



HAL
open science

From high throughput 454 GS FLX data analysis process of 16S RNA gene sequences using barcoding to bacterial community exploration

Laurent Cauquil, Jérôme Mariette, Christine Rousseau, Asma Zened, Francis Enjalbert, Annabelle Troegeler-Meynadier, Christophe Klopp, Sylvie Combes

► To cite this version:

Laurent Cauquil, Jérôme Mariette, Christine Rousseau, Asma Zened, Francis Enjalbert, et al.. From high throughput 454 GS FLX data analysis process of 16S RNA gene sequences using barcoding to bacterial community exploration. 4th Congress of European Microbiologists, Jun 2011, Genève, Switzerland. pp.0, 2011. hal-04041509

HAL Id: hal-04041509

<https://hal.science/hal-04041509>

Submitted on 22 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/Eprints> ID : 5189

To cite this version :

Cauquil, Laurent and Mariette, J and Rousseau, C and Zened, Asma and Enjalbert, Francis and Troegeler, Annabelle and Klopp, C and Combes, Sylvie *From high throughput 454 GS FLX data analysis process of 16S RNA gene sequences using barcoding to bacterial community exploration.* (2011) In: 4th Congress of European Microbiologists, Juin 2011, Genève (Suisse).

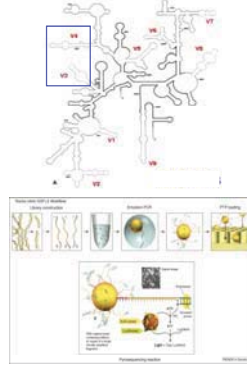
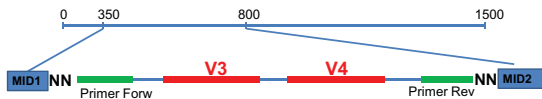
Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr.

From high throughput 454 GS FLX data analysis process of 16S RNA gene sequences using barcoding to bacterial community exploration

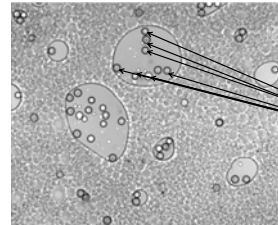
L. Cauquil¹, J. Mariette², C. Rousseau³, A. Zened¹, F. Enjalbert¹, A. Troegeler, ¹ C. Klopp² and S. Combes¹
¹UMR 1289 Tandem, INRA / INPT-ENSAT / ENVT, Castanet-Tolosan, France ²Plateforme bio-informatique Genotoul, Biométrie et Intelligence Artificielle, INRA, Castanet-Tolosan, France, ³INPT-ENSAT, Castanet-Tolosan, France
 Laurent.Cauquil@toulouse.inra.fr



The bacterial community of 97 cow rumen samples was explored using a 440 bp V3-V4 16S RNA amplicon. Each sample is barcoded using a couple of MIDs.



Since sequences might be randomly over represented within a run when sequencing is done using the Roche 454 pyrosequencing platform (Gomez-Alvarez et al. 2009), random nucleotide dimers were introduced in the sequence.



Eight beads instead of one

Pipeline

Rawdata : 889 029 reads



16Spyrocleaner (python)

- Length > 150 nt
- Sequence with one N are removed
- Homopolymere > 10 are removed

-19.7%

MOTHUR (<http://www.mothur.org>)

- Alignment (Silva base)
- Distance matrix
- Assigning sequences to OTU



NNpyrocleaner (python)

- Remove over represented sequences (-0.24%)

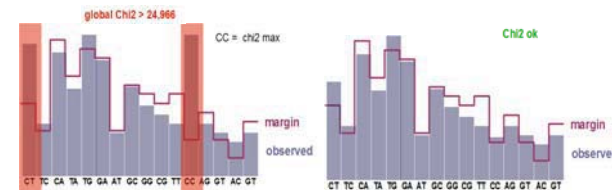
MOTHUR (<http://www.mothur.org>)

- Alignment
- Bad aligned seq were removed (-1%)
- Distance matrix
- Assigning sequences to OTU

706 155 reads
7249 reads/sample

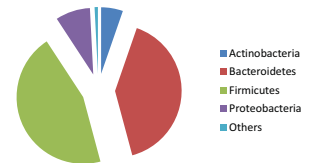
- Taxonomic assignment (bootstrap 60%)

- 868 OTU /sample (cutoff 0.1)
- Coverage of 95.9%
- Diversity indices



Taxonomy levels	Groups identified	% Affiliation
Phylum	15	100
Class	24	98
Order	51	95
Family	114	84
Genus	293	50

1.4% of reads not affiliated to any phylum were removed.

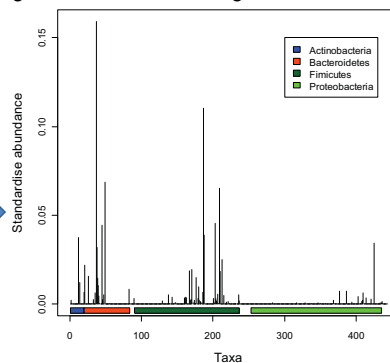


Mean distribution of major bacterial groups at the phylum level

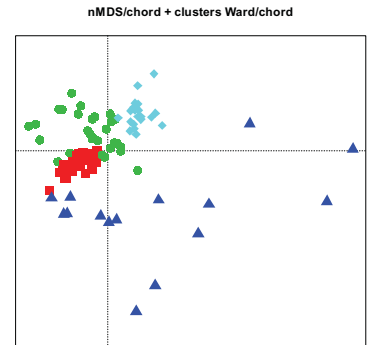
Analysis

From the 97 samples, we identified 430 different taxa leading to a taxonomic ecological matrix. The resulting data set can be used for ordination and clustering analysis (after normalisation) using R software.

X - matrix	Taxon 1	Taxon 2	Taxon 3	Taxon 7	Taxon 8	Taxon 9	...	Taxon 430
Ruminal fluid 1	0	0	0	0	57	0	...	0
Ruminal fluid 2	1	0	3	0	53	0	...	0
Ruminal fluid 3	0	0	0	0	4	0	...	0
Ruminal fluid 4	0	0	1	0	92	0	...	0
Ruminal fluid 5	0	0	1	0	25	0	...	0
Ruminal fluid 6	0	2	3	2	49	0	...	0
...
Ruminal fluid 97	0	0	4	0	75	1	...	0



Plot of the mean of the standardise abundance of each taxa upon the 97 samples



Ordination plot (nMDS) of the 97 samples and classification in 4 clusters