DTU Metagenomics_Guidelines_ DTU_v2	**Metagenomics Projects Guidelines**	

Guidelines for Metagenomic Sequencing (Microbiome) Projects

Date 4. Nov. 2017

Prepared by: *Sünje Johanna Pamp*

Contact: *sjpa@food.dtu.dk*

Institution: *Research Group for Genomic Epidemiology, Technical University of Denmark*

Version: 2

The most recent version can be found at Figshare:

https://figshare.com/articles/Guidelines_for_Metagenomic_Microbiome_Sequencing_Projects_-_Item_K_/5245141

Reference

Kirstahler P, Bjerrum SS, Friis-Møller A, la Cour M, Aarestrup FM, Westh H., and Pamp SJ. (2018) Genomics-Based Identification of Microorganisms in Human Ocular Body Fluid. **Scientific Reports,** 8(1):4126. doi: 10.1038/s41598-018-22416-4.

Sünje J. Pamp
Technical University of Denmark

Kemitorvet
Building 204
2800 Kgs. Lyngby
Denmark

Tel +45 35 88 63 54

sjpa@food.dtu.dk
www.food.dtu.dk

Guidelines for Metagenomic Sequencing Projects

This list provides an overview of **considerations** and **controls** relevant for metagenomic sequencing projects. Project design, sample collection, and sample storage are discussed elsewhere. It is assumed that all samples of a particular project are being collected and stored in the same way.

For entirely new projects/samples it is recommended to perform a pilot experiment with a few representative samples to get insight into microbiome complexity and extent of differences between samples. Potentially, DNA extraction methods need to be optimized before extracting the main samples. For samples with a high diversity it may be considered to sequence a few samples fairly deep to estimate sequencing depth required for the main samples.

Before starting a metagenomic sequencing project, discuss with your immediate supervisor(s) and/or someone with experience in microbiome/metagenomics analysis, which considerations and controls would be most relevant for your particular project.

1. **Whenever possible, physically separate lab areas for sample preparation, DNA/RNA isolation, PCR amplification / library preparation, and analysis. In addition, consider maintaining a physical separation** between the **DNA/RNA isolation** from **high microbial biomass samples (e.g. fecal, sewage, soil)** and **low microbial biomass samples (e.g. swab samples, human/animal body fluids, tissue)** to avoid cross contamination**.**

2. **Regularly clean surfaces and equipment with 1% (vol/vol) sodium hypochloride solution, and/or UV exposure in addition to cleaning with 70% (vol/vol) ethanol.**

3. **Ideally, extract all samples around the same time, using DNA/RNA isolation kits and supplies purchased around the same time to avoid batch effects.**

4. **For each extracted sample, record the date, lot number of the DNA/RNA isolation kit, name of individual who performed the DNA/RNA isolation, and any other aspects that may impact on inferred microbial community composition.**

5. **DNA/RNA-extraction controls: i)** Include at least one **DNA/RNA extraction (blank/negative) control** in each batch of DNA/RNA-extractions. **ii)** If you use swabs, filters, or other devices for sample collection or processing, **extract DNA/RNA from unused swabs, filters, or other**. **iii) Perform biological and technical replicate DNA/RNA extractions** relevant for your project.

6. **If several sequencing runs are required for your project, distribute the samples randomly across these runs.** For example, do <u>not</u> include all case samples on one run, and then all samples from healthy controls on the other run.

7. **Use different barcodes in the library preparation than the ones that were used in the previous sequencing run** to limit carry-over contamination.

8. **Include the DNA/RNA-extraction controls in the sequencing runs.**

Sünje J. Pamp
Technical University of Denmark
Kemitorvet
Building 204
2800 Kgs. Lyngby
Denmark
Tel +45 35 88 63 54
sjpa@food.dtu.dk
www.food.dtu.dk

9. **Consider including a barcode control per run.** I.e. take one sample and label it with two separate sets of barcodes.

10. **Include one (a few) selected microbiome sample(s) on all runs belonging to the same project, and/or:**

11. **Include one sample of known composition (e.g. a mock community) on each sequencing run.**

12. **Process the sequencing data from all samples in the same way.**

13. **Consider using "clean" reference genomes sequences for data analysis from which contaminant (human, microbial) sequences have been removed to the extent possible.**
E.g. ftp://ftp.cbs.dtu.dk/public//CGE/databases/CuratedGenomes

14. **Examine the samples for the presence of: i) human host DNA sequences, ii) ambiguous/contaminant DNA in reference genome sequences, iii) environmental background DNA** possibly originating from DNA/RNA kit reagents and supplies, additional laboratory supplies and equipment, gloves, skin, other.

15. **Analyze your data using relevant statistical approaches.**

16. **Deposit your sequencing data and associated sample information in a public repository,** such as the International Nucleotide Sequence Database Collaboration (INSDC), **DDBJ / EMBL-EBI / NCBI**.

17. **When you share results from your project**, such as through **bioRxiv and/or a peer-reviewed scientific open-access journal**, make sure to **include detailed descriptions** about sample collection, handling, pre-processing, DNA/RNA isolation, sequencing, sequencing data processing, and statistical analysis.

18. Consider **including** in the publication the **code, R Notebooks, Markdown, Jupyter Notebook files** used for the data analysis.

Reference

Kirstahler P, Bjerrum SS, Friis-Møller A, la Cour M, Aarestrup FM, Westh H., and Pamp SJ. (2018) Genomics-Based Identification of Microorganisms in Human Ocular Body Fluid. **Scientific Reports,** 8(1):4126. doi: 10.1038/s41598-018-22416-4.

Sünje J. Pamp
Technical University of Denmark
Kemitorvet
Building 204
2800 Kgs. Lyngby
Denmark
Tel +45 35 88 63 54
sjpa@food.dtu.dk
www.food.dtu.dk