



Training on Galaxy: Metabarcoding

October 2024 - Webinar

FROGS Practice on ITS data and Workflow creation — *correction* and *checkpoints*

LUCAS AUER, MARIA BERNARD, LAURENT CAUQUIL, MAHENDRA MARIADASSOU, GÉRALDINE PASCAL & OLIVIER RUÉ

Correction ITS

HISTORY : ITS_FORMATION_2023_PREV50



FROGS_1 Pre-process

Tool Parameters

Input Parameter	Value
Sequencer	illumina
Input type	archive
TAR archive file	1 : ITS_fast.tar.gz
Are reads already merged ?	paired
Reads 1 size	250 
Reads 2 size	250
Mismatch rate	0.1
Merge software	vsearch
Would you like to keep unmerged reads?	Yes, unmerged reads will be artificially combined.
Minimum amplicon size	180
Maximum amplicon size	490 
Do the sequences have PCR primers?	true
5' primer	CTTGGTCATTTAGAGGAAGTAA 
3' primer	GCATCGATGAAGAACGCAGC



FROGS_2 Clustering swarm

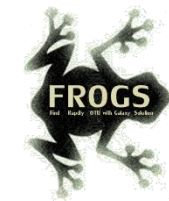
Tool Parameters

Input Parameter	Value
Sequences file	4 : FROGS_1 Pre-process: dereplicated.fasta
Count file	5 : FROGS_1 Pre-process: count.tsv
FROGS guidelines version	3.2
Aggregation distance clustering	1
Refine clustering	Yes, refine clustering with --fastidious swarm option

FROGS_3 Remove chimera

Tool Parameters

Input Parameter	Value
Sequences file (format: FASTA)	7 : FROGS_2 Clustering swarm: seed_sequences.fasta
Abundance type	biom
Abundance file (format: BIOM)	8 : FROGS_2 Clustering swarm: clustering_abundance.biom



FROGS_4 Cluster filters

Tool Parameters

Input Parameter	Value
Sequence file	11 : FROGS_3 Remove chimera: non_chimera.fasta
Abundance file	12 : FROGS_3 Remove chimera: non_chimera_abundance.biom
Minimum prevalence method	replicate
File of replicated sample names	3 : ITS_fast_replicates.tsv ★
Minimum prevalence	0.5 ★
Minimum cluster abundance as proportion or count. We recommend to use a proportion of 0.00005.	proportion
Minimum proportion of sequences abundance to keep cluster	5e-05 ★
N biggest clusters	Not available.
Search for contaminant clusters.	server
Contaminant databank	phiX



FROGS ITSx[★]

Tool Parameters

Input Parameter	Value
Sequence file	16 : FROGS_4 Cluster filters: clusterFilters_sequences.fasta
Abundance file	15 : FROGS_4 Cluster filters: clusterFilters_abundance.biom
Trim conserved sequence (SSU, 5.8S, LSU) ?	yes
Choose pertinent organisms to scan:	Fungi

FROGS_5 Taxonomic affiliation

Tool Parameters

Input Parameter	Value
Using reference database	ITS_UNITE_Fungi_8.3 [★]
Also perform RDP assignation?	No
Taxonomic ranks	Domain Phylum Class Order Family Genus Species
Sequence file	20 : FROGS ITSx: ITS_sequence.fasta
Abundance file	21 : FROGS ITSx: itsx_abundance.biom



FROGS_6_Affiliation_Stat

Tool Parameters

Input Parameter	Value
Abundance file	23 : FROGS_5 Taxonomic affiliation: affiliation_abundance.biom
Taxonomic ranks	Domain Phylum Class Order Family Genus Species
Rarefaction ranks	Class Order Family Genus Species
Affiliation processed	FROGS_blast



FROGS Tree

Tool Parameters

Input Parameter	Value
Sequence file	20 : FROGS ITSx: ITS_sequence.fasta
Biom file	23 : FROGS_5 Taxonomic affiliation: affiliation_abundance.biom



FROGS BIOM to TSV

Tool Parameters

Input Parameter	Value
Abundance file	23 : FROGS_5 Taxonomic affiliation: affiliation_abundance.biom
Sequences file (optional)	20 : FROGS ITSx: ITS_sequence.fasta
Extract multi-alignments	Yes

FROGSSTAT Phyloseq Import Data

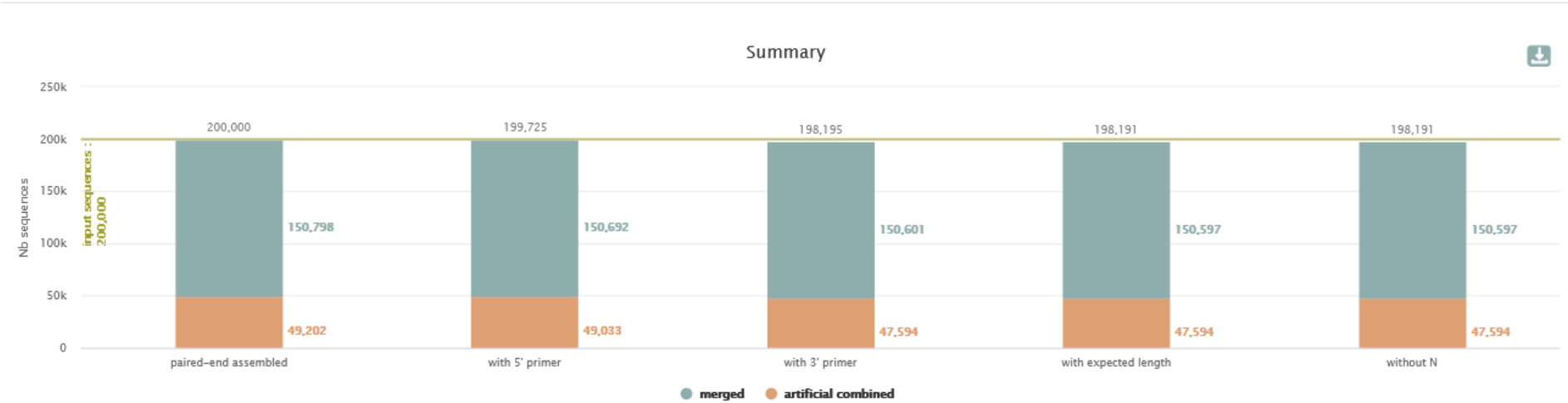
Tool Parameters

Input Parameter	Value
Abundance biom file with taxonomical metadata (format: BIOM)	23 : FROGS_5 Taxonomic affiliation: affiliation_abundance.biom
Metadata associated to samples (format: TSV)	2 : ITS_fast_metadata.tsv 
Taxonomic tree file (format: Newick)	26 : FROGS Tree: tree.nwk
Names of taxonomic levels	Kingdom Phylum Class Order Family Genus Species
Do you want to normalise your data ?	Yes, subsample abundances to the smallest sample size. 

Points de vigilance

FROGS_1 Pre-process

Preprocess summary



Une quantité importante de artificial combined

Details on merged sequences

Show

10

entries

Search:

CSV

Samples

before process

% kept

paired-end assembled

with 5' primer

with 3' primer

with expected length

without N

Ph414

10,000

64.01

6,408

6,405

6,401

6,401

Ph417

10,000

69.83

6,990

6,987

6,984

6,983

Ph407

10,000

71.81

7,188

7,187

7,181

7,181

7,181

Samples

before process

% kept

paired-end assembled

with 5' primer

with 3' primer

with expected length

without N

Ph217

10,000

80.48

8,061

8,052

8,048

8,048

Ph237

10,000

79.59

7,967

7,964

7,959

7,959

7,959

Ph203

10,000

79.42

7,954

7,948

7,942

7,942

7,942

Des échantillons d'unmerged

Et d'autres à ~2

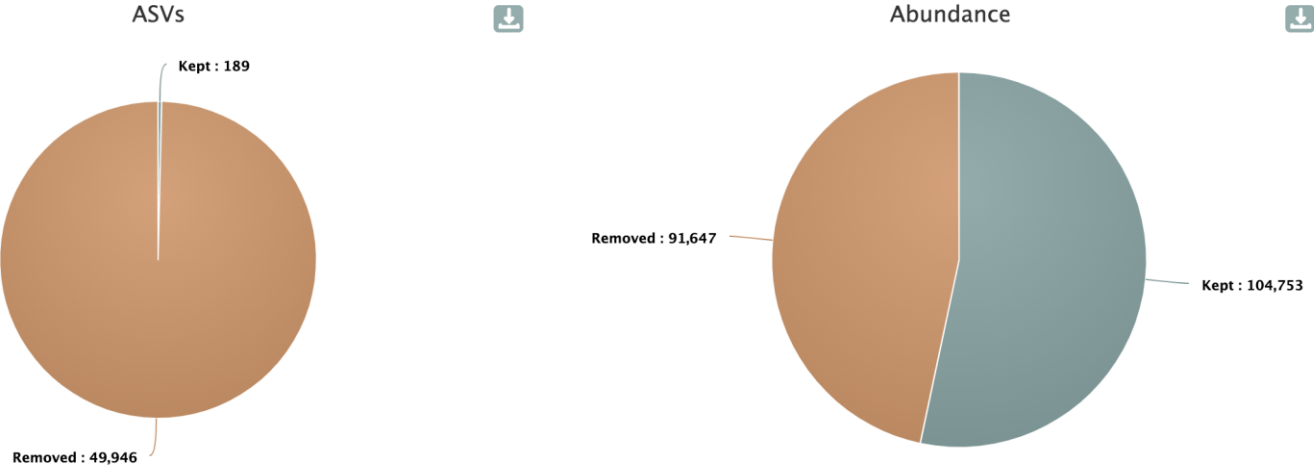
Des échantillons avec >30% d'unmerged

Et d'autres à ~20%

FROGS_4 Cluster filters

Filtres à 50% de prévalence par groupe de réplicats :

Filters summary



Grosse perte de séquences, filtre à ajuster ou modifier?

Filters intersections

Draw a Venn to see which ASVs had been deleted by the filters chosen (Maximum 6 options):

- ☐ Present in less than 50.0% of replicates of all replicate groups.
- ☐ Abundance < 0.005% (i.e 10 sequences)
- ☐ Present in databank of contaminants

 Venn

FROGSSTAT Phyloseq Import Data

FROGSSTAT Phyloseq Import Data



[Summary](#) [Ranks Names](#) [Sample metadata](#) [Plot tree](#)

```
phyloseq-class experiment-level object
otu_table() OTU Table:      [ 186 taxa and 20 samples ]
sample_data() Sample Data:  [ 20 samples by 8 sample variables ]
tax_table() Taxonomy Table: [ 186 taxa by 14 taxonomic ranks ]
phy_tree()  Phylogenetic Tree: [ 186 tips and 185 internal nodes ]
```

Number of sequences in each sample after normalisation: 1454

[Summary](#) [Ranks Names](#) [Sample metadata](#) [Plot tree](#)

Show

Sample variables: kept, Replicas, Incubation, Nitrogen, Forest_management, Quality, Treatment, SampleID

Show

kept : 79.76, 77.64, 80.26, 78.65, 77.18, 79.68, 78.7, 76.38, 76.37, 77.37, 72.52, 64.98, 78.13, 71.17, 75.2, 73.48, 73.21, 74.01, 74.15, 73.77

Replicas : 3, 2, 5, 1, 4

Incubation : T4

Nitrogen : Nitrogen_supplementation, Control

Forest_management : Control, OMR

Quality : Low degradability

Treatment : Control_with_N, Control, OMR_with_N, OMR

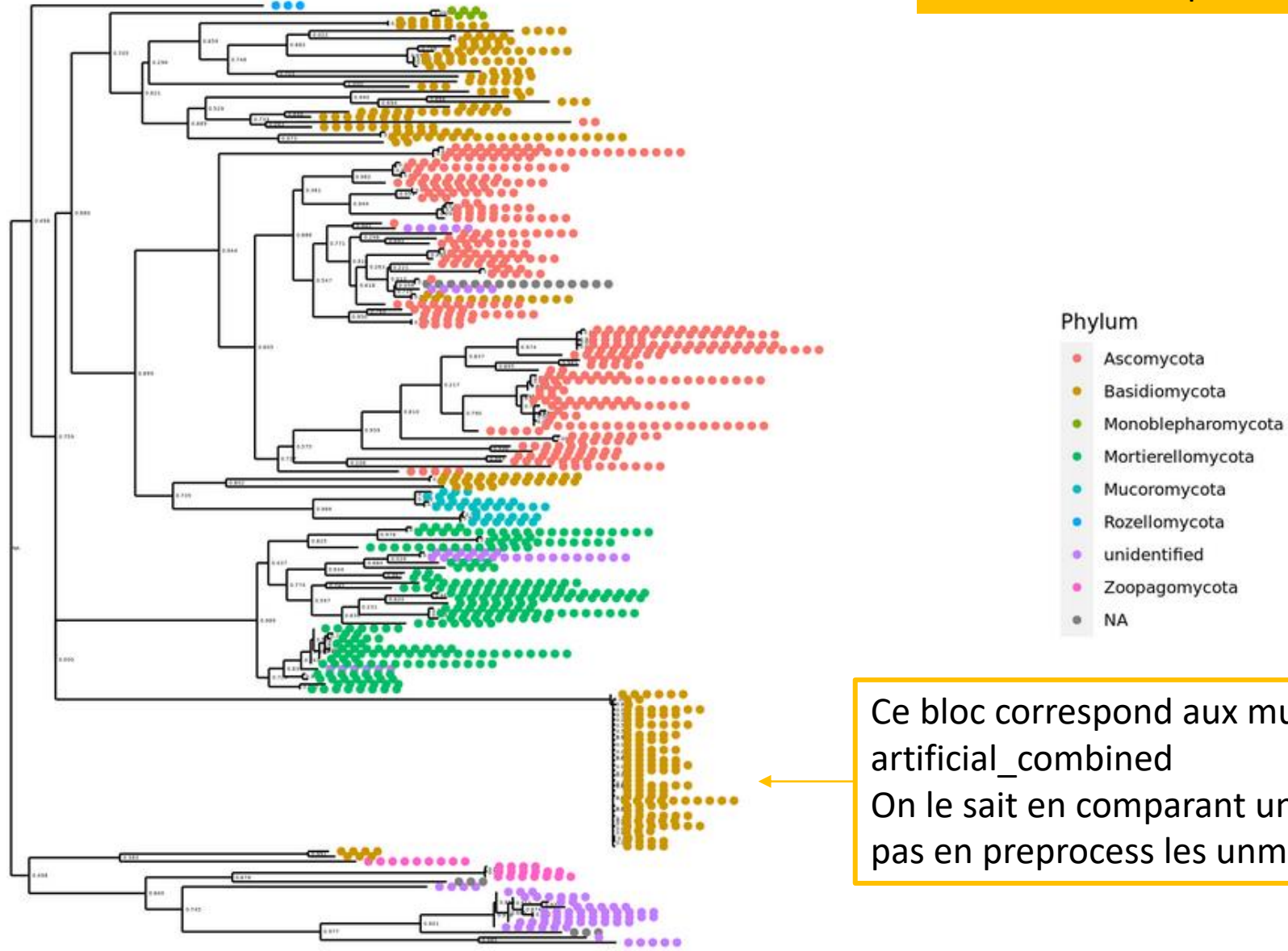
SampleID : Ph203, Ph212, Ph217, Ph222, Ph224, Ph237, Ph241, Ph243, Ph246, Ph250, Ph407, Ph414, Ph415, Ph417, Ph423, Ph428, Ph433, Ph434, Ph439, Ph449

[Summary](#) [Ranks Names](#) [Sample metadata](#) [Plot tree](#)

Show

Rank names : Kingdom, Phylum, Class, Order, Family, Genus, Species, Rank2, Rank3, Rank4, Rank5, Rank6, Rank7, Rank1

Phylogenetic tree colored by Phylum



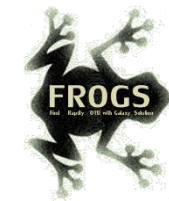
Le phylum des Basidiomycota est éclaté en plusieurs endroits de l'arbre, données Unifrac à considérer avec précautions

Ce bloc correspond aux multiples clusters artificial_combined
On le sait en comparant un historique n'acceptant pas en preprocess les unmerged.

Pistes stat pour ITS

SANS POST-AFFILIATION (ET DONC SANS REGROUPEMENT DE CERTAINS ARTIFICIAL_COMBINED)

HISTORY ITS_FORMATION_2023_PREV50



FROGSSTAT Phyloseq Composition Visualisation

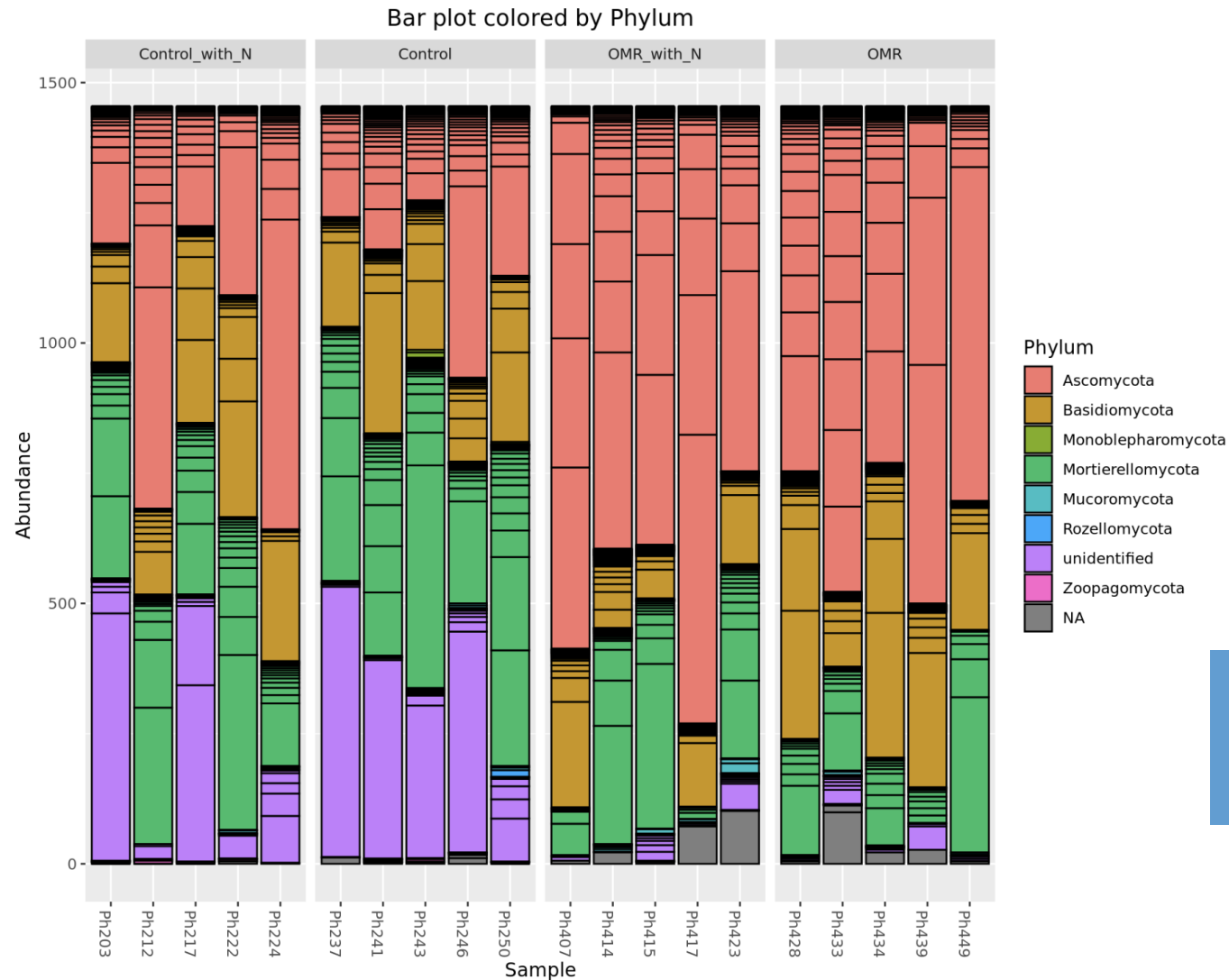
Tool Parameters

Input Parameter	Value
Phyloseq object (format rdata)	30 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
Grouping variable	Treatment
Taxonomic level to filter your data	Kingdom
Taxa (at the above taxonomic level) to keep in the dataset	Fungi
Taxonomic level used for aggregation	Phylum
Number of most abundant taxa to keep	9

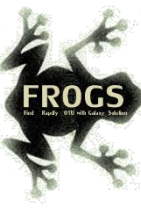
Interprétation des barplots :

La plupart des outils de visualisation FROGS n'autorisent qu'une seule variable de tri.

Utiliser la variable Treatment (qui combine les deux variables Forest_management et Nitrogen) permet de « tricher » pour afficher deux variables dans les outils de visualisation



Différences évidentes entre Control et OMR
(augmentation des Ascomycota)
Moins évident entre with N et sans



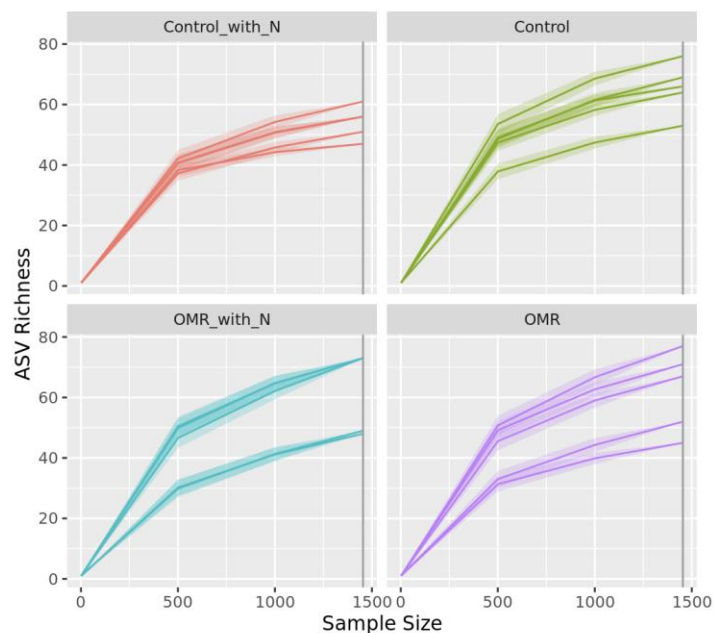
FROGSSTAT Phyloseq Alpha Diversity

Tool Parameters

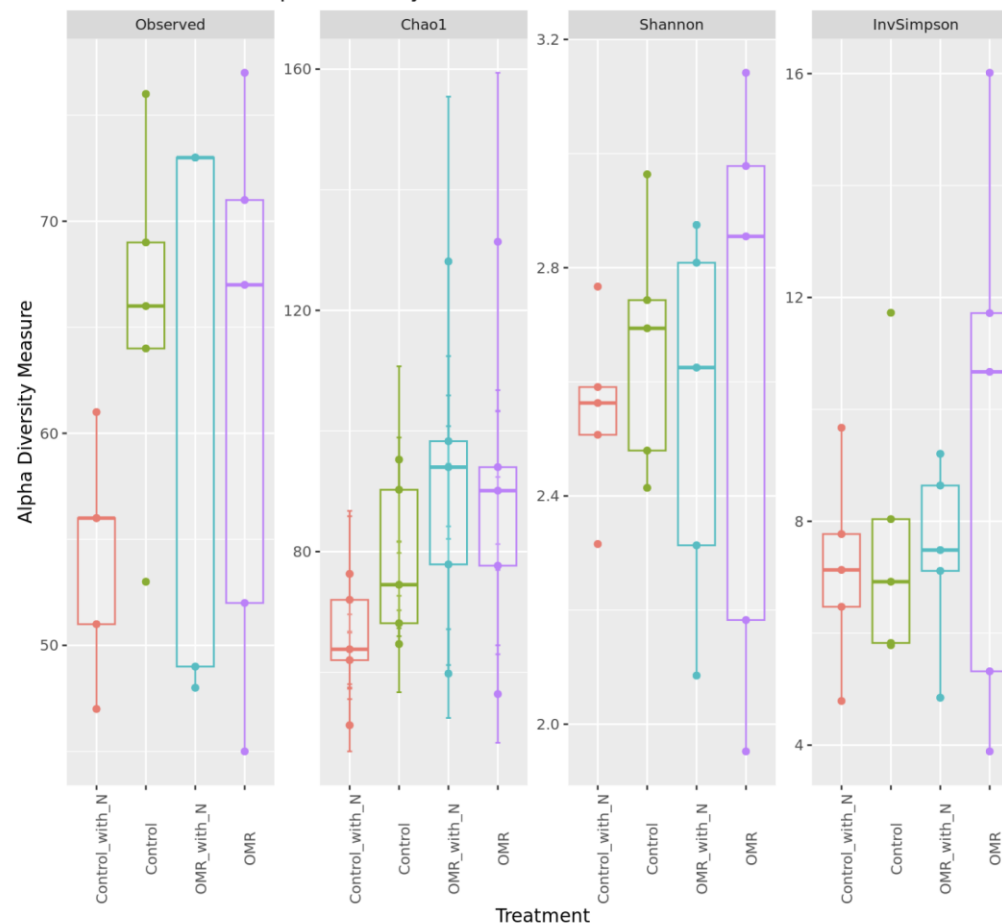
Input Parameter	Value
Phyloseq object (format: RData)	30 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
Experiment variable	Treatment
The alpha diversity indices to compute	Observed Chao1 Shannon InvSimpson

[Show](#)

Rarefaction curves



Alpha diversity distribution in function of Treatment



Richness plot with boxplot

Alpha Diversity Indice Anova Analysis

Treatment

- Control_with_N
- Control
- OMR_with_N
- OMR

```
#####  
#Perform ANOVA on Observed, which effects are significant  
anova.Observed <-aov( Observed ~ Depth + Treatment, anova_data)  
summary(anova.Observed)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	368.5	122.8	1.071	0.389
Residuals	16	1836.0	114.8		

```
#####  
#Perform ANOVA on Chao1, which effects are significant  
anova.Chaol <-aov( Chaol ~ Depth + Treatment, anova_data)  
summary(anova.Chaol)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	2258	752.7	1.795	0.189
Residuals	16	6707	419.2		

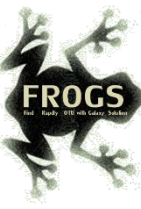
```
#####  
#Perform ANOVA on Shannon, which effects are significant  
anova.Shannon <-aov( Shannon ~ Depth + Treatment, anova_data)  
summary(anova.Shannon)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	0.049	0.01632	0.142	0.933
Residuals	16	1.841	0.11508		

```
#####  
#Perform ANOVA on InvSimpson, which effects are significant  
anova.InvSimpson <-aov( InvSimpson ~ Depth + Treatment, anova_data)  
summary(anova.InvSimpson)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	17.06	5.688	0.623	0.61
Residuals	16	145.98	9.124		

Pas d'effet significatif du traitement sur aucun indice.
Profondeur de séquençage très faible (sous-échantillon pour le TP
et effets des filtres à 50%)

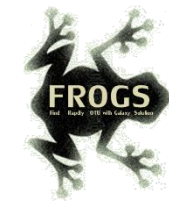


FROGSSTAT Phyloseq Beta Diversity

Tool Parameters

Input Parameter	Value
Phyloseq object (format: RData)	30 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
Grouping variable	Treatment
The methods of beta diversity	Unifrac Weighted Unifrac Bray-Curtis Jaccard (as cc method in betadiver vegan fonction)
Other method	Not available.

Cet outil sert à produire des matrices pour ensuite faire la clusterisation et la visualisation

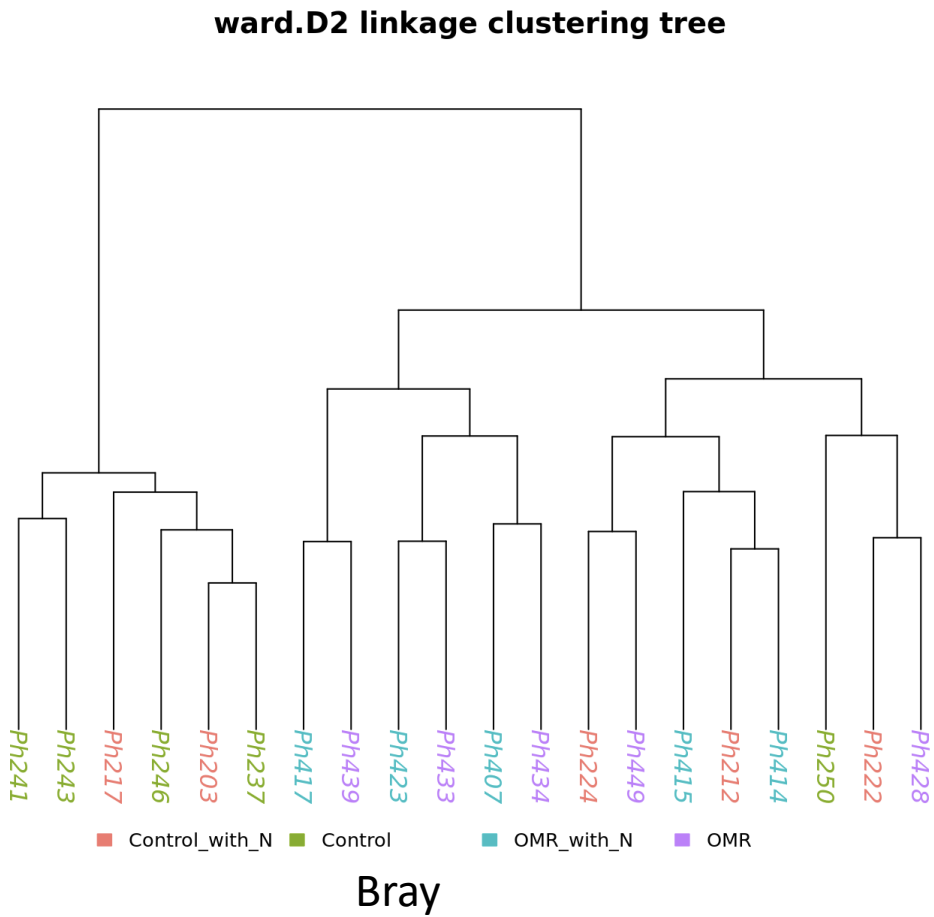


FROGSSTAT Phyloseq Sample Clustering

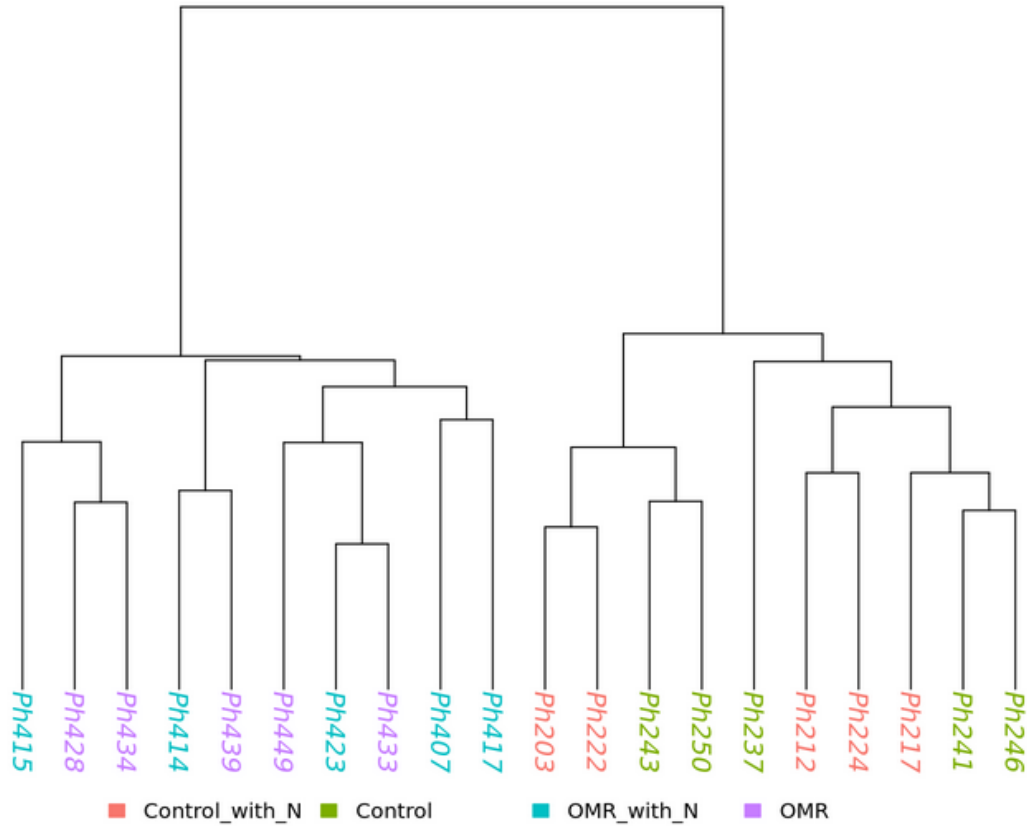
Tool Parameters

Input Parameter	Value
Phyloseq object (format: RData)	30 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
The beta diversity distance matrix file	39 : FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (cc.tsv)
Experiment variable	Treatment

Diversité bêta – clustering :

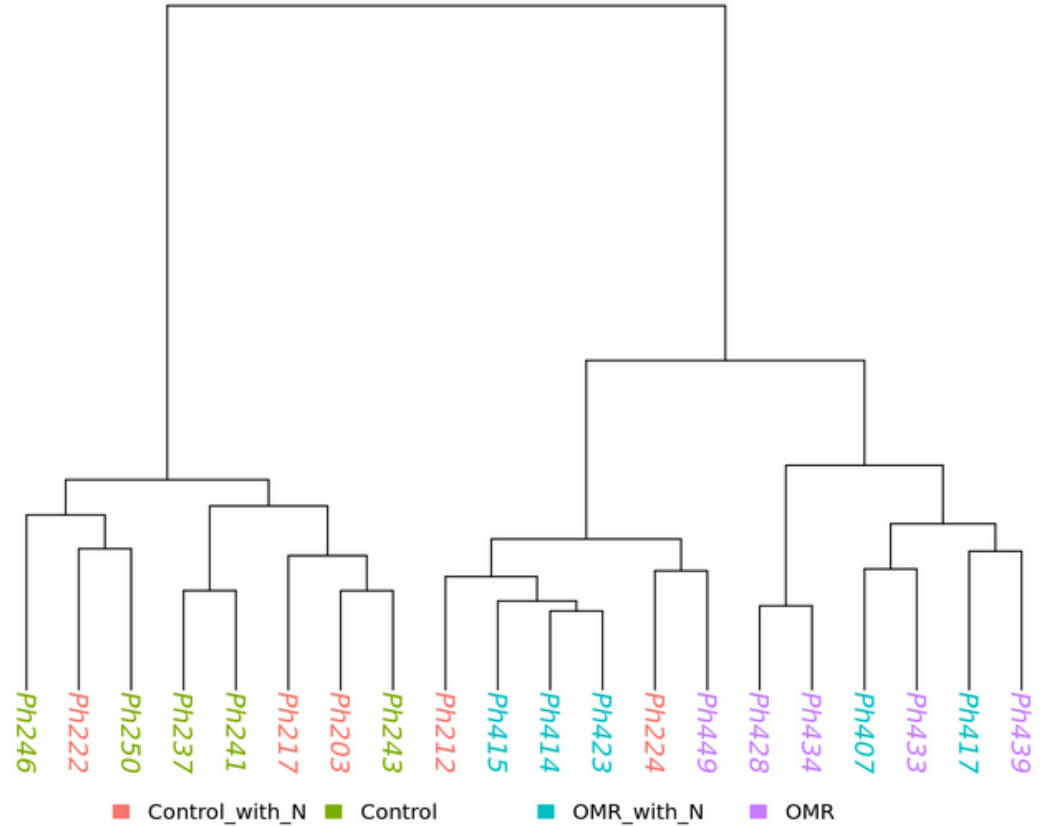


ward.D2 linkage clustering tree



Unifrac

ward.D2 linkage clustering tree



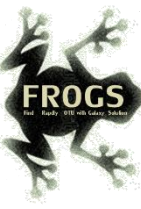
wUnifrac

Différences entre Control et OMR mais pas avec l'ajout d'azote.

Meilleur clustering avec les méthodes qualitatives qu'avec les méthodes quantitatives :

les différences sont plutôt dues à des ASV peu abondants

(quand on prend en compte les abondances les échantillons se ressemblent davantage)

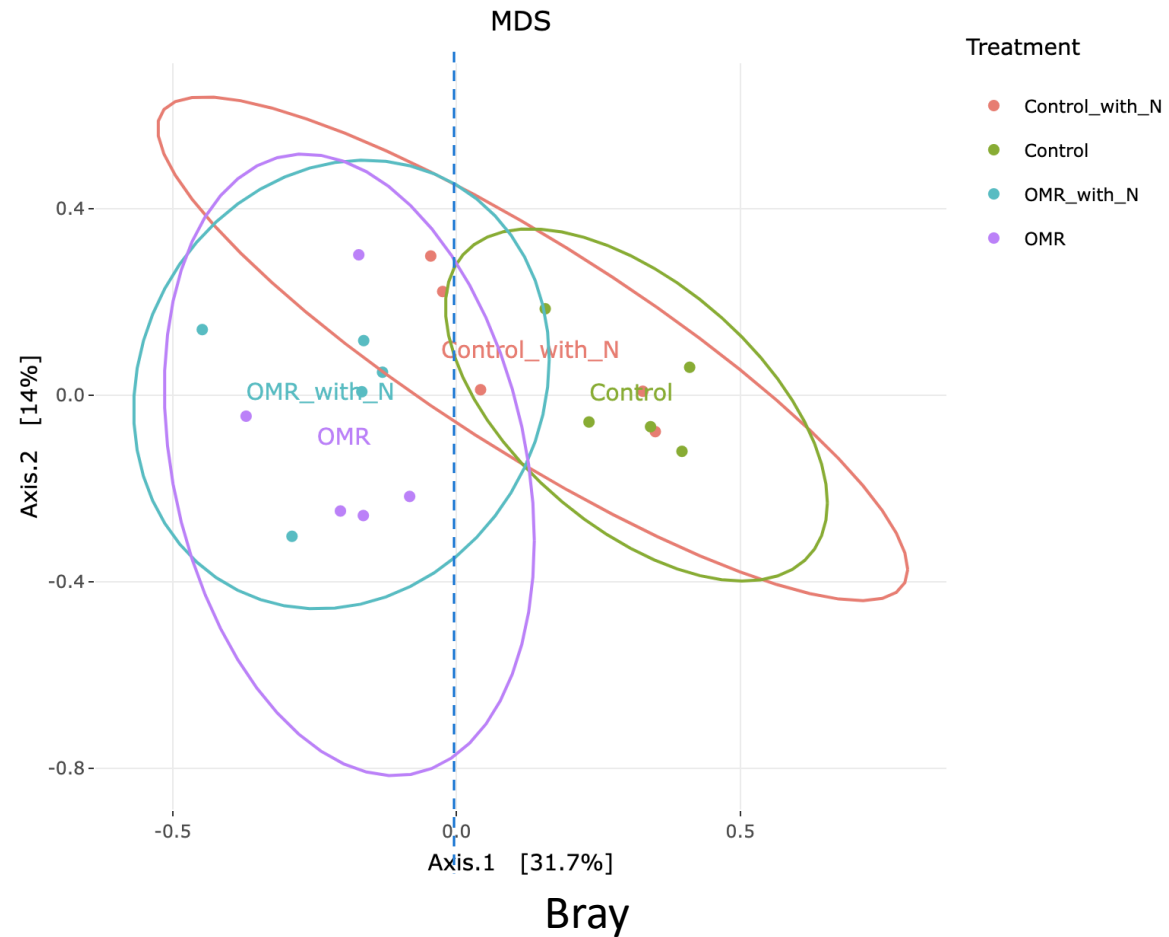
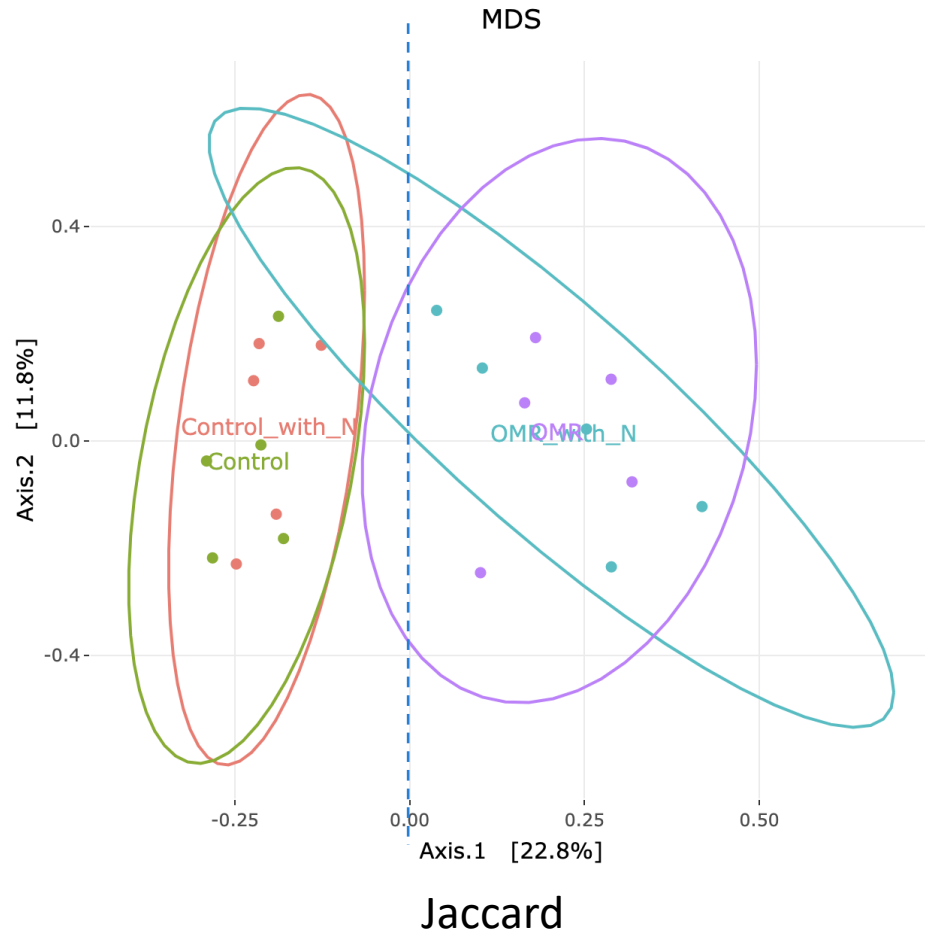


FROGSSTAT Phyloseq Structure Visualisation

Tool Parameters

Input Parameter	Value
Phyloseq object (format rdata)	30 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
The beta diversity distance matrix file	41 : FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (wunifrac.tsv)
Experiment variable	Treatment
Ordination method	MDS/PCoA

Diversité bêta – ordination :



Différences entre Control et OMR mais pas avec l'ajout d'azote.

Meilleure séparation (sur l'axe1) avec les méthodes qualitatives qu'avec les méthodes quantitatives :
les différences sont plutôt dues à des ASV peu abondants
(quand on prend en compte les abondances les échantillons se ressemblent davantage)



FROGSSTAT Phyloseq Multivariate Analysis Of Variance

Tool Parameters

Input Parameter	Value
Phyloseq object (format: RData)	30 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
The beta diversity distance matrix file	37 : FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (cc.tsv)
Experiment variable	Forest_management*Nitrogen

Diversité bêta – PERMANOVA :

Il est ici possible de donner plusieurs variables séparées par des + ou * (effets additifs ou avec interaction), et donc plus pertinent d'utiliser Nitrogen et Forest_management plutôt que Treatment

```
adonis(formula = dist ~ Forest_management * Nitrogen, data = metadata,
        permutations = 9999)
```

Permutation: free
Number of permutations: 9999

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	0.9571	0.95706	4.5861	0.20329	0.0001 ***
Nitrogen	1	0.2424	0.24239	1.1615	0.05149	0.2254
Forest_management:Nitrogen	1	0.1694	0.16944	0.8119	0.03599	0.7190
Residuals	16	3.3390	0.20869		0.70924	
Total	19	4.7079			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Jaccard

Les PERMANOVA confirment avec tous les indices un effet significatif de la variable Forest_management mais pas de la variable Nitrogen.

Permutation: free
Number of permutations: 9999

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	1.0190	1.01898	5.6430	0.23817	0.0001 ***
Nitrogen	1	0.1912	0.19117	1.0587	0.04468	0.3715
Forest_management:Nitrogen	1	0.1791	0.17907	0.9917	0.04185	0.4236
Residuals	16	2.8892	0.18057		0.67529	
Total	19	4.2784			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Bray

Permutation: free
Number of permutations: 9999

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	0.69752	0.69752	13.5132	0.41357	0.0001 ***
Nitrogen	1	0.09117	0.09117	1.7663	0.05406	0.1514
Forest_management:Nitrogen	1	0.07199	0.07199	1.3948	0.04269	0.2245
Residuals	16	0.82588	0.05162		0.48968	
Total	19	1.68657			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Permutation: free
Number of permutations: 9999

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	0.68667	0.68667	7.0531	0.28644	0.0003 ***
Nitrogen	1	0.08504	0.08504	0.8735	0.03547	0.5329
Forest_management:Nitrogen	1	0.06784	0.06784	0.6968	0.02830	0.7330
Residuals	16	1.55772	0.09736		0.64979	
Total	19	2.39727			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Unifrac

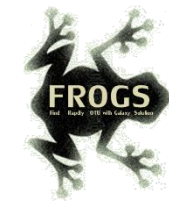
wUnifrac



FROGSSTAT Phyloseq Import Data

Tool Parameters

Input Parameter	Value
Abundance biom file with taxonomical metadata (format: BIOM)	23 : FROGS_5 Taxonomic affiliation: affiliation_abundance.biom   
Metadata associated to samples (format: TSV)	2 : ITS_fast_metadata.tsv   
Taxonomic tree file (format: Newick)	26 : FROGS Tree: tree.nwk   
Names of taxonomic levels	Kingdom Phylum Class Order Family Genus Species
Do you want to normalise your data ?	No, keep abundance as it is.



FROGSSTAT DESeq2 Preprocess

Tool Parameters

Input Parameter	Value
Type of analysis	ASV
Phyloseq object	56 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
Experimental variable	Forest_management
Do you want to correct a confounding factor?	false

FROGSSTAT DESeq2 Visualisation

Tool Parameters

Input Parameter	Value
Type of analysis	ASV
Data object (format: data.RData)	56 : FROGSSTAT Phyloseq Import Data: asv_data.Rdata
DESeq2 object (format: dds.RData)	58 : FROGSSTAT DESeq2 Preprocess: asv_dds.Rdata
Experimental variable	Forest_management
The experimental variable is it quantitative or qualitative?	qual
Condition 1 considered as reference	Control
Condition 2 to be compared to the reference	OMR
Adjusted p-value threshold	0.05

Seule la variable Forest_management est pertinente à analyser puisque Nitrogen n'a pas d'effet significatif.

On pourrait aussi utiliser Treatment et faire des comparaisons entre groupes pertinents, mais les effectifs (en nombre d'échantillons dans chaque groupe) seraient plus petits.

Plusieurs clusters différemment abondants, dans un sens ou dans l'autre.
Et beaucoup de FROGS_combined !!!

Rq: ces FROGS_combined ont le même comportement = peut-être la même espèce = peut-être utile de faire une post-affiliation

Sans un paramétrage adapté ITS lors du FROGS_1_preprocess ces clusters aurait été éliminés pendant le traitement et n'auraient donc pas été identifiés



Pistes stat pour ITS

AVEC POST-AFFILIATION (ET DONC AVEC REGROUPEMENT DE CERTAINS ARTIFICIAL_COMBINED)

HISTORY : ITS_FORMATION_2023_WITH_AFFI_POSTPROCESS



A la suite de FROGS ITSx et FROGS_5 Taxonomic affiliation FROGS_Affiliation_Stat lancement de FROGS Affiliation postprocess

Tool Parameters	
Input Parameter	Value
Sequence file	20 : FROGS ITSx: ITS_sequence.fasta
Abundance file	23 : FROGS_5 Taxonomic affiliation: affiliation_abundance.biom
Is this an amplicon hyper variable in length?	Yes
Using reference database	UNITE_7.1 ITS1
Minimum identity for aggregation	99.0
Minimum coverage for aggregation	99.0



Cluster_40

Cluster_75_FROGS_combined Cluster_121_FROGS_combined Cluster_137_FROGS_combined Cluster_144_FROGS_combined
Cluster_149_FROGS_combined Cluster_150_FROGS_combined Cluster_155_FROGS_combined Cluster_156_FROGS_combined
Cluster_167_FROGS_combined Cluster_168_FROGS_combined Cluster_169_FROGS_combined Cluster_185_FROGS_combined
Cluster_186_FROGS_combined Cluster_187_FROGS_combined Cluster_188_FROGS_combined Cluster_190_FROGS_combined
Cluster_191_FROGS_combined Cluster_192_FROGS_combined Cluster_193_FROGS_combined Cluster_198_FROGS_combined
Cluster_212_FROGS_combined Cluster_236_FROGS_combined Cluster_249_FROGS_combined Cluster_261_FROGS_combined
Cluster_262_FROGS_combined Cluster_263_FROGS_combined Cluster_264_FROGS_combined Cluster_265_FROGS_combined
Cluster_266_FROGS_combined Cluster_267_FROGS_combined Cluster_268_FROGS_combined Cluster_269_FROGS_combined

Cluster_38

Cluster_55

Cluster_43

Cluster_142

Cluster_120

Cluster_143_FROGS_combined

Cluster_163

Cluster_357_FROGS_combined

Cluster_228

Cluster_274

Cluster_298

Cluster_302

Cluster_304

Cluster_224

Cluster_271_FROGS_combined



FROGS Tree

Tool Parameters

Input Parameter	Value
Sequence file	27 : FROGS Affiliation postprocess: sequences.fasta
Biom file	26 : FROGS Affiliation postprocess: affiliation_abundance.biom

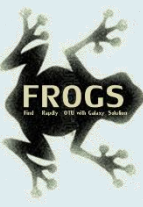
FROGS BIOM to TSV

Tool Parameters	
Input Parameter	Value
Abundance file	26 : FROGS Affiliation postprocess: affiliation_abundance.biom
Sequences file (optional)	27 : FROGS Affiliation postprocess: sequences.fasta
Extract multi-alignments	Yes

FROGSSTAT Phyloseq Import Data

Tool Parameters	
Input Parameter	Value
Abundance biom file with taxonomical metadata (format: BIOM)	26 : FROGS Affiliation postprocess: affiliation_abundance.biom
Metadata associated to samples (format: TSV)	2 : ITS_fast_metadata.tsv 
Taxonomic tree file (format: Newick)	29 : FROGS Tree: tree.nwk
Names of taxonomic levels	Kingdom Phylum Class Order Family Genus Species
Do you want to normalise your data ?	Yes, subsample abundances to the smallest sample size. 

FROGSSTAT Phyloseq Import Data



[Summary](#) [Ranks Names](#) [Sample metadata](#) [Plot tree](#)

```
phyloseq-class experiment-level object
otu_table() OTU Table:      [ 236 taxa and 20 samples ]
sample_data() Sample Data:  [ 20 samples by 8 sample variables ]
tax_table() Taxonomy Table: [ 236 taxa by 14 taxonomic ranks ]
phy_tree()  Phylogenetic Tree: [ 236 tips and 235 internal nodes ]
```

Number of sequences in each sample after normalisation: 4303

[Summary](#) [Ranks Names](#) [Sample metadata](#) [Plot tree](#)

Show

Sample variables: kept, Replicas, Incubation, Nitrogen, Forest_management, Quality, Treatment, SampleID

Show

kept : 79.76, 77.64, 80.26, 78.65, 77.18, 79.68, 78.7, 76.38, 76.37, 77.37, 72.52, 64.98, 78.13, 71.17, 75.2, 73.48, 73.21, 74.01, 74.15, 73.77

Replicas : 3, 2, 5, 1, 4

Incubation : T4

Nitrogen : Nitrogen_supplementation, Control

Forest_management : Control, OMR

Quality : Low degradability

Treatment : Control_with_N, Control, OMR_with_N, OMR

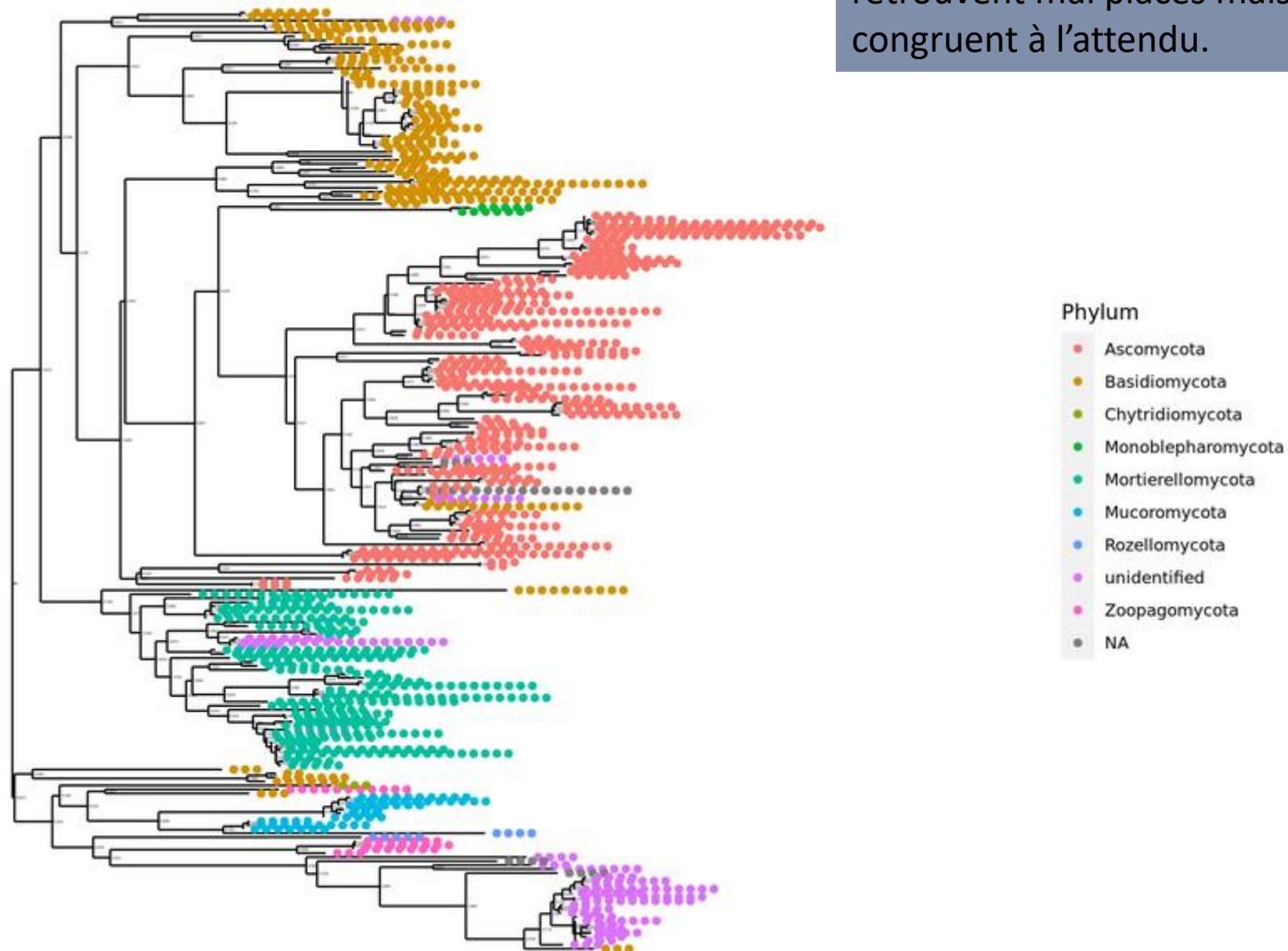
SampleID : Ph203, Ph212, Ph217, Ph222, Ph224, Ph237, Ph241, Ph243, Ph246, Ph250, Ph407, Ph414, Ph415, Ph417, Ph423, Ph428, Ph433, Ph434, Ph439, Ph449

[Summary](#) [Ranks Names](#) [Sample metadata](#) [Plot tree](#)

Show

Rank names : Kingdom, Phylum, Class, Order, Family, Genus, Species, Rank2, Rank3, Rank4, Rank5, Rank6, Rank7, Rank1

Phylogenetic tree colored by Phylum



Attention quelques Basidiomycota qui se retrouvent mal placés mais arbre globalement congruent à l'attendu.

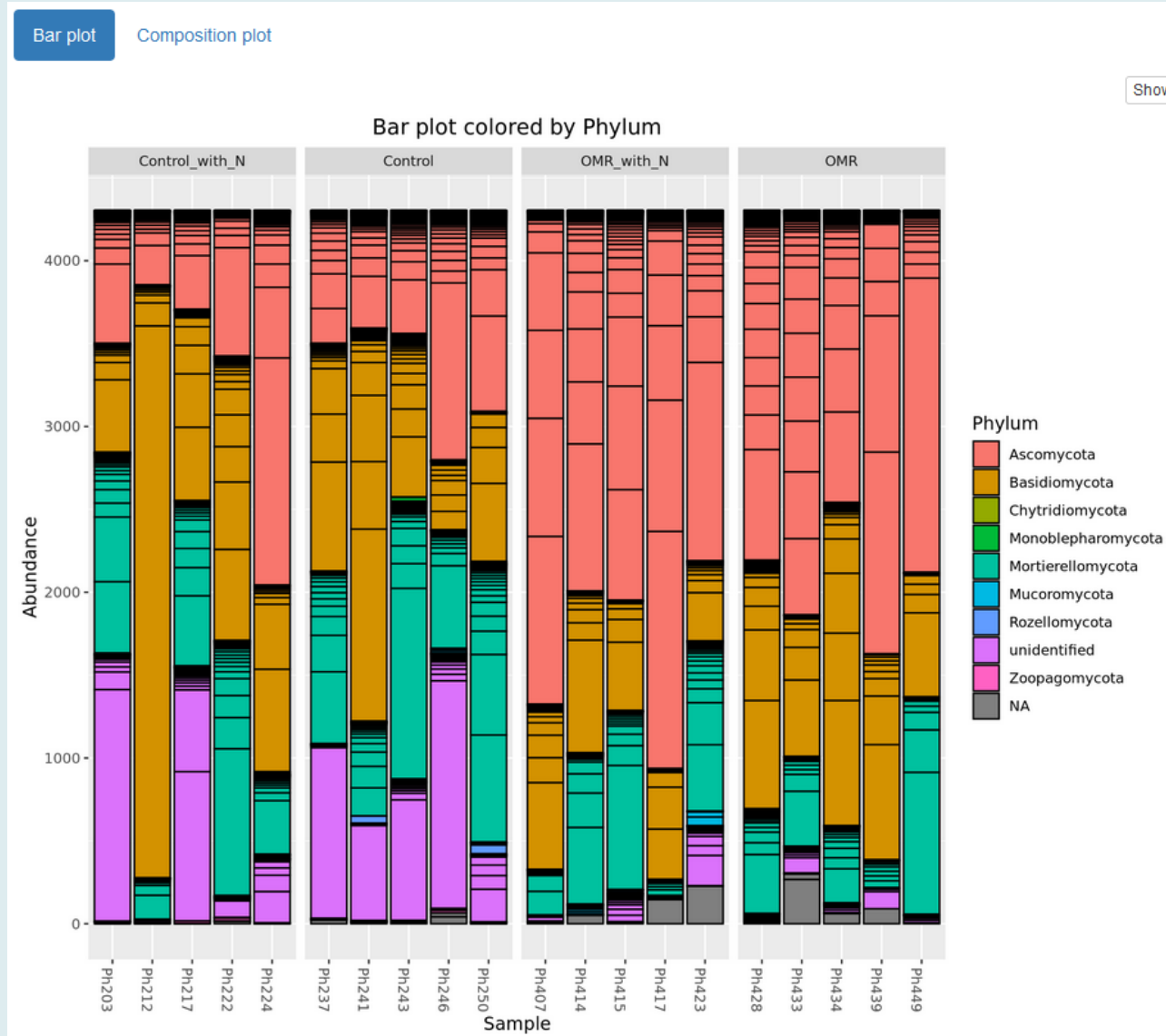
FROGSSTAT Phyloseq Composition Visualisation

Tool Parameters	
Input Parameter	Value
Phyloseq object (format rdata)	33 : FROGSSTAT RAREFIED Phyloseq Import Data: asv_data.Rdata
Grouping variable	Treatment
Taxonomic level to filter your data	Kingdom
Taxa (at the above taxonomic level) to keep in the dataset	Bacteria
Taxonomic level used for aggregation	Phylum
Number of most abundant taxa to keep	9

Interprétation des barplots :

La plupart des outils de visualisation FROGS n'autorisent qu'une seule variable de tri.

Utiliser la variable Treatment (qui combine les deux variables Forest_management et Nitrogen) permet de « tricher » pour afficher deux variables dans les outils de visualisation



Différences évidentes entre Control et OMR
(augmentation des Ascomycota)
Moins évident entre with N et sans

FROGSSTAT Phyloseq Alpha Diversity

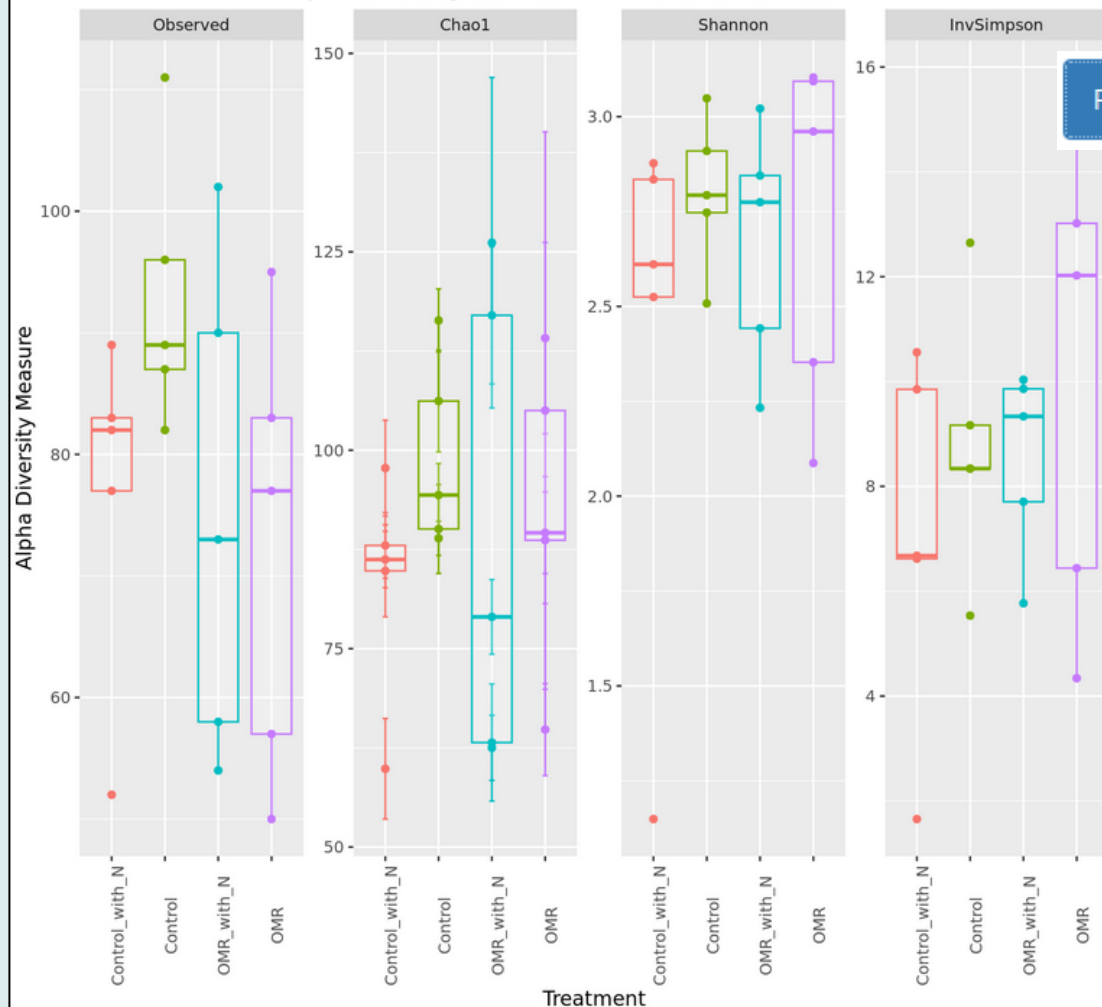
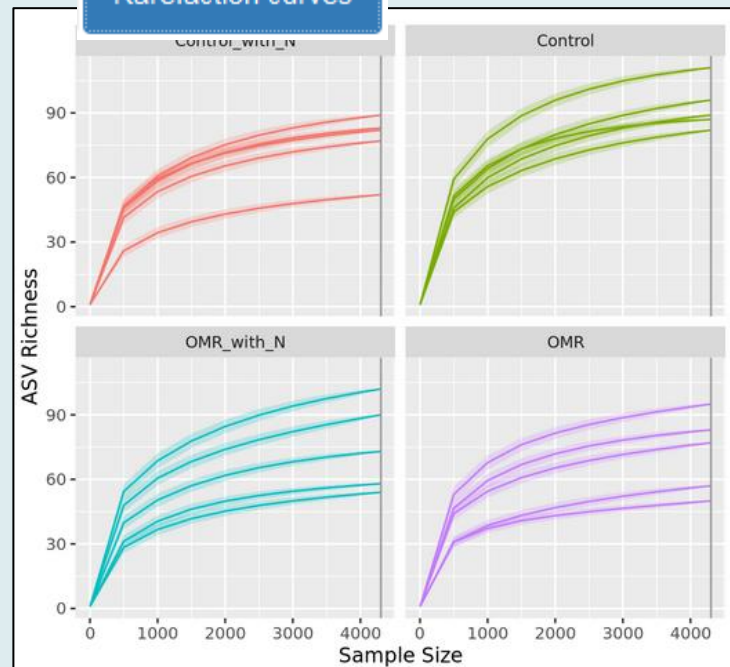
Tool Parameters

Input Parameter	Value
Phyloseq object (format: RData)	33 : FROGSSTAT RAREFIED Phyloseq Import Data: asv_data.Rdata
Experiment variable	Treatment
The alpha diversity indices to compute	Observed Chao1 Shannon InvSimpson



Alpha diversity distribution in function of Treatment

Rarefaction curves



Richness plot with boxplot

Alpha Diversity Index Anova Analysis

Pas d'effet significatif du traitement sur aucun indice.
Profondeur de séquençage très faible (sous-échantillon pour le TP et effets des filtres à 50%)

```
#####
#Perform ANOVA on Observed, which effects are significant
anova.Observed <-aov( Observed ~ Depth + Treatment, anova_data)
summary(anova.Observed)
      Df Sum Sq Mean Sq F value Pr(>F)
Treatment  3  1289   429.6    1.56  0.238
Residuals 16   4406   275.4

#####
#Perform ANOVA on Chao1, which effects are significant
anova.Chao1 <-aov( Chao1 ~ Depth + Treatment, anova_data)
summary(anova.Chao1)
      Df Sum Sq Mean Sq F value Pr(>F)
Treatment  3    649   216.4    0.542  0.661
Residuals 16   6393   399.6

#####
#Perform ANOVA on Shannon, which effects are significant
anova.Shannon <-aov( Shannon ~ Depth + Treatment, anova_data)
summary(anova.Shannon)
      Df Sum Sq Mean Sq F value Pr(>F)
Treatment  3   0.453   0.1510    0.692  0.57
Residuals 16   3.490   0.2182

#####
#Perform ANOVA on InvSimpson, which effects are significant
anova.InvSimpson <-aov( InvSimpson ~ Depth + Treatment, anova_data)
summary(anova.InvSimpson)
      Df Sum Sq Mean Sq F value Pr(>F)
Treatment  3   26.62    8.874    0.791  0.517
Residuals 16  179.56   11.222
```

FROGSSTAT Phyloseq Beta Diversity

Tool Parameters	
Input Parameter	Value
Phyloseq object (format: RData)	33 : FROGSSTAT RAREFIED Phyloseq Import Data: asv_data.Rdata
Grouping variable	Treatment
The methods of beta diversity	Unifrac Weighted Unifrac Bray-Curtis Jaccard (as cc method in betadiver vegan fonction)
Other method	Not available.

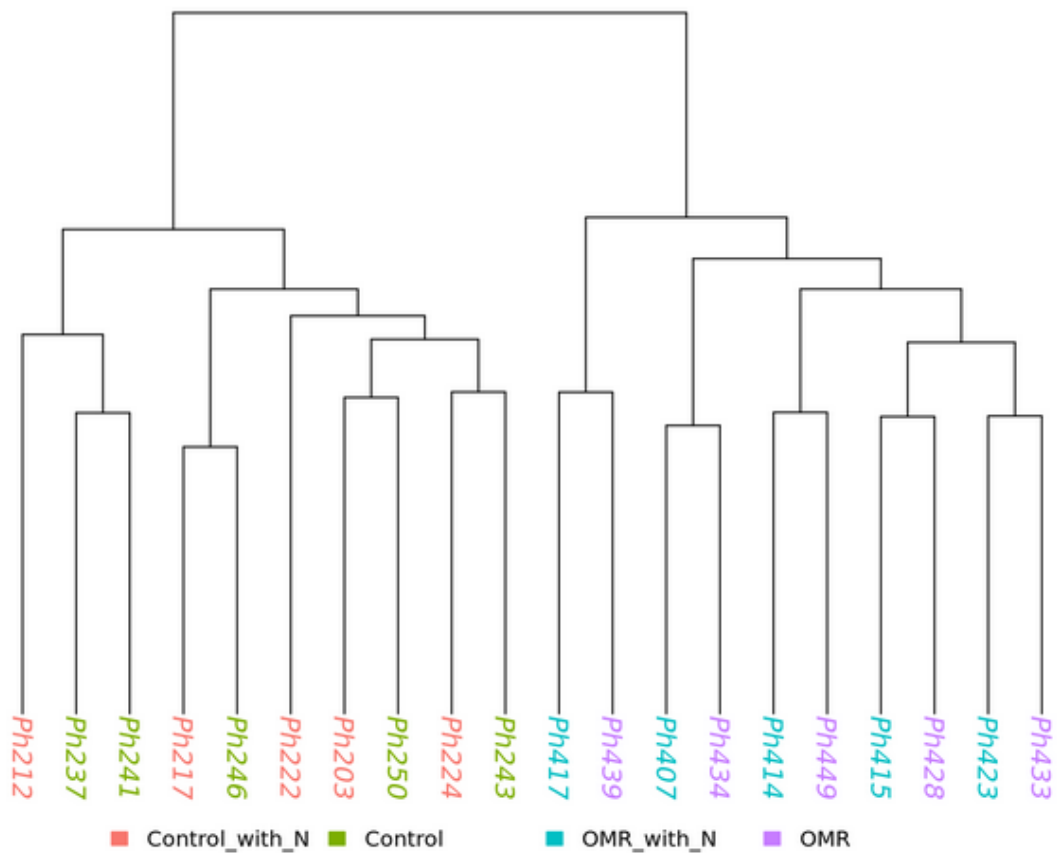
Cet outil sert à produire des matrices pour ensuite faire la clusterisation et la visualisation

FROGSSTAT Phyloseq Sample Clustering

Tool Parameters	
Input Parameter	Value
Phyloseq object (format: RData)	33 : FROGSSTAT RAREFIED Phyloseq Import Data: asv_data.Rdata   
The beta diversity distance matrix file	44 : FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (cc.tsv)   
Experiment variable	Treatment

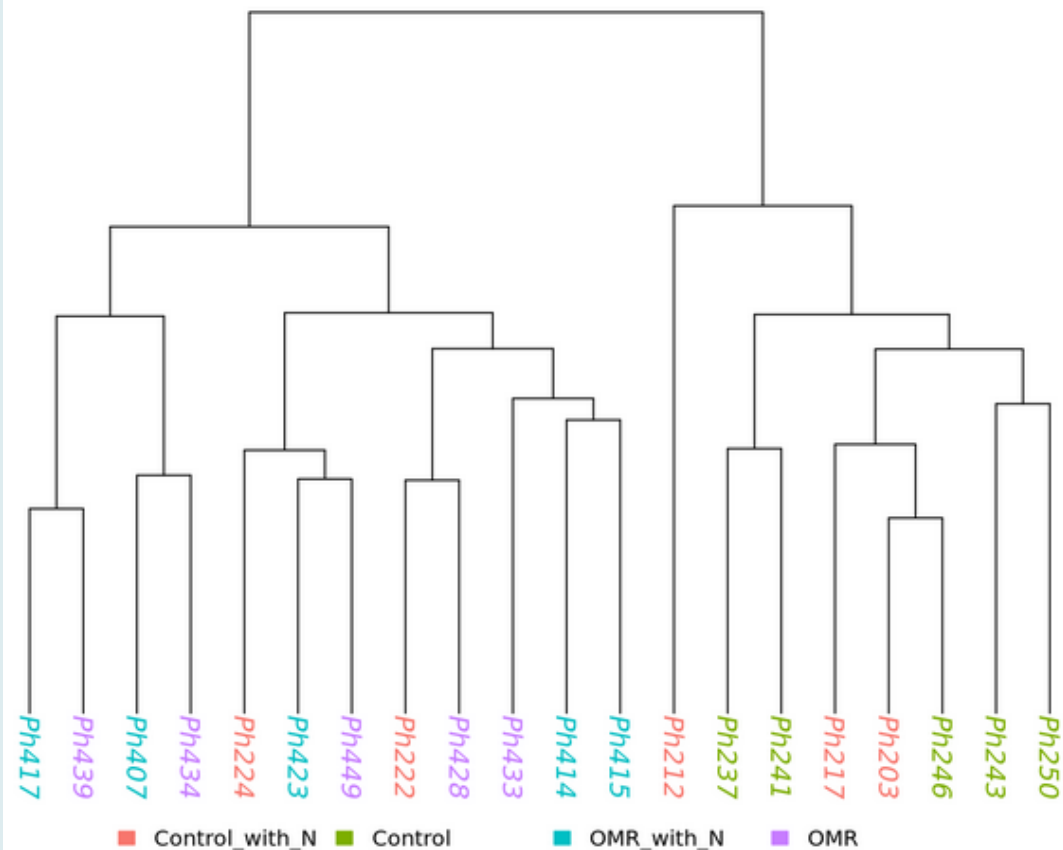
Pour chacune des tables: cc, bray, unirac et wunifrac

ward.D2 linkage clustering tree



Jaccard

ward.D2 linkage clustering tree



Bray



Unifrac



wUnifrac

Différences entre Control et OMR mais pas avec l'ajout d'azote.

Meilleur clustering avec les méthodes qualitatives qu'avec les méthodes quantitatives :
les différences sont plutôt dues à des ASV peu abondants

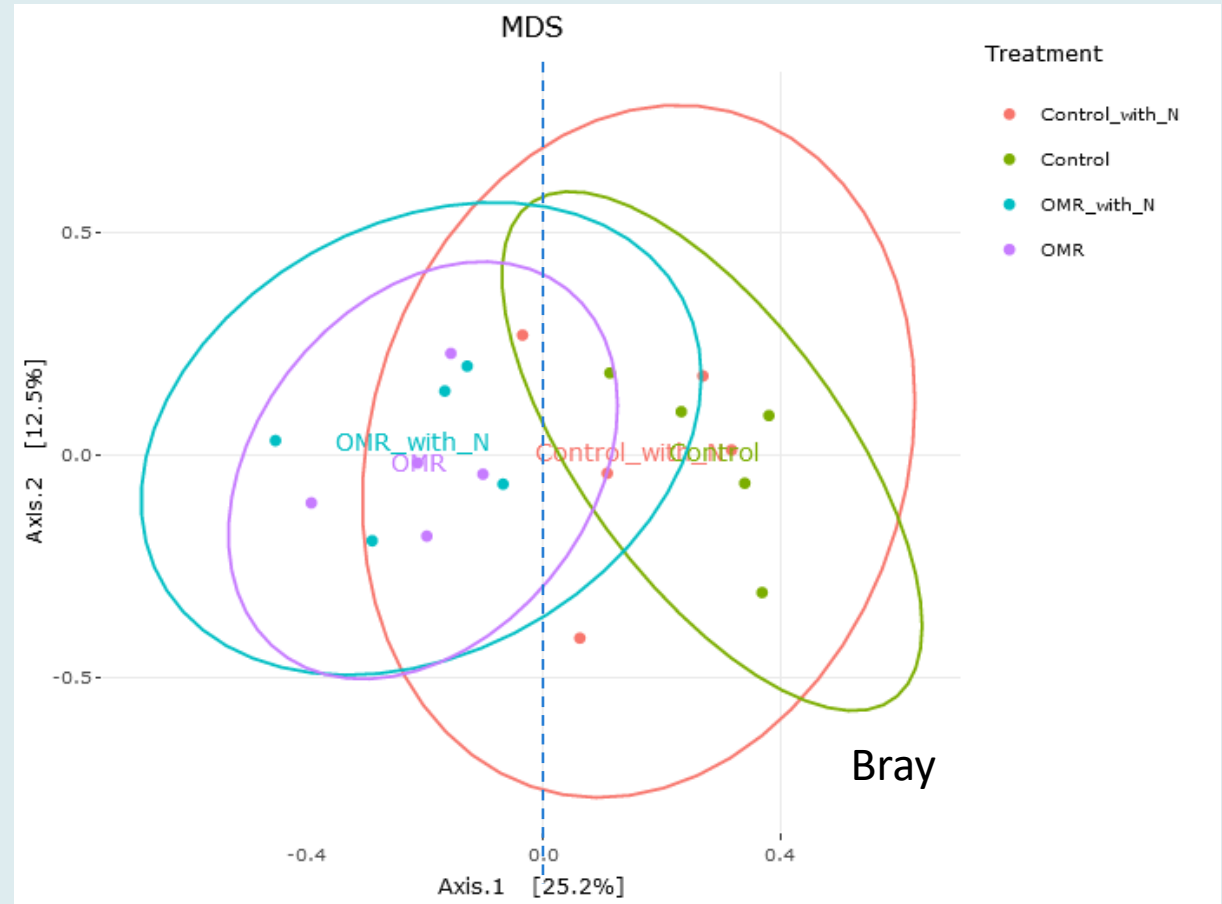
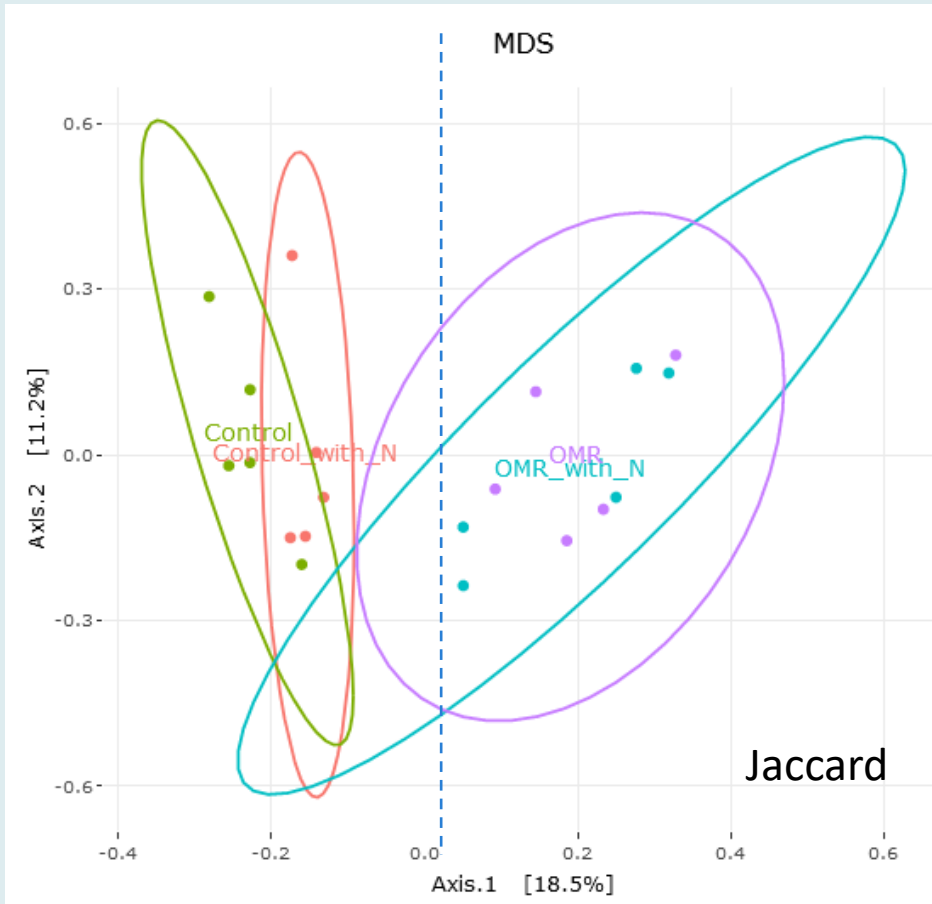
(quand on prend en compte les abondances les échantillons se ressemblent davantage)

FROGSSTAT Phyloseq Structure Visualisation

Tool Parameters	
Input Parameter	Value
Phyloseq object (format rdata)	33 : FROGSSTAT RAREFIED Phyloseq Import Data: asv_data.Rdata
The beta diversity distance matrix file	44 : FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (cc.tsv)
Experiment variable	Treatment
Ordination method	MDS/PCoA

Pour chacune des tables: cc, bray, unirac et wunifrac

Diversité bêta – ordination :



Différences entre Control et OMR mais pas avec l'ajout d'azote.

Meilleure séparation (sur l'axe1) avec les méthodes qualitatives qu'avec les méthodes quantitatives : les différences sont plutôt dues à des ASV peu abondants (quand on prend en compte les abondances les échantillons se ressemblent davantage)

FROGSSTAT Phyloseq Multivariate Analysis Of Variance

Tool Parameters	
Input Parameter	Value
Phyloseq object (format: RData)	33 : FROGSSTAT RAREFIED Phyloseq Import Data: asv_data.Rdata
The beta diversity distance matrix file	44 : FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (cc.tsv)
Experiment variable	Treatment

Diversité bêta – PERMANOVA :

Il est ici possible de donner plusieurs variables séparées par des + ou * (effets additifs ou avec interaction), et donc plus pertinent d'utiliser Nitrogen et Forest_management plutôt que Treatment

```
adonis(formula = dist ~ Forest_management * Nitrogen, data = metadata,  
        permutations = 9999)
```

Jaccard

```
Permutation: free  
Number of permutations: 9999
```

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	0.7867	0.78674	3.6153	0.16713	0.0001 ***
Nitrogen	1	0.2368	0.23682	1.0883	0.05031	0.3039
Forest_management:Nitrogen	1	0.2020	0.20204	0.9284	0.04292	0.5598
Residuals	16	3.4818	0.21761		0.73964	
Total	19	4.7074			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Permutation: free  
Number of permutations: 9999
```

Bray

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	0.9895	0.98953	4.3792	0.19820	0.0002 ***
Nitrogen	1	0.1638	0.16379	0.7248	0.03281	0.7741
Forest_management:Nitrogen	1	0.2240	0.22399	0.9913	0.04486	0.4284
Residuals	16	3.6154	0.22596		0.72413	
Total	19	4.9927			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Permutation: free  
Number of permutations: 9999
```

Unifrac

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	0.62321	0.62321	6.3868	0.26046	0.0001 ***
Nitrogen	1	0.10364	0.10364	1.0621	0.04331	0.3450
Forest_management:Nitrogen	1	0.10463	0.10463	1.0722	0.04373	0.3393
Residuals	16	1.56125	0.09758		0.65250	
Total	19	2.39272			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Les PERMANOVA confirment avec tous les indices un effet significatif de la variable Forest_management mais pas de la variable Nitrogen.

```
Permutation: free  
Number of permutations: 9999
```

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Forest_management	1	0.67636	0.67636	9.1232	0.33917	0.0001 ***
Nitrogen	1	0.06781	0.06781	0.9147	0.03400	0.4357
Forest_management:Nitrogen	1	0.06382	0.06382	0.8608	0.03200	0.4707
Residuals	16	1.18617	0.07414		0.59482	
Total	19	1.99416			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

wUnifrac

FROGSSTAT Phyloseq Import Data

Tool Parameters	
Input Parameter	Value
Abundance biom file with taxonomical metadata (format: BIOM)	26 : FROGS Affiliation postprocess: affiliation_abundance.biom
Metadata associated to samples (format: TSV)	2 : ITS_fast_metadata.tsv
Taxonomic tree file (format: Newick)	29 : FROGS Tree: tree.nwk
Names of taxonomic levels	Kingdom Phylum Class Order Family Genus Species
Do you want to normalise your data ?	No, keep abundance as it is.



FROGSSTAT DESeq2 Preprocess

Tool Parameters	
Input Parameter	Value
Type of analysis	ASV
Phyloseq object	35 : FROGSSTAT NON RAREFIED Phyloseq Import Data : asv_data.Rdata
Experimental variable	Forest_management
Do you want to correct a confounding factor?	false

FROGSSTAT DESeq2 Visualisation

Tool Parameters	
Input Parameter	Value
Type of analysis	ASV
Data object (format: data.RData)	35 : FROGSSTAT NON RAREFIED Phyloseq Import Data : asv_data.Rdata
DESeq2 object (format: dds.RData)	41 : FROGSSTAT DESeq2 Preprocess: asv_dds.Rdata
Experimental variable	Forest_management
The experimental variable is it quantitative or qualitative?	
Condition 1 considered as reference	Control
Condition 2 to be compared to the reference	OMR
Adjusted p-value threshold	0.05

Seule la variable Forest_management est pertinente à analyser puisque Nitrogen n'a pas d'effet significatif.

On pourrait aussi utiliser Treatment et faire des comparaisons entre groupes pertinents, mais les effectifs (en nombre d'échantillons dans chaque groupe) seraient plus petits.

Plusieurs clusters différenciellement abondants, dans un sens ou dans l'autre.
Et 2 ASV FROGS_combined !!!

Sans un paramétrage adapté ITS lors du FROGS_1_preprocess ces clusters aurait été éliminés pendant le traitement et n'auraient donc pas été identifiés

