



## 24 **Abstract**

25

26 It is a central goal of human microbiome studies to see the roles of the microbiome as a mediator  
27 that transmits environmental, behavioral, or medical exposures to health or disease outcomes. Yet,  
28 mediation analysis is not used as much as it should be. One reason is because of the lack of  
29 carefully planned routines, compilers and automative computing systems for microbiome  
30 mediation analysis to perform a series of data processing, diversity calculation, data normalization,  
31 downstream data analysis and visualizations. Many researchers in various disciplines (e.g.,  
32 clinicians, public health practitioners, biologists) are not also familiar with related statistical  
33 methods and programming languages on command-line interfaces. Thus, here we introduce a web  
34 cloud computing platform, named MiMed, that enables comprehensive microbiome mediation  
35 analysis on user-friendly web interfaces. We applied MiMed to the study on the mediating roles  
36 of oral microbiome in subgingival niches between e-cigarette smoking and gingival inflammation,  
37 and found significant mediating effects from two phyla (*Proteobacteria*, *Spirochaetes*), two  
38 classes (*Flavobacteriia*, *Betaproteobacteria*), four orders (*Flavobacteriales*, *Burkholderiales*,  
39 *Neisseriales*, *Cardiobacteriales*), five families (*Flavobacteriaceae*, *Burkholderiaceae*,  
40 *Neisseriaceae*, *Cardiobacteriaceae*, *Enterococcaceae*), and eight genera (*Bergeyella*,  
41 *Capnocytophaga*, *Actinomyces*, *Haemophilus*, *Kingella*, *Burkholderia*, *Cardiobacterium*,  
42 *Enterococcus*).

43

## 44 **Importance**

45

46 The main features of MiMed are as follows. First, MiMed can survey the microbiome in various  
47 spheres (1) as a whole microbial ecosystem using different ecological measures (e.g., alpha- and  
48 beta-diversity indices) or (2) as individual microbial taxa (e.g., phyla, classes, orders, families,  
49 genera, species) using different data normalization methods. Second, MiMed enables covariate-  
50 adjusted analysis to control for potential confounding factors (e.g., age, gender), which is essential  
51 to enhance the causality of the results especially for observational studies. Third, MiMed enables  
52 a breadth of statistical inferences in both mediation effect estimation and significance testing.  
53 Fourth, MiMed provides flexible and easy-to-use data processing and analytic modules and creates  
54 nice graphical representations. Finally, MiMed employs ChatGPT to search for what have been  
55 known about the microbial taxa that are found significantly as mediators using AI technologies.  
56 MiMed is freely available on our web server (<http://mimed.micloud.kr>).

57

58 **Keywords.** Causal mediation analysis, Microbiome data analysis, Web cloud computing, Causal  
59 inference, Human microbiome

60

## 61 **Introduction**

62

63 The human microbiome is the totality of all microbes that live on and inside various organs (e.g.,  
64 gut, mouth, skin, nose) of the human body. The advances in massively parallel metagenomic  
65 sequencing have dramatically lowered the cost of microbiome profiling with substantial increase  
66 in accuracy. Then, the microbiome field has not only become an active area of research, but has  
67 also rapidly grown in industry with the aim of identifying new ways to diagnose, treat and prevent  
68 human diseases.

69 Researchers have revealed a sophisticated interplay between microbiome and its host in various  
70 aspects. For instance, microbiome diversity and its taxonomic composition have been related to a  
71 variety of environmental, behavioral or medical exposures (e.g., diet [1], residence [2], smoking  
72 [3], preterm birth [4], delivery mode [5, 6], antibiotic/probiotic use [7, 8]). Researchers have also  
73 found that microbiome dysbiosis can lead to numerous disorders (e.g., obesity [9, 10], intestinal  
74 disease [11, 12, 13], cancers [14, 15, 16], diabetes [8, 17], brain disorders [18, 19]). However,  
75 beyond such separate discoveries, it is essential to understand if the microbiome transmits the  
76 effects of environmental, behavioral or medical exposures (say, treatment) to health or disease  
77 outcomes (say, outcome) as a mediator (Fig. 1), which can be surveyed through causal mediation  
78 analysis [20].

79 Mediation analysis aims to comprehend the underlying mechanism in an observed relationship  
80 between a treatment and an outcome through a third hypothetical variable, known as a mediator,  
81 indirectly. That is, in human microbiome studies, mediation analysis surveys two links jointly, (1)  
82 the effect of a treatment on microbiome (denoted as ‘treatment - microbiome’) and (2) the effect  
83 of microbiome on an outcome conditional on treatment status (denoted as ‘microbiome - outcome’)  
84 (Fig. 1). If we lose any one of these two links, microbiome does not serve as a mediator. That is,  
85 if we have ‘treatment - microbiome’ but do not have ‘microbiome - outcome’, the treatment alters  
86 microbiome, but the altered microbiome has no effect on the outcome. This means that the effect  
87 of the treatment on the outcome was made ‘directly’ or by some other unknown pathways, not  
88 through the microbiome. Similarly, if we do not have ‘treatment - microbiome’ but have  
89 ‘microbiome - outcome’, the treatment does not alter the microbiome, but only the variability in  
90 microbiome due to some other unknown sources influences the outcome. Thus, the roles of the  
91 microbiome as a mediator are satisfied only when we have both links [20], which we refer as the

92 presence of ‘indirect’ or ‘mediation’ effect. It substantially matters in a clinical context because if  
93 the microbiome is not in a causal pathway, any medical interventions to the microbiome do not  
94 fundamentally treat or prevent human diseases.

95       However, in human microbiome studies, mediation analysis is not used as much as it should be.  
96 One reason is because of the lack of carefully planned routines, compilers and automative  
97 computing systems [21] for microbiome mediation analysis to perform a series of data processing,  
98 diversity calculation, data normalization, downstream data analysis and visualization. The  
99 microbiome data are highly complex and demand many data processing and analytic procedures.  
100 Many researchers in various disciplines (e.g., clinicians, public health practitioners, biologists) are  
101 not also familiar with related statistical methods and programming languages on command-line  
102 interfaces. Moreover, there are many other important issues that need to be addressed for  
103 microbiome causal mediation analysis as follows. First, we can view the microbiome as a whole  
104 community in an ecological context (referred in this paper for ‘community-level analysis’) or can  
105 focus on individual microbial taxa at various taxonomic hierarchies (i.e., phyla, classes, orders,  
106 families, genera, species) (referred in this paper for ‘taxonomy-level analysis’). Researchers  
107 usually survey the former using different ecological measures (e.g., alpha- and beta-diversity  
108 indices) [22, 23] and the latter using different data normalization methods (e.g., centered-log ratio  
109 (CLR) [24], arcsine-root [25]). Second, covariate-adjusted analysis is needed to control for  
110 potential confounding factors (e.g., age, gender), which is especially necessary for observational  
111 studies to enhance the causality of the results. Third, both mediation effect estimation and  
112 significance testing are important portions of statistical inference for better interpretability. Fourth,  
113 we need flexible and easy-to-use data processing and analytic modules as well as high-quality  
114 visualizations to be included in an academic paper. Finally, we need to figure out what have been

115 known about the microbes that we discovered as significant mediators. However, it is not easy in  
116 practice to figure it out all manually since there are too many microbial taxa [26] and related prior  
117 studies. Hence, we may need a well-trained artificial intelligence (AI) machine that can do such a  
118 job for us.

119 To tackle all those critical issues described above, here we introduce a web cloud computing  
120 platform, named MiMed, that enables comprehensive microbiome mediation analysis on user-  
121 friendly web interfaces. MiMed is the first web cloud computing platform for microbiome causal  
122 mediation analysis, which is distinguished from our prior platforms: (1) MiCloud for association  
123 analysis in cross-section or longitudinal microbiome studies [27]; (2) MiPair for design-based  
124 comparative analysis with paired microbiome data [28]; and (3) MiSurv for microbiome data  
125 analysis with survival responses [29]. Interestingly, MiMed is also built-in the popular AI language  
126 model, ChatGPT, to easily search for what have been known about the microbial taxa that are  
127 found significantly as mediators.

128 In the following *Results* section, we describe all the data processing and analytic modules one  
129 by one using an example study to see the mediating roles of oral microbiome between e-cigarette  
130 smoking and gingival inflammation [30]. Then, in the *Discussion* section, we summarize and  
131 discuss all the features and implications of MiMed. Finally, in the *Methods* section, we discuss the  
132 methodological ideas of causal mediation analysis methods, and describe our web server and local  
133 GitHub repository. MiMed is freely available on our web server (<http://mimed.micloud.kr>) or can  
134 alternatively run on a user's local computer (<https://github.com/yj7599/MiMedGit>).

135

## 136 **Results**

137

## 138 **Application Note: On the Roles of Oral Microbiome Between E-** 139 **cigarette Smoking and Gingival Inflammation**

140 To ease our demonstration, we use example data to survey the mediating roles of oral microbiome  
141 between e-cigarette smoking and gingival inflammation [30]. The original data are huge and can  
142 motivate a broad range of study orientations. We refer the original study paper [30] for all the  
143 details on study subjects, sample collection/processing, and sequencing/quantification procedures.

144 To describe only the portion of the data we use, the data are 16S oral microbiome data in  
145 subgingival niches obtained at the baseline visit of the subjects between 18 and 34 years in age.  
146 We employed a bioinformatic pipeline, QIIME2 [31], based on the expanded human oral  
147 microbiome database (eHOMD) [32] for raw sequence data processing, denoising, feature  
148 extraction/quantification, taxonomic annotation, and phylogenetic tree construction. We added  
149 detailed description on the use of each module using these example data at the end of each  
150 following section (see *Application Note*).

151

### 152 **Data Processing: Data Input**

153 Microbiome data can be composed of three data components: (1) a feature table (i.e., count data  
154 for operational taxonomic units (OTUs) or amplicon sequence variants (ASVs)), (2) a taxonomic  
155 table (i.e., taxonomic annotations at various taxonomic hierarchies, kingdom, phylum, class, order,  
156 family, genus, species), and (3) a phylogenetic tree (i.e., a rooted phylogenetic tree for evolutionary  
157 relationships across features, that are OTUs or ASVs). Of course, in addition to microbiome data,  
158 metadata on a treatment variable (e.g., environmental, behavioral or medical exposures), an  
159 outcome variable (e.g., health or disease status), and possibly covariates (e.g., age, gender) for

160 study subjects are needed. If we have all these data components, we can conduct microbiome  
161 causal mediation analysis comprehensively using all available functions of MiMed. However,  
162 researchers do not always have all these data components, but even in such a case, they can still  
163 want to conduct at least some parts of the analysis. Thus, we made the Data Input module flexible  
164 as follows.

165 (1) To fully perform all community-level analyses for all non-phylogenetic and phylogenetic  
166 alpha- and beta-diversity indices as well as all taxonomy-level analyses, users should  
167 upload a feature table, a taxonomic table, a phylogenetic tree, and metadata.

168 (2) To perform community-level analyses for only non-phylogenetic alpha- and beta-diversity  
169 indices as well as all taxonomy-level analyses, users can upload only a feature table, a  
170 taxonomic table, and metadata.

171 (3) To perform only community-level analyses for all non-phylogenetic and phylogenetic  
172 alpha- and beta-diversity indices, users can upload only a feature table, a phylogenetic tree,  
173 and metadata.

174 (4) To perform only community-level analyses for only non-phylogenetic alpha- and beta-  
175 diversity indices, users can upload only a feature table and metadata.

176 Users can upload their data components in a widely used unified format, called phyloseq [33],  
177 or as separate files.

178 **Application Note:** The example data we use can be downloaded in the Example Data section on  
179 the Data Input module. To help users to easily understand data components and their corresponding  
180 data analytic modules as described above, we uploaded four different sets of data components: (1)  
181 a feature table, a taxonomic table, a phylogenetic tree, and metadata; (2) a feature table, a  
182 taxonomic table, and metadata; (3) a feature table, a phylogenetic tree, and metadata; (4) a feature



183 table and metadata. Though since we aim in this paper to describe all available functions of MiMed,  
184 we uploaded the one with all data components (i.e., a feature table, a taxonomic table, a  
185 phylogenetic tree, and metadata).

186

## 187 **Data Processing: Quality Control**

188 MiMed performs quality controls (QCs) just as in MiCloud [27] and MiPair [28]. That is, users  
189 need to select (1) a kingdom of interest (default: Bacteria), (2) a minimum library size (i.e., total  
190 read count) for the study subjects to be rescued (default: 3,000), (3) a minimum mean relative  
191 abundance (i.e., proportion) for the features (OTUs or ASVs) to be rescued (default: 0.002%), and  
192 (4) erroneous taxonomic names in the taxonomic table to be removed.

193 MiMed displays the sample size, the number of features (OTUs or ASVs), the number of phyla,  
194 the number of classes, the number of orders, the number of families, the number of genera and the  
195 number of species using summary boxes before and after QCs. MiMed also visualizes library sizes  
196 across study subjects as well as mean proportions across features using interactive histograms and  
197 box plots before and after QCs.

198 **Application Note:** We simply clicked the Run button to apply the default QC settings. Then, 147  
199 subjects with 2,328 features, 11 phyla, 23 classes, 34 orders, 52 families, 99 genera and 215 species  
200 were retained in the following analyses (Fig. 2).

201

## 202 **Community-level Analysis: Diversity Calculation**

203 As in MiCloud [27], MiPair [28] and MiSurv [29], MiMed calculates 9 alpha-diversity indices  
204 (i.e., 8 non-phylogenetic indices: Observed, Shannon [34], Simpson [35], Inverse Simpson [35],  
205 Fisher [36], Chao1 [37], abundance-based coverage estimator (ACE) [38], incidence-based

206 coverage estimator (ICE) [39]; 1 phylogenetic index: phylogenetic diversity (PD) [40]) and 5 beta-  
207 diversity indices (i.e., 2 non-phylogenetic indices: Jaccard dissimilarity [41], Bray-Curtis  
208 dissimilarity [42]; 3 phylogenetic indices: Unweighted UniFrac distance [43], Generalized  
209 UniFrac distance [44], Weighted UniFrac distance [45]). For reference, users can download all the  
210 calculated alpha- and beta-diversity indices.

211 **Application Note:** We simply clicked the Run button to calculate all the alpha- and beta-diversity  
212 indices.

213

## 214 **Community-level Analysis: Alpha Diversity**

215 This module analyzes if a treatment alters alpha-diversity, and then the altered alpha-diversity, in  
216 turn, influences an outcome, where the alpha-diversity can be surveyed using each of the 9 alpha-  
217 diversity indices. Users first need to select (1) a treatment variable (e.g., diet, residence, smoking,  
218 preterm birth, delivery mode, antibiotic/probiotic use), (2) an outcome variable (e.g., health or  
219 disease status), (3) to include an interaction term between a treatment and a mediator (alpha-  
220 diversity) in the model or not, and (4) covariates (e.g., age, gender) to be adjusted for. We set the  
221 interaction term to be included (yes) as default since it is more natural to assume that the effect of  
222 microbiome on an outcome can be modulated by a treatment. That is, in other words, the effect of  
223 microbiome on an outcome can be different by treatment status. Ignoring the presence of such  
224 interaction effects may cause potential bias in mediation analysis [46, 47]. The only available  
225 analytic method that can address interaction effect is the Imai method [48]) [Table 1]. The Imai  
226 method [48] in addition allows covariate adjustments, estimates mediation effects in both point  
227 and interval estimation, and reports a *P*-value for significance testing. The other available analytic  
228 methods are two traditional (but still in wide use) methods, the Sobel test [49] and Preacher-Hayes

229 approach [50, 51], and one recent method, named Divide-Aggregate Composite-null Test (DACT)  
230 [52]. We organized available functions for different mediation analysis methods in [Table 1]. We  
231 discuss them further later in the *Methods* section. MiMed visualizes the results from its alpha-  
232 diversity analysis using forest plots.

233 **Application Note:** We selected e-cigarette smoking as a treatment variable, gingival inflammation  
234 as an outcome variable, and age, sex and the frequency of brushing teeth as covariates to be  
235 adjusted for in the presence of interaction between e-cigarette smoking and alpha-diversity. Then,  
236 we found significant results using the Imai method [48] as e-cigarette smoking alters alpha-  
237 diversity of the oral microbiome in subgingival niches, and the altered alpha-diversity, in turn,  
238 influences gingival inflammation according to Observed, Shannon [34], InvSimpson [35], Fisher  
239 [36], Chao1 [37], ACE [38] and ICE [39] indices (Fig. 3).

240

## 241 **Community-level Analysis: Beta Diversity**

242 This module analyzes if a treatment alters beta-diversity, and then the altered beta-diversity, in  
243 turn, influences an outcome, where the beta-diversity can be surveyed using each of the 5 beta-  
244 diversity indices. As for alpha-diversity analysis, users need to select (1) a treatment variable (e.g.,  
245 diet, residence, smoking, preterm birth, delivery mode, antibiotic/probiotic use), (2) an outcome  
246 variable (e.g., health or disease status), and (3) covariates (e.g., age, gender) to be adjusted for.  
247 MedTest [53] is currently the only available analytic method that can conduct causal mediation  
248 analysis for beta-diversity [Table 1]. While MedTest [53] allows covariate adjustments and reports  
249 a *P*-value for significance testing, it is purely a test for significance with no facilities for mediation  
250 effect estimation [Table 1]. Furthermore, it does not allow any interaction term to be included

251 [Table 1]. MiMed visualizes the results from its beta-diversity analysis using principal coordinate  
252 analysis (PCoA) plots [54].

253 **Application Note:** We selected e-cigarette smoking as a treatment variable, gingival inflammation  
254 as an outcome variable, and age, sex and the frequency of brushing teeth as covariates to be  
255 adjusted for. Then, we found significant results using MedTest [53] as e-cigarette smoking alters  
256 beta-diversity of the oral microbiome in subgingival niches, and the altered beta-diversity, in turn,  
257 influences gingival inflammation according to Jaccard dissimilarity [41], Bray-Curtis dissimilarity  
258 [42], Generalized UniFrac distance [44] and Weighted UniFrac distance [45] (Fig. 4).

259

## 260 **Taxonomy-level Analysis: Data Normalization**

261 MiMed normalizes taxonomic relative abundances using CLR [24] and arcsine-root [55]  
262 transformations. The CLR transformation is the most widely used normalization method in the  
263 microbiome field to relax the compositional constraint of the data [24]. The arcsine-root  
264 transformation is a traditional approach to stabilize the variance of relative abundances [55]. The  
265 arcsine-root transformation has also recently been often used in the microbiome field [25]. For  
266 reference, users can download all the original count and normalized taxonomic data for microbial  
267 taxa at various taxonomic hierarchies (i.e., phyla, classes, orders, families, genera, species).

268 **Application Note:** We simply clicked the Run button to normalize taxonomic relative abundances.

269

## 270 **Taxonomy-level Analysis: Taxonomic Analysis**

271 This module analyzes if a treatment alters microbial taxa, and then the altered microbial taxa, in  
272 turn, influence an outcome. Users first need to select a data format, CLR [24] or arcsine-root  
273 transformed data. There has been a long debate on which data normalization method is the best,

274 but it is beyond the scope of this paper to make any resolute judgement on it. We set CLR [24] as  
275 default and arcsine-root as a user option based on their popularities. As for alpha-diversity analysis,  
276 users then need to select (1) a treatment variable (e.g., diet, residence, smoking, preterm birth,  
277 delivery mode, antibiotic/probiotic use), (2) an outcome variable (e.g., health or disease status), (3)  
278 to include an interaction term between a treatment and a mediator (taxon) in the model or not, and  
279 (4) covariates (e.g., age, gender) to be adjusted for. Again, the only available analytic method that  
280 can address interaction effect is the Imai method [48] [Table 1]. Importantly, the Imai method [48]  
281 is a non-parametric method based on a bootstrap approach [56]. Thus, it is highly robust against  
282 the high skewness of microbiome data, especially the rare microbial taxa with excessive zeros [48].  
283 The other available analytic methods are two parametric methods, the Sobel test [49] and DACT  
284 [52] [Table 1]. We set the Imai method [48] as default and the Sobel test [49] and DACT [52] as  
285 user options [Table 1], which is because of the robust performance of the Imai method [48] as well  
286 as its broad range of functionalities [Table 1]. We discuss methodological details further later again  
287 in the *Methods* section. To control for false discovery rates (FDR), MiMed applies the Benjamini-  
288 Hochberg (BH) procedure [57] to each taxonomic hierarchy. MiMed visualizes the results from  
289 its taxonomic analyses using forest plots and dendrograms.

290 **Ask ChatGPT:** In this sub-module, users can ask ChatGPT a question: What is known about  
291 (discovered taxon) on (treatment) and (outcome)? For this, users first need to insert a ChatGPT  
292 APT key that can be freely obtained on the website (<https://platform.openai.com/account/api-keys>).  
293 Then, users need to select a taxonomic rank (i.e., phylum, class, order, family, genus, species) and  
294 a taxon that is discovered as a significant mediator, and then rename the treatment and outcome  
295 variables using a human language (not code names). Then, ChatGPT will answer your question.

296 **Application Note:** We selected CLR as a normalization method, e-cigarette smoking as a  
297 treatment variable, gingival inflammation as an outcome variable, and age, sex and the frequency  
298 of brushing teeth as covariates to be adjusted for. Then, we found 21 significant taxa at the  
299 taxonomic hierarchies from phylum to genus (i.e., two phyla: *Proteobacteria*, *Spirochaetes*, two  
300 classes: *Flavobacteriia*, *Betaproteobacteria*, four orders: *Flavobacteriales*, *Burkholderiales*,  
301 *Neisseriales*, *Cardiobacteriales*, five families: *Flavobacteriaceae*, *Burkholderiaceae*,  
302 *Neisseriaceae*, *Cardiobacteriaceae*, *Enterococcaceae*, and eight genera: *Bergeyella*,  
303 *Capnocytophaga*, *Actinomyces*, *Haemophilus*, *Kingella*, *Burkholderia*, *Cardiobacterium*,  
304 *Enterococcus*) using the Imai method [48] as e-cigarette smoking alters their relative abundances,  
305 and the altered relative abundances, in turn, influence gingival inflammation (Fig. 5 and Fig. 6).  
306 We also asked ChatGPT a question, “What is known about *Bergeyella* on e-cigarette and gingival  
307 inflammation?”, selecting a taxonomic rank as genus and a discovered taxon as *Bergeyella*, and  
308 renaming the treatment and outcome variables as e-cigarette and gingival inflammation. Then,  
309 ChatGPT aided to reconfirm our results answering as “*Bergeyella* is a gram-negative bacteria that  
310 is commonly found in the oral microbiome. It has been associated with various oral diseases  
311 including gingivitis and periodontitis. There is some preliminary evidence suggesting that e-  
312 cigarettes may contribute to gingival inflammation and changes in the oral microbiome, including  
313 an increase in the abundance of *Bergeyella*. A study published in Scientific Reports in 2019  
314 reported that e-cigarette users had higher levels of gingival inflammation and an altered oral  
315 microbiota compared to non-smokers or smokers of traditional cigarettes. Specifically, the study  
316 found that e-cigarette users had increased levels of *Bergeyella*, as well as other potentially  
317 pathogenic bacteria, in their oral microbiota. While the relationship between *Bergeyella* and e-

318 cigarettes is still being explored and more research is needed to fully understand the implications,  
319 these findings suggest that e-cigarette use may have negative effects on oral health.” (Fig. 7).

320

## 321 **Discussion**

322

323 Researchers are interested in discovering causal mechanisms through which environmental,  
324 behavioral or medical exposures influence health or disease outcomes. A promising approach has  
325 been to use mediation analysis, though it is highly demanding in the human microbiome field. The  
326 microbiome data are huge and highly complex, and many researchers are not familiar with dealing  
327 with such microbiome data. Thus, we need a well-designed ‘software’ that enables user-friendly  
328 operations for microbiome causal mediation analysis.

329 In this paper, we introduced MiMed that is the first web cloud computing platform for  
330 microbiome causal mediation analysis. MiMed enables a long sequence of data processing and  
331 analytic operations on user-friendly web interfaces with widely extended flexibility and  
332 functionality. MiMed surveys the microbiome in various spheres as a whole ecosystem or as  
333 individual microbial taxa at various taxonomic hierarchies. MiMed also enables covariate-adjusted  
334 analysis and a breadth of statistical inferences in both mediation effect estimation and significance  
335 testing. MiMed also provides step-by-step data processing and analytic modules, and creates high-  
336 quality visualizations. Interestingly, MiMed is also built-in the recent popular chatbot, ChatGPT,  
337 to easily search for prior knowledge on discovered taxa using AI technologies. The ChatGPT  
338 module can be useful to re-confirm the analysis results from MiMed.

339 MiMed is comprehensive and built with many data processing and analytic approaches. It is  
340 usual in the human microbiome field that there is no consensus on which approach is always the

341 best. That is, there is not anything that is superior to the others in all contexts and situations. We  
342 are also curious about many different approaches. Thus, we left much room for our users to freely  
343 explore through many user options, while making a series of recommendations, as a developer,  
344 through default settings. For user’s convenience, MiMed also displays a list of references for the  
345 approaches that they use.

346 The human microbiome field is rapidly emerging, and the microbiome data are recently  
347 flooded. Yet, the microbiome data are demanding, and we are all so busy. Thus, MiMed can be  
348 attractive and useful in practice because it is user-friendly. MiMed will also provide new insights  
349 to the human microbiome field through causal mediation analysis that is “too important to abandon”  
350 [48].

351

## 352 **Methods**

353

### 354 **Statistical Methods**

355 This section is devoted to describing the methodological aspects of the causal mediation analysis  
356 methods that are available in MiMed. We describe only the conceptual ideas and terms to help our  
357 users to easily understand them, while referencing the original papers for all technical details.

358 To begin with the Sobel test [49], Preacher-Hayes approach [50, 51]), and DACT [52], the  
359 Baron and Kenny’s two regression models [20] below can first be considered.

$$M_i = \alpha_0 + \alpha_1 T_i + \varepsilon_i \quad (1)$$

$$Y_i = \beta_0 + \beta_1 M_i + \beta_1 T_i + v_i \quad (2)$$

360 , where  $T_i$  is a treatment,  $M_i$  is a mediator (e.g., an alpha-diversity index or a microbial taxon),  $Y_i$   
361 is a health or disease outcome,  $\alpha_0$  and  $\beta_0$  are intercepts,  $\alpha_1$  and  $\beta_1$  are slopes, and  $\varepsilon_i$  and  $v_i$  are



362 independently distributed random errors with mean zero and variance  $\sigma^2$  for the units  $i = 1, \dots, n$ .  
363 To ease our demonstration, we suppose in addition that  $T_i$  is a binary treatment variable ( $T_i = 0$  for  
364 control &  $T_i = 1$  for treatment), and  $Y_i$  is a continuous health or disease outcome variable. Yet,  
365 more extensions are available [Table 1]. Then, the null and alternative hypotheses below are  
366 considered.

$$H_0: \alpha_1\beta_1 = 0 \text{ vs. } H_1: \alpha_1\beta_1 \neq 0 \quad (3)$$

367 Here,  $\alpha_1$  represents the effect of the treatment ( $T_i$ ) on the mediator ( $M_i$ ) as in Eq. (1), and  $\beta_1$   
368 represents the effects of the mediator ( $M_i$ ) on the outcome ( $Y_i$ ) conditional on treatment status ( $T_i$ )  
369 as in Eq. (2). Then, the null hypothesis,  $H_0: \alpha_1\beta_1 = 0$ , states that at least one of  $\alpha_1$  and  $\beta_1$  equals  
370 to zero indicating no mediation effect, while the alternative hypothesis,  $H_1: \alpha_1\beta_1 \neq 0$ , states that  
371 both  $\alpha_1$  and  $\beta_1$  are non-zero indicating the presence of mediation effect. The Sobel test [49]  
372 conducts significance testing for Eq. (3) using a parametric approach that assumes that  $\varepsilon_i$  and  $v_i$   
373 in Eq. (1) and Eq. (2) are normally distributed. In contrast, the Preacher-Hayes approach [50, 51]  
374 does it non-parametrically using a bootstrap method [56] without the normality assumption. As for  
375 the Sobel test [49], DACT [52] is a parametric approach, but considers the null hypothesis,  $H_0:$   
376  $\alpha_1\beta_1 = 0$ , in Eq. (3) as a composite hypothesis that  $H_0: (1) \alpha_1 = 0 \ \& \ \beta_1 \neq 0; (2) \alpha_1 \neq 0 \ \& \ \beta_1 = 0;$   
377 or (3)  $\alpha_1 = 0 \ \& \ \beta_1 = 0$ ; to improve statistical power while rejecting  $H_0$  for at least one of the three  
378 sub-statements.

379 As for DACT [52], MedTest [53] considers the null hypothesis as a composite hypothesis, but  
380 it is a non-parametric significance test based on a permutation method. A more important  
381 distinction is that MedTest [53] formulates the mediator ( $M_i$ ) in Eq. (1) and Eq. (2) as a function  
382 of beta-diversity (say,  $f(M)_i$ , where  $f(\cdot)$  is a function that transforms microbiome into a beta-  
383 diversity index); as such, it enables causal mediation analysis for beta-diversity [Table 1].

384 We can classify the Sobel test [49], Preacher-Hayes approach [50, 51], DACT [52] and MedTest  
385 [53] as ‘product-of-coefficients’ methods because of their shared hypothesis of Eq. (3) in the form  
386 of  $\alpha_1\beta_1$  (i.e., the product of coefficients from Eq. (1) and Eq. (2)). However, the Imai method [48]  
387 in contrast is based on a potential outcomes framework of causal inference [58]  $Y_i(T_i, M_i(T_i))$ ,  
388 where the level of health or disease outcome is a function of a treatment status (i.e.,  $T_i$ ) and the  
389 level of the mediator under a treatment status (i.e.,  $M_i(T_i)$ ). Then, the unit-level ‘total treatment  
390 effect’ can be defined as Eq. (4), the unit-level ‘direct effect (DE)’ on the mediator can be defined  
391 for each treatment status ( $t = 0$  for control or  $t = 1$  for treatment) as Eq. (5), and finally the unit-  
392 level ‘indirect effect or causal mediation effect (CME)’ can be defined for each treatment status ( $t$   
393  $= 0$  for control or  $t = 1$  for treatment) as Eq. (6),

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \quad (4)$$

$$\zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t)) \quad (5)$$

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \quad (6)$$

394 Here, the unit-level total treatment effect in Eq. (4) was formulated by subtracting the level of  
395 health or disease outcome for the unit under control and the level of the mediator under control  
396 from the level of health or disease outcome for the same unit under treatment and the level of the  
397 mediator under treatment. The unit-level DE for each treatment status (i.e., for control or treatment)  
398 in Eq. (5) was formulated by subtracting the level of health or disease outcome for the unit with  
399 under control from the level of health or disease outcome for the same unit under treatment. Finally,  
400 the unit-level CME for each treatment status (i.e., for control or treatment) in Eq. (6) was  
401 formulated by subtracting the level of health or disease outcome for the unit with the level of the

402 mediator under control from the level of health or disease outcome for the same unit with the level  
403 of the mediator under treatment.

404 Then, the overall ‘average direct effect (ADE)’ can be found by the average between the  
405 average direct effect with the level of mediator under control,  $\frac{1}{n} \sum_{i=1}^n \zeta_i(0)$ , the average direct effect  
406 with the level of mediator under treatment,  $\frac{1}{n} \sum_{i=1}^n \zeta_i(1)$ . Finally, the overall ‘average causal  
407 mediation effect (ACME)’, that is the main result in causal mediation analysis, can be found by  
408 the average between the average causal mediation effect for control,  $\frac{1}{n} \sum_{i=1}^n \delta_i(0)$ , and the average  
409 causal mediation effect for treatment,  $\frac{1}{n} \sum_{i=1}^n \delta_i(1)$ .

410 The Imai method [48] conducts interval estimation for ACME (overall) (as well as ACME  
411 (control), ACME (treatment), ADE (overall), ADE (control), ADE (treatment)) using a bootstrap  
412 method [56] non-parametrically, and its significance testing follows accordingly.

413 There has been a long debate on parametric vs. non-parametric, but it is also beyond the scope  
414 of this paper to make any resolute judgement on it. However, it is usual that non-parametric  
415 approaches are more robust to highly skewed data (e.g., rare taxa with excessive zeros), while  
416 parametric approaches are well suited to less skewed data (e.g., alpha-diversity indices or common  
417 taxa). However, so long as the sample size is large, the skewness does not also substantially matter  
418 for parametric approaches. However, it does not also mean that non-parametric approaches are not  
419 suited to a large sample size. Parametric approaches are not well suited to high skewed data with  
420 a small sample size. Since the microbiome data are usually highly skewed (Fig. 2), we set non-  
421 parametric approaches as default, but we do not discourage the use of parametric approaches,  
422 which are also widely used and reasonable approaches for a large sample size [Table 1].

## 423 **Web Server and Local GitHub Repository**

424 As for MiCloud [27], MiPair [28] and MiSurv [29], we wrote all the user interfaces and server  
425 functions using R shiny (<https://shiny.rstudio.com>). We then developed our web server using  
426 ShinyProxy (<https://www.shinyproxy.io>) and Apache2 (<https://httpd.apache.org>) on the operating  
427 system, Ubuntu 20.04 (<https://ubuntu.com>). The web server currently runs on a computer with the  
428 specifications of Intel Core i7-12700T (12-core) processor and 36 GB DDR4 memory, and takes  
429 up to ten concurrent users. In case that the web server is busy, we also developed a local GitHub  
430 repository to enable to run MiMed using a user's local computers. As usual, we, as a host, are  
431 responsible for and devoted to maintaining our web server and local GitHub repository reliable.

432

## 433 **Data Availability**

434 We used public microbiome data, where the raw sequence data are deposited at the NCBI Gene  
435 Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under access number GSE201949. The  
436 processed data can also be found in the Example Data section on the Data Input module of MiMed  
437 (<http://mimed.micloud.kr>).

438

## 439 **Code Availability**

440 MiMed is freely available on our web server (<http://mimed.micloud.kr>) or can alternatively run on  
441 a user's local computer (<https://github.com/yj7599/MiMedGit>).

442

## 443 **Author Contributions**

444 HJ wrote the programs, developed the shiny application, web server and local GitHub repository,  
445 performed data analysis, and contributed to the visualization and design. SP wrote the programs,  
446 developed the shiny application, performed data analysis, and contributed to the visualization and  
447 design. HK conceptualized and initiated the study, curated the data, contributed to the  
448 methodological aspects, overall architecture, and design, performed data analysis, and wrote the  
449 manuscript. All authors reviewed, revised, and approved the final manuscript.

450

### 451 **Competing Interests**

452 The authors declare that they have no competing interests.

453

### 454 **Acknowledgements**

455 The work was supported by the National Research Foundation of Korea (NRF) grant funded by  
456 the Korean government (MSIT) (2021R1C1C1013861).

457

458

459

460

461

462

463

464

465

466

## 467 **References**

468

- 469 1. Singh RK, Chang HW, Yan D, Lee KM, Ucmak D, Wong K, Abrouk M, Farahnik B,  
470 Nakamura M, Zhu TH, Bhutani T, Liao W. 2017. Influence of diet on the gut microbiome  
471 and implications for human health. *J Transl Med* 15:1-7.
- 472 2. Oduaran OH, Tamburini FB, Sahibdeen V, Brewster R, Gómez-Olivé FX, Kahn K, Norris  
473 SA, Tollman SM, Twine R, Wade AN, Wagner RG. 2020. Gut microbiome profiling of a  
474 rural and urban South African cohort reveals biomarkers of a population in lifestyle  
475 transition. *BMC Microbiol* 20:1-7.
- 476 3. Gui X, Yang Z, Li MD. 2021. Effect of cigarette smoke on gut microbiota: state of  
477 knowledge. *Front Physiol* 12:673341.
- 478 4. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz  
479 TJ, Edupuganti L, Glascock AL, Xu J. 2019. The vaginal microbiome and preterm birth.  
480 *Nat Med* 25:1012-21.
- 481 5. Reyman M, van Houten MA, van Baarle D, Bosch AA, Man WH, Chu ML, Arp K,  
482 Watson RL, Sanders EA, Fuentes S, Bogaert D. 2019. Impact of delivery mode-associated  
483 gut microbiota dynamics on health in the first year of life. *Nat Commun* 10:4997.
- 484 6. Zhang C, Li L, Jin B, Xu X, Zuo X, Li Y, Li Z. 2021. The effects of delivery mode on the  
485 gut microbiota and health: state of art. *Front Microbiol* 12:4084.
- 486 7. Hemarajata P, Versalovic J. 2013. Effects of probiotics on gut microbiota: mechanisms of  
487 intestinal immunomodulation and neuromodulation. *Therap Adv Gastroenterol* 6:39-51.

- 488 8. Zhang XS, Li J, Krautkramer KA, Badri M, Battaglia T, Borbet TC, Koh H, Ng S, Sibley  
489 RA, Li Y, Pathmasiri W. 2018. Antibiotic-induced acceleration of type 1 diabetes alters  
490 maturation of innate intestinal immunity. *Elife* 7:e37816.
- 491 9. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An  
492 obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*  
493 444:1027-1031.
- 494 10. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V,  
495 Henrissat B, Bain JR, Muehlbauer MJ. 2013. Gut microbiota from twins discordant for  
496 obesity modulate metabolism in mice. *Science* 341:1241214.
- 497 11. Touw K, Ringus DL, Hubert N, Wang Y, Leone VA, Nadimpalli A, Theriault BR, Huang  
498 YE, Tune JD, Herring PB, Farrugia G. 2017. Mutual reinforcement of pathophysiological  
499 host-microbe interactions in intestinal stasis models. *Physiol Rep* 5:e13182.
- 500 12. De Palma G, Lynch MD, Lu J, Dang VT, Deng Y, Jury J, Umeh G, Miranda PM, Pigrau  
501 Pastor M, Sidani S, Pinto-Sanchez MI. 2017. Transplantation of fecal microbiota from  
502 patients with irritable bowel syndrome alters gut function and behavior in recipient mice.  
503 *Sci Transl Med* 9:eaaf6397.
- 504 13. Johnsen PH, Hilpüsch F, Cavanagh JP, Leikanger IS, Kolstad C, Valle PC, Goll R. 2018.  
505 Faecal microbiota transplantation versus placebo for moderate-to-severe irritable bowel  
506 syndrome: a double-blind, randomised, placebo-controlled, parallel-group, single-centre  
507 trial. *Lancet Gastroenterol Hepatol* 3:17-24.

- 508 14. Frankel AE, Coughlin LA, Kim J, Froehlich TW, Xie Y, Frenkel EP, Koh AY. 2017.  
509 Metagenomic shotgun sequencing and unbiased metabolomic profiling identify specific  
510 human gut microbiota and metabolites associated with immune checkpoint therapy  
511 efficacy in melanoma patients. *Neoplasia* 19:848-855.
- 512 15. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, Prieto  
513 PA, Vicente D, Hoffman K, Wei SC, Cogdill AP. 2018. Gut microbiome modulates  
514 response to anti-PD-1 immunotherapy in melanoma patients. *Science* 359:97-103.
- 515 16. Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML, Luke JJ, Gajewski TF.  
516 2018. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic  
517 melanoma patients. *Science* 359:104-108.
- 518 17. Livanos AE, Greiner TU, Vangay P, Pathmasiri W, Stewart D, McRitchie S, Li H, Chung  
519 J, Sohn J, Kim S, Gao Z. 2016. Antibiotic-mediated gut microbiome perturbation  
520 accelerates development of type 1 diabetes in mice. *Nat Microbiol* 1:1-3.
- 521 18. Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, Challis C, Schretter  
522 CE, Rocha S, Gradinaru V, Chesselet MF. 2016. Gut microbiota regulate motor deficits  
523 and neuroinflammation in a model of Parkinson's disease. *Cell* 167:1469-1480.
- 524 19. Kang DW, Adams JB, Gregory AC, Borody T, Chittick L, Fasano A, Khoruts A, Geis E,  
525 Maldonado J, McDonough-Means S, Pollard EL. 2017. Microbiota transfer therapy alters  
526 gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study.  
527 *Microbiome* 5:1-6.



- 528 20. Baron RM, Kenny DA. 1986. The moderator-mediator variable distinction in social  
529 psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc*  
530 *Psychol* 51:1173-82.
- 531 21. Tukey JW. 1958. The teaching of concrete mathematics. *Amer Math Monthly* 65:1-9.
- 532 22. Zhao N, Chen J, Carroll IM, Ringel-Kulka T, Epstein MP, Zhou H, Zhou JJ, Ringel Y, Li  
533 H, Wu MC. 2015. Testing in microbiome-profiling studies with MiRKAT, the microbiome  
534 regression-based Kernel Association Test. *Am J Hum Genet* 96:797-807.
- 535 23. Wilson N, Zhao N, Zhan X, Koh H, Fu W, Chen J, Li H, Wu MC, Plantinga AM. 2021.  
536 MiRKAT: kernel machine regression-based global association tests for the microbiome.  
537 *Bioinformatics* 37:1595-1597.
- 538 24. Aitchison J. 1982. The statistical analysis of compositional data. *J R Stat Soc Series B*  
539 44:139-160.
- 540 25. Zhu Z, Satten GA, Hu YJ. 2022. Integrative analysis of relative abundance data and  
541 presence-absence data of the microbiome using the LDM. *Bioinformatics* 38:2915-2917.
- 542 26. Leviatan S, Shoer S, Rothschild D, Gorodetski M, Segal E. 2022. An expanded reference  
543 map of the human microbiome reveals hundreds of previously unknown species. *Nat*  
544 *Commun* 13:3863.
- 545 27. Gu W, Moon J, Chisina C, Kang B, Park T, Koh H. 2022. MiCloud: A unified web  
546 platform for comprehensive microbiome data analysis. *PLoS One* 17:e0272354.

- 547 28. Jang H, Koh H, Gu W, Kang B. 2022. Integrative web cloud computing and analytics  
548 using MiPair for design-based comparative analysis with paired microbiome data. *Sci Rep*  
549 12:20465.
- 550 29. Gu W, Koh H, Jang H, Lee B, Kang B. 2023. MiSurv: An integrative web cloud platform  
551 for user-friendly microbiome data analysis with survival responses. *Microbiol Spectr*  
552 11:e05059-22.
- 553 30. Park B, Koh H, Patatianian M, Reyes-Caballero H, Zhao N, Meinert J, Holbrook JT,  
554 Leinbach LI, Biswal S. 2023. The mediating roles of the oral microbiome in saliva and  
555 subgingival sites between e-cigarette smoking and gingival inflammation. *BMC Microbiol*  
556 23:1-18.
- 557 31. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H,  
558 Alm EJ, Arumugam M, Asnicar F, Bai Y. 2019. Reproducible, interactive, scalable and  
559 extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852-857.
- 560 32. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. 2018. New insights into  
561 human nostril microbiome from the expanded human oral microbiome database  
562 (eHOMD): a resource for the microbiome of the human aerodigestive tract. *MSystems*  
563 3:e00187-18.
- 564 33. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive  
565 analysis and graphics of microbiome census data. *PLoS One* 8:e61217.
- 566 34. Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J* 27:379-  
567 423.

- 568 35. Simpson EH. 1949. Measurement of diversity. *Nature* 163:688.
- 569 36. Fisher RA, Corbet AS, Williams CB. 1943. The relation between the number of species  
570 and the number of individuals in a random sample of an animal population. *J Anim Ecol*  
571 12:42-58.
- 572 37. Chao A. 1984. Non-parametric estimation of the number of classes in a population. *Scand*  
573 *J Stat* 11:265-70.
- 574 38. Chao A, Lee SM. 1992. Estimating the number of classes via sample coverage. *J Am Stat*  
575 *Assoc* 87:210-217.
- 576 39. Lee SM, Chao A. 1994. Estimating population size via sample coverage for closed  
577 capture-recapture models. *Biometrics* 50:88-97.
- 578 40. Faith DP. 1992. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1-10.
- 579 41. Jaccard P. 1912. The distribution of the flora in the alpine zone. *New Phytol* 11:37-50.
- 580 42. Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern  
581 Wisconsin. *Ecol Monogr* 27:325-49.
- 582 43. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing  
583 microbial communities. *Appl Environ Microbiol* 71:8228-8235.
- 584 44. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman  
585 FD, Li H. 2012. Associating microbiome composition with environmental covariates using  
586 generalized UniFrac distances. *Bioinformatics* 28:2106-2113.

- 587 45. Lozupone CA, Hamady M, Kelley ST, Knight R. 2007. Quantitative and qualitative beta  
588 diversity measures lead to different insights into factors that structure microbial  
589 communities. *Appl Environ Microbiol* 73:1576-1585.
- 590 46. Richiardi L, Bellocco R, Zugna D. 2013. Mediation analysis in epidemiology: methods,  
591 interpretation and bias. *Int J Epidemiol* 42:1511-1519.
- 592 47. Valeri L, VanderWeele TJ. 2013. Mediation analysis allowing for exposure-mediator  
593 interactions and causal interpretation: theoretical assumptions and implementation with  
594 SAS and SPSS macros. *Psychol Methods* 18:137-150.
- 595 48. Imai K, Keele L, Tingley D. 2010. A general approach to causal mediation analysis.  
596 *Psychol Methods* 15:309-334.
- 597 49. Sobel ME. 1982. Asymptotic confidence intervals for indirect effects in structural equation  
598 models. *Sociol Methodol* 13:290-312.
- 599 50. Preacher KJ, Hayes AF. 2004. SPSS and SAS procedures for estimating indirect effects in  
600 simple mediation models. *Behav Res Methods* 36:717-731.
- 601 51. Preacher KJ, Hayes AF. 2008. Asymptotic and resampling strategies for assessing and  
602 comparing indirect effects in multiple mediator models. *Behav Res Methods* 40:879-91.
- 603 52. Liu Z, Shen J, Barfield R, Schwartz J, Baccarelli AA, Lin X. 2022. Large-scale hypothesis  
604 testing for causal mediation effects with applications in genome-wide epigenetic studies. *J*  
605 *Am Stat Assoc* 117:67-81.
- 606 53. Zhang J, Wei Z, Chen J. 2018. A distance-based approach for testing the mediation effect  
607 of the human microbiome. *Bioinformatics* 34:1875-1883.

608 54. Torgerson WS. 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* 17:  
609 401-419.

610 55. Warton DI, Hui FK. 2011. The arcsine is asinine: the analysis of proportions in ecology.  
611 *Ecology* 92:3-10.

612 56. Efron B. 1979. Bootstrap methods: another look at the jackknife. *Ann Statist* 7:1-26.

613 57. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and  
614 powerful approach to multiple testing. *J R Stat Soc Series B* 57:289-300.

615 58. Rubin DB. 1974. Estimating causal effects of treatments in randomized and  
616 nonrandomized studies. *J Educ Psychol* 66:688-701.

617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629

630 **Tables**

631

632 **Table 1. Descriptive table for the functionalities of causal mediation analysis methods: Imai method, Sobel test, Preacher-Hayes approach,**  
633 **DACT and MedTest.**

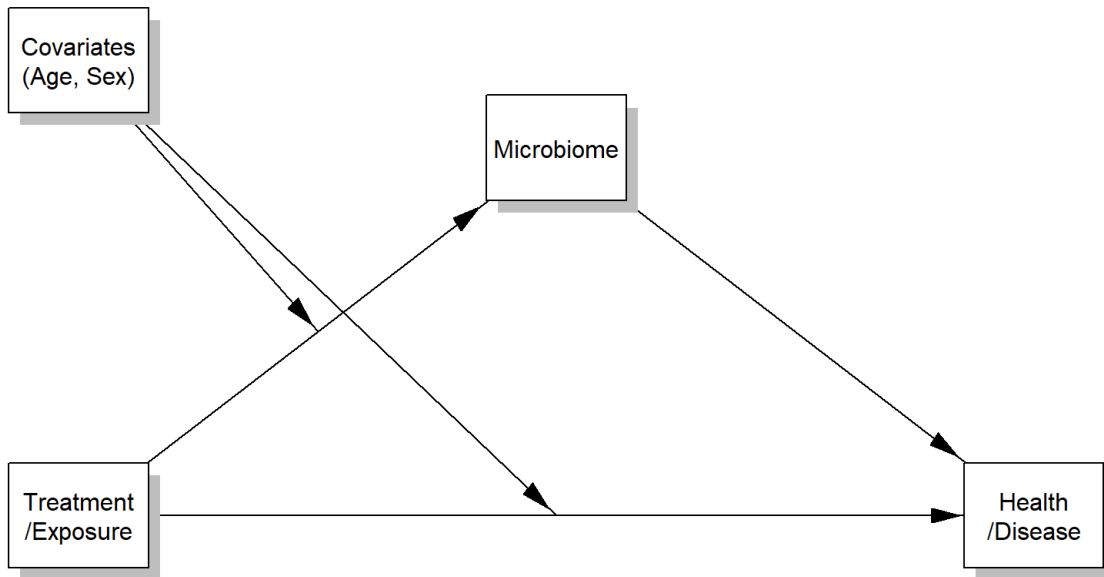
			Community-level Analysis					Taxonomy-level Analysis		
			Alpha Diversity				Beta Diversity			
Treatment Variable	Outcome Variable		Imai (Default)	Sobel	Preacher-Hayes	DACT	MedTest (Default)	Imai (Default)	Sobel	DACT
Binary	Binary	Interaction	O	X	X	X	X	O	X	X
		Covariates	O	X	O	O	O	O	X	O
		Point Estimation	O	X	O	O	X	O	X	O
		Interval Estimation	O	X	O	X	X	O	X	X
		P-value	O	X	X	O	O	O	X	O
	Continuous	Interaction	O	X	X	X	X	O	X	X
		Covariates	O	X	O	O	O	O	X	O
		Point Estimation	O	O	O	O	X	O	O	O
		Interval Estimation	O	X	O	X	X	O	X	X
		P-value	O	O	X	O	O	O	O	O
Continuous	Binary	Interaction	O	X	X	X	X	O	X	X
		Covariates	O	X	O	O	O	O	X	O
		Point Estimation	O	X	O	O	X	O	X	O
		Interval Estimation	O	X	O	X	X	O	X	X
		P-value	O	X	X	O	O	O	X	O
	Continuous	Interaction	O	X	X	X	X	O	X	X
		Covariates	O	X	O	O	O	O	X	O
		Point Estimation	O	O	O	O	X	O	O	O
		Interval Estimation	O	X	O	X	X	O	X	X
		P-value	O	O	X	O	O	O	O	O

634



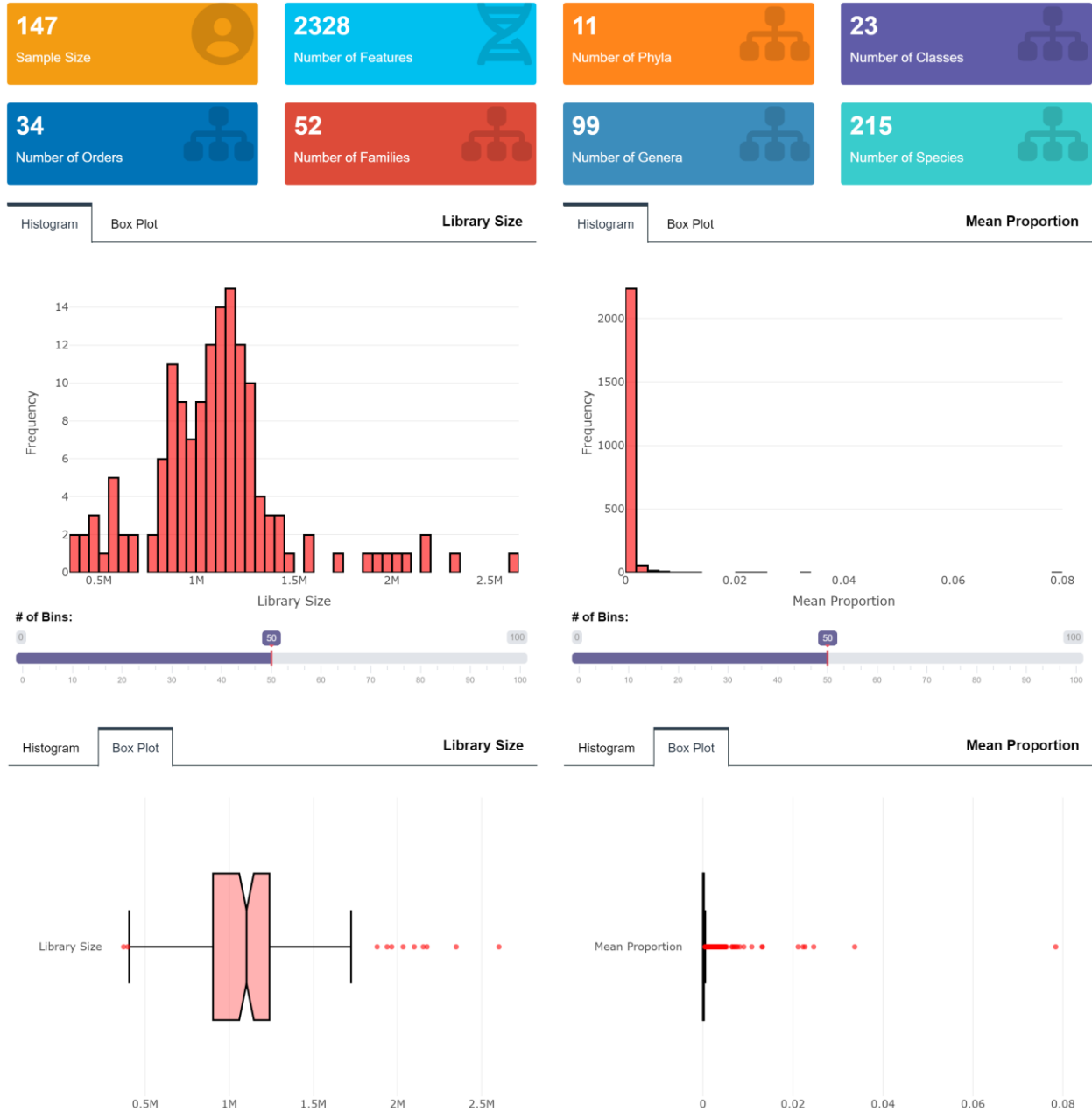
## Figures

**Fig. 1. A conceptual illustration for the roles of the microbiome as a mediator between a treatment/exposure and a health or disease outcome with potential covariate effects.**

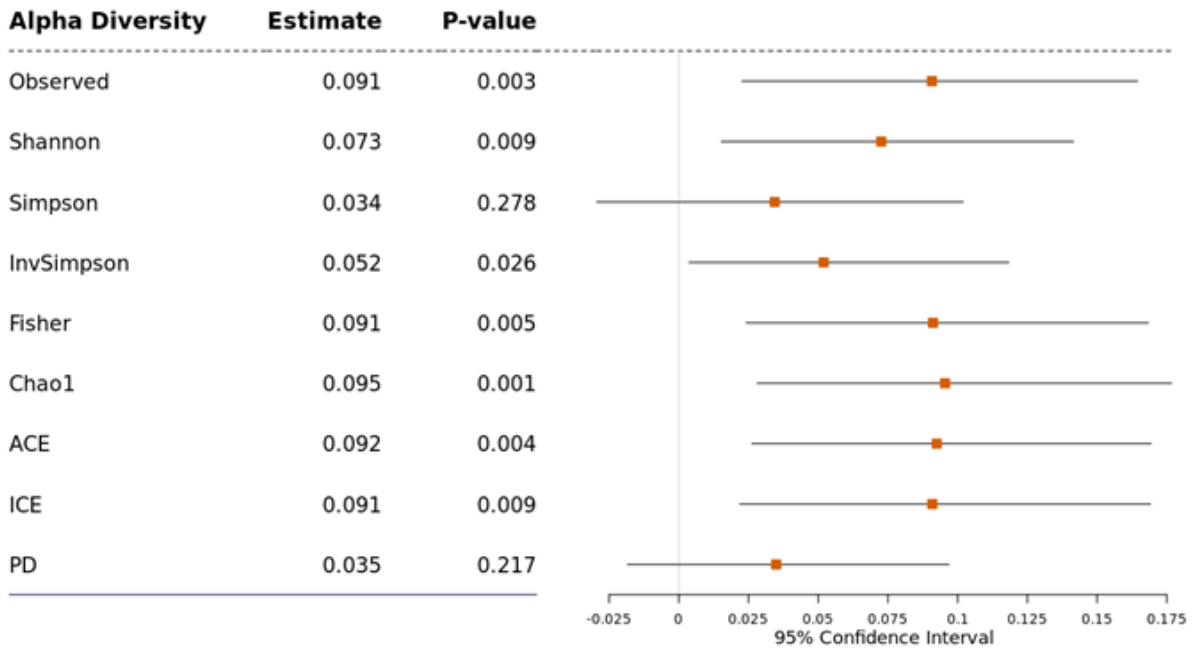




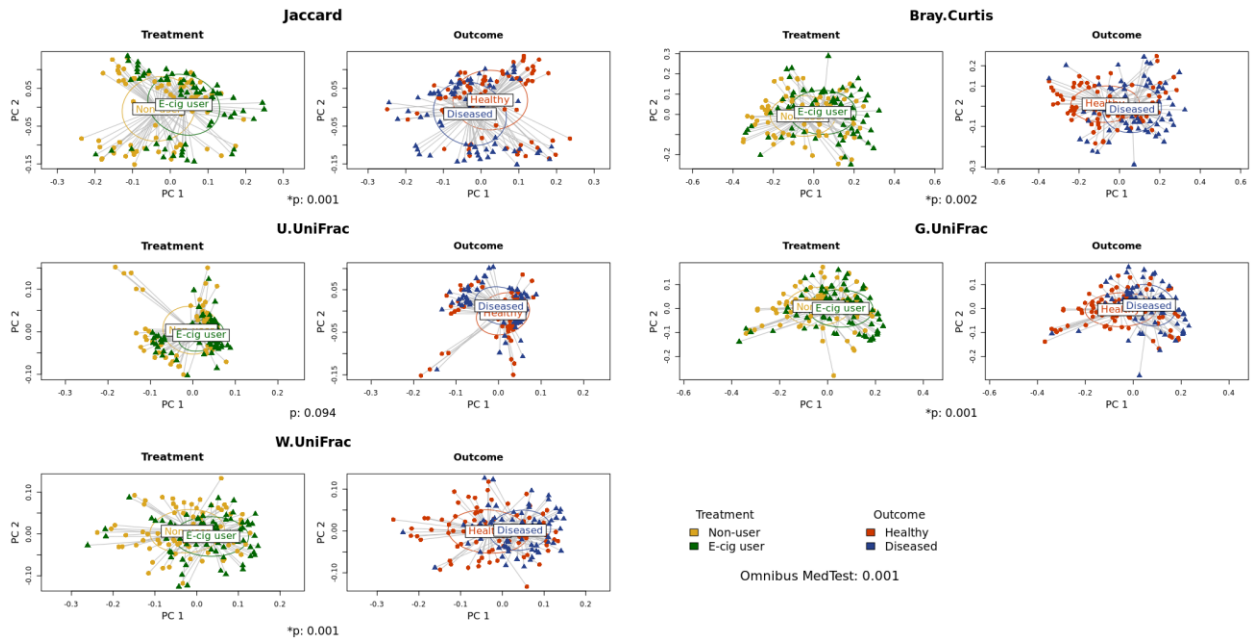
**Fig. 2. The status of the microbiome data after QCs.** The summary boxes below display the sample size, the number of features, the number of phyla, the number of classes, the number of orders, the number of families, the number of genera and the number of species after QCs. The histograms and box plots below visualize the library sizes across study subjects and the mean proportions across features.



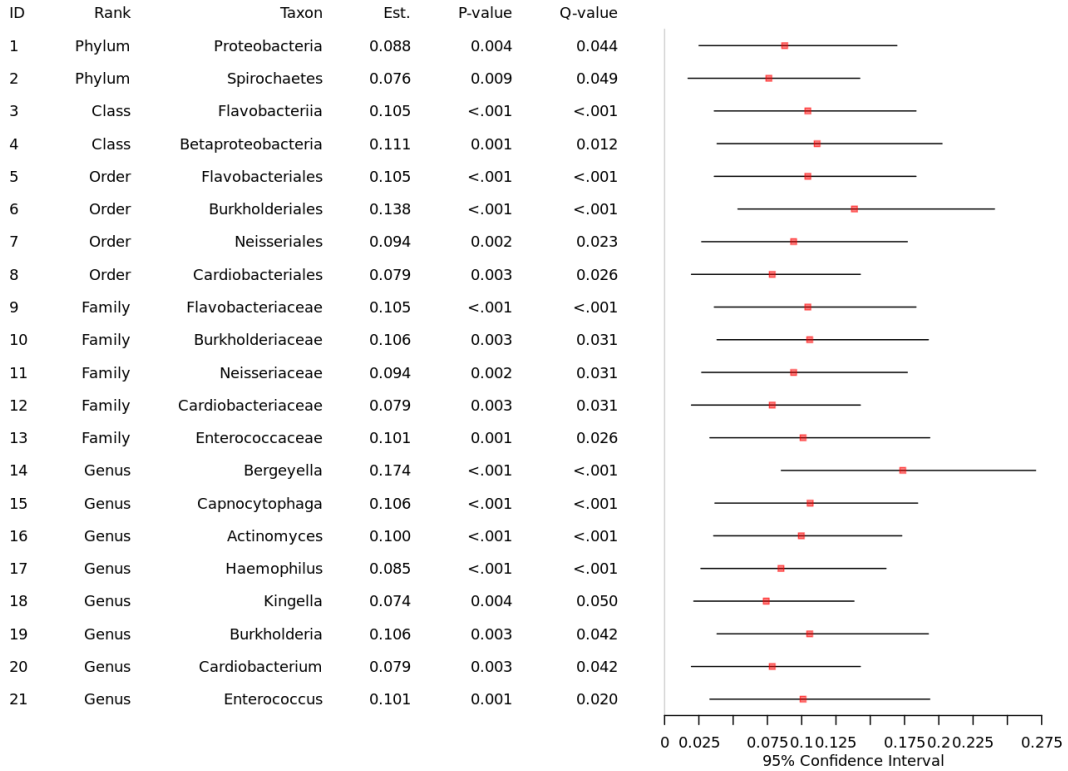
**Fig. 3. The results for alpha-diversity.** We surveyed if e-cigarette smoking alters alpha-diversity of the oral microbiome in subgingival niches, and the altered alpha-diversity, in turn, influences gingival inflammation, adjusting for age, sex and the frequency of brushing teeth.



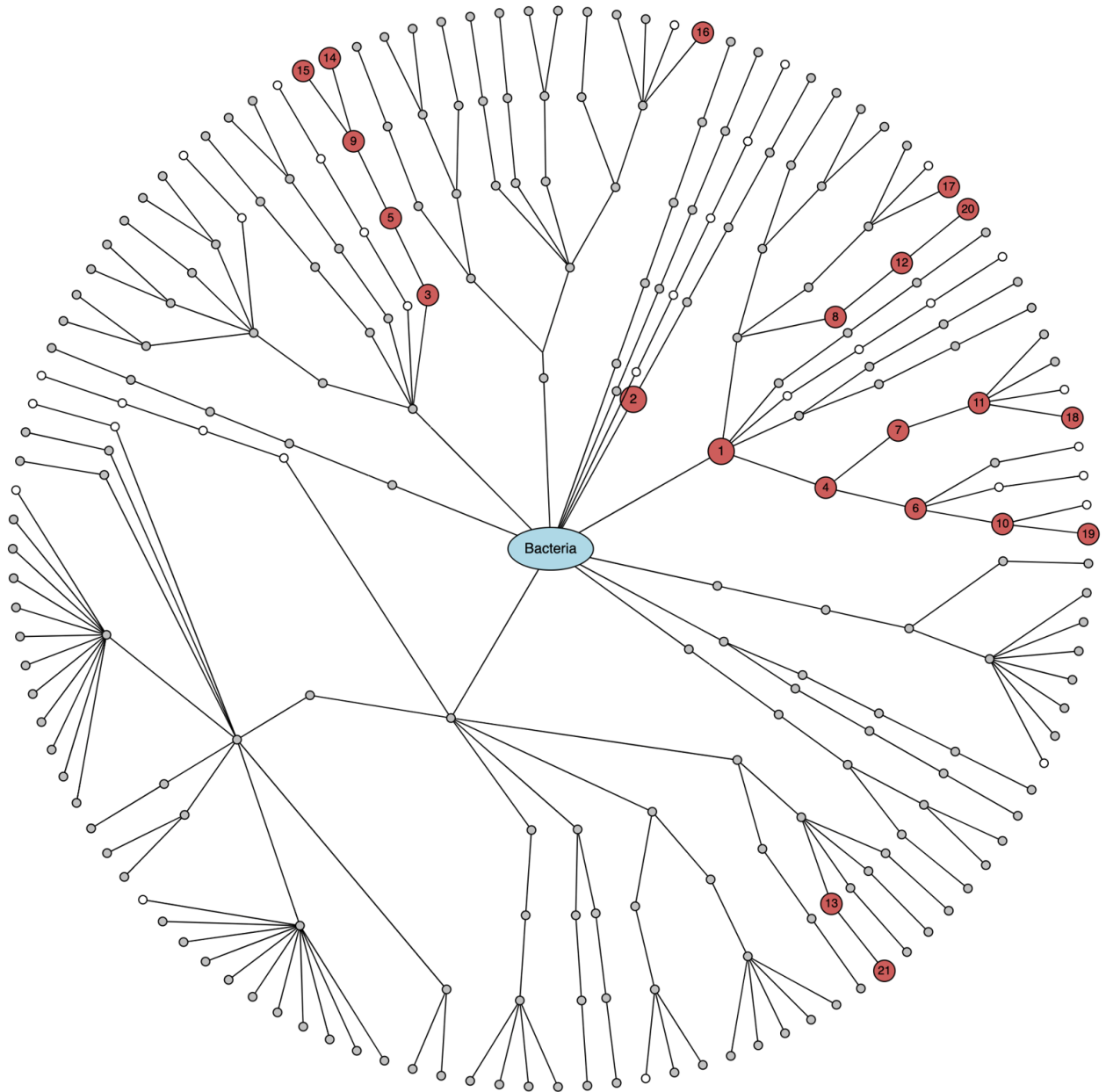
**Fig. 4. The results for beta-diversity.** We surveyed if e-cigarette smoking alters beta-diversity of the oral microbiome in subgingival niches, and the altered beta-diversity, in turn, influences gingival inflammation, adjusting for age, sex and the frequency of brushing teeth.



**Fig. 5. The results for microbial taxa.** We surveyed if e-cigarette smoking alters the microbial taxa of the oral microbiome in subgingival niches, and the altered microbial taxa, in turn, influence gingival inflammation, adjusting for age, sex and the frequency of brushing teeth.



**Fig. 6. A hierarchical visualization for the taxonomic discoveries.** The numbers in circles are matched with the IDs in Fig. 5.



**Fig. 7. The screenshot of the Ask ChatGPT module.** We asked ChatGPT a question, “What is known about *Bergeyella* on e-cigarette and gingival inflammation?”. Then, ChatGPT answered the question.

**Ask ChatGPT**

What is known about (discovered taxon) on (treatment) and (outcome)?

Insert your private ChatGPT API key

sk-g0t8KwHIV1YubKsllMn1ET3BtkFJLZgABJxq4I0HPAKI2Wd

You can visit <https://platform.openai.com/account/api-keys> to download your private ChatGPT API Key.

Select a taxonomic rank

Genus

Select a discovered taxon

Bergeyella

Rename the selected taxon

Bergeyella

You can rename it using a human language, for instance, deleting possible codes, special symbols or numbers (e.g., from 'Erysipelotrichaceae\_[G-1]' to 'Erysipelotrichaceae')

Rename the treatment variable

e-cigarette

You can rename the treatment variable using a human language (e.g., from 'ecig\_status' to 'e-cigarette').

Rename the outcome variable

gingival inflammation

You can rename the outcome variable using a human language (e.g., from 'gingival\_inflammation' to 'gingival inflammation').

Ask!

---

What is known about *Bergeyella* on e-cigarette and gingival inflammation ?

*Bergeyella* is a gram-negative bacteria that is commonly found in the oral microbiome. It has been associated with various oral diseases including gingivitis and periodontitis. There is some preliminary evidence suggesting that e-cigarettes may contribute to gingival inflammation and changes in the oral microbiome, including an increase in the abundance of *Bergeyella*. A study published in Scientific Reports in 2019 reported that e-cigarette users had higher levels of gingival inflammation and an altered oral microbiota compared to non-smokers or smokers of traditional cigarettes. Specifically, the study found that e-cigarette users had increased levels of *Bergeyella*, as well as other potentially pathogenic bacteria, in their oral microbiota. While the relationship between *Bergeyella* and e-cigarettes is still being explored and more research is needed to fully understand the implications, these findings suggest that e-cigarette use may have negative effects on oral health.