

MiSurv: an Integrative Web Cloud Platform for User-Friendly Microbiome Data Analysis with Survival Responses

Won Gu,^a DHyunwook Koh,^a Hyojung Jang,^a Byungho Lee,^a Byungkon Kang^b

Microbiology Spectrum

AMERICAN SOCIETY FOR MICROBIOLOGY

^aDepartment of Applied Mathematics and Statistics, The State University of New York, Korea, Incheon, South Korea ^bDepartment of Computer Science, The State University of New York, Korea, Incheon, South Korea

Won Gu and Hyunwook Koh contributed equally to this work as first authors. Author order was determined alphabetically.

ABSTRACT Investigators have studied the treatment effects on human health or disease, the treatment effects on human microbiome, and the roles of the microbiome on human health or disease. Especially, in a clinical trial, investigators commonly trace disease status over a lengthy period to survey the sequential disease progression for different treatment groups (e.g., treatment versus placebo, new treatment versus old treatment). Hence, disease responses are often available in the form of survival (i.e., time-to-event) responses stratified by treatment groups. While the recent web cloud platforms have enabled user-friendly microbiome data processing and analytics, there is currently no web cloud platform to analyze microbiome data with survival responses. Therefore, we introduce here an integrative web cloud platform, called MiSurv, for comprehensive microbiome data analysis with survival responses.

IMPORTANCE MiSurv consists of a data processing module and its following four data analytic modules: (i) Module 1: Comparative survival analysis between treatment groups, (ii) Module 2: Comparative analysis in microbial composition between treatment groups, (iii) Module 3: Association testing between microbial composition and survival responses, (iv) Module 4: Prediction modeling using microbial taxa on survival responses. We demonstrate its use through an example trial on the effects of antibiotic use on the survival rate against type 1 diabetes (T1D) onset and gut microbiome composition, respectively, and the effects of the gut microbiome on the survival rate against T1D onset. MiSurv is freely available on our web server (http://misurv.micloud.kr) or can alternatively run on the user's local computer (https://github.com/wg99526/MiSurvGit).

KEYWORDS microbiome data analysis, web cloud platform, survival analysis, association analysis, prediction modeling

The human microbiome is the entirety of the microorganisms that reside in and on the human body. The advances in high-throughput sequencing have greatly lowered the cost for microbiome profiling and improved the accuracy of the microbiome data. Then, investigators have found that the dysbiosis of the human microbiome is linked to numerous human diseases, for example, through studies on the effects of a medical treatment on human microbiome (1–6), the roles of the microbiome on human health or disease (1, 2, 4, 7–14), and so forth.

There have also been many analytic pipelines and tools that are developed for microbiome data processing and downstream data analysis. Some example microbiome profiling pipelines are QIIME (15), QIIME2 (16), MG-RAST (17), Mothur (18), and Nephele (19) for 16S rRNA sequencing, and MEGAN (20) and MetaPhlAn (21) for shot-gun metagenomics. Some example downstream microbiome data analytic tools are aMiAD (22) for alpha-diversity analysis, PERMANOVA (23–25) and MiRKAT (26–28) for beta-diversity analysis, and metagenomeSeq (29) and ANCOM (30) for taxonomic

Editor Jonathan L. Jacobs, American Type Culture Collection

Copyright © 2023 Gu et al. This is an openaccess article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Hyunwook Koh, hyunwook.koh@stonybrook.edu.

The authors declare no conflict of interest.

Received 8 December 2022 Accepted 12 March 2023 Published 11 April 2023 differential abundance analysis. Among those, we especially pay attention to the web cloud platforms, such as Nephele (19), Qiita (31), PUMAA (32), MicrobiomeAnalyst (33), METAGENassist (34), EzBioCloud (35), and MiCloud (36). These web cloud platforms have had the advantages of (i) user-friendly operation, (ii) streamlined analytic pipeline, and (iii) cloud computing service over the past software packages based on command-line interfaces. Thereby, investigators can easily deal with highly complex microbiome data, with no professional programming skills, not spending too much time and effort for a long sequence of data processing and analytic tasks, and with no fancy computer to process a huge amount of information.

However, for downstream microbiome data analysis, there is currently no web cloud platform available to analyze microbiome data with survival (i.e., time-to-event) responses. Investigators have been conducting numerous clinical trials, for example, to probe for the treatment effects on human microbiome (2–6), the treatment effects on human health or disease (2, 4), and the roles of the microbiome on human health or disease (1, 2, 4, 7–14). It is also common in a clinical trial to trace disease status over a lengthy period to survey the sequential progress of disease or recovery for different treatment groups (e.g., treatment versus placebo, new treatment versus old treatment), not restricted to the disease status at a single specific time point (1, 2, 9, 12). Hence, disease responses are often available in the form of survival (i.e., time-to-event) responses that are stratified by different treatment groups.

We therefore introduce here an integrative web cloud platform, called MiSurv, for microbiome data analysis with survival responses. MiSurv is a user-friendly step-bystep analytic pipeline like our previous web cloud platform, MiCloud (36). However, MiSurv is designed to deal with survival responses while MiCloud (36) is for continuous or binary responses. MiSurv performs microbiome data processing, analytic and graphical procedures comprehensively. MiSurv consists of a data processing module and its following four data analytic modules. Module 1: Comparative survival analysis between treatment groups. Module 2: Comparative analysis in microbial composition between treatment groups. Module 3: Association testing between microbial composition and survival responses. Module 4: Prediction modeling using microbial taxa on survival responses (Fig. 1). MiSurv is freely available on the web server (http://misurv.micloud .kr) or can alternatively run on the user's local computer (https://github.com/wg99526/MiSurvGit).

In the following Results section, we delineate all the details for each module, while demonstrating them step-by-step through an example trial on the effects of antibiotic use on the survival rate against type 1 diabetes (T1D) onset (i.e., time-to-T1D onset) and gut microbiome composition, respectively, as well as the roles of the gut microbiome on the survival rate against T1D onset (4) (see Real data application). In the Discussion section, we summarize the results and finish with concluding remarks. Finally, in Materials and Methods, we describe the underlying machinery and web architecture of MiSurv.

RESULTS

Data processing: data input. As in MiCloud (36), users need to upload four necessary data components: a feature table (i.e., count table for operational taxonomic units [OTUs] or amplicon sequence variants [ASVs]), a taxonomic table (i.e., taxonomic assignments at seven taxonomic levels, kingdom, phylum, class, order, family, genus, species), metadata (i.e., survival time, censored/event, demographics, etc. for the study subjects) and a phylogenetic tree (i.e., a rooted phylogenetic tree relating features [OTUs or ASVs]) (Fig. 1). Users can upload them as separate files or in a single unified format, called phyloseq (37) (Fig. 1).

(i) Real data application. We used the gut microbiome data for the study of antibiotic-induced acceleration of T1D (4). T1D is an autoimmune disease which is increasing in incidence globally with decreasing age of onset. To survey if the antibiotic use (i.e., a pulsed macrolide antibiotic [known as tylosin] treatment) accelerates T1D onset through gut microbiome perturbation, Zhang et al. conducted a randomized clinical trial using non-obese diabetic mice. Here, we use these data to demonstrate the use of



FIG 1 Overall workflow. MiSurv consists of a data processing module and its following four data analytic modules.

MiSurv step-by-step. We described the raw sequence data repository and processing procedures in Materials and Methods. The final processed data (4) are also stored as example data in the phyloseq (37) format and also as four individual data files on these data processing module.

At this module, we uploaded these data clicking the "Upload" button.

Data processing: quality control. MiSurv performs quality controls for the survival data in follow-up period (Fig. 1). For this, users first need to select a variable for survival time (i.e., follow-up time) and a variable for censored or event (i.e., a binary indicator for censored [0] or event [1]). Then, users can adjust the follow-up period (start time and end time) of interest using a slide bar (or leave it as it is to survey the entire follow-up period just as given in the data), where the censored/event indicator is also automatically adjusted to be suited to the selected follow-up period. The rationale behind this quality control user option is that investigators often (but not always) set up a very long period of follow-up at a design stage because they do not exactly know how long they need to follow up in advance. A very long period of follow-up also tends to result in no disparity in survival rate between treatment groups because treatment effects tend to disappear (and/or converge to the same level) at a very late time in the end.

As in MiCloud (36), MiSurv also performs quality controls for the microbiome data in four criteria: kingdom (i.e., a kingdom of interest, default: Bacteria); library size (i.e., a minimum total read count for the study subjects to have, default: 3,000); mean proportion (i.e., a minimum mean proportion for the features [OTUs or ASVs] to have, default: 0.002%); taxonomic names (i.e., erroneous taxonomic names to be removed in the taxonomic table) (Fig. 1). MiSurv visualizes the microbiome data before and after quality controls using summary boxes, histograms and box plots. The microbiome data before and after quality controls can also be downloaded.

(i) Real data application. We selected "T1Dweek" as the variable for the follow-up time, and "T1D" as the binary indicator for censored (0) or event (1), where the censored (0) represents T1D free and the event (1) represents T1D onset. We did not change the follow-up period as we wanted to survey the entire follow-up period just as given in the data. For the microbiome data, we applied the default quality control settings; as such, 294 features, 7 phyla, 14 classes, 18 orders, 21 families, 30 genera and 13 species for 172 subjects were finally retained (Fig. 2A).

Data processing: data transformation. As in MiCloud (36), for the ecological analysis, MiSurv calculates (1) nine alpha-diversity indices: Observed, Shannon (38), Simpson (39), Inverse Simpson (39), Fisher (40), Chao1 (41), abundance-based coverage estimator (ACE) (42), incidence-based coverage estimator (ICE) (43), and phylogenetic diversity (PD) (44); and (2) five beta-diversity indices: Jaccard dissimilarity (45), Bray-Curtis dissimilarity (46), Unweighted UniFrac distance (47), Generalized UniFrac distance (48), and Weighted UniFrac distance (49) (Fig. 1). For the taxonomical analysis, MiSurv transforms the original count data into four data forms: (i) count (rarefied) (50), (ii) proportion, (iii) centered log ratio (CLR) (51), and (iv) arcsine-root for each taxonomic rank (phylum, class, order, family, genus, species). While there is no consensus on which data format works best, the most popular microbiome data transformation method is the CLR transformation to deal with the compositionality of the data (51). Hence, we set CLR as the default for all following taxonomical analysis modules. All the calculated alpha-and beta-diversity indices as well as all the original and transformed taxonomic abundance data can also be downloaded.

(i) **Real data application.** We simply clicked the "Run" button to calculate the alpha- and beta-diversity indices, and transform the original count data into the four different data forms.

Module 1: comparative survival analysis between treatment groups (not involving microbiome data). Module 1 is for some basic exploratory data analysis, in which the microbiome data are not involved, to compare the survival rates between two treatment groups (e.g., treatment versus placebo, or new treatment versus old treatment), coded as 0 for the reference group and 1 for the comparison group (Fig. 1).





FIG 2 The screenshot after the quality controls using Data Processing (A) and the screenshot after performing comparative survival analysis between the control and tylosin groups adjusting for sex using Module 1 (B). *HR represents the estimated hazard ratio (i.e., the hazard rate for the tylosin group divided by the hazard rate for the control group).

For univariate analysis, the Kaplan-Meier analysis (default) (52) with the log-rank test (default) (53) or the Wilcoxon test (54), or the Cox model (55) can be employed. For covariate-adjusted analysis, only the Cox model (55) can be employed. The Kaplan-Meier analysis (52) and the Cox model (55) are the two most commonly used survival analysis methods. MiSurv visualizes the results using the Kaplan-Meier or covariate-adjusted survival curves. MiSurv also visualizes the sample sizes for different treatment groups and the sample sizes for the censored and events using summary boxes.

(i) **Real data application.** We selected "Antibiotics" as the treatment variable, where "Control" represents the group of normal controls and "Tylosin" represents the group of the subjects treated by a macrolide antibiotic, called tylosin (4). We selected "Sex" as a covariate, and then employed the Cox model (55) to perform covariate-adjusted analysis. We found a significant difference in survival rate between the control and tylosin groups (*P*-value: 0.005), while estimating that the hazard rate is greater for the tylosin group than the control group (hazard ratio [HR]: 1.692 > 1) (Fig. 2B). This

MiSurv

indicates that the subjects treated by tylosin progresses to T1D onset more rapidly than the normal controls. The sample sizes for the control and tylosin groups were 88 and 84, respectively, while the sample sizes for the censored and events were 56 and 116, respectively (Fig. 2B).

Module 2: comparative analysis in microbial composition between treatment groups (not involving survival data). Module 2 is for other basic exploratory data analysis, in which the survival data are not involved, to compare the microbial composition between two treatment groups (e.g., treatment versus placebo, or new treatment versus old treatment), coded as 0 for the reference group and 1 for the comparison group (Fig. 1). MiSurv performs such comparative analysis with respect to alpha-diversity, beta-diversity and taxonomic abundance as in MiCloud (36). For the alpha-diversity analysis, the Welch ttest or the Wilcoxon rank-sum test (default) (56) with box plots can be employed for univariate analysis, while the linear regression model with forest plots can be employed for covariate-adjusted analysis. For the beta-diversity analysis, MiRKAT (26, 27) can be employed for both univariate and covariate-adjusted analyses, yet only the principal coordinate analysis (PCoA) plots (57) are available to visualize the disparity between treatment groups in a univariate trend not reflecting covariate adjustments. For the taxonomic differential abundance analysis, four different data forms of the CLR (default) (51), count, proportion and arcsineroot are available. For the CLR (default) and arcsine-root, the Welch t test or the Wilcoxon rank-sum test (default) (56) with box plots can be employed for univariate analysis, while the linear regression model with forest plots can be employed for covariate-adjusted analysis. For the count, the Welch t test or the Wilcoxon rank-sum test (default) (56) using rarefied counts (50) with box plots can be employed for univariate analysis, while the negative binomial regression model using original counts, including the library size as an off-set variable with forest plots can be employed for covariate-adjusted analysis. For the proportion, the Welch t test or the Wilcoxon rank-sum test (default) (56) with box plots can be employed for univariate analysis, while the beta regression model with forest plots can be employed for covariate-adjusted analysis. Users can survey from phylum to genus (default) (e.g., for 16S rRNA sequencing) or from phylum to species (e.g., for shotgun metagenomics). For any taxonomic differential abundance analysis, MiSurv applies the Benjamini-Hochberg (BH) procedures (58) to each taxonomic rank to control for the false discovery rate (FDR) at 5%. MiSurv also visualizes the hierarchical taxonomic discovery status using a dendrogram.

(i) Real data application. We selected "Antibiotics" as the treatment variable, where "Control" represents the group of normal controls and "Tylosin" represents the group of the subjects treated by a macrolide antibiotic, called tylosin (4). We selected "Sex" as a covariate, and then performed covariate-adjusted analyses using the linear regression model for the alpha-diversity analysis (Fig. 3A), using MiRKAT (26, 27) for the beta-diversity analysis (Fig. 3B), and using the linear regression model with the CLR (default) (51) transformed data for the taxonomic differential abundance analysis (Fig. 4 and 5). Then, we found a significant difference in alpha-diversity between the control and tylosin groups for all the nine alpha-diversity indices (Fig. 3A), estimating that the alpha-diversity is smaller for the tylosin group than the control group for all nine alpha-diversity indices (Fig. 3A). We also found a significant difference in beta-diversity between the control and tylosin groups for all tylosin groups for all five beta-diversity indices (Fig. 3B). Finally, we discovered 66 taxa (from phylum to genus) that have significant disparity in relative abundance between the control and tylosin groups (Fig. 4 and 5).

Module 3: association testing between microbial composition and survival responses. Module 3 involves both microbiome data and survival data and performs association testing between microbial composition and survival responses with respect to alpha-diversity, beta-diversity and taxonomic abundance. Each submodule for the alpha-diversity, beta-diversity or taxonomic abundance begins with subsetting the data for subgroup analysis. That is, users can select a subgroup of interest using a category of a nominal variable (usually the treatment variable) to perform subgroup analysis. If they select a subgroup, only the subjects in that subgroup are retained in the following analysis. Otherwise, all the subjects just as given in the data are retained. The rationale behind this subgroup analysis user option is that the effects of microbiome on survival responses are

Microbiology Spectrum



FIG 3 The results from alpha-diversity (A) and beta-diversity (B) analyses to compare the microbial compositions between the control and tylosin groups adjusting for sex using Module 2, and the results from alpha-diversity (C) and beta-diversity (D) analyses to test the association with survival responses for the control group adjusting for sex using Module 3.

usually different between treatment groups, indicating an interaction between microbiome and treatment.

For the alpha-diversity analysis, the Cox model (55) with forest plots can be employed for both univariate and covariate-adjusted analyses. For the beta-diversity analysis, MiRKAT-S

| ID | Rank | Taxon | Est. | P-value | Q-value | | |
|----|--------|------------------------|----------------|---------|---------|------------------|--------------|
| 1 | Phylum | Verrucomicrobia | 3.062 | 0.000 | 0.000 | | |
| 2 | Phylum | Proteobacteria | 3.916 | 0.000 | 0.000 | | |
| 3 | Phylum | Cyanobacteria | -0.808 | 0.024 | 0.024 | | |
| 4 | Phylum | Actinobacteria | -2.046 | 0.000 | 0.000 | | |
| 5 | Phylum | Firmicutes | 0.293 | 0.007 | 0.008 | - | • |
| 6 | Phylum | Tenericutes | -3.265 | 0.000 | 0.000 | | |
| 7 | Phylum | Bacteroidetes | -1.153 | 0.000 | 0.000 | | |
| 8 | Class | Verrucomicrobiae | 2.780 | 0.000 | 0.000 | | |
| 9 | Class | Gammaproteobacteria | 4.485 | 0.000 | 0.000 | | |
| 10 | Class | Betaproteobacteria | 1.089 | 0.000 | 0.000 | | |
| 11 | Class | 4C0d-2 | - 0.941 | 0.006 | 0.011 | | |
| 12 | Class | Actinobacteria | -1.320 | 0.000 | 0.000 | | |
| 13 | Class | Alphaproteobacteria | 0.589 | 0.034 | 0.047 | | - |
| 14 | Class | Bacilli | -0.477 | 0.023 | 0.035 | | |
| 15 | Class | Mollicutes | -2.662 | 0.000 | 0.000 | | |
| 16 | Class | Erysipelotrichi | -2.614 | 0.000 | 0.000 | | |
| 17 | Class | Bacteroidia | -1.435 | 0.000 | 0.000 | | |
| 18 | Order | Verrucomicrobiales | 2.772 | 0.000 | 0.000 | | |
| 19 | Order | Enterobacteriales | 4.477 | 0.000 | 0.000 | | |
| 20 | Order | Burkholderiales | 1.137 | 0.000 | 0.000 | | |
| 21 | Order | YS2 | -0.949 | 0.007 | 0.013 | | |
| 22 | Order | Bifidobacteriales | -1.090 | 0.000 | 0.000 | | |
| 23 | Order | Lactobacillales | -0.464 | 0.026 | 0.042 | | |
| 24 | Order | Anaeroplasmatales | -2.353 | 0.000 | 0.000 | | |
| 25 | Order | Erysipelotrichales | -2.484 | 0.000 | 0.000 | | |
| 26 | Order | Rickettsiales | 1.253 | 0.000 | 0.000 | | |
| 27 | Order | RF39 | -1.349 | 0.000 | 0.000 | | |
| 28 | Order | Bacteroidales | -1.443 | 0.000 | 0.000 | | |
| 29 | Family | Verrucomicrobiaceae | 2.742 | 0.000 | 0.000 | | |
| 30 | Family | Enterobacteriaceae | 4.447 | 0.000 | 0.000 | | |
| 31 | Family | Alcaligenaceae | 1.113 | 0.000 | 0.000 | | |
| 32 | Family | Bifidobacteriaceae | -1.155 | 0.000 | 0.000 | | |
| 33 | Family | Lactobacillaceae | -1.503 | 0.000 | 0.000 | | |
| 34 | Family | Enterococcaceae | 2.707 | 0.000 | 0.000 | | |
| 35 | Family | Anaeroplasmataceae | -2.367 | 0.000 | 0.000 | | |
| 36 | Family | Lacnnospiraceae | 0.996 | 0.000 | 0.000 | | + |
| 37 | Family | Ruminococcaceae | -3.887 | 0.000 | 0.000 | | |
| 38 | Family | Erysipelotrichaceae | -2.539 | 0.000 | 0.000 | | |
| 39 | Family | mitochondria | 1.241 | 0.000 | 0.000 | _ | |
| 40 | Family | Clostridiaceae | -1.090 | 0.000 | 0.000 | | |
| 41 | Family | S24-7 | -1.705 | 0.000 | 0.000 | | |
| 42 | Family | Akkormonoio | 1.103 | 0.000 | 0.000 | | - |
| 43 | Conus | Dontooo | 2.000 | 0.000 | 0.000 | | |
| 44 | Conus | Eschorichia | 0.035 | 0.000 | 0.000 | | <u> </u> |
| 40 | Conus | Trabulsiolla | 0.304 | 0.000 | 0.000 | | |
| 40 | Conus | Klobeiello | 3 459 | 0.000 | 0.000 | | |
| 48 | Genus | Sutterella | 0.923 | 0.000 | 0.000 | | |
| 19 | Genus | Bifidobacterium | 1 3/18 | 0.000 | 0.001 | | |
| 50 | Genus | Lactobacillus | -1.687 | 0.000 | 0.000 | | |
| 51 | Genus | Enterococcus | 2 309 | 0.000 | 0.000 | | |
| 52 | Genus | Bilophila | 1 170 | 0.000 | 0.000 | | + |
| 53 | Genus | Anaeroplasma | -2.570 | 0.000 | 0.000 | | |
| 54 | Genus | [Ruminococcus] | -1.499 | 0.000 | 0.000 | | |
| 55 | Genus | Blautia | 4.178 | 0.000 | 0.000 | | |
| 56 | Genus | Oscillospira | -3.302 | 0.000 | 0.000 | | |
| 57 | Genus | Ruminococcus | -3.515 | 0.000 | 0.000 | | |
| 58 | Genus | Clostridium * | -0.960 | 0.000 | 0.000 | | |
| 59 | Genus | Coprococcus | 0.835 | 0.000 | 0.000 | | |
| 60 | Genus | Dorea | -1.731 | 0.000 | 0.000 | - | |
| 61 | Genus | Clostridium * * | 1.920 | 0.000 | 0.000 | | |
| 62 | Genus | Allobaculum | -1.308 | 0.000 | 0.000 | | |
| 63 | Genus | Coprobacillus | -1.626 | 0.000 | 0.000 | | |
| 64 | Genus | Desulfovibrio | -1.289 | 0.000 | 0.000 | | |
| 65 | Genus | Candidatus Arthromitus | -1.288 | 0.000 | 0.000 | | |
| 66 | Genus | Prevotella | 0.919 | 0.000 | 0.000 | | |
| | | | | | | | |
| | | | | | | -5 -4 -3 -2 -1 0 | 1 2 3 4 5 |
| | | | | | | 95% Confide | nce Interval |

FIG 4 The forest plot for taxonomic differential abundance analyses adjusting for sex using Module 2.



FIG 5 The dendrogram for taxonomic differential abundance analyses adjusting for sex using Module 2. The numbers in circles are the taxon IDs in the forest plot (Fig. 4). Red: positive association. Blue: negative association. Gray: non-significance. White: no association testing for non-taxonomic annotation.

(27, 59, 60) with PCoA plots (57) can be employed for both univariate and covariate-adjusted analyses. Here, again, the PCoA (57) plots do not reflect covariate adjustments and display the disparity in microbiome composition between the censored and events only at the end of the follow-up. For the taxonomic association analysis, four different data forms of the CLR (default) (51), count (rarefied) (50), proportion and arcsine-root are available. Then, the Cox model (55) with forest plots can be employed for both univariate and covariate-adjusted analyses. As in Module 2, users can survey from phylum to genus (default) (e.g., for 16S rRNA sequencing) or from phylum to species (e.g., for shotgun metagenomics). MiSurv applies the BH procedures (58) to each taxonomic rank to control for the FDR at 5%. MiSurv also visualizes the hierarchical taxonomic discovery status using a dendrogram.

(i) Real data application. We performed subgroup analysis while selecting the control group in the treatment variable as the subgroup of interest; as such, only the normal controls (excluding tylosin subjects) were retained in the following analyses. We selected "Sex" as a covariate to perform covariate-adjusted analyses. For the alpha-diversity analysis, we did not find any significant association between alpha-diversity and survival responses for any alpha-diversity index (Fig. 3C). For the beta-diversity analysis, we found a significant association between beta-diversity and survival responses for the Jaccard dissimilarity (P value: 0.046) (Fig. 3D). Finally, we discovered 5 phyla (Verrucomicrobia, Proteobacteria, Actinobacteria, Firmicutes, Tenericutes) that have significant association with the survival responses (Fig. 6). We estimated that the hazard rate is greater with increasing CLR transformed relative abundance for Verrucomicrobia (HR: 1.253 > 1), Proteobacteria (HR: 1.236 > 1) and Firmicutes (HR: 1.262 > 1), indicating the subjects that have more Verrucomicrobia, Proteobacteria and Firmicutes in their gut progresses to T1D onset more rapidly. On the contrary, we estimated that the hazard rate is smaller with increasing CLR transformed relative abundance for Actinobacteria (HR: 0.759 < 1) and Tenericutes (HR: 0.796 < 1), indicating the subjects that have less Actinobacteria and Tenericutes in their gut progress to T1D onset more slowly (Fig. 6).

Module 4: prediction modeling using microbial taxa on survival responses. Module 4 involves both microbiome data and survival data, and performs prediction modeling using microbial taxa at the taxonomic ranks of phylum, class, order, family, genus and species, respectively, on survival responses. As in Module 3, users can select a subgroup of interest using a category of a nominal variable (usually the treatment variable) to perform subgroup analysis. As in Module 3, users first need to select a data form among CLR (default) (51), count (rarefied) (50), proportion and arcsine-root. Then, they can survey from phylum to genus (default) (e.g., for 16S rRNA sequencing) or from phylum to species (e.g., for shotgun metagenomics). For prediction modeling, the random survival forests (61, 62) can be employed. The random survival forests (61, 62) is an ensemble learning method that averages the predictions on survival rates resulting from a multitude of decision trees constructed by randomly selected inputs. Hence, the random survival forests can robustly suit various linear or non-linear patterns while decorrelating taxa and adapting to various sparsity levels. Users can select the number of trees (5,000 [default] or 10,000) to be ensembled. Users can also select the number of taxa at each taxonomic rank to be displayed on the variable importance plot (61, 62) that ranks taxa in the amount of increase in the residual sum of squares. MiSurv also visualizes the out-of-bag error rates by the number of trees to be ensembled.

(i) Real data application. We performed subgroup analysis while selecting the control group in the treatment variable as the subgroup of interest; as such, only the normal controls (excluding tylosin subjects) were retained in the following analysis. We selected 10,000 as the number of trees to be ensembled and 10 as the number of taxa at each taxonomic rank to be displayed on the variable importance plot. We listed top 10 genera that have the most influential predictions on the survival rates in (Fig. 7A). We also observed that the out-of-bag error rates well converge by 10,000 of trees ensembled in (Fig. 7B).

DISCUSSION

In this paper, we introduced an integrative web cloud platform, MiSurv, for user-friendly step-by-step microbiome data analysis with survival responses. MiSurv performs extended data processing, analytic and graphical procedures to efficiently deal with microbiome data with survival responses. We described in detail a data processing module and its following four data analytic modules of MiSurv along with the example gut microbiome trial that involves a treatment (antibiotics) variable and survival responses on T1D onset.

In the Data Processing module, we added the quality control to adjust the followup period because investigators often set up a very long period of follow-up at the stage of experimental design just to survey the disease progression sufficiently. Though, it is usually not the one intended in a later data analysis. Module 1 and



FIG 6 The forest plot (A) and dendrogram (B) for taxonomic association analyses between each taxon and survival responses adjusting for sex using Module 3. The numbers in circles (Fig. 6B) are the taxon IDs in the forest plot (Fig. 6A). Red: positive association. Blue: negative association. Gray: non-significance. White: no association testing for non-taxonomic annotation. *HR represents the estimated hazard ratio (i.e., the relative increase in hazard rate with one unit increase in the CLR transformed relative abundance).

Module 2 perform some basic exploratory data analysis, not involving microbiome and survival data, respectively, for comparative survival analysis between treatment groups and comparative analysis in microbial composition between treatment groups, respectively. On the contrary, Module 3 and 4 link the microbial composition to survival responses with respect to association testing and prediction modeling, respectively. At the beginning of Module 3 and 4, we added the subgroup analysis user option because it is often assumed that the effects of microbiome on survival responses are

MiSurv



FIG 7 The importance plot (A) and error rates (B) for prediction modeling using genera on survival responses using Module 4.

different between treatment groups, which indicates an interaction between microbiome and treatment, which in turn indicates the need for subgroup analysis. We believe that not all the facilities in MiSurv are just ours. Indeed, they are mostly based on other researchers' methodologies; hence, we devoted to adding all related references to each of the data processing and analytic steps on our web platform.

In summary, MiSurv is easy to use and provides comprehensive facilities for microbiome data processing and downstream data analysis with survival responses. MiSurv can serve as a useful tool in the rapidly growing field of human microbiome for numerous trials to study the treatment effects on survival responses, the treatment effects on human microbiome, and the roles of the microbiome on survival responses.

MATERIALS AND METHODS

Overall architecture. We made all the user interfaces and server functions using R Shiny (https:// shiny.rstudio.com). The computer server currently has Intel Core i7-12700T processor (12 cores,1.40 to 4.70 GHz) and 36 GB DDR4 memory and runs on the Ubuntu 20.04 server. We will improve the specification if it gets more popular and busier. A web server, in theory, is any software that can interact with another entity over a network. The overall architecture of our web server follows a typical client-server model, deploying it through two layers, (1) ShinyProxy (https://www.shinyproxy.io) to maintain our service stable by means of the docker and container, and (2) Apache2 (https://httpd.apache.org) to provide easier access to our service by means of the domain name. Our web server currently supports up to 10 simultaneous connections. We also host our platform through GitHub (https://github.com) repository for the users who wish to run it on their local computer.

Underlying machinery. We wrote all the data processing, analytic and graphical procedures in R language [Key Resource Table]. MiSurv calculates the alpha-diversity indices (38–44) using "estimate_r-ichness" in the "phyloseq" package and "pd" function in the "picante" package. MiSurv calculates the beta-diversity indices (45–49) using the "dist" function in the "proxy" package, the "bcdist" function in the "ecodist" package, and the "GUniFrac" in the "GUniFrac" package. MiSurv conducts the CLR (51) transformation using the "clr" function in the "compositions" package and the rarefaction (50) using the "rarefy_even_depth" in the "phyloseq" package. MiSurv performs the Kaplan-Meier analysis (52) with the log-rank test (53) and Wilcoxon test (54) using the "Surv" and "survdiff" functions in the "survival"

package. MiSurv fits the Cox model (55) using the "coxph" function in the "survival" package. MiSurv performs MiRKAT (26, 27) and MiRKAT-S (27, 59, 60) with PCoA plots (57) using the "D2K," "MiRKAT," and "MiRKATS" functions in the "MiRKAT" package and the "pcoa" function in the "ape" package. MiSurv performs the Welch *t* test using the "t.test" function in the "stats" package and the Wilcoxon rank-sum test (56) using the "wilcox.test" function in the "stats" packages. MiSurv fits the linear regression model using the "lm" and "glm" function in the "stats" package, the negative binomial regression model using the "glm.nb" function in the "MASS" package, and the beta regression model using the "betareg" function in the "betareg" package. MiSurv fits the random survival forests (61, 62) using the "rfsrc" function in the "randomForestSRC" package.

Raw sequence data processing procedures. Zhang et al. collected fecal samples from non-obese diabetic mice from Jackson Laboratory (4). Then, they extracted intestinal microbiota DNA using the PowerLyzer PowerSoil DNA isolation kit (MoBio, Carlsbad, CA) and the PowerSoil-htp 96-Well Soil DNA isolation kit (MoBio). Then, they acquired the amplicon library of V4 regions of the 16S rRNA genes through triplicate PCR with barcoded fusion primers, quantification with the Qubbit 2.0 Fluorometer (Life Technologies, Carlsbad, CA). and combination of each DNA sample at equal concentrations as in (2). Then, they sequenced the library using Illumina MiSeq 2×150 bp paired end platform (Illumina, San Diego, CA).

We used QIIME 2.0 [16] to obtain the feature table, taxonomic annotations, and phylogenetic tree. We retained only the reads with >75% of the original length while removing the reads with more than three consecutive low-quality bases (Phred score < 20). We constructed (1) the feature table using the 97% sequence similarity through open reference piking based on the Greengenes database (63), (2) the taxonomic annotations using the RDP classifier while removing chimeras using ChimeraSlayer (64), and (3) the phylogenetic tree using FastTree (65).

Data and code availability. We used public microbiome data for our real data applications (4). The raw sequence data are deposited in QIITA (https://qiita.ucsd.edu) with the ID number 10508 (https://qiita.ucsd.edu/study/description/10508). All the final processed data are also stored in the Data Processing: Data Input module on the web server (http://misurv.micloud.kr) and also in the Data folder on the GitHub page (https://github.com/wg99526/MiSurvGit) with the for the "phyloseq" (37) object (see the file named "biom.Rdata") and also with four individual files (see the files named "otu.tab.txt," "sam.dat.txt," "tax.tab.txt," and "tree.tre").

MiSurv can be run through our web server (http://misurv.micloud.kr) or on the user's local computer through the GitHub repository (https://github.com/wg99526/MiSurvGit). Either of these platforms comes with a detailed software manual on implementation, example data (4), required or optional inputs or settings, numeric or graphical outputs, related references, and so forth.

ACKNOWLEDGMENTS

We declare that we have no competing interests. H.K. was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2021R1C1C1013861). B.K. was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2022-00165660).

H.K. developed the overall architecture and conceptualized methodological ideas. W.G. and H.K. developed the data processing and analytic platforms and wrote the programs and manuscript. H.J. and B.K. developed the overall architecture of the web server and client model and wrote the programs. B.L. contributed to visualizations and wrote the programs. All authors reviewed and approved the manuscript.

REFERENCES

- Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, Moore BB, White ES, Flaherty KR, Huffnagle GB, Martinez FJ, COMET Investigators. 2014. Association between lung microbiome and disease progression in IPF; a prospective cohort study. Lancet Respir Med 2:548–556. https://doi.org/10 .1016/S2213-2600(14)70069-4.
- Livanos AE, Greiner TU, Vangay P, Pathmasiri W, Stewart D, McRitchie S, et al. 2016. Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. Nat Microbiol 1:1–13. https://doi .org/10.1038/nmicrobiol.2016.140.
- Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, Abrouk M, Farahnik B, Nakamura M, Zhu TH, Bhutani T, Liao W. 2017. Influence of diet on the gut microbiome and implications for human health. J Transl Med 15:1–17. https://doi.org/10.1186/s12967-017-1175-y.
- Zhang XS, Li J, Krautkramer KA, Badri M, Battaglia T, Borbet TC, et al. 2018. Antibiotic-induced acceleration of type 1 diabetes alters maturation of innate intestinal immunity. Elife 7:e37816. https://doi.org/10.7554/eLife .37816.
- Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. 2020. Impact of commonly used drugs on the composition and metabolic function

of the gut microbiota. Nat Commun 11:1-11. https://doi.org/10.1038/s41467 -019-14177-z.

- Gui X, Yang Z, Li MD. 2021. Effect of cigarette smoke on gut microbiota: state of knowledge. Front Physiol 12:1–14. https://doi.org/10.3389/fphys .2021.673341.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444:1027–1031. https://doi.org/10.1038/nature05414.
- Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. 2014. Human genetics shape the gut microbiome. Cell 159:789–799. https://doi.org/ 10.1016/j.cell.2014.09.053.
- Frankel AE, Coughlin LA, Kim J, Froehlich TW, Xie Y, Frenkel EP, Koh AY. 2017. Metagenomic shotgun sequencing and unbiased metabolomic profiling identify specific human gut microbiota and metabolites associated with immune checkpoint therapy efficacy in melanoma patients. Neoplasia 19: 848–855. https://doi.org/10.1016/j.neo.2017.08.004.
- Liu M, Koh H, Kurtz ZD, Battaglia T, PeBenito A, Li H, Nazzal L, Blaser MJ. 2017. Oxalobacter formigenes-associated host features and microbial community structures examined using the American Gut Project. Microbiome 5:1–17. https://doi.org/10.1186/s40168-017-0316-0.

- Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, et al. 2018. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science 359:97–103. https://doi.org/10 .1126/science.aan4236.
- Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML, et al. 2018. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science 359:104–108. https://doi.org/10.1126/ science.aao3290.
- Sharma S, Tripathi P. 2019. Gut microbiome and type 2 diabetes: where we are and where to go? J Nutr Biochem 63:101–108. https://doi.org/10 .1016/j.jnutbio.2018.10.003.
- Glassner KL, Abraham BP, Quigley EM. 2020. The microbiome and inflammatory bowel disease. J Allergy Clin Immunol 145:16–27. https://doi.org/ 10.1016/j.jaci.2019.11.003.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/ nmeth.f.303.
- 16. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/ s41587-019-0209-9.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinform 9:1–8.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM .01541-09.
- Weber N, Liou D, Dommer J, MacMenamin P, Quiñones M, Misner I, Oler AJ, Wan J, Kim L, Coakley McCarthy M, Ezeji S, Noble K, Hurt DE. 2018. Nephele: a cloud platform for simplified, standardized and reproducible microbiome data analysis. Bioinformatics 34:1411–1413. https://doi.org/ 10.1093/bioinformatics/btx617.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. Genome Res 17:377–386. https://doi.org/10.1101/gr.5969107.
- Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 9:811–814. https://doi.org/10.1038/nmeth .2066.
- Koh H. 2018. An adaptive microbiome α-diversity-based association analysis method. Sci Rep 8:1–12. https://doi.org/10.1038/s41598-018-36355-7.
- Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol 26:32–46.
- McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8: e61217. https://doi.org/10.1371/journal.pone.0061217.
- Tang ZZ, Chen G, Alekseyenko AV. 2016. PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. Bioinformatics 32:2618–2625. https://doi.org/10 .1093/bioinformatics/btw311.
- Zhao N, Chen J, Carroll IM, Ringel-Kulka T, Epstein MP, Zhou H, Zhou JJ, Ringel Y, Li H, Wu MC. 2015. Testing in microbiome-profiling studies with MiRKAT, the microbiome regression-based kernel association test. Am J Hum Genet 96:797–807. https://doi.org/10.1016/j.ajhg.2015.04.003.
- Wilson N, Zhao N, Zhan X, Koh H, Fu W, Chen J, Li H, Wu MC, Plantinga AM. 2021. MiRKAT: kernel machine regression-based global association tests for the microbiome. Bioinformatics 37:1595–1597. https://doi.org/10 .1093/bioinformatics/btaa951.
- Koh H, Li Y, Zhan X, Chen J, Zhao N. 2019. A distance-based kernel association test based on the generalized linear mixed model for correlated

microbiome studies. Front Genet 10:1–14. https://doi.org/10.3389/fgene .2019.00458.

- Paulson JN, Stine OC, Bravo HC, Pop M. 2013. Differential abundance analysis for microbial marker-gene surveys. Nat Methods 10:1200–1202. https://doi.org/10.1038/nmeth.2658.
- Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 26:27663. https://doi .org/10.3402/mehd.v26.27663.
- Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. 2018. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods 15:796–798. https://doi.org/10.1038/s41592-018-0141-9.
- Mitchell K, Ronas J, Dao C, Freise AC, Mangul S, Shapiro C, et al. 2020. PUMAA: a platform for accessible microbiome analysis in the undergraduate classroom. Front Microbiol 11:1–15. https://doi.org/10.3389/fmicb.2020.584699.
- Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. 2017. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res 45:W180–W188. https://doi.org/10.1093/nar/gkx295.
- Arndt D, Xia J, Liu Y, Zhou Y, Guo AC, Cruz JA, Sinelnikov I, Budwill K, Nesbø CL, Wishart DS. 2012. METAGENassist: a comprehensive web server for comparative metagenomics. Nucleic Acids Res 40:W88–W95. https:// doi.org/10.1093/nar/gks497.
- 35. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.
- Gu W, Moon J, Chisina C, Kang B, Park T, Koh H. 2022. MiCloud: a unified web platform for comprehensive microbiome data analysis. PLoS One 17: e0272354. https://doi.org/10.1371/journal.pone.0272354.
- McArdle BH, Anderson MJ. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82:290–297. https://doi.org/10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2.
- Shannon CE. 1948. A mathematical theory of communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
- 39. Simpson EH. 1949. Measurement of diversity. Nature 163:688. https://doi .org/10.1038/163688a0.
- Fisher RA, Corbet AS, Williams CB. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. J Anim Ecol 12:42–58. https://doi.org/10.2307/1411.
- 41. Chao A. 1984. Non-parametric estimation of the number of classes in a population. Scand J Stat 11:265–270.
- 42. Chao A, Lee SM. 1992. Estimating the number of classes via sample coverage. J Am Stat Assoc 87:210–217. https://doi.org/10.1080/01621459.1992 .10475194.
- Lee SM, Chao A. 1994. Estimating population size via sample coverage for closed capture-recapture models. Biometrics 50:88–97. https://doi.org/10 .2307/2533199.
- 44. Faith DP. 1992. Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10. https://doi.org/10.1016/0006-3207(92)91201-3.
- 45. Jaccard P. 1912. The distribution of the flora in the alpine zone. New Phytol 11:37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.
- Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27:325–349. https://doi.org/10.2307/ 1942268.
- Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235. https://doi .org/10.1128/AEM.71.12.8228-8235.2005.
- Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman FD, Li H. 2012. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics 28:2106–2113. https://doi.org/10.1093/bioinformatics/bts342.
- Lozupone CA, Hamady M, Kelley ST, Knight R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 73:1576–1585. https:// doi.org/10.1128/AEM.01996-06.
- 50. Sanders HL. 1968. Marine benthic diversity: a comparative study. Am Nat 102:243–282. https://doi.org/10.1086/282541.
- Aitchison J. 1982. The statistical analysis of compositional data. J R Stat Soc Series B 44:139–160. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x.

- Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481. https://doi.org/10.1080/01621459.1958 .10501452.
- 53. Mantel N, Haenszel W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719–748.
- Peto R, Peto J. 1972. Asymptotically efficient rank invariant test procedures. J R Stat Soc Series A 135:185–207. https://doi.org/10.2307/2344317.
- 55. Cox DR. 1972. Regression models and life-tables. J R Stat Soc Series B 34: 187–220. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.
- Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60. https://doi.org/10.1214/aoms/1177730491.
- Torgerson WS. 1952. Multidimensional scaling: I. Theory and method. Psychometrika 17:401–419. https://doi.org/10.1007/BF02288916.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B 57: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
- Plantinga A, Zhan X, Zhao N, Chen J, Jenq RR, Wu MC. 2017. MiRKAT-S: a community-level test of association between the microbiota and survival times. Microbiome 5:1–13. https://doi.org/10.1186/s40168-017-0239-9.

- Koh H, Livanos AE, Blaser MJ, Li H. 2018. A highly adaptive microbiomebased association test for survival traits. BMC Genomics 19:1–13. https:// doi.org/10.1186/s12864-018-4599-8.
- Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. 2008. Random survival forests. Ann App Statist 2:841–860.
- 62. Mogensen UB, Ishwaran H, Gerds TA. 2012. Evaluating random forests for survival analysis using prediction error curves. J Stat Softw 50:1–23. https://doi.org/10.18637/jss.v050.i11.
- 63. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05.
- 64. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Petrosino JF, Knight R, Birren BW, Human Microbiome Consortium. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrose-quenced PCR amplicons. Genome Res 21:494–504. https://doi.org/10.1101/gr.112730.110.
- Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26: 1641–1650. https://doi.org/10.1093/molbev/msp077.