Metagenomics approaches for molecular identification of pathogens in human samples

Slavica Stanojcic [1], Bridlin Barckmann [1], Manon Tauzin [1], Tizaanuo Hien [1], Christophe Ravel [1,2], Yvon Sterkers [1-3]

[1] **MiVEGEC**, Université de Montpellier, CNRS, IRD et **Département de Parasitologie-Mycologie**, CHU de Montpellier, Montpellier, France
[2] **Centre National de Référence Leishmania**, Montpellier, France
[3] **Centre National de Référence Toxoplasmose, Pôle Biologie Moléculaire**, Montpellier, France

Background

Classical PCR Sanger sequencing method

NGS-based method

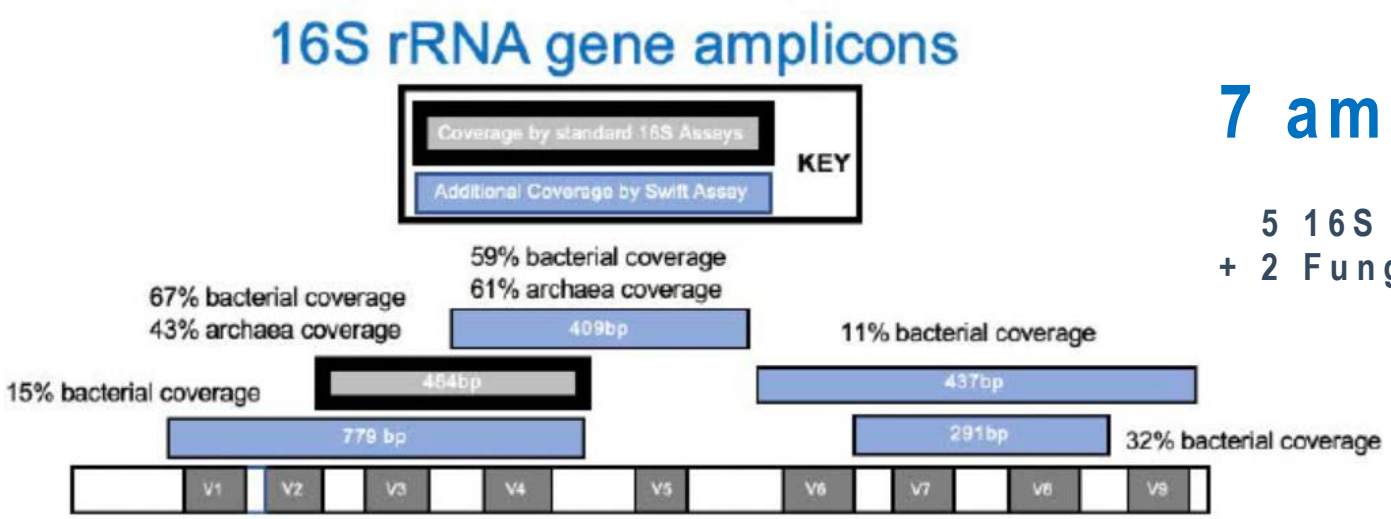
2 main advantages:

not target specific
allows the detection of several microbial agents in the same sample

Material and Methods

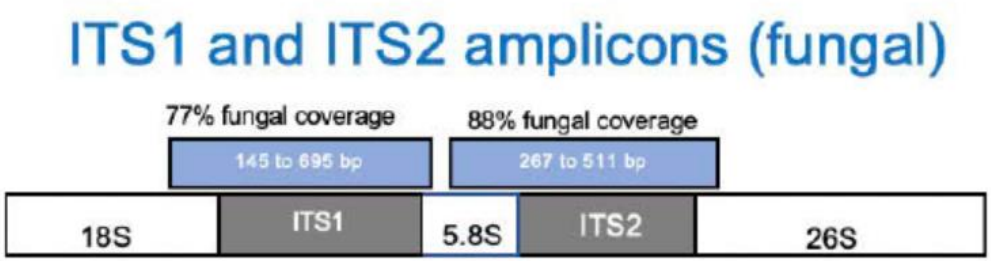
SWIFT AMPLICON™ 16S+ ITS PANEL

Cat. No. AL-51648 Visit swiftbiosci.com/protocols for updates



7 amplicons

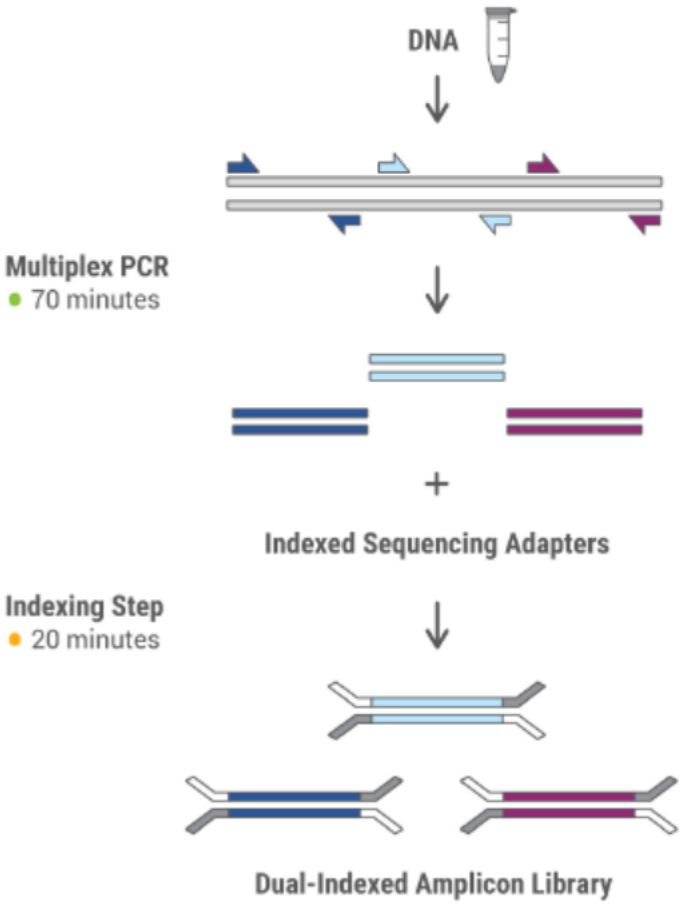
5 16S rRNA
+ 2 Fungal (ITS1 and ITS2)



Average amplicon size: 475 bp

Material and Methods

3 steps: amplification, indexing (library preparation), purification and controls

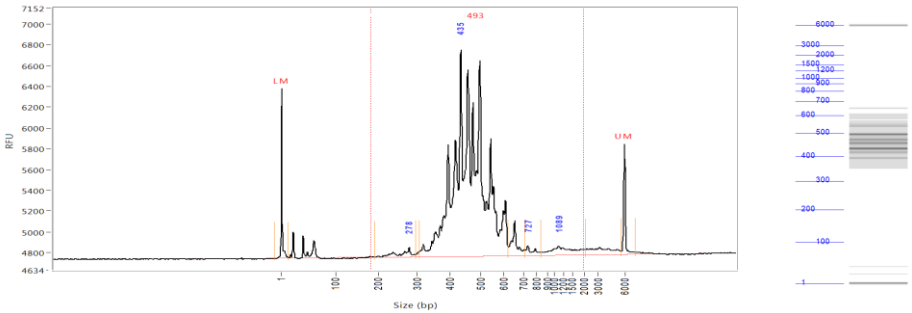


1. input DNA concentration: 10pg-50ng

2. optimization: increasing the nb of PCR cycles

3. Quantification, purification on beads

Bioanalyzer



Material and Methods

Sequencing

MGX

Flow cell MiniSeq Mid-Output paired-end
Paired-end **150 nts** sequencing

Data pre-processing and QC

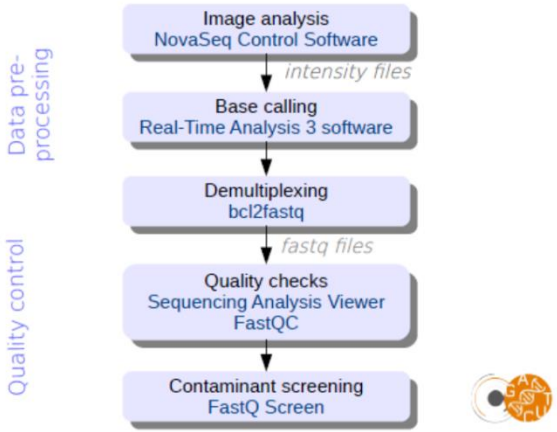
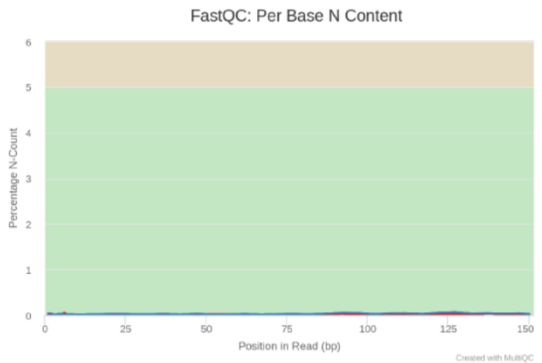
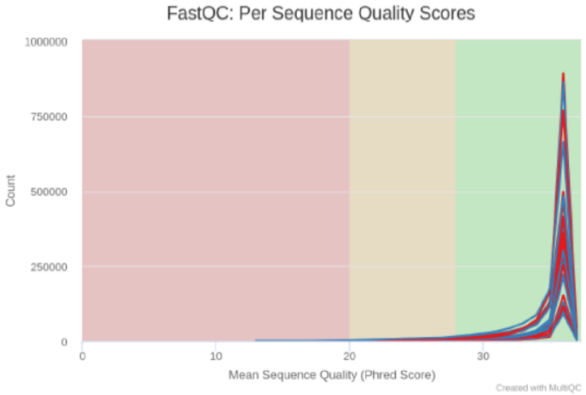
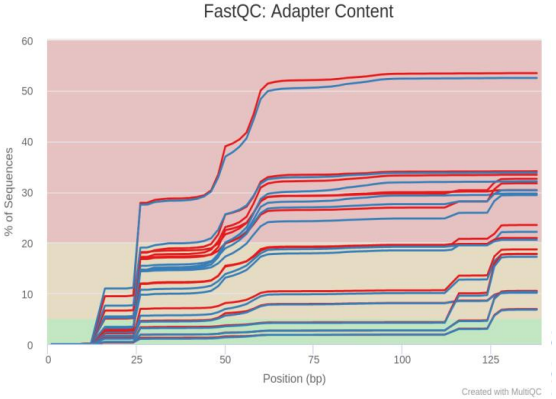


FIGURE 1 – Etapes de traitement et contrôle qualité des données

Score Phred	Erreurs pendant le base calling	Précision
10	1/10	90%
20	1/100	99%
30	1/1000	99,9%
40	1/10000	99,99%





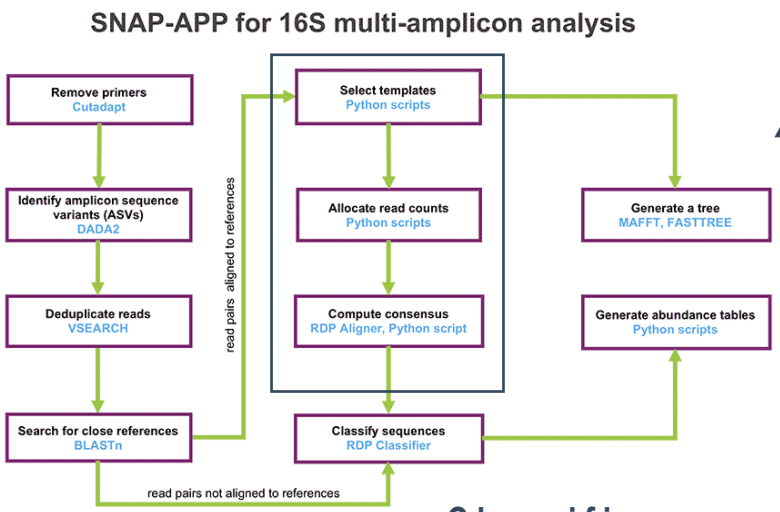


Material and Methods

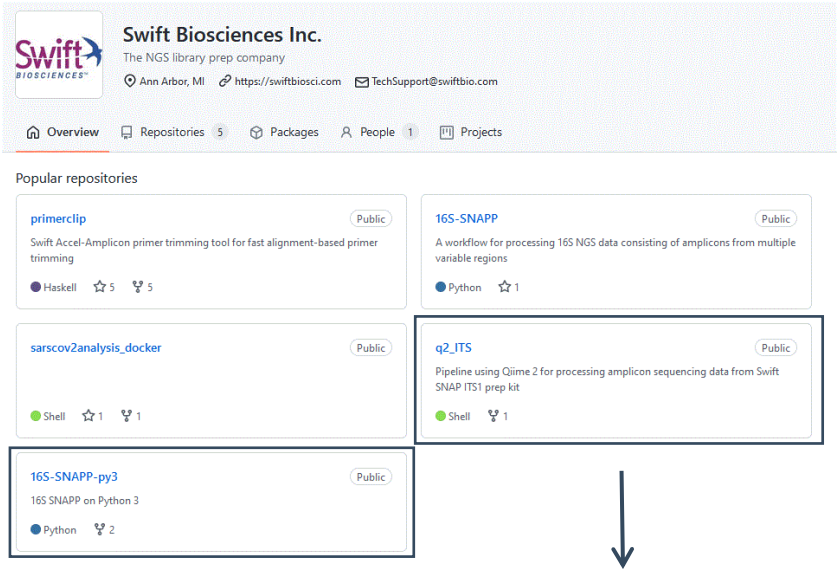
Bioinformatic pipeline Bioswift
Open source: Github

16S: 16S-SNAPP

18S: q2_ITS



Classifier:
RDP11.5



ITS analysis workflow using Qiime 2

Classifier:
UNITE ver.8 all eucaryotes

Université de Montpellier
Faculté de Médecine
Montpellier-Nîmes

Material and Methods

Commercial positive control

Clinical samples from routine practices

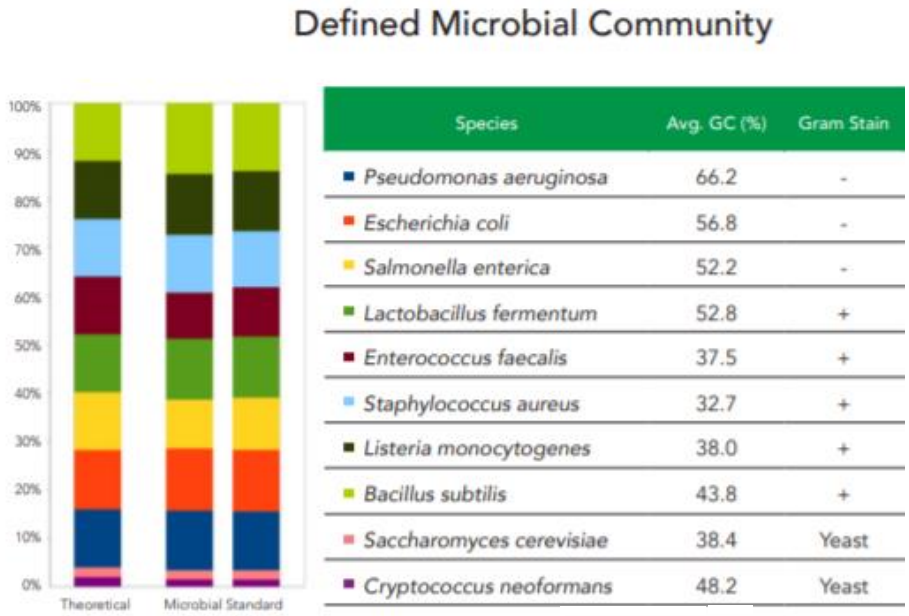
- Negative sample
- Samples found positive by specific PCR Sanger sequencing

Spiked samples

- *E. coli* positive spiked sample in a human blood matrix
- *C. albicans* spiked sample in a human blood matrix
- Equimolar mixture of *E. coli* and *C. albicans* positive samples
- Serial tenfold dilution assay in a human blood matrix

Results

Commercial positive control : ZymoBIOMICS® Microbial Community Standard



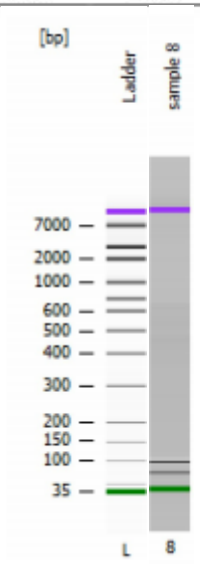
Species	Avg. GC (%)	Gram Stain
Pseudomonas aeruginosa	66.2	-
Escherichia coli	56.8	-
Salmonella enterica	52.2	-
Lactobacillus fermentum	52.8	+
Enterococcus faecalis	37.5	+
Staphylococcus aureus	32.7	+
Listeria monocytogenes	38.0	+
Bacillus subtilis	43.8	+
Saccharomyces cerevisiae	38.4	Yeast
Cryptococcus neoformans	48.2	Yeast

We found all the microrganisms present in this standard

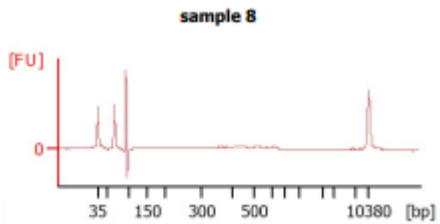
including *Cryptococcus neoformans*

but *Saccharomyces cerevisiae*

Negative control



2021-05-12



Results

Clinical cases (positive by classical methods)

Case 1: we were able to detect *Histoplasma capsulatum* in a positive bone marrow

Case 2-5: we were able to detect Leishmania spp. (*Leishmania infantum, L. major* and *L. guyanensis*) in blood, bone marrow and skin samples

- Issue 1: not with the Bioswift bioinformatic pipeline
- Issue 2: we were not able to determine the species

Spiked samples and serial dilution assay

We were able to detect *E.coli* and *C. albicans*

- alone
- mixed together
- in a serial dilution assay 1/10 and 1/100

Conclusion

r-DNA amplicons NGS is able to detect
- a large spectrum of microbial agents
- several microbial agents in a mixture

Sensitive

Issues
- not turn key, necessites optimization and a skilled staff
- taxonomic issues
- quantitative? Not assessed yet

Perspectives
- other protozoan (*Toxoplasma gondii*)
- new version of this kit by Bioswift, other kits
- home-made primer pairs, home-made pipeline, others databases
- WGS

Remerciements

Laboratoire de Parasitologie-Mycologie, CHRU de Montpellier

Yvon Sterkers
Slavica Stanojcic (Ingénieure UM)
Bridlin Barckmann (Bioinformaticienne)
Manon Tauzin (Master 1)
Tizaanuo Hien (5eA Pharmacie)
Christophe Ravel

Laurence Lachaud
Maude F Lévêque
Nathalie Bourgeois
Emmanuelle Varlet-Marie
Sahar Albaba
Grégoire Pasquier
Patrick Bastien

MGX-Montpellier GenomiX - @MGX_Montpellier
c/o Institut de Génomique Fonctionnelle - 141 rue de la Cardonille - 34 094 Montpellier Cedex
Tel. Exp : 04 34 35 92 61 / Bioinfo : 04 34 35 92 39

Biologie moléculaire : Mathilde ESTEVEZ-VILLAR, Simon GEORGE, Anne-Alicia GONZALEZ, Elise GUERET, Hugues PARRINELLO, Dany SEVERAC
Bioinfo : Emeric DUBOIS, Anaïs LOUIS, Xavier MIALHE, Stéphanie RIALLE
Mail: mgx@mgx.cnrs.fr Web: http://www.mgx.cnrs.fr

Sylvie Douzou, Martine Brun, France Joullié

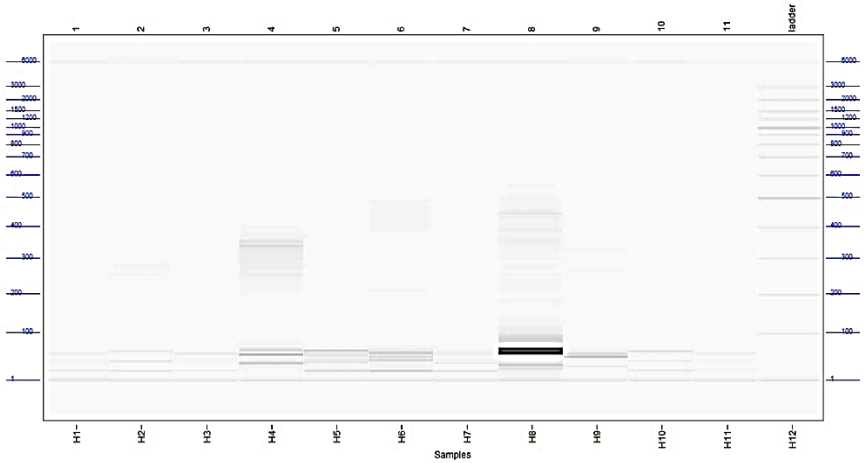
Optimization

Different number of PCR amplification cycles influence the quantity of libraries

(we can mesure the quantity of total DNA (mixture of human's DNA and DNA from microbiota), but it is difficult to estimate the quantity of target DNA (DNA from microbiota) in each sample)

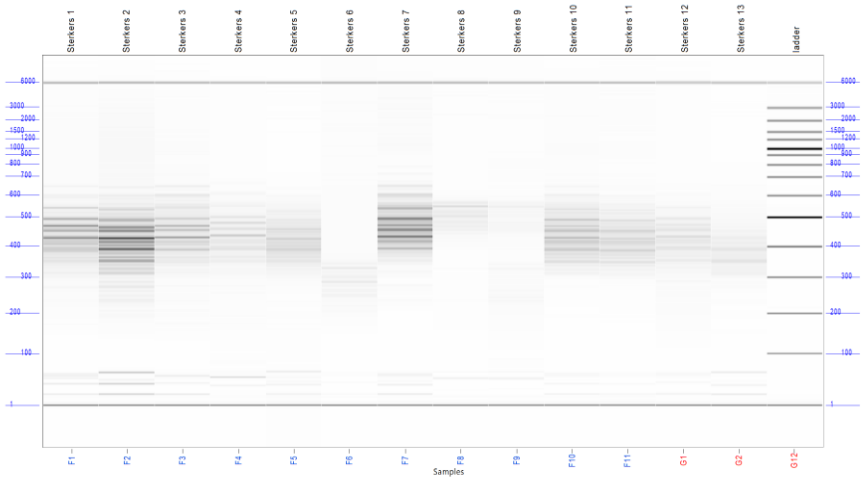
25 PCR cycles

according to kit's protocol
In many libraries not enough of DNA to perform sequencing



35 PCR cycles

In almost all libraries enough DNA to perform sequencing



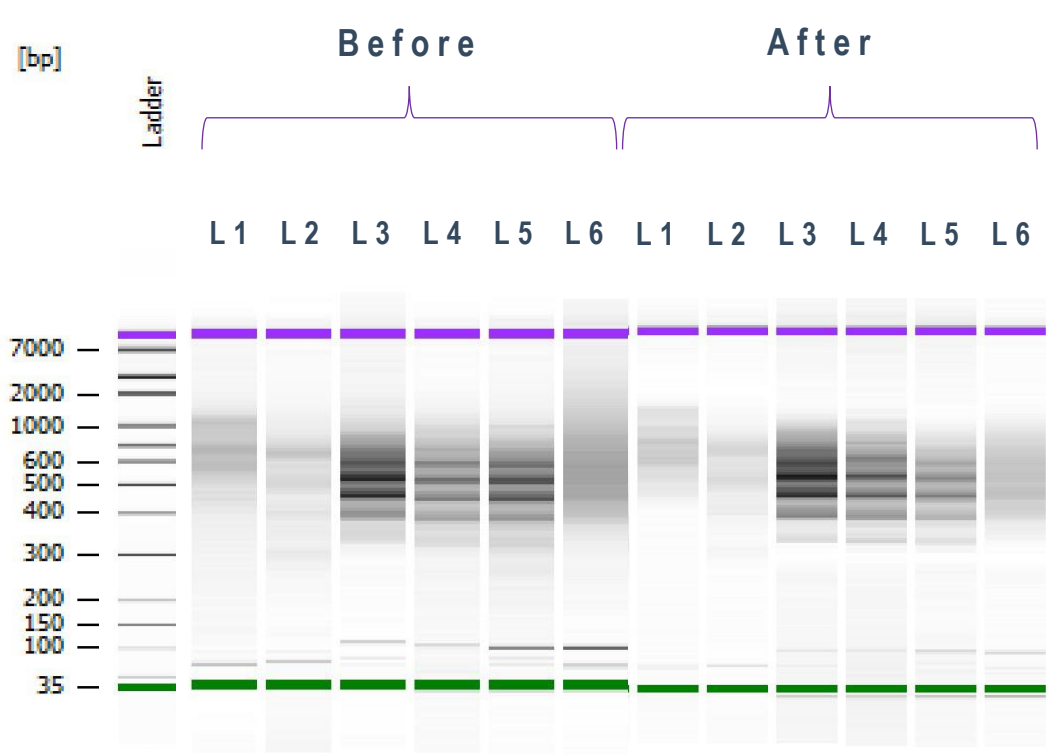
UNIVERSITÉ DE MONTPELLIER

Université de Montpellier
Faculté de Médecine
Montpellier-Nîmes

Optimization

Purification with magnetic beads remove the contaminating bands like primer dimers and adapter dimers

It is very important to minimize the percentage of primer dimer molecules in the libraries to less than 0.5% because adapter dimers are very competitive molecules during sequencing process



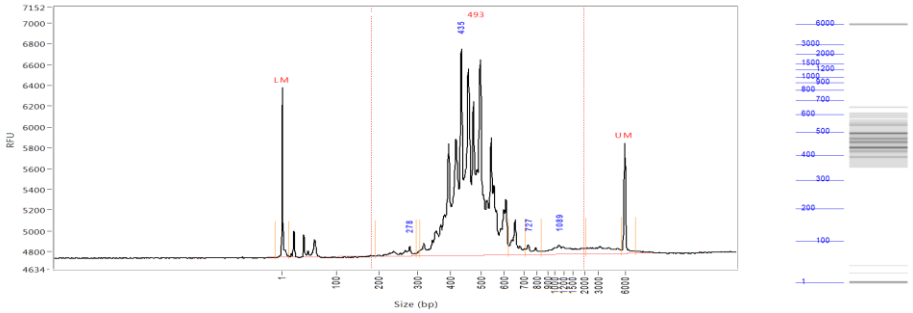
Purification with magnetic beads

Adapter dimers

Optimization

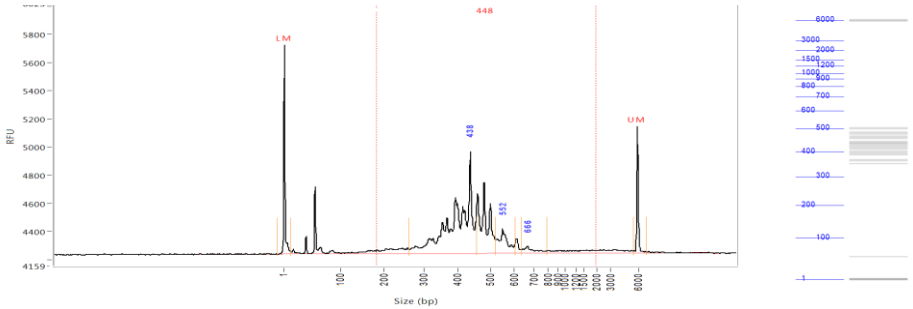
The libraries can be of different quality

Library of good quality

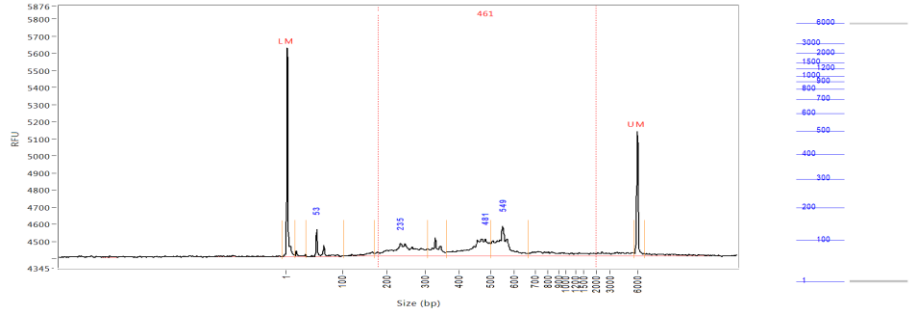


Percentage of adapter dimers low compare to amplified microbiota DNA

Library of medium quality



Library of bad quality



Percentage of adapter dimers high compare to amplified microbiota DNA

Data pre-processing and QC

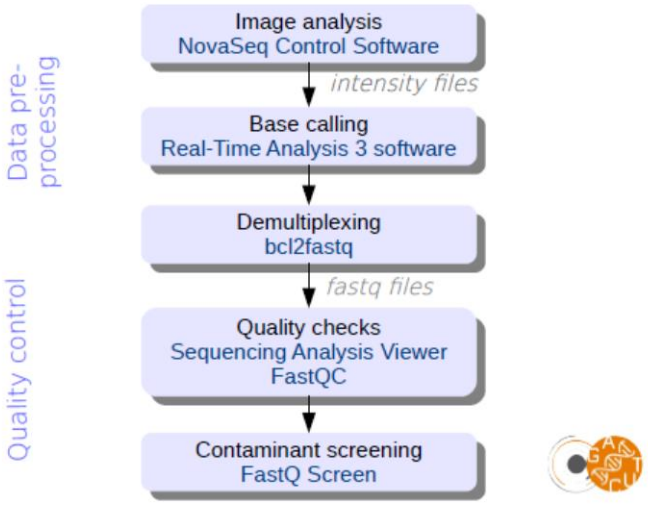


FIGURE 1 – Etapes de traitement et contrôle qualité des données

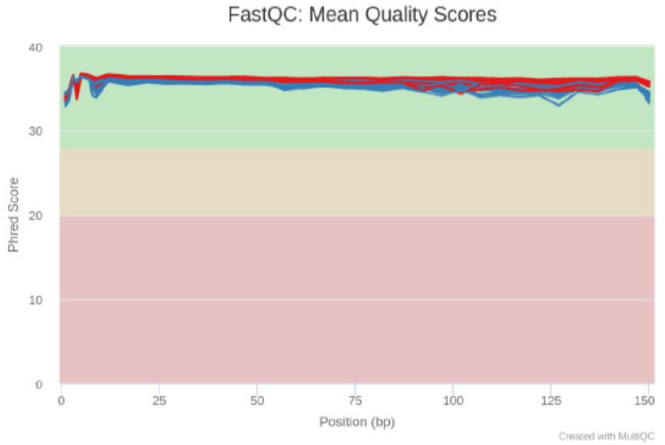


FIGURE 2 – Qualité par base pour tous les échantillons

Score Phred	Erreurs pendant le base calling	Précision
10	1/10	90%
20	1/100	99%
30	1/1000	99,9%
40	1/10000	99,99%

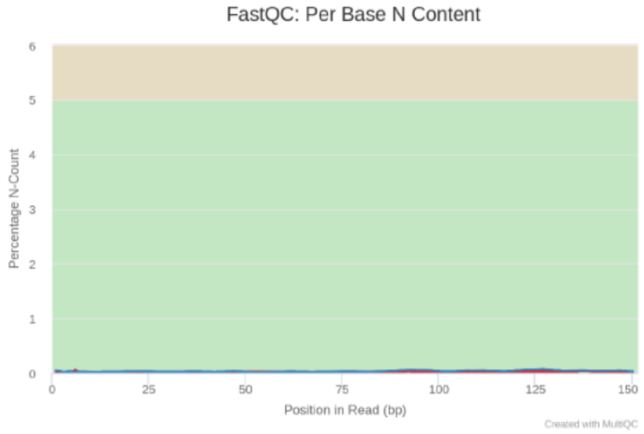


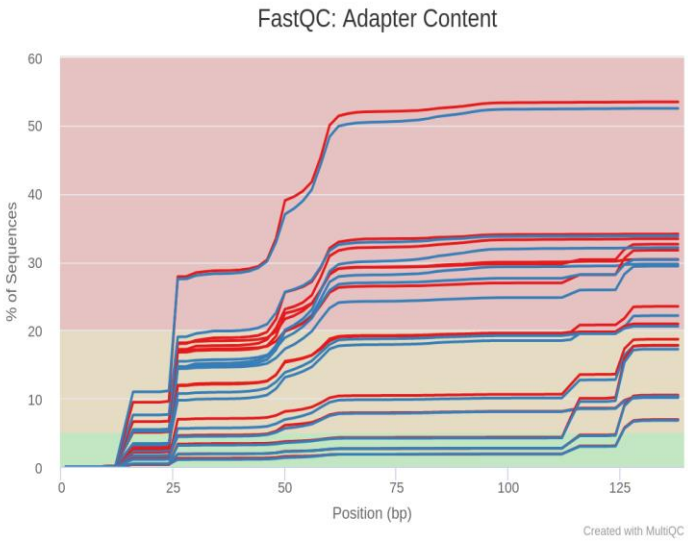
FIGURE 4 – Pourcentage de bases "N" par cycle

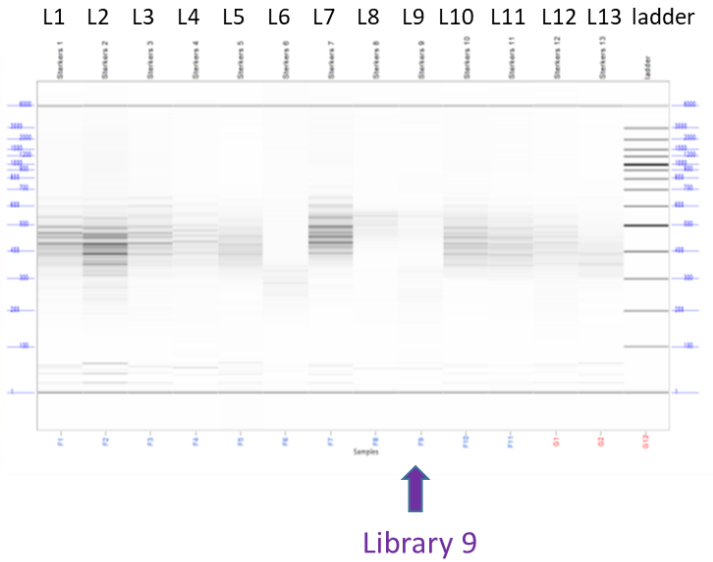
FIGURE 3 – Distribution des scores moyens de qualité par séquence pour tous les échantillons

Quality control of data after sequencing

Echantillon	PF
1	613 681
2	345 629
3	569 337
4	463 929
5	353 835
6	1 268 266
7	471 291
8	803 649
9	1 176 547
10	200 491
11	160 920
12	388 625
13	150 985
Total	6 967 185

a large number of sequences of very good quality.
Nevertheless, there is a strong imbalance in terms of number of sequences between the samples,
but they all have more than 100 000 sequences, which was desired





Percentage of adapters per cycle

The figure shows the percentage of adapter sequences detected at each position of the reads.
Up to 53.38% adapters are observed at end of reads (Library 9).