

AUG 10, 2023

OPEN ACCESS



**DOI:**  
[dx.doi.org/10.17504/protocols.io.rm7vzbe32vx1/v1](https://dx.doi.org/10.17504/protocols.io.rm7vzbe32vx1/v1)

**Protocol Citation:** Luis Enrique Vazquez-Morado, Melany J Cervantes-Echeverría, Fernanda Cornejo-Granados, Luigui Gallardo-Becerra, Filiberto Sánchez-López, Adrian Ochoa-Leyva 2023. Protocol for the Experimental and Bioinformatic Analysis for Gut Microbiota Profiling of *Litopenaeus vannamei*.

**protocols.io**  
<https://dx.doi.org/10.17504/protocols.io.rm7vzbe32vx1/v1>

**License:** This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

# Protocol for the Experimental and Bioinformatic Analysis for Gut Microbiota Profiling of *Litopenaeus vannamei*

Luis Enrique Vazquez-Morado<sup>1</sup>, Melany J Cervantes-Echeverría<sup>1</sup>,  
 Fernanda Cornejo-Granados<sup>1</sup>, Luigui Gallardo-Becerra<sup>1</sup>,  
 Filiberto Sánchez-López<sup>1</sup>, Adrian Ochoa-Leyva<sup>1</sup>

<sup>1</sup>Departamento de Microbiología Molecular, Instituto de Biotecnología (IBT), Universidad Nacional Autónoma de México (UNAM) Av. Universidad #2001, Col. Chamilpa, Cuernavaca, Morelos 62210, México.



Luis Enrique Vazquez-Morado

## ABSTRACT

The Pacific white shrimp (*Litopenaeus vannamei*, *Lv*) is a non-model organism of significant global economic importance, constituting 51.7% of crustaceans produced through aquaculture. Consequently, scientific research has focused on various approaches to enhance the production of this crustacean, such as understanding the dynamics of its microbiota. Microbiota refers to the collection of microorganisms inhabiting a host organism, which can be essential for the host's healthy development. The gut is one of the most critical organs for shrimp development and survival. The current protocol outlines the methodology for adequately extracting gut samples from *Lv*. It describes the DNA extraction method, amplification, and purification of the V3 region of the 16S ribosomal gene and the construction of libraries for subsequent sequencing. Furthermore, it also includes a bioinformatic protocol for sequence analysis to describe the microbiota composition.

## IMAGE ATTRIBUTION

Microbiomics Lab. IBT UNAM.

## MATERIALS

### ### Shrimp Gut Sample Extraction ###

- Eppendorf Tubes of 1.5 ml
- RNeasy® or Glycerol at 20%
- Scalpel
- Dissection Forceps

**Protocol status:** Working  
We use this protocol and it's working

**Created:** Mar 17, 2023

**Last Modified:** Aug 10, 2023

**PROTOCOL integer ID:**  
78998

**Keywords:** Shrimp Gut Microbiome, DNA Extraction, Microbiome Bioinformatics, Shrimp Gut Extraction, 16s Gene Amplification, Amplicon Purification

- Two 100 ml. Beaker
- Milli-Q Water
- 10% chlorine Solution
- One Rack for Eppendorf Tubes
- Latex gloves
- Cooler with Ice

### ### DNA Extraction from Shrimp Intestine Samples - Amplification and Purification of the V3 Region of the 16s Ribosomal Gene ###

- Quick-DNA Fecal/Soil Microbe MiniPrep Kit, Zymo Research
- 10% Chlorine Solution
- Milli-Q Water
- Dissection Forceps
- Vortex Genie Pulse for Eppendorf Tubes
- Eppendorf Tubes of 1.5 ml
- phosphate buffer at pH 7.4
- One Rack for Eppendorf Tubes
- Cooler with Ice
- Refrigerated Centrifuge for 1.5 ml Eppendorf Tubes
- Two 100 ml. Beaker
- Agarose Gels at 1.5% and 2%
- 6X DNA Loading Dye, Thermo Scientific
- TBE buffer 1X
- GelRed® Nucleic Acid Stainer
- Nuclease-Free Water, Thermo Fisher Scientific
- Gel Electrophoresis Chamber for agarose Gels
- Qubit™ Instrument
- Qubit dsDNA HS Assay Kit, Invitrogen™
- UV transilluminator
- Forward and reverse oligonucleotides for the V3 region of the 16s ribosomal gene
- Q5® Hot Start High-Fidelity 2X Master Mix
- Thermocycler
- 80% Ethanol
- Ampure XP Beads
- DynaMag™-2 Magnet

### ### Analysis of the shrimp microbiome using DADA2 by Qiime2 ###

It is necessary to install the following programs on a computer with a Linux distribution or on a Linux terminal enabled on a Windows computer.

- Update the package list on your linux system:  
\$ sudo apt-get update

```
-Anaconda3 - https://repo.anaconda.com/archive/
$ cd /tmp
$ curl https://repo.anaconda.com/archive/Anaconda3-2023.07-2-Linux-x86\_64.sh --
output anaconda.sh
$ bash anaconda.sh
$ yes
confirm the output location
→ Press Enter
Conda init?
$yes
$ source ~/.bashrc
```

```
- FastQC v0.12.1
https://github.com/s-andrews/FastQC
$ sudo apt install fastqc
$ conda install -c bioconda fastqc
```

```
-Cutadapt 4.0
https://github.com/marcelm/cutadapt
$ Sudo apt install cutadapt
$ conda install -c bioconda cutadapt
```

```
-Prinseq-Lite 0.20.4
https://github.com/uwb-linux/prinseq
$ sudo apt install prinseq-lite
```

```
-Trimmomatic v0.39
https://github.com/usadellab/Trimmomatic
$ sudo apt install trimmomatic
$ conda install -c bioconda trimmomatic
```

```
-16s gene Silva 138.1 Database
https://www.arb-silva.de
```

```
-Qiime2 2023_5
https://docs.qiime2.org/2023.5/install/
$ conda install -c qiime2 qiime2
```

```
-qzv Archives Visualizer
https://view.qiime2.org
```

**### Primers for the 16s ribosomal gene ###**

### V3 region

Forward 338F:

ACTCCTACGGGAGGCAGCAG

Reverse 533R:

TTACCGCGGCTGCTGGCAC

### V3-V4 region

Forward 341F:

CCTACGGGNGGCWGCAG

Reverse 805R:

GACTACHVGGGTATCTAATCC

## SAFETY WARNINGS



In case of direct contact with any of the reagents presented in this protocol, please rinse abundantly with water. Avoid consuming them and follow the basic safety protocols of your laboratory. We also recommend following the manufacturer's recommendations for the recommended kits.

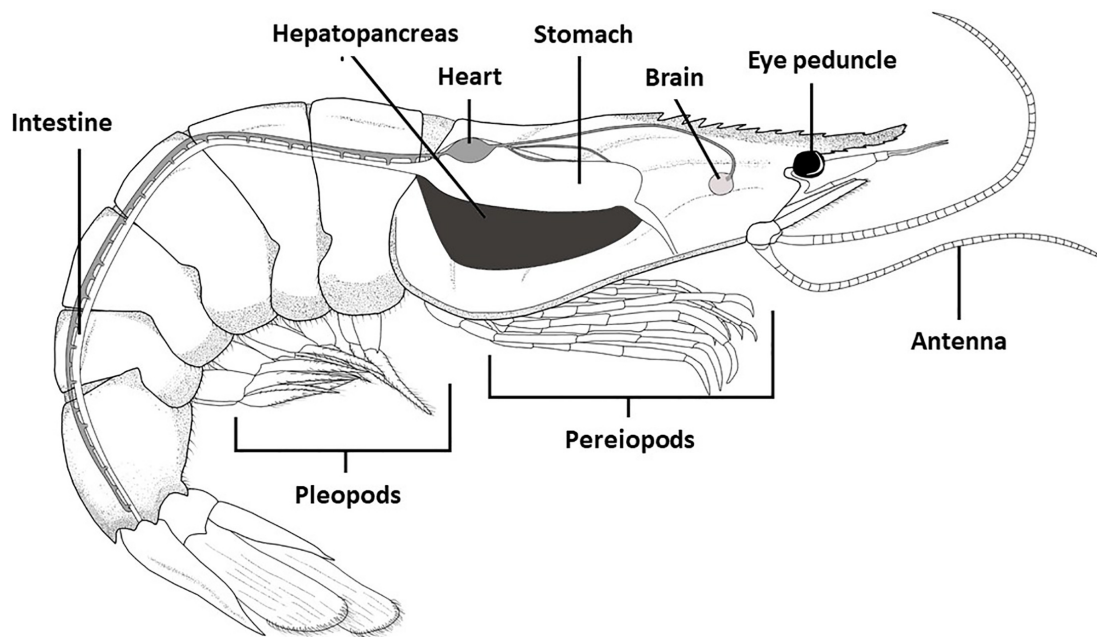
## BEFORE START INSTRUCTIONS

Before starting, it is necessary to have a thoroughly disinfected workspace. Use 10% sodium hypochlorite solutions, 70% ethanol, distilled water, and absorbent paper towels to clean the area. It is advisable to collect intestine samples from shrimp that have been recently taken out of the pond. Otherwise, it is essential to verify the integrity of this organ and confirm that it is not undergoing lysis. Additionally, a basic understanding of Linux and related working environments is recommended for processing the final reads.

## Shrimp Gut Sample Extraction

- 1 Before beginning any sample collection, it is crucial to prepare the working environment properly. Start by rinsing the surgical forceps with a 10% chlorine solution, then rinse thoroughly with sterile Milli-Q water. Next, place 1.5 ml Eppendorf tubes with 1 mL of RNAlater or 20% glycerol as preservation reagent in a rack for easy access during sampling. If collecting samples outside the laboratory, be sure to have a container of liquid nitrogen on hand to freeze the samples of 20% glycerol and a cooler with ice to keep the samples with RNAlater for transport. By taking these simple precautions, you can ensure the integrity of your samples and obtain accurate results.

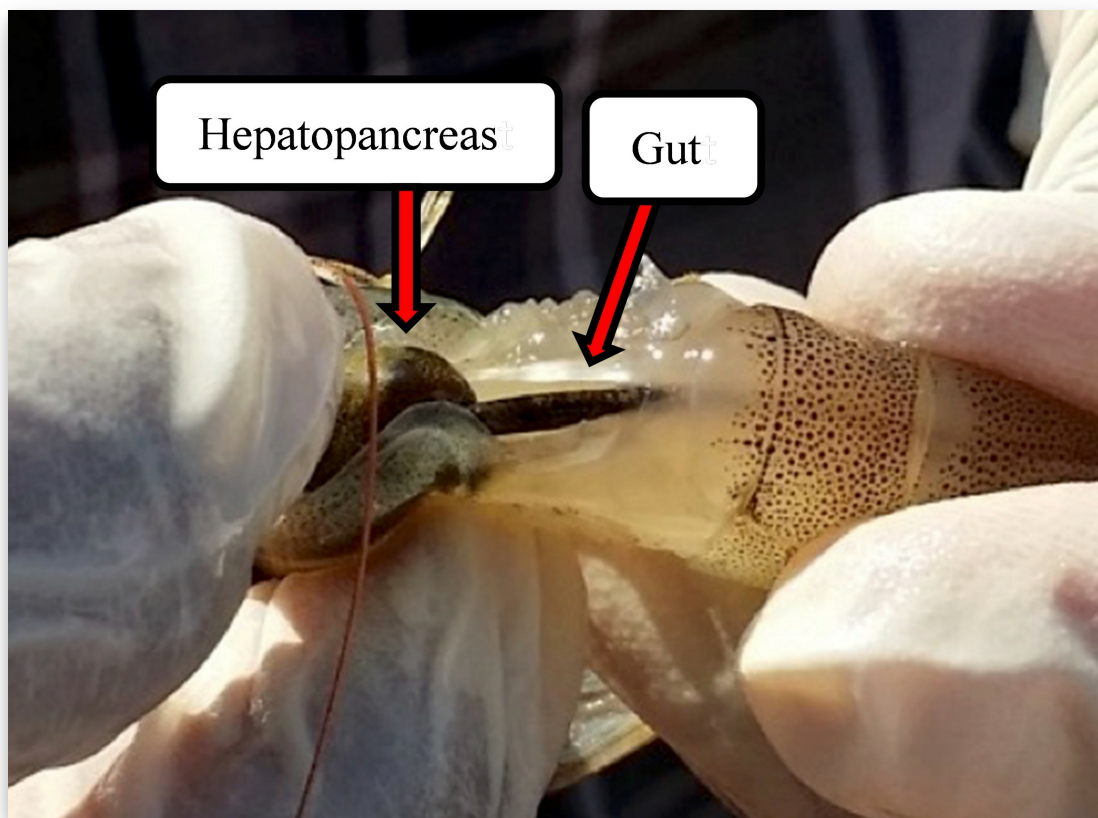
- 2 First, to collect a shrimp sample, put on sterile latex gloves and remove the exoskeleton without detaching the cephalothorax zone. Then, hold the shrimp with your dominant hand and apply measured pressure to expose the hepatopancreas and gut. Next, cut along the shrimp following the same line as the gut, careful not to damage the organ. Cut the gut connection with the hepatopancreas, and open the shrimp like a butterfly following the previously made cut. Finally, transfer the gut using surgical forceps to a 1.5 mL Eppendorf tube with preservation reagent. Keep the RNA later sample at 4°C for 48 hours and posteriorly maintain it at -70 C until use. Contrary, save the glycerol sample directly in liquid nitrogen and keep it at -70 until use. Remember to take the necessary precautions to ensure the integrity of your samples and obtain accurate results.



**Figure 1. Anatomy of *Litopenaeus vannamei*.**

Duarte y col., 2020.

<https://doi.org/10.1371/journal.pone.0231310>



**Figure 2. Gut (Intestine) and Hepatopancreas of *Penaeus vannamei***

Image from Microbiomics Laboratory. IBT, UNAM.

#### Note

The cephalothorax region encompasses from the rostrum (the small horn on the shrimp's forehead) to the heart region (see Figure 1).

The butterfly-style incision is recommended to facilitate gut extraction since it is connected to the anus. If the gut extraction is forced could be ruptured.

## DNA Extraction from Shrimp Gut Samples

- 3 Take the previously collected sample from the freezer (-70° C) and let it thaw on ice to extract the genomic DNA. Once the sample is thawed, centrifuge the tube 8,000 g for 5 min. Then, discard the supernatant, add 1 ml of phosphate buffer at pH 7.4, and let it sit on ice for 0.5-1 minute. This is to remove as much of the conservation reagent as possible.

- 4 From now on, use the genomic DNA extraction kit Quick-DNA Fecal/Soil Microbe MiniPrep Zymo Research. You can opt to use the ZR Bashing Bead Lysis Tube (mechanical lysis) or use the Genomic Lysis Buffer directly (chemical lysis).

5

#### Note

Extraction Kit Manual:

[https://files.zymoresearch.com/protocols/\\_d6010\\_quick-dna\\_fecalsoil\\_microbe\\_miniprep\\_kit.pdf](https://files.zymoresearch.com/protocols/_d6010_quick-dna_fecalsoil_microbe_miniprep_kit.pdf)

This kit was chosen because shrimp gut samples often contain traces of sediment, affecting DNA quality and final concentration if non-specialized kits are used. Additionally, the kit uses chemical and mechanical lysis which helps recover greater bacterial richness.

## DNA Extraction from Shrimp Gut Samples

- 6 To only use the chemical lysis skip to step 10 and add 1,200 µl of the Genomic Lysis Buffer to the sample. Continue with the following steps as specified.  
For the mechanical lysis transfer the gut sample in PBS buffer to a ZR Bashing Bead Lysis Tube and add 700 µl of BashingBead Buffer. Secure the tube in a Vortex Genie2 and agitate for 5 min at maximum speed.

7

#### Note

If you have a beat beater device other than the Vortex Genie2, refer to the Appendix in the Extraction Kit Manual:

[https://files.zymoresearch.com/protocols/\\_d6010\\_quick-dna\\_fecalsoil\\_microbe\\_miniprep\\_kit.pdf](https://files.zymoresearch.com/protocols/_d6010_quick-dna_fecalsoil_microbe_miniprep_kit.pdf)

## DNA Extraction from Shrimp Gut Samples

- 8 Centrifuge the ZR BashingBead™ Lysis Tube in a microcentrifuge at  $\geq 10,000 \times g$  for 1 minute.
- 9 Transfer up to 400  $\mu\text{l}$  supernatant to a Zymo-Spin™ III-F Filter in a collection tube and centrifuge at  $8,000 \times g$  for 1 minute.
- 10 Add 1,200  $\mu\text{l}$  of the Genomic Lysis Buffer to the filtrate in the collection Tube from the previous step. Mix well.
- 11 Transfer 800  $\mu\text{l}$  of the mixture from the previous step to a Zymo-Spin™ IICR column in a collection tube and centrifuge at  $10,000 \times g$  for 1 minute.
- 12 Discard the flow through from the collection tube and repeat the previous step.
- 13 Add 200  $\mu\text{l}$  DNA Pre-wash buffer to the Zymo-Spin™ IICR Column in a new collection tube and centrifuge at  $10,000 \times g$  for 1 minute
- 14 Add 500  $\mu\text{l}$  g-DNA wash buffer to the Zymo-Spin™ IICR Column and centrifuge at  $10,000 \times g$  for 1 minute.
- 15 Transfer the Zymo-Spin™ IICR column to a clean 1.5 ml microcentrifuge tube and add 100  $\mu\text{l}$  (50  $\mu\text{l}$  minimum) DNA elution buffer directly to the column matrix. Centrifuge at  $10,000 \times g$  for 30 seconds to elute the DNA.



**16** Place a Zymo-Spin™ III-HRC filter in a clean collection Tube and add 600 µl prep solution. Centrifuge at 8,000 x g for 3 minutes.

**17** Transfer the eluted DNA to a prepared Zymo-Spin™ III-HRC Filter in a clean 1.5 ml microcentrifuge tube and centrifuge at exactly 16,000 x g for 3 minutes.

The purified DNA should be stored at -20°C until use. For long storage periods ranging from months to years, -80 °C is recommended.

**18** To quantify DNA recovered in previous extraction, use a fluorometric method as Qubit dsDNA HS Assay with the manufacturer's recommended conditions.

Manual:

[https://assets.fishersci.com/TFS-Assets/LSG/manuals/MAN0017455\\_Qubit\\_1X\\_dsDNA\\_HS\\_Assay\\_Kit\\_UG.pdf?\\_ga=2.202384017.1462401010.1681754434-2071138171.1681754434](https://assets.fishersci.com/TFS-Assets/LSG/manuals/MAN0017455_Qubit_1X_dsDNA_HS_Assay_Kit_UG.pdf?_ga=2.202384017.1462401010.1681754434-2071138171.1681754434)

#### Note

If you're looking to quantify samples in your laboratory, you can use various methods. However, for sequencing purposes, you should use the Qubit platform. This method is known for its higher precision, sensitivity, and measurement range, making it an excellent choice for those looking to get the most accurate results possible.

**19** It is necessary to assess the quality of the extracted DNA extracted by agarose gel (2%) electrophoresis using fluorescent staining of dsDNA. An appropriate DNA size marker should be loaded along with the samples. The gel should be run at a constant voltage of 100 volts for 30 minutes. Finally, to visualize the gel in UV-transilluminator.

## Amplification and Purification of the V3 region of the 16s R...

**20** Prepare the following PCR reaction using the DNA extracted from shrimp gut:

V-Add the required amount of sample and make up to 20 µl with water.

Reaction component	Volume in µl
--------------------	--------------

Reaction component	Volume in $\mu\text{l}$
DNA Template (~12.5 ng)	V
Forward primer rRNA 16s Gene (10 pmol/ $\mu\text{l}$ )	1
Reverse primer rRNA 16s Gene (10 pmol/ $\mu\text{l}$ )	1
Reagent Q5 Hot Start Master Mix	10
H <sub>2</sub> O	x
Final Volume	20

### Preparation of the PCR Reaction for the Amplification of Regions of the 16s Ribosomal Gene

#### Note

If necessary, in the materials section, we have oligonucleotides for the amplification of the V3-V4 region. The use of the V3 or V3-V4 region of the 16S ribosomal gene will depend on the needs of your study. It has been noted that using the V3 region is ideal for descriptive studies, where taxonomic detection at the family or genus level is sufficient. On the other hand, the V3-V4 region is suitable for studies requiring greater resolution, as this strategy allows us to define at the genus and species level.

## 21 Amplify under the following conditions in the thermocycler:

A: Temperature ( $^{\circ}\text{C}$ )

B: Time

C: Cycles

Steps	Cycle step	Temperature ( $^{\circ}\text{C}$ )	Time	Number of cycles
A	Initial denaturation	98	1 m	1
B	Denaturation	98	30 s	30
B	Annealing	55	30 s	30
B	Extension	72	30 s	30
C	Final extension	72	5 m	1
D	Hold	4	$\infty$	1

### Thermalcycler Conditions

#### Note

The thermocycler conditions depend on the polymerase used. If you are using a different High Fidelity polymerase other than the Q5 Hot Start Master Mix, refer to the enzyme manual and adjust the thermocycler conditions accordingly.

If extremely necessary, it is possible to increase to 35 cycles to increase the amplicon concentration.

We recommend quantifying the concentration of amplified DNA and perform 1.5% agarose electrophoresis to assess for the presence of the amplified region.

The expected length of the V3 region is 232 nucleotides.

- 22** In this step, purify the amplicons using AMPure XP beads. Before beginning, bring the beads to room temperature with at least one hour before use, prepare 80% ethanol solution and keep it refrigerated at -20 °C.
- 23** Transfer the entire Amplicon PCR product to a 1.5 ml Eppendorf tube and add nuclease-free water until reaching a final volume of 100 µl.
- 24** Vortex the AMPure XP beads to homogenize them. Add 50 µl of AMPure XP beads to each sample tube. Gently homogenize the mixture by vortexing at a low intensity, and then incubate it at room temperature for 5 minutes.
- 25** Place the tube with the sample on a DynaMag™-2 magnetic rack until the mixture clears. Then, recover the supernatant by transferring it to a new 1.5 mL Eppendorf tube.
- 26** Add 15 µL of Ampure XP beads to the obtained supernatant. Then, homogenize in vortex for 1 min at low intensity and incubate the sample at room temperature for 5 minutes.

- 27** Place the tube with the sample in the magnetic rack until it clears and discard the supernatant.
- 28** While the tube is on the magnetic rack, add 200 µl of 80% ethanol to wash the beads without homogenizing. Remove the ethanol carefully and repeat this step once more for a total of two wash steps.
- 29** Maintain the tube with the sample on the magnetic rack and allow to air-dry for a maximum of 10 min until the ethanol evaporates.
- 30** Add 22 µl of nuclease-free water to the tube, homogenize, and incubate at room temperature for 5 minutes.  
Place the tube with the sample in the magnetic rack until it clears and recover the supernatant. Store the purified amplicon at -80 °C until use.

#### Note

The final volume may vary. This will depend on the sequencing service unit that will be used. We suggest referring to the Illumina manual for selecting the final elution volume.

[https://support.illumina.com/documents/documentation/chemistry\\_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf](https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)

- 31** Verify the integrity of the amplicon in a 2% agarose gel and quantify the final DNA obtained.

## V3 Metagenomic Library Preparation.

### **32** Index addition

The next step is to attach dual indices and Illumina sequencing adapters using the Nextera XT Index Kit

1- Transfer 5 µl from purified amplicon to a new Eppendorf tube.

#### Note

Nextera provides different sets of primers to prepare a different number of samples.  
Refer to the manual of your available set for reference.

**33** Add 5 µl Index 1 to the tube with purified amplicon.

**34** Add 5 µl Index 2 to the tube with purified amplicon.

**35** Set up the following reaction of DNA:

Reaction	Volume
Reagent Q5 Hot Start Master Mix 2X	25 µl
PCR Grade Water	25 µl
Nextera XT Index 1 Primers (N7XX)	5 µl
Nextera XT Index 2 Primers (S5XX)	5 µl
DNA template	5 µl
Final Volume	50 µl

**36** Gently pipette up and down 10 times to mix.

**37** Perform PCR on a thermal cycler using the following program:

Step	Step Cycle	Temperature	Time	Cycles
A	Initial denaturation	98	1 m	1
B	Denaturation	98	30 s	12
B	Annealing	55	30 s	
B	Extension	72	30 s	
C	Final extension	72	5 m	1
D	Hold	4	∞	1

### 38 **PCR Clean-Up**

This step uses AMPure XP beads to clean up the final library before quantification.

Centrifuge the Index PCR tube at 280 x g at 20 °C for 1 minute to collect condensation.

**39** Vortex the AMPure XP beads for 30 seconds to make sure that the beads are evenly dispersed. Add an appropriate volume of beads to a trough.

**40** Add 56 µl of AMPure XP beads to the Index PCR tube.

**41** Gently pipette mix up and down 10 times.

**42** Incubate at room temperature without shaking for 5 minutes.

**43** Place the tube on a magnetic stand for 2 minutes, or until the supernatant has cleared.

- 44** With the Index PCR tube on the magnetic stand, use a pipette to remove and discard the supernatant.
- 45** With the Index PCR tube on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:
- Using a multichannel pipette, add 200 µl of freshly prepared 80% ethanol to the tube.
  - Incubate the tube on the magnetic stand for 30 seconds.
  - Carefully remove and discard the supernatant.
- 46** With the Index PCR tube on the magnetic stand, perform a second ethanol wash as follows:
- Using a multichannel pipette, add 200 µl of freshly prepared 80% ethanol to the tube.
  - Incubate the tube on the magnetic stand for 30 seconds.
  - Carefully remove and discard the supernatant.
  - Use a pipette with fine pipette tips to remove excess ethanol.
- 47** With the Index PCR tube still on the magnetic stand, allow the beads to air-dry for 10 minutes.
- 48** Remove the Index PCR tube from the magnetic stand.
- 49** Using a pipette, add 27.5 µl of H<sub>2</sub>O to the Index PCR tube.
- 50** Incubate at room temperature for 2 minutes.
- 51** Place the tube on the magnetic stand for 2 minutes, or until the supernatant has cleared.

**52** Using a pipette, carefully transfer 25 µl of the supernatant from the Index PCR tube to a new eppendorff tube.

### **53 Library Quantification**

Illumina recommends quantifying your libraries using a fluorometric quantification method that uses dsDNA binding dyes.

Calculate DNA concentration in nM, based on the size of DNA amplicons as determined by an Agilent Technologies 2100 Bioanalyzer trace:

$$(\text{concentration in ng/}\mu\text{l}) \times 106 = (660 \text{ g/mol} \times \text{average library size})$$

For example:

$$15 \text{ ng/}\mu\text{l} \times 106 = (660 \text{ g/mol} \times 500)$$

concentration in nM

45 nM

#### **Note**

Library quantification is typically calculated as part of the sequencing facility services. However, if you need to perform it yourself, follow this step as usual.

The final amount of purified library necessary for sequencing depends on the sequencing depth required for each sample and the sequencing platform. Please refer to the person in charge of the sequencing service to confirm the requirements.

**54** Sequencing the sample in 2x150 pb for the V3 region is necessary, with a recommended total number of reads of 100k to 200k per sample. The amount of DNA recommended in this protocol may vary between service units.

## **Analysis of the shrimp microbiome using DADA2 by Qiime2**

**55** After sequencing, data files in a FASTQ format containing the nucleotide sequences or reads are generated. A computer or server with Ubuntu 22.04 operating system or some other linux distribution to the user's taste should be used for the analysis of the reads. The sequences can be worked in the same way in macOS terminals or in Ubuntu terminals installed in windows with the WSL options enabled.





#### Note

Please review the materials section of the protocol where you will find the official pages of the programs to be used, as well as the suggested commands for their installation.

**56** Before starting, create a directory dedicated solely for the analysis of the reads.

```
$ mkdir workspace
```

Subsequently, enter the directory created and save the raw reads.

```
$ cd workspace  
$ mkdir raw_seqs
```

Raw reads should be stored in the "raw\_seqs" folder.

#### Note

The names of all the work folders from this point onwards are at the user's choice, the ones mentioned are examples to continue with the analysis. All commands from this point onwards will begin with the **\$** symbol.

**57** Perform a quality visualization of the reads from FASTQ file with fastQC.

```
$ fastqc Archive_R1.fastq Archive_R2.fastq
```

or

```
$ fastqc Archive*fastq
```

## Note

You should view the result in HTML format in the web browser of your choice. In it, it is necessary to note the average quality of the sequences, the range of sizes of the sequences, the sectors of worst quality in the sequences, the presence of adapters, and the repetition of nucleotides in the readings, as this will be important at the moment of Cleaning.

## 58 Remove NEXTERA adapters with Cutadapt.

```
$ cutadapt -b "file:adapter_list.fasta" -o Archive_no_adapters_R1.fastq Archive_R1.fastq
$ cutadapt -b "file:adapter_list.fasta" -o Archive_no_adapters_R2.fastq Archive_R2.fastq
```

The file with the Nextera adapters list (adapter\_list.fasta) needs to be created with the following sequences:

```
>PrefixNX/1
AGATGTGTATAAGAGACAG
>PrefixNX/2
AGATGTGTATAAGAGACAG
>Trans1
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
>Trans1_rc
CTGTCTCTTATACACATCTGACGCTGCCGACGA
>Trans2
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
>Trans2_rc
CTGTCTCTTATACACATCTCCGAGCCCACGAGAC
```

## Note

In the FastQC results, you can determine whether your sequences contain adapters and what they are. With this information, you can search for lists of oligonucleotides and create a fasta file that you will use to feed the program. If you only need to remove a known oligo, it is possible to run the program in the following way:

```
$ cutadapt -b ACGTACGT -o Archive_no_adapters_R1.fastq Archive_R1.fastq
$ cutadapt -b ACGTACGT -o Archive_no_adapters_R2.fastq Archive_R2.fastq
```

- 59** Use the Trimmomatic tool to quality-trim your reads. Before selecting any program option, carefully review the characteristics of your reads in the fastQC results. The files can be used in a compressed .gz format.

```
$ TrimmomaticPE -phred33 -threads 8 Archive_no_adapters_R1.fastq
Archive_no_adapters_R2.fastq Archive_trimmo_results_R1.fastq unpaired_reads_R1.fastq
Archive_trimmo_results_R2.fastq unpaired_reads_R2.fastq OPTION1:<value> OPTION2:
<value>
```

The resulting unpaired files will not be used in the following steps, only the "trimmo" paired files.

## Note

### Recommendations:

If the quality at the end of the sequence is less than 20, you can perform a sequence cut using the "CROP:" option, where you should enter the value of the maximum length to which the sequences should be trimmed. For example, if your sequences are 150 nts, and you want to remove the last 12 nts, enter CROP:138. Later, for sequences with these characteristics, it is recommended to use the "SLIDEWINDOW:" command to perform cuts by quality windows. For example, if we add the command "SLIDEWINDOW:5:15", the program will analyze every 5 nucleotides of the sequence looking for an average quality of 15, and if it is lower, it will cut those nucleotides from the sequence.

If your reads are of good quality, you can simply use the "AVGQUAL:" command with recommended values of 18, 20, or 22, as this option will only retain the reads that exceed this average quality value. If you have low quality at the beginning of your reads, you can perform a cut using the "HEADCROP:" option and add the number of nucleotides to be cut.

Additionally, you can filter by the minimum sequence size you want for your reads using the "MINLEN:" option, where it is recommended to use a value of 100 for reads of 2x150, 150 for reads of 2x200, and 200 for reads of 2x250. The values for each program option should be used according to the characteristics of your reads and at your discretion.

**Optionally**, although it is recommended, you can use the Prinseq-lite program with the entropy filtering method to discard low complexity sequences from your reads. These include sequences with 1, 2 or 3 nucleotide repeats throughout the sequence. The result is found in the file Archive\_R1-2\_good\_xxx.fastq. The command is as follows:

```
$ prinseq-lite -fastq Archive_trimmo_results_R1.fastq -fastq2  
Archive_trimmo_results_R2.fastq -lc_method entropy -lc_threshold 68 -out_format 3
```

If you wish, this program can also be used to perform quality processing of your sequences. To do so, please refer to the manual on the website, which you can visit using the link described in the materials section of this protocol, or alternatively by using the command "prinseq-lite -h".

60

Once your sequences have been quality-trimmed, modify the file names of your reads according to the following structure (this is mandatory):

### Name Structure:

1                      2                      3           4   5   6

SampleName\_AAAAAAAAA\_L001\_R1\_001.fastq

- 1- Name of the sample.
- 2- 8-letter code that must be unique for each sample (it must be the same for the R1 and R2 pair).
- 3- This should be as shown for all sequences.
- 4- Sequence R1 or R2.
- 5- This should be as shown for all sequences.
- 6- The format in fastq.

Then the sequences should be compressed to "gz" format.

- 61** Create a metadata.txt file in a text table format that includes the sample names, labels, sample origins, and any comparisons you need to develop. Here's an example template you can use:

A	B	C	D	E
#SampleID	BarcodeSequence	Group	Treatment	Organ
sample_01	AAAAAAAAA	1	1	Gut
sample_02	AAAAAAAC	2	2	Gut
sample_03	AAAAAAAG	3	3	Gut

### Metadata Structure

#### Note

**Only the column title "#SampleID" should not be modified.** You can add more rows for additional samples and fill in the corresponding information accordingly. Make sure to separate the columns using tabs or spaces. Once you have filled in the metadata information for your samples and comparisons, save the file in a text format, such as a ".txt" or ".tsv" file, and it will be ready to use with Qiime2.

- 62** Now, let's start using the Qiime2 tool.

Activate the conda environment of qiime2:

```
$ conda activate qiime2-2023.5
```

Import the reads to the qiime2 program:

```
$ qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path  
quality_reads/ --input-format CasavaOneEightSingleLanePerSampleDirFmt --output-path  
work_space/demux-paired-end.qza
```

Performs the process of denoising the reads to obtain the ASVs:

```
$ cd work_space/  
$ qiime dada2 denoise-paired --i-demultiplexed-seqs demux-paired-end.qza --p-trunc-len-f 0 --p-  
trunc-len-r 0 --o-table PE-table.qza --o-representative-sequences PE-rep-seqs.qza --o-  
denoising-stats PE-stats.qza
```

#### Note

##### Recommendations:

If the result of the read merging is low, you can modify the number of nucleotides that should overlap using the option:

```
p-min-overlap <value>
```

By default, the program takes a value of 12. We do not recommend selecting values lower than 6 for overlap.

If the quality of the reads continues to be a problem, you can modify the values of the maximum expected error filter with the following option:

```
--p-max-ee-f <value> --p-max-ee-r <value>
```

By default, the program takes a value of 2.0. We recommend initially using a value of 5.0 for the reverse reads, as these sequences usually have a lower average quality compared to their counterpart.

Converts the resulting files to "qzv" format, which can be directly viewed on the page <https://view.qiime2.org>

```
$ qiime metadata tabulate --m-input-file PE-stats.qza --o-visualization PE-stats.qzv
```

```
$ qiime feature-table summarize --i-table PE-table.qza --o-visualization PE-table.qzv --m-sample-metadata-file metadata.txt
```

```
$ qiime feature-table tabulate-seqs --i-data PE-rep-seqs.qza --o-visualization PE-rep-seqs.qzv
```

#### Note

The PE-stats.qzv file shows step-by-step statistics of the samples during their processing to obtain ASVs. The PE-table.qzv file indicates the characteristics of the samples, of which the most important data is the size of the smallest sample, as it will be a value to be used when performing rarefactions of them. Finally, the PE-rep-seqs.qzv file shows the list of unique ASVs present in the experiment and their statistics.

## 63 From here on, we will proceed with the taxonomic annotation of your ASVs.

Train a classifier for the 16S ribosomal gene. Obtain the necessary sequences directly from the SilvaDB website. We recommend using the nr99\_ref sequences, always in their latest version.

```
$ qiime feature-classifier fit-classifier-naive-bayes --i-reference-reads  
silva_version_nr99_ref_seqs.qza --i-reference-taxonomy silva-version-99-tax.qza --o-classifier  
silva_version_nr99_classifier.qza
```

### Note

We use the Silva 138.1 database version for this protocol.

It is possible to delimit the database to the region that is going to be studied, which is recommended to reduce the computation time. For this, the step to follow before generating the classifier is the following:

```
$ qiime feature-classifier extract-reads --i-sequences silva_version_nr99_ref_seqs.qza --p-f-primer 16s_region_forward --p-r-primer 16s_region_reverse --o-reads silva_version_region_nr99_ref_seqs.qza
```

X- It is the size of the fragment corresponding to the product of the oligonucleotides of the region to be analyzed.

Perform the annotation of your ASVs with the following commands:

```
$ qiime feature-classifier classify-sklearn --i-classifier silva_version_nr99_classifier.qza --i-reads PE-rep-seqs.qza --o-classification PE-taxonomy_silva_version.qza
```

```
$ qiime taxa collapse --i-table PE-table.qza --i-taxonomy PE-taxonomy_silva_version.qza --p-level 7 --o-collapsed-table table_taxa_PE.qza
```

```
$ qiime taxa barplot --i-table PE-table.qza --i-taxonomy PE-taxonomy_silva_version.qza --m-metadata-file metadata.txt --o-visualization PE-taxa-bar-plots_silva_version.qzv
```

### Note

The file PE-taxa-bar-plots\_silva\_version.qzv can be viewed at <https://view.qiime2.org>, where you can compare the taxonomy based on the groups you have created in the metadata. You can also download these groups in tsv format for further processing as needed.

## 64 From this step onwards, we will begin calculating alpha and beta diversity.

Create a rooted tree of the ASVs using the following commands:

```
$ qiime alignment mafft --i-sequences PE-rep-seqs.qza --o-alignment PE_alignment_taxonomy.qza
```

```
$ qiime alignment mask --i-alignment PE_alignment_taxonomy.qza --o-masked-alignment PE_masked_alignment.qza
```



```
$ qiime phylogeny fasttree -i-alignment PE_masked_alignment.qza --o-tree  
PE_unrooted_tree.qza  
$ qiime phylogeny midpoint-root -i-tree PE_unrooted_tree.qza --o-rooted-tree  
PE_rooted_tree.qza
```

Calculate the different beta diversity parameters with the following command:

```
$ qiime diversity core-metrics-phylogenetic --i-phylogeny PE_rooted_tree.qza --i-table PE-  
table.qza --p-sampling-depth <90% of the number of reads of the smallest sample in PE-  
table.qzv> --m-metadata-file metadata.txt --output-dir core_metrics_results
```

#### Note

To calculate the significance test per group for Faith\_PD and Evenness, the following procedure is followed:

```
$ qiime diversity alpha-group-significance --i-alpha-diversity  
core_metrics_results/faith_pd_vector.qza --m-metadata-file metadata.txt --o-visualization  
core_metrics_results/faith_pd_group_significance.qzv  
$ qiime diversity alpha-group-significance --i-alpha-diversity  
core_metrics_results/evenness_vector.qza --m-metadata-file metadata.txt --o-visualization  
core_metrics_results/evenness_group_significance.qzv
```

Perform the necessary statistical tests for the beta diversity analyses you require using the following command:

```
$ qiime diversity beta-group-significance --i-distance-matrix core_metrics_results/beta_test.qza --  
m-metadata-file metadata.txt --m-metadata-column <Column of the metadata with which you  
want to search for statistical differences> --o-visualization  
core_metrics_results/beta_test_significance_statistics.qzv --p-pairwise --p-method <statistic  
test>
```

To calculate alpha diversity, use the following command:

```
$ qiime diversity alpha-rarefaction --i-table PE-table.qza --i-phylogeny PE_rooted_tree.qza --p-max-  
depth <90% of the number of reads of the smallest sample in PE-table.qzv> --m-metadata-file  
metadata.txt --o-visualization Alfa_test.qzv --p-metrics <alpha diversity test name> --p-iterations  
<number of iterations to perform>
```

## Note

For these tests, we recommend a minimum of 1000 iterations, although a quick analysis could work correctly with 100 iterations. To see all the tests that can be realized, we recommend viewing the manual by typing the command "\$ qiime diversity alpha-rarefaction".

All TSV format files from the tests performed can be downloaded from the <https://view.qiime2.org> website, starting from the qzv files. Likewise, you can also decompress the qzv files and extract the data yourself.

For differential expression analysis, we recommend using the programs Deseq2 or LDA score with LEfSE.

<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>

<https://huttenhower.sph.harvard.edu/galaxy/>