Ref. Ares(2023)1949248 - 17/03/2023

Deliverable 4.2
Standardized protocol for real-time microbiome monitoring in fish experimental facilities.

Version 1

WP 4
Deliverable 4.2
Lead Beneficiary: DTU
Call identifier: Biological and Medical Sciences - Advanced Communities: Research infrastructures in aquaculture
Topic: INFRAIA-01-2018-2019
Grant Agreement No: 871108
Dissemination level: PU
Date: 27.02.2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 871108 (AQUAEXCEL3.0). This output reflects only the author's view and the European Commission cannot be held responsible for any use that may be made of the information contained therein.

Contents

Objective

The purpose of this report is to describe the protocols successfully used to conduct fast microbiome monitoring in different types of fish experimental facilities using Oxford Nanopore Technologies (ONT) sequencing. This sequencing method was chosen in order to ensure fast result turn over, with the potential of obtaining results within 24 hours from sampling.

In this deliverable we aim to show easy methods for implementation, optimization, and harmonization of a microbiome monitoring protocol in different kind of fish experimental facilities (RAS, flow through, hatcheries), sample kinds (water, skin, mucus, intestine, gills, larvae and eggs), and fish species (rainbow trout, sea bass, sea bream).

The protocols and methodologies developed for this deliverable include all the steps in the process of microbiome monitoring, from sampling to sequencing and data analysis.

The present task is part of the work package that aims to develop “Technological tools for improved experimental procedures.”

Background

Fish trials generally involve a large number of live animals and resources, and usually run for a long time in specialized experimental tank facilities. Although a number of physical and chemical parameters can easily be monitored daily (pH, temperature, oxygen in water, CO2 concentration, feed amount etc), biological parameters such as the microbiota associated with different experimental conditions are neglected, in part due to the difficulties inherent to this kind of monitoring in the field. In most cases, monitoring of the microbiome in water, fish, filters, etc., is carried out within the frame of research projects, where a number of samples are collected, and results are only obtained after weeks to months from the original sampling. This time lapse in part due to the sequencing technology used (Illumina), which involves sequencing in specialized laboratories or companies, where a large number of samples need to be analysed simultaneously in order to reduce costs. This kind of methodology, although very efficient, provides results in a retrospective manner, so that what was happening in the experimental facilities at the time of sampling is only known months after. Although this is of high value in some research experiments, there may be other cases where faster monitoring of the microbiome in experimental facilities is required. The protocols developed for this deliverable are aiming to do just that, to implement technologies that allow a fast turn over between sampling and sequencing results, enabling more frequent (even daily) monitoring of the microbiome. To do so, we are using sequencing technologies that move away from the highly specialized sequencing centres, towards sequencing in regular wet laboratories, using relatively cheap and portable equipment that are still cost effective if only a few samples are sequenced at a time. The possibility of monitoring the microbiota in an experimental facility, with the subsequent increased yield of experimental data, contributes to increase standardisation across experiments, potentially allowing extrapolation of data to feed models and algorithm of virtual experiments, overall contributing to the 3Rs principals.

To fulfil this purpose, the platform of choice was the portable, 3rd generation sequencing platform Oxford Nanopore Technology MinION, which can generate data that can be examined in real time

during the sequencing run. The collaboration between the partners involved in this task resulted in a protocol considering various species, developmental stages, and setups, after testing and comparing different methods and samples from experimental tanks. The sample types include environmental and host-associated microbial communities, and among the latter, lethal and non-lethal samples. Therefore, the resulting protocol provides detailed guidelines aimed at giving instructions and common operating procedures that can be relevant for a range of experimental designs.

One limitation of the present deliverable is that no protocol standardization for sample collection and DNA extraction was achieved among participant laboratories. Instead, each laboratory have tested a number of procedures that could be easily used with their current equipment and routines. In this way we have identified protocols that provide DNA of enough quality for ONT library preparation. Once the DNA samples of good quality is obtained, a common standardized protocol for PCR amplification, library construction and sequencing was followed.

Methods

In the elaboration of these protocols we have included samples collected in different fish experimental facilities and from different fish species and developmental stage (Table 1). For each system/species, different kinds of biological samples were included.

Table 1. Species, system type and developmental stages for which ONT microbiome sequencing was tested.

Partner	Species	System	Freshwater/seawater	Developmental stages
CCMAR	Gilthead sea bream & European sea bass	RAS (Hatcheries)	Seawater	Embryo (eggs), larvae (Flexion, mid-metamorphosis)
CSIC	Sea bream	Flow through	seawater	juvenile, adult
DTU	Rainbow trout	RAS & Flow-through	Freshwater	Fingerling, Adult
WU	European sea bass	RAS	Seawater	Juvenile

Sample collection

Collection and DNA extraction strongly depends on the kind of samples of interest. Here we worked with environmental samples (water and tank sediment samples) and with gut mucosa and digesta, as well as skin mucus. Standardization among all the participants was not possible regarding sample collection and DNA extraction, as different laboratories work with different conditions, consumables and equipment. Here we are indicating protocols that we have tested with good results in the different laboratories (**for more detailed information and step by step protocols from the different laboratories see Appendix 1**), and from which we obtained DNA of the required quality and quantity to library construction and sequencing using Oxford Nanopore Technologies.

Microbiome samples were collected according to the following guidelines as depicted in Figure 1. Although these methods were developed with a focus on experimental tank facilities, they can be modified in the case of field sampling. In particular, the preservation of the samples by snap-freezing could be challenging when sampling outside of the laboratory. For this reason, it should be considered the use of storage solutions (e.g., RNAlater, DNA/RNA Shield) and transportation at 4°C until storage at –80°C or extraction is possible. As for every other step, consistency in the transportation and preservation method is crucial.

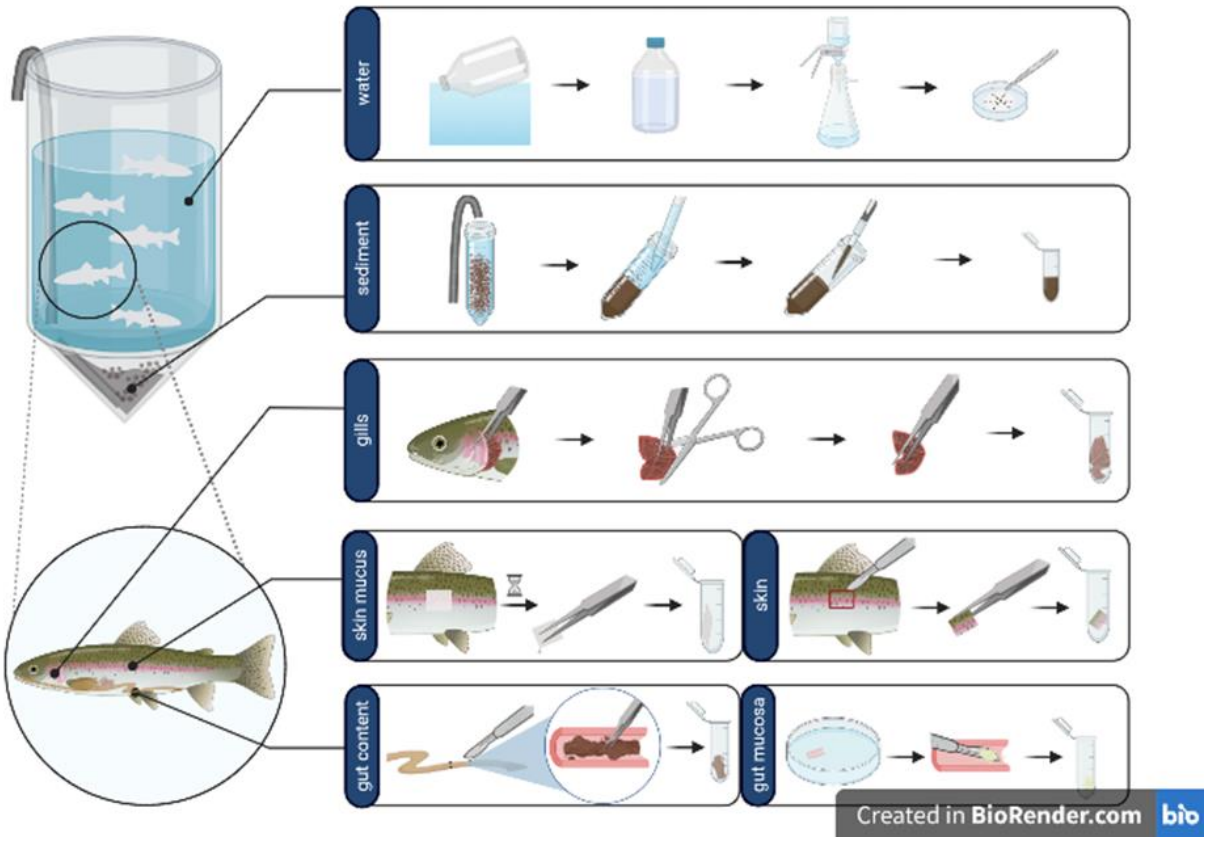

Figure 1. Schematic representation of various sample types and respective collection methods from experimental tank facilities. From top to bottom: water, sediment, gill, skin, and gut samples collection.

Water samples (filters) - 10 minutes to 1 hour per sample

Filtering times are dependent on the available filtration system, pump, turbidity and volume of the filtered water. At least two duplicates per tank are recommended.

Up to 1 liter of water is filtered per sample. In some systems the filter clogs faster and no more than 200 to 500 ml can be filtered. This does not affect results.

Filtration is done using a vacuum system through a PVDF membrane filter of 47mm size and 0.22 µm pore size.

Filters are either placed in a Petri dish or cut into smaller pieces using sterile scissors and placed in sterile tubes.

For long-term storage, filters are placed at -80°C.

Water samples (bacterial pellet) – 20 minutes

400 ml – 800 ml seawater was collected into sterile 500-ml flasks and centrifuged for 10 min at 16,100 x g and 4ºC.

The pellet for each sample was resuspended in 2 to 5 ml of RNA-later preserving solution, transferred to a 2-ml microcentrifuge tube, and stored at -20ºC until analysis.

Sediment samples – 15 minutes per tank

Organic sediment is collected from the bottom of the tank and 1.5 mL of sediment is placed in 2 mL Eppendorf tubes in triplicates.

Samples are either immediately processed or stored at -80°C.

Samples of gut mucosa and digesta – 2 hours for 10-12 samples

After euthanizing, the fish is placed on a flame-sterilized surface. Open the abdominal cavity and cut a section of the gut of interest (see Appendix 1 for details). Size of the section depends on fish size, if possible, a 2 to 4 cm piece of the gut should be collected.
On a sterile Petri dish, open the intestine and carefully remove any remaining content. If interested in the transient gut microbiome, collect the digesta in a sterile 2 mL tube and snap freeze.

For adherent microbiota (mucosa), wash the intestinal section with cold, sterile PBS using a plastic Pasteur pipette to remove non-adherent bacteria or any remaining material. Place the washed intestine on a clean Petri dish. Gently scrape the intestinal epithelium with the blunt end of a clean/flamed scalpel to obtain the mucus. Place the mucus directly from the scalpel in a sterile Eppendorf tube and snap freeze.

Alternatively, gut mucus can be collected using Isohelix swabs.

Samples of skin mucus – 5 minutes per sample

Skin mucus can be collected gently scrapping the side of the fish in the direction of the scales using either sterile cell scrapers or a coverslip. The accumulated mucus is collected with pipettes and placed in labeled sterile Eppendorf tubes.

Alternatively, a piece of absorptive material (e.g., microscope cleaning wipes) is placed on the chosen sampling area. After absorption, the mucus sample is lifted with sterile tweezers and placed in a sterile 2 mL tube.

In case this is not possible, samples should be stored at -80ºC.

Fish tissue samples (skin and gills) – 5 minutes per sample

Small pieces of tissue of interest – here we have tested skin and gill tissue- are placed in a sterile 2 mL tube which is snap-frozen. Alternatively, a preservation reagent such as RNA later can be used.

Fish eggs and larvae – 20 minutes per sample

Sea bass and sea bream larvae (ranging in age from 5 to 77 dph) were collected with a sterile beaker from their tanks.

The water containing the larvae and was filtered through a 20 µm filter, rinsed with sterile filtered sea water (0.22 µm filter) and collected into 15-ml Falcon tubes containing 10 volumes of RNA-later. Samples were stored at -20ºC until extraction.

Eggs are collected from the surface of the tank water with a clean beaker (500 ml) and pouring the seawater containing eggs, through a clean plankton net previously cleaned with ethanol and then sterile seawater.

The eggs are thoroughly rinsed with several volumes of sterile, filtered seawater and transferred using a sterile Pasteur pipette into 10 volumes of RNA later (1 volume of eggs/10 volumes of RNA later). Samples were stored at -20ºC until extraction.

Control samples

Negative control samples (reagents only) were extracted along with experimental samples to assess presence of contamination. Additionally, a mock sample to assess DNA extraction bias was extracted in each extraction batch. In this case ZymoBIOMICS™ Microbial Community Standard (cat. D6300) was used, but alternative mock community standards can be used.

DNA extraction

Adequate DNA extraction is crucial for successful sequencing. DNA recovery was evaluated in yield, integrity, and purity. In table 2 it is possible to see the different kits used to recover good quality DNA from the different kind of samples tested. Some samples can be extracted using more than one kit, and some kits can be used for more than one sample type.

Table 2. DNA extraction method used for obtaining DNA in enough quantity and quality for successful library construction and sequencing.

Sample kind	Sample collection	DNA extraction
Water	200 ml to 1 l water filtered through 0.22 µM pore size filter	MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems) + lyzosyme digestion OR DNeasy® PowerSoil® Pro Kit (QIAGEN)
Sediment	Sediment in 2 ml collection in tubes	MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems) + lyzosyme digestion
Gut mucosa & digesta (transient or adherent microbiota)	Mucus collected from foregut or hind gut or Digesta	MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems) + lyzosyme digestion OR High Pure PCR Template Preparation Kit (Roche) + lysozyme digestion
Skin mucus	Collected by scrapping or using absorptive paper.	MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems) + lyzosyme digestion OR High Pure PCR Template Preparation Kit (Roche) + lysozyme digestion
Fish tissue (gills and skin)	Small pieces collected in a 2 ml tube.	MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems) + lyzosyme digestion

Eggs and larvae	30 mg of eggs and larvae (pools depending of species and age)	DNeasy Blood and Tissue Kit (Qiagen) + Lysozyme digestion

Protocol 1 – DNA extraction using the MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems) plus lysozyme digestion.

Duration: 2 hours for up to 96 samples.

Starting material:

300 mg of starting material is used, this could be tank sediment, gut mucus and digesta, or fish tissue.

For water samples a whole filter is used. For skin mucus collected in absorbent paper the whole paper (2-4 cm2) is used. In both cases the filter or paper is cut in small pieces or folded into a 2 mL Eppendorf tube.

1. For all samples 800 mL of lysis buffer and zirconia beads (included in the kit) is added. VARIATION FOR TISSUE SAMPLES: together with the zirconia beads, add a 5 mm metal bead.
2. Disruption is done by bead beating at 25 Hz for 5 minutes (TissueLyser II or similar).
3. The sample is digested with lysozyme 0.9 mg/mL (Sigma cat. 62970) for 1 hour at 37°C on a thermal block.
4. The lysates are centrifuged at 14000 g for 10 minutes and the supernatant is collected in a new sterile tube.
5. Proceed to DNA extraction according to the manufacturer's instructions.

   Note: this protocol was tested on a KingFisherTM Flex System (ThermoFisher cat. 5400630), using the MagMAX_Microbiome_Soil_Flex program.
6. Elution volume was changed from 50 µL to 75 µL.

The quality and quantity of the DNA is assessed, and samples are stored at -20 °C if not prepared for sequencing immediately.

Protocol 2- DNA extraction from water samples using the DNeasy® PowerSoil® Pro Kit (Qiagen)

Duration: 2h

Starting material:

Water samples: Filters are cut in small pieces and put into a PowerBead Pro Tube (no more than 1 filter/tube). Add 800 µl of solution CD1 and mix by vortexing. Homogenize by bead beating (Fast Prep24 homogenizer (MP Biomedicals) cycles of 30s at 6 m/s).

Gut and skin mucus in Isohelix swabs: place the whole swab into a PowerBead Pro Tube. Add 800 µl of solution CD1 and mix by vortexing. Homogenize by bead beating at 3 m/sec for 3min (VWR BeadMill) repeated 3 times.

Centrifuge the PowerBead Pro Tube at 15000 xg 1 min and transfer 500-600 µl of supernatant to a clean 2 ml tube. No problem if some filter/beads particles are transferred to the clean tube at this step.

From this point onwards follow manufacturer's instructions (for detailed instructions see Appendix 1).

For filter samples elute in a final volume of 100 µl of elution buffer.

For swabs samples, elute in 50 µl of nuclease-free water preheated at 50°C.

The quality and quantity of the DNA is assessed, and samples are stored at -20°C if not prepared for sequencing immediately.

Protocol 3 - DNeasy Blood and Tissue Kit (Qiagen) with Lysozyme digestion

Duration: 4 hours for up to 24 samples

Starting material:

a) Fish eggs: 30 mg of fish eggs (a pool of 10 – 15 eggs, this will vary with species).

   Fish larvae: 10 – 30 mg of larvae (the number of larvae varied with age).

   Pools of eggs or larvae are added to 2 ml microcentrifuge tubes containing 400 µl of lysis buffer and two iron beads (Qiagen stainless steel beads of 5 mm).

b) Bacterial pellet (from water samples):

   400ul of lysis mix to bacterial pellet is added to the bacterial pellet and put into a PowerBead Pro Tube, resuspend by vortexing.

Procedure:

1. <u>For egg and larvae samples only:</u> Mechanical disruption is done by bead beating with 3 cycles of 30 sec at 30 Hz; (Tissue Lyser, Qiagen). The iron beads are removed after the first disruption.

<u>Common for all samples:</u>

2. 400 mg of 0.1 mm zirconia/silica beads per tube are added. Sample is kept on ice until disruption.
3. Mechanical disruption by bead beating with 3 cycles of 30 sec at 30 Hz; (Tissue Lyser, Qiagen) is performed.
4. Samples are incubated for 30 min at 37 ºC with lysozyme (80 µl of 100 mg/ml lysozyme per sample).
5. Lysates are centrifuged for 1 min at 4,300 g and the supernatant is collected into a clean microcentrifuge tube.
6. Samples are treated with RNase (10 µl of 10 mg/ml) for 10 min at room temperature.
7. Continue with manufacturer's instructions (Details in Appendix 1).
8. Final elution is done in 170 µl of Tris-HCl (10 Mm, pH 8)

The quality and quantity of the DNA is assessed, and samples are stored at -20 °C if not prepared for sequencing immediately.

Note: for bacterial pellet from the water samples the DNA concentration is low (< 5-10 ng/ µl). To increase the gDNA concentration, use the QIAamp DNA Micro kit (Qiagen) following the supplier's protocol and eluted with 20 µl of Tris-HCl (10Mm, pH 8).

Protocol 4- DNA extraction from digesta, gut mucus and skin mucus samples using the Pure PCR Template Preparation Kit (Roche) + lysozyme digestion

Duration: 2h

Starting material: 200 µl of intestinal mucus, skin mucus samples or digesta.

Samples are treated with 250 µg/mL of lysozyme (Sigma) for 15 min at 37°C.

DNA is extracted following the manufacturer's instructions (for detailed instructions see Appendix 1). Final elution in 100 µl.

The quality and quantity of the DNA is assessed, and samples are stored at -20 °C if not prepared for sequencing immediately.

PCR amplification, library construction and sequencing

Procedures for library construction and sequencing are common for all kinds of samples. Library construction and sequencing has been tested with DNA from water, sediment, fish gut mucosa and skin mucus samples. Step by step details for library construction and sequencing can be found in Appendix 1.

Quality control and preparation for 16S rRNA sequencing

DNA purity was checked by Nanodrop measurement and assessment of absorbance ratios. DNA was considered sufficiently pure if 260/280 ratio was between 1.8 and 2.0, and 260/230 ratio was over 2.

Genomic DNA integrity was visually inspected by electrophoresis on an agarose gel.

DNA concentration was measured either by Qubit dsDNA HS Assay using a Qubit 4 Fluorometer or by Quant-iT™ PicoGreen™ dsDNA (Invitrogen).

PCR amplification and library preparation - 3 hours

Using the 16S Barcoding Kit 1-24 (SQK-16S024, Nanopore) PCR amplification and library preparation are performed together. Standard protocols were followed with some modifications:

1) The initial quantity of DNA was raised to 50 ng.

2) PCR amplification cycles were raised to 30.

All procedures were performed including a blank PCR tube (negative control) and a tube with a control mock community (10 ng DNA from Mock community ZymoBIOMICS Log II).

Amplicon products are cleaned with 30 µL AMPure XP beads and quantified (in these protocols we have used either the Qubit dsDNA HS or the Quant-iT™ PicoGreen™ dsDNA).

We pooled the barcoded libraries to a total of 50-100 fmoles in 10 µL of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl and added 1 µL of RAP.

After incubation at room temperature for 5 min, the libraries were placed on ice until loading.

Sequencing

Priming and loading of the MinION flow cell is performed exactly as described in the Nanopore protocol (16S Barcoding Kit 1-24 (SQK-16S024)).

Libraries are sequenced using an R9.4/FLO-MIN106 flow cell and by using the MinKNOW v21.10.4 and Guppy v5.0.11 for basecalling and demultiplexing.

It is crucial to know the complexity of the samples in order to determine the target number of reads/sample. To this purpose, a test sequencing run can be performed and rarefaction curves can be fitted (with e.g. R package vegan). If rarefaction curves plateau, the number of reads is adequate for the sample.

Of course, the number of reads needed to reach plateau in the rarefaction curves analysis is higher for more complex samples.

The maximum yield in terms of bases is 50 Gb for each flow cell if sequencing is run until flow cell exhaustion (e.g., 72 h). Therefore, the maximum number of libraries that can be multiplexed and the sequencing time will change depending on the target number of reads/sample.

Concerning sequencing time, as an example, we estimated that the time to reach 200,000 total reads in a R9.4/FLO-MIN106 flow cell can span between 1 and 4 hours.

Bioinformatic pipelines for Nanopore sequences

Three alternative pipelines were tested. Results are consistent among pipelines. Details of each pipeline and parameters used can be found in Appendix 1. Time estimated for the different pipelines is 3-4 hours, but it may vary depending on the available computing resources. When choosing the bioinformatic pipeline to use, if speed is a major concern we recommend using the fast basecalling protocol. The high accuracy basecalling protocol takes significantly more time to complete, and largely depends on the GPU type.

Pipeline 1: MinKNOW v21.10.4 and Guppy v5.0.11 were used for basecalling and demultiplexing. FASTQ reads were processed using Porechop v0.2.4 (https://github.com/rrwick/Porechop) for removing sequencing adapters from reads, NanoFilt[1] for filtering reads below 1,200 base pairs (bp), and > above 1,800 bp and yacrd[2] for chimeras detection and removal. Sequences were subsequently aligned for taxonomy assignment with Minimap2[3], using the SILVA[4] v 138.1 as the reference database. Then, the results were filtered and summarized in a bacterial amplicon sequence variants (ASVs) table. ASV tables are then imported and manipulated using R[5] to plot rarefaction curves, perform diversity metrics analyses and statistical tests (phyloseq[6] package, vegan[7] package).

Pipeline 2: Reads are retrieved and all FAST5 files are basecalled with GPU accelerated High Accuracy basecalling (Guppy v6.0.1). At the same time, barcodes are detected and trimmed, as well as primer and adapters, and reads are demultiplexed. -- *detect_mid_strand_barcodes* parameter is used to detect chimeras. The default minimum q-score (9) was used. The reads are filtered by length (awk text processing utility). Emu taxonomic classifier[8] (v3.3.1) is used against the default emu database. Species tables are then imported and manipulated using R[5] to plot rarefaction curves (vegan[7] package), perform diversity metrics analyses and statistical tests (phyloseq[6] package).

Pipeline 3: Reads are retrieved and all FAST5 files are basecalled with GPU accelerated with High Accuracy basecalling (Guppy v6.0.1). At the same time, barcodes are detected and trimmed, as well as primer and adapters, and reads are demultiplexed. -- *detect_mid_strand_barcodes* parameter is used to detect chimeras. The default minimum q-score (9) was used. The METONTIIME platform was used to process the raw reads[9]. Sequences were aligned for taxonomy assignment with BLAST, using the SILVA[4] v 132 as the reference database. Species tables are then imported and manipulated using R[5] to plot rarefaction curves (vegan[7] package), perform diversity metrics analyses and statistical tests (microeco[10] package).

Results and Discussion

Every partner collaborating in the work package has a different field of expertise in terms of experimental settings, developmental stages, and species they work with. Although such heterogeneity revealed the lack of a unique protocol to process all sample types, it allowed to develop specific guidelines for each kind of sample.

Aquaculture and fish microbiome studies in general represent a relatively new object of research. Thus, unlike in human gut microbiome research[11] or environmental ecology[12], no standardization efforts have been put in place yet, leading to challenges when considering meta-analyses studies.

First, tested sampling methods were described for various specimens. Seen the challenges posed by host DNA contamination, the host-associated microbial communities sampling steps try to exclude as much host tissue as possible. Amplification of the 16S rRNA gene (V1-V9) can be impaired when sampling too much host tissue; for instance, PCR amplification can be inhibited when using a full section of intestinal tract instead of mucosal scrapings.

Subsequently, pre-processing of samples before extraction was optimised. It involves homogenization and microbial cells disruption by bead beating or enzymatic lysis by lysozyme treatment, or a combination of both. Depending on the sample type, the homogenization and disruption step was differentiated.

The choice of extraction kit is important since it was shown that the bias introduced by the extraction kit reagents is significant[13] in microbiome analyses. Other than the kits that were tested in this work (see Methods), there are various commercial solutions aimed at extracting total microbial nucleic acid content. Generally, only testing and subsequent quality control steps will inform on the adequacy of the extraction kit for the sample of interest. Another essential QC step for informed extraction kit choice consists in testing PCR amplification prior library preparation (see Methods).

Regarding the alternative library preparation kit, except the standard 16S Barcoding Kit 1-24 (ONT, cat. SQK-16S024), we tested the PCR Barcoding kit (SQK-LSK109 WITH EXP-PBC096). In comparison with the 16S barcoding kit, the PCR barcoding kit requires additional time for the library preparation, of approximately 1.5 h, due to the barcoding step. Despite that, this can be considered as a cost-effective alternative as more samples can be sequenced at the same time, thus reducing the costs per sample (from 28 Euros with 24 barcodes to 9 Euros with 96). The same water sample was compared

using the two different kits, and the results will be shared at a later stage with all the task partners, as bioinformatic analysis are still in progress at the date or writing the present deliverable.

When it comes to bioinformatic analysis from Nanopore sequences, specific protocols and tools are being continuously developed. However, compared to short reads sequencing, there is still a lack of standardization regarding complete bioinformatic pipelines since ONT is a relatively new technology. ONT offers its own user-friendly workflows, contained in EPI2ME Labs, which enable less-experienced users to perform the bioinformatics analysis independently, even without computing capacity or familiarity with command line-based programs.

Although ready-to-use and faster, EPI2ME 16S rRNA analysis workflow comes with limitations in terms of parameters customization. Therefore, we tested two alternative bioinformatics pipelines, both based on Minimap2[3], for taxonomic classification. We also tested two different databases (SILVA[4] and a custom database constituted of NCBI-RDP[8]). Raw sequencing results coming from different laboratories and sample types (water, sediment, gills sequencing data) were shared among partners to test the workflows. Surprisingly, the two pipelines led to substantially analogous results in terms of both taxonomical classification and abundance estimations. The major observable differences were due to alternative nomenclatures of the same taxa in the different databases (Figure 2). Hence, the two pipelines were considered equally valuable for pre-processing and taxonomical classification of ONT 16S rRNA sequencing data.

The aim of the work package put focus on speed of the optimised protocol, as the final goal is to gain fast information on the fluctuations of microbial communities in the system. All collaborators estimated the time required from sampling to results to be approximately 24 hours. This time span is significantly less time than any second-generation sequencing workflow developed so far.

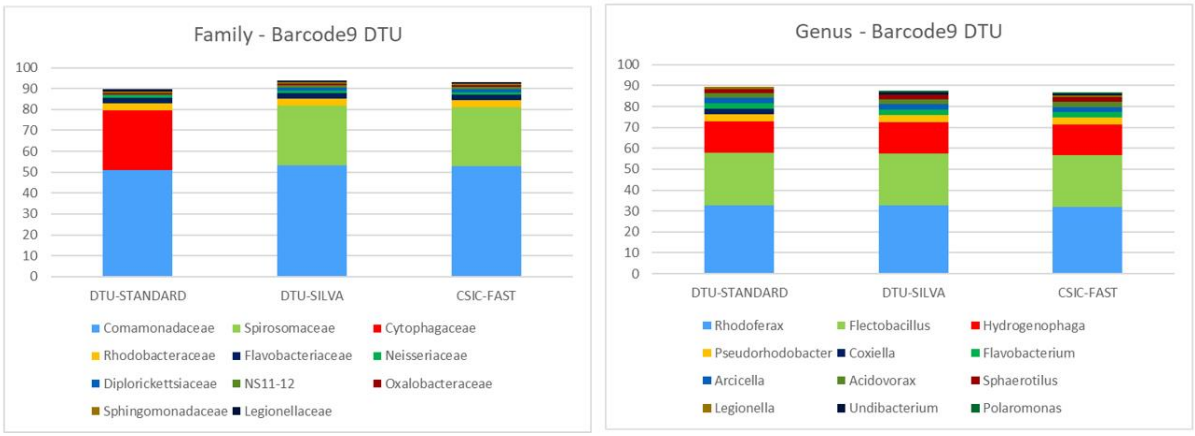

Figure 2. Bacterial composition of a water sample analysed with two different bioinformatic pipelines. Bar plots depicting family level (left) and genus level (right) bacterial compositions from one sample used for pipeline comparison. The three bars represent, from left to right, results from: Pipeline 2 and RDP-NCBI database; Pipeline 2 and SILVA database; Pipeline 1 and SILVA database. Bacterial compositions at both taxonomical levels seem to be the same using either pipeline. The

visible differences (e.g., *Cytophagaceae* vs *Spirosomaceae* in the left plot) are due to database annotations. Therefore, we concluded that the main source of differences is the database used for taxonomical assignments.

Conclusion

It is widely recognized across research fields (human, animal or environmental microbiomes research) that experimental methods should be standardized when comparing microbiome compositions in the same study. Although it is not possible to completely harmonize the methods needed for all sample types, we propose specific methods that are efficient in the identification and characterisation of various microbial communities that are relevant to aquaculture research.

We tested and validated the protocols from sampling to library preparation and sequencing for a diverse range of samples comprising both environmental (non-lethal) and host-associated specimens (both lethal and non-lethal). We took advantage of the differences in species, setups and developmental stages that constitute the field of expertise of each partner involved in the work package to build a comprehensive set of guidelines, that are therefore applicable in various experimental settings.

Although we focused on monitoring microbiomes in experimental tank facilities, the aforementioned methods can be easily implemented in the field with minor modifications.

References

1. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
2. Marijon, P., Chikhi, R. & Varré, J.-S. yacrd and fpa: upstream tools for long-read genome assembly. *Bioinforma. Oxf. Engl.* **36**, 3894–3896 (2020).
3. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma. Oxf. Engl.* **34**, 3094–3100 (2018).
4. Yilmaz, P. *et al.* The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res.* **42**, D643-648 (2014).
5. R: The R Project for Statistical Computing. https://www.r-project.org/.
6. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**, e61217 (2013).
7. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
8. Curry, K. D. *et al.* Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* **19**, 845–853 (2022).
9. Maestri, S. Development of novel bioinformatic pipelines for MinION-based DNA barcoding. (Universita' degli Studi di Verona, 2021).
10. Liu, C., Cui, Y., Li, X. & Yao, M. microeco: an R package for data mining in microbial community ecology. *FEMS Microbiol. Ecol.* **97**, fiaa255 (2021).
11. Tourlousse, D. M. *et al.* Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. doi:10.1186/s40168-021-01048-3.
12. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

13. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).

Appendixes

1) Amplicon sequencing protocol for aquaculture samples
2) Bioinformatic pipelines for 16S rRNA gene sequencing data

Document Information

EU Project	No 871108	Acronym	AQUAEXCEL3.0
Full Title	AQUAculture infrastructures for EXCELlence in European fish research 3.0		
Project website	www.aquaexcel.eu		

Deliverable	N°	D4.2	Title	Standardized protocol for real-time microbiome monitoring in fish experimental facilities.
Work Package	N°	4	Title	Technological tools for improved experimental procedures
Work Package Leader	Argelia Cuenca (DTU)			
Work Participants	Giulia Zarantonello (DTU), Niccolò Vendramin (DTU), Fotini Kokou (WU), CSIC: Carla Piazzon (carla.piazzon@csic.es) (CSIC), Socorro Toxqui (socorro.toxqui@csic.es) (CSIC), Fernando Naya-Català (fernando.naya@iats.csic.es)(CSIC), Jaume Pérez-Sánchez (jaime.perez.sanchez@csic.es) (CSIC), Ariadna Sitjà-Bobadilla (ariadna.sitja@csic.es) (CSIC)			

Lead Beneficiary	DTU (18)
Authors	Giulia Zarantonello, DTU, giuz@aqua.dtu.dk Argelia Cuenca, DTU, arcun@aqua.dtu.dk
Reviewers	Name, Organisation, Email address

Due date of deliverable	28.02.2023
Submission date	31.01.2023
Dissemination level	PU[1]
Type of deliverable	Other[2]

Version log			
Issue Date	Revision N°	Author	Change
03.02.2023	1	Finn Olav Bjørnson	first version

[1]Dissemination level (DELETE ACCORDINGLY): **PU:** Public, **CO:** Confidential, only for members of the consortium (including the Commission Services), set out in Model Grant Agreement, **CL:** Classified, information as referred to in Commission Decision 2001/844/EC

[2] Nature of deliverable (DELETE ACCORDINGLY): **R:** Report, **DEM:** Demonstration, pilot, prototype, plan design, **DEC:** Website, patent filing, market studies, press & media, videos, **Other:** Software, technical diagram, etc., **Ethics:** Ethics deliverable

17.02.2023	2	Wiegertjes, Geert	Second review