**Microbial research**

Environmental metagenomics is the study of organisms in a microbial community based on analysing the DNA or RNA within an environmental sample. Examples include profiling microbial populations in water samples taken from wastewater, deep ocean vents, or in soil samples from human-made environments like active mineral mines. Environmental metagenomics study data are used for agricultural microbiome analysis, ecological remediation, population disease monitoring or other biological investigations.

**Nucleic acids consideration**

When designing projects, it is worth considering which nucleic acid would be most useful for analysis. RNA is degraded quickly in the environment and may accurately reflect active cells and RNA viruses. The DNA/RNA Shield™ reagent (Zymo Research, available from Cambridge Biosciences, UK) is a DNA and RNA stabilisation solution which preserves the genetic integrity and expression profiles of samples at ambient temperatures, whilst completely inactivating infectious agents including viruses, bacteria, fungi and parasites. The solution prevents degradation from freeze-thaw cycling and enables shipping at ambient temperature.

DNA is more stable than RNA so dormant and dead cells may contribute to DNA isolated from the environment. A propidium monoazide (PMA) pretreatment step before PCR/library preparation can effectively exclude DNA from nonviable cells, as PMA can selectively diffuse through compromised cell membranes and intercalate with DNA to form DNA–PMA complex upon light exposure. The complex strongly inhibits the amplification of the bound DNA in PCR, and thus, only cells with intact cell membranes are detected.

**Isolation of Nucleic Acids**

Standard methods can be used for DNA or RNA isolation however the impact of the chosen method can only be determined if internal controls are used. Quality control standards can be included to estimate potential sources of error and bias from sample collection to sequencing. The potential for variation at each step in the microbiomic or metagenomic workflow is enormous. Microbial cell standards can be used to assess the DNA isolation bias, whereas Isolated DNA standards can be used to assess the library preparation and sequencing bias. <https://www.bioscience.co.uk/products/zymobiomics-microbial-community-standards>

In addition, nternal Spike-in for DNA or RNA library preparation can be used to calculate relative abundance and provide an additional level of control.

Correct selection of controls will ensure:

1. General benchmarking and provide a positive control for microbiome profiling
2. Assess detection limit and sensitivity
3. Absolute quantification of bacterial load (separate controls for high or low loads)

Please include negative controls at the nucleic acid isolation stage. These are easily overlooked, and many labs have contaminating species in unexpected places!

**Amplicon sequencing or whole metagenomic profiling**

Amplicon sequencing using conserved regions of rRNA gene as primer annealing sites from which to amplify variable regions of rRNA genes is a cost-effective method for family or species identification. However, over-amplification during PCR with increase chimera formation, whereas under-amplification may not capture the full community diversity. Using replicate PCR’s for each sample (3-4) will allow variability to be monitored and optimisation to be effective. There are reports that a single PCR per sample is sufficient for community profiling but a cautious approach should be taken when applying this under different experimental conditions.

The majority of amplicon sequencing use primers selected during previous research, such as the Earth Microbiome Project <https://earthmicrobiome.org/> which uses:

[Illumina 16S V4 Primer Constructs (515F–806R)](ftp://ftp.microbio.me/emp/protocols/primers/EMP_16S_V4_515F_806R_Parada_Apprill.xlsx)

[Illumina 16S V4–V5 Primer Constructs (515F–926R)](ftp://ftp.microbio.me/emp/protocols/primers/EMP_16S_V4V5_515F_926R_Parada_Quince.xlsx)

[Illumina ITS Primer Constructs (ITS1f-ITS2)](ftp://ftp.microbio.me/emp/protocols/primers/ITS1f_ITS2_Nov2016.xlsx)

[Illumina 18S Primer Constructs (Euk1391f-EukBr)](ftp://ftp.microbio.me/emp/protocols/primers/Euk1391f_EukBr_Nov2016.xlsx)

The protocol generates Illumina-barcoded libraries from a single PCR reaction.

Alternatively, Illumina amplicon library preparation uses a two-step PCR to generate barcoded sequencing libraries from 16S V3-V4 region (Klindworth A, Pruesse E, Schweer T, Peplles J, Quast C, et al. (2013). <https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf>

This is our standard method which has been modified to capture V1-V3, V4 alone, fungal ITS1 and ITS2, plantITS and insectITS for different projects. If you require replicates these will be treated as addition samples. The amount of reads required for amplicon sequencing can be as low as 20,000 per sample, although some researchers prefer to use more (100,000).

A third method used to prepare amplicon libraries for sequencing involves directly amplifying the region of interest using primers with barcodes at the 3’ end, pooling the barcoded PCR products before ligating Illumina adapters in the standard library preparation method. In this instance some consideration and optimisation may need to be undertaken to ensure the barcoded primers don’t introduce bias.

Nanopore technology is rapidly replacing Illumina for amplicon sequencing as the full 16s region is amplified and sequenced ~1.5kb will enable more accurate species identification, phylogenetic analysis and pathovar detection. On a per sample cost basis, it is highly competitive, although the amount of data generated is less.

Whole metagenomic sequencing (WMS) provides a complete picture of the entire metagenome and can be used for species identification, pathway analysis and antimicrobial resistance analysis. The high number of reads required for this will depend on the complexity of the community and the type of analysis required; for low coverage faecal samples 1Gb data / 3M 150 PE reads may provide an overview but sequencing of 60 M reads/sample is not unusual for deeper analysis. Nanopore sequencing may prove advantageous in this field with many single reads of 5kb-15kb or more, the amount of coverage required for assembly and analysis will be less. The disadvantage of WMG using Nanopore is the requirement for much higher quality and quantity of input DNA (1-5 ug).

General costs for Illumina and Nanopore sequencing can be found at <http://biosciences.exeter.ac.uk/sequencing/pricing/internal/>

Total costs are a combination of library preparation and sequencing so once you have a better idea of exactly what options you prefer, we can draw up some costs for you. We would want to know which library preparation you prefer and how much sequencing (Gb data or million paired end reads) is required.