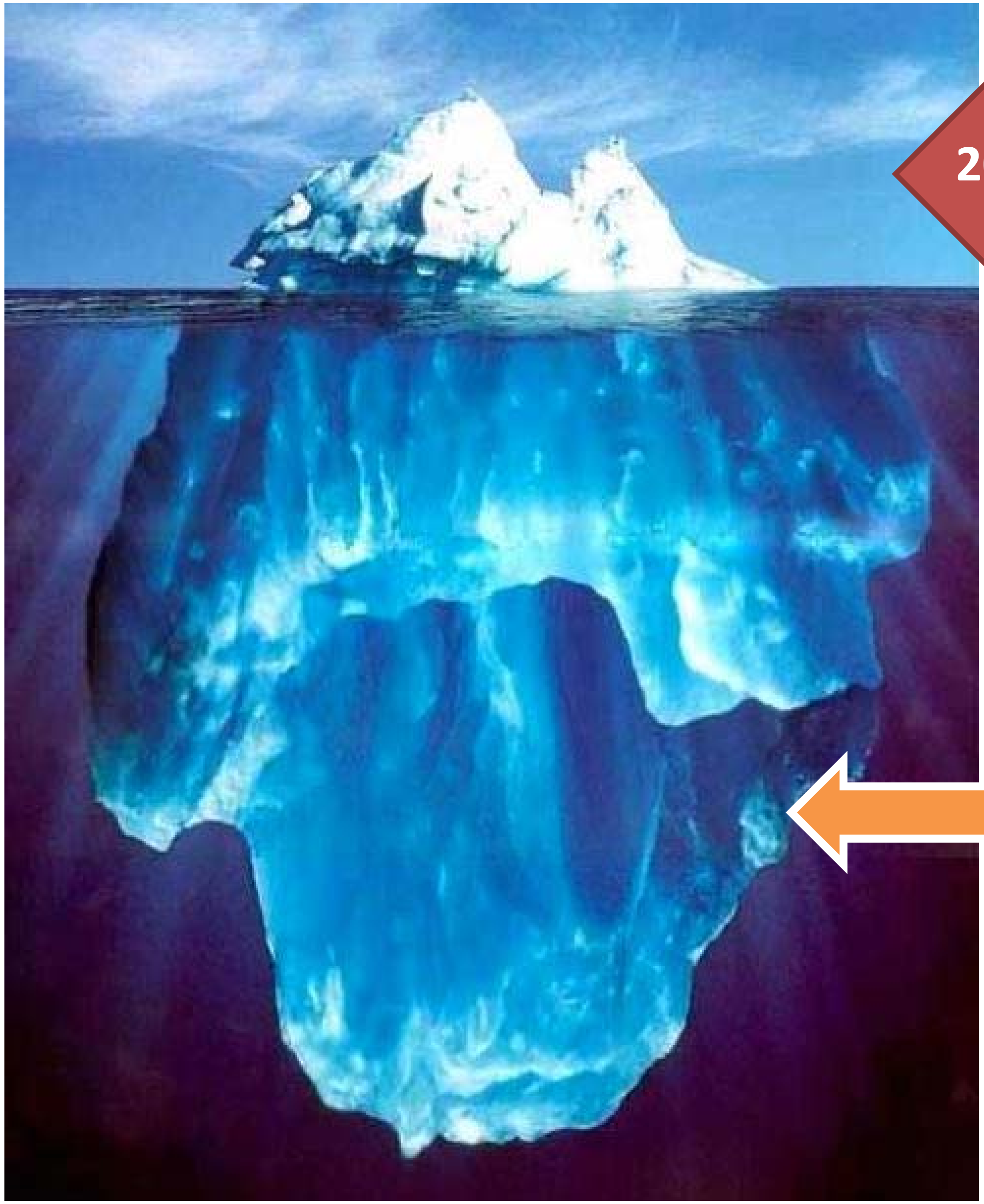
Analyse de communautés bactériennes complexes: du 454 au MiSeq

Sylvie COMBES, Laurent CAUQUIL, Olivier ZEMB, Béatrice GABINAUD, Géraldine PASCAL

Olivier BOUCHEZ, Frédéric ESCUDIER, Sophie VALIERE

30 Sept 2013

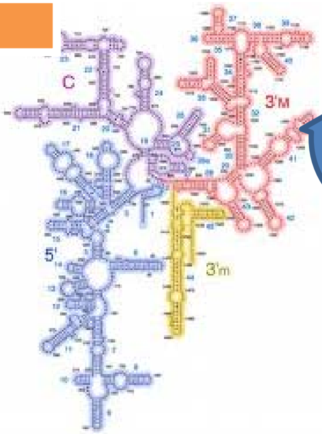
Microbiote dominant

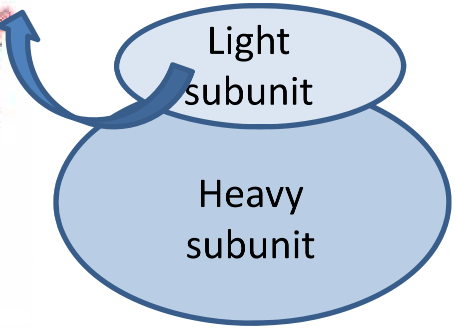


20 to 40%

Re-évaluation par approches moléculaires indépendantes de la culture

16S RNA

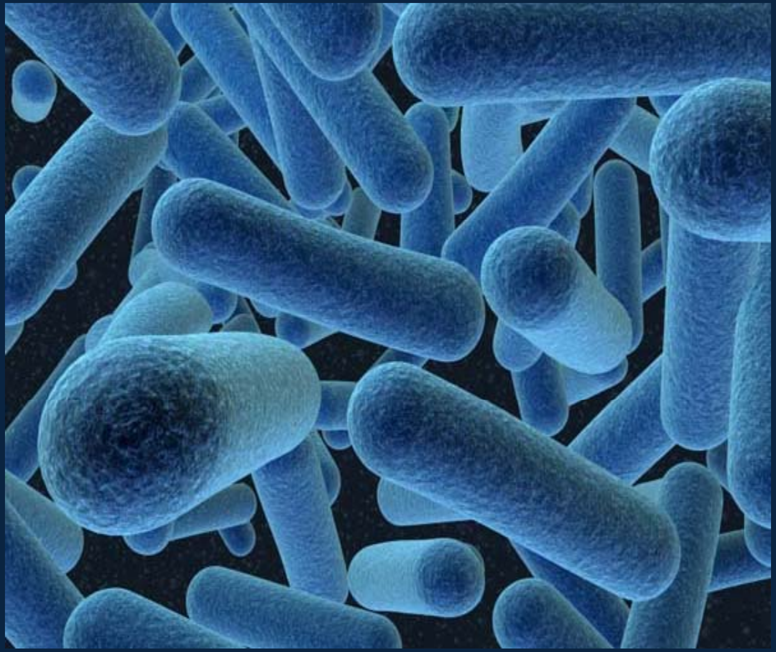




Light subunit

Heavy subunit

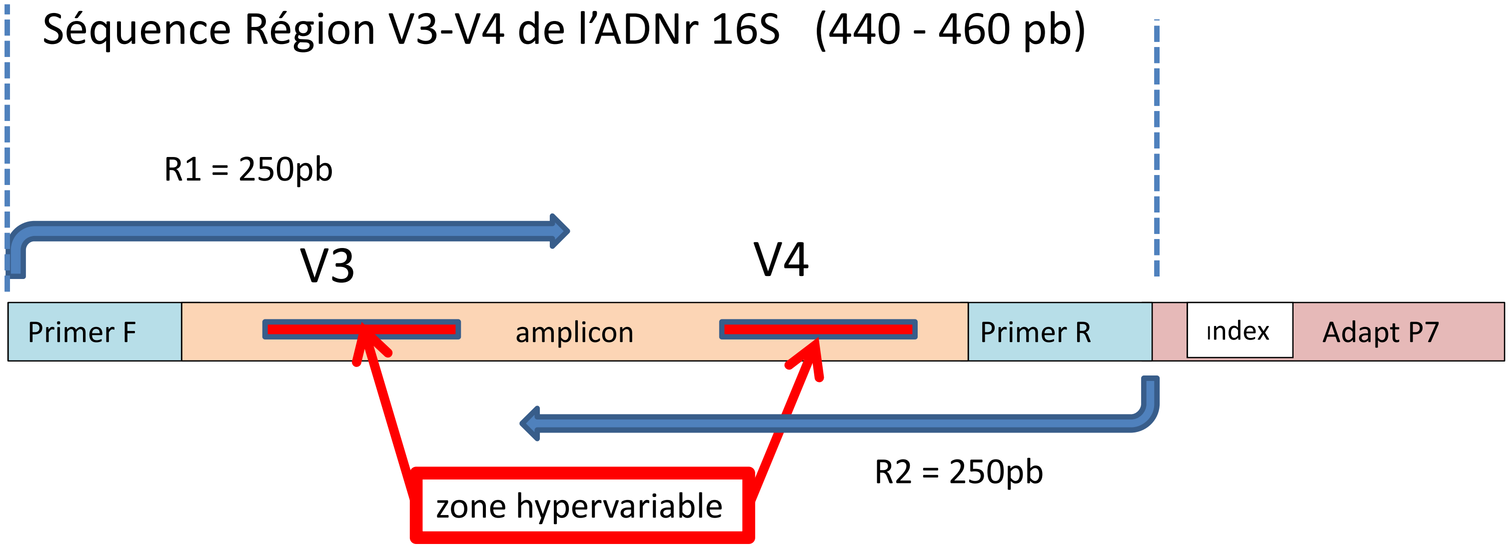
Miseq: mode d'emploi



CONSTRUCTION

Séquençage paire-end 2x250nt
Géolocalisé (R1 & R2 associés)

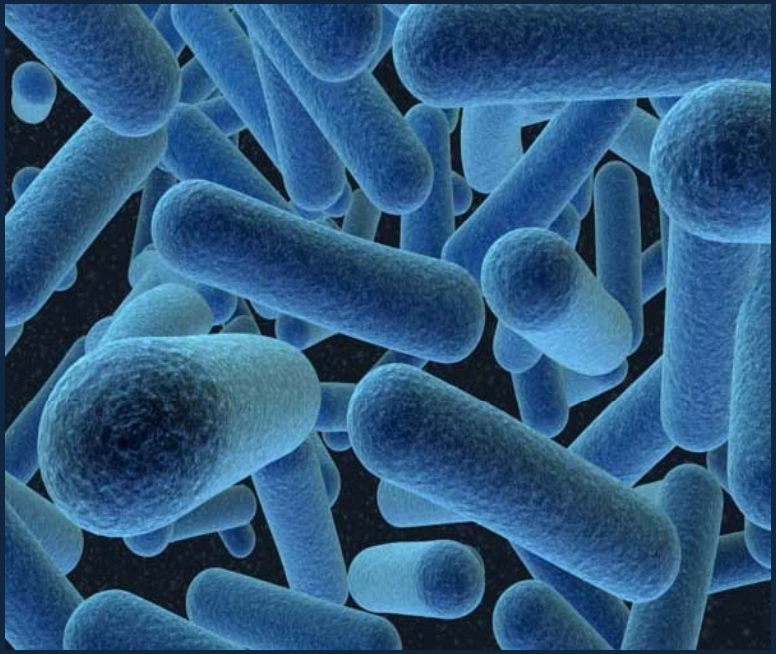
Séquence Région V3-V4 de l'ADNr 16S (440 - 460 pb)



Run de 104 échantillons (index de 6 nt):

- Communauté digestive lapin (n=80)
- Communauté digestive porc (n=21)
- Plasmides contenant la séquence ADNr 16S de bacterie connues (n=2)
- Mélange 4 bactéries marines (n=1)

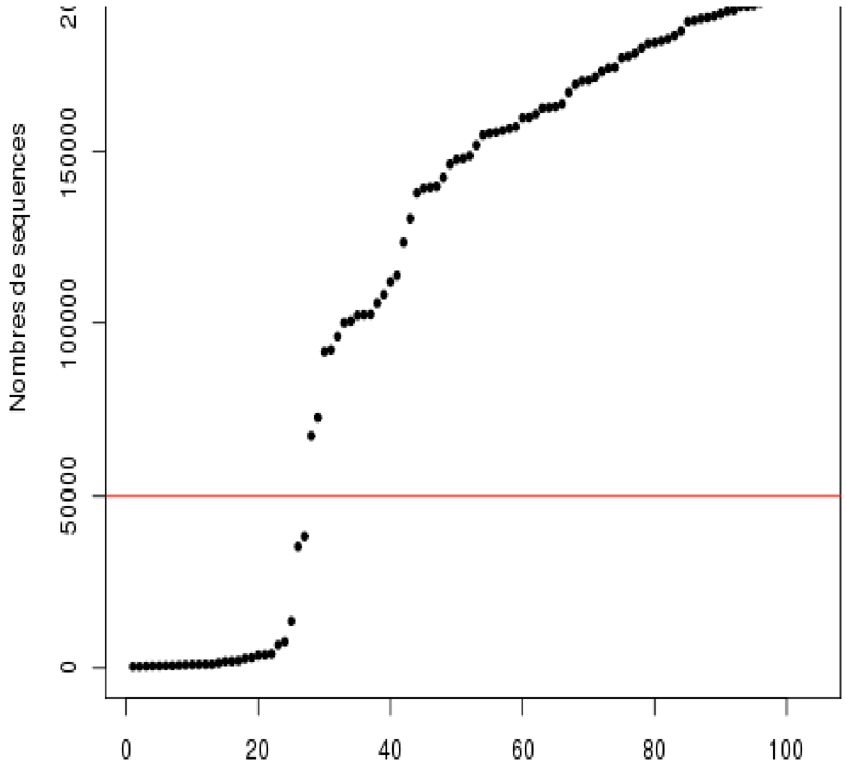
Miseq: les données



Nombre de reads par echantillon

Nombre de reads après filtrage Illumina: 24 854 086
(8 % de reads supprimées)

soit
12 427 043 **R1**
12 427 043 **R2**



REPARTITION DES SEQUENCES DANS LES 104 ECHANTILLONS

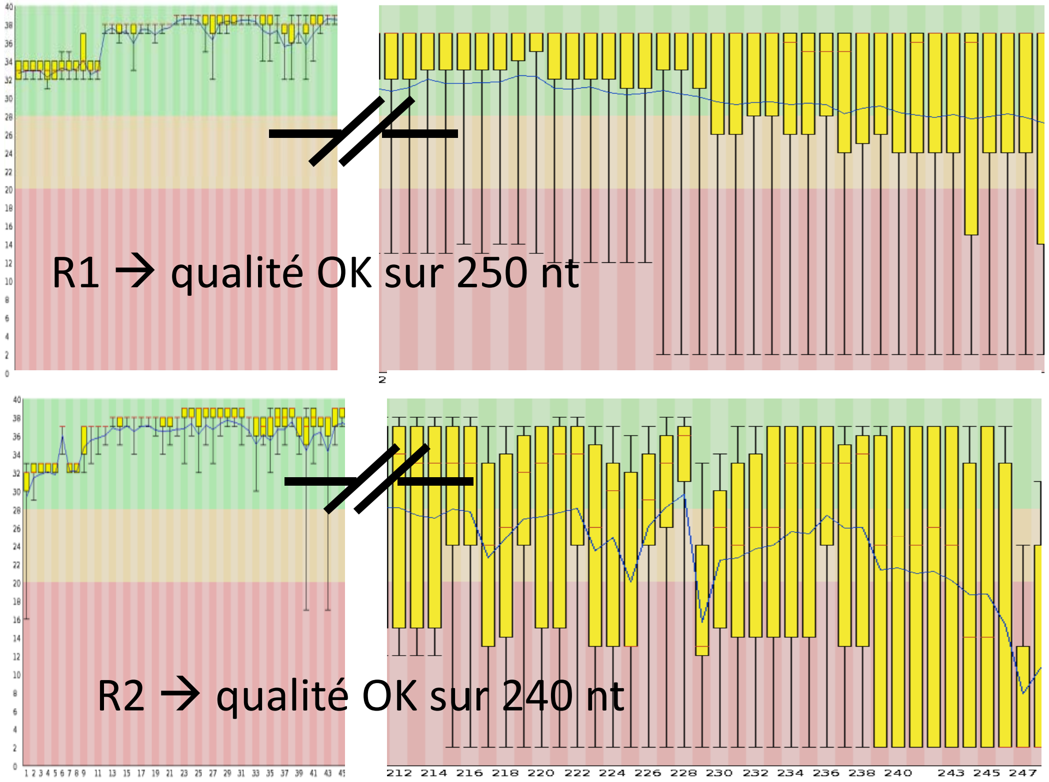
Problème technique:

27 échantillons ont moins de 50 000 séquences après filtrage Illumina

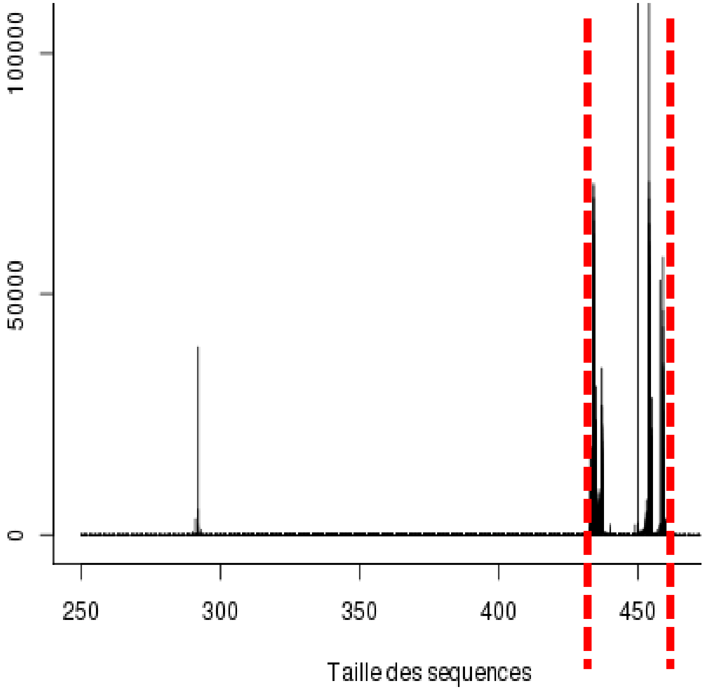
Problème d’amplification sur PCR1 → présence d’inhibiteur

Suite de la présentation sur 104-27 = 77

CONTIGAGE PAR FLASH

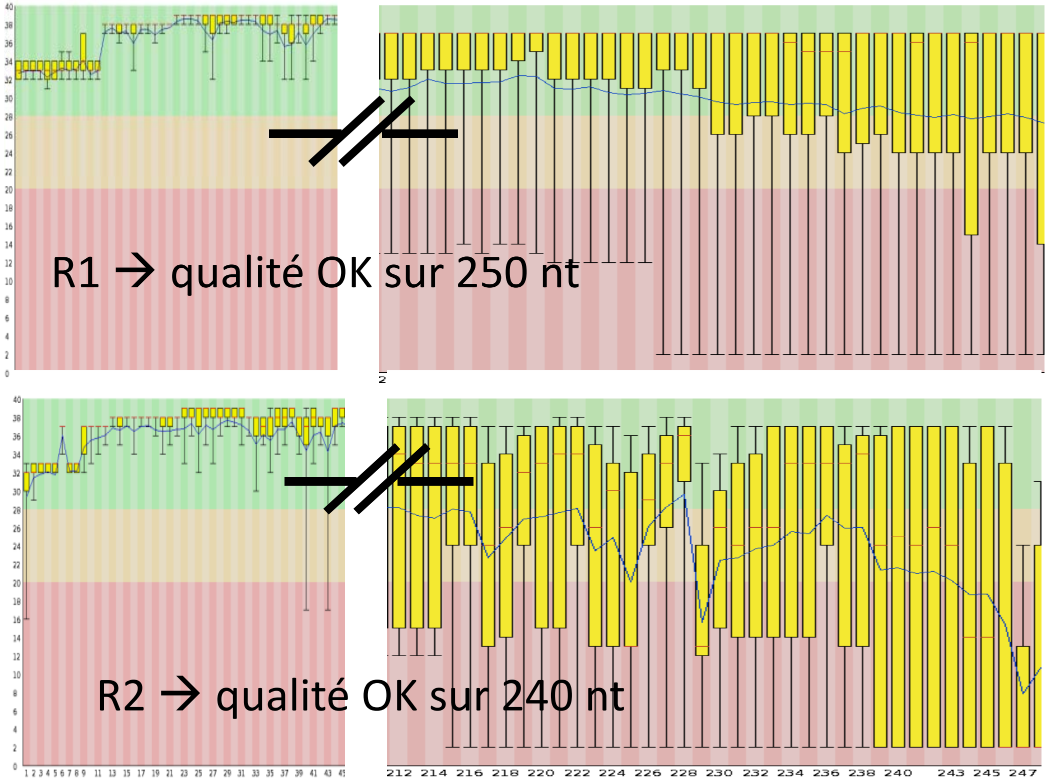

R1 → qualité OK sur 250 nt

R2 → qualité OK sur 240 nt

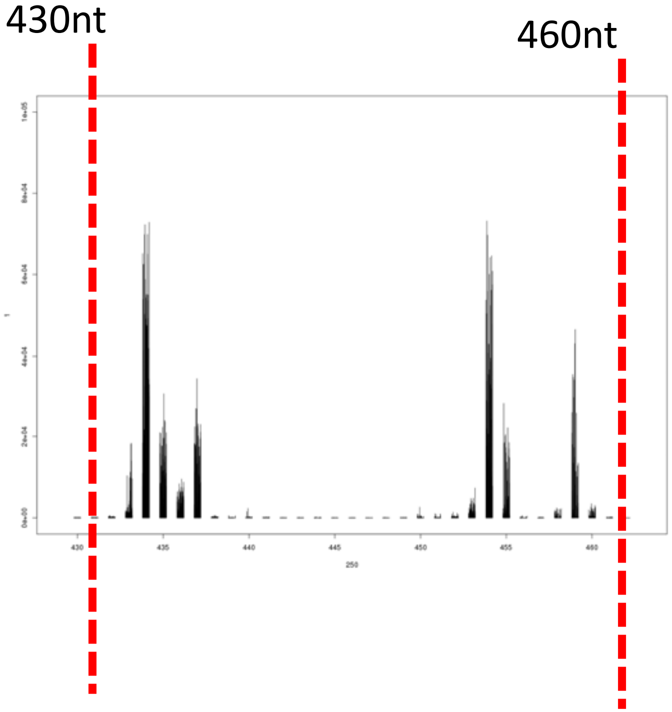

430nt 460nt

- ➢ overlap minimum 10 bases→ notre construction nécessite un overlap de min 20 et max 35 nt

CONTIGAGE PAR FLASH


R1 → qualité OK sur 250 nt

R2 → qualité OK sur 240 nt

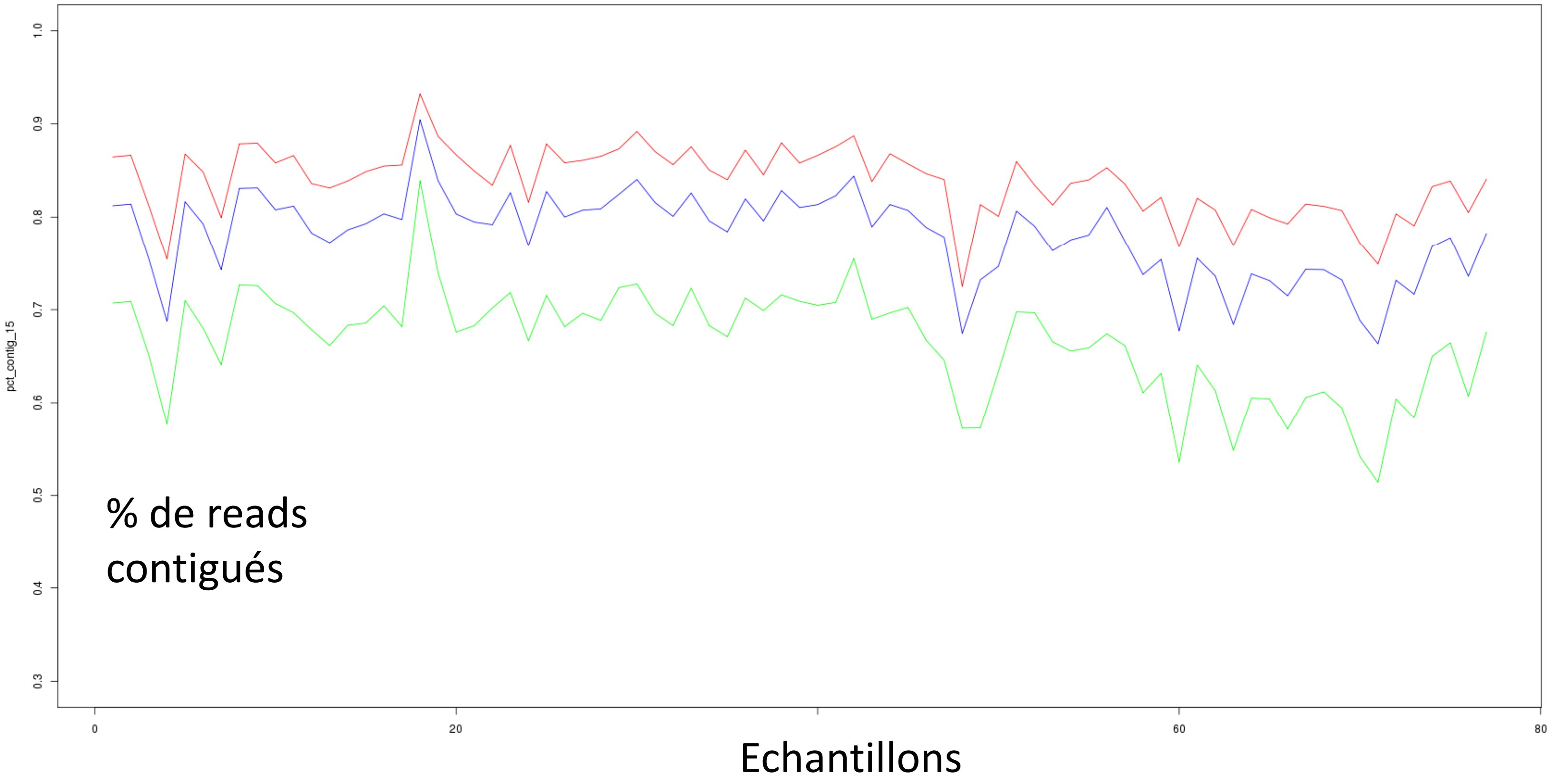


- ➢ overlap minimum 10 bases→ notre construction nécessite un overlap de min 20 et max 35 nt
- ➢ élimine une grande partie des reads contenant des "N" mais pas tous

CONTIGAGE PAR FLASH

3 seuils de mismatch testés: 5%, 10%, 15%

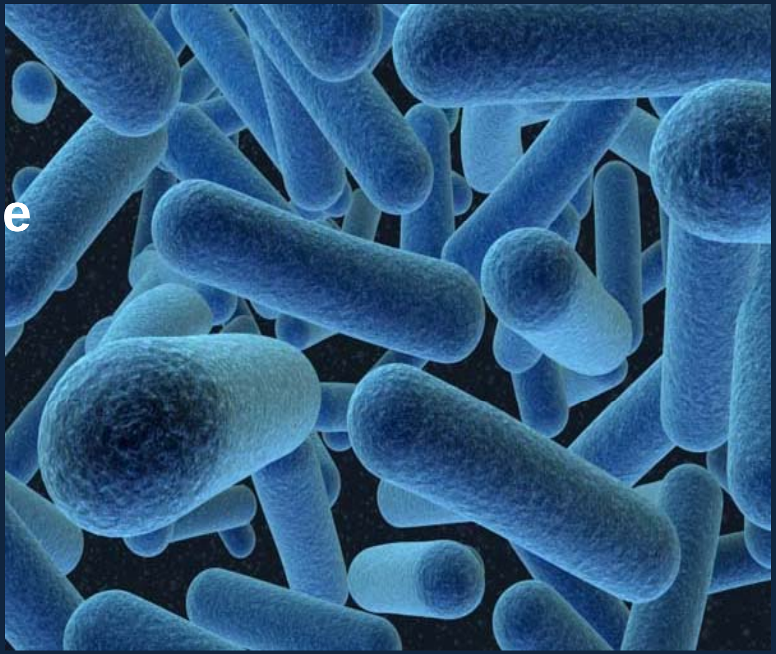
- A 5% → 65 % des couples R1 R2 sont contigués → 8 millions
- A 10% → 75 %
- A 15% → 80 %



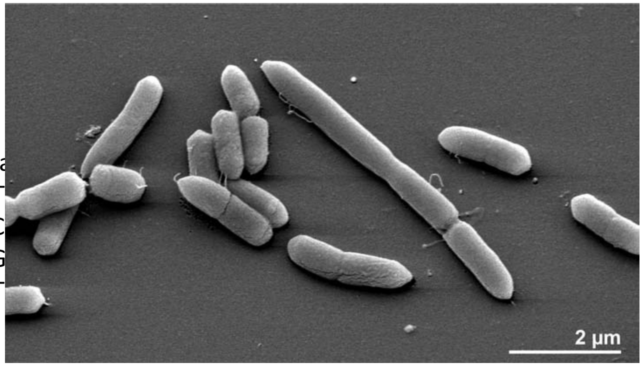
% de reads contigués

LES PLASMIDES

Vérifier
la qualité du contigage
et du processus de séquençage



Plasmides

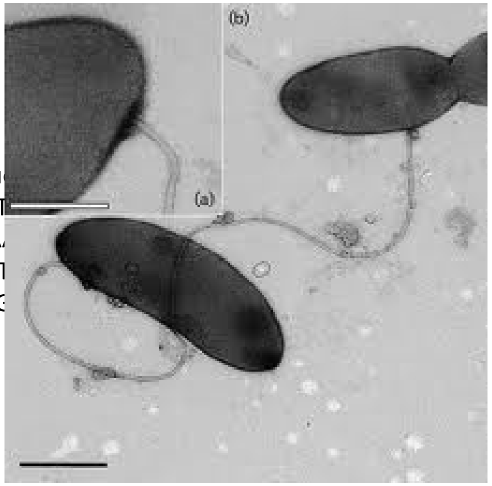


Prevotella bryantii (Bacteroides): NED4B6 (454nt)

```
>gi|129594480|gb|EF445235.1| Uncultured bacterium clone NED4B6 16S ribosomal RNA gene, pa
TACGGGAGGCAGCAGTGAGGAATATTGGTCAATGGACGGAAGTCTGAACCAGCCAAGTAGCGTGCAGGAT
ATAAAGTTAGCCACGTGTGGTTATTTGCATGTACCCTACGAATAAGGACCGGCTAATTCCGTGCCAGCAGCC
ATTGGGTTTAAAGGGAGCGCAGGCCGTTTGGTAAGCGTGTTGTGAAATGTCCGGGCTCAACCTGGGCACTG
GGCGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATATCACGAAGAACTCCAATTGCGAAGGCAGCTCGCT
TCGAACAGGATTAGATACCCTGGTA
```

208 824 R1 ou R2 brut

137 622 contigs (66%)



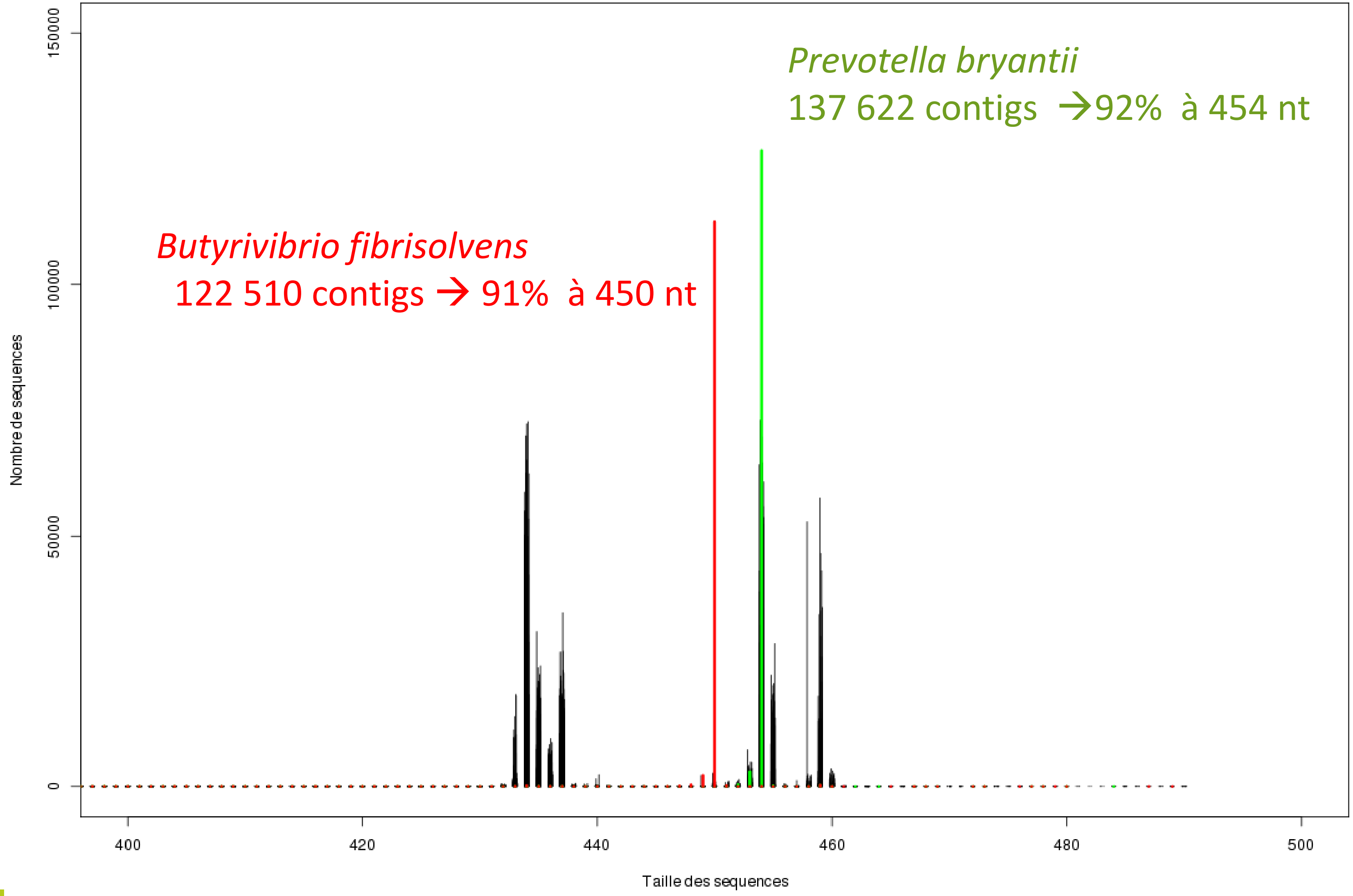
Butyrivibrio fibrisolvens (Firmicutes): NED4E2 (450nt)

```
>gi|129594507|gb|EF445262.1| Uncultured bacterium clone NED4E2 16S ribosomal RNA gene, partial sequ
TACGGGAGGCAGCAGTGGGGGATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTAT
GAAGAAAGACCTCGTAAGAGGGGATGACGGTACCTGAGTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTA
ATTTACTGGGTGTAAAGGGAGCGTAGACGGTGTATCAAGTCTGAAGTGAAACCCCACGGCTCAACCGTGGGCTTGCTTT
GAGGTAAGCGGAATTCCTGGTGTAGTAGTGAAATGCGTAGATATCAGGAAGAACATCAGTGGCGAAGGCGGCTTACTG
GGCGTGGGGAGCA           AGATACCCTGGT
```

181 734 brut

122 510 contigs (67%)

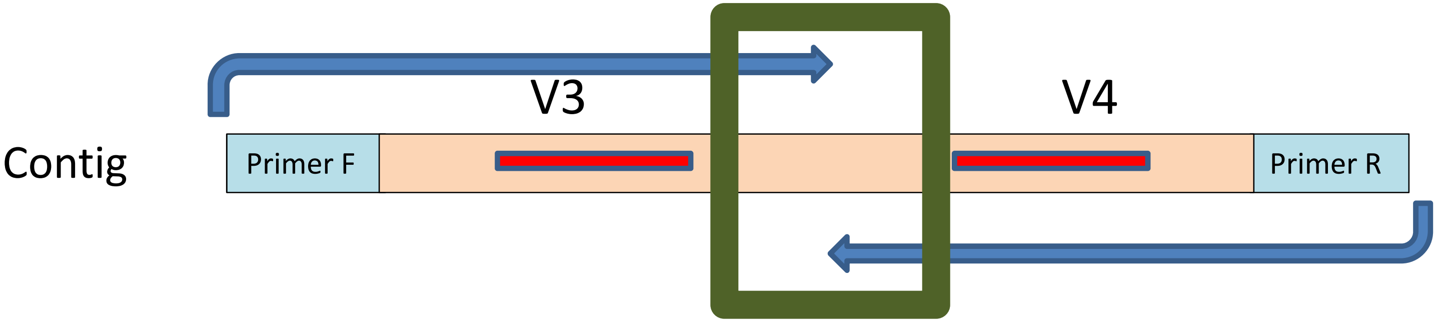
Plasmides



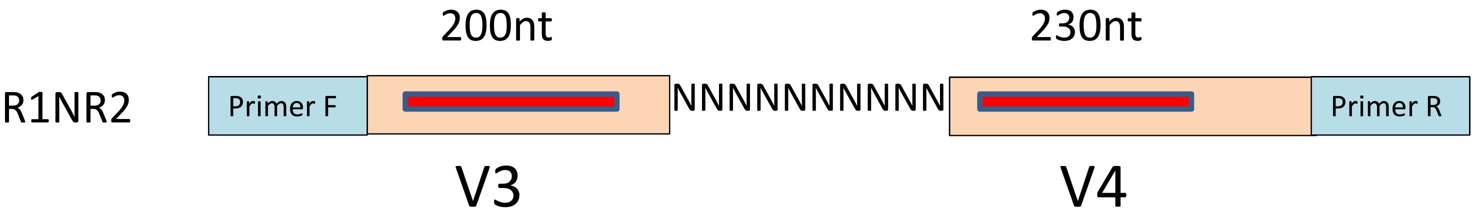
Prevotella bryantii
137 622 contigs →92% à 454 nt

Butyrivibrio fibrisolvens
122 510 contigs → 91% à 450 nt

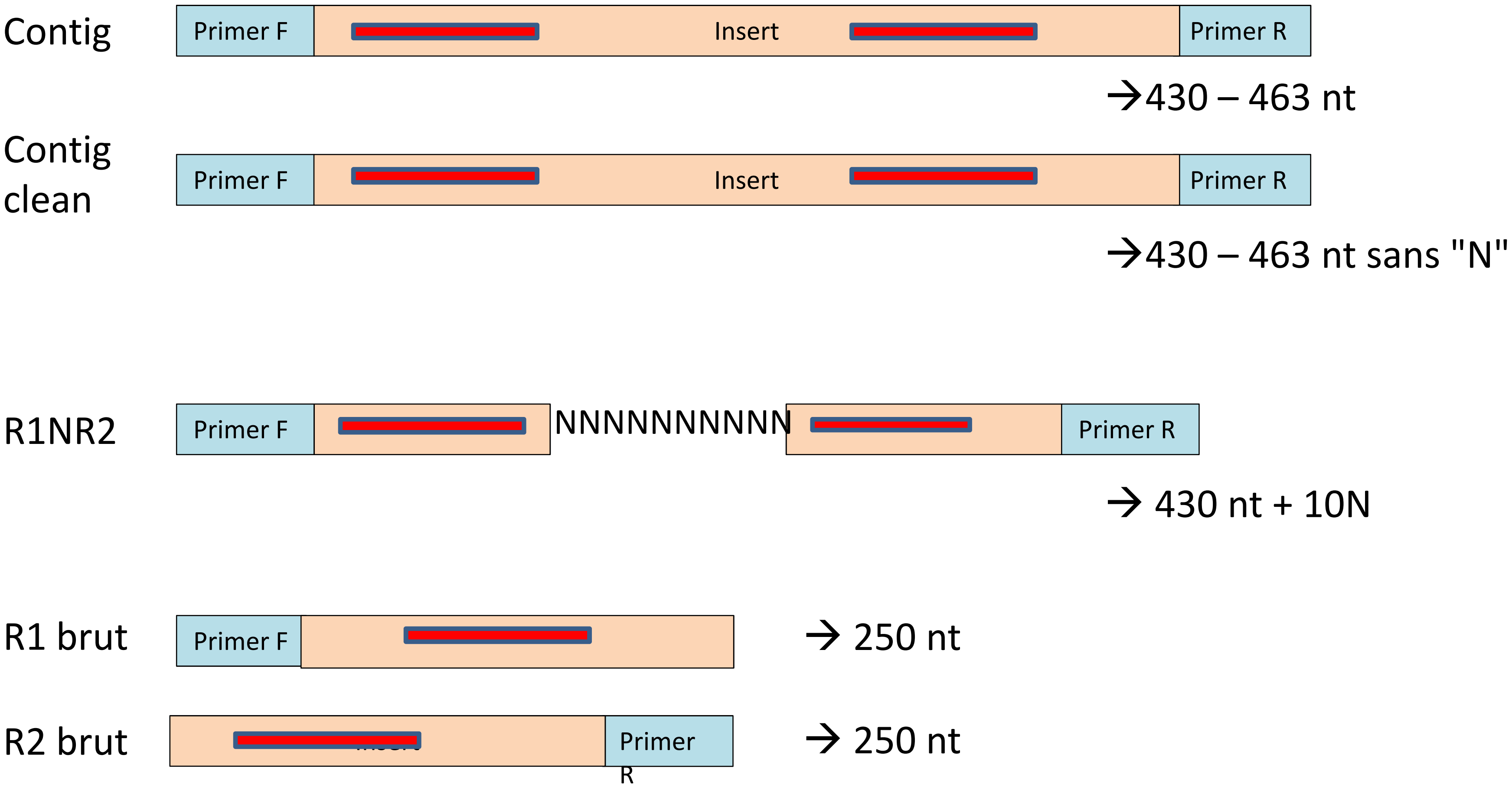
Plasmides les constructions comparées



Zone de contig ?



Plasmides les constructions comparées



Contig

→430 – 463 nt

Contig clean

→430 – 463 nt sans "N"

R1NR2

→ 430 nt + 10N

R1 brut

→ 250 nt

R2 brut

→ 250 nt

2 façons d'analyser les communautés

Reads

Assignation taxonomique

Mothur (RDP classifier) + Base LTP111

1 read= 1 taxon (bootstrap 85- 95 %)
Regroupement par chaine de caractères

Clusterisation = regroupement en OTU

Mothur (RDP classifier) + Base LTP111

+ Esprit Tree

Chaque read est classé dans 1 OTU puis assignation taxonomique des OTU

Table d'abondance

100 - 400 taxons connus ou «unclassified»

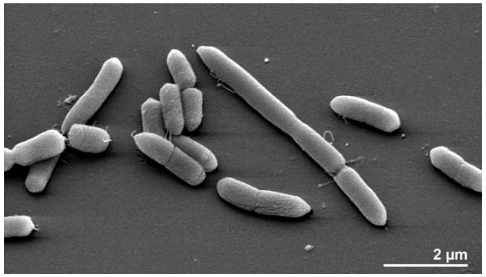
Y- matrix	Taxon 1	Taxon 2	Taxon 3	…	Taxon 196	Taxon n	TOTAL
sample 1	0	22	0	…	826	0	5399
sample 3	0	1	0	…	53	0	3155
sample 4	0	5	1	…	5694	0	7791
…	…	…	…	…	…	…	
sample 100	0	2	4	…	5943	1	11421

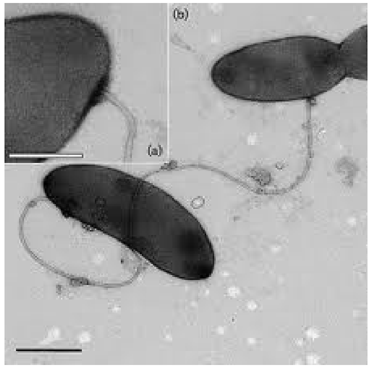
90 000 OTU

Y - matrix	OTU 1	OTU 2	OTU3	…	OTU3176	OTU n	TOTAL
sample 1	0	12	0	…	2320	0	7928
sample 3	0	64	0	…	76	28	16616
sample 4	0	748	36	…	52	0	5436
…	…	…	…	…			
sample 100	0	48	4	…	380	0	3248

Affiliation taxonomique (RDP classifier – LTP111)

	Prevotella bryantii				
Boot	Contig	ontig clean	R1NR2	1 (V3)	R2 (V4)
95	94.5	99.2	93.4	91.2	88.4
100	94.4	99.0	90.9	86.0	83.0





	Butyrivibrio fibrisolvens				
Boot	Contig	ontig clean	R1NR	1 (V3)	R2 (V4
95	94.2	99.1	90.4	88.9	80.8
100	88.2	91.2	74.1	72.4	31.0

Clusterisation en OTU (Esprit-tree) sur contig

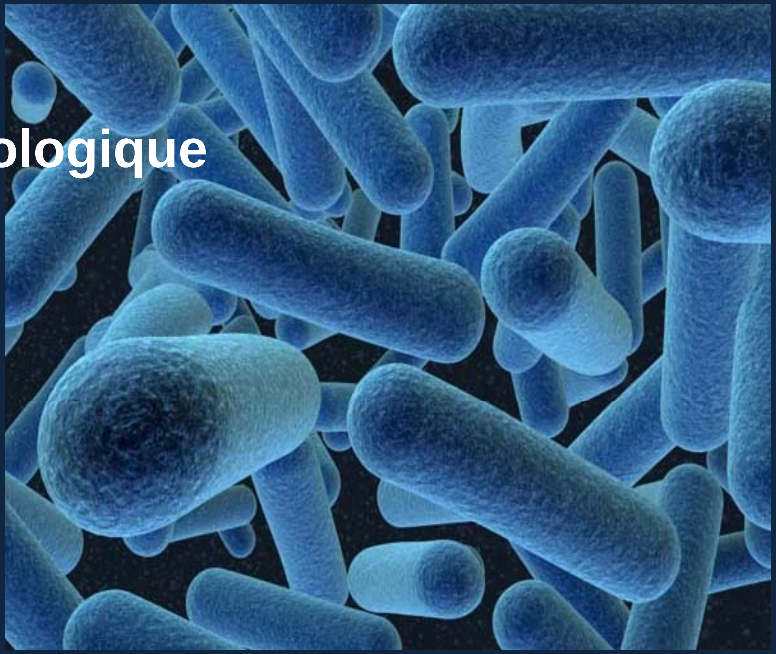
	Prevotella bryantii	***Butyrivibrio fibrisolvens***
Contig	137 622	122 510
Nb OTU attendu	1	1
Nb total OTU	1397	998
OTU 1 (maj)	93.6 %	93.6 %
OTU 2	0.2 %	0.2 %

To be continued…

- Contig clean
- R1NR2
- R1
- R2

Comparaison 454 vs MiSeq

Comparer les profondeurs
L’impact sur l’interprétation biologique



Comparaison de communautés digestives de porcs 454 vs Miseq

15 échantillons
3 lots : propre, sale, t1 (n=5)

454 (USA)

61 158 reads

Miseq (PLAGE)

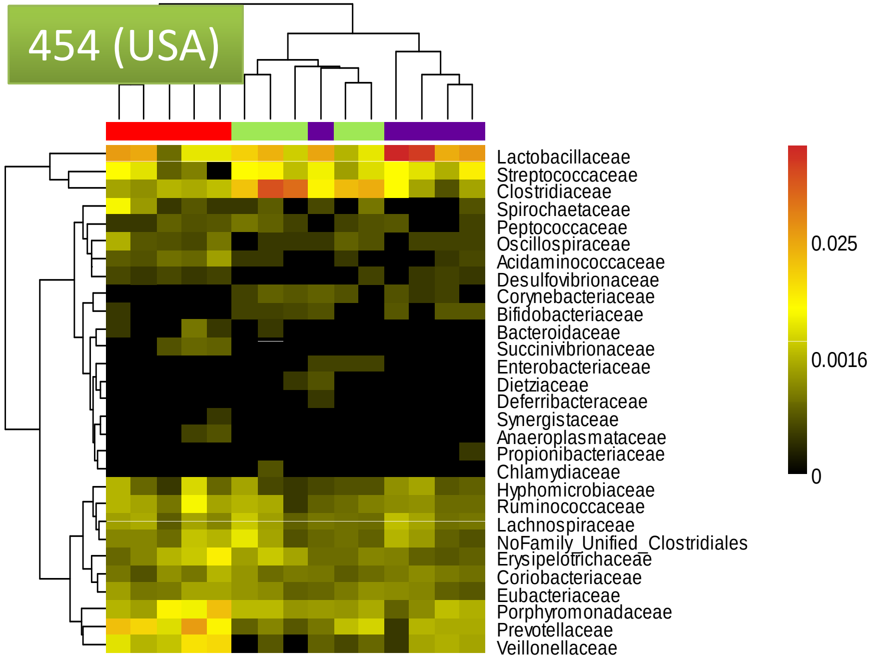
1 211 554 reads

Classification:
RDP classifier base LTP 111
2 niveaux taxonomiques :

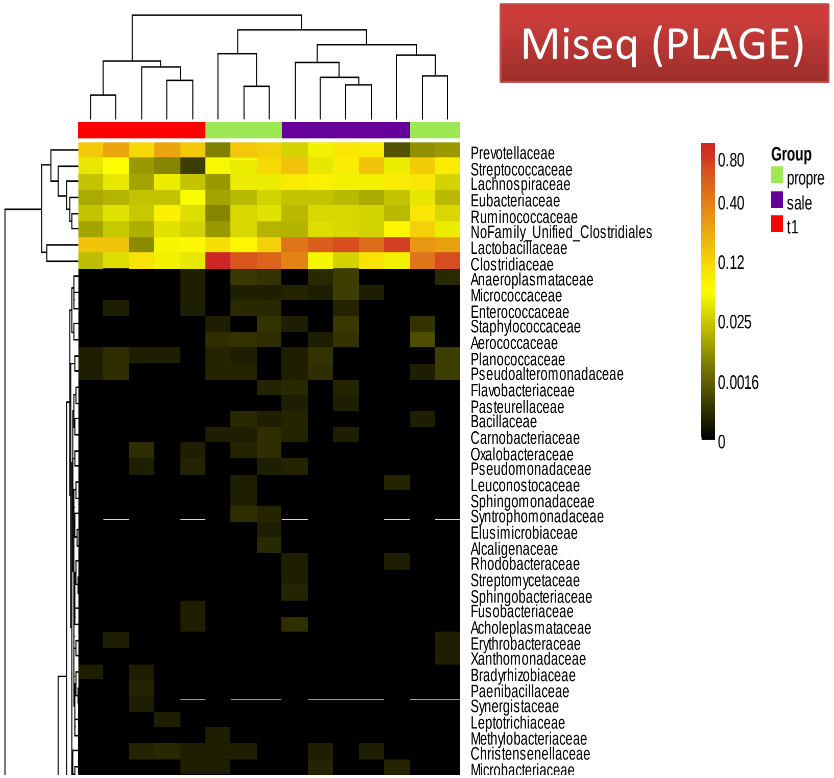
- famille (85%)
- genre (95%)

Clusterisation:
Esprit-tree
Avec préclusterisation par classification RDP

454 (USA)

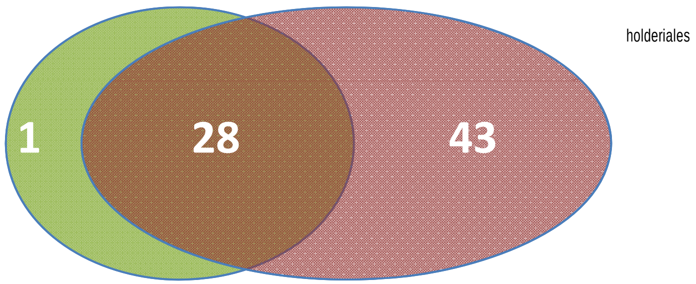


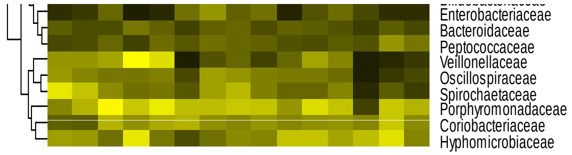
Miseq (PLAGE)



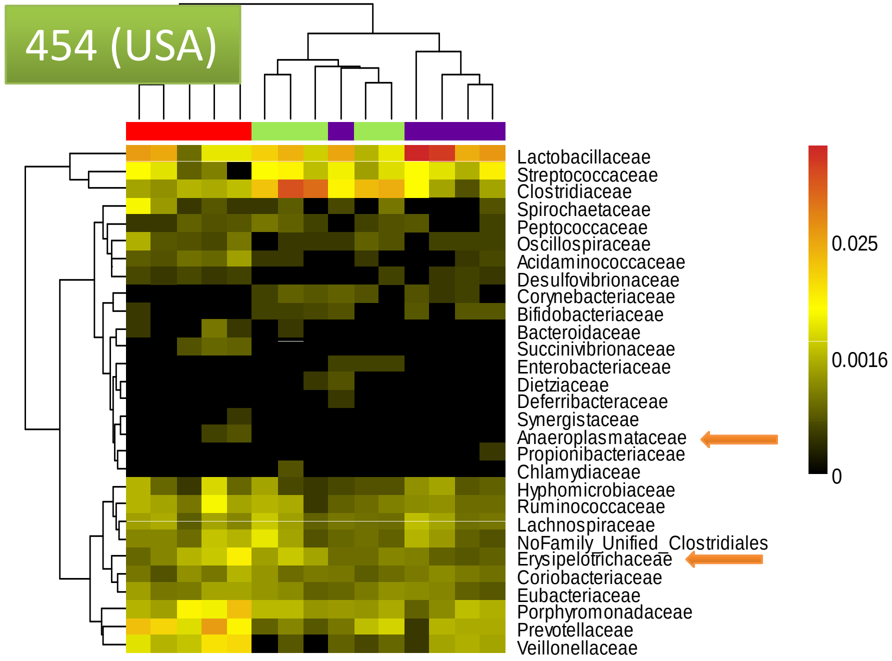
Classification (RDP classifier – LTP111) niv famille

	454	MISEQ
Nb seq tot	61158	1211554
Tx d'affiliation	85%	71%
Nb taxons	29 taxons	71 taxons

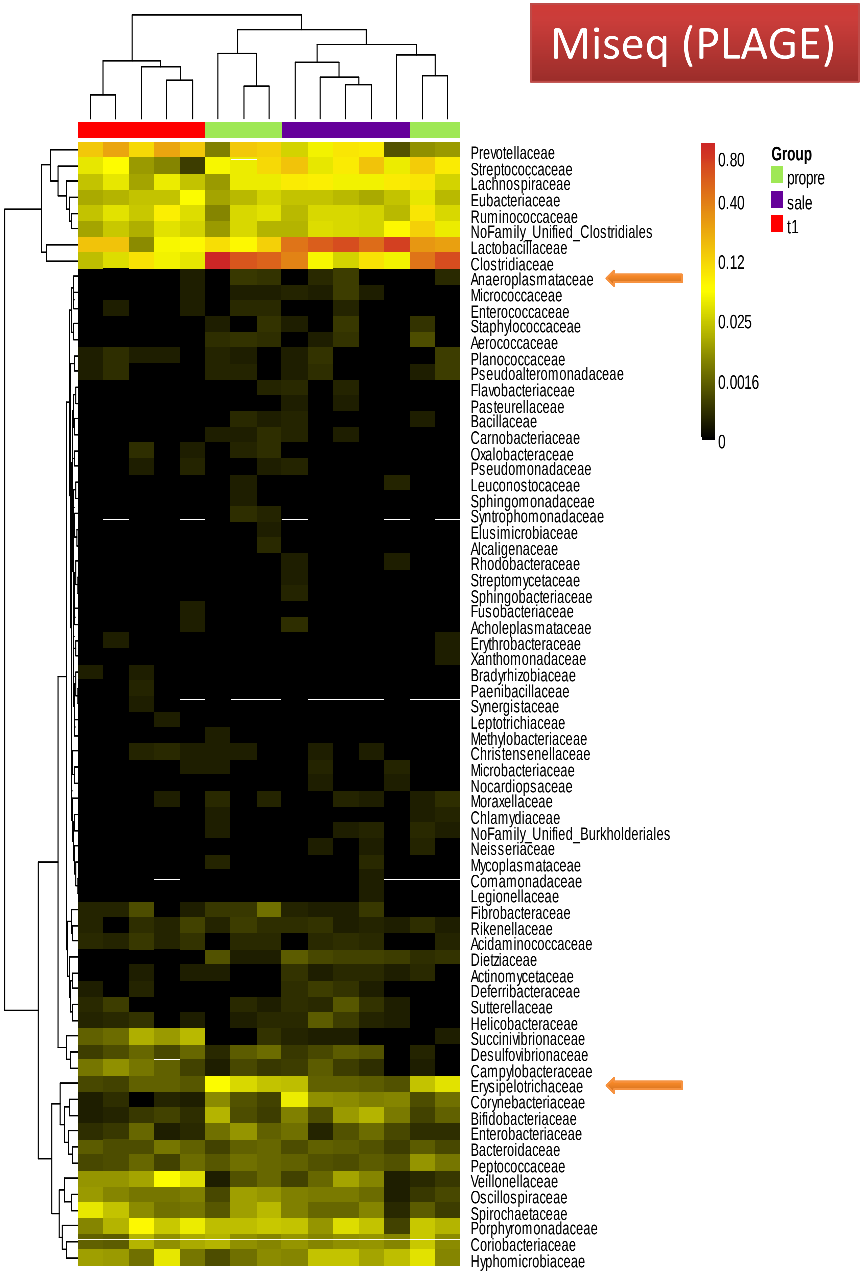


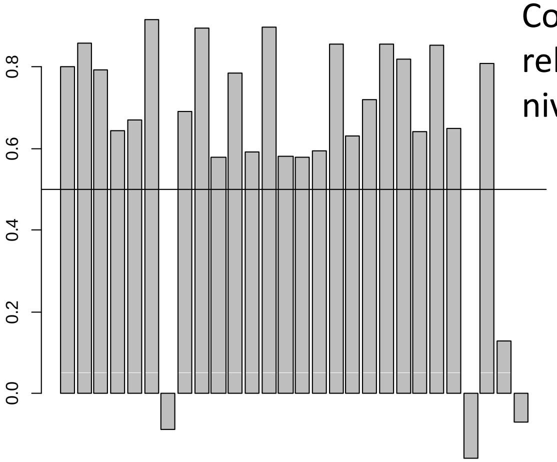


454 (USA)



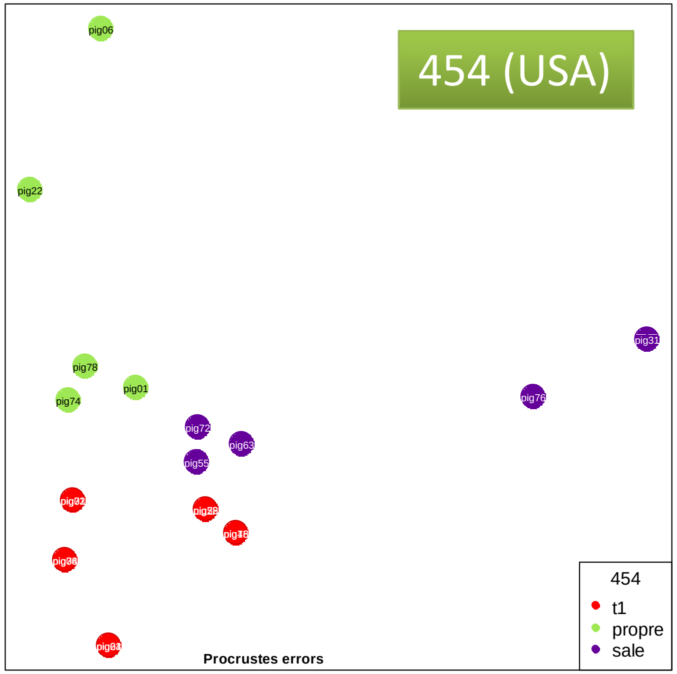
Miseq (PLAGE)



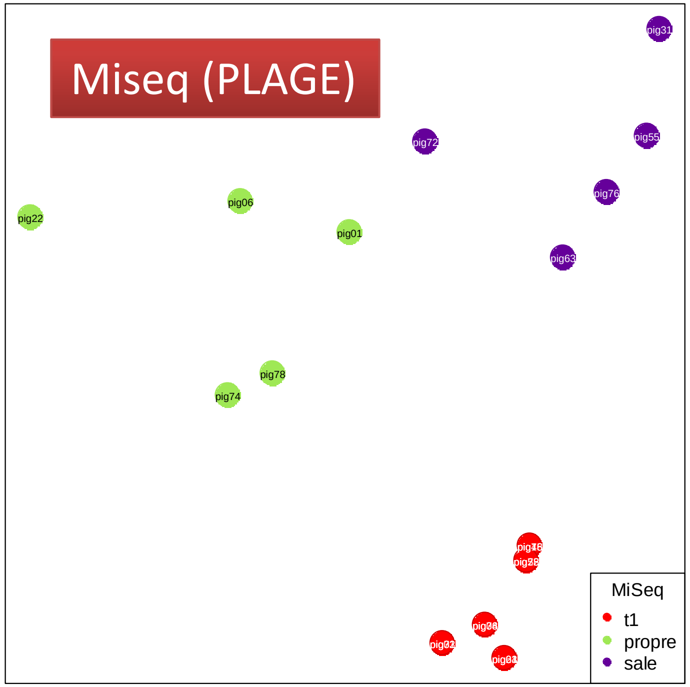


Corrélation abondance relative 454 vs MiSeq niveau famille

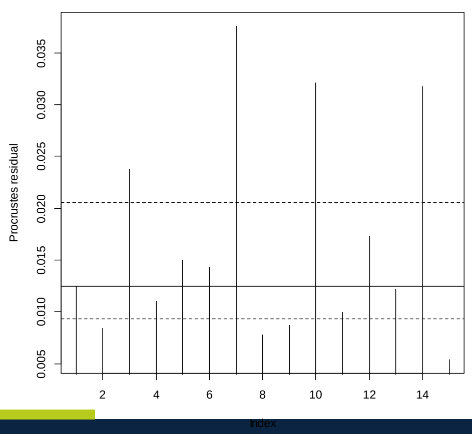
Comparaison des profiles taxonomiques par nMDS



454 (USA)

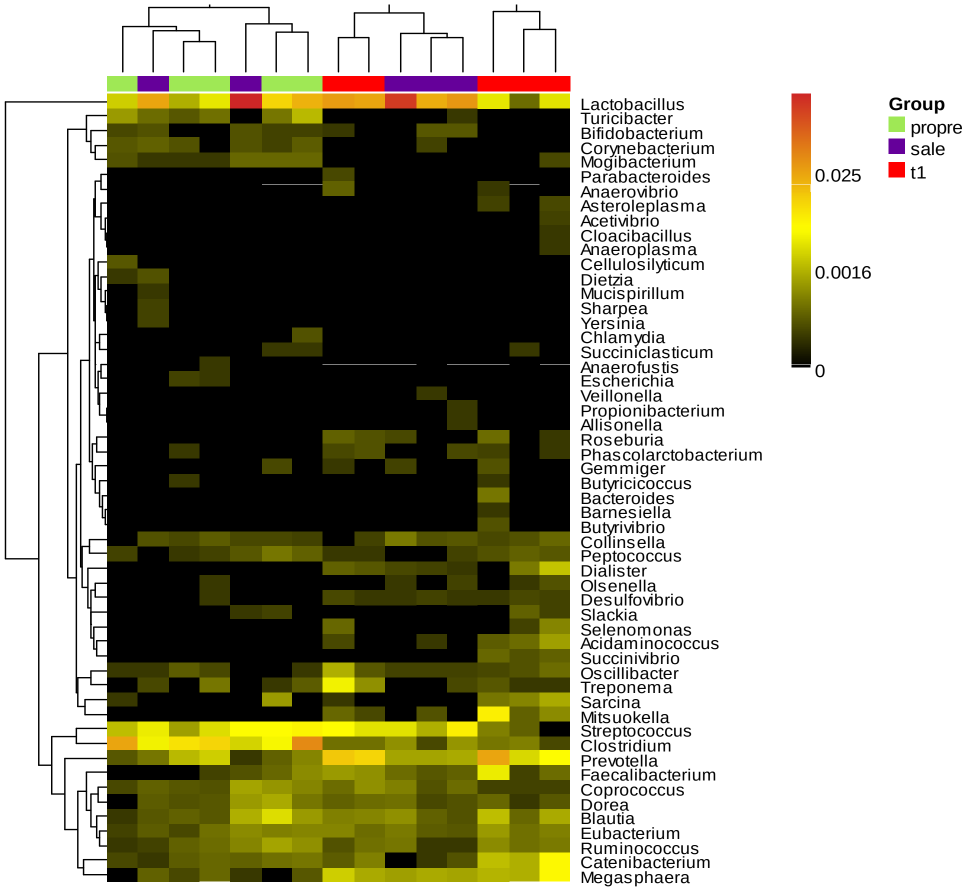


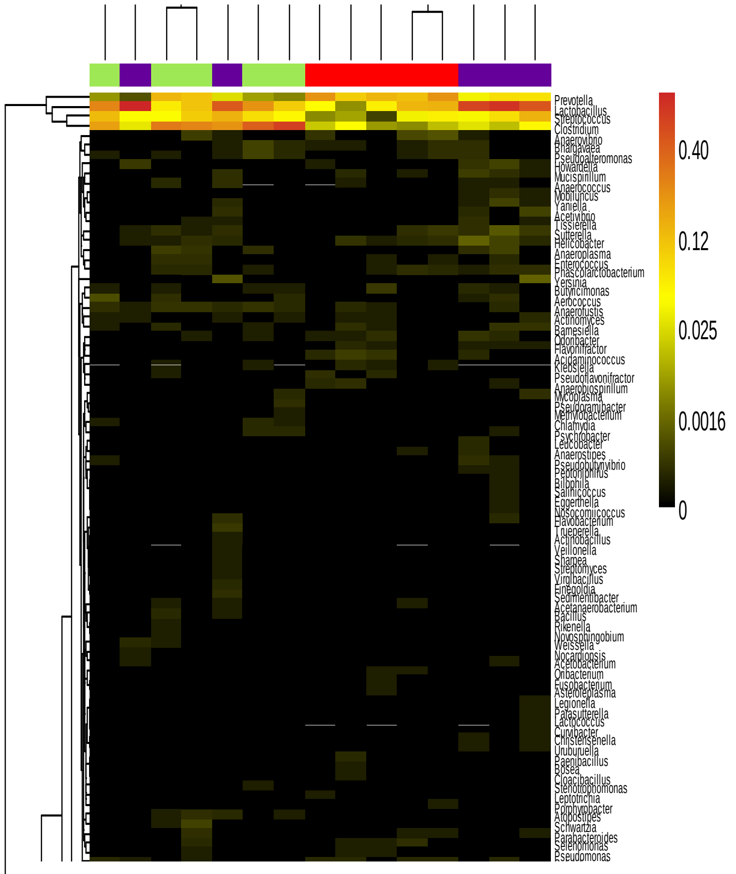
Miseq (PLAGE)



Classification (RDP classifier – LTP111)
niv famille

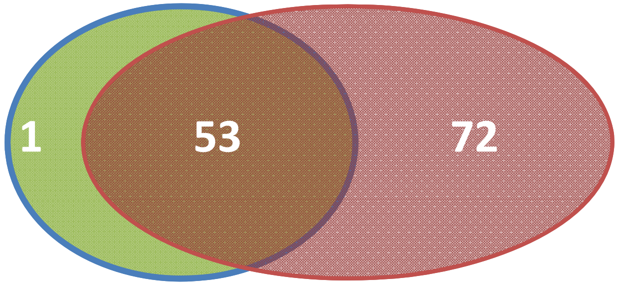
Représentation robuste
→ 4 individus ont un positionnement significativement différent

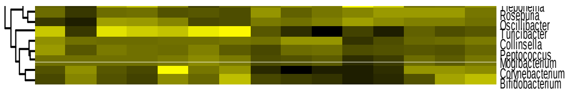


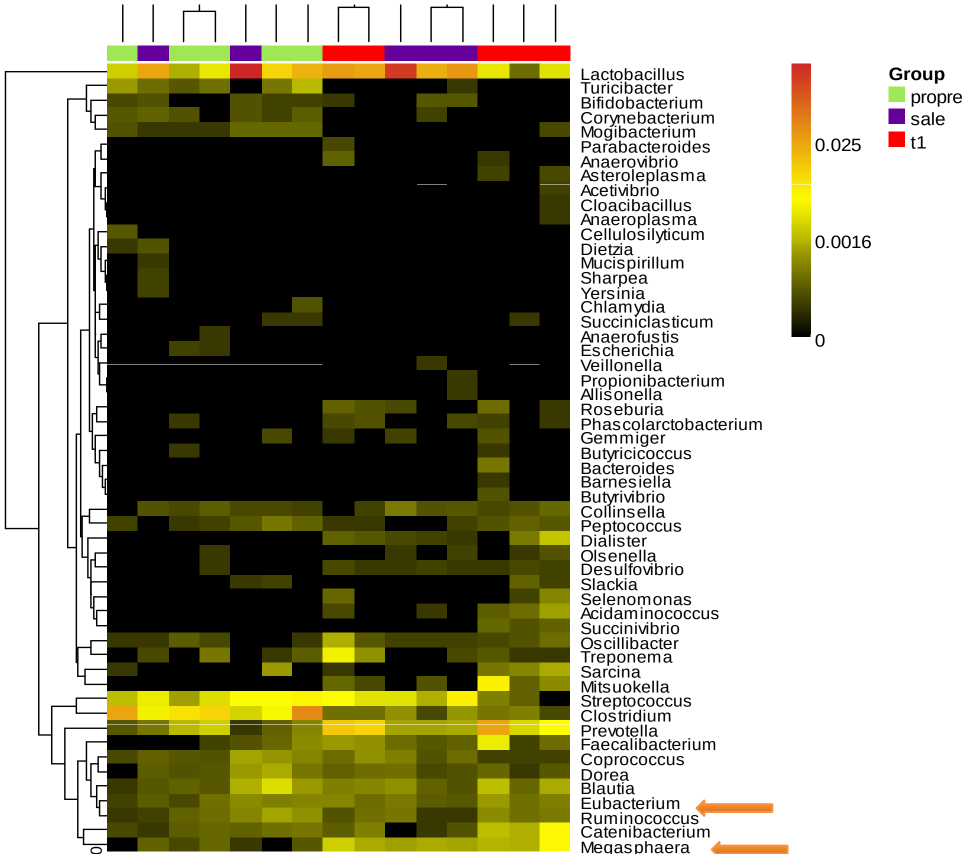


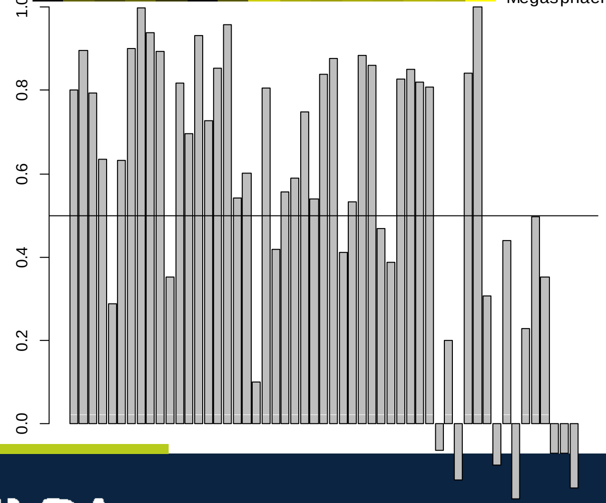
Classification (RDP classifier – LTP111) niv genre

	454	MISEQ
Nb reads	61 158	1 211 554
Tx d'affiliation	67 %	53 %
Nb taxons	54 taxons	125 taxons



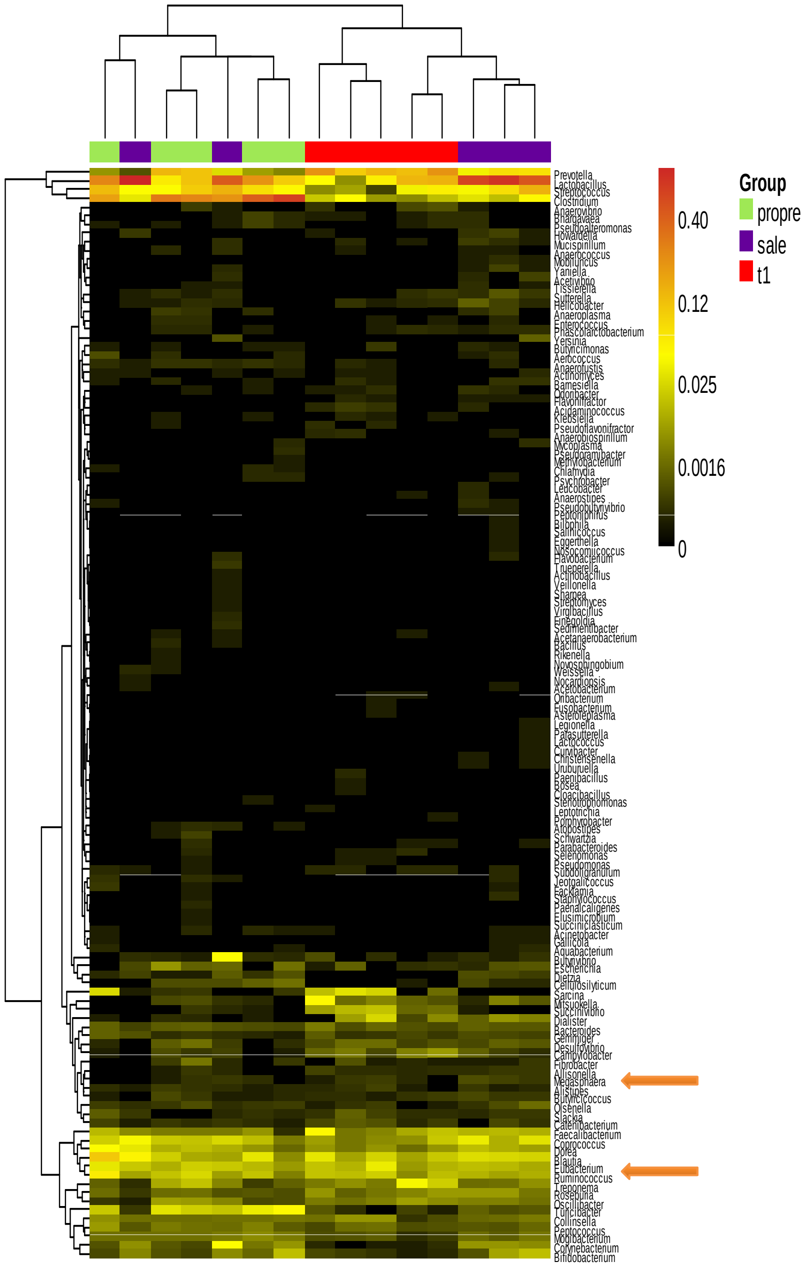






Megasphaera

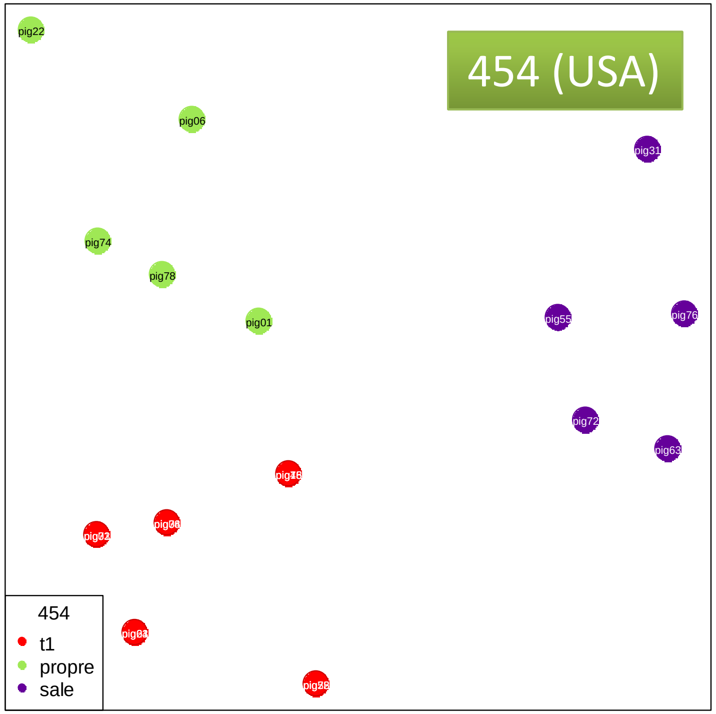
Eubacterium

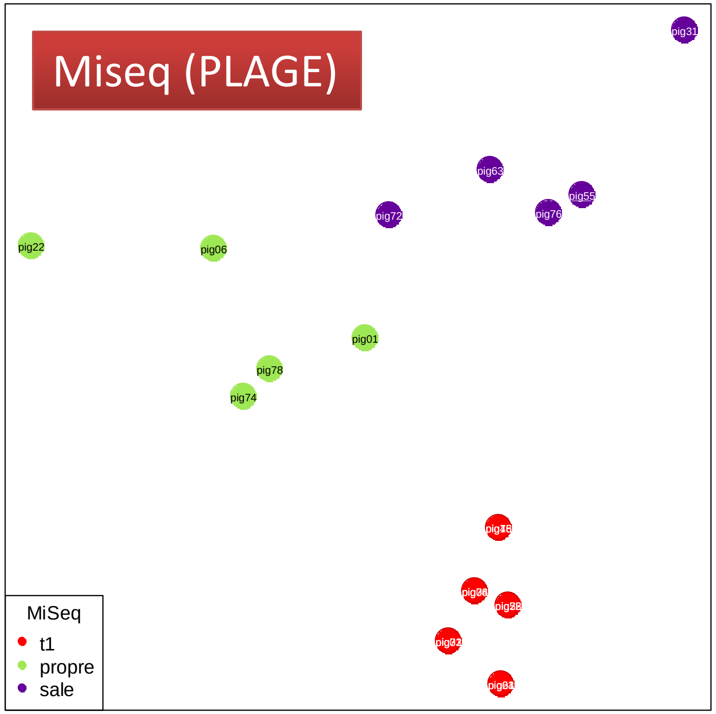


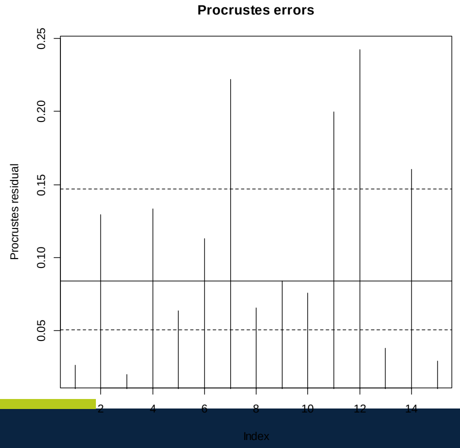
Clusterisation : Esprit Tree (0.03)

	454	MISEQ
Nb seq tot	61 392	1 227 651
Nb OTU	4 346	75 636
Singleton	56%	61%

Comparaison des profiles taxonomiques par nMDS







Clusterisation : Esprit Tree (0.03)

Représentation robuste
→ 4 individus ont un positionnement significativement différent

Conclusion

Miseq (PLAGE)

Qualité des séquences satisfaisantes
Contigage possible, suppression des bases N
Profondeur accrue
Moindre coût
Multiplexage plus important
Résultats cohérents d’un point de vue composition communauté

Mais

Difficulté de traitement dû à l’augmentation du nb de reads
→ Pipeline à revoir

Thanks for your attention

Merci pour votre attention

Thanks to each of my team members


Laurent Cauquil


Carole Bannelier


Valérie Lozano


Béatrice Gabinaud


Olivier Zemb


Muriel Segura


Vergnes Jean Philippe


Corine Bayourthe


Viviane Bataillier


Francis Enjalbert


Géraldine Pascal


Annabelle Troegeler


Marie Luce Schmidt

PECTOUL


Thierry Gidenne


Laurence Fortun-Lamothe


Christine Julien


Vargas Anita