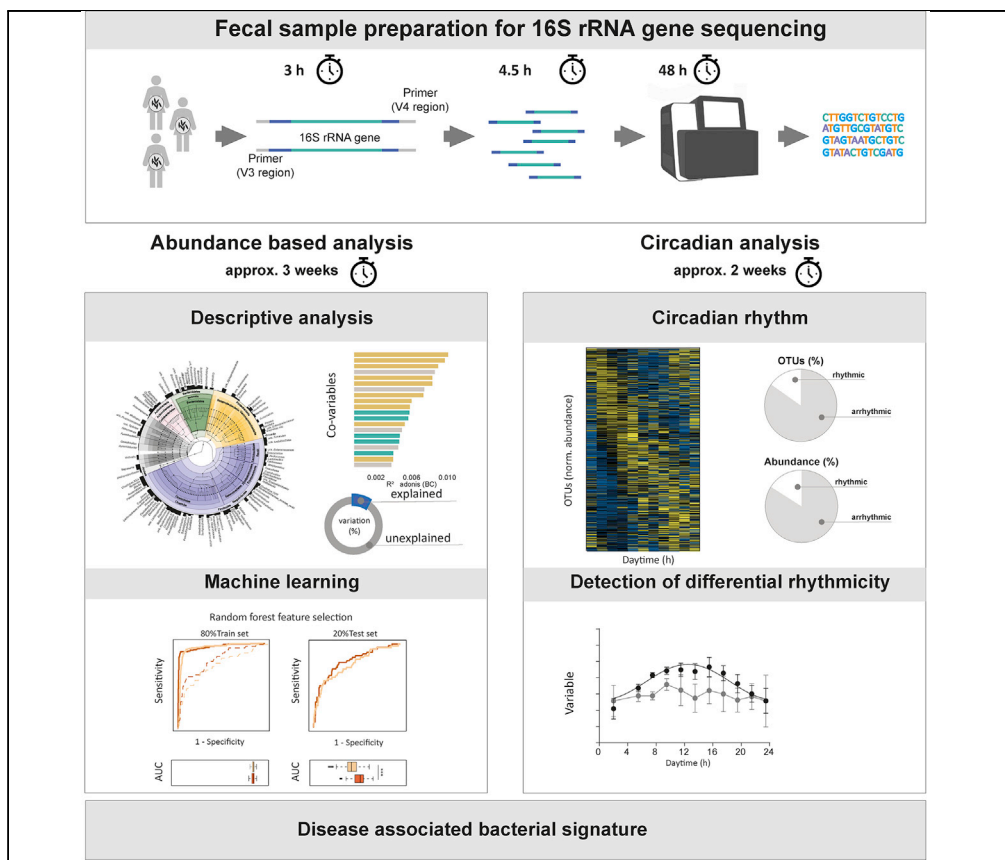


Protocol

Comparing Circadian Rhythmicity in the Human Gut Microbiome



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.

Sandra Reitmeier,
Silke Kiessling, Klaus
Neuhaus, Dirk Haller
dirk.haller@tum.de

HIGHLIGHTS
Walkthrough of
sample preparation
for 16S rRNA gene
sequencing for
human stool samples

Determine disease-
associated microbial
features based on
machine learning

Circadian analysis to
identify presence of
rhythms in a
population-based
cohort study

Define bacterial
signatures by
differences in
rhythmicity within/
between cohorts

Reitmeier et al., STAR
Protocols 1, 100148
December 18, 2020 © 2020
The Author(s).
<https://doi.org/10.1016/j.xpro.2020.100148>



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 Haller^{1,2,5,*}¹ZIEL - Institute for Food & Health, Technical University of Munich, 85354 Freising, Germany²Chair of Nutrition and Immunology, Technical University of Munich, Gregor-Mendel-Str. 2, 85354 Freising, Germany³These authors contributed equally⁴Technical Contact⁵Lead 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.



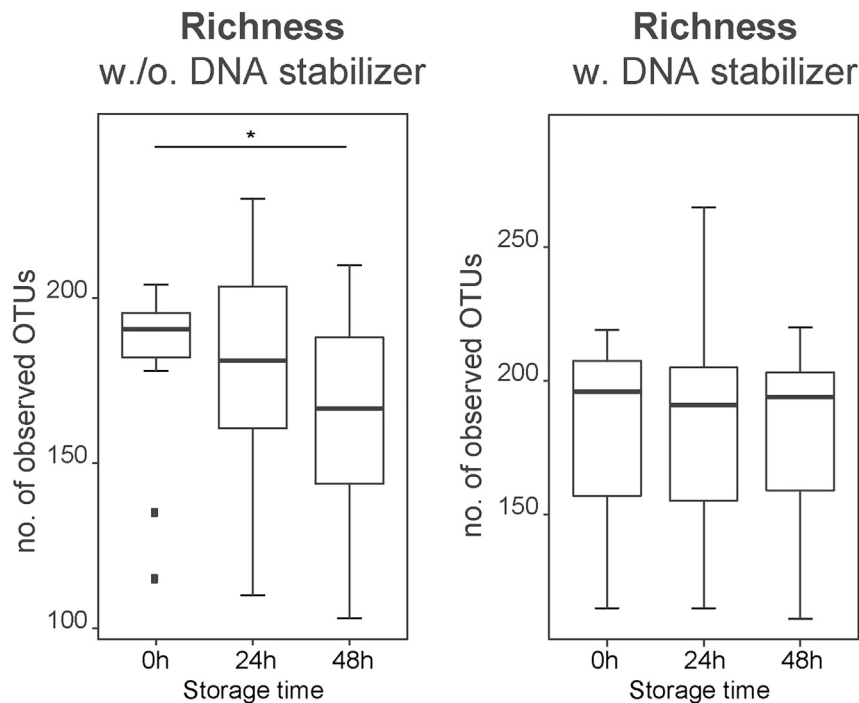


Figure 1. Influence of Storage Time and Number of Observed OTUs (Richness)

Boxplots shows either the changes without DNA stabilizer (left boxplot) or with DNA stabilizer (right boxplot) over time as indicated.

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°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°C .

△ **CRITICAL:** It is important to have unique labels for each sample. We recommend a barcode 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# L5125 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 × 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 RRID:SCR_015058

(Continued on next page)

Continued

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	Lagkourdos 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	Kanehisa 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	Lagkourdos et al. (2017)	https://github.com/Lagkourdos/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	Clindworth et al. (2013)	N/A
785R-ovh Primer: GACTACHVGGGTATCTAATCC	Clindworth 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- μ L tips with barrier	beckman	Cat# B01124
50- μ L tips with barrier	beckman	Cat# A21586
250- μ L tips with barrier	beckman	Cat# 717253
250- μ L tips without barrier	beckman	Cat# 717252
AMPure XP beads	beckman	Cat# A63881

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deep-well Plate (AB-1127)	Fisher Scientific	Cat# 10243223
PCR Tubes, 0.5 mL for Qubit (AXYGEN) PCR-05-C	Fisher Scientific	Cat# 11331974
Tips GP LTS, 20 μ L	Mettler-Toledo	Cat# 30389274
Tips GP LTS, 200 μ L	Mettler-Toledo	Cat# 30389276
Tips GP LTS, 1,000 μ L	Mettler-Toledo	Cat# 30389272
10/20 μ L RPT XL Graduated Filter Tip (Sterile)	StarLab	Cat# S1180-3710-C
0.2 m: 8-Strip "Non-Flex" Natural PCR Tubes, Ind-Attached Flat Caps (Xtra-Clear)	StarLab	Cat# I1402-3700
200 μ L RPT Graduated Filter Tip (Sterile)	StarLab	Cat# S1180-8710-C
10/20 μ L RPT XL Graduated Filter Tip (Sterile)	StarLab	Cat# S1180-3710-C
FastPrep-24	MP Biomedicals	Cat# 15260488
CoolPrep adapter	MP Biomedicals	Cat # 6002528
Biomek 4000 Automated Liquid Handler	Beckman coulter	Cat # C23350
Biometra TAdvanced	Analytik Jena AG	Cat # 846-x-070-211

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 μ L DNA Stabilizer from Invitex, 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 μ L 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 μ L 4 M guanidinium thiocyanate to the sample. This step is necessary to denature proteins.
5. Add 500 μ L 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 \times g and 4°C.

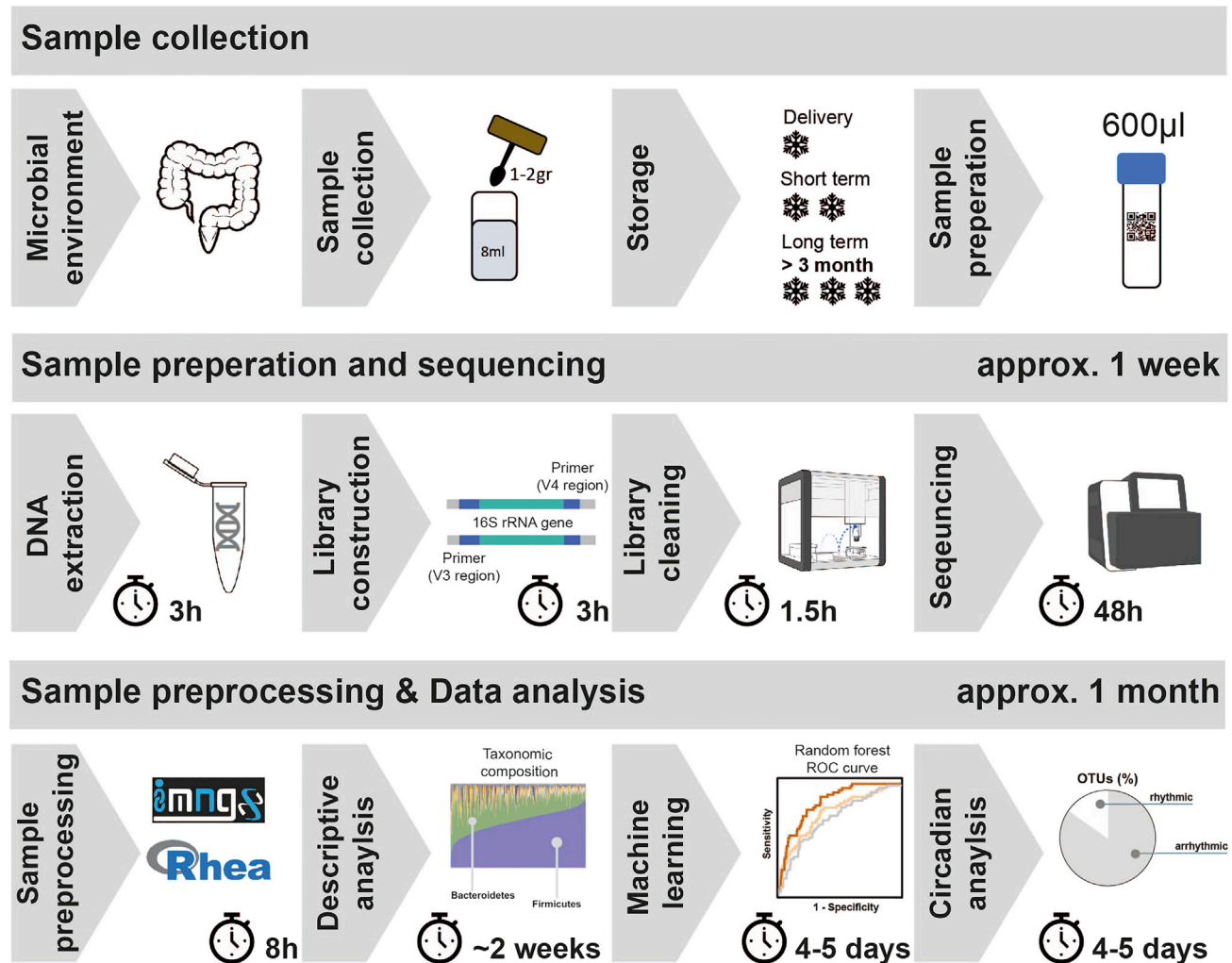


Figure 2. .Workflow of 16S rRNA Gene Sequencing Preparation and Analysis

Steps are structured into three sections: the sample collection and storage, the sample preparation and sequencing, and the sample preprocessing and data analysis. The given time for each step can be seen as a point of reference.

- 500 µL of the supernatant is transferred to a new 2-mL tube.
- 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).

- Add 1500 µL Binding Buffer and vortex for 5 s.
- 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.
- 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.
- Dry the silica membrane by centrifuging the columns for 1 min (11,000 × g) and discard the collection tube.
- 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

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

PCR Cycling Conditions			
Steps	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	∞	

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 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

Table 4. Primer selection for 2nd PCR

Forward primer	341-ovh-HTS-SB501-508
	341-ovh-HTS-SA502-509
	341-ovh-HTS-SD501, 502, 505, 508
	341-ovh-HTS-SC502, 505, 507, 508
	341-ovh-HTS-i5_1-16
Reverse primer	785r-ovh-HTS-SA701-712
	785r-ovh-HTS-SB701-711
	785r-ovh-HTS-SC701, 703, 704, 706-7012
	785r-ovh-HTS-SD703, 705-712
	785r-ovh-HTS-i7_02-06, 08-12, 15-18, 20-24

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:

$$\text{concentration in nM} = \frac{(\text{concentration in ng/}\mu\text{l}) \times 10^6}{(\text{average library size in bp} * 660 \text{ g/mol})}$$

		Qubit-1 (ng/µL)	Qubit-2 (ng/µL)	Qubit-3 (ng/µL)	Qubit-4 (ng/µL)	Mean (ng/µL)	nM
Final Pool	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 × 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 × 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.

Table 5. Settings for 2nd PCR. Rows in gray are performed for ten cycles.

PCR Cycling Conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	98	30 s	1
Denaturation	98	5 s	10
Annealing	55	10 s	
Extension	72	10 s	
Final Extension	72	2 min	1
Hold	10	∞	

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 μ L 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 ([Lagkourdos 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 ([Lagkourdos 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 *lm* from the R package *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 *randomForest* from the R package *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 *rfcv* from R package *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.

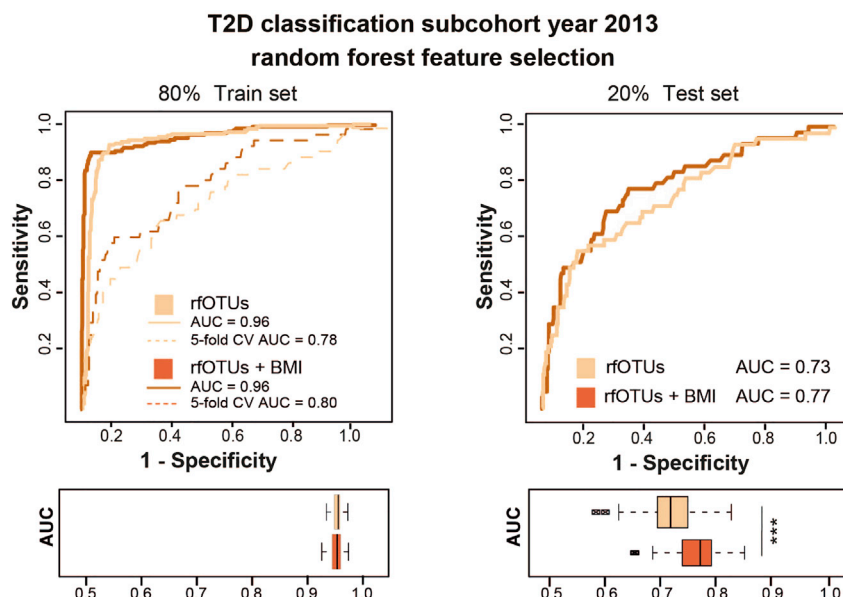


Figure 3. Random Forest Model for T2D Classification

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).

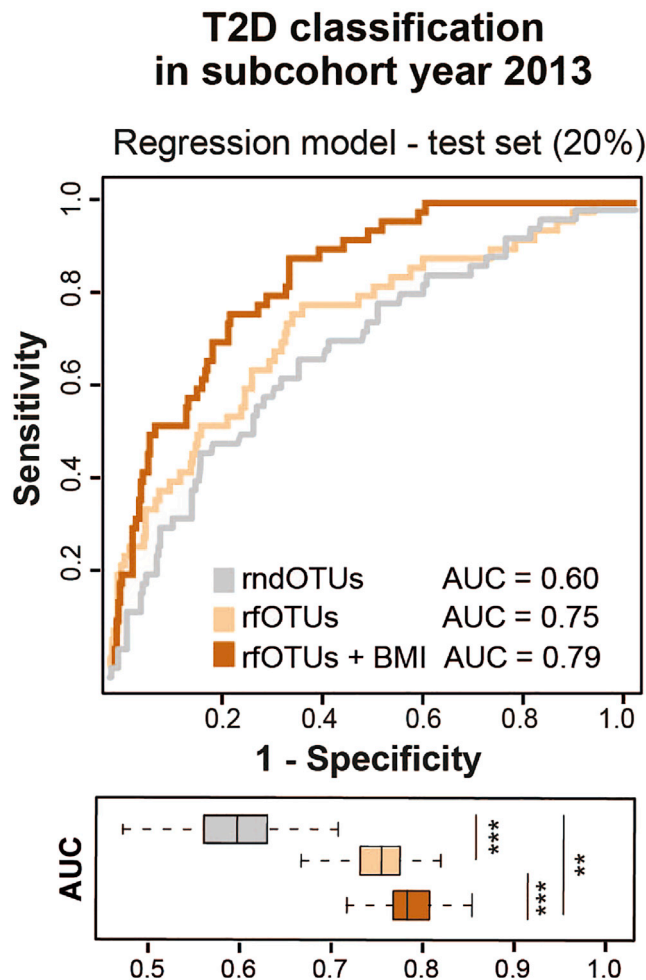


Figure 4. .Generalized Linear Model

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.

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
...

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):

$$y = baseline + amplitude * \cos\left(2 * \pi \left(x - \frac{phase\ shift}{24}\right)\right)$$

or a double harmonic cosine wave equation:

$$y = baseline + \left(amplitude\ A * \cos\left(2 * \pi \left(\frac{x - phase\ shift\ A}{24}\right)\right)\right) + \left(amplitude\ B * \cos\left(4 * \pi \left(\frac{x - phase\ shift\ B}{24}\right)\right)\right)$$

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

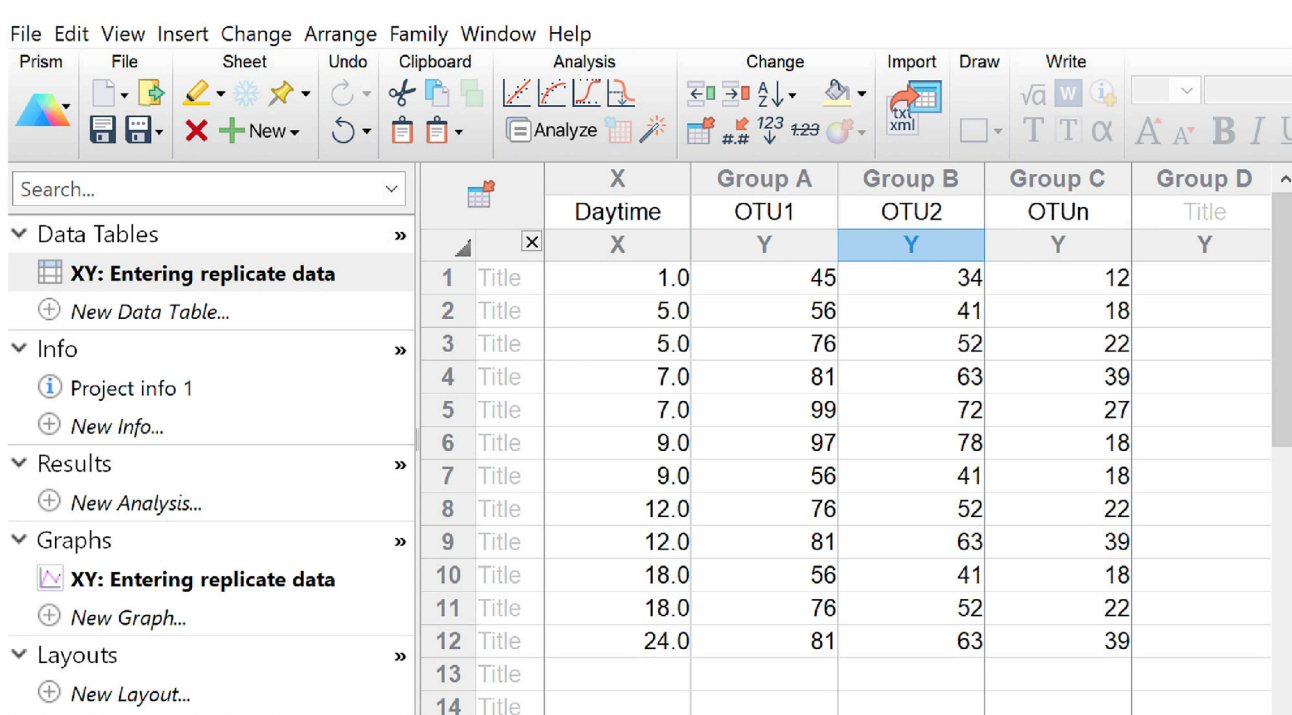


Figure 5. .OTU Table in GraphPad

The Excel sheet is transferred in an XZ/Sheet in GraphPad Prism for further analysis.

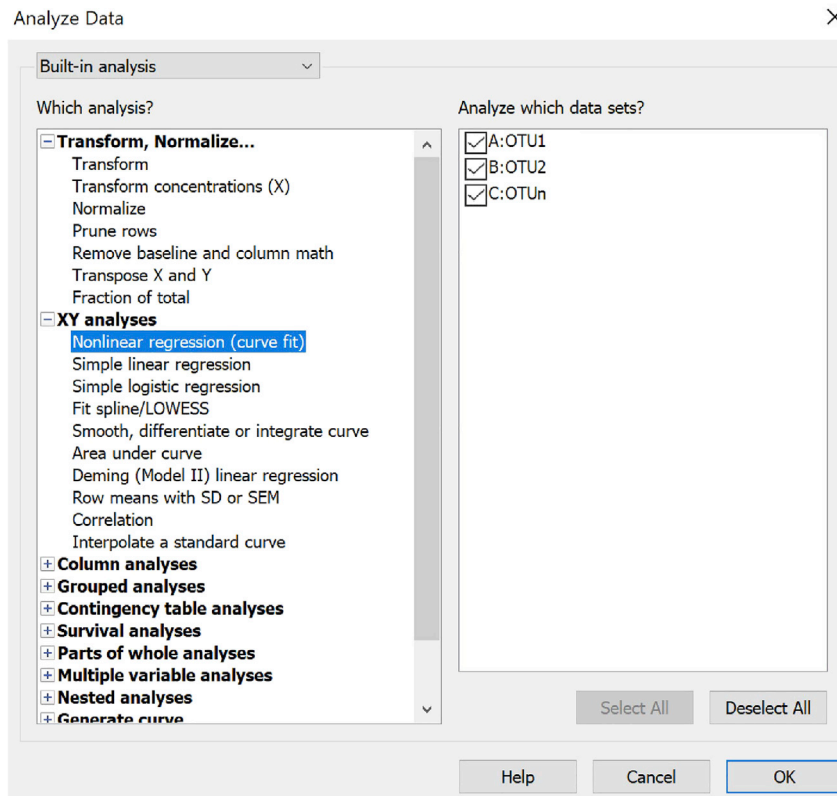


Figure 6. .OTU Nonlinear Regression Analysis in GraphPad

- 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 $p \leq 0.05$.
- 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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Daytime	1	1	1	1	1	1	4	4	4	4	4	4	7	7	7	n
2	OTU_1	0.560717	3.737978	2.490024	2.290998	2.454	0.786036	1.126928	4.50975	10.55799	4.335034	1.559364	1.566732	5.642944	7.376644	14.89387	8.171206
3	OTU_2	0.321588	0.673787	0.352384	0.426985	0.740908	1.106808	0.551382	0.637649	0.928894	0.691859	0.563155	0.600075	3.247811	1.29978	1.370805	1.757331
4	OTU_3	4.515969	6.153545	6.326573	5.421243	8.831536	7.429777	3.89482	4.682254	4.507098	3.293799	2.940302	2.979359	1.74099	2.265074	1.636428	1.469952
5	OTU_4	0.005497	0.081078	0.016336	0.005889	0.002047	0.028898	0	0.613006	0.281285	0.059762	0.003717	0.049033	0.040725	0.938356	1.422981	0.623077
6	OTU_5	1.060964	0.002796	0.39439	0.444654	2.908369	1.231072	1.29388	0	2.525021	0.507976	2.237752	0.768189	11.28589	0.040408	3.853907	1.380941
7	OTU_6	0.324336	1.685864	0.275373	0.170794	0.777748	0.968096	0.68758	2.646089	0.95506	0.273525	0.624489	0.434295	2.558033	2.675886	0.934424	0.640879
8	OTU_7	0.101699	0.430552	0.193694	0.173739	0.270165	0.410357	0.140591	0.264917	0.500425	0.404542	0.044606	0.231157	0.203625	0.25367	0.208704	0.432339
9	OTU_8	0	2.21427	0.004667	0	0.083915	0.026009	0.002197	0.57604	0.009812	0.002299	0.005576	0.464649	0	0.52081	0.104352	0.010173
10	OTU_9	0.590952	2.191903	0.557747	0.176684	2.214536	2.1616	1.245552	1.925269	1.749853	0.429826	1.408817	1.246848	6.098554	2.801598	2.670461	0.98675
11	OTU_10	0.997746	0.318721	1.577559	0.471156	4.375857	4.620853	0.89627	0.033885	0.147184	0.353974	7.545907	5.844307	0.142537	0.07857	0.189731	0.040691
12	OTU_11	0.046726	0.162156	0.035005	0.026503	0.182157	0.17339	0.125214	0.160182	0.143913	0.025284	0.113374	0.091062	0.468336	0.260405	0.215819	0.10427
13	OTU_12	1.36606	1.780921	1.197172	0.883418	1.784727	2.641313	1.383946	1.047346	0.627985	0.462005	1.183927	1.800224	1.356648	0.974274	0.322542	0.223799
14	OTU_13	0.51399	0	0.002334	0	1.966884	0.627095	0.487676	0	0.006542	0	0.996208	0.408611	0.987579	0.013469	0	0
15	OTU_14	0.291353	0.617871	0.163357	0	0.583311	0.468154	0.314134	1.349228	0.575653	0	0.544569	0.165779	2.026064	1.750999	0.747065	0
16	OTU_15	0.45627	0.005592	0.186693	0.070673	2.576803	0.794706	0.773252	0	0.73919	0.11033	1.353059	0.375922	5.352779	0.00449	0.9866	0.261947
17	OTU_16	0.665164	0.05312	0.532076	0.562443	1.643505	1.505606	0.858925	0.237193	1.076078	1.133177	0.966471	0.828897	3.907405	0.305302	1.413495	2.344803
18	OTU_17	0	0.732498	0	0	0.501443	0.00578	0.002197	1.666513	0	0	0.193294	0.315214	0	1.360392	0.007115	0
19	OTU_18	0.065967	0.114628	0.056008	0.008834	0.114616	0.124263	0.092263	0.101654	0.147184	0.032179	0.091071	0.058373	0.391977	0.215508	0.154156	0.05595
20	OTU_19	0.239129	1.478976	0.945135	1.319238	0.501443	0.826494	0.380036	3.462403	3.578204	2.482416	0.267638	0.688802	2.397679	4.004849	5.160678	5.788256
21	OTU_20	0.002749	0.10624	0.023337	0.008834	0.026607	0.020229	0.008787	0.135539	0.13083	0.101135	0.014869	0.007005	0.012727	0.118978	0.109095	0.116986
22	OTU_21	0.021989	0	0.018669	0.008834	0.251745	0.054907	0.079083	0	0.068686	0.009194	0.159839	0.037359	0.427611	0.002245	0.085379	0.027975
23	OTU_22	1.511737	0.749273	1.796924	1.590153	2.756913	2.675991	1.168666	0.323445	1.200366	1.342344	0.858672	1.316895	0.822134	0.298568	0.367603	0.455227
24	OTU_n	2.941015	1.274883	1.850599	0.126623	2.214536	2.3321	1.673916	1.518652	0.317263	0.01609	6.070181	4.086112	0.404704	0.397342	0.182616	0.002543

Figure 7. Transposed OTU Table

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 Westermarck, 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 consider 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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	Names	BH.Q	ADJ.P	PER	LAG	AMP	X1	X1.1	X1.2	X1.3	X1.4	X1.5	X4	X4.1	X4.2	X4.3	X4.4	
2	OTU_55	0.001026	7.13E-05	24		0	0.146802	0.41504	0.248826	1.470211	0.59778	0.454369	0.419027	0.713941	0.138619	0.225682	0.491886	0.799197
3	OTU_37	0.00108	7.83E-05	24	12	0.506144	0.087956	0.603892	0.333714	0.382815	0.765468	0.190729	0.318527	0.902566	1.540525	1.710109	0.353134	
4	OTU_237	0.001137	8.59E-05	24	21	0.021994	0.057721	0.100649	0.023337	0.147236	0.038887	0.060687	0.048328	0.030804	0.003271	0.034478	0.046465	
5	OTU_176	0.001315	0.000103	24	0	0.033344	0.115442	0.064303	0.09568	0.10601	0.055261	0.106924	0.09446	0.036965	0.003271	0.041374	0.078061	
6	OTU_20	0.001466	0.000124	24	22.5	0.496222	1.511737	0.749273	1.796924	1.590153	2.756913	2.675991	1.168666	0.323445	1.200366	1.342344	0.858672	
7	OTU_132	0.011374	0.001924	24	21	0.071446	0.164917	0.125811	0.555413	0.574222	0.112569	0.18495	0.197707	0.098574	0.134101	0.402243	0.124526	
8	OTU_12	0.034674	0.008066	24	22.5	0.324536	1.36606	1.780921	1.197172	0.883418	1.784727	2.641313	1.383946	1.047346	0.627985	0.462005	1.183927	
9	OTU_1007	0.040809	0.009863	24	10.5	0.006694	0.002749	0.013979	0.014002	0.005889	0.002047	0.00867	0.004393	0.024643	0.045791	0.020687	0.007434	
10	OTU_416	0.042541	0.010539	24	0	0.046444	0.164917	0.072691	0.231033	0.21791	0.130989	0.161831	0.186723	0.027724	0.04252	0.07815	0.226749	
11	OTU_209	0.042541	0.010539	24	22.5	0.041665	0.085207	0.081078	0.151688	1.092494	0.104382	0.106924	0.063705	0.055448	0.104664	0.34478	0.072485	
12	OTU_143	0.084029	0.026402	24	22.5	0.052669	0	0.223664	0.371053	0.391649	0.081868	0.072246	0	0.400456	0.049061	0.096538	0.327113	
13	OTU_156	0.08494	0.027201	24	22.5	0.015939	0.049475	0.011183	0.277706	0.032392	0.053214	0.00289	0.043935	0.006161	0.101393	0.004597	0.078061	
14	OTU_281	0.08494	0.027201	24	22.5	0.011751	0.024738	0.013979	0.126018	0.170794	0.01228	0.037568	0.019771	0.009241	0.032708	0.160897	0.022303	
15	OTU_99	0.08937	0.02889	24	15	0.046933	0.104447	0.355066	0.123684	0.167849	0.376594	0.121373	0.160362	0.280319	0.300909	0.248242	0.144971	
16	OTU_11	0.093141	0.030672	24	12	0.852279	0.590952	2.191903	0.557747	0.176684	2.214536	2.1616	1.245552	1.925269	1.749853	0.429826	1.408817	
17	OTU_1522	0.093141	0.030672	24	0	0.017823	0.09895	0.016775	0.04434	0.126623	0.038887	0.132933	0.079083	0.006161	0.003271	0.105733	0.031596	
18	OTU_1262	0.120239	0.043591	24	0	0.02689	0.049475	0.114628	0.191361	0.032392	0.038887	0.144492	0.20869	0.107815	0.029437	0.011493	0.059475	
19	OTU_1797	0.120239	0.043591	24	0	0.021564	0.131933	0.142586	0.046673	0.020613	0.077775	0.020229	0.116427	0.021563	0.068686	0.013791	0.10594	
20	OTU_146	0.125245	0.046163	24	0	0.058265	0.563465	0.128607	0.301043	0.05595	0.190344	0.288984	0.705154	0.095493	0.032708	0.011493	0.237901	
21	OTU_402	0.125245	0.046163	24	19.5	0.006753	0	0	0.02567	0.014724	0.014327	0.040458	0	0	0.003271	0.01609	0.011152	
22	OTU_108	0.131509	0.048869	24	22.5	0.043684	0.087956	0.142586	0.247369	0.577167	0.096195	0.089585	0.076886	0.150941	0.124289	0.436721	0.128243	
23	OTU_174	0.138046	0.051715	24	21	0.023297	0.382057	0.011183	0.102681	0.058895	0.040934	0.332332	0.316331	0	0.016354	0.029881	0.029738	
24	OTU_21	0.149031	0.056281	24	0	0.436921	2.941015	1.274883	1.850599	0.126623	2.214536	2.3321	1.673916	1.518652	0.317263	0.01609	6.070181	

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).

- 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 nonT2D 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).

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.

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

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 ± 5 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 \pm standard deviation (SD) and sample size (n) calculated, e.g., in Microsoft Excel.

c. As described in paragraph 4a, 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 8. OTU Table Group B

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	0.007819	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	0.02134	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	0.017795	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	1.134686	0.421816	0.11389	0.177163	0.054836
5	0.549429	0.759849	0	0.027277	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.

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).

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

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

Interval	Group A			Group B		
	Average	SD	n	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

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.

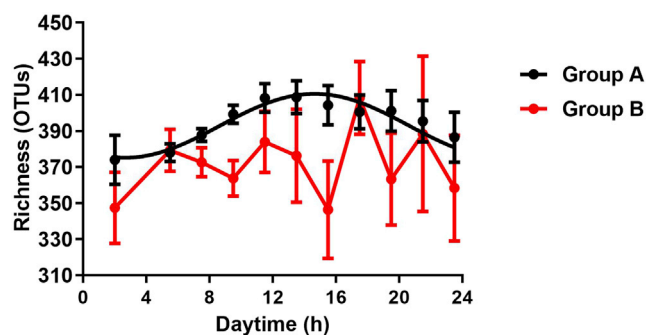


Figure 9. Illustration of Cosine-Wave Regression

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 μ L.

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.

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., Klaenhammer, 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.
- Dominianni, C., Wu, J., Hayes, R.B., and Ahn, J. (2014). Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiol.* *14*, 103.
- Franzosa, E.A., McIver, L.J., Rahnava, G., Thompson, L.R., Schirmer, M., Weingart, 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., Habouzit, F., and Moletta, R. (1997). Molecular microbial diversity of an anaerobic digester 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). Evolviz v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* *44*, W236–W241.
- Hughes, M.E., DiTacchio, L., Hayes, K.R., Vollmers, C., Pulivarthy, S., Baggs, J.E., Panda, S., and Hogenesch, J.B. (2009). Harmonics of circadian gene transcription in mammals. *PLoS Genet.* *5*, e1000442.
- Hughes, M.E., Hogenesch, J.B., and Kornacker, K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. Biol. Rhythms* *25*, 372–380.
- Ilett, E.E., Jorgensen, M., Noguera-Julian, M., Daugaard, G., Murray, D.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.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* *28*, 27–30.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* *41*, e1.
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* *79*, 5112–5120.
- Lagkouvardos, I., Fischer, S., Kumar, N., and Clavel, T. (2017). Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ* *5*, e2836.
- Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., and Clavel, T. (2016). IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci. Rep.* *6*, 33721.
- Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News* *2*, 18–22.
- Lomb, N.R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* *39*, 447–462.
- Luck, S., Thurley, K., Thaben, P.F., and Westermark, P.O. (2014). Rhythmic degradation explains and unifies circadian transcriptome and proteome data. *Cell Rep.* *9*, 741–751.
- Reitmeier, S., Kiessling, S., Clavel, T., List, M., Almeida, E.L., Ghosh, T.S., Neuhaus, K., Grallert, H., Linseisen, J., Skurk, T., et al. (2020). Arrhythmic gut microbiome signatures predict risk of type 2 diabetes. *Cell Host Microbe* *28*, 258–272.e6.
- Revelle, W. (2020). psych: Procedures for Psychological, Psychometric, and Personality Research (Northwestern University), R package version 2.0.7.
- Segata, N., Bornigen, D., Morgan, X.C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* *4*, 2304.
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (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). Latitude in sample handling and storage for infant faecal microbiota studies: the elephant in the room? *Microbiome* *4*, 40.
- Thaben, P.F., and Westermark, P.O. (2014). Detecting rhythms in time series with RAIN. *J. Biol. Rhythms* *29*, 391–400.
- Thaben, P.F., and Westermark, P.O. (2016). Differential rhythmicity: detecting altered rhythmicity in biological data. *Bioinformatics* *32*, 2800–2808.
- Wu, G., Anafi, R.C., Hughes, M.E., Kornacker, K., and Hogenesch, J.B. (2016). MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics* *32*, 3351–3353.
- Yang, R., and Su, Z. (2010). Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics* *26*, i168–i174.
- Yoon, S.H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., and 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.