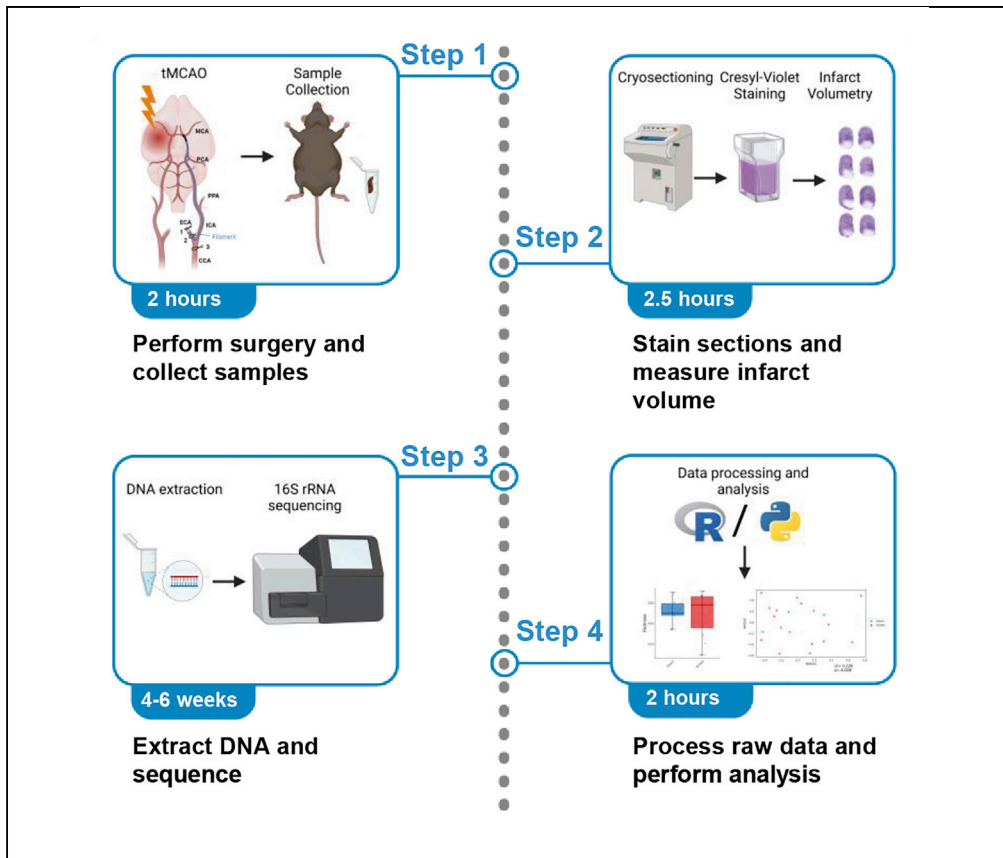


## Protocol

# Protocol for microbiota analysis of a murine stroke model



Adam Sorbie, Rosa Delgado Jiménez, Monica Weiler, Corinne Benakis

adam.sorbie@med.uni-muenchen.de

### Highlights

Detailed instructions and walkthrough of the tMCAO murine stroke model

Simple and reproducible infarct volumetry and visualization

Sample collection and DNA extraction for 16S-based microbiota analysis

Easy-to-use, bioinformatic pipeline, with a choice of programming language (R/Python).

Investigations on the microbiota in neurological diseases such as stroke are increasingly common; however, stroke researchers may have limited experience with designing such studies. Here, we describe a protocol to conduct a stroke microbiota study in mice, from experimental stroke surgery and sample collection to data analysis. We provide details on sample processing and sequencing and provide a reproducible data analysis pipeline. In doing so, we hope to enable researchers to conduct robust studies and facilitate identification of stroke-associated microbial signatures.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Sorbie et al., STAR Protocols 4, 101969  
March 17, 2023 © 2022 The Author(s).  
<https://doi.org/10.1016/j.xpro.2022.101969>



## Protocol

## Protocol for microbiota analysis of a murine stroke model

Adam Sorbie,<sup>1,2,3,\*</sup> Rosa Delgado Jiménez,<sup>1</sup> Monica Weiler,<sup>1</sup> and Corinne Benakis<sup>1</sup><sup>1</sup>Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany<sup>2</sup>Technical contact<sup>3</sup>Lead contact\*Correspondence: [adam.sorbie@med.uni-muenchen.de](mailto:adam.sorbie@med.uni-muenchen.de)  
<https://doi.org/10.1016/j.xpro.2022.101969>**SUMMARY**

Investigations on the microbiota in neurological diseases such as stroke are increasingly common; however, stroke researchers may have limited experience with designing such studies. Here, we describe a protocol to conduct a stroke microbiota study in mice, from experimental stroke surgery and sample collection to data analysis. We provide details on sample processing and sequencing and provide a reproducible data analysis pipeline. In doing so, we hope to enable researchers to conduct robust studies and facilitate identification of stroke-associated microbial signatures.

For complete details on the use and execution of this protocol, please refer to Sorbie et al. (2022).<sup>1</sup>

**BEFORE YOU BEGIN**

The below protocol describes the specific steps for performing transient middle cerebral artery occlusion in 8- to 12-week-old male C57BL/6 mice, and subsequent 16S rRNA sequencing and analysis. However, the experimental stroke and 16S rRNA sections are generalizable and could be applied to other research questions or experimental designs. For example, our extraction protocol is best suited for stool samples, however with minor modifications it could also be used to process cecal or ileal content. Similarly, mice of different age or sex could be used to answer different research questions. Importantly, we provide here a detailed bioinformatic pipeline for processing 16S rRNA sequencing data (trimming, filtering, chimera removal and taxonomic assignment), as well as statistical analysis (alpha and beta-diversity and differential abundance analysis) allowing investigators to conduct and understand the process themselves, facilitating reproducibility and open science.

**Institutional permissions**

For mouse experiments, researchers are required to obtain permission from the relevant authorities before beginning. In this study, experimental protocols were approved by the responsible governmental department (Regierung von Oberbayern, Munich, Germany).

**Transient middle cerebral artery occlusion (tMCAO)**

⌚ Timing: approximately 2 weeks (acclimatization: 1 week, stroke surgery: 1 h 30 min, 3 days recovery and sample collection/processing: 1 day)

1. Conduct a power-analysis or consult a statistician to determine an appropriate sample-size for your study.



**Note:** Pilot or previous data from the investigators' own laboratory are preferable for determining an appropriate sample size, however we acknowledge that this is not always possible and would recommend making use of our dataset for this purpose. Processed data is hosted here: <https://zenodo.org/record/6821707>.

2. Obtain male mice, approximately 8–12 weeks of age.
  - a. Ensure mice are from the same vendor/facility as this can impact microbiota composition and stroke outcome.<sup>2</sup>
    - i. Mice purchased externally should be allowed to acclimatize to their new environment for at least 1 week to reduce animal stress and stabilize microbiota composition.<sup>3</sup>
    - ii. Groups should be randomized and mice from both groups (sham and stroke) cohoused until surgery.
3. Prepare materials for transient middle cerebral artery occlusion:
  - a. Sterilize all surgical instruments using a bead sterilizer.
  - b. Prepare analgesic and anesthetic:
    - i. Analgesic – 4 mg/kg Carprofen with 0.1 mg/kg Buprenorphine.
    - ii. Anesthetic – medical air (78% nitrogen and 21% oxygen) with a concentration of up to 3% by volume isoflurane in medical air.
  - c. Prepare solutions for surgery:
    - i. 70% ethanol solution to be used as an antiseptic.
    - ii. Dexpanthenol (Bepanthen, Bayer) eye ointment.
    - iii. 0.9% saline solution to hydrate site of incision.
  - d. Turn on and set heat blanket to 37°C.
  - e. Turn on the laser Doppler device (PeriFlux System 5000).
4. (On the day of sample collection) Prepare solutions for sacrifice/sample collection:
  - a. PBS-heparin (2 mg/mL).
  - b. Ketamine (120 mg/kg) and Xylazine (16 mg/kg).

**Note:** The protocols described here (tMCAO to microbiota sequencing) have been validated in 8- to 12-week-old male C57BL/6 mice in the acute phase response to stroke (3 days). They can, however, be used in other mouse models, in female mice, with different ages or strains or at different time points after stroke with minor adaptations. Estrogen and estradiol have been shown to be neuroprotective after tMCAO,<sup>4</sup> aged mice have worse stroke outcome and higher mortality,<sup>5</sup> different mouse strains<sup>6</sup> and C57BL/6 mice from different commercial breeders<sup>2</sup> have different susceptibility to stroke and a different microbiota composition, respectively. Therefore, the timing of the MCA occlusion may need to be adjusted according to the model and experimental question. There is clear evidence that microbiota dysbiosis occurs between 24 and 72 h after stroke in rodents. However, it is not well defined how quickly changes in microbiota composition take place and how long they persist after stroke.<sup>7–10</sup> In addition, it is not well known whether the microbiota is altered in female mice subjected to stroke. Hence, although the pipeline for microbiota analysis remains the same as presented here for these different mouse models, microbiota changes after stroke may be different.

### Brain coronal sectioning

⌚ Timing: 10 min for transfer and labelling

5. Transfer the brains stored at –80°C to –20°C, the day before cutting and label microscope slides.

### Cresyl-violet staining solution and ethanol solutions

⌚ Timing: 30 min to prepare solutions

6. Prepare cresyl-violet staining solutions:
  - a. To 500 mL of warm distilled water, add 2.5 g cresyl-violet acetate and 5 mL 100% glacial acetic acid.
  - b. Filter solution before use.
  - c. Pre-warm cresyl-violet solution to 65°C.
7. Prepare a series of ethanol solution: 50%, 70%, 95% and 100%.

### Infarct volumetry

⌚ Timing: 15 min for image acquisition

8. Obtain high quality digital images of slides using a dedicated slide scanner or high-resolution copier.

### DNA extraction from fecal samples & sequencing

⌚ Timing: approximately 1 week (accounting for the time to find a sequencing vendor and receive a quote)

9. Choose an appropriate DNA extraction kit and PCR/sequencing protocol.
10. Arrange sequencing in-house if possible or find a reputable vendor/core facility which offers 16S rRNA sequencing services.
  - a. In-house protocols using Illumina MiSeq™ and HiSeq™ machines are compatible with this protocol without any adjustments.
  - b. A non-exhaustive list of reputable and competitively priced commercial vendors, organized by region can be found [here](#).

**Note:** The choice of DNA extraction kit and protocol is mostly dependent on user choice and availability; however, it is important that any kit/protocol includes a mechanical lysis step.<sup>11</sup> In our laboratory we use a QIAGEN DNeasy® PowerLyzer PowerSoil® Kit. Check the lysis protocol on your tissue homogenizer before starting. Refer to the manufacturer's instructions for the correct settings. We use a QIAGEN TissueLyser LT set to 50 oscillations/sec for 6 min for stool samples. While arranging sequencing, ensure all samples are sequenced by the same vendor/core facility, using the same hypervariable region and platform and in as few batches as possible. Choice of hypervariable region is also mostly based on researcher preference, however V1/V2, V3, V3/V4 and V4 are amongst the most widely used.<sup>12</sup>

### Software installation

⌚ Timing: Approximately 15 min to install all required software, depending on internet connection speed

#### ImageJ

11. Download and install the correct version of [ImageJ](#) for your platform:

#### Windows subsystem for Linux

12. Install Windows subsystem for Linux, version 2 (WSL2).
13. Run the following command in a PowerShell terminal/command prompt as administrator: `wsl – install`.
14. See the official Microsoft guide for further details if necessary: <https://docs.microsoft.com/en-us/windows/wsl/install>.

*R/R packages*

These steps are specific for the R-based pipeline.

15. Download and install R and RStudio (see Table below).

Platform	macOS	Windows (windows subsystem for Linux 2)/Linux
R	<ol style="list-style-type: none"> <li>1. Install homebrew by typing the following command into a terminal and provide password if required:  <code>/bin/bash -c `\$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)`</code></li> <li>2. Install R using homebrew b: <code>brew install --cask r</code></li> </ol>	<ol style="list-style-type: none"> <li>1. R can be installed using the bash script provided. Type the following into a terminal: <code>sudo bash install_R_linux.sh</code></li> </ol>
Rstudio	<ol style="list-style-type: none"> <li>1. Install Rstudio using homebrew. Type the below command into a terminal window: <code>brew install --cask rstudio</code></li> </ol>	<ol style="list-style-type: none"> <li>1. Rstudio is also installed using the provided bash script. Type the below command into a terminal window: <code>sudo bash install_Rstudio.sh</code></li> </ol>

**Note:** The instructions for WSL2 and Linux assume you are using Ubuntu as your Linux distribution.

16. Install pipeline dependencies:

- a. Clone git [repository](#) which contains the processing and analysis scripts.
  - i. This will create a directory called "Stroke\_Microbiota\_reproducibility" which contains all of the scripts, tutorials, and example data.
- b. Install extra software tools (mafft and fasttree) used in pipeline.

macOS	Windows (windows subsystem for Linux 2)/Linux
<pre># install git if not already installed brew install git # clone repository git clone https://github.com/adamsorbie/Stroke_Microbiota_reproducibility.git # install dependencies for phylogenetic tree construction brew install mafft # fasttree not supported by brew so # install with conda in next step</pre>	<pre># install git if not already installed sudo apt update sudo apt upgrade sudo apt install git # clone repository git clone https://github.com/adamsorbie/Stroke_Microbiota_reproducibility.git # install dependencies for phylogenetic tree construction sudo apt install mafft sudo apt install mafft</pre>

- c. Install Miniconda to facilitate installation of additional dependencies:

macOS	Windows (windows subsystem for Linux 2)/Linux
<pre># install with brew brew install miniconda # dependencies for read QC and fasttree conda install -c bioconda fastqc multiqc fasttree</pre>	<pre># Download installer wget https://repo.anaconda.com/miniconda/Miniconda3-py38_4.10.3-Linux-x86_64.sh # run installation script and follow on-screen instructions bash Miniconda3-py38_4.10.3-Linux-x86_64.sh # cleanup rm Miniconda3-py38_4.10.3-Linux-x86_64.sh # install dependencies for read QC conda install -c bioconda fastqc multiqc</pre>

**Note:** The packages installed above are necessary to perform quality control of reads and for alignment of sequences and subsequent phylogenetic tree construction.

### QIIME2 installation

For those who would prefer to use the QIIME2/python-based pipeline, follow the steps below to install the necessary tools.<sup>13</sup>

17. Install Miniconda as above.
18. Run the following commands to install QIIME2.

macOS	Windows (windows subsystem for Linux 2)/Linux
<pre># Download installation template # Replace qiime2-2022.8 with desired version. 2022.8 is current as of 21.10.22 wget https://data.qiime2.org/distro/ core/qiime2-2022.8-py38-osx-conda.yml # create new conda environment conda env create -n qiime2-2022.8 -file qiime2-2022.8-py38-osx-conda.yml # remove installation template rm qiime2-2022.8-py38-osx-conda.yml # add short alias to activate qiime environment quickly echo "alias q2='conda activate qiime2-2022.8'" &gt;&gt; ~/.zshrc &amp;&amp; source ~/.zshrc</pre>	<pre># Download installation template # Replace qiime2-2022.8 with desired version. 2022.8 is current as of 21.10.22 wget https://data.qiime2.org/distro/core/qiime2-2022. 8-py38-linux-conda.yml # create new conda environment conda env create -n qiime2-2022.8 -file qiime2-2022. 8-py38-linux-conda.yml # remove installation template rm qiime2-2022.8-py38-linux-conda.yml # add short alias to activate qiime environment quickly echo "alias q2='conda activate qiime2-2022.8'" &gt;&gt; ~/.bashrc &amp;&amp; source ~/.bashrc</pre>

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, peptides, and recombinant proteins</b>		
Dexpanthenol (Bepanthen)	Bayer	81552983
Cresyl-violet acetate	Sigma-Aldrich	C5042-10G
Acetic acid 100%	Roth	3738.1
Eukitt Quick-hardening mounting medium	Sigma-Aldrich	03989
Isoflurane 1000 mg/mL	Dechra	ATC: QN01AB06
Ketamine 100 mg/mL	bela-pharm	ATC: QN01AX03
Xylazine 20 mg/mL	Serumwerk	ATC: QN05CM92
Carprofen 50 mg/mL	Zoetis	ATC: QM01AE91
Buprenorphine 0.3 mg/mL	Indivior	ATC: N02AE01
O.C.T. Compound (TissueTek)	Sakura	Ref. 4583
<b>Critical commercial assays</b>		
DNeasy® PowerLyzer PowerSoil® Kit	QIAGEN	Cat. No. / ID: 12855-50
<b>Experimental models: Organisms/strains</b>		
C57BL6/J (male mice 6–10 weeks old)	Charles River Laboratories	000664
<b>Deposited data</b>		
Raw sequencing data	Sorbie et al. <sup>1</sup>	PRJEB48735
Processed data and metadata	Sorbie et al. <sup>1</sup>	<a href="https://github.com/adamsorbie/Stroke_Microbiota_reproducibility/tree/main/data">https://github.com/adamsorbie/Stroke_Microbiota_reproducibility/tree/main/data</a> <a href="https://zenodo.org/record/6821707">https://zenodo.org/record/6821707</a>
<b>Software and algorithms</b>		
Windows subsystem for Linux	Microsoft	<a href="https://docs.microsoft.com/en-us/windows/wsl/install">https://docs.microsoft.com/en-us/windows/wsl/install</a>
FastQC/MultiQC	Andrews/Ewels et al. <sup>14,15</sup>	<a href="https://github.com/s-andrews/FastQC">https://github.com/s-andrews/FastQC</a> <a href="https://multiqc.info">https://multiqc.info</a>
Cutadapt	Martin <sup>16</sup>	<a href="https://cutadapt.readthedocs.io/en/stable/">https://cutadapt.readthedocs.io/en/stable/</a>
R	The R foundation	<a href="https://www.r-project.org">https://www.r-project.org</a>
RStudio	RStudio	<a href="https://www.rstudio.com">https://www.rstudio.com</a>

(Continued on next page)

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
DADA2	Callahan et al. <sup>17</sup>	<a href="https://benjjneb.github.io/dada2/index.html">https://benjjneb.github.io/dada2/index.html</a>
QIIME2	Boleyn et al. <sup>13</sup>	<a href="https://qiime2.org">https://qiime2.org</a>
phyloseq	McMurdie and Holmes <sup>18</sup>	<a href="https://joey711.github.io/phyloseq/">https://joey711.github.io/phyloseq/</a>
DECIPHER - IDTAXA	Murali et al. <sup>19</sup>	<a href="http://bioconductor.org/packages/release/bioc/html/DECIPHER.html">http://bioconductor.org/packages/release/bioc/html/DECIPHER.html</a>
picante	Kembel et al. <sup>20</sup>	<a href="https://cran.r-project.org/web/packages/picante/index.html">https://cran.r-project.org/web/packages/picante/index.html</a>
phangorn	Schliep <sup>21</sup>	<a href="https://cran.r-project.org/web/packages/phangorn/index.html">https://cran.r-project.org/web/packages/phangorn/index.html</a>
vegan	Oksanen et al. <sup>22</sup>	<a href="https://cran.r-project.org/web/packages/vegan/index.html">https://cran.r-project.org/web/packages/vegan/index.html</a>
ANCOM-BC	Lin and Peddada <sup>23</sup>	R: <a href="https://www.bioconductor.org/packages/release/bioc/html/ANCOMBC.html">https://www.bioconductor.org/packages/release/bioc/html/ANCOMBC.html</a> QIIME2: <a href="https://github.com/mortonjt/q2-ancombc">https://github.com/mortonjt/q2-ancombc</a>
GUniFrac	Chen et al. <sup>24</sup>	<a href="https://cran.r-project.org/web/packages/GUniFrac/index.html">https://cran.r-project.org/web/packages/GUniFrac/index.html</a>
ImageJ	NIH, US	<a href="https://imagej.nih.gov/ij">https://imagej.nih.gov/ij</a>
BioRender	BioRender	<a href="https://biorender.com">https://biorender.com</a>
<i>Other</i>		
Chow	Ssniff	V1574-300
Monofilament silicon coated	Doccol	602223PK10Re
Surgical glue COSMOPLAST 515	Weiss Chemie	COSMO CA-500.130
Insta-Set accelerator	BSI	BSI-151
Superfrost® Plus Adhesion microscope slides	Menzel GmbH	J1800AMNZ
Forceps	Fine Science Tools	11253-25
Scissors	Fine Science Tools	14090-09
Laser Doppler - PeriFlux System 5000	Perimed	5000-PF5010
QIAGEN® TissueLyser LT	QIAGEN	85600
EpreDia™ CryoStar™ NX70	Thermo Scientific	957040

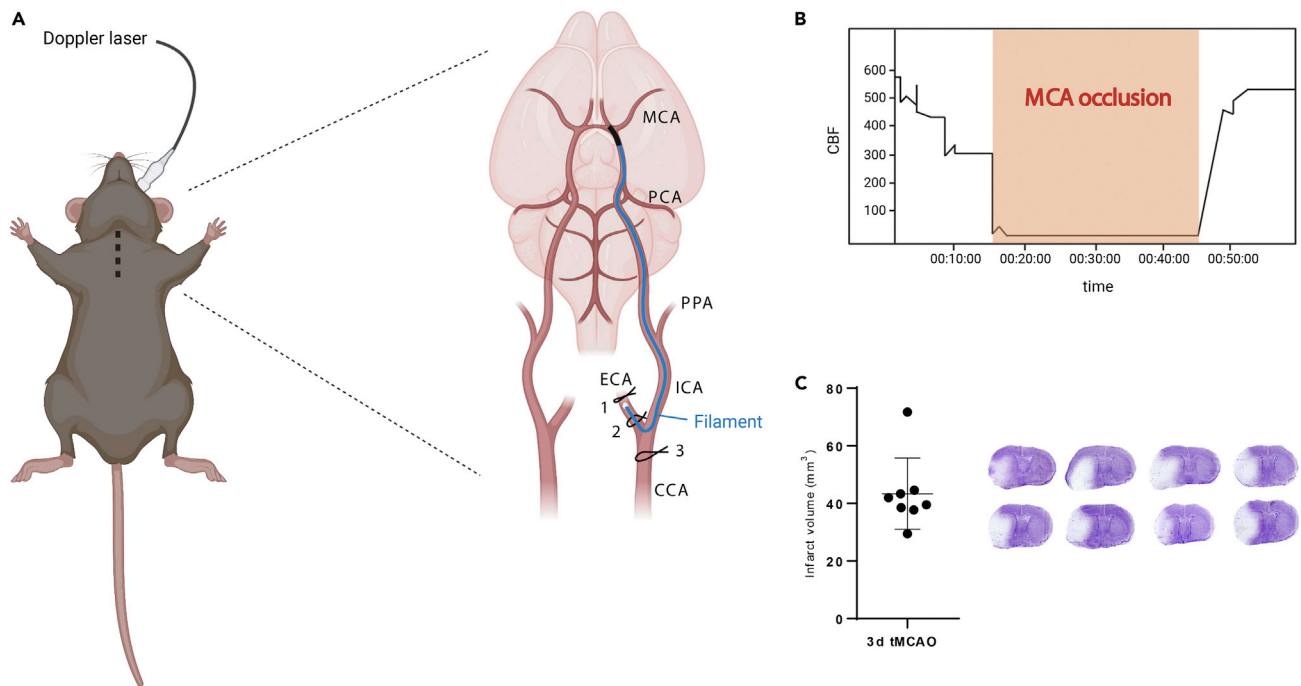
## STEP-BY-STEP METHOD DETAILS

### Transient middle cerebral artery occlusion (tMCAO)

⌚ Timing: 110 min total per mouse: 30 min for the induction of analgesia; 20 min for filament insertion and 15 min for removal, 45 min for the occlusion time

Transient middle cerebral artery occlusion (tMCAO) is used to model ischemic stroke in patients by temporary occlusion of the MCA using a filament inserted through the common carotid artery, as described previously<sup>25</sup> and in details below and [Figure 1](#). For our experiments, we perform 45 min of occlusion resulting in cortical and striatal lesions and 5%–10% mortality, as well as a clear shift in microbiota composition.<sup>9</sup>

1. Analgesia/Anesthesia:
  - a. Inject the animal with analgesic (4 mg/kg Carprofen and 0,1 mg/kg Buprenorphine) intraperitoneally, 30 min prior to surgery.
  - b. Anesthetize the animal with isoflurane.
    - i. Place the mouse in a chamber, maintaining isoflurane concentration at 3% for a further minute, before reducing to 2%.
2. Surgical preparation:



**Figure 1. Schematic representation of the surgical procedure of transient filament occlusion of the left middle cerebral artery**

(A) The filament is inserted between the first tight suture (1) and the second loose knot (2). The filament is redirected into the ICA and advanced until reaching the MCA origin.

(B) Laser Doppler flowmetry graph of blood flow before, during and after occlusion. An acute reduction in the blood flow is observed when the filament occludes the MCA origin. There is a restoration of the blood flow to steady-state level after filament removal.

(C) Infarct volume analysis and representative cresyl-violet staining of brain sections.

- a. Place the mouse in a prone position over a heating pad with the nose in the anesthetic cone (2% isoflurane)
- b. Monitor and maintain body temperature at 37°C with a rectal probe taped underneath the animal.

**Note:** Be careful not to insert the temperature monitoring probe beyond the rectum to avoid any damage of the GI tract and contamination of the animal's microbiota.

- c. To protect the ocular surface of the animal during surgery and prevent drying out, apply dexpanthenol eye-drops to both eyes.
  - d. Disinfect the area surrounding the left part of the animal's head with a 70% ethanol solution, and make a 1 cm long excision, revealing the skull underneath.
3. Laser doppler probe:
- a. Attach the laser Doppler probe to the left MCA territory to the parietal bone 2 mm posterior and 5 mm lateral to bregma with a drop of glue, apply a drop of accelerator to harden the glue and secure the probe. Wait 1 min until the probe is firmly attached.
    - i. Turn the animal over into a supine position and fix paws with tape.
    - ii. Fix the fiberoptic probe to the probe holder connected to the laser Doppler flowmeter to measure cortical perfusion (Periflux System 5010, Perimed, Sweden).
    - iii. Write down the value of the steady state cerebral blood flow.

**Note:** If the value on the laser Doppler flowmeter is < 100, carefully detach and re-position the probe, before taking another reading.



4. tMCAO – Arterial ligation:
  - a. Disinfect the area surrounding the neck and chest and make a 2 cm incision in the neck.
  - b. Using a pair of forceps, gently separate the skin and submandibular glands.
  - c. Fix the sternomastoid muscle using retractors allowing the common carotid artery (CCA) to be visualized underneath.
    - i. Dissect the external carotid artery (ECA) and make a permanent knot at the most distal visible section.
    - ii. Make another suture on the ECA, close to the bifurcation with the CCA, and tie an additional loose knot for later use.
    - iii. Free the CCA from the surrounding connective tissue and nerves and perform a transient ligation immediately before the bifurcation.
  - d. Dissect the internal carotid artery (ICA), removing surrounding fat and connective tissue and place a microvascular clip, 5 mm beyond the bifurcation, making sure not to damage the vagal nerve.
  - e. Using spring scissors, make a small incision in the ECA between the tight and the loose ligations, being careful not to cut through the ECA.
    - i. Introduce the filament into the lumen of the ECA and move it towards the CCA.
    - ii. Tighten the previously made loose ligation on the ECA to secure the filament in that position and avoid bleeding while removing the microvascular clip.
    - iii. Remove the microvascular clip and redirect the filament into the ICA until the origin of the MCA is reached. This is determined by the observation of a stark reduction (>80%) in cerebral blood flow.
    - iv. Secure the filament in this position by further tightening the loose ligation around the ECA.
5. Confirmation of occlusion:
  - a. Record laser Doppler values before (steady state) and after filament insertion (start of occlusion).
  - b. Remove the retractors and return the sternomastoid muscle and the submandibular glands to their original position before suturing the wound.
  - c. Detach the Doppler probe from the holder.
  - d. Place the animal in a recovery chamber at 37°C for 45 min (until filament removal).
6. tMCAO – Reperfusion:
  - a. 5 min before the filament is removed, anesthetize the animal in the anesthetic chamber with isoflurane as above.
  - b. Lay the animal in a prone position in the surgical area and place the snout into the anesthesia mask, maintaining isoflurane concentration at 2%.
    - i. Apply dexpanthenol eye-drops to both eyes and again fix the paws with tape.
  - c. Place the laser Doppler probe into the holder to allow reperfusion to be monitored.
  - d. Remove the wound suture, using forceps to pull the skin and the submandibular glands apart. Gently pull the sternomastoid muscle and expose the surgical field, using retractors.
  - e. Loosen the ECA suture securing the filament in place, and gently pull on the filament, being careful to avoid damage to the silicone-rubber coating during removal.
  - f. Tie the ECA suture tightly.
  - g. Confirm the increase in the cerebral blood flow in the laser Doppler device (>80% of the initial value before reperfusion).
    - i. Record laser Doppler values before and after filament removal.
  - h. Open the transient ligation before the bifurcation from the CCA.
7. Recovery:
  - a. Remove the retractor and relocate the sternomastoid muscle and the submandibular glands before suturing the neck and head wounds. Place the animal in a recovery chamber, maintained at a constant temperature of 37°C for 1 h to recover. Afterward, return the mice to their cages.
  - b. House stroke and sham mice separately, with a minimum of two mice and a maximum of 5 mice per cage, according to the regulations of the responsible governmental department.

△ **CRITICAL:** The total operation time should not exceed 20 min. Be careful not to disturb the vagal nerve and avoid excessive bleeding. If after several attempts cerebral blood flow is not reduced below 20%, the experiment should be terminated.

### 8. Sham Surgery.

- a. For sham surgery, conduct the procedures as above, including the ligation of arteries without inserting the filament, and keeping the animal under anesthesia for a similar amount of time as stroke mice (35 min in total). Sham mice are not awakened during this procedure.
- b. Suture the wound and place the animal in a chamber at 37°C for 1 h to recover from anesthesia.
- c. After recovery, return the mice to new cages in a temperature-controlled room.

**Note:** Sham mice should be housed in a different cage than stroke mice. Because sham surgery alone can influence the composition of the gut microbiota,<sup>9</sup> the use of naïve mice as a control group is not recommended, except as an additional control if desired/warranted by a different experimental design.

### 9. Aftercare:

- a. Provide analgesia every 12 h for 3 days post-surgery.
- b. Animals must be closely monitored after surgery. Record weight every day, monitor for signs of sickness such as poor grooming and test neurological deficit using the Neuroscore scale.<sup>25,26</sup> Perform the Neuroscore at the same time of the day and use surgical scrubs to maintain a neutral scent. Additionally, the same investigator should examine each set of animals. [Troubleshooting 1](#).

**Note:** After surgery mice may consume less food and liquid. This can be partially alleviated by moistening food pellets placed in the cage floor, providing hydrogel to the mice and by daily subcutaneous injections of saline, if needed. Food consumption and weight loss can also be monitored to control for any potential effects on microbiota composition.

### 10. Stool Sample collection, 3 days post-surgery.

- a. On day 3 collect a fecal sample in sterile Eppendorf tubes from each mouse in the experiment and immediately snap-freeze on dry ice. Store at –80°C until DNA extraction. [Troubleshooting 2](#).

### 11. Sample collection and intracardiac perfusion, 3 days post-surgery:

- a. Prepare a 20 mL syringe containing phosphate-buffered saline and heparin (2 U/mL) and fix 1 m above the bench to facilitate gravity-dependent perfusion.
  - i. Inject 100 µL of ketamine and xylazine (120 and 16 mg/kg body weight, respectively) intraperitoneally to euthanize the animal. After 5 min, confirm cessation of spontaneous body movement.
  - ii. Fix the animal in a supine position with tape and disinfect the abdomen with 70% ethanol.
  - iii. Make a 3 cm-long incision, cutting through the diaphragm, ribs, and sternum to visualize the heart.
  - iv. Make a small incision in the right atrium and insert a cannula into the left ventricle.
  - v. Perfuse with 20 mL of PBS-heparin.
- b. Decapitate the mouse and remove the entire brain for infarct volumetry.
  - i. Freeze brain on dry ice and store at –80°C until further use.

## Brain coronal sectioning

⌚ **Timing:** 30 min per brain

12. For each animal in the experiment, cut 20 µm brain coronal sections every 400 µm using a cryotome (CryoStar NX70, Thermo Scientific).

- a. Set the blade and specimen head at  $-20^{\circ}\text{C}$  (CryoStar NX70, Thermo Scientific)
  - b. Mount the brain on the specimen chuck using O.C.T.Compound (Tissue Tek, Sakura) and freeze using the quick-freeze function of the Cryobar.
  - c. Trim at  $50\ \mu\text{m}$  until the infarct area appears and cut  $20\ \mu\text{m}$  of brain coronal sections every  $400\ \mu\text{m}$  until the infarct area is not visible anymore.
13. Collect the brain sections on Superfrost Plus Adhesion (Menzel GmbH) microscope slides and store at  $-80^{\circ}\text{C}$  until further use.

### Cresyl-violet staining of brain sections

⌚ Timing: 1 h 30 min

Cresyl-violet staining is used here to stain serial sections of brain tissue from mice subjected to tMCAO, enabling measurement of the infarct size.<sup>27</sup>

14. Thaw the brain sections at room temperature for 30 min.
  - a. Place slides into a slide holder.
15. Immerse slides in a descending series of ethanol solutions to remove lipids and fixatives from the tissue:
  - a. 95% ethanol for 15 min.
  - b. 70% ethanol for 1 min.
  - c. 50% ethanol for 1 min.
16. Wash slides in distilled water twice, for 2 min and then for 1 min.
17. Stain with cresyl-violet solution (pre-heated to  $65^{\circ}\text{C}$ ) for 15 min.
18. Wash slides in distilled water twice, 1 min each.
19. Dehydrate sections before mounting:
  - a. 95% ethanol for 2 min.
  - b. 100% ethanol twice (5 min then 2 min).
  - c. Mount sections by adding  $85\ \mu\text{L}$  Eukitt Quick-hardening mounting medium (Sigma-Aldrich) to the section and gently placing a coverslip on top. [Troubleshooting 3](#).

**⚠ CRITICAL: Cresyl-violet acetate solution and Eukitt mounting medium are potentially hazardous, therefore these steps should be performed in a fume-hood with appropriate laboratory PPE.**

### Infarct volumetry

⌚ Timing: Approximately 30 min per brain

Infarct volumetry is performed here using the Swanson method, allowing the indirect infarct volume to be determined accounting for swelling.<sup>28,29</sup>

20. Scan slides using a scanner set at 600 dpi. Open the .jpg file with ImageJ and set the scale in the settings (Analyze -> Set Scale) to 23.62 pixels/mm and check the "global" option.
21. Adjust the contrast and/or brightness settings so infarcts are clearly visible (Image -> Adjust -> Brightness/Contrast -> Auto -> Apply).
22. For each brain section, using the polygon tool, draw an outline around the infarct area.
  - a. Measure and record the area (Analyze -> Measure).
  - b. Repeat the above steps for the ipsilateral and contralateral hemisphere.
  - c. Copy all the mean area values (three per section) and paste them into the provided [excel template](#).
  - d. Quantify the infarct volume by multiplying individual areas by the intersection distance ( $400\ \mu\text{m}$ ). Infarct volume is corrected for edema (swelling) by subtracting the difference

between the total ipsilateral and contralateral hemisphere volumes from the measured infarct volume.

### DNA extraction from fecal samples

⌚ Timing: Approximately 2 h for 20 samples

23. Weigh up to 250 mg stool (~5–10 fecal pellets) samples and transfer to QIAGEN PowerBead Tubes. Include at least one blank extraction control, with no sample added.
  - a. Let the stool samples thaw in tubes at room temperature.
  - b. Add 750  $\mu$ L PowerBead Solution and 60  $\mu$ L Solution 1 to each sample, vortex.
24. Place the tubes in a PowerLyser LT(Qiagen) for 6 min at 50 oscillations per second.
  - a. Check the homogenization of the samples, if necessary, repeat for a further 2 min.
25. Proceed with DNA extraction according to the manufacturer's protocol.
26. Measure and record DNA concentration and purity on a Nanodrop or similar instrument. [Troubleshooting 4](#).
27. Optional: Most DNA extraction kit protocols already include a drying step after washing, where the column is centrifuged once more before elution, without the addition of any solution/buffer. This step is not necessary for all kit protocols, however, it can improve DNA purity in some cases, as any additional alcohol contamination is removed.

⏸ Pause point: Extracted genomic DNA can be stored at  $-20^{\circ}\text{C}$  (weeks to months) for short-term storage or  $-80^{\circ}\text{C}$  for long-term (months to years).

### PCR and sequencing

⌚ Timing: Approximately 1 week to 1 month

Here we provide a brief overview of the PCR and sequencing steps. In our laboratory and indeed many others in the stroke-microbiota field, these steps are outsourced to a core facility or commercial vendor, thus it is not within the scope of this protocol to describe these steps in detail here. Nevertheless, a basic comprehension of these steps is crucial for understanding downstream data processing. This overview only pertains to Illumina based sequencing (MiSeq™ and HiSeq™), however, as these machines are the most commonly used for 16S rRNA sequencing. For researchers who have the means to conduct the PCR and sequencing in house, we refer them to Reitmeier et al.<sup>30</sup> for a full protocol detailing these steps.

28. After arrival at the sequencing facility, samples are diluted to an equal concentration and subjected to two rounds of PCR.
  - a. The first PCR amplifies genomic DNA using locus-specific primers targeting a particular hypervariable region of the 16S rRNA gene e.g., V4 region, using the 515f and 806r primer pair.
  - b. The second PCR step adds adapter/index sequences. This allow large numbers of samples to be pooled and sequenced simultaneously in a single run. These index sequences act as a barcode allowing reads from a given sample to be identified.
29. PCR products are then pooled and cleaned up using magnetic beads, leaving the purified amplicon sequences.
30. Purified DNA is then diluted to equimolar concentration and denatured before being spiked with PhiX control.

**Note:** The PhiX control is a DNA library derived from PhiX bacteriophage genome, which is used for calibration/quality control in Illumina sequencing, as well as adding base diversity which is useful for amplicons low base diversity targets like 16S.

31. The DNA is then loaded onto the sequencer cartridge and sequenced.
  - a. Samples can be sequenced in either single-end or paired-end mode. In single-end mode reads are sequenced in a single direction, whereas in paired-end mode reads are sequenced in both directions, generating overlapping reads which can later be merged. Paired-end mode is currently the most commonly used option for 16S rRNA sequencing.
  - b. Different sequencing platforms also have differing capabilities regarding read length. Illumina recommends  $2 \times 301$  bp if using a MiSeq™ and  $2 \times 151$  bp for HiSeq™ or other machines. Regardless, when using paired-end sequencing, the total read length should be long enough to allow significant overlap between the forward and reverse reads, to allow them to be successfully merged during data processing.
  - c. Output files are converted into demultiplexed (per sample) FASTQ files using the Illumina software bcl2fastq. Data is usually delivered in this format.

**Note:** This protocol assumes raw FASTQ files are returned demultiplexed, i.e., two files (forward and reverse reads), or in the case of single-end sequencing a single file per sample. This is usually the case; however, some facilities may still provide reads in multiplexed format, along with an index file(s) containing the barcode sequences. If FASTQ files are returned in multiplexed format, cutadapt<sup>16</sup> can be used to demultiplex reads.

## EXPECTED OUTCOMES

After tMCAO there is a failure rate of approximately 20% for an experienced investigator. This failure rate comprises the base mortality rate of 10%–15% as well as 5%–10% of animals which meet termination criteria early including indication of hemorrhage, no infarct as measured by cresyl-violet staining and early termination criteria (according to the responsible governmental department). Animals subjected to tMCAO will likely display acute changes in behavior/general health; mice should be excluded if they reach neuroscore values over 10 points for more than one day and experience a reduction over 20% of initial weight. At 3 days post-tMCAO, the ischemic area comprises the cortex (in most cases) and striatal regions (always). The indirect infarct volumes are approximately  $40 \pm 12$  mm<sup>3</sup>. Mice with hemorrhage or no infarct should be excluded from further analysis.

DNA isolation using commercial kits with bead-beating based lysis from stool samples is expected to yield high purity genomic DNA, with an approximate concentration range of 50–200 ng/μL. After sequencing, researchers should receive raw reads in FASTQ format. These may be combined (multiplexed) including index files, or per sample (demultiplexed). Sequencing on Illumina MiSeq™ or HiSeq™ machines should generally produce high quality (>Q30) reads, however a drop-off in quality towards the end of the reads is common and not unexpected. These files can then be processed to generate ASV tables.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Here we provide two protocols to process and analyze 16S rRNA sequencing data, one written mostly in R and the other using QIIME2 (Bolyen et al.<sup>13</sup>) - a python-based program and full-analysis suite for microbiota data – in combination with python itself. The aim of providing two protocols is to allow researchers to choose a method in a language they are comfortable with. Both protocols mirror each other as closely as possible, however due to minor differences in the code of the underlying algorithms, results may differ slightly. The main conclusions, however, should not, and results generated from either pipeline should be comparable.

The following steps comprise data pre-processing, including quality control, trimming/filtering, chimera removal and the generation of a table of amplicon sequence variants (ASV) with taxonomic assignment and finally a simple statistical analysis of the resulting data to examine differences between groups.

**Note:** If users do not currently have their own 16S rRNA sequencing data, the full dataset used in this example can be downloaded [here](#). For those wishing just to test the pipeline, a subset of this dataset can also be found [here](#). Note that all steps in both the R and QIIME2-based pipeline assume the scripts can be found in your current working directory or are added to the path variable.

### R-based pipeline

This section describes an example workflow from the R-based pipeline, processing the 16S dataset from Sorbie et al.<sup>1</sup> Users analyzing their own data should replace the highlighted paths/parameters to suit. Skip to step 12 for the QIIME2-based pipeline.

#### Data pre-processing

**Note:** Steps where the user has to replace the example parameters with their own machine specific parameters e.g., file paths, are highlighted in red text.

1. Perform quality control (QC) of raw reads using the provided bash script "qc.sh".
  - a. Determine read quality using FASTQC/MultiQC.<sup>14,15</sup>
    - i. Provide the script with the full path to FASTQ files ("-p" flag) and specify an output folder ("-o" flag). The "-t" flag is optional and specifies the number of threads to use while running FASTQC.

```
bash qc.sh -p ~/data_win/16S/iScience-review/data/fastq/ -o ~/data_win/16S/iScience-review/data/fastqc
```

- ii. Open and inspect the resulting html file: "Multiqc.html". Quality profiles should be similar to that shown in [Figure 2. Troubleshooting 5](#).
2. Remove non-biological sequences from reads (primer and adapter sequences).
    - a. Trim and filter reads using cutadapt.<sup>16</sup>
      - i. Specify the forward and reverse primer sequence (including ambiguous nucleotides) with the "-g" and "-G" flags, respectively.
      - ii. Provide a minimum ("-m") and maximum ("-M") read length.

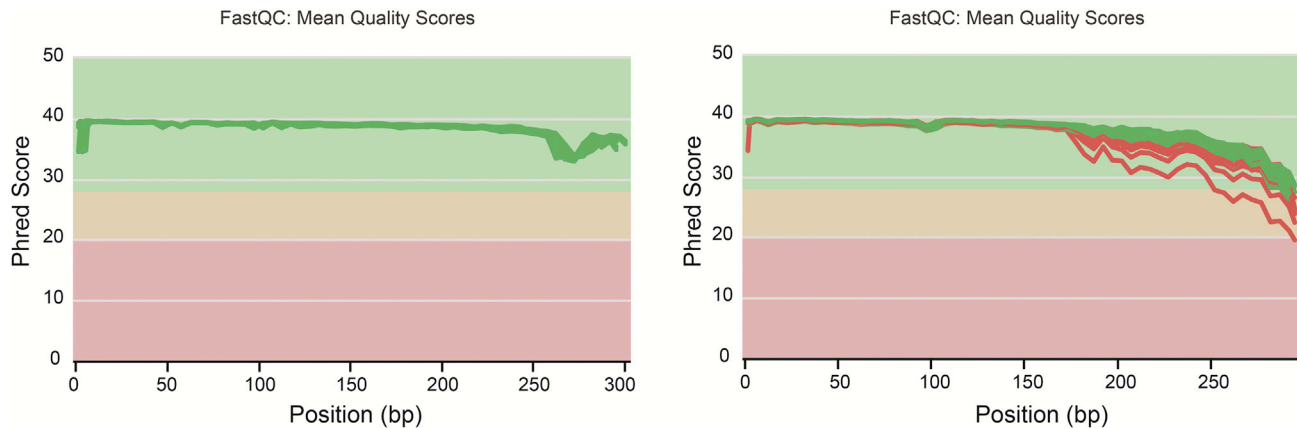
**Note:** Reads smaller or larger than this will be filtered out. We recommend using 10% of the expected read length as a cutoff in either direction.

- iii. Specify the path where FASTQ files are stored with the "-p" flag.

```
bash run_cutadapt.sh -g GTGCCAGCMGCCGCGGTAA -G GGACTACHVGGGTWTCTAAT -m 250 -M 300 -p ~/data_win/16S/iScience-review/data/fastq/
```

**Note:** This script automatically generates an output folder called "trimmed\_primer" in the same folder where the raw FASTQ files are stored.

⚠ **CRITICAL:** Ensure primers are properly removed during this step, as their presence can cause issues with merging of sequences and taxonomic assignment downstream. Cutadapt outputs a detailed log file "cutadapt\_primer\_trimming\_stats.txt" which can be inspected to check if primers were successfully removed.



**Figure 2. FASTQC read quality, for forward (left) and reverse (right) reads, generated as part of the R-based pipeline**

3. Determine optimal trimming parameters.
  - a. Run the “dada2\_qc.R” script to identify optimal parameters.

**Note:** This is essentially an additional QC step, using the DADA2 function “plotQuality-profile” internally.

- b. Provide the path (“-p”) to the trimmed reads from the previous step (step 2) and the name of the desired output folder (“-o”).
- c. Examine the output – for both forward and reverse reads if working with paired-end data - (Figure 3) and note down the approximate location where the aggregated quality begins to drop off (<Q30).

```
Rscript dada2_qc.R -p ~/data_win/16S/iScience-review/data/fastq/trimmed_primer/ -o ~/data_win/16S/iScience-review/data/dada2_qc_out/
```

4. Run DADA2 to denoise reads and generate ASV tables using the “run\_dada2\_PE.R” or “run\_dada2\_SE.R” for paired or single-end reads respectively.
  - a. Provide the path to the trimmed reads with “-p” flag.
  - b. To remove low quality tails and facilitate merging (if working with paired-end reads), reads must be trimmed. Trimming parameters are determined by the investigator in step 3c. Paired-end reads must retain at least 20–30 bp overlap to successfully merge.
    - i. Determine using the formula: Total read length – expected read length e.g., Here the expected amplicon length is 291 bp (Based on the primer binding sites: 806 – 515 = 291). If using 240 and 220 bp for the forward and reverse reads respectively, then the total overlap is 169 [(240 + 220) – 291].

**Note:** Quality on the reverse reads is often worse overall or drops off earlier thus commonly a lower value is selected.

- ii. Provide the identified values for the forward and the reverse reads using the “-f” and “-r” flag respectively.
- iii. Provide an output path with the “-o” flag.

If working with single-end data, use the “run\_DADA2\_SE.R” script. This is essentially the same as the above script, minus the flags specifying parameters for reverse reads.

```
Rscript run_dada2_PE.R -p ~/data_win/16S/iScience-review/data/fastq/trimmed_primer/ -f
240 -r 220 -o ~/data_win/16S/iScience-review/data/dada2_out/

Rscript run_dada2_SE.R -p ~/data_win/16S/iScience-review/data/fastq/trimmed_primer/ -f
240 -o ~/data_win/16S/iScience-review/data/dada2_out/
```

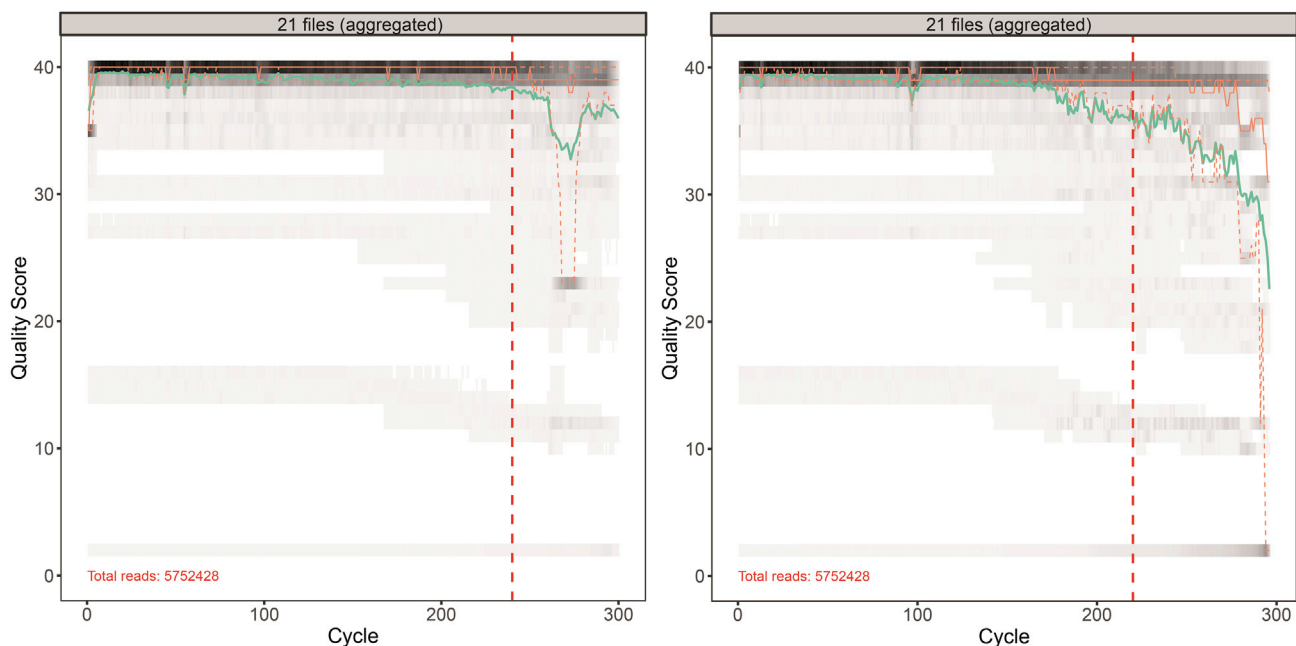
- Open and inspect the log file ("study\_stats.txt") to ensure the majority of reads passed each step. [Troubleshooting 6](#).

**Note:** In our experience ~60%–80% percent of reads are retained after all of the processing steps ("perc\_reads\_retained" column).

**Optional:** Additional flags specifying the number of expected errors on the forward and reverse reads respectively can also be specified using the "-n" and "-N" flags. To increase computation time, multithreading is also available and can be specified with the "-t" flag.

### Analysis

- Load analysis functions and data:
  - Source the provided analysis functions script "analysis\_functions.R". [Troubleshooting 7](#).
  - Manually remove blank extraction control from the metadata file. When importing to phyloseq this sample will be automatically removed from the ASV table as well.
  - Provide the path of the ASV table and phylogenetic tree generated in step 4, as well as the sample metadata to the "import\_as\_pseq" function. This creates a phyloseq object, used as input for downstream analyses.



**Figure 3. Example of DADA2 QC report depicting aggregated quality score across all reads, both forward (left) and reverse (right) after trimming of primers and adapter sequences**

Chosen trimming locations are highlighted by a red dotted line for each set of reads.



```
source("../scripts/analysis_functions.R")
ps <- import_as_pseq(asvtab = "../data/ASV_seqtab_tax.tab",
  mapping = "../data/Metadata-16S-sequenced_wo_ctrls.txt",
  tree = "../data/ASV_tree.tre")
```

7. Inspect taxonomy table to ensure taxonomic assignment was successful. [Troubleshooting 8](#).

```
View(tax_table(ps))
```

**Note:** Depending on environment, the majority (in our experience >80%) of ASVs should have taxonomy assigned to at least Family level.<sup>19,31</sup> Many unidentified ASVs or assignments of Bacteria at the Kingdom level only, are usually indicative of an issue with taxonomic assignment.

8. Normalize ASV table:

- a. Normalization is necessary to correct for differences in sequencing depth between samples.
  - i. Provide the phyloseq object to the "transform" function. The transform parameter accepts the arguments "mss" and "relative" to perform minimum sum scaling and relative abundance transformation respectively.

```
# Scaling to lowest read count total
ps_norm <- transform(ps, transform = "mss")

# Relative abundance
ps_rel <- transform(ps, transform = "relative")

# Check normalization output
colSums(otu_table(ps_norm))
colSums(otu_table(ps_rel))
```

9. Calculate alpha and beta-diversity metrics.

- a. Use the "calc\_alpha" function, providing the normalized phyloseq object as input to calculate Richness, Shannon effective diversity and Faith's Phylogenetic diversity.
- b. Extract metadata from the phyloseq object and merge with the alpha diversity results.
- c. Calculate beta diversity, using the "calc\_betadiv" function, providing the normalized phyloseq object, a distance metric (e.g., Unifrac) and finally an ordination method (e.g., NMDS). Supported metrics are: Unifrac, Weighted Unifrac, Generalized Unifrac and Bray-Curtis denoted as "unifrac", "wunifrac", "gunifrac" and "bray" respectively. Supported ordination metrics include: NMDS and PCoA, denoted as "NMDS" and "PCoA"/"MDS".

**Note:** Shannon effective diversity (calculated as the exponent of the Shannon index) is used here as unlike the Shannon index, Shannon effective diversity is linear, which facilitates more intuitive comparisons between groups as well as visualization.



```
## Alpha-diversity calculation and plotting
alpha_div <- calc_alpha(ps_norm)
meta <- meta_to_df(ps_norm)
# merge alpha diversity dataframe and metadata by rownames (0)
alpha_div_meta <- merge(alpha_div, meta, by=0)
## Beta-diversity calculation and plotting
betadiv <- calc_betadiv(ps_norm, dist = "gunifrac", ord_method = "NMDS")
```

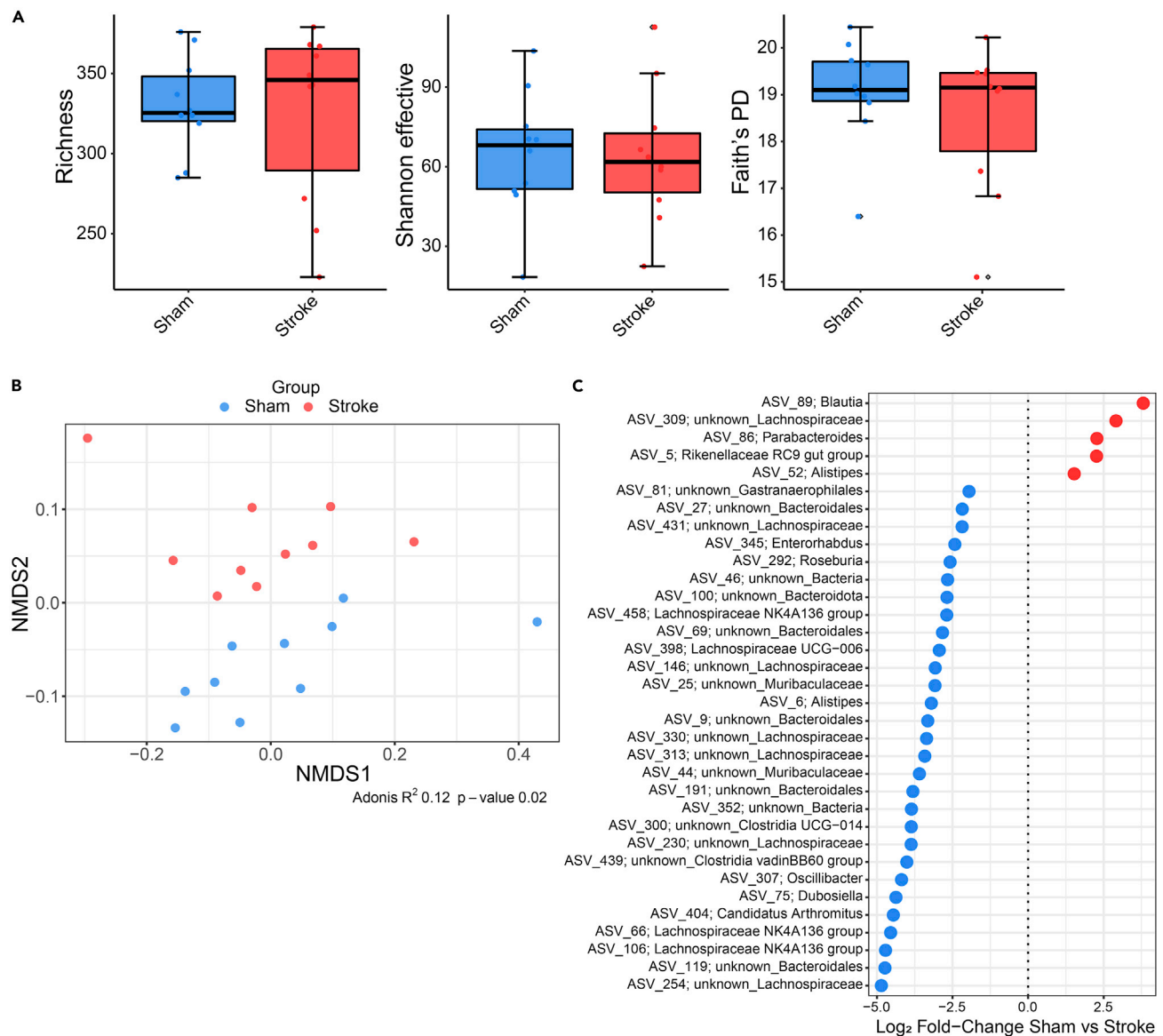
### 10. Plot alpha and beta-diversity metrics:

- a. Use the “plot\_boxplot” function to plot alpha diversity and perform statistical analysis in a single command (Figure 4A).
  - i. Define a color palette and a list of statistical comparisons.
  - ii. Provide the combined alpha diversity results and metadata to the “df” parameter, the name of the column name containing group information to the “variable\_col” parameter and the alpha-diversity metric to the “value\_col” parameter.
- b. Ordination plots of beta-diversity can be plotted using the “plot\_beta\_div” function (Figure 4B).

```
# Define colour palette
colours <- c("Sham" = "dodgerblue2", "Stroke" = "firebrick1")
# Define comparisons for statistical analysis
comps <- list(c("Stroke", "Sham"))
# Plot alpha-diversity
# Richness - for other metrics replace Richness with name of desired metric
plot_boxplot(df = alpha_div_meta, variable_col = "Group",
             value_col = "Richness", comparisons_list = comps,
             fill_var = "Group", xlab = "Group",
             ylab = "Richness", p_title = "Richness Stroke vs Sham",
             cols = colours, group.order = c("Sham", "Stroke"))
# Plot beta-diversity
plot_beta_div(ps_norm, dist_matrix = betadiv$Distance_Matrix,
             ordination = betadiv$Ordination, group_variable = "Group",
             cols = colours) +
xlab("NMDS1") +
ylab("NMDS2")
```

### 11. Identify differentially abundant taxa using ANCOM-BC.<sup>23</sup>

- a. Calculate differentially abundant taxa using the “ancom\_da” function.
  - i. Provide the normalized phyloseq object as the first argument.
  - ii. Define a formula to be tested.



**Figure 4. Representative figures depicting alpha, beta-diversity and differential abundance results from the R-based pipeline**

(A) Boxplots of Richness, Shannon effective and Faith's phylogenetic diversity comparing sham and stroke. Differences between means are tested using a Mann-Whitney U test. Error bars represent the median  $\pm$  1.5 multiplied by the IQR. Outliers are highlighted by an empty square.

(B) Non-metric multidimensional scaling plot of Generalized UniFrac distance comparing sham and stroke samples. Differences between groups are tested using PERMANOVA: Adonis  $R^2$  = 0.12,  $p$  value = 0.013.

(C) Log<sub>2</sub> fold-change of differentially abundant taxa between sham (blue) and stroke (red) as identified by ANCOM-BC.

**Note:** Confounding variables can be tested here using R formula notation e.g., for the example above: "Group + Age + Cage".

- iii. Provide the grouping variable. This is required for detection of structural zeroes.
- iv. Specify the desired group order using the "ord" parameter.
- b. Plot results using the "plot\_da" function (Figure 4C).
  - i. Provide the function with the ANCOM-BC results, a vector of the groups and a color palette if desired.

```
res_ancom <- ancom_da(ps_norm, formula = "Group", group = "Group", ord=c("Stroke", "Sham"))  
p <- plot_da(res_ancom, groups = c("Sham", "Stroke"), cols=colours)
```

**Note:** ANCOM-BC is preferred here based on the recommendation of Lin and Peddada<sup>23</sup> and Cappellato et al.,<sup>32</sup> being one of the few compositional differential-abundance methods with a good balance between statistical power and control of the false-discovery rate.<sup>23,32</sup>

### QIIME2-based pipeline

This section describes an example workflow from the QIIME2/Python-based pipeline, again processing the same data from the linked publication.<sup>1</sup>

#### Data processing

12. Activate the QIIME2 conda environment.

```
q2
```

13. Perform quality control (QC) of raw reads using the provided bash script "qiime2\_qc.sh".

**Note:** Reads are imported, and QC is performed here using the "qiime demux" command.

- a. Provide the script with the path to the read files using the "-p" flag.
- b. Specify an output folder with the "-o" flag.
- c. This should produce two files: "demux-paired-end.qza" and "demux-paired-end.qzv".

```
bash qiime2_qc.sh -p ~/data_win/16S/iScience-review/data/fastq/phred33 -g GTGCCAGCMGCCG  
CGGTAA -G GGACTACHVGGGTWTCTAAT -l 250 -o ~/data_win/16S/iScience-review/data/q2_out
```

d. Visualize the results with the "qiime tools view" command:

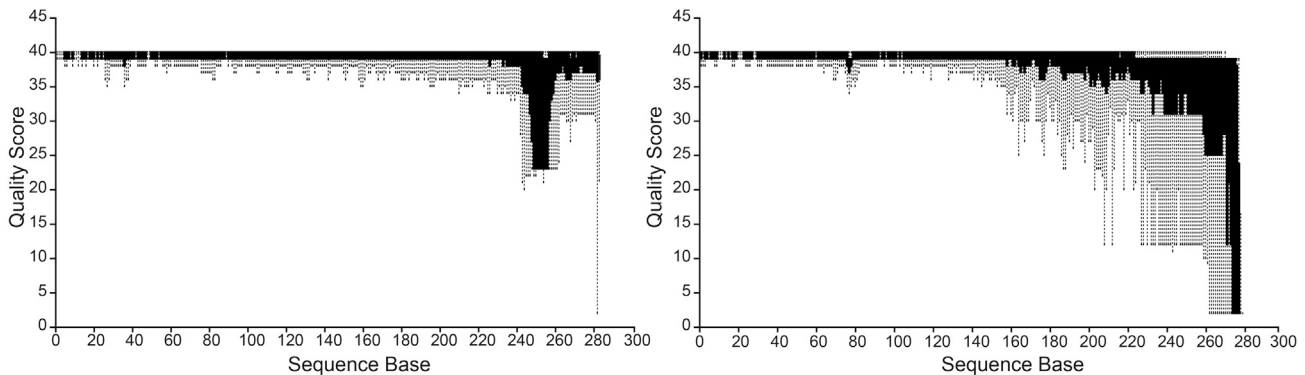
```
# Provide demux qzv file as argument  
qiime tools view demux-paired-end.qzv
```

e. Select the interactive quality plot tab to inspect read quality. The result should be similar to [Figure 5. Troubleshooting 5](#).

14. Run the pipeline script "qiime2\_pipeline.sh" to denoise reads and produce feature tables.

This script removes primer and adapter sequences, denoises reads and generates ASV tables and a phylogenetic tree, using the QIIME2 wrapper around cutadapt, DADA2 and FastTree respectively.

- a. Provide the path to the FASTQ files using the "-p" flag.
- b. Provide the path of the metadata table with the "-m" flag.
- c. Supply both the forward and reverse primer sequence, using "-g" and "-G" flags respectively.



**Figure 5. Screenshots of the interactive quality plots generated by QIIME2 for forward (left) and reverse (right) reads**  
Chosen trimming locations are highlighted by a red dotted line for each set of reads.

- d. Provide values for minimum read length for Cutadapt (“-l”) and forward (“-f”) and reverse (“-r”) read truncation length. These values are determined from step 13d.

```
bash qiime2_pipeline.sh -p ~/data_win/16S/iScience-review/data/q2_out/ -m ~/data_win/16S/iScience-review/data/Metadata-16S-sequenced_wo_ctrls.txt -f 240 -r 220
```

15. Inspect the output file “denoising-stats.qzv” to ensure reads were successfully merged.  
a. 60%–80% percent of reads should be retained after all processing steps (“percentage of input non-chimeric” column). [Troubleshooting 6](#).

If working with single-end data, use the “qiime2\_pipeline\_SE.sh” script. This script is modified to process single-end reads, however its usage is exactly the same as above, without the flags specifying parameters for reverse reads.

```
bash qiime2_pipeline_SE.sh -p ~/data_win/16S/iScience-review/data/q2_out/ -m ~/data_win/16S/iScience-review/data/Metadata-16S-sequenced_wo_ctrls.txt -f 240
```

**△ CRITICAL:** Ensure primers are properly removed during this step, as their presence can cause issues with merging of sequences and taxonomic assignment downstream. The “qiime cutadapt” command outputs a detailed log file “cutadapt\_primer\_trimming\_stats.txt” which can be inspected to check if primers were successfully removed.

**Note:** In some cases, reads may be in Phred64 format which must be converted to Phred33 before they can be imported by QIIME2. This information is normally specified by your sequencing provider; however, it is also possible to check by running the following command:

```
# check Phred format of a single file
vsearch -fastq_chars RD118_S1_L001_R1_001.fastq
```

If the reads are in Phred64 format, convert them by installing [bbmap<sup>33</sup>](#) and running the following command:

```
# install bbmap
conda install -c bioconda bbmap

# convert Phred format
bash convert_phred.sh -p ~/data_win/16S/iScience-review/data/fastq
```

### Analysis

16. Check processing was successful and prepare data for downstream analysis.
  - a. Inspect taxonomy table produced by the “qiime2\_pipeline.sh” script to ensure taxonomic assignment was successful. [Troubleshooting 8](#).
  - b. Remove the blank extraction control sample from the qiime artifact.
    - i. Open the metadata in a text editor/Microsoft excel and manually remove row containing negative control metadata. Save the file under a new name.
    - ii. Provide the edited file to the metadata argument of the “qiime feature-table filter-samples” command.
  - c. Normalize the feature table with the “normalize.sh” script.
    - i. Provide the filtered feature table from the previous step as input.
    - ii. This will output a file called “feature-table-norm.qza”, normalized using minimum sum scaling.

```
# View and inspect taxonomy
qiime tools view taxonomy.qzv

# remove from metadata manually and provide edited file as metadata argument. qiime feature-table filter-samples -i-table feature-table-filt.qza -m-metadata-file Metadata-16S-sequenced_wo_ctrls.txt -o-filtered-table feature-table-filt-nocontrols.qza

# normalize feature table using minimum sum scaling
bash normalize.sh feature-table-filt-nocontrols.qza

# outputs normalized feature table: feature-table-norm.qza as well as log file
```

17. Calculate alpha and beta-diversity metrics using the following commands:
  - a. Calculate core metrics (this command includes phylogenetic methods e.g., Faith’s PD and Unifrac and non-phylogenetic e.g., Shannon effective diversity and Bray-Curtis for alpha and beta diversity respectively).
  - b. Calculate GUniFrac distance.

**Note:** GUniFrac is not included in the core phylogenetic methods and thus must be calculated separately.

```
# Core phylogenetic methods
qiime diversity core-metrics-phylogenetic -i-table feature-table-norm.qza -i-phylogeny rooted-tree.qza -m-metadata-file Metadata-16S-sequenced_wo_ctrls.txt -p-sampling-depth 139010 -output-dir ./core-metrics-phylo-results

# GUniFrac
```

```
qiime diversity beta-phylogenetic -i-table feature-table-norm.qza -i-phylogeny rooted-tree.qza -p-metric 'generalized_unifrac' -p-alpha 0.5 -o-distance-matrix core-metrics-phylo-results/GUniFrac_distance_matrix.qza
```

18. Identify differentially abundant taxa using ANCOM-BC.

- a. Make a directory to store results e.g., "ANCOM".
- b. Identify differentially abundant taxa using the command "qiime ancombc ancombc".

**Note:** ANCOM-BC is not included in the core QIIME2 plugins, however, is provided via an external [plugin](#). ANCOM-BC is preferred for the same reasons detailed above.

```
# Core Non-phylogenetic metrics
qiime ancombc ancombc -i-table feature-table-norm.qza -m-metadata-file Metadata-16S-sequenced_wo_ctrls.txt -p-formula "Group" -p-p-adj-method 'BH' -p-zero-cut 0.7 -p-lib-cut 0 -p-group "Group" -p-struc-zero True -p-neg-lb False -p-tol 1e-05 -p-max-iter 100 -p-conserve True -p-alpha 0.05 -o-differentials ANCOM/ancom-bc_res.qza
```

19. Perform statistical analysis of alpha and beta-diversity and generate plots in python using JupyterNotebook/JupyterLab<sup>34</sup>:

- a. Load the required libraries (analysis\_functions, zipfile, pandas and seaborn) and define a variable containing the path to the QIIME2 output files: "observed\_features\_vector.qza", "shannon\_vector.qza", "faith\_pd\_vector.qza", "GUniFrac\_distance\_matrix.qza", "ancom-bc\_res.qza" and "taxonomy.qza". These files contain the Richness, Shannon effective diversity, Faith's PD, GUniFrac, ANCOM-BC results and taxonomy respectively.
- b. Load data:
  - i. As QIIME2 artifacts are essentially zip files they can be opened using the python zipfile library.
  - ii. Store the contents of the qza file as a "zipfile" object.
  - iii. Open the data files with "pandas.read\_csv", wrapping the "zipfile.open" function to open zipped files and the internal function "get\_zip\_name", to obtain the universally unique identifier (UUID) automatically.
  - iv. Load metadata as normal, using "pandas.read\_csv".

```
# import analysis functions
import analysis_functions as af

# enable in-line plotting
%matplotlib inline

# import libraries used in analysis
import zipfile

import pandas as pd

import seaborn as sns

# set data directory - i.e., where the main output of the qiime commands are stored
data_dir = "/home/adamsorbie/data_win/16S/iScience-review/analysis/qiime2/"
```

```
# Since qiime2 artifacts are just zip files we can get the unzipped contents using Zipfile
# alpha-diversity
r_zf = zipfile.ZipFile(data_dir + "core-metrics-phylo-results/observed_features_vector.qza")
s_zf = zipfile.ZipFile(data_dir + "core-metrics-phylo-results/shannon_vector.qza")
fpd_zf = zipfile.ZipFile(data_dir + "core-metrics-phylo-results/faith_pd_vector.qza")
# beta-diversity
gu_zf = zipfile.ZipFile(data_dir + "core-metrics-phylo-results/GUniFrac_distance_matrix.qza")
# differential abundance
ac_zf = zipfile.ZipFile(data_dir + "ANCOM/ancom-bc_res.qza")
tax_zf = zipfile.ZipFile(data_dir + "taxonomy.qza")
# Pandas read.csv combined with Zipfile lets us open the tsv files as pandas dataframe
# alpha-diversity
richness = pd.read_csv(r_zf.open(af.get_zip_name(r_zf) + '/data/alpha-diversity.tsv'),
sep="\t", index_col=0)
shannon = pd.read_csv(s_zf.open(af.get_zip_name(s_zf) + '/data/alpha-diversity.tsv'),
sep="\t", index_col=0)
faiths = pd.read_csv(fpd_zf.open(af.get_zip_name(fpd_zf) + '/data/alpha-diversity.tsv'),
sep="\t", index_col=0)
# beta-diversity
gunifrac = pd.read_csv(gu_zf.open(af.get_zip_name(gu_zf) + '/data/distance-matrix.tsv'),
sep="\t", index_col=0)
# differential abundance
ancom_res = pd.read_csv(ac_zf.open(af.get_zip_name(ac_zf) + '/data/differentials.tsv'),
names=["beta", "se", "W", "p-value", "q-value"], sep="\t", index_col=0, header=0)
taxonomy = pd.read_csv(tax_zf.open(af.get_zip_name(tax_zf) + '/data/taxonomy.tsv'), sep=
"\t", index_col=0)
# metadata
metadata = pd.read_csv(data_dir + "Metadata-16S-sequenced_wo_ctrls.txt", sep="\t", index_col=0)
```

20. Test and plot differences in alpha diversity (Figure 6A).

- a. Convert Shannon to Shannon effective.
- b. Merge the metadata with the alpha diversity results for plotting.
- c. Define pairs for statistical testing and the alpha diversity metrics to be tested.
- d. Define plotting options such as figure size and colors.
- e. Calculate alpha diversity using the "plot\_alpha" function.
  - i. Define a dictionary of desired axis labels.
  - ii. Provide the alpha diversity results merged with metadata, the metrics to be tested, the group column name, a list of colors, the dictionary of axis labels and finally the pairs for statistical testing.



```

# Convert shannon to shannon effective since Qiime2 does not calculate this by default.

# See discussion here for calculation: https://forum.qiime2.org/t/how-are-diversity-measures-calculated-in-qiime/6973/4

shannon["Shannon.E"] = 2** shannon.shannon_entropy

# Here we join all of the alpha diversity results and metadata into one dataframe

alphadiv = pd.concat([richness, shannon, faiths, metadata], axis=1)

# define pairs for stats

pairs=[("Sham", "Stroke")]

# plotting

# define metrics

alpha_metrics = ["observed_features", "Shannon.E", "faith_pd"]

# increase font size on plots since default is small

sns.set(font_scale=1.25)

# set fig size

sns.set(rc={'figure.figsize': (6.5, 5)})

# set background colour

sns.set_style("white")

# define colours

cols = ["#1e90ff", "#ff3030"]

# here we create a dictionary of labels to use as axis titles for each metric

labs = {"observed_features": "Richness", "Shannon.E": "Shannon effective", "faith_pd": "- Faith's PD"}

# use the calc_alpha function to plot and calculate stats

af.plot_alpha(alphadiv, alpha_metrics, "Group", cols, labs, pairs)

```

21. Test whether GUniFrac distance differs between groups with.
  - a. Calculate ordination with the "ordinate" function.
    - i. Defining two variables to store the ordination combined with sample metadata and the permutational multivariate analysis of variance (PERMANOVA) results, provide the distance matrix, metadata dataframe and group column name.
  - b. Plot beta diversity ordination using the "plot\_beta\_div" function (Figure 6B).
    - i. Set plotting options.
    - ii. Provide the merged ordination results and metadata and PERMANOVA results from above, as well as the group column name and desired ordination method (only NMDS and PCoA are currently supported).

```

# calculate betadiversity ordination and permanova

nmds_metadata, permanova = af.ordinate(gunifrac, metadata, group="Group")

# plot beta-diversity

# set styles and figure size

sns.set(rc={'figure.figsize': (8, 5)})

```

```
sns.set_style("white")  
af.plot_beta_div(nmds_metadata, permanova, group="Group", ord_method="NMDS", cols=cols)
```

22. Plot differentially abundant taxa:

- a. Plot ANCOM-BC results using the "plot\_ancom" function.
  - i. Set plotting options.
  - ii. Provide the ANCOM-BC results dataframe, metadata and taxonomy, a list of the groups and the colors.

```
# change figure size  
sns.set(rc={'figure.figsize': (6, 16)})  
  
# set style again  
sns.set_style("whitegrid")  
  
# plot ancom results  
ancom_plot = af.plot_ancom(ancom_res, metadata, taxonomy, groups=["Sham", "Stroke"], cols=  
reversed(cols))
```

## LIMITATIONS

The tMCAO model necessitates a complex surgical procedure, which requires significant training to ensure robust and reproducible results. With respect to the microbiota analysis, 16S rRNA sequencing has limited resolution and can generally only provide taxonomic annotations up to genus-level. As such, no information regarding species or strain level taxonomy and function can be determined. The provided pipelines, particularly the R-based pipeline are relatively fast and should run on most newer laptops, however very large datasets (>200 samples) or those that have been very deeply sequenced (>~100,000 reads per sample), may require access to more substantial computing infrastructure.

## TROUBLESHOOTING

### Problem 1

Poor survival of animals after stroke (mortality rate > ~10%–15%) (steps 1–9 under [step-by-step method details](#)).

### Potential solution

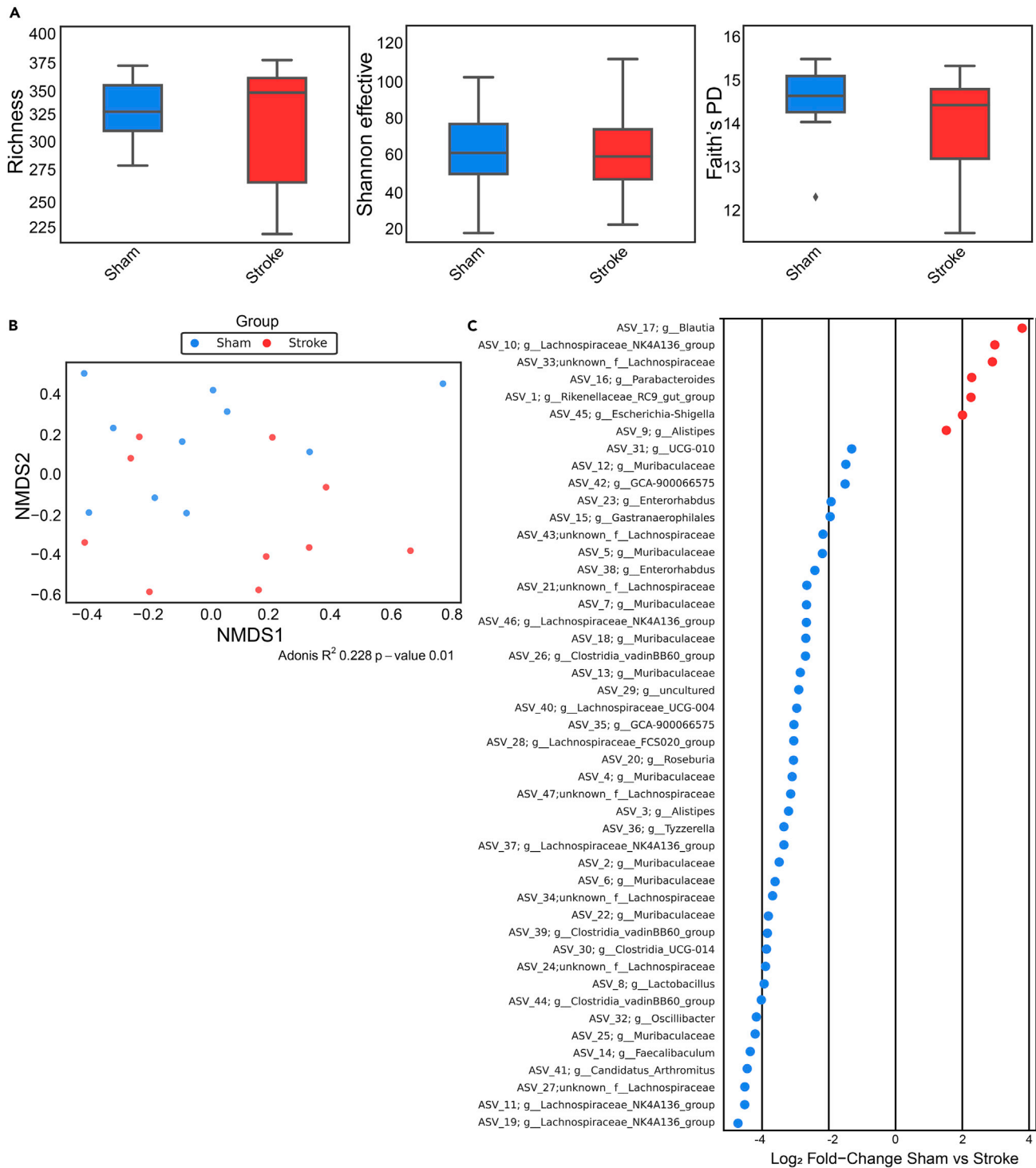
- Surgical procedures should take approximately 20 min as longer times is associated to high mortality rate due possibly to bleeding or impaired reperfusion.
- Subcutaneous injections of saline after surgery until sacrifice can also aid recovery.
- Housing experimental animals in a heated recovery chamber for the first two days after surgery for 2–4 h can facilitate recovery.
- Cleaning the animals daily with gauze and saline can prevent infections due to lack of grooming.

### Problem 2

Difficulty obtaining stool sample(s) (step 10 under [step-by-step method details](#)).

### Potential solution

- Gently massaging the abdomen of the animal can promote defecation.
- Place the mouse in a separate cage without bedding to collect stool sample.



**Figure 6. Representative figures depicting alpha, beta-diversity and differential abundance results from the QIIME2/python-based pipeline**

(A) Boxplots of Richness, Shannon effective and Faith's phylogenetic diversity comparing sham and stroke. Differences between means are tested using a Mann-Whitney U test. Error bars represent the median  $\pm$  1.5 multiplied by the IQR. Outliers are highlighted by a filled diamond.

(B) Non-metric multidimensional scaling plot of Generalized UniFrac distance comparing sham and stroke samples. Differences between groups are tested using PERMANOVA: Adonis  $R^2$  = 0.12,  $p$  value = 0.013.

(C) Log<sub>2</sub> fold-change of differentially abundant taxa between sham (blue) and stroke (red) as identified by ANCOM-BC.

- Spend a maximum of 15 min attempting to obtain a sample. If no sample can be obtained after this time, exclude the animal from the experiment.
- Stool collection may be easier in the morning as mice are nocturnal and consume the majority of their food during the dark period.

### Problem 3

Cresyl-violet staining problem: low quality staining (steps 14–19 under [step-by-step method details](#)).

### Potential solution

- Ensure cresyl-violet staining solution is stored in the dark and at room temperature.
- Keep the cresyl-violet staining solution protected from light during filtration.
- Replace the cresyl-violet staining solution every 3 months.

### Problem 4

DNA extraction – Low yield (steps 23–27 under [step-by-step method details](#)).

### Potential solution

- Heat elution buffer to 70°C before elution and/or increase incubation time for elution step.
- Decrease elution volume, for example from 50µL to 30µL.
- Increase starting material amount. We recommend starting with at least 20 mg and pooling multiple pellets from the same mouse and time point if necessary.

### Problem 5

Poor quality reads (steps 28–31 under [step-by-step method details](#)).

### Potential solution

- Contact sequencing provider to resolve any issues with sequencing/arrange re-sequencing.
- If resequencing not possible, using only forward reads (if of sufficient quality) is a valid approach.

### Problem 6

Low number of reads passing filtering or merging steps (step 4 for R-based pipeline and step 14 for QIIME2-based pipeline respectively, under [quantification and statistical analysis](#)).

### Potential solution

- Ensure overlap between paired reads is at least ~30 nucleotides.
- Data quality may be poor. In some cases, proceeding with forward reads only can solve this issue, at the expense of sequencing depth.

### Problem 7

Problems when installing R packages (step 15 under [before you begin](#) section).

### Potential solution

- Many R packages are compiled from source. Ensure you have the correct compiler installed for your system.

### Problem 8

Poorly assigned taxonomy (step 4 for R-based pipeline and step 14 for QIIME2-based pipeline respectively, under [quantification and statistical analysis](#)).

### Potential solution

- Check reads do not still contain non-biological sequences such as primers.
- Check low quality sections of reads have been properly removed.
- While we suggest SILVA, in some cases other databases such as RDP or GTDB, may have more coverage for certain taxa and using these either alone or in combination with SILVA to create a custom database may improve taxonomic assignment.<sup>35–37</sup>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests should be directed to and will be fulfilled by the lead/technical contact, Adam Sorbie ([adam.sorbie@med.uni-muenchen.de](mailto:adam.sorbie@med.uni-muenchen.de)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- Raw sequencing data were deposited in the European Nucleotide Archive under project number: [PRJEB48735](https://www.ebi.ac.uk/ena/browser/view/PRJEB48735) and are publicly available.
- Processed data, and subsampled raw data for testing is also available at the following [URL](#).
- Scripts used for data processing and analysis are publicly available on [GitHub](#).
- Any additional information required is available from the [lead contact](#) upon request

### ACKNOWLEDGMENTS

Figures were created using BioRender (Toronto). This study was supported by the European Commission H2020MSCA-IF 2016 753893 MetaBiota (C.B.).

### AUTHOR CONTRIBUTIONS

Conceptualization, A.S., C.B.; data curation, A.S.; formal analysis, A.S.; investigation, A.S., R.D., M.W.; methodology, A.S., R.D., M.W.; supervision, A.S., C.B.; visualization, A.S.; writing original draft, A.S.; funding acquisition, C.B.; project administration, C.B.; manuscript review, C.B.; editing, C.B.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### REFERENCES

1. Sorbie, A., Delgado Jiménez, R., and Benakis, C. (2022). Increasing transparency and reproducibility in stroke-microbiota research: a toolbox for microbiota analysis. *iScience* 25, 103998. <https://doi.org/10.1016/j.isci.2022.103998>.
2. Sadler, R., Singh, V., Benakis, C., Garzetti, D., Brea, D., Stecher, B., Anrather, J., and Liesz, A. (2017). Microbiota differences between commercial breeders impacts the post-stroke immune response. *Brain Behav. Immun.* 66, 23–30. <https://doi.org/10.1016/j.bbi.2017.03.011>.
3. Montonye, D.R., Ericsson, A.C., Busi, S.B., Lutz, C., Wardwell, K., and Franklin, C.L. (2018). Acclimation and institutionalization of the mouse microbiota following transportation. *Front. Microbiol.* 9, 1085. <https://doi.org/10.3389/fmicb.2018.01085>.
4. Manwani, B., Bentivegna, K., Benashski, S.E., Venna, V.R., Xu, Y., Arnold, A.P., and McCullough, L.D. (2015). Sex differences in ischemic stroke sensitivity are influenced by gonadal hormones, not by sex chromosome complement. *J. Cereb. Blood Flow Metab.* 35, 221–229. <https://doi.org/10.1038/jcbfm.2014.186>.
5. Manwani, B., Liu, F., Scranton, V., Hammond, M.D., Sansing, L.H., and McCullough, L.D. (2013). Differential effects of aging and sex on stroke induced inflammation across the lifespan. *Exp. Neurol.* 249, 120–131. <https://doi.org/10.1016/j.expneurol.2013.08.011>.
6. Fujii, M., Hara, H., Meng, W., Vonsattel, J.P., Huang, Z., and Moskowitz, M.A. (1997). Strain-related differences in susceptibility to transient forebrain ischemia in SV-129 and C57black/6 mice. *Stroke* 28, 1805–1810. , discussion: 1811. <https://doi.org/10.1161/01.str.28.9.1805>.
7. Brichacek, A.L., Nwafor, D.C., Benkovic, S.A., Chakraborty, S., Kenney, S.M., Mace, M.E., Jun, S., Gambill, C.A., Wang, W., Hu, H., et al. (2020). Experimental stroke induces chronic gut dysbiosis and neuroinflammation in male mice. Preprint at bioRxiv. <https://doi.org/10.1101/2020.04.29.069575>.
8. Stanley, D., Moore, R.J., and Wong, C.H.Y. (2018). An insight into intestinal mucosal microbiota disruption after stroke. *Sci. Rep.* 8, 568. <https://doi.org/10.1038/s41598-017-18904-8>.
9. Singh, V., Roth, S., Llovera, G., Sadler, R., Garzetti, D., Stecher, B., Dichgans, M., and Liesz, A. (2016). Microbiota dysbiosis controls the neuroinflammatory response after stroke. *J. Neurosci.* 36, 7428–7440. <https://doi.org/10.1523/JNEUROSCI.1114-16.2016>.
10. Houlden, A., Goldrick, M., Brough, D., Vizi, E.S., Lénárt, N., Martinecz, B., Roberts, I.S., and Denes, A. (2016). Brain injury induces specific changes in the caecal microbiota of mice via altered autonomic activity and mucoprotein

- production. *Brain Behav. Immun.* 57, 10–20. <https://doi.org/10.1016/j.bbi.2016.04.003>.
11. Nearing, J.T., Comeau, A.M., and Langille, M.G.I. (2021). Identifying biases and their potential solutions in human microbiome studies. *Microbiome* 9, 113. <https://doi.org/10.1186/s40168-021-01059-0>.
  12. Abellan-Schneyder, I., Machado, M.S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., List, M., and Neuhaus, K. (2021). Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. *mSphere* 6, e01202-20. <https://doi.org/10.1128/mSphere.01202-20>.
  13. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
  14. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].
  15. Ewels, P., Magnusson, M., Lundin, S., and Källér, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
  16. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10. <https://doi.org/10.14806/ej.17.1.200>.
  17. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. <https://doi.org/10.1038/nmeth.3869>.
  18. McMurdie, P.J., and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8, e61217. <https://doi.org/10.1371/journal.pone.0061217>.
  19. Murali, A., Bhargava, A., and Wright, E.S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome* 6, 140. <https://doi.org/10.1186/s40168-018-0521-5>.
  20. Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464. <https://doi.org/10.1093/bioinformatics/btq166>.
  21. Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. <https://doi.org/10.1093/bioinformatics/btq706>.
  22. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szocs, E., et al. (2022). Vegan: Community Ecology Package.
  23. Lin, H., and Peddada, S.D. (2020). Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* 11, 3514. <https://doi.org/10.1038/s41467-020-17041-7>.
  24. Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., Collman, R.G., Bushman, F.D., and Li, H. (2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* 28, 2106–2113. <https://doi.org/10.1093/bioinformatics/bts342>.
  25. Llovera, G., Simats, A., and Liesz, A. (2021). Modeling stroke in mice: transient middle cerebral artery occlusion via the external carotid artery. *JoVE*, e62573. <https://doi.org/10.3791/62573>.
  26. Llovera, G., Hofmann, K., Roth, S., Salas-Pédomo, A., Ferrer-Ferrer, M., Perego, C., Zanier, E.R., Mamrak, U., Rex, A., Party, H., et al. (2015). Results of a preclinical randomized controlled multicenter trial (pRCT): anti-CD49d treatment for acute brain ischemia. *Sci. Transl. Med.* 7, 299ra121. <https://doi.org/10.1126/scitranslmed.aaa9853>.
  27. McGee-Russell, S.M., Brown, A.W., and Brierley, J.B. (1970). A combined light and electron microscope study of early anoxic-ischaemic cell change in rat brain. *Brain Res.* 20, 193–200. [https://doi.org/10.1016/0006-8993\(70\)90288-X](https://doi.org/10.1016/0006-8993(70)90288-X).
  28. Jackman, K., Kunz, A., and Iadecola, C. (2011). Modeling focal cerebral ischemia in vivo. In *Neurodegeneration Methods in Molecular Biology*, G. Manfredi and H. Kawamata, eds. (Humana Press), pp. 195–209. [https://doi.org/10.1007/978-1-61779-328-8\\_13](https://doi.org/10.1007/978-1-61779-328-8_13).
  29. Swanson, R.A., Morton, M.T., Tsao-Wu, G., Savalos, R.A., Davidson, C., and Sharp, F.R. (1990). A semiautomated method for measuring brain infarct volume. *J. Cereb. Blood Flow Metab.* 10, 290–293. <https://doi.org/10.1038/jcbfm.1990.47>.
  30. Reitmeier, S., Kiessling, S., Neuhaus, K., and Haller, D. (2020). Comparing circadian rhythmicity in the human gut microbiome. *STAR Protoc.* 1, 100148. <https://doi.org/10.1016/j.xpro.2020.100148>.
  31. Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
  32. Cappellato, M., Baruzzo, G., and Di Camillo, B. (2022). Investigating differential abundance methods in microbiome data: a benchmark study. *PLoS Comput. Biol.* 18, e1010467. <https://doi.org/10.1371/journal.pcbi.1010467>.
  33. Bushnell, B. (2014). BMAP: A Fast, Accurate, Splice-Aware Aligner.
  34. Kluwyer, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., et al. (2016). Jupyter Notebooks—A Publishing Format for Reproducible computational Workflows. *Positioning and Power in (Academic Publishing: Players, Agents and Agendas, IOS Press)*, pp. 87–90.
  35. Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. <https://doi.org/10.1093/nar/gkt1244>.
  36. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38, 5315–5316. <https://doi.org/10.1093/bioinformatics/btac672>.
  37. Robeson, M.S., O'Rourke, D.R., Kaehler, B.D., Ziemski, M., Dillon, M.R., Foster, J.T., and Bokulich, N.A. (2021). RESCRIPt: reproducible sequence taxonomy reference database management. *PLoS Comput. Biol.* 17, e1009581. <https://doi.org/10.1371/journal.pcbi.1009581>.