Targeted sequencing of 16S rRNA genes enables the analysis of microbiomes. Here, we describe a protocol for the collection, storage, and preparation of fecal samples. We describe how we cluster similar sequences and assign bacterial taxonomies. Using diversity analysis and machine learning, we can extract disease-associated features. We also describe a circadian analysis to identify the presence or absence of rhythms in taxonomies. Differences in rhythmicity between cohorts can contribute to determining disease-associated bacterial signatures.
Protocol
Comparing Circadian Rhythmicity in the Human Gut Microbiome

Sandra Reitmeier,1,2,3,4 Silke Kiessling,1,2,3 Klaus Neuhaus,1 and Dirk Haller1,2,5,*

1ZIEL - Institute for Food & Health, Technical University of Munich, 85354 Freising, Germany
2Chair of Nutrition and Immunology, Technical University of Munich, Gregor-Mendel-Str. 2, 85354 Freising, Germany
3These authors contributed equally
4Technical Contact
5Lead Contact
*Correspondence: dirk.haller@tum.de
https://doi.org/10.1016/j.xpro.2020.100148

SUMMARY
Targeted sequencing of 16S rRNA genes enables the analysis of microbiomes. Here, we describe a protocol for the collection, storage, and preparation of fecal samples. We describe how we cluster similar sequences and assign bacterial taxonomies. Using diversity analysis and machine learning, we can extract disease-associated features. We also describe a circadian analysis to identify the presence or absence of rhythms in taxonomies. Differences in rhythmicity between cohorts can contribute to determining disease-associated bacterial signatures.
For complete details on the use and execution of this protocol, please refer to Reitmeier et al. (2020).

BEFORE YOU BEGIN
The Study Centre informs participants about the aims of the study and provides a material box which includes everything necessary guaranteeing a clean and sanitized sample collection (e.g., gloves, tearproof stool collector).

The participants are asked to collect the sample at the appointment day and store it in the fridge (4°C) until then. The stool collector should be used to avoid any contamination.

The questionnaire comprises questions regarding fecal stool collection (date, time, problems etc.), about personal information including health status (age, medication, disease etc.) and about dietary habits.

Each participant gets a postal package including:

• One questionnaire
• An instruction manual
• 2 collection containers – one being empty and one containing DNA Stabilizer from Invitek (Stool DNA Stabilizer – Catalog No 1038111100). The collection tubes have unique QR codes each
• 1 pair of disposable gloves
• 2 stool collector (one being a replacement)
• Participants are asked to collect samples, if possible, on the day of the appointment at the Study Centre or earliest 1 day before.
Fecal Sample Collection

1. The study center prepares sample collection kits which are handed out to the study participants.
   a. The kit includes a sample collection instruction guiding the participants through the procedure.
   b. According to the instructions the participant is asked to use all provided disposals.
2. Samples should be stored in the household's fridge at 4°C as short as possible. For a storage >36 h, we recommend storing samples at −20°C.
3. Transport of the two 8-mL tubes (including the sample).
   a. Deliver samples to the center during visit (which is the preferred transport).
   b. Send sample as soon as possible by postal mailing (a prepaid and addressed envelope might be provided).

Samples collected in DNA Stabilizer are stable for at least 3 days at ambient temperature and at least 7 days at 4°C. It was shown that short- and long-term storage have an effect on microbial DNA stability (Carroll et al., 2012; Dominianni et al., 2014) with some bacteria tend to be more sensitive than others (Shaw et al., 2016). DNA stabilization liquid has advantages for preservation of the DNA and facilitates the process of sample collection and storage in studies (Ilett et al., 2019). In a small in-house study, we analyzed the influence of storage (at 20°C–22°C) and showed that samples including DNA stabilizer have increased stability over time (0 h, 24 h, and 48 h) compared to samples without (Figure 1).

Arrival at the Study Center

4. The QR code of the 8-mL tubes with the stabilizer liquid is scanned and the tubes are stored at −20°C.
**Note:** For a long-time storage (more than 3 months) it is recommended to store samples at 
\(-80^\circ\text{C}\) (Goodrich et al., 2014).

5. The QR code of the 5-mL tubes without the stabilizer liquid are scanned and the tubes are stored 
at \(-80^\circ\text{C}\).

⚠️ **CRITICAL:** It is important to have unique labels for each sample. We recommend a bar-code system which helps in proper sample and data handling. For human studies, an anonymization system with restricted access may be important as well. Information about storage, arrival time, and additional information should be noted in a database. Questionnaires need to be electronically recorded (e.g., scanned for future reference).

⚠️ **CRITICAL:** Variables, names and information (included in the database) should be formatted smartly in advance in order to avoid later re-formatting of, e.g., identifiers for subsequent analyses (e.g., statistics).

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Chemicals, Peptides, and Recombinant Proteins** | | |
| polyvinylpyrrolidone (PVPP) | Sigma | Cat# 77627 100 g |
| guanidinium thiocyanate | Sigma | Cat# G9277 500 g |
| N-lauroylsarcosine sodium | Sigma | Cat# LS125 100 g |
| Phusion Hot Start II High fidelity | Thermo Fisher | Cat# F-549L |
| HF Puffer Pack | Thermo Fisher | Cat# F-518L |
| dNTP Mix, 10 mM each, 2 × 0.5 mL | Biozym | Cat# 331520 |
| 100 bp DNA Ladder | NEB | Cat# N3231S |
| GelRed Nucleic Acid Gel Stain, 10,000 in water; 0.5 mL | VWR | Cat# 41003 |
| dNTPs | Sigma | Cat# D7295 20 × 0.2 mL |
| Agarose | Sigma | Cat# A9539 500 g |
| DMSO | Sigma | Cat# D2650 5 × 10 mL |
| Lysing Matrix B | MP Biomedicals | Cat# 116911500 |
| 16S rRNA gene Illumina sequencing primers (V3V4) | Kozich et al., 2013 | 341F-ovh and 785r-ovh |
| AMPure XP beads | Beckman | Cat# A63881 |
| PhiX Control v3 Library | Illumina | FC-110-3001 |
| RNase A | Thermo Fisher | Cat# EN0531 |
| **Critical Commercial Assays** | | |
| Nucleo Spin gDNA clean-up (250) | Machery-Nagel | Cat# 740230250 |
| Binding Buffer DB | Machery-Nagel | Cat# 740323.1 |
| Qubit 1 x dsDNAhs Kit 500 assays REF Q32854 (Life Technologies) | Fisher Scientific | Cat# 15860210 |
| MiSeq® Reagent Kit v3 (600 cycle) | Illumina Inc | Cat# MS-102-3003 |
| Mock community | ZymoBIOMICS | Cat# D6300 |
| **Software and Algorithms** | | |
| bcl2fastq | bcl2fastq | [https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html](https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) RRID:SCR_015058 |

(Continued on next page)
| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| GraphPad Prism v8.0.2 | Graphpad Software | https://www.graphpad.com/scientific-software/prism/ RRID:SCR_002798 |
| RStudio | RStudio | https://rstudio.com/products/rstudio |
| BLAST | Altschul et al. (1990) | https://blast.ncbi.nlm.nih.gov RRID:SCR_007190 |
| IMNGS | Lagkouvardos et al. (2016) | https://www.imngs.org/ |
| EvoView | He et al. (2016) | https://www.evolgenius.info/ |
| FASTQC | | http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ RRID:SCR_014583 |
| EzBioCloud | Yoon et al. (2017) | https://www.ezbiocloud.net/ |
| KEGG | Kanehsa and Goto (2000) | https://www.genome.jp/kegg/ RRID:SCR_001120 |
| Heatmapper | Babicki et al. (2016) | http://www.heatmapper.ca; |
| GraPhlAn | Segata et al. (2013) | https://github.com/biobakery/graphlan |
| Rhea | Lagkouvardos et al. (2017) | https://github.com/Lagkouvardos/Rhea |
| JTK_CYCLE | Hughes et al. (2010) | https://www.r-project.org/ |
| HUMAnN2 | Franzosa et al. (2018) | https://github.com/biobakery/humann |
| Psych R package | Revelle (2020) | https://cran.r-project.org/web/packages/psych/index.html |
| randomForest R package | Liaw and Wiener (2002) | https://cran.r-project.org/web/packages/randomForest/ randomForest.pdf RRID:SCR_015718 |
| metaphlan2 | Segata et al. (2012) | https://github.com/biobakery/metaphlan RRID:SCR_004915 |

**Oligonucleotides**

- 341F-ovh Primer: CCTACGGGNGGCWGCAG | Klindworth et al. (2013) | N/A |
- 785R-ovh Primer: GACTACHVGGGTATCTAATCC | Klindworth et al. (2013) | N/A |

**Biological Samples**

- Healthy adults (N = 8), stool samples (n = 24) for the analysis of storage effect | Technical University Munich, Chair of Nutrition and Immunology | Available upon request |

**Other**

- DNA-Stool-Stabilizer | INVITEK | Cat# 1038111100 |
- Stool Collection Tubes with Stabilizer | INVITEK | Cat# 1038111300 |
- Combitips advanced, 5 mL | diagonal | Cat# 30089812 |
- Combitips advanced, 25 mL | diagonal | Cat# 30089839 |
- Micro tube, 2.0 mL, SafeSeal | sarstedt | Cat# 72695400 |
- Micro tube, 1.5 mL, SafeSeal | sarstedt | Cat# 72706400 |
- Micro tube, 2.0 mL, PP | sarstedt | Cat# 72693005 |
- 96-Well Skirted PCR Plate | 4ti-tude | Cat# 4ti-0960 |
- PCR Foil Seal | 4ti-tude | Cat# 4ti-0550 |
- Microplate Seals for Aqueous Sample Storage | 4ti-tude | Cat# 4ti-0510 |
- Adhesive Seals for PCR Plates | 4ti-tude | Cat# 4ti-0500 |
- 1,000-µL tips with barrier | beckman | Cat# B01124 |
- 50-µL tips with barrier | beckman | Cat# A21586 |
- 250-µL tips with barrier | beckman | Cat# 717253 |
- 250-µL tips without barrier | beckman | Cat# 717252 |
- AMPure XP beads | beckman | Cat# A63881 |

(Continued on next page)
**STEP-BY-STEP METHOD DETAILS**

Sample processing is divided into four main steps: DNA isolation, library construction by PCR, amplicon cleaning and dilution, and sequencing (Figure 2).

### DNA Isolation

**Timing:** approx. 3 h for 24 samples

DNA is isolated with a modification of the protocol by Godon et al. (1997). A blank sample, consisting of 600 μL DNA Stabilizer from Invitrogen, is processed in every second DNA isolation batch (i.e., one blank sample for each 47 samples).

1. Thaw fecal samples (ca. 2 g in 8 mL DNA stabilizer) for approximately 2 h at 20°C–22°C.
2. Vortex until the sample is fully homogenized and let stand for 3 min to sediment debris.
3. For each sample, a volume of 600 μL fecal slurry is transferred into a 2-mL screw cap tube containing 0.1 mm silica beads. Use autoclaved hand-cut blue tips that allow pipetting even in the presence of remaining debris. This aliquot is processed immediately.

   **Note:** The remaining sample is frozen at –80°C for long-term storage.
4. Add 250 μL 4 M guanidinium thiocyanate to the sample. This step is necessary to denature proteins.
5. Add 500 μL 5% N-lauroylsarcosine sodium salt, which is an ionic surfactant that separates all cellular components from each other.
6. Incubate the samples for 60 min at 70°C while shaking at 700 rpm.
7. Lyse remaining microbial cells by using a FastPrep-24 fitted with a CoolPrep adapter (filled with a handful of dry ice). The FastPrep instrument performs the lysis of biological samples by using an optimized motion to disrupt cells through beating of beads on the sample material.
   a. Program: 5
   b. Cycles: 40 s; 6.5 m/s
   c. 3 rounds (add more dry ice between each round)
8. Add 15 mg polyvinylpyrrolidone (PVPP), a polymer used for removing phenolics and other fecal contaminants.
9. After vortexing, centrifuge for 3 min at 15,000 × g and 4°C.
10. 500 μL of the supernatant is transferred to a new 2-mL tube.
11. Add 5 μL RNase A and incubate for 20 min at 37°C while shaking at 700 rpm.

The DNA is then purified using a silica membrane-based approach following the manufacturer’s instructions of the kit used (NucleoSpin gDNA Clean-up Kit, REF 740230.250 Machery-Nagel).

12. Add 1500 μL Binding Buffer and vortex for 5 s.
13. Transfer each sample to one column: this is performed in three steps with each 650 μL. After each transfer, columns are centrifuged for 30 s (11,000 × g); discard the flow-through.
14. Wash columns by adding 700 μL Washing Buffer. After 2 s vortex, columns are centrifuged for 30 s (11,000 × g); discard the flow-through. Washing is performed three times.
15. Dry the silica membrane by centrifuging the columns for 1 min (11,000 × g) and discard the collection tube.
16. Add 50 μL Buffer DE to elute the DNA. Incubate for 1 min and centrifuge for 1 min (as before). Repeat the elution step and pool the flow-through to obtain a final volume of 100 μL with the isolated DNA.
After DNA purification, nucleic acid concentrations are measured by using a NanoDrop.

**Note:** Use a DNA solution of known concentration and measure serial dilutions thereof to check for the accuracy of the NanoDrop.

### Library Construction by Polymerase Chain Reaction

© Timing: approx. 3 h for 96 samples

17. Dilute isolated DNA of each sample to a final concentration of 12 ng/µL in 20 µL water into a 96-well skirted plate.
18. Prepare the Master Mix (Table 1) for the first (1st) PCR.
19. Transfer 27 µL of the prepared Master Mix (per well) and add 3 µL of the sample (per well) to a new 96-well skirted plate. The well plate with 30 µL sample per well is covered with a foil seal and is centrifuged for 30 s at low speed to collect the liquid at the bottom.
20. Put the plate into the cycler (Biometra TAdvanced) and run the first (1st) PCR program for 15 cycles following the time and temperature settings shown in (Table 2).
21. Prepare the Master Mix (Table 3) for the second (2nd) PCR including forward index primer. For each 96-well plate, 6 different forward primer and 16 different reverse primer are used. The reverse primer is not included in the Master Mix, they are divided in strips which are placed in the robot working area as well. For each of the six forward primer one separate Master Mix is to be prepared.
22. After the first PCR the plate returns to the robot.
23. Mix 2 µL of the DNA from 1st PCR, 45.5 µL of the Master Mix (Table 3), and 2.5 µL of one reverse index primer. Primer are combined in order to insert a double index in each sample following the method introduced by Kozich et al. (2013). It is possible to select from 38 forward and 60 reverse primer Table 4.
24. The plate is covered again with a PCR foil seal and is centrifuges for 30 s as before.
25. The second PCR starts by putting the covered plate into the cycler (Biometra TAdvanced). Run the program for ten cycles following the time and temperature settings shown in (Table 5).

|| Pause Point: After the second PCR, the plate can be stored at 4°C for 1 day.
26. Pool the final PCR products of both plates after the second PCR, which results in a total volume of 100 µL per sample.

**Note:** Fifteen µL can be used for quality control issues (e.g., gel electrophoresis).

### Library Cleaning

© Timing: approx. 1 h 30 min for 96 samples

### Table 1. Master Mix for 1st PCR

| Reagents                                      | Volume µL/Sample |
|-----------------------------------------------|------------------|
| Phusion® HF Buffer (without Dye)              | 6                |
| dNTPs (20 µmol)                               | 0.6              |
| 341F-ovh Primer (20 µM)                       | 0.1875           |
| 785r-ovh Primer (20 µM)                       | 0.1875           |
| Phusion® High-Fidelity DNA Polymerase Hotstart| 0.15             |
| DMSO (100%)                                   | 2.25             |
| Water (for molecular biology, DEPC-treated and filter-sterilized) | 17.625 |
PCR purification is performed with AGENCOURT AMPure XP Beads (Beckman Coulter) and again fully automatized using Beckman Coulter Biomek 4000 robot.

27. Prior to the library cleaning
   a. Remove the AMPure XP beads from 4°C storage and let stand for at least 30 min to bring to 20°C–22°C.
   b. Vortex the AMPure XP beads until they are well dispersed.
28. Add 1.8 μL AMPure XP beads per 1.0 μL PCR product. Using a P1000 multi-channel pipette, the robot gently pipettes the entire volume up and down 10-times to mix thoroughly.
   Note: For stool samples, the standard settings are 85 μL PCR product and 153 μL AMPure XP beads resulting in a total volume of 238 μL.
29. Incubate at 20°C–22°C for 5 min.
30. Put the well plate in the magnetic rack and let stand at 20°C–22°C for 5 min or until the liquid becomes clear in appearance. The robot removes the all of the clear supernatant using a P1000 multi-channel pipette.
31. The fragment is bound to the beads and 200 μL freshly prepared 70% EtOH is added to each well using a P250 without barrier.
32. Leave at 20°C–22°C for 30 s and discard the supernatant. Take extra care not to disturb the beads.
33. Steps 31 and 32 are repeated once more, for a total of two 70% EtOH washes.
34. Let the 96-well plate at 20°C–22°C for 4-5 min for drying, and then remove from the magnetic rack.
35. Re-suspend the bead pellet in each well in 80 μL BE Elution (recommended volume of AMPure standard protocol). The robot gently pipettes the entire volume up and down 10-times to mix thoroughly using a P250 multi-channel pipette.
   △ CRITICAL: The amount of added Elution Buffer depends on the DNA yield of the PCR product. Low amounts of PCR product, i.e., weak bands on the gel, should be re-suspend with amount at or below 20 μL BE Elution.

### Table 2. Settings for 1st PCR. Rows in gray are performed for 15 cycles.

| PCR Cycling Conditions | in °C | Time   | Cycles |
|------------------------|-------|--------|--------|
| Initial Denaturation   | 98    | 30 s   | 1      |
| Denaturation           | 98    | 5 s    | 15     |
| Annealing              | 55    | 10 s   |        |
| Extension              | 72    | 10 s   |        |
| Final Extension        | 72    | 2 min  | 1      |
| Hold                   | 10    | ∞      |        |

### Table 3. Master Mix for 2nd PCR

| Reagents                      | Volume in μL/Sample |
|-------------------------------|---------------------|
| Phusion® HF Buffer (without Dye) | 10                  |
| dNTPs (20 μmol, Bioline BIO-39043) | 1                  |
| Forward primer (e.g., 341-ovh-HTS- SC501 Primer (20 μM)) | 0.313               |
| Phusion® High-Fidelity DNA Polymerase Hotstart | 0.2                |
| DMSO (100%)                   | 1.5                 |
| Water (for molecular biology, DEPC-treated and filter-sterilized) | 32.487              |
36. Incubate the 96-well plate at 20°C–22°C for 2 min.
37. Place the 96-well plate on the magnetic rack at 20°C–22°C for 2 min or until the liquid becomes clear in appearance. Seventy μL of the clear supernatant from each well are transferred to an XP plate. Eight μL are transferred to a second plate for DNA measurements by fluorimetry (Qubit measurement according to the manufacturer’s instructions).

**Note:** If not enough volume is available, the total amount is transferred manually.
38. Samples are diluted to a concentration of 2 nM and finally diluted to a concentration of 0.5 nM.
39. From each sample of the 96-well plate, 5 μL are transferred to a low binding tube (pool of all samples of one plate).

¶ **Pause Point:** After the library cleaning the plate can be stored at 4°C for 1 day.

### Prepare Samples for 16S rRNA Gene Sequencing

40. Calculate molarity of each sample based on measured Qubit concentrations for a mean over four measurements:

\[
\text{concentration in nM} = \frac{\text{(concentration in ng/μL)} \times 10^6}{\text{(average library size in bp} \times 660 \text{ g/mol)}}
\]

|        | Qubit-1 (ng/μL) | Qubit-2 (ng/μL) | Qubit-3 (ng/μL) | Qubit-4 (ng/μL) | Mean (ng/μL) | nM  |
|--------|-----------------|-----------------|-----------------|-----------------|--------------|-----|
| Final  | Pool_1          | 0.18            | 0.18            | 0.19            | 0.18         | 0.18 | 0.49 |
|        | Pool_2          | 0.17            | 0.18            | 0.18            | 0.18         | 0.18 |     |
|        | Pool_3          | 0.17            | 0.17            | 0.18            | 0.18         | 0.18 |     |

**Note:** For V3V4, the average library size is 572 bp.

Following steps are necessary to denature the DNA and set to a concentration of 20 pM.

41. Create a fresh 0.2 nM NaOH solution and a 0.2 nM Tris HCl solution.
42. Add 40 μL of the 0.5 nM DNA pool and 40 μL of the 0.2 N NaOH solution to a 1.5-mL tube.
43. Vortex the sample and centrifuge for 1 min (280 x g). Leave it in a stand for 5 min at 20°C–22°C.
   Add 40 μL of the 0.2 nM Tris HCl solution. Vortex the mixture and centrifuge for 1 min (280 x g).
44. Incubate for 5 min at 95°C and for 5 min at 4°C.
45. Add 880 μL cooled HT1-Buffer to the denaturated DNA pool to generate a 20 pM library.
46. Dilute the DNA to get the final pM concentration of 10 pM final library concentration that was spiked-in using 20% (v/v) PhiX. PhiX DNA in a ready to sequence library (Illumina PhiX Control v3, FC-110-3001) is added in order to increase complexity for the first few bases sequenced. Otherwise, the sequencer miscalculates the amount of the dominating base and the sequencing fails.

47. Six-hundred µL of the final pool is transferred to the Illumina MiSeq cartridge v3 with 600 cycles.

**EXPECTED OUTCOMES**

After sequencing, the demultiplexed FASTQ files (forward and reverse file for each sample, Illumina bcl2fastq software) are transformed into Operational Taxonomic Unit (OTU) tables using the IMNGS (Lagkouvardos et al., 2016) platform which is based on the UPARSE approach for sequence quality check, chimera filtering, and cluster formation. To avoid spurious OTUs, we recommend a filtering threshold of 0.25% to remove artificial species Reitmeier et al. (2020).

For downstream analysis, the generated OTU table is normalized by using the fully modular R pipeline Rhea (Lagkouvardos et al., 2017). The pipeline also provides information about alpha-diversity (within-sample diversity), beta-diversity (between-sample diversity) and generates a taxonomic classification.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**Quality Control**

The quality of the sequencing run is evaluated by FASTQC, which provides a modular set of quality control analysis. Graphical illustration about the quality scores over reads (bp) is used to show any problems occurred during the sequencing run.

For human stool samples, it is intended to have for each sample about 10,000 reads (or more) after trimming, filtering, and chimera checking. Samples with too low number of reads should be excluded. However, the exact minimum threshold of reads depends on the studied environment and sequencing technology.

Samples with total processed reads below the determined threshold should be re-sequenced.

**Statistical Analysis**

1. Descriptive analysis and data handling
   a. Handling sparsity in microbial datasets. For the analysis of 16S rRNA gene sequencing data of the large population-based cohort studies, we excluded OTUs with a relative abundance <0.1% and a prevalence <10%.
   b. Adjust for confounding and determine effect modifier.
      i. Confounders and effect modifiers are determined via a permutational multivariate analysis of variance using a distance matrix. For the confounders, the function is applied on the
Bray-Curtis distance matrix considered as independent variable. The dependent variables are known confounding factors for which the data should be stratified anyway and the outcome of interested (e.g., Type 2 Diabetes).

ii. Effect modifiers help to explain the variation of the underlying microbial ecosystem. They are not considered as confounders but as contributors to the total variation. Therefore, co-variables are individually tested and ranked according to their significant explained variation.

iii. Statistical analysis to determine differences between groups/samples is obtained via linear regression model using \texttt{lm} from the R package \texttt{vegan} adjusted for the previously determined confounding factors.

2. Machine learning – tool for classification and prediction

A random forest model is used to classify binary outcome variables based on a combination of BMI and microbial composition with a 5-fold cross validation by using \texttt{randomForest} from the R package \texttt{randomForest} v4.6-14.

To receive a robust and generalizable classification model, the machine-learning algorithm is applied 100-times iteratively assigning randomly individuals to either the training (80%) or test set (20%). For the training set, a subset of equally distributed T2D and nonT2D cases is taken to train the model. The model is further validated on the 20% test set. Based on out-of-bag error rates and Gini index, the most important features are selected for each iteration using \texttt{rfcv} from R package \texttt{randomForest} v4.6-14. Features, which appear in at least 50% of all 100 random forest models, are considered as classification feature for the final model (Figure 3)

⚠ CRITICAL: To avoid overfitting of the classifier the data input needs to be reduced in advance, for example, based on a predefined cutoff for minimum relative abundance and prevalence.

**Figure 3. Random Forest Model for T2D Classification**

Curves of receiver operating characteristics (ROC) for a random forest model using a training set (train set) of 80% of the data (dashed lines in the left panel) as well as using a test set with the remaining 20% of the data (ROC curves in the right panel). The mean AUC over 100 random data splits is shown. The boxplots below the curve panels show the distribution of AUCs across all generated models for the corresponding training and test sets, respectively.

Reused figure from Reitmeier et al. (2020); permission obtained from the corresponding author.
3. Implementation of a Generalized Linear Model

For the risk prediction of T2D, a generalized linear model (GLM) for binomial distribution and binary outcome (logit) is generated using the previously selected features based on arrhythmic OTUs including BMI as additional variable. Therefore, two approaches are followed. First, the model is tested in a nested 80% - training and 20% - test scenario as described in the previous section for the random forest model. To verify the importance of the selected features, a generalized linear model for control OTUs is implemented repetitively 100-times (Figure 4).

4. Circadian analysis of human stool samples
   a. Identify rhythmic OTUs’ “Pre-filtering”
      i. Collection daytime needs to be converted into a 24-h time scale ranging from 0 to 23:59 h (see “Time point” in Table 6).
      ii. The Raw OTU table including “Time point” need to be transferred in GraphPad Prism using an XY table with single Y values for each time point (Figure 5).

---

**Figure 4. Generalized Linear Model**

ROC curves for classification of T2D. The distribution of AUCs are shown by boxplots and are significantly different between the types of models. Results showed that the classification of T2D in the 20% blind test set performed comparable as the 5-fold cross validated data.

Reused figure from Reitmeier et al. (2020); permission obtained from the corresponding author.
iii. A cosine-regression can be applied for each single OTU by using the Analyze button.

A nonlinear regression needs to be applied with the following equation (Figure 6):

\[ y = \text{baseline} + \frac{\text{amplitude}}{C_3} \cos\left(\frac{2 \pi x}{C_{18}}\right) \]

or a double harmonic cosine wave equation:

\[ y = \text{baseline} + \left(\frac{\text{amplitude } A \cos\left(\frac{2 \pi x}{C_{18}}\right)}{C_{3}}\right) + \left(\frac{\text{amplitude } B \cos\left(\frac{4 \pi x}{C_{18}}\right)}{C_{3}}\right) \]

on alpha-diversity and relative abundance, with a fixed 24-h period.

---

**Table 6. Example of a Raw OTU Table with Assigned Time Points and Intervals**

| Time | Interval | Group | Subject ID | OTU 1 | OTU 2 | OTU 3 | OTU X |
|------|----------|-------|------------|-------|-------|-------|-------|
| Daytime | Time Point |       |            |       |       |       |       |
| 00:01 | 0        | 23.5  | A          | XXX1  | A1    | A2    | -     | A4    |
| 00:05 | 0        | 23.5  | B          | YYY1  | B1    | -     | B3    | B4    |
| 00:10 | 0        | 23.5  | C          | XXX2  | A1    | A2    | A3    | A4    |
| 01:10 | 1        | 1.5   | A          | XXX3  | A1    | A2    | A3    | A4    |
| 04:20 | 4        | 3.5   | C          | YYY2  | -     | B2    | B3    | -     |
| 11:20 | 11       | 11.5  | B          | YYY3  | B1    | B2    | B3    | B4    |
| ...   | ...      | ...   | ...        | ...   | ...   | ...   | ...   | ...

---

**Figure 5. OTU Table in GraphPad**

The Excel sheet is transferred in an XZ/Sheet in GraphPad Prism for further analysis.
iv. The goodness of fit needs to be corrected for multiple comparisons and the significance determined using an F-test. Each p value needs to be Bonferroni-adjusted for multiple testing. A statistically significant difference can be assumed when p <= 0.05.

v. Most circadian rhythm detection algorithms were developed to assess the significance of rhythms in large data sets obtained from gene expression analysis (e.g., microarray, in situ hybridization) with relatively low sampling rates (~1 sample/h). Thus, microbiota data collected throughout the 24-h day need to be combined in hourly intervals to be analyzed with different methods. Alternatively to the cosine wave regression fit, which can handle high sampling rates, the rhythmicity detection algorithm named JTK_CYCLE (Hughes et al., 2010) can be used. JTK_CYCLE employs a non-parametric algorithm, detecting sinusoidal signals and, therefore, is more reliable when data are not normally distributed. Importantly, 4-h sampling intervals are a minimum and JTK_CYCLE is not working well with only one daily cycle. Nevertheless, JTK_CYCLE presents the highest false negative rates (Hughes et al., 2009). For example, the OTU table in Table 6 can be transposed as illustrated in Figure 7.

Although microbiota sequencing data are predominantly sinus shaped, the analysis may certainly profit from adding harmonics in order to describe a more complex microbiota profile. Harmonics are integrated, e.g., in CircWave or Harmonic cosine wave regression. CircWave is different from JTK_CYCLE in that it uses a parametric approach, e.g., an F-tested forward harmonic regression procedure similar to the Cosine- or Harmonic cosine wave regression, except that CircWave automatically detects how many harmonics can be added by F-test criterion (step forward regression style). Thus, it is likely more powerful to detect rhythmicity in normal distributed data compared to JTK_CYCLE.
Unfortunately, in comparison to gene expression data, human sequencing data are particular in multiple ways: (i) the prevalence of OTUs can vary between groups and between individuals within one group, and (ii) the distribution of fecal samples in a human population study varies dramatically over the 24-h day. In particular, defecation occurs in 70% of the people between 5 and 11 am. Consequently, an algorithm assuming equally distributed samples over the course of the day, such as CircWave, would need optimization. A method which works independently of the sample size per time point and which can handle missing values, are the Cosine and Harmonic cosine wave regression. Both are parametric analyses similar to CircWave, which can integrate up to two harmonics. Other possibilities are represented by the online tool Nitecap (unpublished) or RAIN (Thaben and Westermark, 2014), which, similarly to JTK_CYCLE, represents a non-parametric method for the detection of rhythms in biological data sets and, thus, can detect arbitrary wave forms. Nevertheless, RAIN requires a fairly powerful computer, which, at least in our case, must be able to handle data from more than 2,000 subjects.

In summary, we highly recommend identifying rhythms in microbiome data sets with multiple tools, including parametric and non-parametric, non-harmonic and harmonic logarithms, depending on the microbiome data set available.

Note: There are various analysis tools available, which combine multiple methods, such as MetaCycle (Wu et al., 2016), incorporating JTK_CYCLE, ARSER (Yang and Su, 2010), and Lomb-Scargle (Lomb, 1976). Nevertheless, ARSER does not considers replicates and cannot cope with missing data that are likely present with microbiome data.

Note: Time points are named from Row 1B onwards. Single OTU names are found in Column 1B downwards. When saved as txt file, JTK_CYCLE identifies significantly rhythmic OTUs with a p value corrected for multiple regression as illustrated in yellow in the output file (Figure 8).

vi. Importantly, the circadian analysis needs to be performed separately for group A and group B. Thereby, the amount of rhythmic OTUs in group A can be compared to the amount of rhythmic OTUs in group B. However, to compare rhythmicity of a specific OTU directly between the two groups, further analysis, as described in 4a, is necessary.
b. Detection of differential rhythmicity of specific OTUs, e.g., comparing rhythmicity of different genotypes, treatments, or phenotypes

i. The relative abundance of each OTU was assessed for a 24-h rhythmicity in the pre-filtering step 4a using the cosine wave regression, JTK_CYCLE or any other circadian analysis software for each group examined (such as non-T2D or T2D) separately. With this pre-filtering method, the amount of OTUs from all OTUs analyzed will be identified as significantly rhythmic in group A and independently in group B. However, these rhythmic OTUs can differ between the groups. Therefore, all OTUs rhythmic in at least one group need to be further analyzed for differential 24-h time-of-day patterns comparing data from group A with group B using the Detection of Differential Rhythmicity (DODR) R packages (Thaben and Westermark, 2016).

ii. One OTU table per group needs to be generated in txt format. Importantly, the same OTUs need to be listed in both group A and in the file from group B as illustrated in Tables 7 and 8.

iii. In the DODR output table, the results from all applied analysis (described in detail by Thaben and Westermark (2016) are indicated for every specific OTU (see Table 9) including the p value for the robustDODR analysis.

iv. Resulting DODR p-values need to be corrected for multiple comparisons and for significant OTUs that have a corrected p value ≤0.05, a significance level can be identified, e.g., which OTUs appear rhythmic in group A, but show a differential rhythmicity in group B.

v. To address what kind of difference appears between the two groups, such as amplitude or phase, differences can be analyzed by an additionally R package called “HarmonicRegression” (Luck et al., 2014).

---

**Figure 8. JTK Output Table**

Columns are referring to adjusted q (BH Q) and p value (ADJ. P), period (PER), phase (LAG) and amplitude (AMP) as well as the relative abundance values of the corresponding OTU (rows).

---

**Note:** These results will determine whether an OTU, which appears rhythmic in group A, also (1) exhibit circadian oscillation, (2) shows a different rhythmicity (i.e., phase and amplitude), or (3) lacks rhythmicity in group B and vice versa.

ii. One OTU table per group needs to be generated in txt format. Importantly, the same OTUs need to be listed in both group A and in the file from group B as illustrated in Tables 7 and 8.

**Note:** The time points may differ between the groups.

iii. In the DODR output table, the results from all applied analysis (described in detail by Thaben and Westermark (2016) are indicated for every specific OTU (see Table 9) including the p value for the robustDODR analysis.

iv. Resulting DODR p-values need to be corrected for multiple comparisons and for significant OTUs that have a corrected p value ≤0.05, a significance level can be identified, e.g., which OTUs appear rhythmic in group A, but show a differential rhythmicity in group B.

v. To address what kind of difference appears between the two groups, such as amplitude or phase, differences can be analyzed by an additionally R package called “HarmonicRegression” (Luck et al., 2014).

5. Illustration of cosine wave-fitted grouped data using GraphPad Prism

Grouping of subjects to predefined time intervals.
a. To receive the highest possible resolution of the curve fit, time intervals need to be predefined with the goals to (i) include an equal number of subjects per interval and (ii) group for further circadian rhythm analysis. The higher the frequency of sample collection, the better the resolution. Next, subjects are to be grouped according to the assigned intervals (i.e., bins; see "Interval" in Table 6). For instance, for 2-h intervals, data from subjects collected within 23:00 p.m. and 0:59 a.m. are merged into one bin referred to as "23.5."

CRITICAL: When time intervals are assigned, group sizes should be equal between time points within one group and between groups. For example, when data are obtained from 360 subjects per group, each of the 12 2-h intervals should include 30 subjects.

b. Results from the different groups need to be averaged per interval within each group (as illustrated in Table 10) to be transferred to GraphPad Prism using, e.g., an XY table with mean (AVE) values ± standard deviation (SD) and sample size (n) calculated, e.g., in Microsoft Excel.

c. As described in paragraph 4aiii, a cosine-wave regression will be applied and the significance of the goodness of fit is evaluated by an F-test. In case the cosine fit is reaching significance, the cosine wave can be illustrated in the graph, whereas a non-significant fit is shown by simply connecting straight lines between data points (see Figure 9).

LIMITATIONS

- The sample preparation strongly influences the outcome and quality of the sequencing, which limits the comparability between studies.

Table 7. OTU Table Group A

| Time Point | OTU 10  | OTU 15     | OTU 100    | OTU 219    | OTU 412    | OTU n    |
|------------|---------|------------|------------|------------|------------|----------|
| 0          | 0       | 0.103057   | 0          | 0          | 0.246193   | 0.080156 |
| 0          | 0.011865| 0          | 0          | 0          | 0.219493   | 0.065255 |
| 0          | 0       | 0.030116   | 0          | 0.007529   | 0.037645   | 0.097877 |
| 0          | 0.011885| 0          | 0          | 0          | 0.178274   | 0        |
| 1          | 0.011277| 0.078935   | 0.033829   | 0.101488   | 0.236806   | 0.045106 |
| 1          | 0.058903| 0          | 0          | 0          | 0.008415   | 0        |
| 1          | 0       | 0.07632    | 0          | 0.010176   | 0.040704   | 0.055968 |
| 1          | 0       | 0.102211   | 0          | 0.016139   | 0.059175   | 0.032277 |
| 2          | 0       | 0.120948   | 0          | 0.02419    | 0.02419    | 0.036284 |
| 2          | 0       | 0.166207   | 0.05084    | 0.015643   | 0.199449   | 0.021509 |
| 2          | 0.30281 | 0          | 0          | 0.072674   | 0.096899   | 0.084787 |
| 3          | 0       | 0          | 0.181148   | 0          | 0.162409   | 0.037479 |
| 3          | 0       | 0.146516   | 0          | 0.070545   | 0.059692   | 0.179075 |
| 3          | 0       | 0          | 0.029483   | 0.041276   | 0.053069   | 0.076655 |
| 4          | 0       | 0.02998    | 0          | 0.059959   | 0.083943   | 0.077947 |
| 4          | 0       | 0          | 0          | 0          | 0.059968   | 0.21322 |
| 4          | 0       | 0          | 0          | 0.139297   | 0.294072   | 0.023216 |
| 5          | 0       | 0.017839   | 0.029732   | 0.011893   | 0.053517   | 0        |
| 5          | 0       | 0.030765   | 0.006153   | 0          | 0.067684   | 0.049225 |
| 5          | 0       | 0.083903   | 0          | 0          | 0.023972   | 0        |
| n          | ...     | ...        | ...        | ...        | ...        | ...      |
Bioinformatic methods, clustering approaches, and filtering can influence the abundance of certain taxonomies.

Taxonomic classification of 16S rRNA gene sequencing data is limited in its accuracy to assign species or even strains correctly. The taxonomic assignment is only based on a short amplicon, which increases the difficulty to determine correctly the bacterial species found. This designation also depends on the used database, which have differences when comparing them against each other.

In human studies, it is almost impossible to cover all daytimes for the analysis of circadian rhythmicity.

A minimum of approx. 300 samples distributed across the full day are required within a single group to achieve a resolution necessary to detect significant circadian rhythms.

**TROUBLESHOOTING**

**Problem 1**

Strikingly low 260/280 values obtained by NanoDrop could be due to mistakes during the DNA cleaning step (e.g., ethanol residuals in cleaning columns) (DNA Isolation, steps 1–16).

| Time point | OTU 10 | OTU 15 | OTU 100 | OTU 219 | OTU 412 | OTU n |
|------------|--------|--------|---------|---------|---------|-------|
| 0          | 0.540106 | 0.700988 | 0       | 0.103425 | 0.022983 | 0.321765 |
| 0          | 2.30002 | 0.006785 | 0.013569 | 0.169618 | 0.122125 | 0.061062 |
| 0          | 0.769108 | 0.376542 | 0.040058 | 0       | 0       | 0.176254 |
| 0          | 1.071233 | 0.680272 | 0.000781 | 0       | 0.179842 | 0.062554 |
| 0          | 3.63199 | 0.95009 | 0.084185 | 0       | 0.481058 | 0.264582 |
| 0          | 2.406787 | 0.629112 | 0.294355 | 0.017315 | 0.086575 | 0.06926 |
| 1          | 6.290906 | 0.009768 | 0.019537 | 0       | 0.048842 | 0.029305 |
| 1          | 0.256082 | 1.099018 | 0       | 0.298762 | 0       | 0.02134 | 0.096031 |
| 1          | 0.47203 | 0.934016 | 0       | 0.371598 | 0.160691 | 0.230993 |
| 2          | 0.197815 | 0.847777 | 0       | 0       | 0.047099 | 0.36737 |
| 2          | 0.011863 | 0.972774 | 0       | 0.213536 | 0       | 0.017795 | 0 |
| 3          | 1.427067 | 0     | 0       | 0       | 0.023395 | 0.666745 |
| 3          | 6.60828 | 0     | 0.009952 | 0       | 0.009952 | 0.159236 |
| 3          | 7.888502 | 0.111498 | 0.515679 | 0       | 0.097561 | 0 |
| 4          | 2.59718 | 0.599349 | 0.313945 | 0.342485 | 0.079913 | 0.022832 |
| 4          | 1.453825 | 0.022028 | 0.016521 | 0       | 0.049562 | 0.033041 |
| 5          | 2.931624 | 1.134666 | 0.421816 | 0.11389 | 0.177163 | 0.054836 |
| 5          | 0.549429 | 0.759849 | 0       | 0.207277 | 0.015587 | 0.066243 |
| 5          | 2.644024 | 0.010748 | 0       | 0       | 0.150473 | 0.042992 |
| N          | ...     | ...    | ...     | ...     | ...     | ...     |

- Bioinformatical methods, clustering approaches, and filtering can influence the abundance of certain taxonomies.
- Taxonomic classification of 16S rRNA gene sequencing data is limited in its accuracy to assign species or even strains correctly. The taxonomic assignment is only based on a short amplicon, which increases the difficulty to determine correctly the bacterial species found. This designation also depends on the used database, which have differences when comparing them against each other.
- In human studies, it is almost impossible to cover all daytimes for the analysis of circadian rhythmicity.
- A minimum of approx. 300 samples distributed across the full day are required within a single group to achieve a resolution necessary to detect significant circadian rhythms.

**Table 8. OTU Table Group B**

| Time point | OTU 10 | OTU 15 | OTU 100 | OTU 219 | OTU 412 | OTU n |
|------------|--------|--------|---------|---------|---------|-------|
| 10         | 0.3543 | 3.33E-16 | 1       | 1.33E-15 | 4.66E-06 | 0.538833 | 2.00E-15 |
| 15         | 0.0042 | 1.11E-13 | 1       | 8.38E-12 | 6.97E-05 | 0.005814 | 6.65E-13 |
| 100        | 0.2147 | 0       | 1       | 0       | 0.000288 | 0.180528 | 0 |
| 219        | 0.0029 | 5.44E-15 | 1       | 1.85E-13 | 0.000567 | 0.030348 | 3.26E-14 |
| 412        | 0.0006 | 1.09E-13 | 1       | 7.21E-12 | 0.00073 | 0.077292 | 6.54E-13 |
| N          | 0.0001 | 9.26E-13 | 1       | 6.59E-11 | 0.002062 | 0.031011 | 5.56E-12 |

**Table 9. DODR Output Table**

| OTU | HANOVA | HarmNoisePred1 | HarmNoisePred2 | HarmScaleTest | robustDODR | robustHarmScaleTest | meta.p.val |
|-----|--------|----------------|----------------|---------------|------------|---------------------|------------|
| 10  | 0.3543 | 3.33E-16 | 1 | 1.33E-15 | 4.66E-06 | 0.538833 | 2.00E-15 |
| 15  | 0.0042 | 1.11E-13 | 1 | 8.38E-12 | 6.97E-05 | 0.005814 | 6.65E-13 |
| 100 | 0.2147 | 0 | 1 | 0 | 0.000288 | 0.180528 | 0 |
| 219 | 0.0029 | 5.44E-15 | 1 | 1.85E-13 | 0.000567 | 0.030348 | 3.26E-14 |
| 412 | 0.0006 | 1.09E-13 | 1 | 7.21E-12 | 0.00073 | 0.077292 | 6.54E-13 |
| N   | 0.0001 | 9.26E-13 | 1 | 6.59E-11 | 0.002062 | 0.031011 | 5.56E-12 |
Potential Solution
If enough starting material (e.g., stool sample) is available, the sample preparation needs to be repeated. Including one additional washing step during the DNA isolation (DNA Isolation, steps 1–16).

Problem 2
Low biomass samples could result in insufficient PCR products (Library construction by Polymerase Chain Reaction, steps 17–26).

Potential Solution
Increase the number of the second PCR cycles and/or the dilution of the sample could help to overcome this problem.

Problem 3
Samples with a low number of reads could be caused by problems during the demultiplexing (e.g., misdisposition of indices) (Library construction by Polymerase Chain Reaction, steps 17–26).

Potential Solution
Double-check the assigned index primers with the sequences provided in the samples sheet. Adjust trimming length of the forward and reverse reads.

Table 10. Results from Different Groups Averaged per Interval within Each Group

| Interval | Group A Average | SD | n | Group B Average | SD | n |
|----------|-----------------|----|---|-----------------|----|---|
| 1.5  | AVE1            | SD1 | 30 | AVE1            | SD1 | 30 |
| 2.5  | AVE2            | SD2 | 30 | AVE2            | SD2 | 30 |
| 5.5  | AVE3            | SD3 | 30 | AVE3            | SD3 | 30 |
| 7.5  | AVE4            | SD4 | 30 | AVE4            | SD4 | 30 |
| 9.5  | AVE5            | SD5 | 30 | AVE5            | SD5 | 30 |
| 11.5 | AVE6            | SD6 | 30 | AVE6            | SD6 | 30 |
| 13.5 | AVE7            | SD7 | 30 | AVE7            | SD7 | 30 |
| 15.5 | AVE8            | SD8 | 30 | AVE8            | SD8 | 30 |
| 17.5 | AVE9            | SD9 | 30 | AVE9            | SD9 | 30 |
| 19.5 | AVE10           | SD10| 30| AVE10           | SD10| 30|
| 21.5 | AVE11           | SD11| 30| AVE11           | SD11| 30|
| 23.5 | AVE12           | SD12| 30| AVE12           | SD12| 30|

Figure 9. Illustration of Cosine-Wave Regression
Diurnal profiles of richness depending of subjects from different groups (red, Group B; black, Group A). Significant rhythms (cosine-wave regression, \( p \leq 0.05 \)) are illustrated with fitted cosine-wave curves; data points connected by straight lines indicate no significant cosine fit curves (\( p > 0.05 \)) and thus no rhythmicity.
Problem 4
Precipitous FASTQ curves could be an indicator for primer dimers, which could be due to poor purification of the sample (e.g., when using magnetic beads) (Library cleaning, steps 27–40).

Potential Solution
Repeat the purification step (Library cleaning, steps 27–40) with the pooled PCR products using AMPure XP magnetic beads with a lower concentration of 0.6 μL.

Problem 5
Insufficient data to calculate rhythmicity (Statistical analysis - 4. Circadian analysis of human stool samples)

Potential Solution
Sample numbers within a group needs to be increased. A study with 80 people requires approximately four samples per person, which results in 320 samples in total distributed across the day, to find diurnal rhythms comparable to results obtained from a cohort with more than 1,900 subjects from whom a single sample per person was collected (Reitmeier et al., 2020).

If an increase in sample size is not possible the distribution of collection times needs to be expanded. For example, the number of samples need to be spread across the daytime, such as 20–30 samples per daytime hours to receive a resolution to detect significant rhythms.

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Prof. Dr. Dirk Haller

Materials Availability
This study did not generate any unique materials or reagents.

Data and Code Availability
Sequence data, analyses, and resources related to the 16S rRNA gene sequencing of human cohort (N = 8), and data from human cohort are available upon request from the corresponding author. Software used to analyze the data are either freely or commercially available. Source code data are available from the corresponding author on request.

ACKNOWLEDGMENTS
The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. The project is embedded in the Collaborative Research Center funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) SFB 1371 (Projektnummer 395357507). The Technical University of Munich provided funding for the ZIEL Institute for Food & Health and technical assistance. The ZIEL Core Facility Microbiome provided outstanding technical support for sample preparation and 16S rRNA gene amplicon sequencing.

AUTHOR CONTRIBUTIONS
D.H. conceived and coordinated the project. S.R. and S.K. performed 16S rRNA gene sequencing data analysis. S.R. performed bioinformatics analysis. D.H. and K.N. supervised the work and data analysis. K.N. supported sample preparation and 16S rRNA gene sequencing analysis. S.R., S.K., K.N., and D.H. wrote the manuscript. All authors reviewed the manuscript.
Protocol

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., and Wishart, D.S. (2016). Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 44, W147–W153.

Carroll, I.M., Ringel-Kulka, T., Siddle, J.P., Kleenhammer, T.R., and Ringel, Y. (2012). Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. PLoS One 7, e46953.

Carroll, I.M., Ringel-Kulka, T., Siddle, J.P., Kleenhammer, T.R., and Ringel, Y. (2012). Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. PLoS One 7, e46953.

Dominazzi, C., Wu, J., Hayes, R.B., and Ahn, J. (2014). Comparison of methods for fecal microbiome biospecimen collection. BMC Microbiol. 14, 103.

Francoza, A.E., McVey, L.J., Rahnavaara, G., Thompson, L.R., Schirmer, M., Wengart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. Nat. Methods 15, 962–968.

Godon, J.-J., Zumstein, E., Dabert, P., Habouzet, F., and Moletta, R. (1997). Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. Appl. Environ. Microbiol. 63, 2802–2813.

Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., Knight, R., and Ley, R.E. (2014). Conducting a microbiome study. Cell 158, 250–262.

He, Z., Zhang, H., Gao, S., Lercher, M.J., Chen, W.H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized genome-scale data sets. J. Biol. Rhythms 25, 372–380.

Illert, E.E., Jorgensen, M., Nogueira-Julian, M., Daugard, G., Muth, M.D., Helleberg, M., Paredes, R., Lundgren, J., Sengelov, H., and MacPherson, C. (2019). Gut microbiome comparability of fresh-frozen versus stabilized-frozen samples from hospitalized patients using 16S rRNA gene and shotgun metagenomic sequencing. Sci. Rep. 9, 13351.

Revellie, W. (2020). psyehh: Procedures for Psychological, Psychometric, and Personality Research (Northwestern University). R package version 2.0.7.

Segata, N., Waldron, L., Ballarini, A., Larimer, M., Naas, J., Fraser, A., et al. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nat. Methods 9, 811–814.

Shaw, A.G., Sim, K., Powell, E., Cornwell, E., Cramer, T., McClure, Z.E., Li, M.S., and Kroll, J.S. (2016). Latitudinal variation in gut microbiome composition: the elephant in the room? Microbiome 4, 40.

Thaden, P.F., and Westermark, P.O. (2014). Detecting rhythms in time series with RAIN. J. Biol. Rhythms 29, 391–400.

Thaden, P.F., and Westermark, P.O. (2016). Differential rhythmicity: detecting altered rhythmicity in biological data. Bioinformatics 32, 2800–2808.

Yang, R., and Su, Z. (2010). Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. Bioinformatics 26, 1165–1174.

Yoon, S.H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., and Chun, J. (2017). Introducing EziBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int. J. Syst. Evol. Microbiol. 67, 1613–1617.