	0.0172
	Sex	0.0390	0.0080
	Age	0.0120	0.0105
Bray-Curtis	Waist Hip Ratio	0.0140	0.0072
	Height	0.0030	0.0118
	Weight	0.0020	0.0096
	Fat Free Mass	0.0040	0.0110
	Waist Size	0.0210	0.0065
	Age	0.0020	0.0106
	Sex	0.0380	0.0059

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

497

498 **Figures:**

499 **Figure 1.** Atlantic Canadian oral microbiome composition is dominated by the genus Veillonella
500 and is relatively similar at the genus level but highly variable at the ASV level. Samples from
501 Atlantic Partnership for Tomorrow's Health project (n=1049). Samples were subsampled to a
502 depth of 5000 reads. A) Genera that had a mean relative abundance less than 1% were grouped
503 into "Other". B) Genera were removed at varying different sample presence cut-offs and the
504 remaining total mean relative abundance of non-filtered genera was then calculated. C) ASVs
505 were removed at varying different sample presence cut-offs and the remaining total mean
506 relative abundance of non-filtered ASVs was then calculated.

507 **Figure 2.** Various anthropometric, dietary and lifestyle features are significantly associated to
508 oral microbiome composition. Saliva samples from Atlantic Partnership for Tomorrow's Health
509 cohort (n=741). Samples were subsampled to a depth of 5000 reads. Two different metrics were
510 tested weighted Unifrac distances (A) and Bray-Curtis dissimilarity (C) using a PERMANOVA
511 test while controlling for differences in DNA extraction and correction for false discovery ($q <$
512 0.1). Arrows point toward the direction each feature correlates along PC1 and PC2 while their
513 size was scaled by the PERMANOVA R^2 value. Panels B and D show the relative rankings of
514 the effect sizes (R^2) of each significant feature.

515

516 **Figure 3**. Various genera and ASVs are associated with features found to influence oral
517 microbiome composition. Genera (A) and ASVs (B) meeting an FDR <0.1 using the Corncob R
518 package which uses beta binomial regressions. Each feature's false discovery rate was corrected

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

519 separately, and each tested to control for differences in DNA extraction and differential
520 variability within that feature. Ordinal variables were converted into a ranked scale for testing,
521 and all features except for sex were scaled. *Sex was treated as a categorical value and therefore
522 the magnitude is not directly comparable to other log odd ratios.
523
524 **Figure 4**. Validation of Genera and ASV association in a smaller Atlantic Partnership for
525 Tomorrow's Health cohort (n=308). Samples that did not complete all questions examined in this
526 study were used to validate previous associations identified in the larger cohort. Testing
527 procedure was done in the same manner as Figure 3 with A representing Genera and B
528 representing ASVs.
529
530
531
532 References

533 1. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy
534 HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA,
535 Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J,
536 Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO,
537 Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi
538 HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V,
539 Bihan M, Blaser MJ, Bloom T, Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam
540 DA, Campbell JL, Canon SR, Cantarel BL, Chain PSG, Chen I-MA, Chen L, Chhibba S,
541 Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

542 Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E,
543 Ding Y, Dooling DJ, Dugan SP, Michael Dunne W, Scott Durkin A, Edgar RC, Erlich RL,
544 Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney
545 LJ, Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H,
546 Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A,
547 Griggs A, Gujja S, Kinder Haake S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA,
548 Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson
549 JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, King NB,
550 Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides NC, La Rosa PS,
551 Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu
552 Y, Lo C-C, Lozupone CA, Dwayne Lunsford R, Madden T, Mahurkar AA, Mannon PJ,
553 Mardis ER, Markowitz VM, Mavromatis K, McCorrison JM, McDonald D, McEwen J,
554 McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ,
555 Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM,
556 Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Pop M, Priest ME, Proctor LM, Qin
557 X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodriguez-
558 Mueller B, Rogers Y-H, Ross MC, Russ C, Sanka RK, Sankar P, Fah Sathirapongsasuti J,
559 Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N,
560 Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA,
561 Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG,
562 Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA,
563 Walker J, Wang L, Wang Z, Ward D V, Warren W, Watson MA, Wellington C,
564 Wetterstrand KA, White JR, Wilczek-Boney K, Wu Y, Wylie KM, Wylie T, Yandava C,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

565 Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD,
566 Birren BW, Gibbs RA, Highlander SK, Methé BA, Nelson KE, Petrosino JF, Weinstock
567 GM, Wilson RK, White O, Consortium THMP. 2012. Structure, function and diversity of
568 the healthy human microbiome. Nature 486:207–214.

569 2. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, Lakshmanan A, Wade
570 WG. 2010. The Human Oral Microbiome. J Bacteriol 192:5002 LP – 5017.

571 3. Wade WG. 2013. The oral microbiome in health and disease. Pharmacol Res 69:137–143.

572 4. Takahashi N, Nyvad B. 2010. The Role of Bacteria in the Caries Process: Ecological
573 Perspectives. J Dent Res 90:294–303.

574 5. Karpiński MT. 2019. Role of Oral Microbiota in Cancer Development. Microorg .

575 6. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent Jr. RL. 1998. Microbial
576 complexes in subgingival plaque. J Clin Periodontol 25:134–144.

577 7. Kumar PS, Griffen AL, Barton JA, Paster BJ, Moeschberger ML, Leys EJ. 2003. New
578 Bacterial Species Associated with Chronic Periodontitis. J Dent Res 82:338–344.

579 8. Murray PA, Prakobphol A, Lee T, Hoover CI, Fisher SJ. 1992. Adherence of oral
580 streptococci to salivary glycoproteins. Infect Immun 60:31 LP – 38.

581 9. Kolenbrander PE, Palmer Jr RJ, Rickard AH, Jakubovics NS, Chalmers NI, Diaz PI. 2006.
582 Bacterial interactions and successions during plaque development. Periodontol 2000
583 42:47–79.

584 10. Shungin D, Haworth S, Divaris K, Agler CS, Kamatani Y, Keun Lee M, Grinde K, Hindy
585 G, Alaraudanjoki V, Pesonen P, Teumer A, Holtfreter B, Sakaue S, Hirata J, Yu Y-H,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

586 Ridker PM, Giulianini F, Chasman DI, Magnusson PKE, Sudo T, Okada Y, Völker U,
587 Kocher T, Anttonen V, Laitala M-L, Orho-Melander M, Sofer T, Shaffer JR, Vieira A,
588 Marazita ML, Kubo M, Furuichi Y, North KE, Offenbacher S, Ingelsson E, Franks PW,
589 Timpson NJ, Johansson I. 2019. Genome-wide analysis of dental caries and periodontitis
590 combining clinical and self-reported data. Nat Commun 10:2773.

591 11. Ahn J, Chen CY, Hayes RB. 2012. Oral microbiome and oral and gastrointestinal cancer
592 risk. Cancer Causes Control 23:399–404.

593 12. Flemer B, Warren RD, Barrett MP, Cisek K, Das A, Jeffery IB, Hurley E, O'Riordain M,
594 Shanahan F, O'Toole PW. 2017. The oral microbiota in colorectal cancer is distinctive
595 and predictive. Gut.

596 13. Fan X, Alekseyenko A V, Wu J, Peters BA, Jacobs EJ, Gapstur SM, Purdue MP, Abnet
597 CC, Stolzenberg-Solomon R, Miller G, Ravel J, Hayes RB, Ahn J. 2018. Human oral
598 microbiome and prospective risk for pancreatic cancer: a population-based nested case-
599 control study. Gut 67:120 LP – 127.

600 14. Porter CM, Shrestha E, Peiffer LB, Sfanos KS. 2018. The microbiome in prostate
601 inflammation and prostate cancer. Prostate Cancer Prostatic Dis 21:345–354.

602 15. Koren O, Spor A, Felin J, Fåk F, Stombaugh J, Tremaroli V, Behre CJ, Knight R,
603 Fagerberg B, Ley RE, Bäckhed F. 2011. Human oral, gut, and plaque microbiota in
604 patients with atherosclerosis. Proc Natl Acad Sci 108:4592 LP – 4598.

605 16. Said HS, Suda W, Nakagome S, Chinen H, Oshima K, Kim S, Kimura R, Iraha A, Ishida
606 H, Fujita J, Mano S, Morita H, Dohi T, Oota H, Hattori M. 2013. Dysbiosis of Salivary
607 Microbiota in Inflammatory Bowel Disease and Its Association With Oral Immunological

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

608 Biomarkers. DNA Res 21:15–25.

609 17. De Filippis F, Vannini L, La Storia A, Laghi L, Piombino P, Stellato G, Serrazanetti DI,
610 Gozzi G, Turroni S, Ferrocino I, Lazzi C, Di Cagno R, Gobbetti M, Ercolini D. 2014. The
611 Same Microbiota and a Potentially Discriminant Metabolome in the Saliva of Omnivore,
612 Ovo-Lacto-Vegetarian and Vegan Individuals. PLoS One 9:e112373.

613 18. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. 2009. Global diversity in the human
614 salivary microbiome. Genome Res 19:636–643.

615 19. Takeshita T, Kageyama S, Furuta M, Tsuboi H, Takeuchi K, Shibata Y, Shimazaki Y,
616 Akifusa S, Ninomiya T, Kiyohara Y, Yamashita Y. 2016. Bacterial diversity in saliva and
617 oral health-related conditions: the Hisayama Study. Sci Rep 6:22164.

618 20. Zaura E, Keijser BJ, Huse SM, Crielaard W. 2009. Defining the healthy "core
619 microbiome" of oral microbial communities. BMC Microbiol 9.

620 21. Li J, Quinque D, Horz H-P, Li M, Rzhetskaya M, Raff JA, Hayes MG, Stoneking M.
621 2014. Comparative analysis of the human saliva microbiome from different climate zones:
622 Alaska, Germany, and Africa. BMC Microbiol 14:316.

623 22. Belstrøm D, Holmstrup P, Nielsen CH, Kirkby N, Twetman S, Heitmann BL, Klepac-
624 Ceraj V, Paster BJ, Fiehn N-E. 2014. Bacterial profiles of saliva in relation to diet,
625 lifestyle factors, and socioeconomic status. J Oral Microbiol 6:23609.

626 23. Renson A, Jones HE, Beghini F, Segata N, Zolnik CP, Usyk M, Moody TU, Thorpe L,
627 Burk R, Waldron L, Dowd JB. 2019. Sociodemographic variation in the oral microbiome.
628 Ann Epidemiol 35:73-80.e2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

24. Peters BA, McCullough ML, Purdue MP, Freedman ND, Um CY, Gapstur SM, Hayes RB, Ahn J. 2018. Association of Coffee and Tea Intake with the Oral Microbiome: Results from a Large Cross-Sectional Study. Cancer Epidemiol Biomarkers & Prev 27:814 LP – 821.

25. Mason MR, Nagaraja HN, Camerlengo T, Joshi V, Kumar PS. 2013. Deep Sequencing Identifies Ethnicity-Specific Bacterial Signatures in the Oral Microbiome. PLoS One 8:e77287.

26. Fan X, Peters BA, Jacobs EJ, Gapstur SM, Purdue MP, Freedman ND, Alekseyenko A V, Wu J, Yang L, Pei Z, Hayes RB, Ahn J. 2018. Drinking alcohol is associated with variation in the human oral microbiome in a large study of American adults. Microbiome 6:59.

27. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, Ma Y, Purdue MP, Jacobs EJ, Gapstur SM, Li H, Alekseyenko A V, Hayes RB, Ahn J. 2016. Cigarette smoking and the oral microbiome in a large study of American adults. Isme J 10:2435.

28. Wu Y, Chi X, Zhang Q, Chen F, Deng X. 2018. Characterization of the salivary microbiome in people with obesity. PeerJ 6:e4458.

29. Yang Y, Cai Q, Zheng W, Steinwandel M, Blot WJ, Shu X-O, Long J. 2019. Oral microbiome and obesity in a large study of low-income and African-American populations. J Oral Microbiol 11:1650597.

30. Hansen TH, Kern T, Bak EG, Kashani A, Allin KH, Nielsen T, Hansen T, Pedersen O. 2018. Impact of a vegan diet on the human salivary microbiota. Sci Rep 8.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

650 31. Sweeney E, Cui Y, DeClercq V, Devichand P, Forbes C, Grandy S, Hicks JMT, Keats M,
651 Parker L, Thompson D, Volodarsky M, Yu ZM, Dummer TJB. 2017. Cohort Profile: The
652 Atlantic Partnership for Tomorrow's Health (Atlantic PATH) Study. Int J Epidemiol
653 46:1762-1763i.

654 32. Comeau AM, Douglas GM, Langille MGI. 2017. Microbiome Helper: a Custom and
655 Streamlined Workflow for Microbiome Research. mSystems 2.

656 33. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
657 reads. EMBnet.journal; Vol 17, No 1 Next Gener Seq Data Anal.

658 34. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander
659 H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A,
660 Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da
661 Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF,
662 Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A,
663 Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen
664 S, Jarmusch AK, Jiang L, Kaehler BD, Kang K Bin, Keefe CR, Keim P, Kelley ST,
665 Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X,
666 Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ,
667 Melnik A V, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias
668 LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen
669 LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song
670 SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ,
671 Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M,
672 Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

673 Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive,
674 scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–
675 857.

676 35. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open
677 source tool for metagenomics. PeerJ 4:e2584.

678 36. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley
679 EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur Rapidly Resolves
680 Single-Nucleotide Community Sequence Patterns. mSystems 2.

681 37. Andersen GL, DeSantis TZ, Liu Z, Knight R. 2008. Accurate taxonomy assignments from
682 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res
683 36:e120–e120.

684 38. MIRARAB S, NGUYEN N, WARNOW T. 2011. SEPP: SATé-Enabled Phylogenetic
685 Placement, p. 247–258. *In* Biocomputing 2012. WORLD SCIENTIFIC.

686 39. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, Winker K, Kado
687 DM, Orwoll E, Manary M, Mirarab S, Knight R. 2018. Phylogenetic Placement of Exact
688 Amplicon Sequences Improves Associations with Clinical Information. mSystems
689 3:e00021-18.

690 40. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA,
691 Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon
692 sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.

693 41. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

694 SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA
695 sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196.

696 42. Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New
697 York.

698 43. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and
699 powerful approach to multiple testing. J R Stat Soc Ser B 57.

700 44. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14.

701 45. Martin BD, Witten D, Willis AD. 2019. Modeling microbial abundances and dysbiosis
702 with beta-binomial regression.

703 46. McMurdie PJ, Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive
704 Analysis and Graphics of Microbiome Census Data. PLoS One 8:e61217.

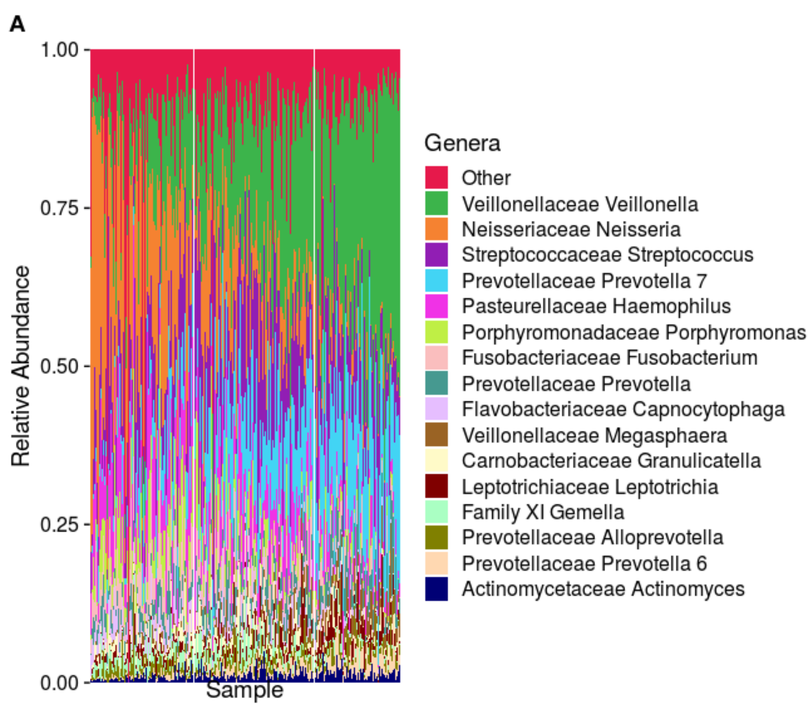
705 47. Raju SC, Lagström S, Ellonen P, de Vos WM, Eriksson JG, Weiderpass E, Rounge TB.
706 2019. Gender-Specific Associations Between Saliva Microbiota and Body Size. Front
707 Microbiol 10.

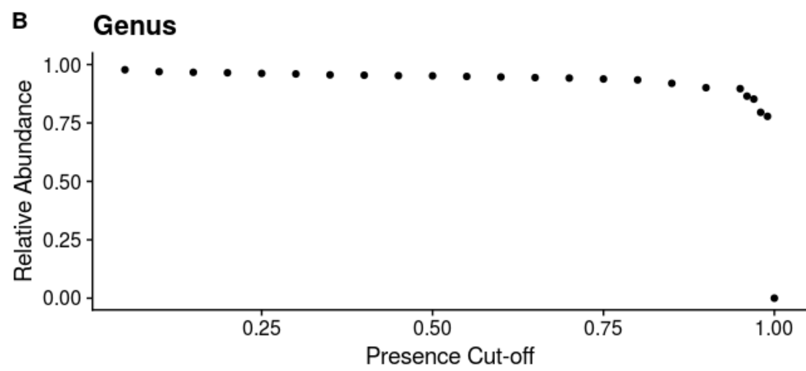
708 48. DeClercq V, Langille MGI, Van Limbergen J. 2018. Differences in adiposity and diet
709 quality among individuals with inflammatory bowel disease in Eastern Canada. PLoS One
710 13:e0200580.

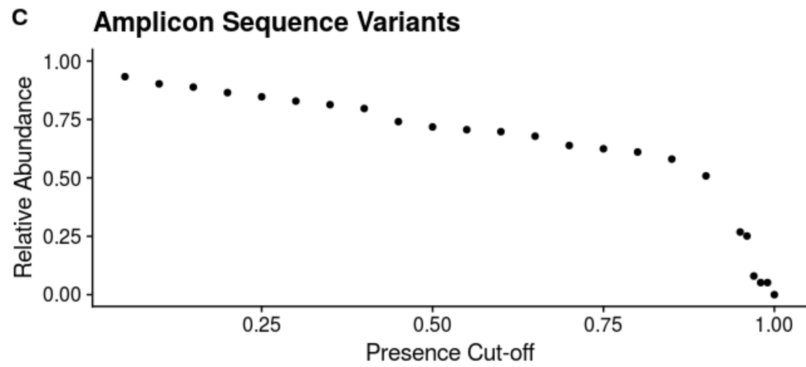
711 49. Xun Z, Zhang Q, Xu T, Chen N, Chen F. 2018. Dysbiosis and ecotypes of the salivary
712 microbiome associated with inflammatory bowel diseases and the assistance in diagnosis
713 of diseases using oral bacterial profiles. Front Microbiol 9.

714

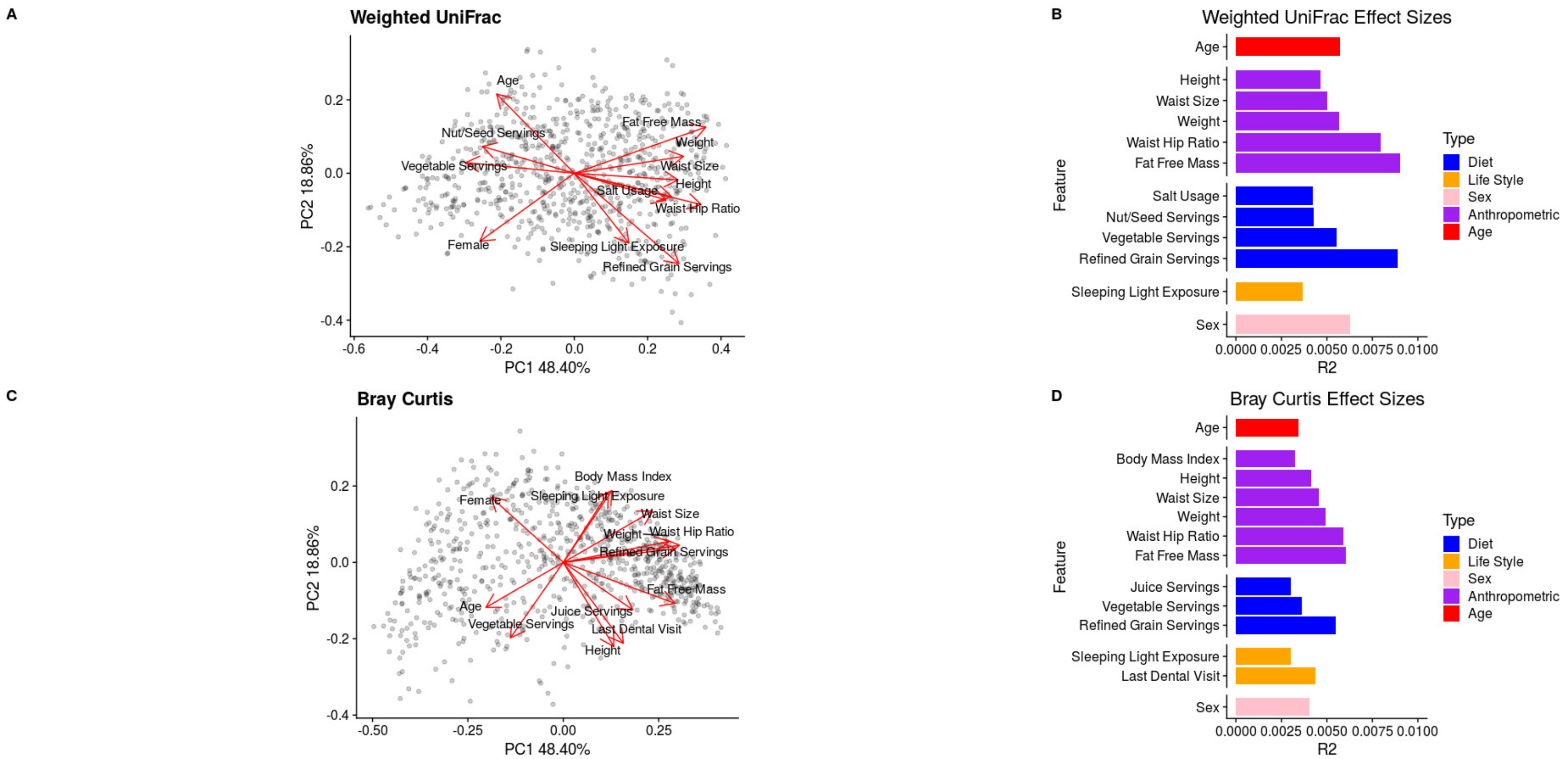
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



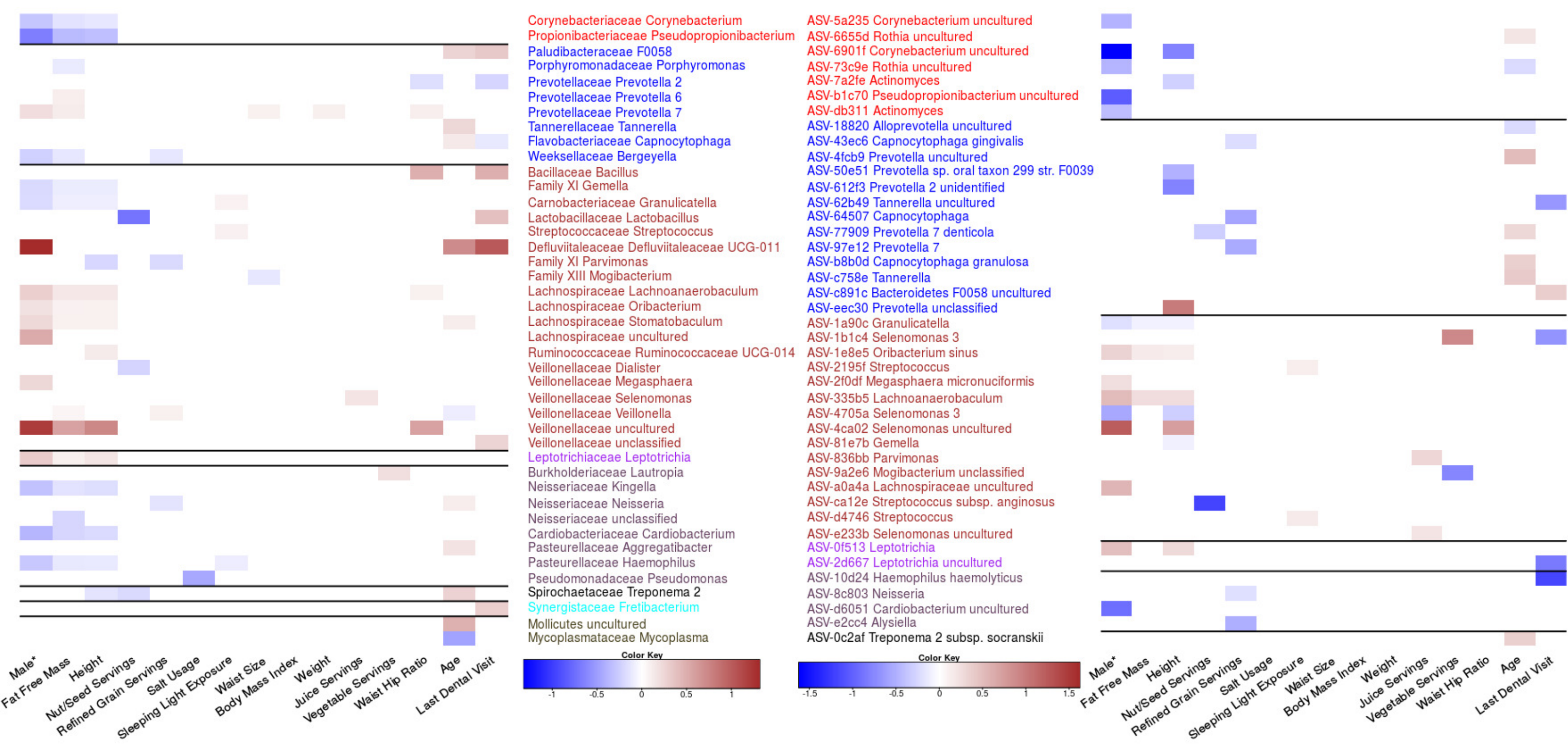




bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.07.083634; this version posted May 9, 2020. 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 4.0 International license.

Corynebacteriaceae Corynebacterium
Propionibacteriaceae Pseudopropionibacterium
Paludibacteraceae F0058
Porphyromonadaceae Porphyromonas
Prevotellaceae Prevotella 2
Prevotellaceae Prevotella 6
Prevotellaceae Prevotella 7
Tannerellaceae Tannerella
Flavobacteriaceae Capnocytophaga
Weeksellaceae Bergeyella
Bacillaceae Bacillus
Family XI Gemella
Carnobacteriaceae Granulicatella
Lactobacillaceae Lactobacillus
Streptococcaceae Streptococcus
Defluviitaleaceae Defluviitaleaceae UCG-011
Family XI Parvimonas
Family XIII Mogibacterium
Lachnospiraceae Lachnoanaerobaculum
Lachnospiraceae Oribacterium
Lachnospiraceae Stomatobaculum
Lachnospiraceae uncultured
Ruminococcaceae Ruminococcaceae UCG-014
Veillonellaceae Dialister
Veillonellaceae Megasphaera
Veillonellaceae Selenomonas
Veillonellaceae Veillonella
Veillonellaceae uncultured
Veillonellaceae unclassified
Leptotrichiaceae Leptotrichia
Burkholderiaceae Lautropia
Neisseriaceae Kingella
Neisseriaceae Neisseria
Neisseriaceae unclassified
Cardiobacteriaceae Cardiobacterium
Pasteurellaceae Aggregatibacter
Pasteurellaceae Haemophilus
Pseudomonadaceae Pseudomonas
Spirochaetaceae Treponema 2
Synergistaceae Fretibacterium
Mollicutes uncultured
Mycoplasmataceae Mycoplasma

Age
Height
Waist Hip Ratio
Fat Free Mass
Male*

Color Key
-0.5 0 0.5

ASV-5a235 Corynebacterium uncultured
ASV-6655d Rothia uncultured
ASV-6901f Corynebacterium uncultured
ASV-73c9e Rothia uncultured
ASV-7a2fe Actinomyces
ASV-b1c70 Pseudopropionibacterium uncultured
ASV-db311 Actinomyces
ASV-18820 Alloprevotella uncultured
ASV-43ec6 Capnocytophaga gingivalis
ASV-4fcb9 Prevotella uncultured
ASV-50e51 Prevotella sp. oral taxon 299 str. F0039
ASV-612f3 Prevotella 2 unidentified
ASV-62b49 Tannerella uncultured
ASV-64507 Capnocytophaga
ASV-77909 Prevotella 7 denticola
ASV-97e12 Prevotella 7
ASV-b8b0d Capnocytophaga granulosa
ASV-c758e Tannerella
ASV-c891c Bacteroidetes F0058 uncultured
ASV-eec30 Prevotella unclassified
ASV-1a90c Granulicatella
ASV-1b1c4 Selenomonas 3
ASV-1e8e5 Oribacterium sinus
ASV-2195f Streptococcus
ASV-2f0df Megasphaera micronuciformis
ASV-335b5 Lachnoanaerobaculum
ASV-4705a Selenomonas 3
ASV-4ca02 Selenomonas uncultured
ASV-81e7b Gemella
ASV-836bb Parvimonas
ASV-9a2e6 Mogibacterium unclassified
ASV-a0a4a Lachnospiraceae uncultured
ASV-ca12e Streptococcus subsp. anginosus
ASV-d4746 Streptococcus
ASV-e233b Selenomonas uncultured
ASV-0f513 Leptotrichia
ASV-2d667 Leptotrichia uncultured
ASV-10d24 Haemophilus haemolyticus
ASV-8c803 Neisseria
ASV-d6051 Cardiobacterium uncultured
ASV-e2cc4 Alysiella
ASV-0c2af Treponema 2 subsp. socranskii

Color Key
-0.5 0 0.5

Height
Fat Free Mass
Sleeping Light Exposure
Male*