SECOND GENOME
THE MICROBIOME COMPANY

MICROBIAL PROFILING REPORT

G3 PHYLOCHIP™ ASSAY

EPAN12_0119

2013-08-22

Second Genome, Inc. 1150 Bayhill Drive, Suite 215 San Bruno, CA 94066 650.440.4606

SECOND GENOME
THE MICROBIOME COMPANY

METHODS

PROJECT: MICHA – MICROBIOME STRUCTURES OF TWO MODULES FROM THE MARS 500 COMPLEX

Principal Investigator: Petra Schwendner, Dipl.-Biol. Univ., German Aerospace Center

EPAN: 12-0119

PROJECT DESCRIPTION

This study aims to examine the microbial community in an enclosed environment that mimics the living arrangement in a space capsule over time. This study also aims to track specific pathogens. Swabs samples from 2 Modules are collected at 7 timepoints. Module 250 samples are a combination of four surfaces, and Module 150 five surfaces. Swabs are combined directly in extraction buffer.

MAIN FINDINGS

SUMMARY

- ✓ **The two modules 150 and 250 exhibit significantly different microbiome structures.**
- ✓ **Module 150 shows a greater intragroup dissimilarity among samples than module 250.**
- ✓ **No significant influence on the whole microbiome structure is identified for factor Time.**
- ✓ **Significant correlations between abundances and Time are found for 56 OTUs of module 150 and 37 OTUs for module 250.**
- ✓ **81 Potential Pathogenic OTUs (PPO) are identified, among them *Enterococcus faecalis*, which also shows a significant correlation with factor Time in module 250.**

METHODS

Sample Manifest

MANIFEST

Short_name	Barcode	Module	Timepoint	Isolation_day	Cfu_surface	Cfu_air	Comp1	Comp2	gDNA_ship	PCR_yield	PCR_hyb
EU-250-06/10	PT12-4511	250	time1	14	443	27	NA	1early	6.75	253	253
EU-250-07/10	PT12-4512	250	time2	44	4748	91	NA	1early	9.00	457	457
EU-250-11/10	PT12-4513	250	time3	169	1660	48	NA	2mid	20.95	420	420
EU-250-03/11	PT12-4514	250	time4	286	528	313	NA	2mid	27.25	530	500
EU-250-07/11	PT12-4515	250	time5	406	5493	66	NA	2mid	5.60	215	215
EU-250-10/11	PT12-4516	250	time6	495	2183	78	NA	3late	3.17	222	222
EU-250-04/12	PT12-4517	250	time7	700	140	62	NA	3late	18.30	213	213
EU-150-06/10	PT12-4518	150	time1	14	3983	151	1early	NA	22.45	118	118
EU-150-07/10	PT12-4519	150	time2	44	8378	398	1early	NA	31.00	322	322
EU-150-11/10	PT12-4520	150	time3	169	4813	279	2mid	NA	15.35	370	370
EU-150-03/11	PT12-4521	150	time4	286	15638	168	2mid	NA	20.25	95	95
EU-150-07/11	PT12-4522	150	time5	406	65190	290	2mid	NA	12.70	415	415
EU-150-10/11	PT12-4523	150	time6	495	6080	83	3late	NA	15.45	410	410
EU-150-04/12	PT12-4524	150	time7	700	8330	54	3late	NA	25.55	602	500
IsoCntl	PT12-4525	NA	NA	NA	NA	NA	NA	NA	0.00	0	0

Notes regarding metadata abbreviations:

1. *The gDNA_ship refers to the genomic DNA mass measured in ng as reported by the client.*
2. *Genomic DNA content measured in ng upon arrival at our laboratories was below detection limit for all samples. This was measured using the PicoGreen® method.*
3. *PCR_yield refers to the amplified DNA mass measured in ng after PCR and amplicon cleanup. For each sample, 500 ng of amplified, labeled product was hybridizied.*
4. *PCR_hyb refers to the amount of amplified DNA loaded onto the microarray. Ideally, this is 500 ng. Due to low amplification rates for some samples, less DNA was loaded onto chips. In order to ensure a biological comparability of the generated results, fluorescence intensities were therefore normalized by rank across probes for each array individually.*

METHODS

Pre-Analysis Methods

15 frozen DNA isolates from swab content (DNA concentration below detection limit) were received in Second Genome's service laboratory on March 7, 2013, and stored at -20°C.

The bacterial 16S rRNA genes were amplified using the degenerate forward primer:
27F.1 5'-AGRGTTTGATCMTGGCTCAG-3'
and the non degenerate reverse primer:
1492R.jgi 5'-GGTTACCTTGTTACGACTT-3'

15 samples amplified to specification in PCR and were moved forward for hybridization.

For each sample, amplified products are concentrated using a solid-phase reversible immobilization method for the purification of PCR products and quantified by electrophoresis using an Agilent 2100 Bioanalyzer®. PhyloChip Control Mix™ is added to each amplified product. Thirty-five cycles of bacterial 16S rRNA gene PCR amplification was performed.

METHODS

Laboratory Analysis Methods

Labeled bacterial products were fragmented, biotin labeled, and hybridized to the PhyloChip™ Array, version G3. PhyloChip arrays were washed, stained, and scanned using a GeneArray® scanner (Affymetrix). Each scan is captured using standard Affymetrix software (GeneChip® Microarray Analysis Suite). Hybridization values, the fluorescence intensity, for each taxon were calculated as a trimmed average, with maximum and minimum values removed before averaging.

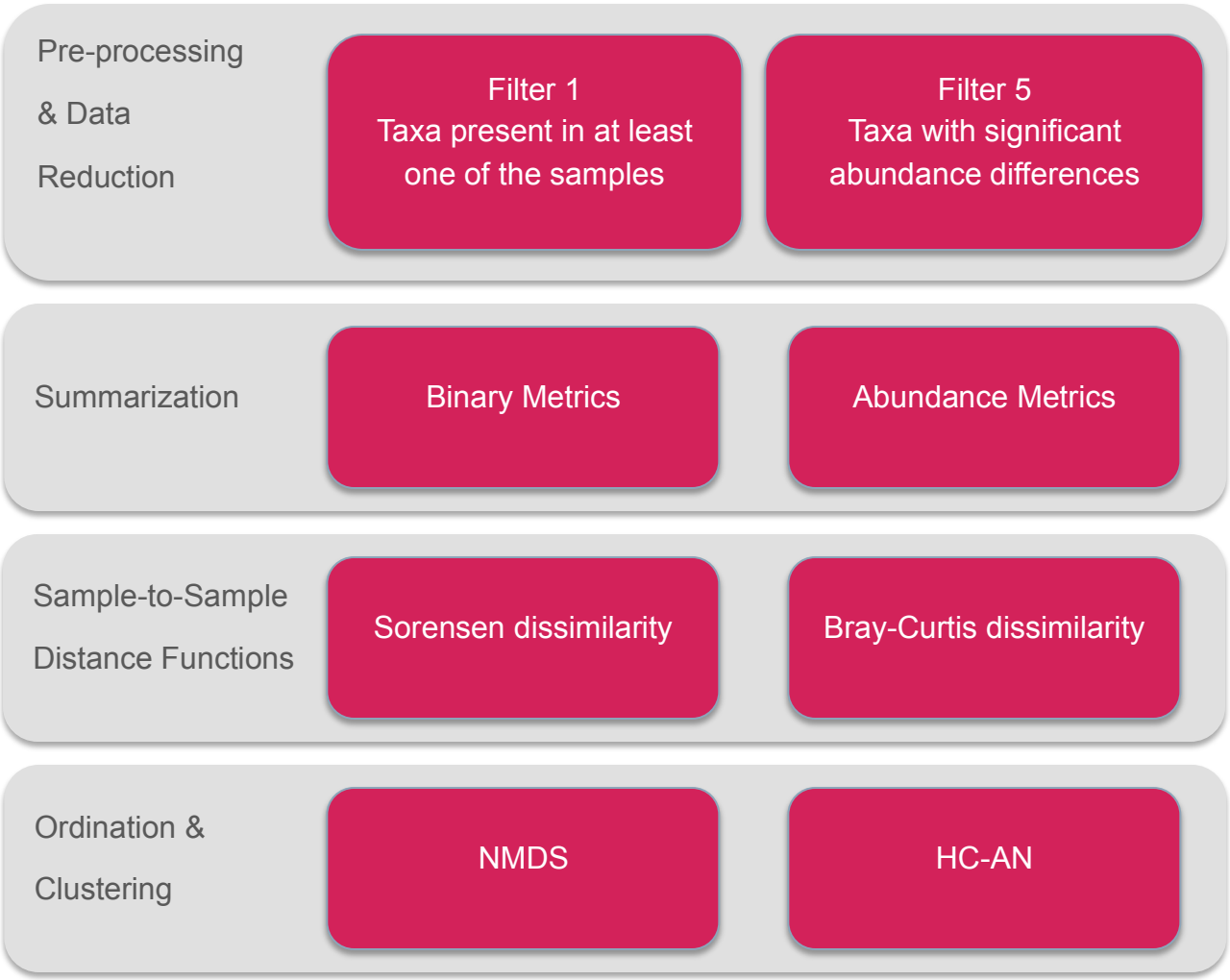
Samples are processed in a Good Laboratory Practices (GLP) compliant service laboratory running Quality Management Systems for sample and data tracking. The laboratory implements detailed SOPs, equipment and process validation, training, audits and document control measures. QC and QA metrics are maintained for all sample handling, processing and storage procedures.

METHODS

Data Analysis Methods

OVERVIEW

The full data analysis pipeline for Second Genome's Microbial Profiling Service incorporates several separate stages: pre-processing and data reduction, summarization, normalization, sample-to-sample distance metrics, ordination/clustering, sample classification, and significance testing. Second Genome's PhyCA-Stats™ analysis software package was used for multivariate statistical analysis of client data.



Data analysis steps incorporated into Second Genome's PhyCA-Stats™ analysis software

METHODS

OTU SELECTION

To calculate the summary fluorescence intensity (FI) for each feature on each array, the central 9 pixels of individual image features were ranked by intensity and the 75% percentile was used. Probe FIs were background-subtracted and scaled to the PhyloChip™ Control Mix. Array FI is collected as integer values ranging from 0 to 65,536 (2^{16}). Fluorescence intensity observed from PM ("perfectly matching") probes were compared to MM ("mis matching") probes and were considered positive if PM/MM >= 1.5 and PM-MM >= 50**N* and *r* >= 0.95 where *N* indicates the array specific noise (DeSantis, 2005), and *r* represents the response score (Hazen, 2010). Only PM FI from probes observed as positive in at least 3 experiments were exported from all experiments then rank normalized and used as input to empirical probe-set discovery. Probes were clustered into probe-sets based on both correlations in FI across all biological samples and taxonomic relatedness. Where multiple clustering solutions were available, higher correlation coefficients were favored over lower, taxonomic relatedness at the species level was favored over higher ranks, and sets composed of more probes were favored over less. All probe sets contained >= 5 probes, with average pair-wise correlation coefficients >= 0.85. The empirical OTU (eOTU) tracked by a probe set was taxonomically annotated from the combination of the 9-mers contained in all probes of the set. Hybridization score (HybScore) are the mean of the ranked probe FIs within each set and are used in abundance-based analysis. eOTUs were considered present if >=80% of their probes were positive.

OTU FILTERS

Taxa are filtered to those present in at least one of the samples (**Filter-1**), or to taxa significantly increased in their abundance in one category compared to the alternate categories (**Filter-5**). For Filter-5, the parametric Welch test was employed to calculate p-values. Additionally, q-values were calculated using the Benjamini-Hochberg procedure to correct p-values, controlling for false discovery rates.

SUMMARIZATION

After the taxa are identified for inclusion in the analysis, the values used for each taxa-sample intersection are populated in two distinct ways. In the first case, the **Abundance metrics** are used directly (**AT**). Note that abundance values (HybScores) for eOTUs that did not achieve >=80% of their probes as positive are not discarded. **Binary metrics,** also referred to as incidence scores, are created where 1's represent presence, 0's indicate absence (**BT**).

SAMPLE-TO-SAMPLE DISTANCE FUNCTIONS

All profiles are inter-compared in a pair-wise fashion to determine a dissimilarity score and store it in a distance dissimilarity matrix. The distance functions are chosen to allow similar biological samples to produce only small dissimilarity scores. Bray-Curtis Distance is a statistic used to quantify the compositional dissimilarity between two different communities. The Bray–Curtis dissimilarity considers differences in the abundance of a species or OTU across two communities. Please note, when the Bray-Curtis index is calculated from incidence values, it equals the Sorensen dissimilarity.

METHODS

ORDINATION, CLUSTERING, AND CLASSIFICATION METHODS

Two-dimensional ordinations and hierarchical clustering maps of the samples in the form of dendrograms were created to graphically summarize the inter-sample relationships. To create dendrograms, the samples from the distance matrix are clustered hierarchically using the **average-neighbor (HC-AN)** method. Non-metric Multidimensional Scaling (NMDS) is a method of two-dimensional ordination plotting that is used to visualize complex relationships between samples. NMDS uses the dissimilarity values to position the points relative to each other in two dimensions.

WHOLE MICROBIOME SIGNIFICANCE TESTING

The **Adonis test** is utilized for finding significant differences among discrete categorical or continuous variables. In this randomization/Monte Carlo permutation test, the samples are randomly reassigned to the various sample categories, and the between-category differences are compared to the true between-category differences. Adonis utilizes the sample-to-sample distance matrix directly, not a derived ordination or clustering outcome.

METHODS

Custom Data Analysis Methods

CORRELATION OF OTU TRAJECTORIES WITH METADATA

Abundances of individual OTUs per sample were correlated with metadata factors using Spearman rank correlation. OTUs with significant p-values (<0.05) were selected and displayed in a heatmap, where either OTU trajectories are clustered based on a Euclidean distance measure or OTUs were ranked by their significance.

RECLASSIFICATION OF 16S RRNA GENE SEQUENCES FROM ISOLATES

16S rRNA gene sequences provided by the client were first quality checked and taxonomically annotated in similar fashion as eOTUs in order to ensure comparability. Quality filtering was performed by a) SINA aligning the sequences against a taxonomic database and removing bases that were not aligned, b) manual trimming of homopolymers at the beginning and end of the sequences, and c) by setting a sequence length cutoff of 700 bps (53 sequences of 959 were removed). For each taxonomic level, confidence scores were calculated. For reliable taxonomic classification confidence score of 0.8 was used as a cutoff, sequences not classified at kingdom level were removed (6 out of 908). 902 sequences passed the quality filtering.

COMPARISON OF PHYLOCHIP DERVIED TAXA WITH ISOLATES

Taxonomic identifications of classified OTUs were compared against the taxonomic identification retrieved from isolate analysis at all taxonomic levels, when classification was retrieved.

IDENTIFICATION OF POTENTIAL PATHOGENIC OTUS (PPO)

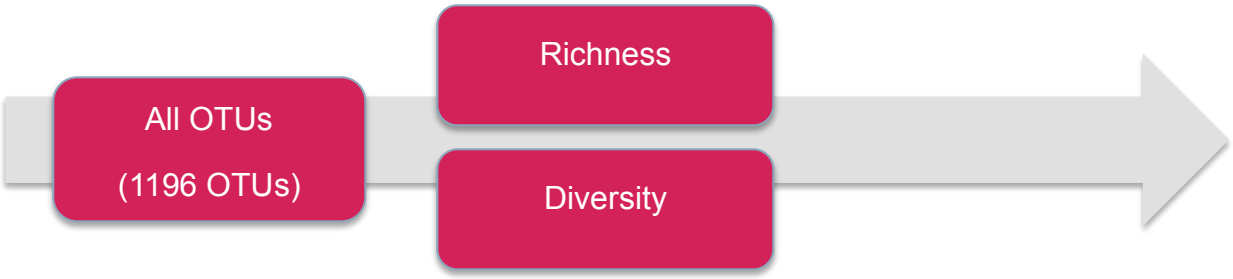
Species classifications of each eOTU was individually compared to a catalogue of pathogens provided by the client. So identified taxa were then analyzed separately concerning the observed microbiome structure, their aggregated hybscores, and used for correlation analysis with time. These methods were identical to those explained above.

RESULTS

Chapter 1: Community Characterization

In this chapter, we address the richness, diversity, and taxonomic composition of each sample. Explicit comparisons between samples are presented in later chapters.

Richness and diversity analysis indicates that the extraction control is an outlier and is excluded from the ordination and HCAN analysis.



SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 1-1	*./Community_Characterization/genus.richness.ps (+pdf)*
Figure 1-2	*./Community_Characterization/family.barchart.ps (+pdf)*
	Additional richness and diversity figures can be found for every taxonomic level (i.e. phylum, class, order, etc.) in the Community_Characterization folder
Figure 1-3	*./Community_Characterization /bt1.bray.NMDS.Module.pdf*
Figure 1-4	*./Community_Characterization /at1.bray.NMDS.Module.pdf*

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./Community_Characterization/genus.bt1.bacteria.richness.table.txt	*Taxon richness at the genus level.*
./Community_Characterization/genus.bt1.archaea.richness.table.txt	*Taxon richness at the genus level.*
./Community_Characterization/family.barchart.table.txt	*Proportions of eOTUs classified at the family rank for top 9 families.*

RESULTS

FIGURE 1-1

(genus.richness.pdf) Archaea and bacteria taxon richness at the genus level, using the bt1 table.

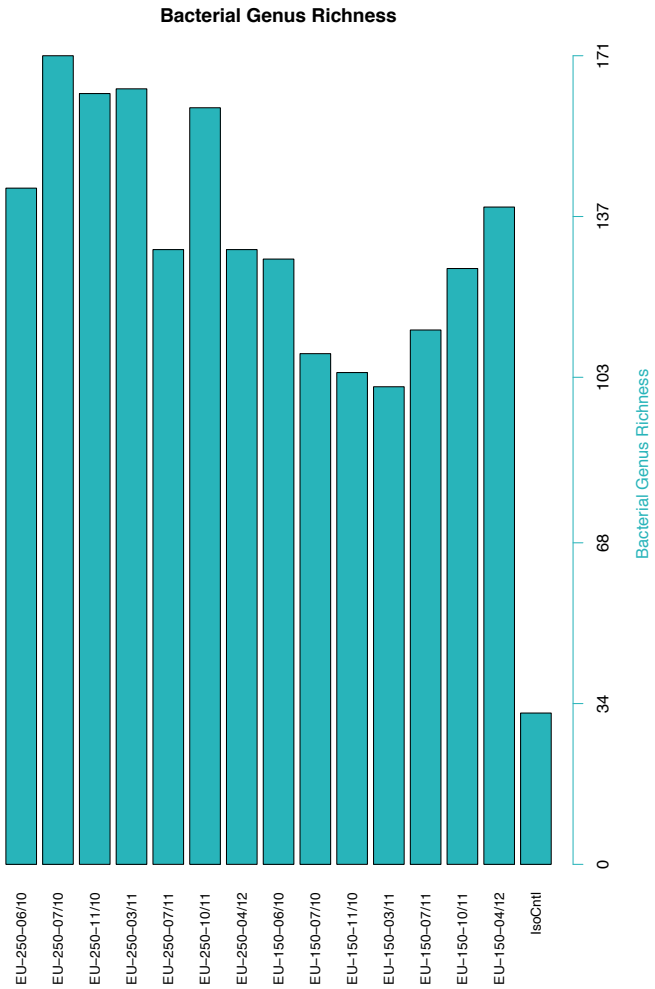


FIGURE 1-1

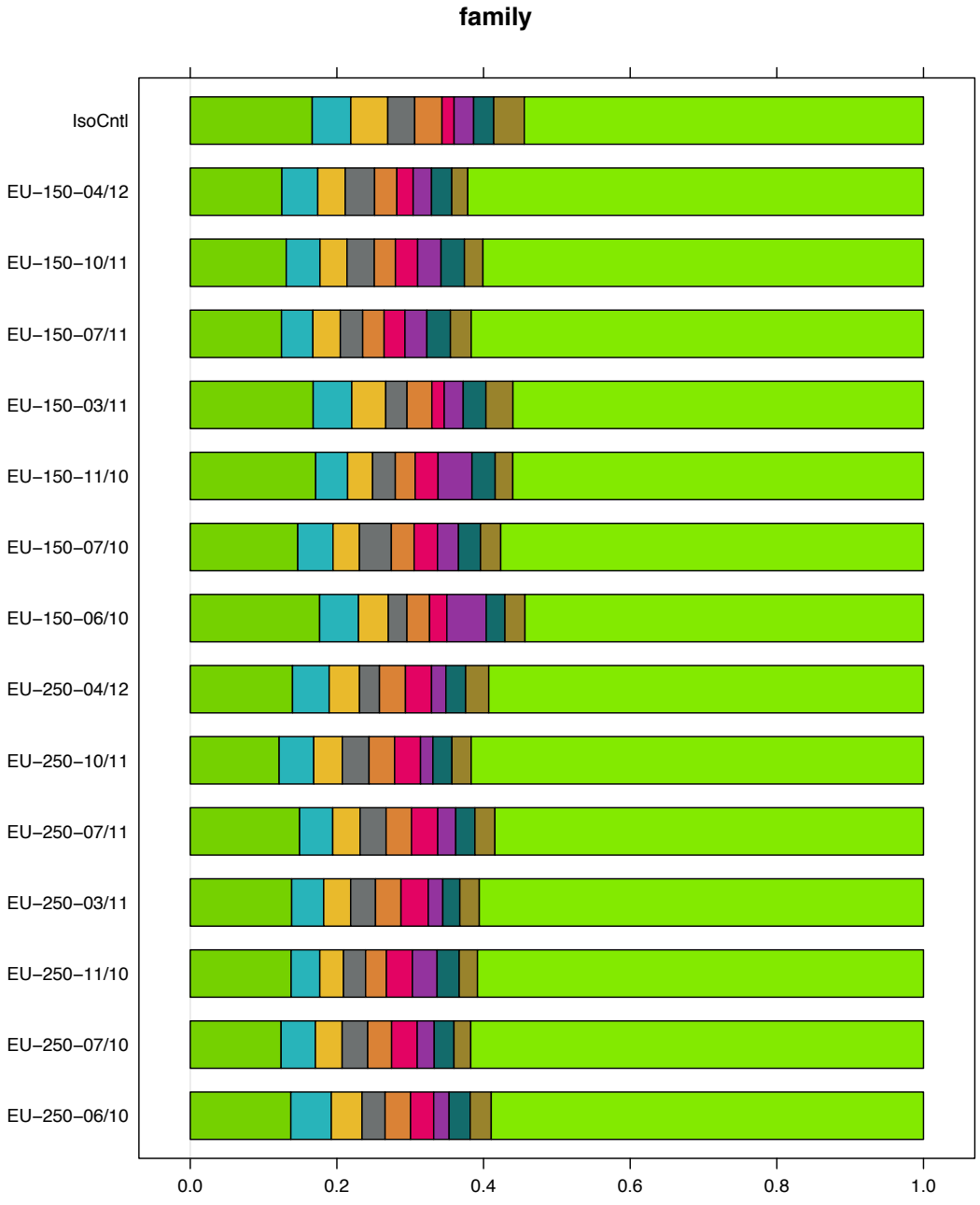
OBSERVATIONS

1. Bacterial genus richness ranges from 32 to 171, whereas none of the samples show presence of archaea.
2. The richness in IsoCntl is the least of all samples.
3. A statistically significant difference (non-paired, heteroscedastic student's t-test) in bacterial richness is observed between Module 250 and Module 150 samples ($p < 0.05$).

RESULTS

FIGURE 1-2

(family.abd.barchart.pdf) Comparison of family-level proportional abundance across samples. The bar chart displays the 9 families with the largest HybScores found by summing the HybScores from the OTUs within the families.



RESULTS

	Domain	Phylum	Class	Order	Family
	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae
	Bacteria	Proteobacteria	unclassified	unclassified	unclassified
	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
	Bacteria	Proteobacteria	Gammaproteobacteria	unclassified	unclassified
	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae
	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae
	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	RikenellaceaeII
	others				

FIGURE 1-2

OBSERVATIONS

1. The top 9 families represent on average 61.6% of proportional abundance of each sample.
2. Family-level richness patterns are non-uniform across the samples. OTUs within the *Lachnospiraceae* and *Comamonadaceae* families comprise the largest proportion of the overall HybScores.

FIGURE 1-3

(bt1.bray.NMDS.Module.pdf) NMDS based on Bray-Curtis distance between samples given presence/absence of 1196 taxa present in at least one sample. Stress=0.1062.

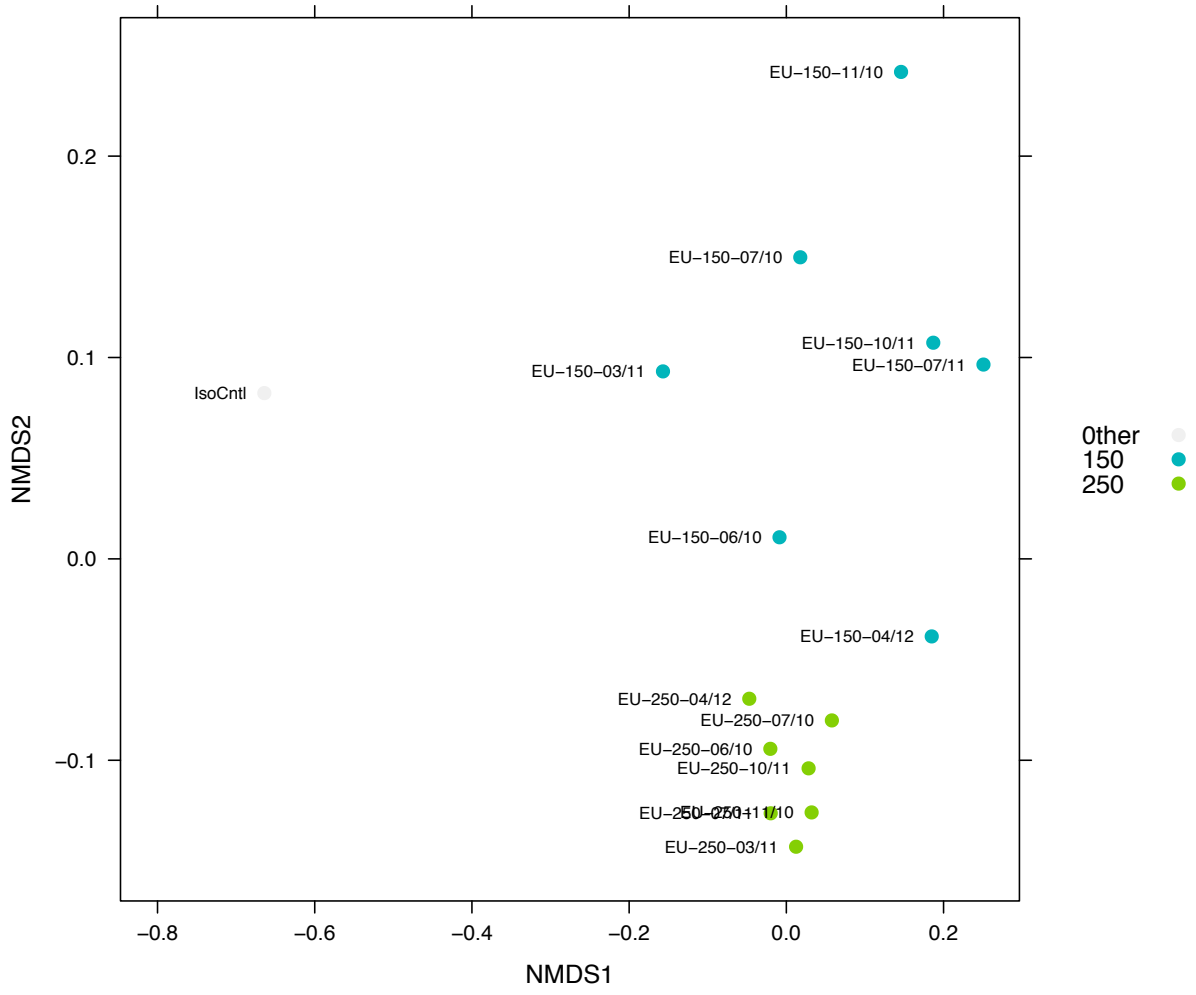


FIGURE 1-3

OBSERVATIONS

1. Ordination analysis based on incidence values of 1196 taxa present in at least one of the samples suggests the sample Isolation Control is a potential outlier sample.

FIGURE 1-4

(at1.bray.NMDS.Module.pdf) NMDS based on Bray-Curtis distance between samples given abundance of 1196 taxa present in at least one sample. Stress=0.1298.

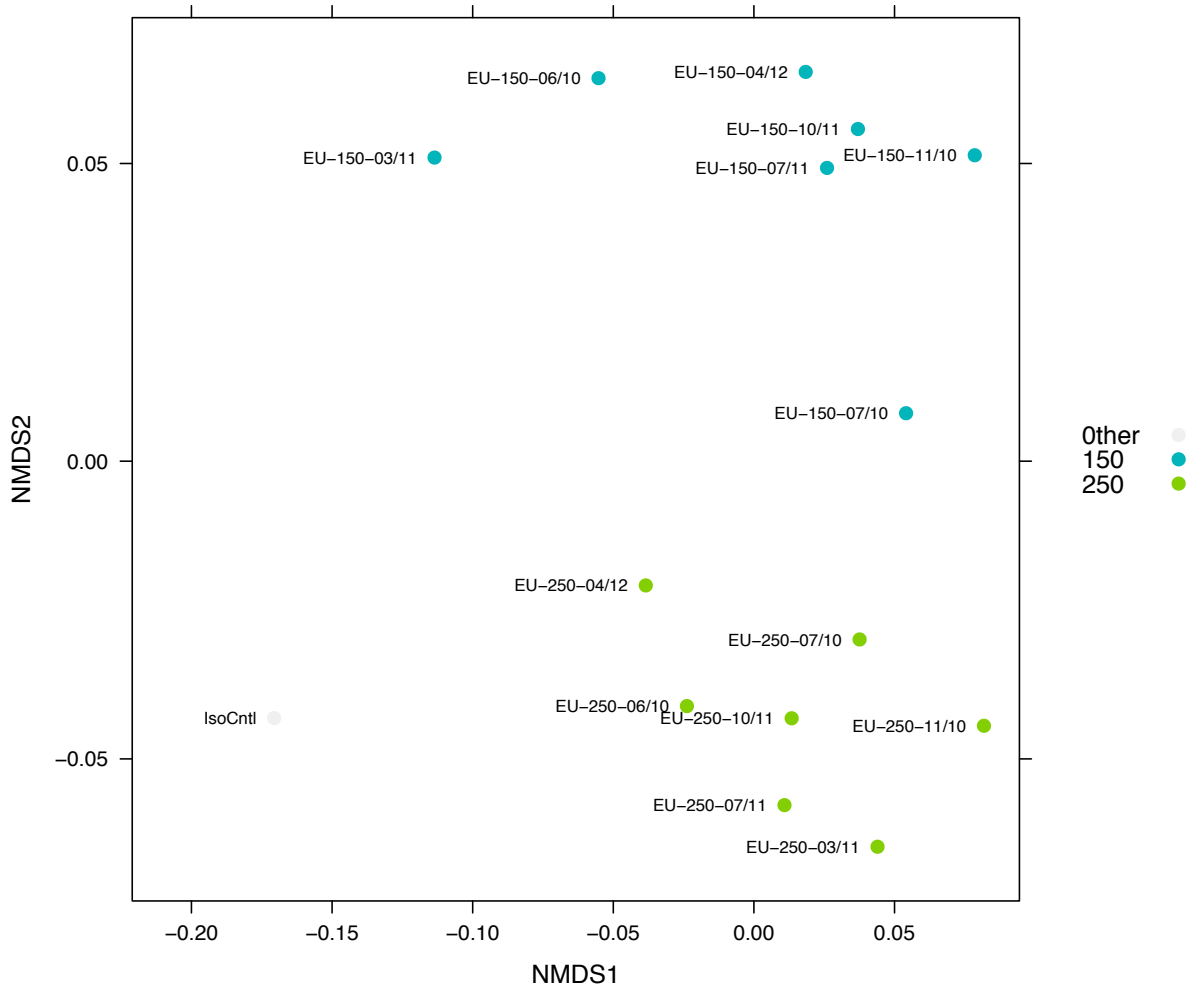


FIGURE 1-9

OBSERVATIONS

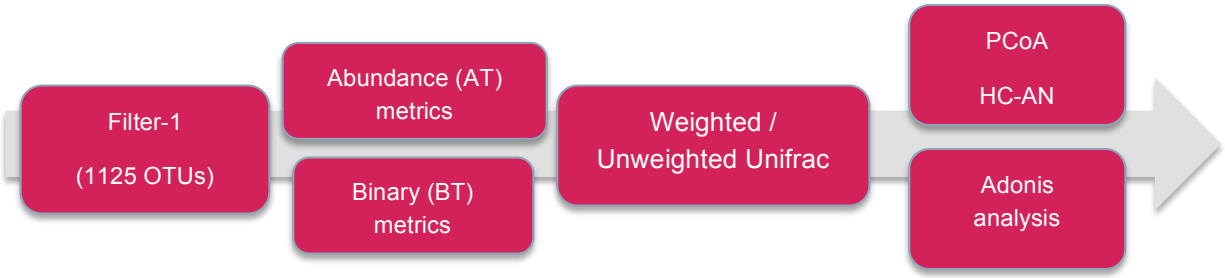
1. Abundance Bray-Curtis distance analysis shows a separation of the Isolation Control from other samples.
2. Sample Isolation Control is removed from the entire analysis in the following chapters, since it is an outlier concerning bacterial richness and ordination analysis.
3. Taxa present in the Isolation Control are removed from further analyses.

Chapter 2: Whole Microbiome Analysis

In this chapter, we consider analyses of beta diversity, or explicit comparisons between samples, considering data from the whole microbiome.

GENERAL OBSERVATIONS

1. Module 150 and 250 samples form separate groups in ordination analysis using abundance and binary Bray-Curtis Distance.
2. Adonis test suggests a significant difference between the two modules' microbiome.



SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 2-1	*./Whole_Microbiome/at1.bray.NMDS.Module.pdf*
Figure 2-2	*./Whole_Microbiome/at1.bray.NMDS.Timepoint.pdf*
Figure 2-3	*./Whole_Microbiome/at1.bray.NMDS.Comp1.pdf*
Figure 2-4	*./Whole_Microbiome/at1.bray.NMDS.Comp2.pdf*
Figure 2-5	*./Whole_Microbiome/at1.bray.HCAN.pdf*
Figure 2-6	*./Whole_Microbiome/bt1.bray.NMDS.Module.pdf*
Figure 2-7	*./Whole_Microbiome/bt1.bray.NMDS.Timepoint.pdf*
Figure 2-8	*./Whole_Microbiome/bt1.bray.NMDS.Comp1.pdf*
Figure 2-9	*./Whole_Microbiome/bt1.bray.NMDS.Comp2.pdf*
Figure 2-10	*./Whole_Microbiome/bt1.bray.HCAN.pdf*

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./whole_microbiome/at1_wunifrac_adonis_table.txt	*Adonis test score for abundance metric.*
./whole_microbiome/bt1_unifrac_adonis_table.txt	*Adonis test score for binary metrics.*

FIGURE 2-1

(at1.bray.NMDS.Module.pdf) NMDS based on Bray-Curtis distance between samples given abundance of 1125 taxa present in at least one sample. Stress=0.1417

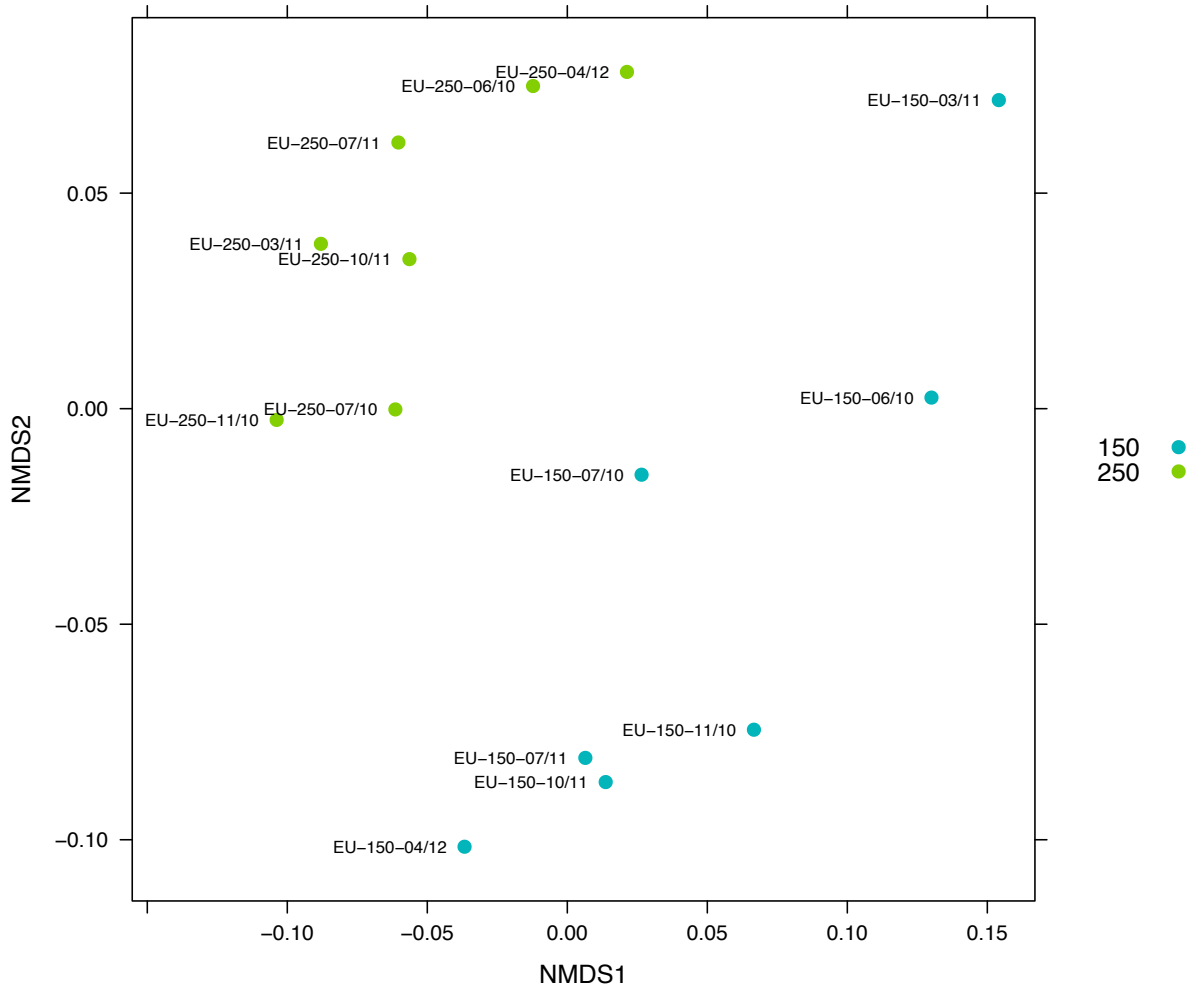


FIGURE 2-1

OBSERVATIONS

1. Abundance Bray-Curtis distance analysis shows a distinct separation of microbiomes according to the factor Module.
2. Using the Bray-Curtis distance on the abundance of 1125 taxa present in at least one sample, the Adonis test yields a p-value of 0.003, indicating a significant microbiome difference is observed between module 150 and 250 samples.

FIGURE 2-2

(at1.bray.NMDS.Timepoint.nolabels) NMDS based on Bray-Curtis distance between samples given abundance of 1125 taxa present in at least one sample. Stress=0.1417

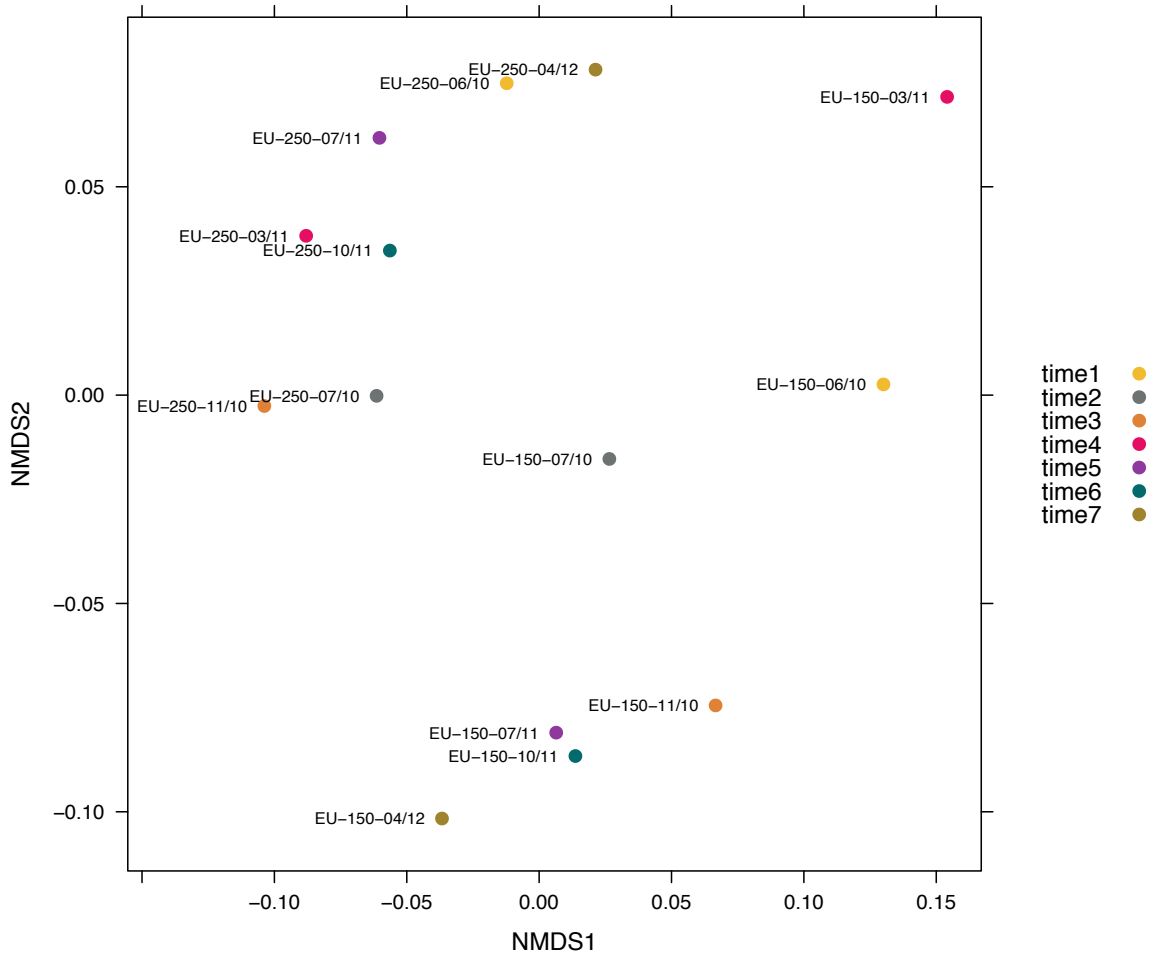


FIGURE 2-2

OBSERVATIONS

1. Abundance Bray-Curtis distance analysis shows no distinct separation of microbiomes according to the factor Timepoint. Samples collected at different time points from module 150 are distinct from module 250 samples.
2. Using the Bray-Curtis distance on the abundance of 1125 taxa present in at least one sample, the Adonis test yields a p-value of 0.748, indicating no significant microbiome difference is observed between different time points.

FIGURE 2-3

(at1.bray.NMDS.Comp1.pdf) NMDS based on Bray-Curtis distance between samples given abundance of 1125 taxa present in at least one sample. Stress=0.1417

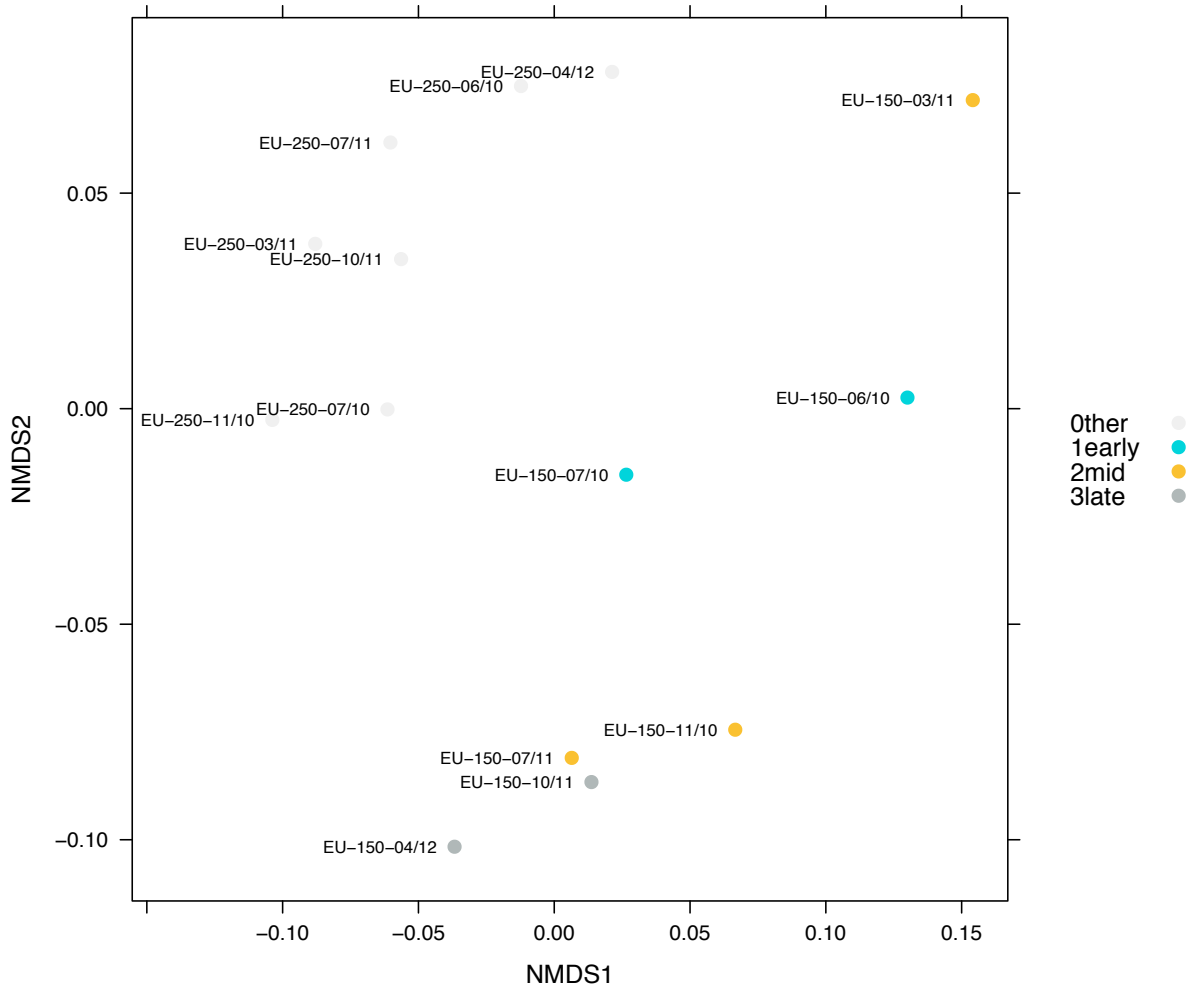


FIGURE 2-3

OBSERVATIONS

1. Abundance Bray-Curtis distance analysis shows no distinct separation of microbiomes according to the factor Comp1.
2. Using the Bray-Curtis distance on the abundance of 1125 taxa present in at least one sample, the Adonis test yields a p-value of 0.302, indicating no significant microbiome difference is observed between different time points.

FIGURE 2-4

(at1.bray.NMDS.Comp2.pdf) NMDS based on Bray-Curtis distance between samples given abundance of 1125 taxa present in at least one sample. Stress=0.1417

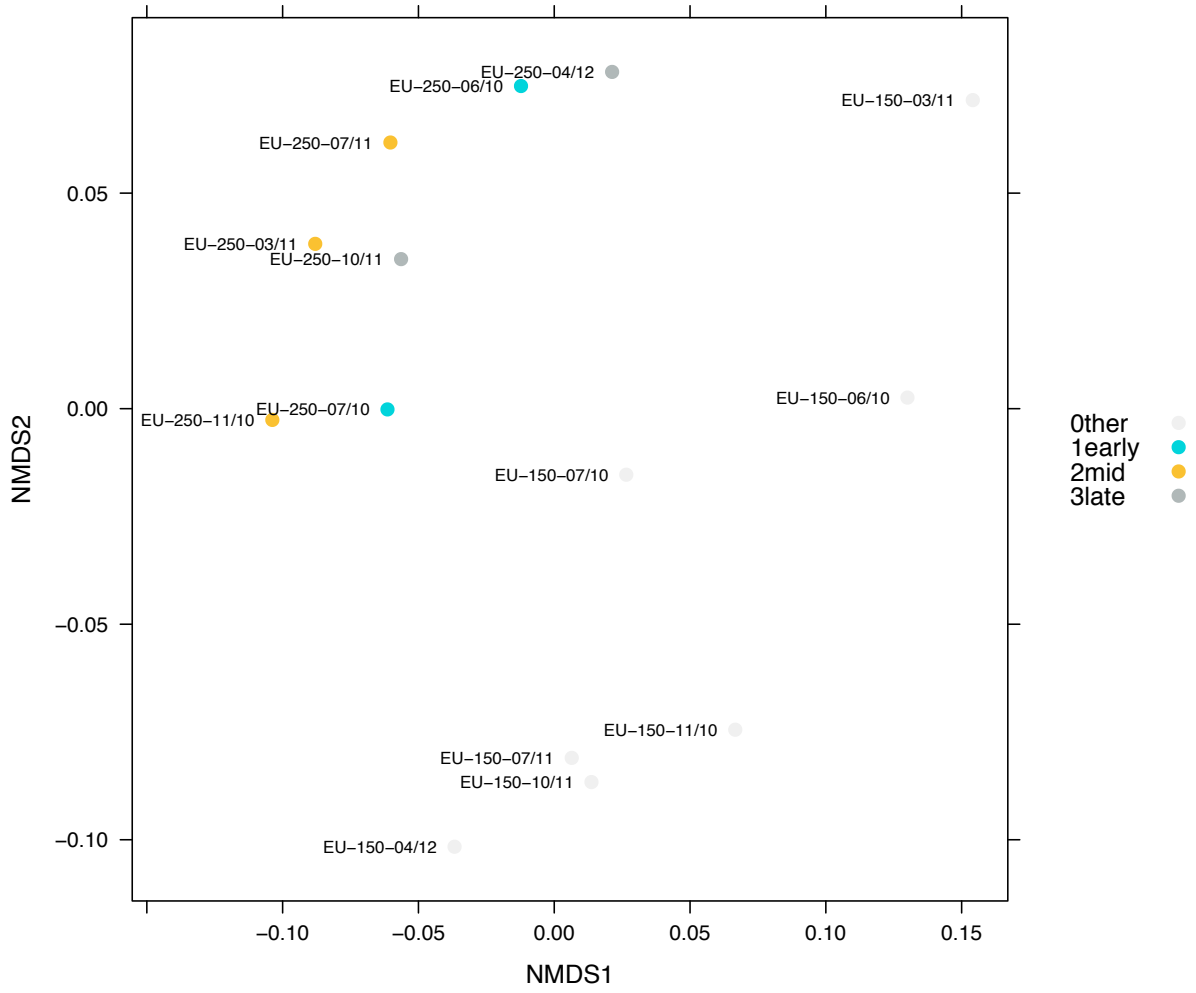


FIGURE 2-4

OBSERVATIONS

1. Abundance Bray-Curtis distance analysis shows no distinct separation of microbiomes of samples early, mid and late from module 250.
2. Using the Bray-Curtis distance on the abundance of 1125 taxa present in at least one sample, the Adonis test yields a p-value of **0.186**, indicating no significant microbiome difference is observed between different sample categories.

FIGURE 2-5

(at1.bray.HCAN) Hierarchical Clustering (average linkage) based on Bray-Curtis distance between samples given abundance of 1125 taxa present in at least one sample.

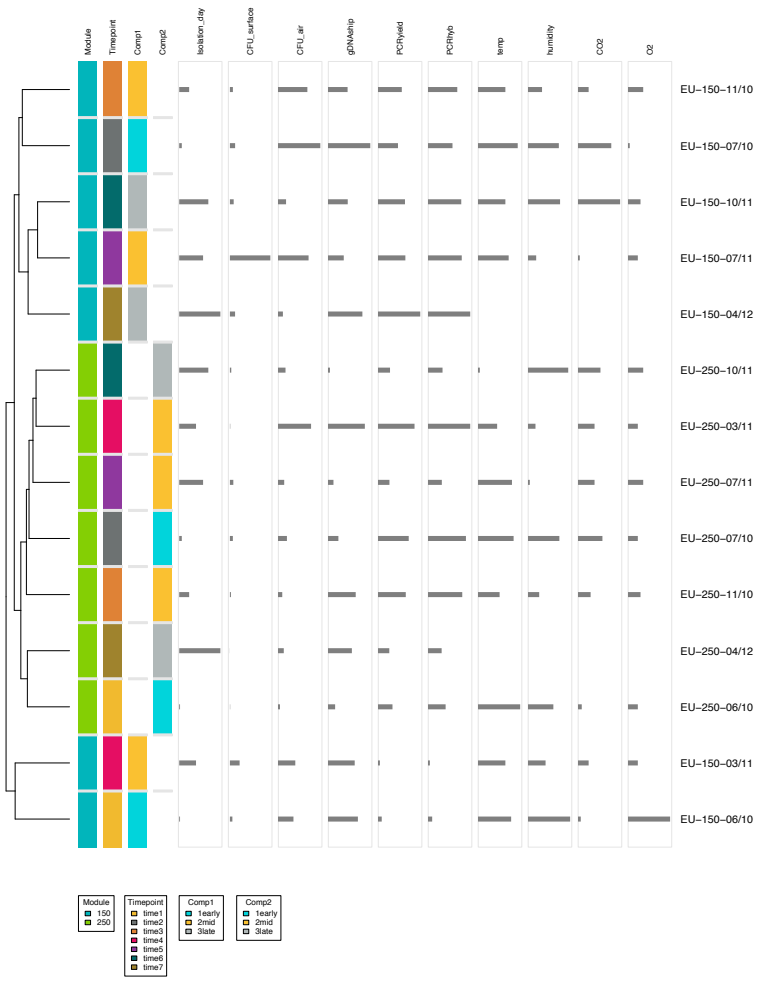


FIGURE 2-5

OBSERVATIONS

1. HC-AN analysis based on abundance metrics of 1125 taxa reveals clusters of 150 and 250 module samples.
2. With regard to any category, no separation is observed.
3. Three major clusters are observed, each of them is comprised uniformly of samples from one module.

FIGURE 2-6

(bt1.bray.NMDS.Module.pdf) NMDS based on Bray-Curtis distance between samples given presence/absence of 1125 taxa present in at least one sample. Stress=0.1415.

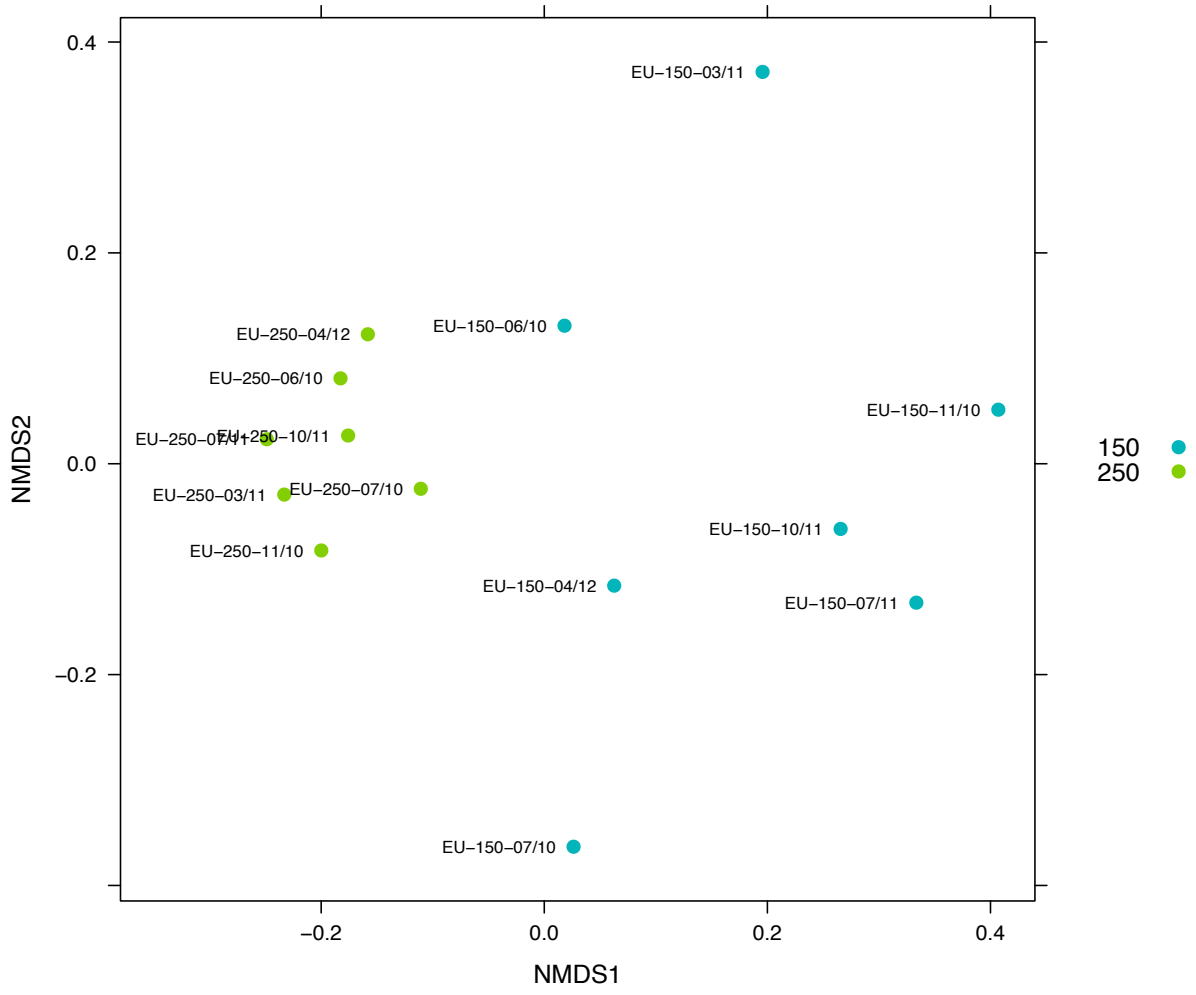


FIGURE 2-6

OBSERVATIONS

1. Tight clustering of module 250 samples is observed.
2. Using the Bray-Curtis distance on the presence/absence of 1125 taxa present in at least one sample, the Adonis test yields a p-value of 0.002, indicating a significant microbiome difference is observed between module 150 and 250 categories.

FIGURE 2-7

(bt1.bray.NMDS.Timepoint.pdf) NMDS based on Bray-Curtis distance between samples given presence/absence of 1125 taxa present in at least one sample. Stress=0.1415.

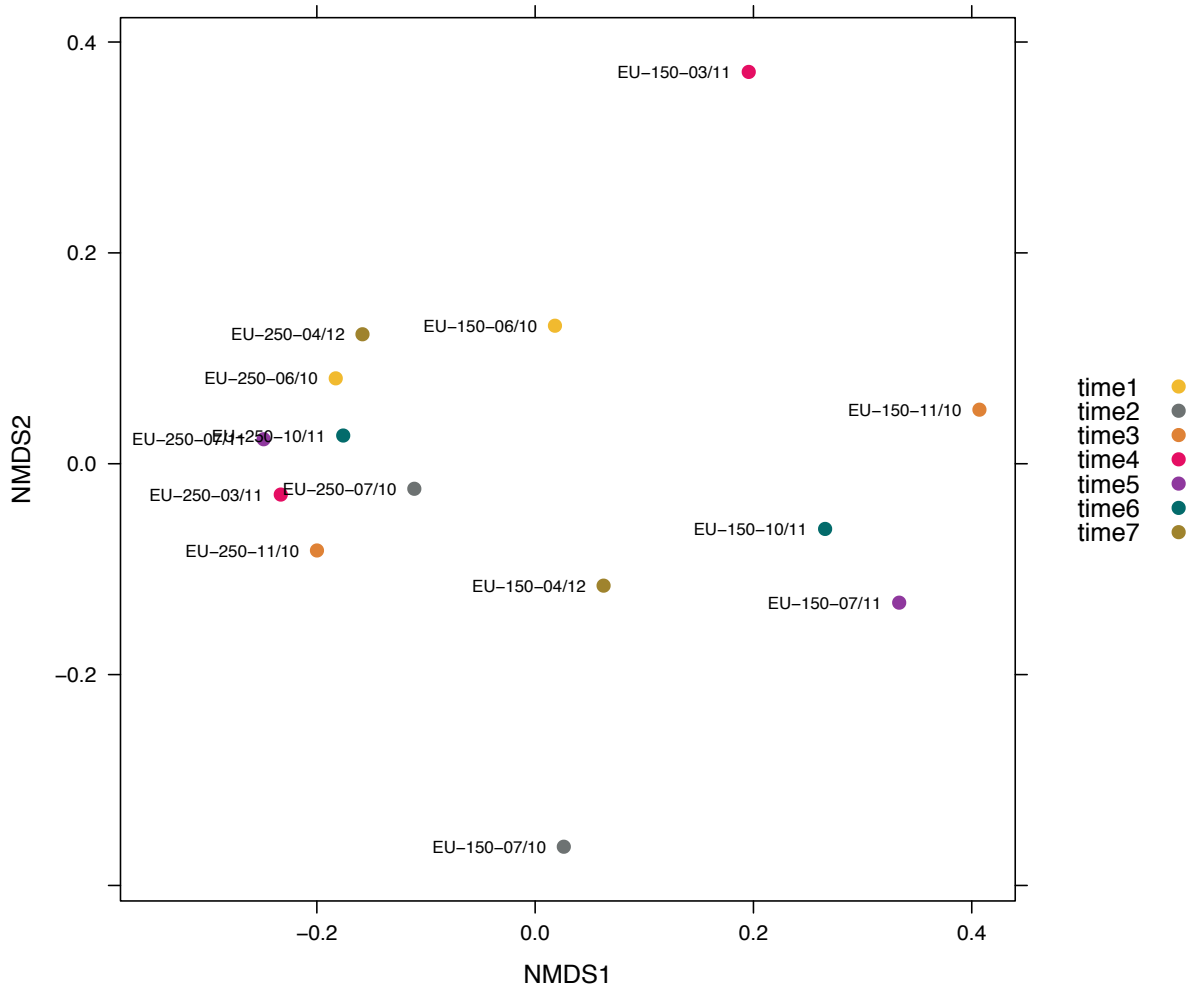


FIGURE 2-7

OBSERVATIONS

1. No significant separation of microbiomes between samples in Timepoint category is observed. Insufficient replicates (two samples per group) are present in each group.
2. Using the Bray-Curtis distance on the presence/absence of 1125 taxa present in at least one sample, the Adonis test yields a p-value of **0.861**, indicating no significant microbiome difference is observed between the Timepoint groups.

FIGURE 2-8

(bt1.bray.NMDS.Comp1.pdf) NMDS based on Bray-Curtis distance between samples given presence/absence of 1125 taxa present in at least one sample. Stress=0.1415.

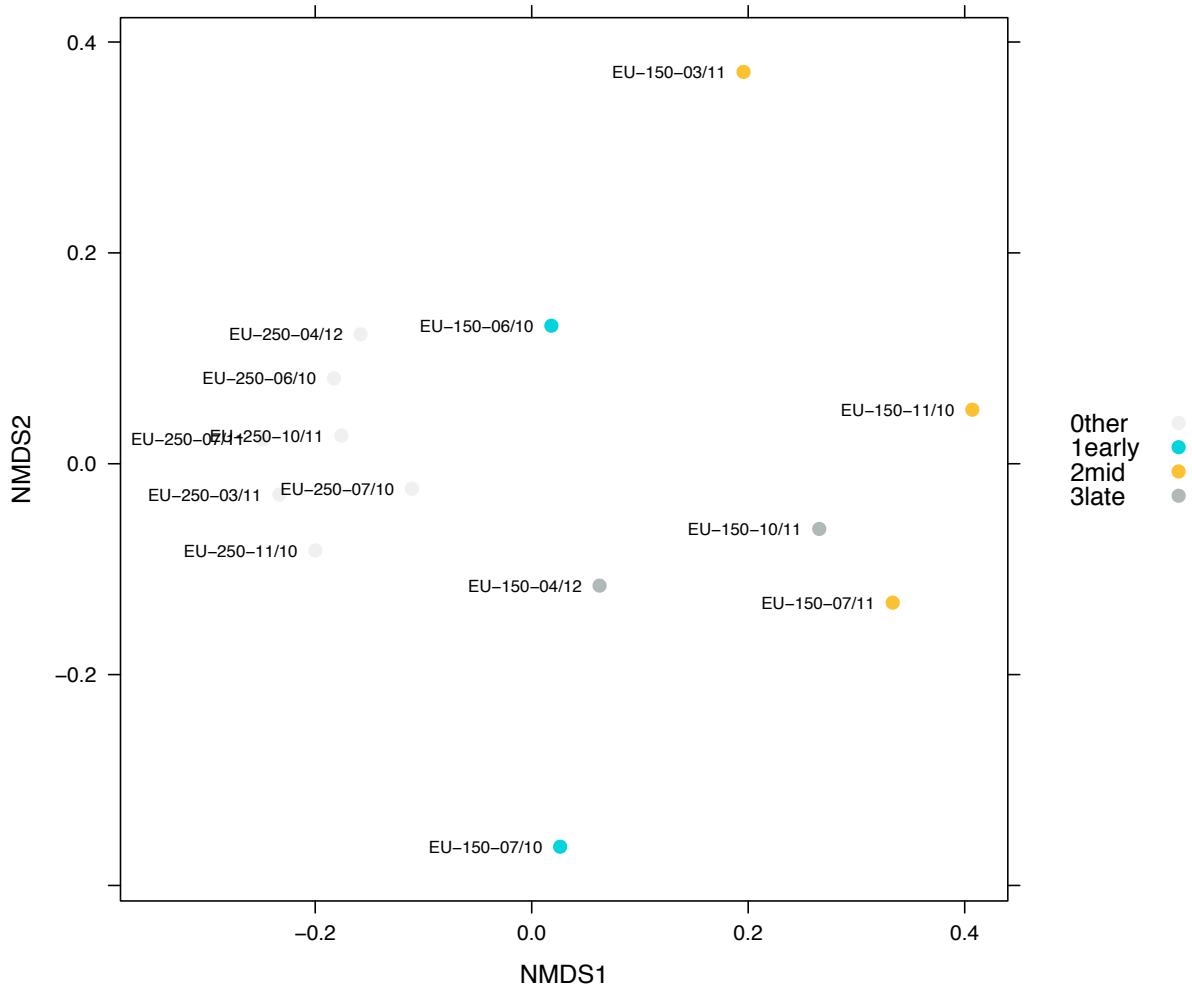


FIGURE 2-8

OBSERVATIONS

1. No significant separation of microbiome communities is observed between module 150 samples early, mid and late.
2. Using the Bray-Curtis distance on the presence/absence of 1125 taxa present in at least one sample, the Adonis test yields a p-value of **0.584**, indicating no significant microbiome difference is observed between sample categories.

FIGURE 2-9

(bt1.bray.NMDS.Comp2.pdf) NMDS based on Bray-Curtis distance between samples given presence/absence of 1125 taxa present in at least one sample. Stress=0.1415.

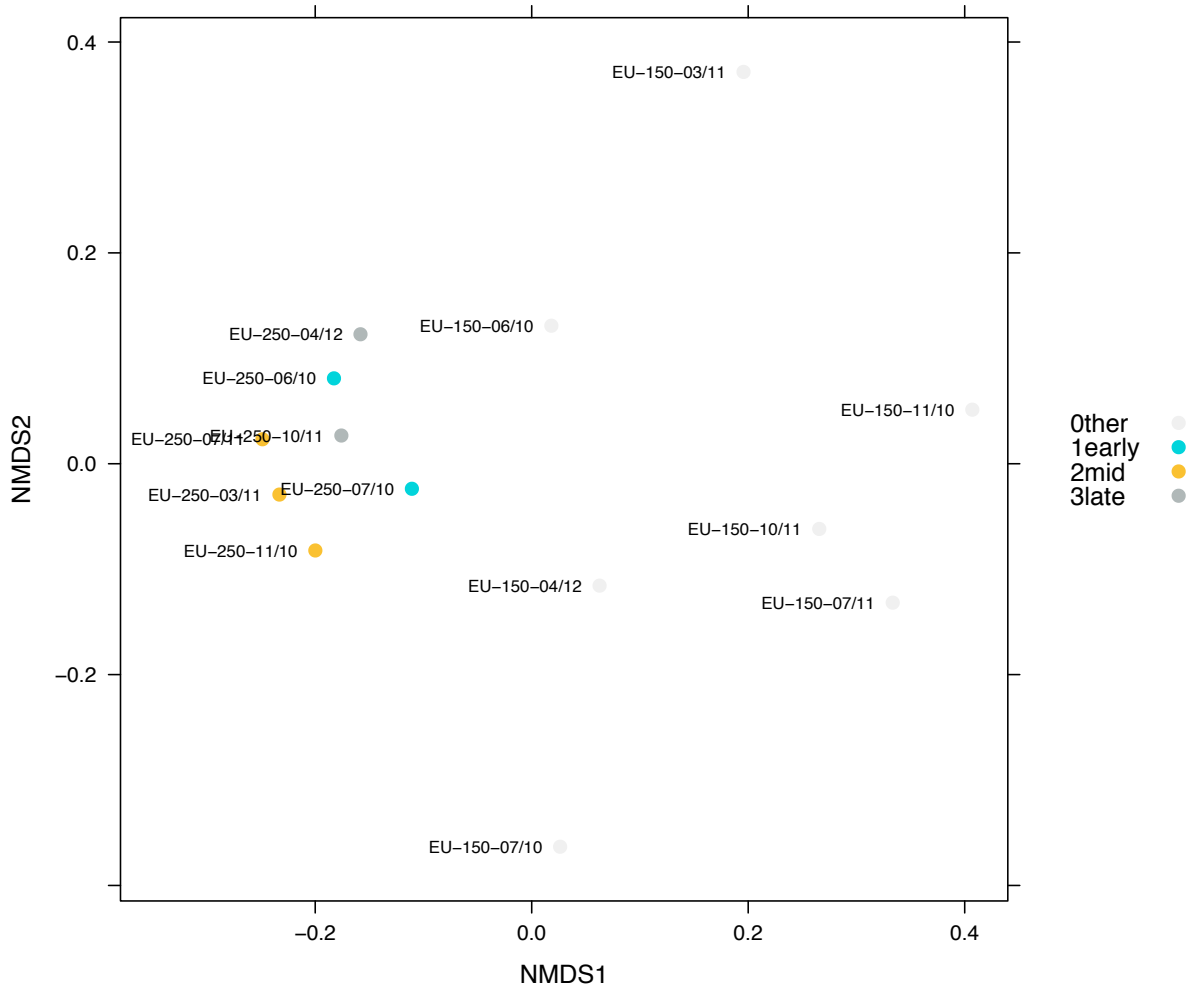


FIGURE 2-9

OBSERVATIONS

1. A trend of microbiome separation is observed between module 250 samples mid from early and late.
2. Using the Bray-Curtis distance on the presence/absence of 1125 taxa present in at least one sample, the Adonis test yields a p-value of 0.031, indicating a significant microbiome difference is observed between at least one of the three sampling categories.

FIGURE 2-10

(bt1.bray.HCAN) Hierarchical Clustering (average linkage) based on Bray-Curtis distance between samples given presence/absence of 1125 taxa present in at least one sample.

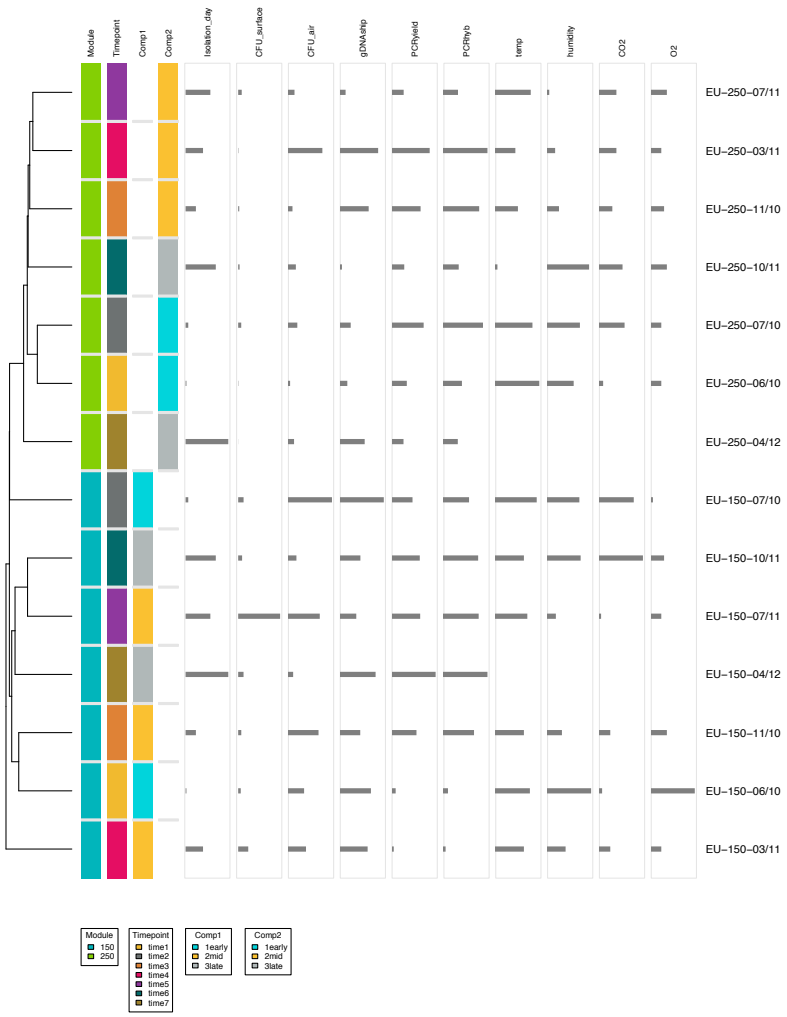


FIGURE 2-10

OBSERVATIONS

1. HC-AN analysis based on presence/absence metrics of 1125 taxa reveals significance clustering of module 150 and 250 samples.
2. Two major clusters are observed. One of them is comprised of module 150 samples only.

TABLE 2-1. ADONIS TEST AND SIGNIFICANCE SUMMARY

Bold font indicates statistical significance reached, p<0.05.

TABLE 2-1. SIGNIFICANCE SUMMARY TABLE.

Category	Bins	Sample Count	BrayCurtisDist at1	BrayCurtisDist bt1	HC-AN Distinct			Differentiating OTU Count
					at1	bt1	at5	at5
Module	150 | 250	7 | 7	**0.003**	**0.002**	Yes	Yes	Yes	279
Timepoint	time1 | time2 | time3 | time4 | time5 | time6 | time7	2 | 2 | 2 | 2 | 2 | 2 | 2	0.748	0.861	No	No	NA	NA
Comp1 (mod150)	1early | 2mid | 3late	2 | 3 | 2	0.302	0.584	No	No	No	62
Comp2 (mod250)	1early | 2mid | 3late	2 | 3 | 2	0.186	**0.031**	No	No	No	53
Isolation_day	continuous	14	0.243	0.485	NA	NA	NA	NA
CFU_surface	continuous	14	0.213	**0.020**	NA	NA	NA	NA
CFU_air	continuous	14	0.619	0.190	NA	NA	NA	NA
gDNAship	continuous	14	0.702	0.567	NA	NA	NA	NA
PCRyield	continuous	14	**0.007**	0.146	NA	NA	NA	NA
PCRhyb	continuous	14	**0.005**	0.121	NA	NA	NA	NA
temp	continuous	12	0.549	0.773	NA	NA	NA	NA
humidity	continuous	12	0.533	0.516	NA	NA	NA	NA
CO2	continuous	12	0.456	0.559	NA	NA	NA	NA
O2	continuous	12	0.231	0.512	NA	NA	NA	NA

OBSERVATIONS

1. A significant microbiome difference is observed in Module factor based on Bray-Curtis distance on the abundance metric, and Bray-Curtis distance on the presence/absence of 1125 taxa.
2. A significant microbiome difference is observed in grouping samples by PCRyield and PCRhyb based on Bray-Curtis distance of the abundance metric, and CFU_surface samples based on Bray-Curtis distance of the presence/absence of 1125 taxa.

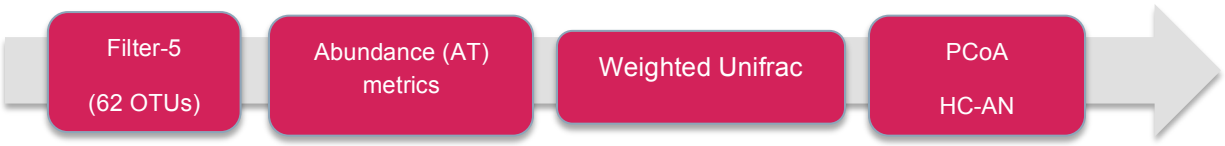
Chapter 3: Comparison between samples early, mid and late of module 150

In this chapter, we performed a parametric Welch test to look for those OTUs that are significantly increased or decreased in samples early, mid, and late of module 150.

Please note, that the Welch test performed did not have sufficient samples per group in two out of three categories.

GENERAL OBSERVATIONS

1. On the bases of at5 on Bray-Curtis distance, given abundance of 62 taxa revealed a distinct microbiome between samples early, mid and late of modules 150.
2. Clustering of microbiomes of early samples from other samples on the basis of HCAN analysis is observed.



SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 3-1	*./Comp1/at5.wunifrac.PCoA.pdf*
Figure 3-2	*./Comp1/at5.wunifrac.HCAN.pdf*
Figure 3-3	*./Comp1/at5.profs.pdf*

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./Comp1/at5_p.table.txt	*HybScores (abundance metrics) of taxa with significant abundance differences across at least one of the categories ($p < 0.05$).*
./Comp1/at5_p.table.txt.annotated.txt	

FIGURE 3-1

(at5.bray.NMDS.Comp1.pdf) NMDS based on Bray-Curtis distance between samples given abundance of 62 taxa with significant abundance differences across at least one of the categories. Stress=0.1175.

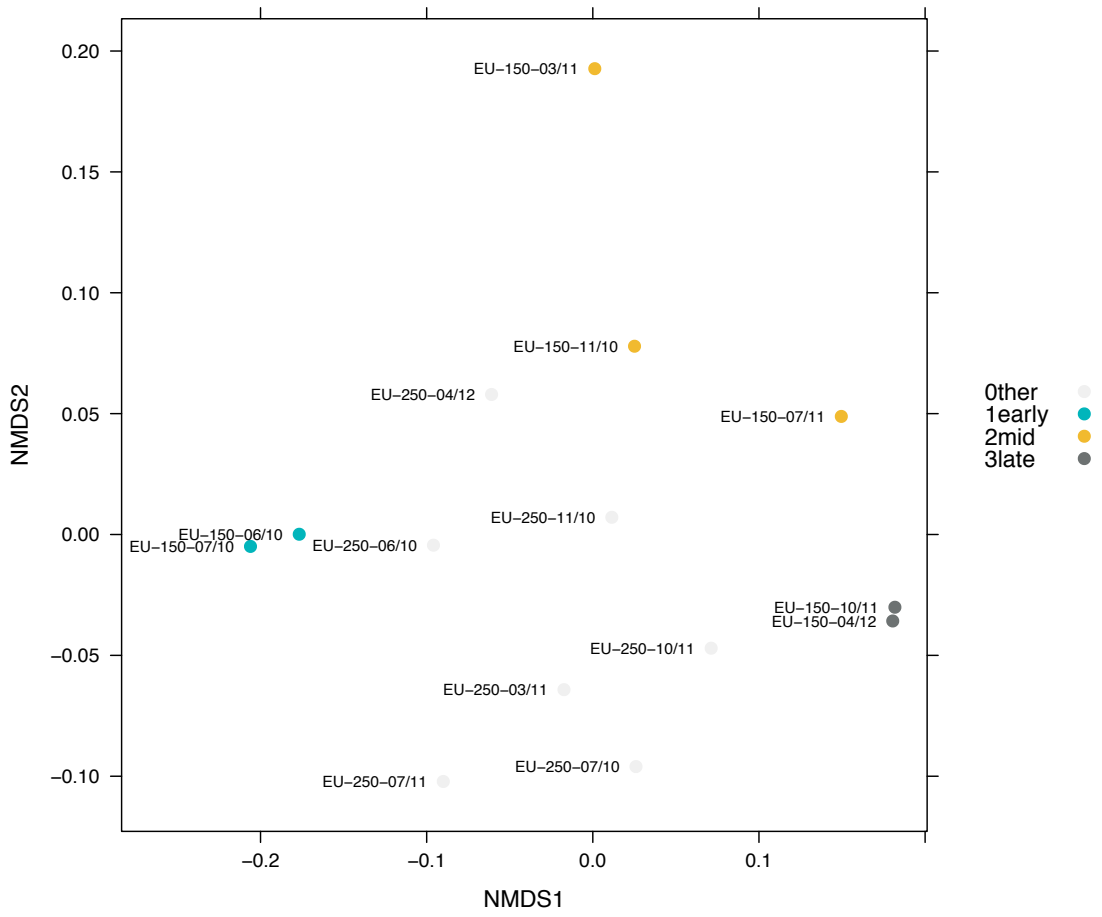


FIGURE 3-1

OBSERVATIONS

1. A significant separation of microbiome communities is observed for module 150 samples from early, mid and late categories.
2. Insufficient numbers of samples (two) are present in early and late categories to adequately power this analysis.

FIGURE 3-2

(at5.bray.HCAN.pdf) Hierarchical Clustering (average linkage) based on Bray-Curtis distance between samples given abundance of 62 taxa with significant abundance differences across at least one of the categories.

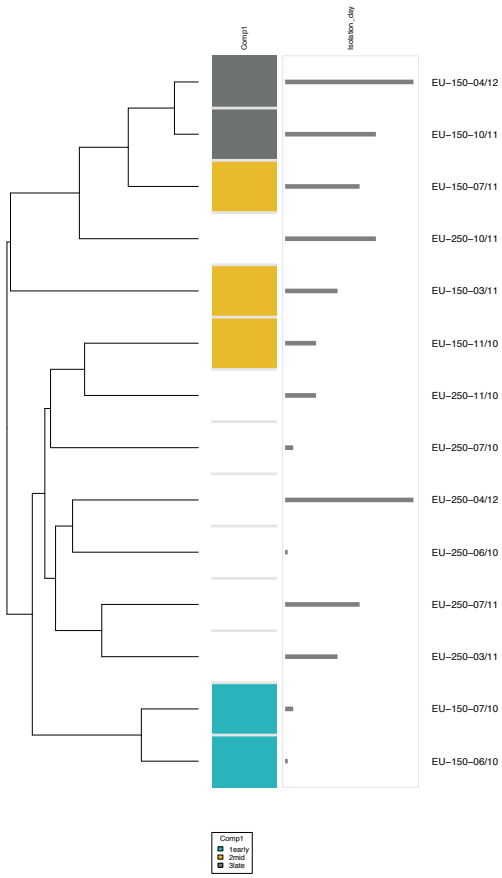


FIGURE 3-2

OBSERVATIONS

1. Based on the abundance of 62 OTUs in the microbiome community characterization, early samples form a distinct cluster from mid and late samples.

FIGURE 3-3

(Comp1.at5.profs.pdf) Profiles of at5 OTUs generating the lowest p-values. P-values shown at top of each OTU plot are unadjusted for multiple testing. The y-axis represents the HybScore. Samples are grouped and colored by category along the x-axis in the following order: EU-150-06/10, EU-150-07/10, EU-150-11/10, EU-150-03/11, EU-150-07/11, EU-150-10/11, EU-150-04/12.

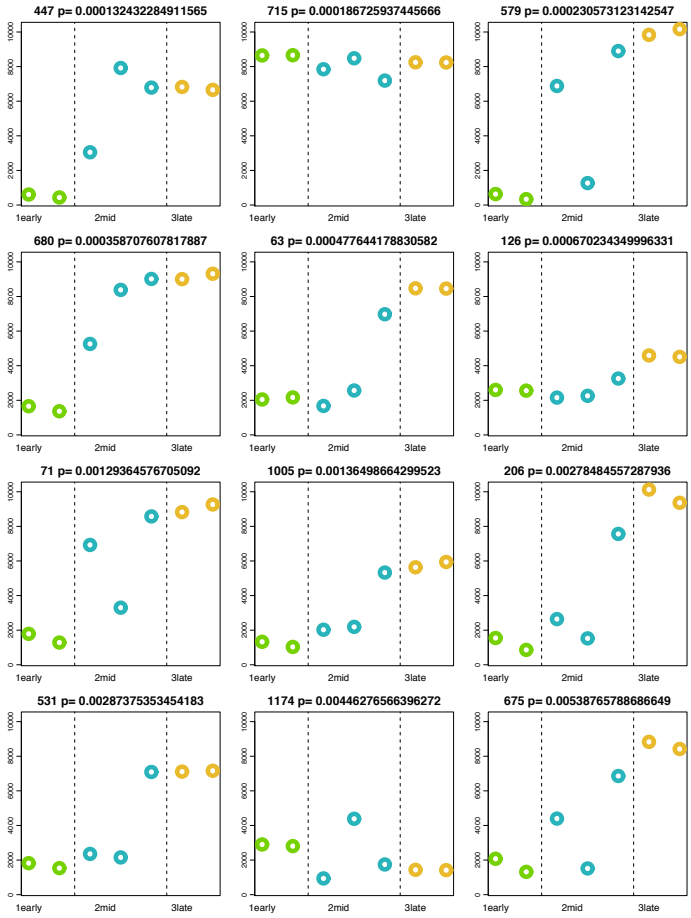


FIGURE 3-3

OBSERVATIONS

1. All of the top 12 selected OTUs belong to 1 of 5 phyla: Proteobacteria (6), Actinobacteria (2), Bacteroidetes (2), Firmicutes (1) and Planctomycetes (1).
2. 9 of 12 selected OTUs display a significant increase in late samples as compared to early and mid samples.

TAXONOMIC ANNOTATIONS

TABLE 3-1. (AT5_P.TABLE.TXT.ANNOTATED.TXT) ANNOTATIONS OF THE OTUS WITH THE LOWEST P-VALUES.

Taxa ID	kingdom	phylum	class	order	family	genus
447	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae	Brucella
715	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter
579	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	unclassified
680	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Haloanella
63	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Roseomonas
126	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	unclassified	unclassified
71	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Paracoccus
1005	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	unclassified
206	Bacteria	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae	Sphingobacterium
531	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium
1174	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	unclassified
675	Bacteria	Planctomycetes	Planctomycea	Gemmatales	Isosphaeraceae	unclassified

Chapter 4: Comparison between samples early, mid and late of module 250

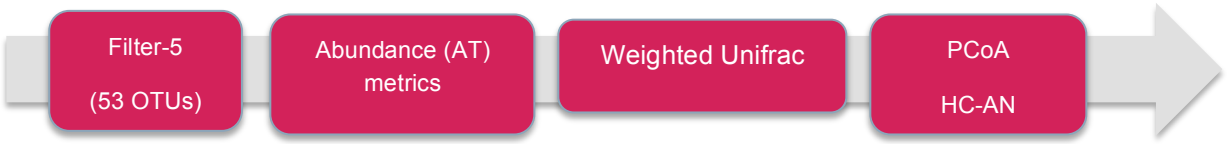
In this chapter, we performed a parametric Welch test to look for those OTUs that are significantly increased or decreased in samples early, mid and late of module 250.

Please note that the Welch test performed does not have sufficient samples per group in two out of three categories.

GENERAL OBSERVATIONS

1. On the basis of at5 on Bray-Curtis distance analysis of abundance of 53 taxa distinguished module 250 samples early, mid and late.

HCAN analysis also revealed distinct clustering.



SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 4-1	*./Comp2/at5.wunifrac.PCoA.pdf*
Figure 4-2	*./Comp2/at5.wunifrac.HCAN.pdf*
Figure 4-3	*./Comp2/at5.profs.pdf*

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./ Comp2/at5_p.table.txt	*HybScores (abundance metrics) of taxa with significant abundance differences across at least one of the categories ($p < 0.05$).*
./ Comp2/at5_p.table.txt.annotated.txt	

FIGURE 4-1

(at5.bray.NMDS.Comp2.pdf) NMDS based on Bray-Curtis distance between samples given abundance of 53 taxa with significant abundance differences across at least one of the categories Stress=0.1441.

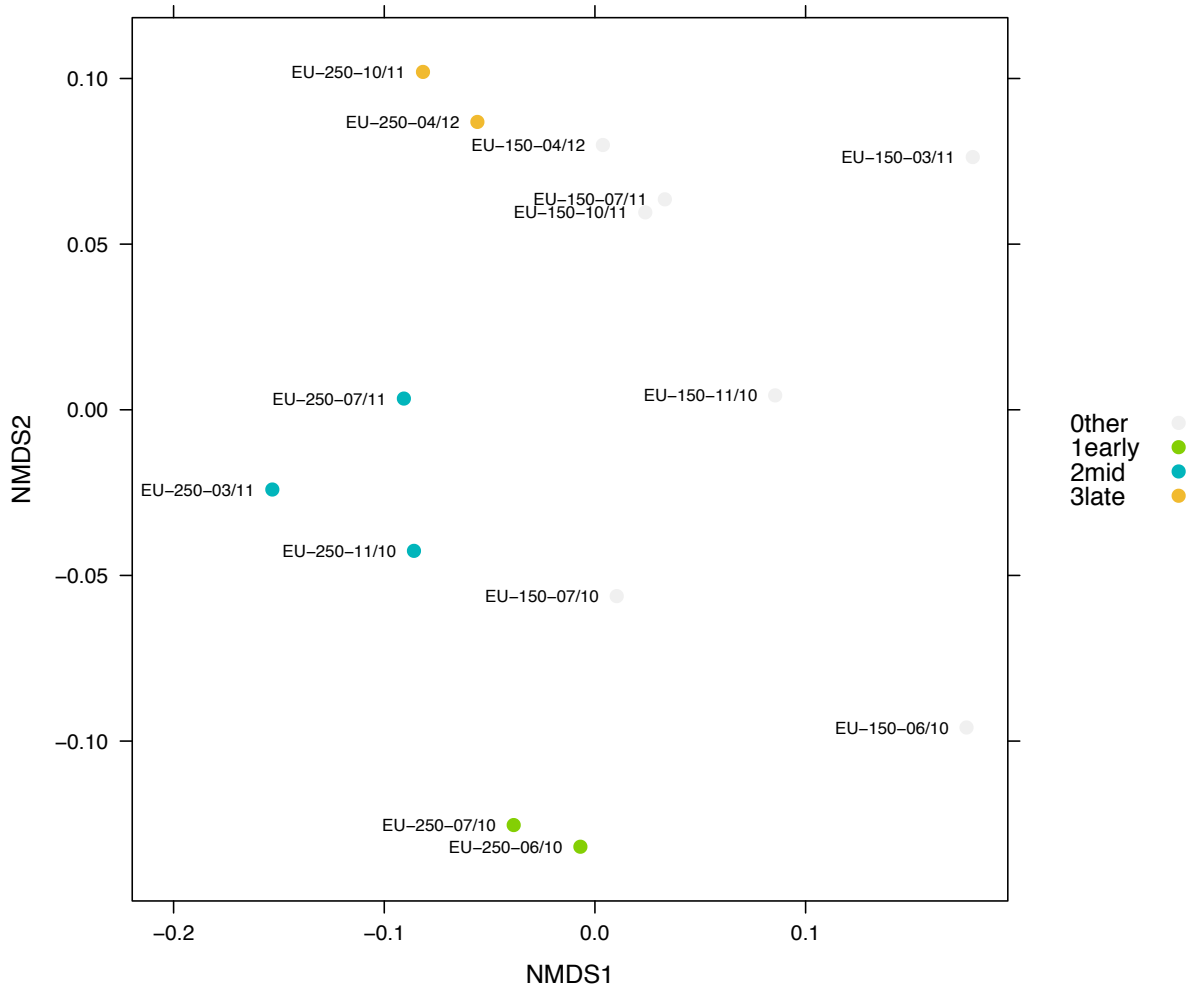


FIGURE 4-1

OBSERVATIONS

1. Module 250 samples in early, mid and late categories display significant microbiome community separation.

FIGURE 4-2

(at5.bray.HCAN.pdf) Hierarchical Clustering (average linkage) based on Bray-Curtis distance between samples given abundance of 53 taxa with significant abundance differences across at least one of the categories.

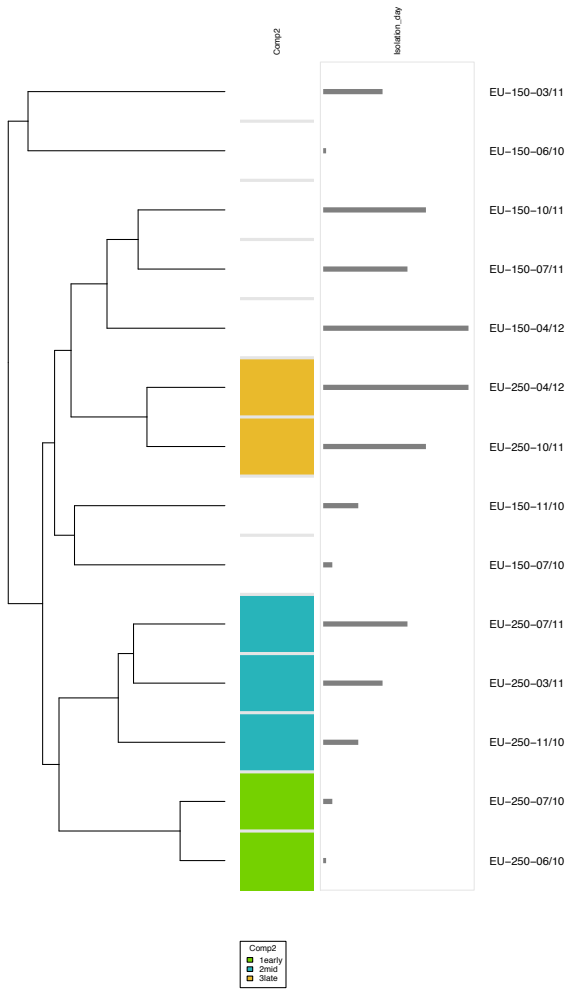


FIGURE 4-2

OBSERVATIONS

1. The microbiomes of module 250 samples in early, mid and late categories form separate clusters based on the 53 OTUs abundance.

FIGURE 4-3

(Comp2.at5.profs.pdf) Profiles of at5 OTUs generating the lowest p-values. P-values shown at top of each OTU plot are unadjusted for multiple testing. The y-axis represents the HybScore. Samples are grouped and colored by category along the x-axis in the following order: EU-250-06/10, EU-250-07/10, EU-250-11/10, EU-250-03/11, EU-250-07/11, EU-250-10/11, EU-250-04/12.

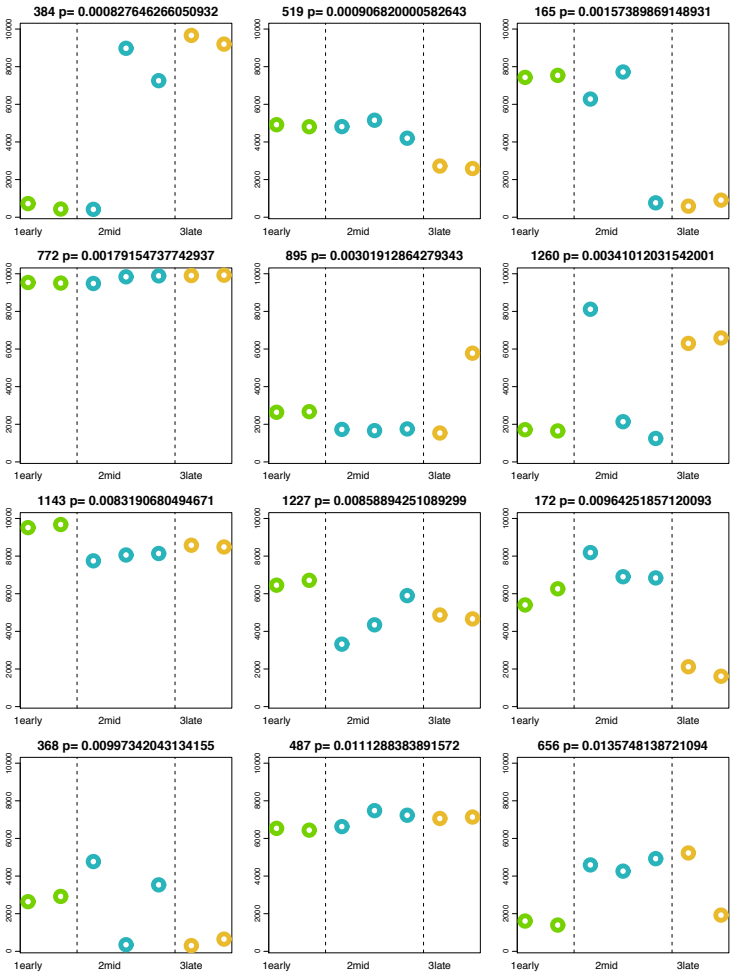


FIGURE 4-3

OBSERVATIONS

3. All of the top 12 selected OTUs belong to 1 of 5 phyla: Proteobacteria (5), Fusobacteria (2), Firmicutes (3), Actinobacteria (1) and Bacteroidetes (1).
4. 4 of 12 selected OTUs display a significant increase in early samples.
5. OTU 1227, *Enterobacteriaceae* is significantly greater in all early samples.

TAXONOMIC ANNOTATIONS

TABLE 4-1. (AT5_P.TABLE.TXT.ANNOTATED.TXT) ANNOTATIONS OF THE OTUS WITH THE LOWEST P-VALUES.

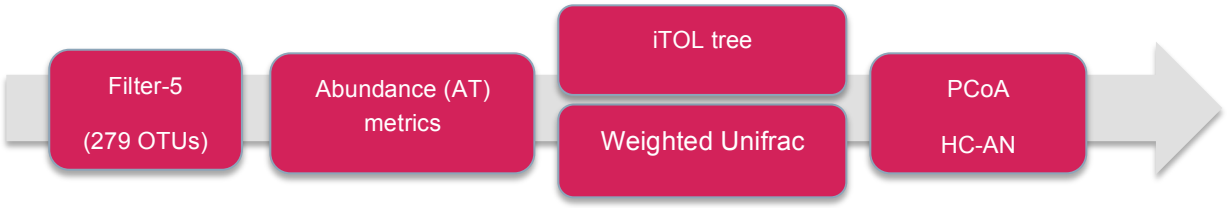
Taxa ID	kingdom	phylum	class	order	family	genus
384	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
519	Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	unclassified
165	Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	Leptotrichia
772	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	unclassified
895	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Nitrospinaceae	unclassified
1260	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Brevibacteriaceae	Brevibacterium
1143	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	unclassified
1227	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	unclassified
172	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella
368	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	unclassified
487	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Veillonella
656	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas

Chapter 5: Comparison between modules 150 and 250

In this chapter, we performed a parametric Welch test to look for those OTUs that are significantly increased or decreased in samples based on the different Modules.

GENERAL OBSERVATIONS

1. On the bases of at5 Bray-Curtis distance, the abundances of 279 taxa reveal distinct microbiomes between modules 150 and 250.



SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 5-1	*./module/at5.bray.PCoA.module.pdf*
Figure 5-2	*./module/at5.bray.HCAN.pdf*
Figure 5-3	*./module/module.at5.profs.pdf*
Figure 5-4	*./module/Circular.tree.pdf*

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./module/at5_p.table.txt.annotated.txt	*HybScores (abundance metrics) of taxa with significant abundance differences across at least one of the categories ($p < 0.05$).*

FIGURE 5-1

(at5.bray.NMDS.Module.pdf) NMDS based on Bray-Curtis distance between samples given abundance of 279 taxa with significant abundance differences across at least one of the categories. Stress=0.0553.

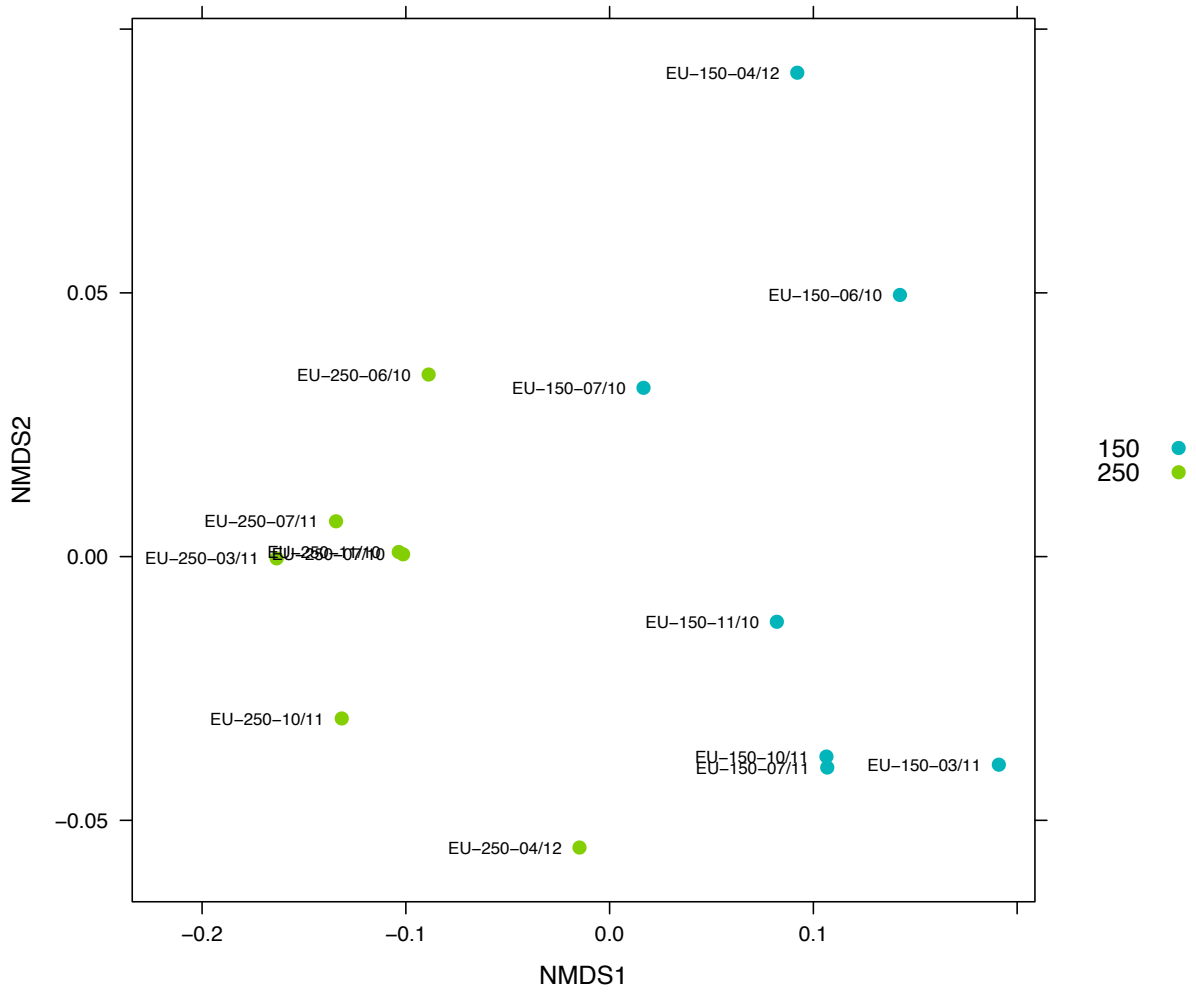


FIGURE 5-1

OBSERVATIONS

1. A significant separation of microbiome between samples from 150 vs. 250 modules is observed.

FIGURE 5-2

(at5.bray.HCAN.pdf) Hierarchical Clustering (average linkage) based on Bray-Curtis distance between samples given abundance of 279 taxa with significant abundance differences across at least one of the categories.

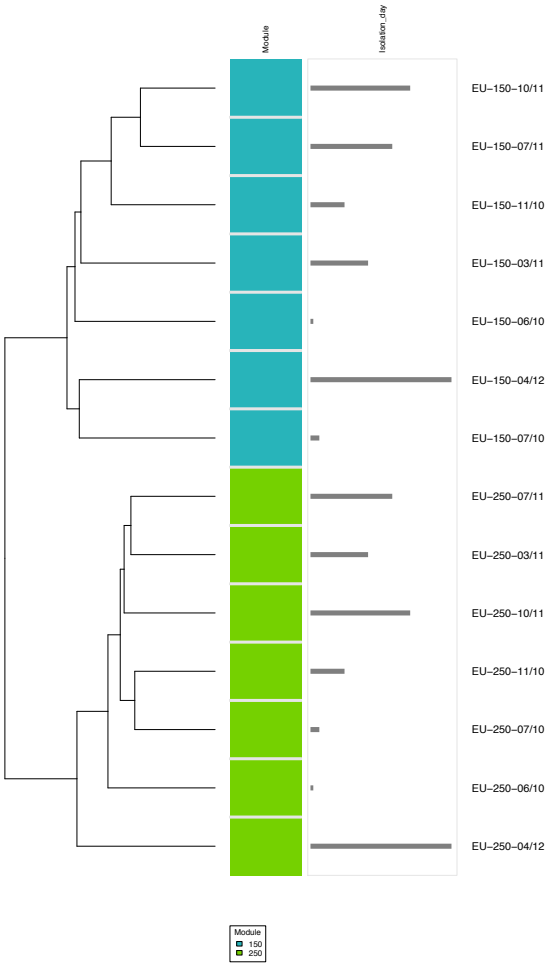


FIGURE 5-2

OBSERVATIONS

1. The microbiome of 150 and 250 samples form distinct clusters based on the 279 OTUs abundance.

FIGURE 5-3

(modules.at5.profs.pdf) Profiles of at5 OTUs generating the lowest p-values. P-values shown at top of each OTU plot are unadjusted for multiple testing. The y-axis represents the HybScore. Samples are grouped and colored by category along the x-axis in the following order: EU-150-06/10, EU-150-07/10, EU-150-11/10, EU-150-03/11, EU-150-07/11, EU-150-10/11, EU-150-04/12, EU-250-06/10, EU-250-07/10, EU-250-11/10, EU-250-03/11, EU-250-07/11, EU-250-10/11, EU-250-04/12.

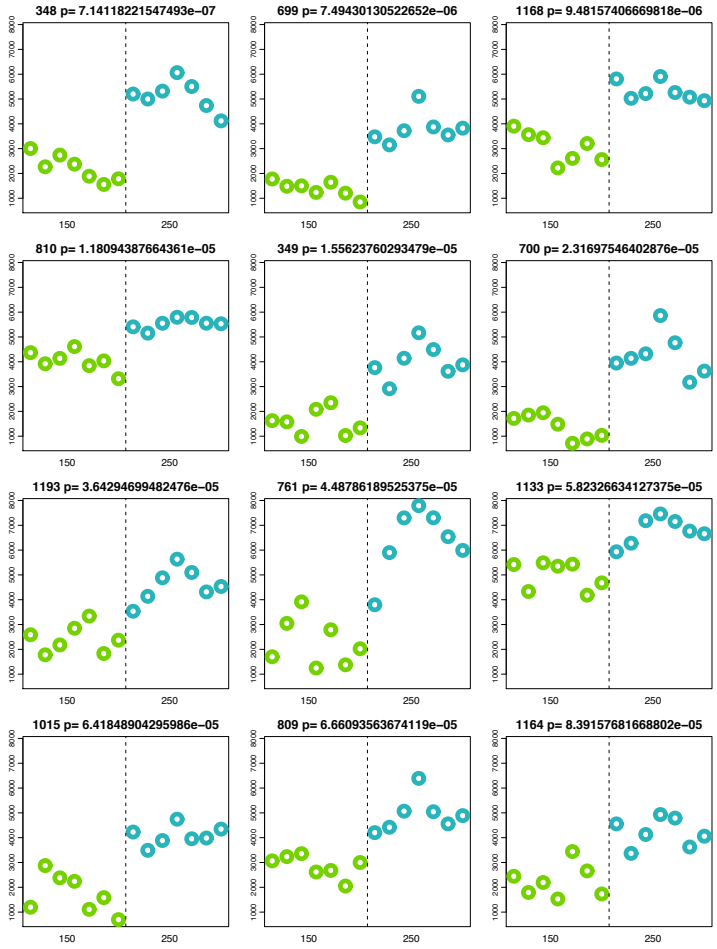


FIGURE 5-3

OBSERVATIONS

1. All of the top 12 selected OTUs belong to one of 2 phyla: Firmicutes (8) and Proteobacteria(4).
2. All 12 selected OTUs display a significant increase in 250 modules.

TAXONOMIC ANNOTATIONS

TABLE 5-1. (AT5_P.TABLE.TXT.ANNOTATED.TXT) ANNOTATIONS OF THE OTUS WITH THE LOWEST P-VALUES.

Taxa ID	kingdom	phylum	class	order	family	genus
348	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	unclassified
699	Bacteria	Firmicutes	unclassified	unclassified	unclassified	unclassified
1168	Bacteria	Firmicutes	Bacilli	Lactobacillales	unclassified	unclassified
810	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	unclassified
349	Bacteria	Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified
700	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	unclassified
1193	Bacteria	Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified
761	Bacteria	Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified
1133	Bacteria	Proteobacteria	Betaproteobacteria	Hydrogenophilales	Hydrogenophilaceae	unclassified
1015	Bacteria	Firmicutes	Clostridia	Clostridiales	unclassified	unclassified
809	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	unclassified
1164	Bacteria	Firmicutes	Clostridia	Clostridiales	unclassified	unclassified

FIGURE 5-4

(./module/circular_tree.pdf) Tree comparing module 250 samples (inner rings) and module 150 samples (outer rings). From the 1125 OTUs present in the study, 279 OTUs (within 69 families) are significantly different (p < 0.05) in one of the comparison groups. The one OTU with the greatest difference between the two group means (250 vs 150 groups) from each family is selected. 13 families contained OTUs with both higher and lower abundance scores in module 250 compared to module 150. In these 13 families both OTUs are selected. A representative 16S rRNA gene from each of the 82 OTUs is aligned and used to infer a phyogenetic tree. The color saturation indicates the degree of difference from the mean module 250 value, where dark blue indicates a ratio of 0.06, white = 1.0 (steady state samples as reference), dark red = 7.5.

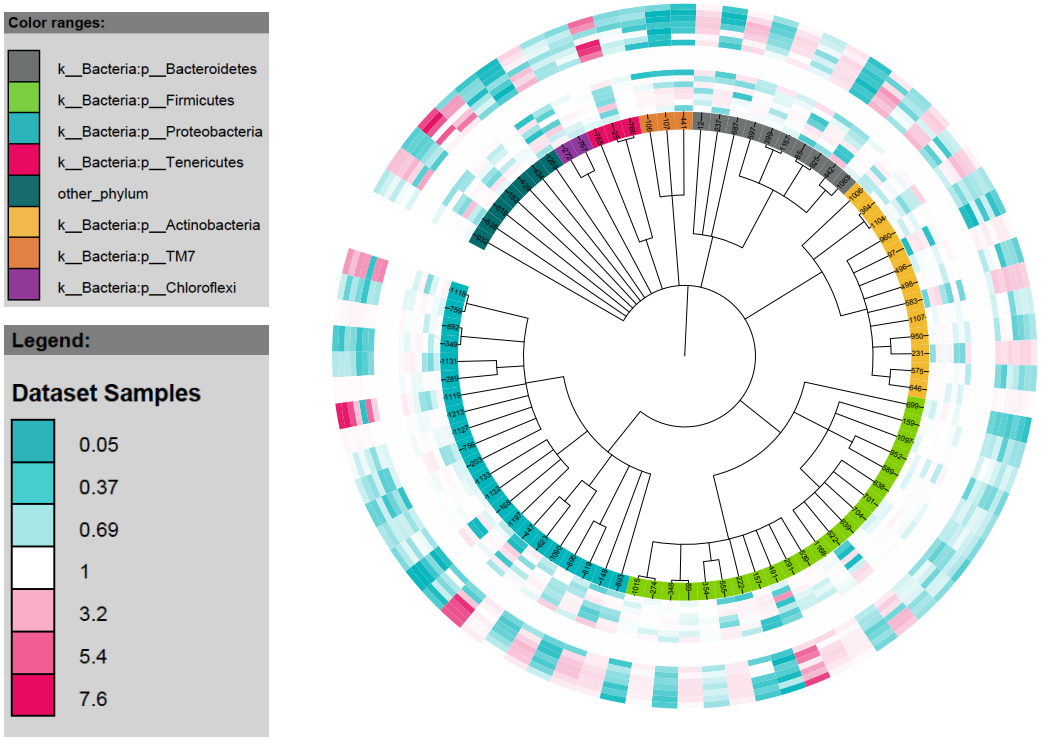


FIGURE 5-4

OBSERVATIONS

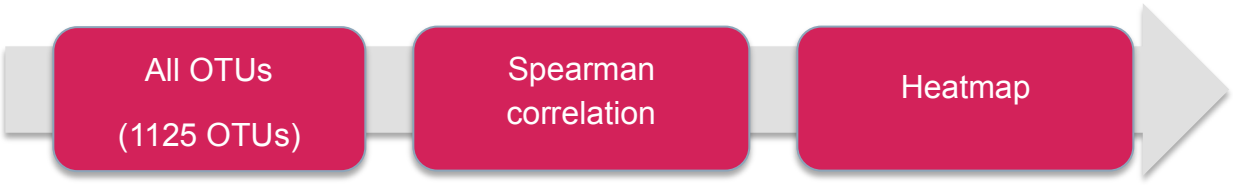
1. Of the OTUs selected, those TM7 group exhibited a decrease in module 150 samples.
2. Selected OTUs of the other phyla exhibited mixed responses.

Chapter 6: Custom Analysis Part A and B – Correlation analysis

In this chapter, we performed a Spearman rank correlation to look for those OTUs that had significant correlations with different metadata factors.

GENERAL OBSERVATIONS

1. 56 taxa are identified to have a significant correlation between OTU abundances and factor time in module 150.
2. 38 taxa are identified to have a significant correlation between OTU abundances and factor time in module 250.
3. Additional correlation analyses have been performed on cfu_surface and cfu_air. Corresponding files and heatmaps are available as supplementary files.



SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 6-1	*./Correlation_Analysis/module150/ at1.module150.txt.annotated.txt_spearman_time_pvals0.05.txt_Heatmap.pdf*
Figure 6-2	*./Correlation_Analysis/module250/ at1.module250.txt.annotated.txt_spearman_time_pvals0.05.txt_Heatmap.pdf*

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./Correlation_Analysis/module150/at1.module150.txt.annotated.txt_spearman_time_pvals0.05.txt	*List of taxa with highly significant correlations including p-values, correlation values and OTU trajectories.*
./Correlation_Analysis/module250/at1.module250.txt.annotated.txt_spearman_time_pvals0.05.txt	*List of taxa with highly significant correlations including p-values, correlation values and OTU trajectories.*

FIGURE 6-1

(at1.module150.txt.annotated.txt_spearman_time_pvals0.05.txt_Heatmap.pdf) Heatmap of OTUs that showed a significant correlation (p<0.05) with the factor time in module 150. The OTUs are ordered by positive and negative correlation and by p-value in an increasing manner.

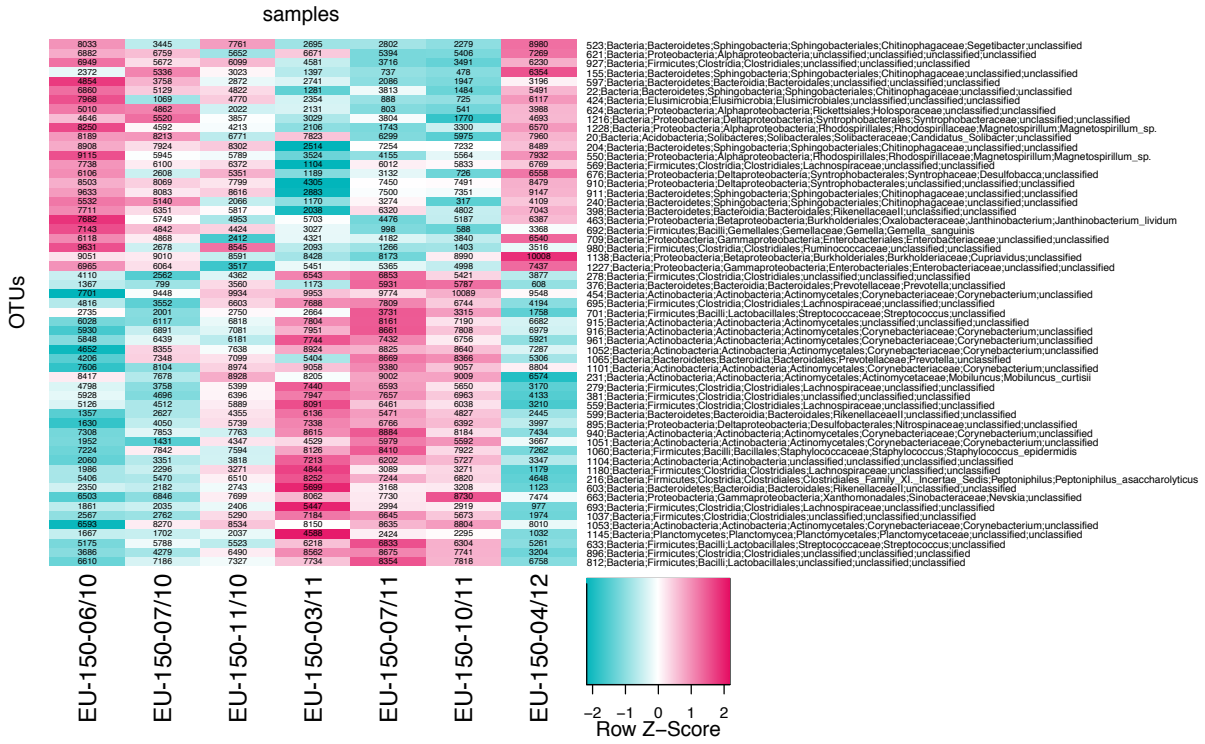


FIGURE 6-1

OBSERVATIONS

1. 57 different OTUs are found to correlate significantly with the factor time.
2. Many taxa unclassified at species level are identified.

FIGURE 6-2

(at1.module250.txt.annotated.txt_spearman_time_pvals0.05.txt_Heatmap.pdf) Heatmap of OTUs that showed a significant correlation (p<0.05) with the factor time in module 150. The OTUs are ordered by positive and negative correlation and by p-value in an increasing manner.

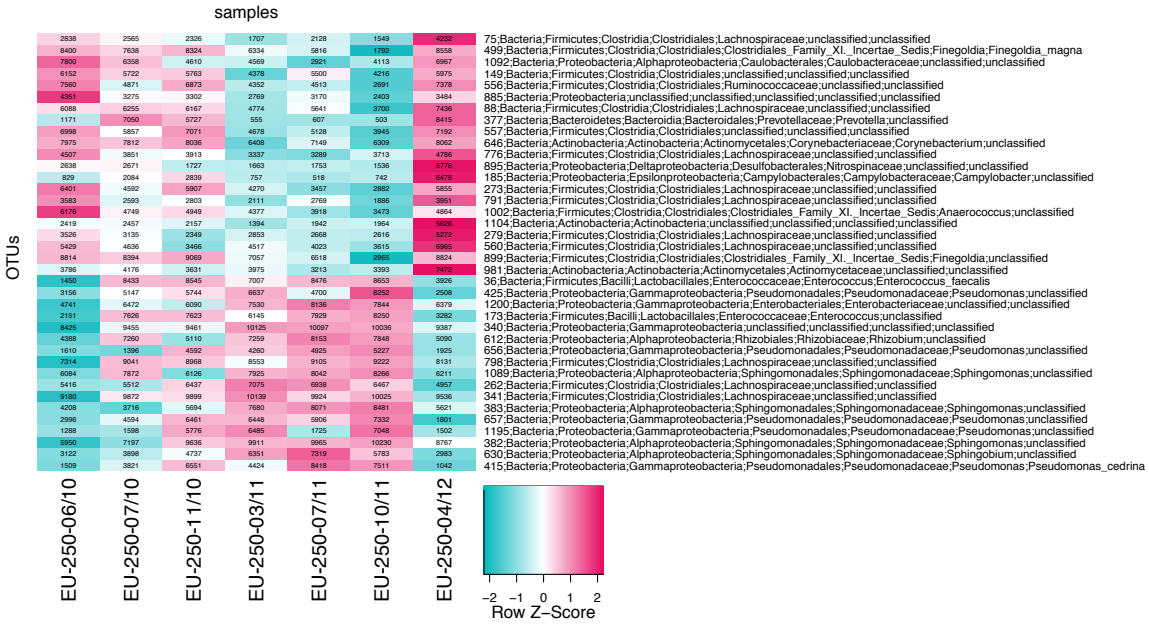


FIGURE 6-2

OBSERVATIONS

1. 38 taxa with significant abundance correlations with factor time are identified.
2. Among many unclassified taxa at species level, *Enterococcus faecalis* is identified.

TAXONOMIC ANNOTATIONS

TABLE 6-1. NUMBER OF OTUS CORRELATING SIGNIFICANTLY WITH DIFFERENT METADATA FACTORS.

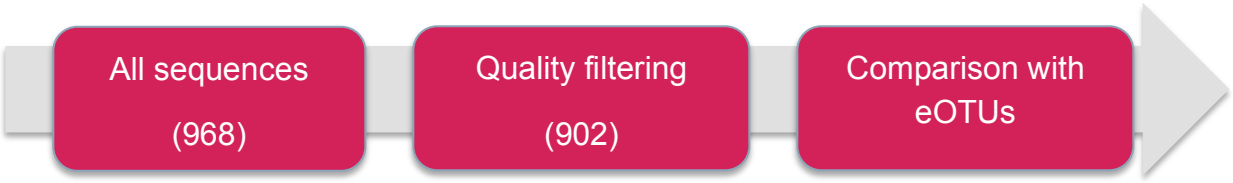
Module	Factor	Number of OTUs
150	Time	56
150	CFU_surface	50
150	CFU_air	40
250	Time	37
250	CFU_surface	32
250	CFU_air	59

Chapter 7: Custom Analysis Part C and D – Integrating isolate sequences in analysis

In this chapter, we quality checked and classified all 16S rRNA genes of isolates provided by the client. Taxonomic affiliations are compared to PhyloChip Assay retrieved taxa. Abundance-based analysis of isolate data is performed using the time points capturedin the PhyloChip assay and from non-heat-shocked samples only (airfilter and swabs).

GENERAL OBSERVATIONS

1. 902 of 968 isolate sequences pass the quality filtering.
2. The PhyloChip Assay complements the culturing techniques on every taxonomic level.
3. Divergence of the two techniques (PhyloChip Assay and Culturing) is observed on species level.
4. Correlation of isolation abundance and aggregated HybScores (PhyloChip Assay) on genus level reveals 3 of 21 genera that correlated significantly.



SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 7-1	*./Isolates/IsolationVSG3.pdf*
Figure 7-2	*./Isolates/IsolatesVSeOTUs.pdf*
Figure 7-3	*./Isolates/Isolates_aggregated_Species_richness.pdf*
Figure 7-4	*./Isolates/Isolates_aggregated_Species.pdf*
Figure 7-5	*./Isolates/abundance_correlation_genus.pdf*

RESULTS

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./Isolates/seqsIDs_for_abundance.txt	*List of isolate IDs that are used in the abundance based comparison*
./Isolates/Isolates_aggregated_Species.csv	*List of species abundance based on isolation frequency in abundance based analysis.*
./Isolates/Isolates_aggregated_Genus.csv	*List of genera abundance based on isolation frequency in abundance based analysis.*
./Isolates/Genus_sumHyb_G3.csv	*Aggregated HybScores of Genera detected in PhyloChip analysis (used for correlation with genus richness from isolation).*
./Isolates/abundance_correlation_genus.txt	*Correlation of aggregated HybScores of Genera detected in PhyloChip analysis with abundance data from isolation strategy.*

FIGURE 7-1

(IsolatesVSG3.pdf) Barchart displaying the number of classified taxa per taxonomic level that could be covered by each method. This analysis included the entire isolation data of various timepoints.

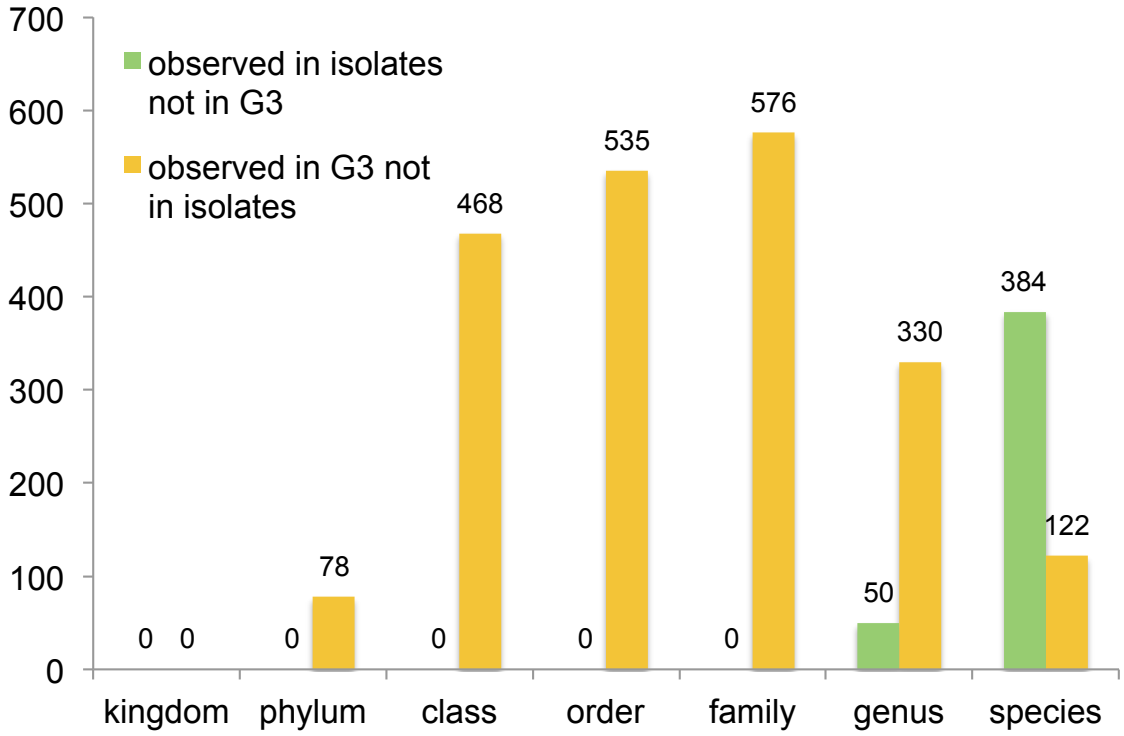


FIGURE 7-1

OBSERVATIONS

1. From kingdom to family level, PhyloChip analysis detects the entire spectrum of isolates.
2. 50 isolate sequences are not detected at genus level, and 384 of 902 on the species level.
3. PhyloChip G3 analysis revealed greater diversity at any taxonomic level.
4. Please note that unclassified taxonomic levels are not considered in this analysis. Most of the eOTUs detected in the PhyloChip Assay are not classified at species level.
5. Please note that every isolate contributes to this analysis regardless of whether taxonomic affiliation is covered multiple times (e.g. *Staphylococcus heamolyticus* is found more than 100 times among the isolates).

FIGURE 7-2

(IsolatesVSeOTUs.pdf) Barchart displaying the number of classified taxonomic levels that were covered by each method. This analysis included the entire isolation data of various timepoints.

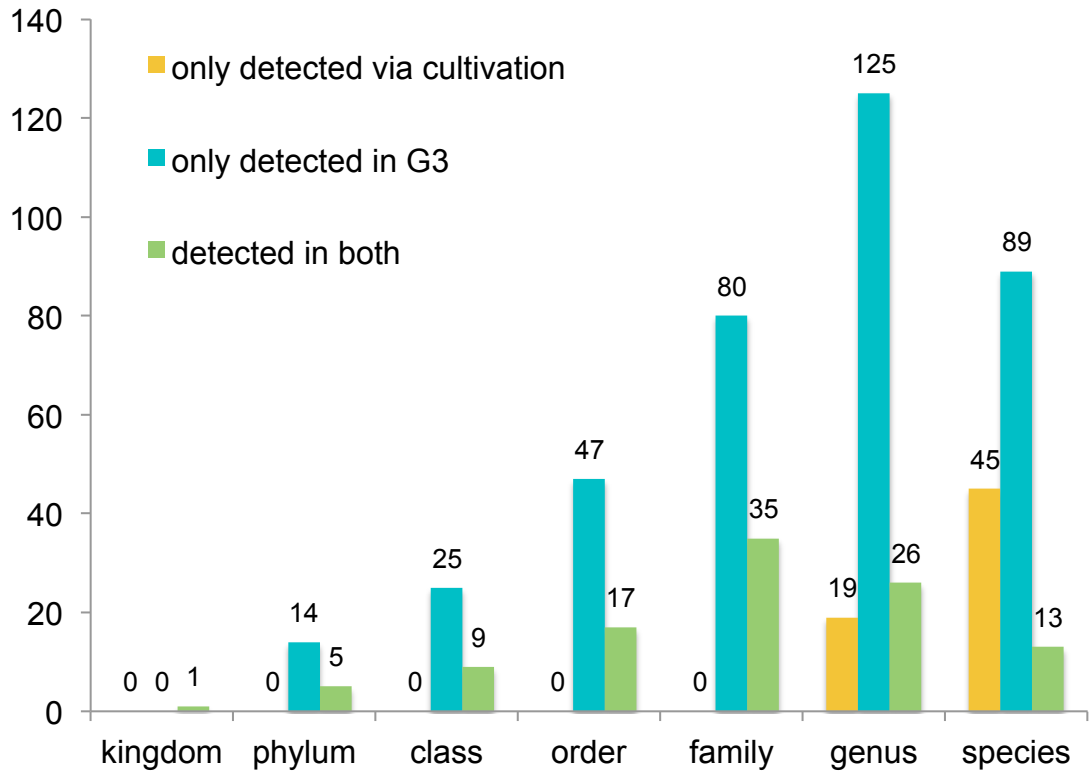


FIGURE 7-2

OBSERVATIONS

1. PhyloChip analysis reveals a greater proportion of classified taxonomic levels than isolation.
2. The divergence of the two profiling methods increases with taxonomic resolution.
3. Please note that unclassified taxa at each level are not considered.

RESULTS

SUPPLEMENTARY FILES

TABLE 7-1. SUPPORTING FILES FOR SEQUENCE QUALITY ANALYSIS.

File	Quality check
seq.fasta	All sequences as submitted by the client
seqs_trimmed.fasta	Sequences after alignment and homopolymer trimming
seqs_trimmed.tax	Taxonomic classification of seqs_trimmed.fasta
seqs_removed.fasta	Sequences that are removed due to length or could not be classified
seqs_qual_filtered.fasta	High quality sequences used for analysis
seqs_qual_filtered.ids	IDs of high quality sequences
seqs_qual_filtered.tax	Taxonomic classification of high quality sequences

FIGURE 7-3

(Isolates_aggregated_Species_richness.pdf) Barchart displaying the species richness retrieved by isolation from selected time points. These time points are congruent with those covered in PhyloChip Assay and include only samples from air filters and samples from swabs without heat-shock treatment.

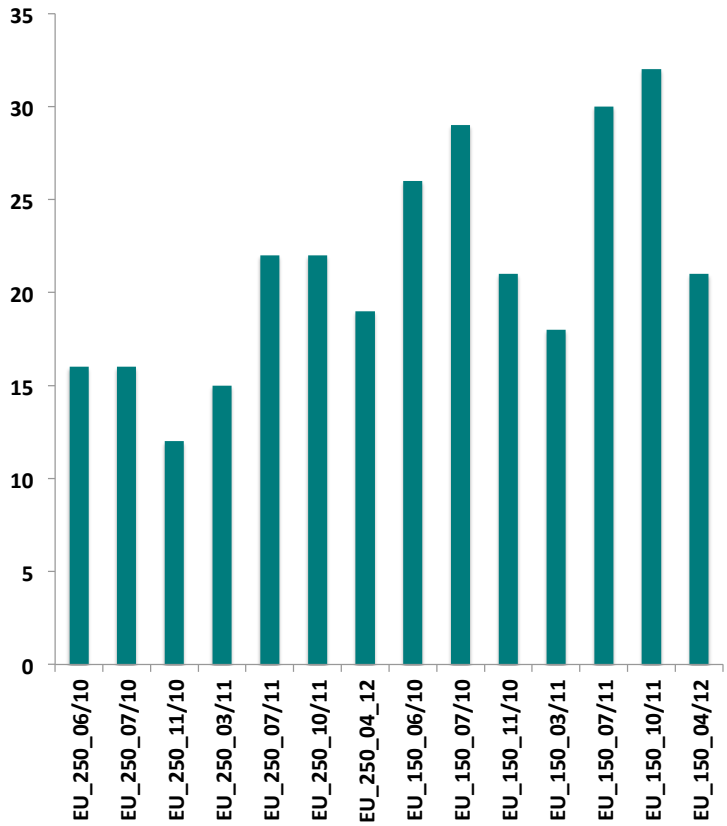


FIGURE 7-3

OBSERVATIONS

1. Species richness of isolates ranges between 12 and 32.
2. Spearman rank correlation of number of different isolates with number of different eOTUs reveals a highly significant negative correlation at Species level (p=0.026, rho=-0.59).
3. Considering a Spearman correlation at Genus level, no significant correlation is found concerning the richness detected via eOTUs and via Isolation (p=0.054, rho=-0.52).

FIGURE 7-4

(Isolates_aggregated_Species.pdf) Comparison of species-level proportional abundance of Isolates across selected timepoints. The bar chart displays the 9 species with the largest number of isolations.

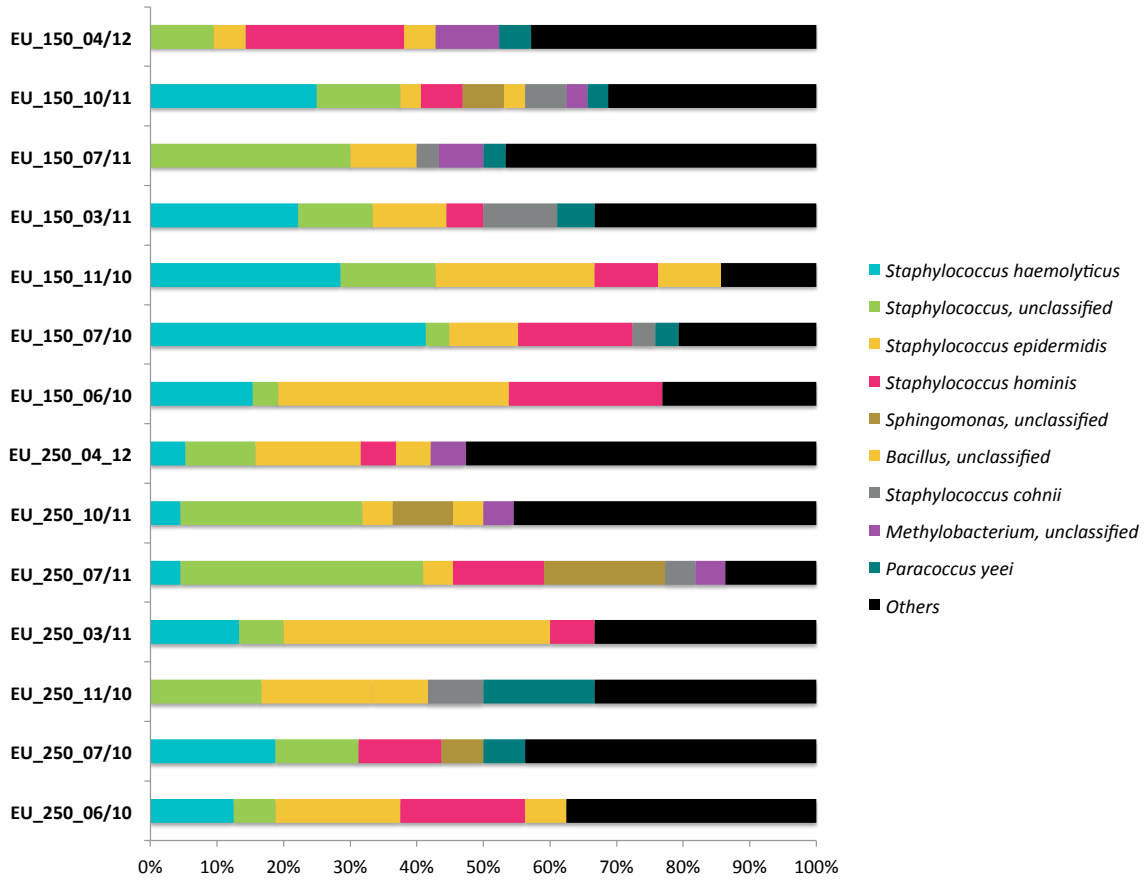


FIGURE 7-4

OBSERVATIONS

1. The most abundant species retrieved is *Staphylococcus haemolyticus*.
2. The composition of the samples based on aggregated Species richness is found to be very heterogeneous.

FIGURE 7-5

(abundance_correlation_genus.pdf) Correlation of aggregated HybScores retrieved from PhyloChip analysis with isolate counts on Genus level (21 genera included). Genera with significant correlation are labeled with an asterisk. Only positive correlations are displayed.

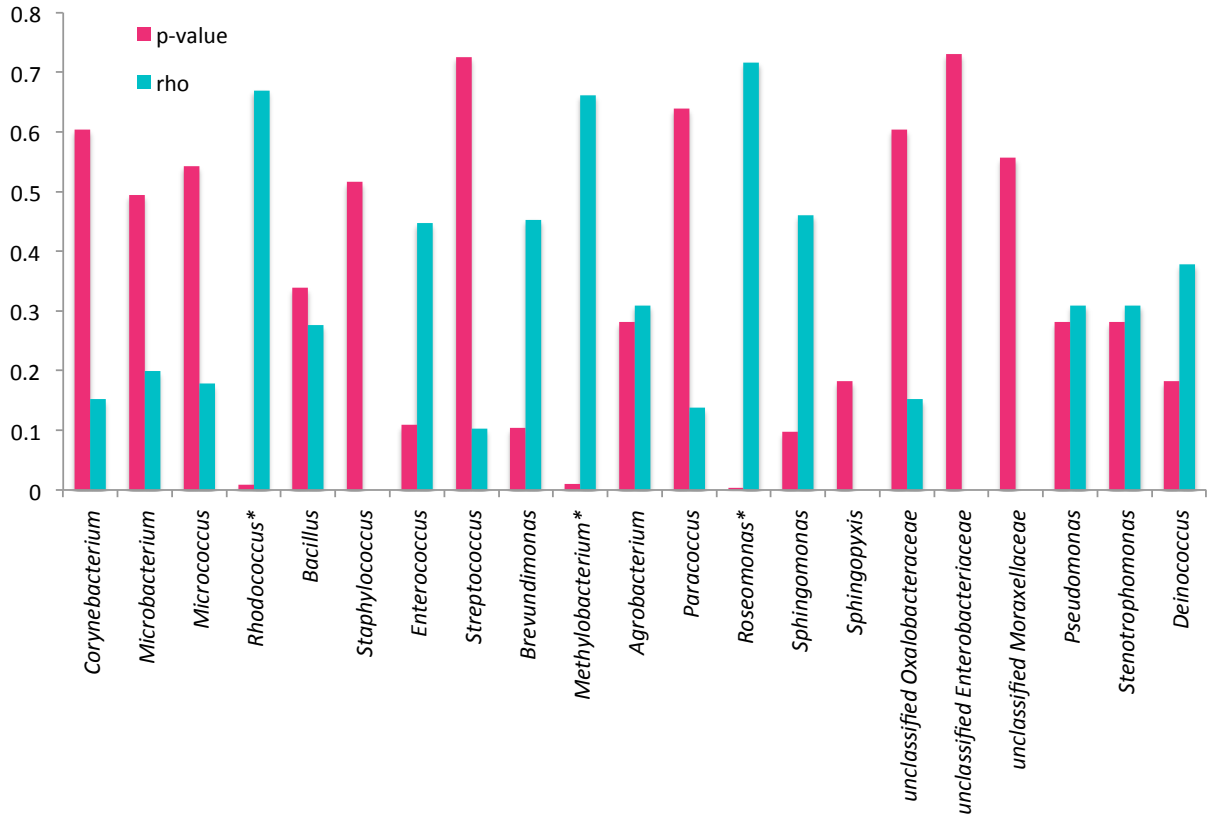


FIGURE 7-5

OBSERVATIONS

1. Isolation abundance of three genera is found to have a significant correlation with aggregated HybScores of the same genus in PhyloChip analysis.
2. PhyloChip analysis reveals 255 different genera to be present in samples from swabs taken at the selected timepoints. Isolation strategies reveal 31 genera from swab and airfilter samples.
3. 21 of the 255 genera detected via PhyloChip analysis are also retrieved in cultivation.

Chapter 8: Custom Analysis Part E – Detection and analysis of Potential Pathogenic OTUs (PPO)

In this chapter, we compared the taxonomic characterization PhyloChip Assay retrieved eOTUs with a reference catalogue of pathogenic organisms provided by the client.

GENERAL OBSERVATIONS

1. The number of pathogenic species is significantly higher in module 250 than module 150.
2. Significant differences in ordination analysis are found between module 150 and module 250.
3. Six PPOs are identified that correlated significantly with time.

SUPPORTING FIGURES

NAME	HIGH-RESOLUTION IMAGE LOCATION
Figure 8-1	*./PPO/species.richness.pdf (+.ps)*
Figure 8-2	*./PPO/species.abd.barchart.pdf (+.ps)*
Figure 8-3	*./PPO/at1.bray.NMDS.Module.pdf (+.ps)*
Figure 8-4	*./PPO/at1.bray.HCAN.pdf (+.ps)*
Figure 8-5	*./PPO/bt1.bray.NMDS.Module.pdf (+.ps)*
Figure 8-6	*./PPO/bt1.bray.HCAN.pdf (+.ps)*
Figure 8-7	*./PPO/PPO_spearman_correlation_time.pdf*

SUPPORTING DATA TABLES

NAME	DESCRIPTION
./Community_Characterization/species.bt1.bacteria.richness.table.txt	*Taxon richness at the species level per PPO.*
./Community_Characterization/species.abd.barchart.pdf	*Proportions of aggregated hybscores per eOTUs classified at the species rank for top 9 species (considering only PPO).*

FIGURE 8-1

(species.richness.ps) Barchart displaying the number of species of PPOs per sample.

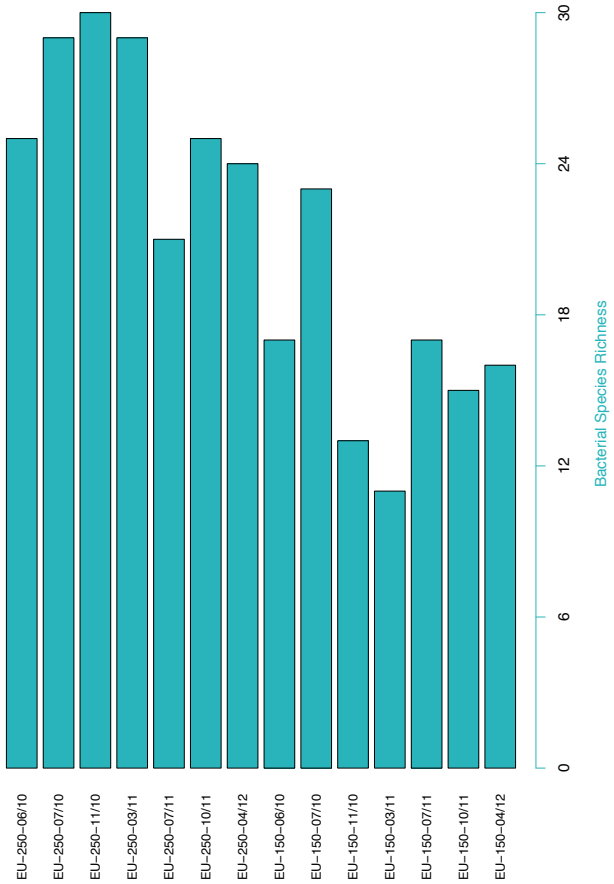


FIGURE 8-1

OBSERVATIONS

1. The number of potential pathogenic taxa at species level per sample varies from 11 to 30.
2. Using a heteroscedastic t-test, a highly significant difference between PPO species richness of module 150 and module 250 is observed (p-value<0.0002).

FIGURE 8-2

(species.abd.barchart.ps) Comparison of species-level proportional abundance across samples considering PPOs only. The bar chart displays the 9 species with the largest HybScores found by summing the HybScores from the OTUs within the families.

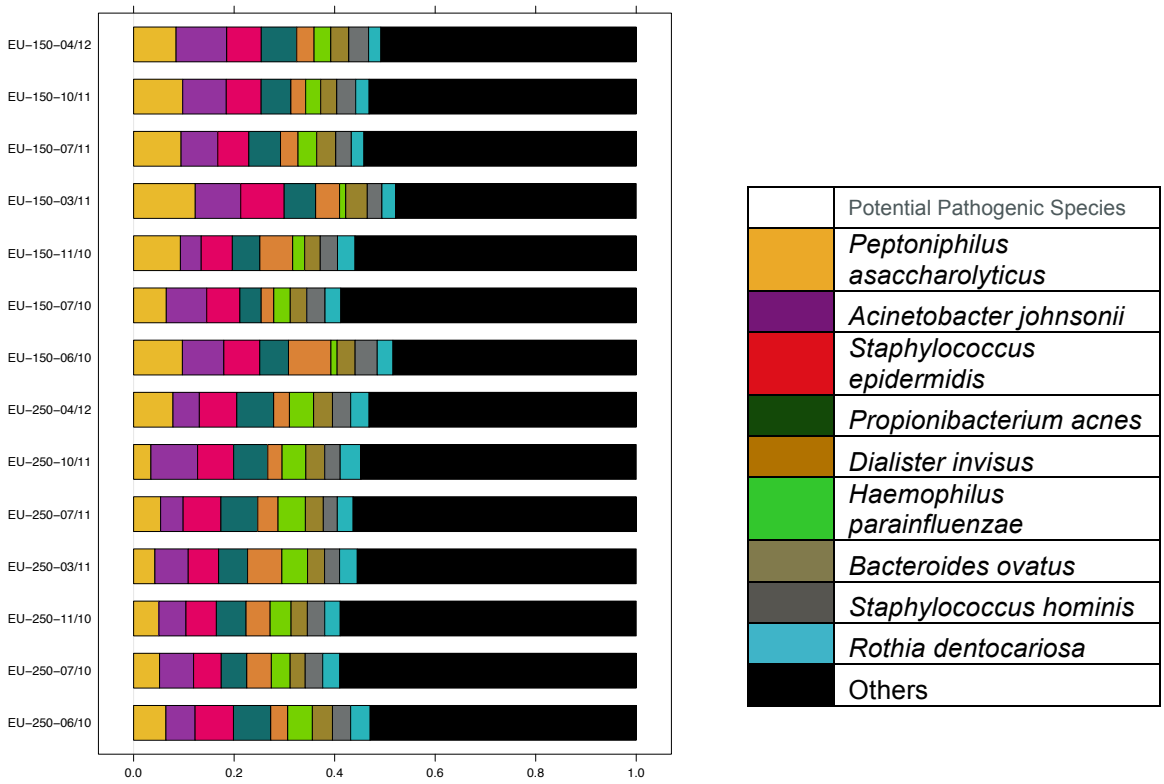


FIGURE 8-2

OBSERVATIONS

1. The highest proportional aggregated hybscore is found for *Peptoniphilus asaccharolyticus*.
2. In general, the variation of the aggregated hybscores per potential pathogenic species is not homogenous.

FIGURE 8-3

(at1.bray.NMDS.Module.ps) NMDS based on Bray-Curtis distance between samples given abundance of 81 PPOs present in at least one sample. Stress=0.1253.

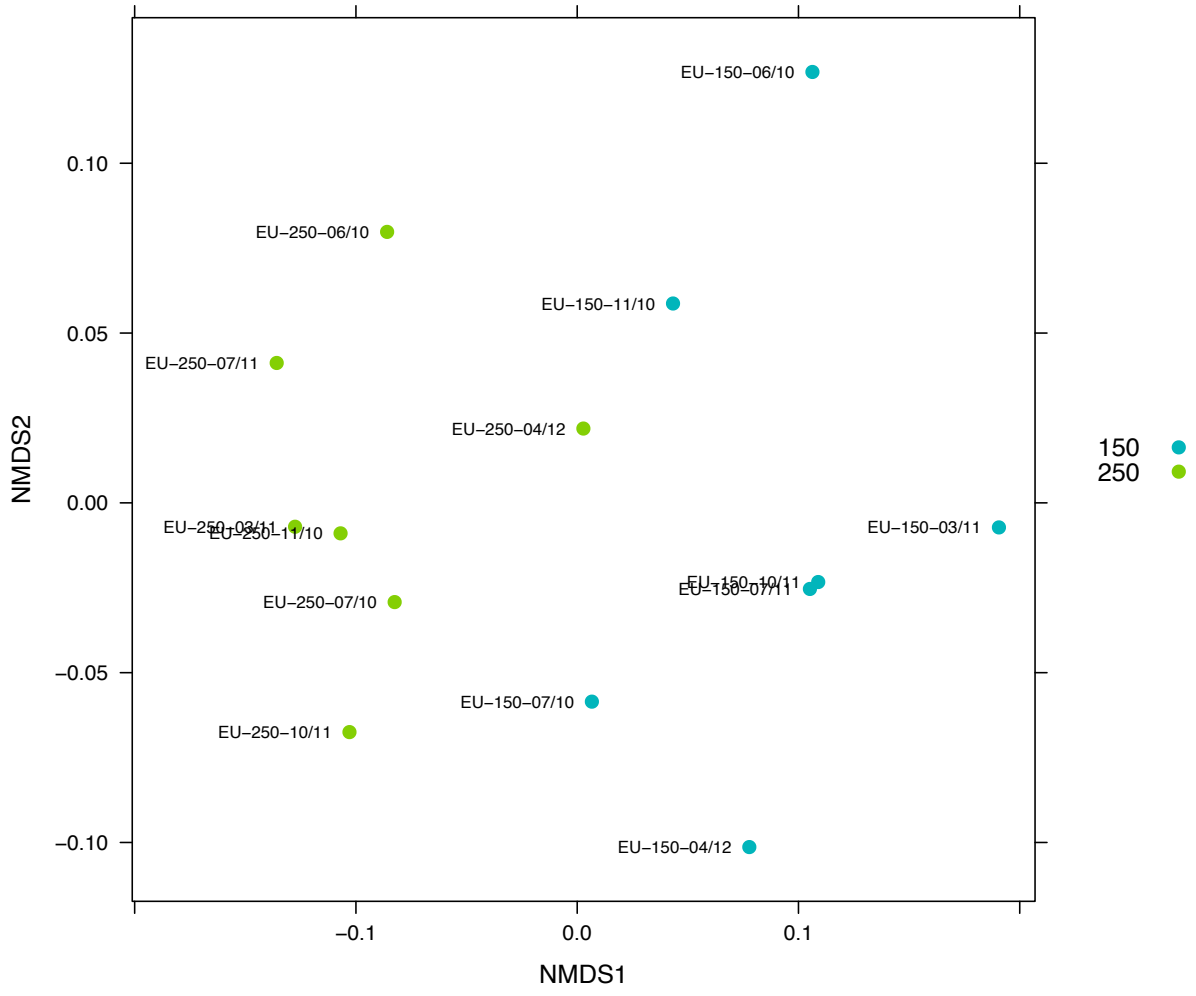


FIGURE 8-3

OBSERVATIONS

1. A trend of separation of the two module microbiomes is observed along NMDS1 axis.

FIGURE 8-4

(at1.bray.HCAN.ps) Hierarchical Clustering (average linkage) based on Bray-Curtis distance between samples given abundance of 81 PPOs present in at least one sample.

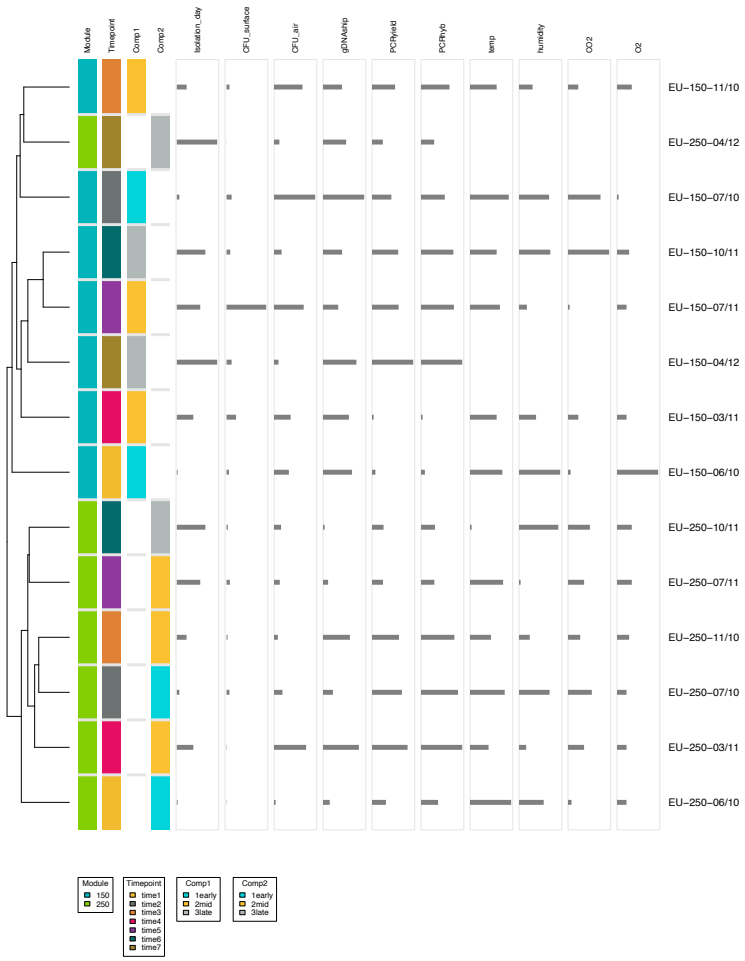


FIGURE 8-4

OBSERVATIONS

1. Two separate clusters of samples are observed, one of them consisting of samples from module 250 only.

FIGURE 8-5

(bt1.bray.NMDS.Module.ps) NMDS based on Bray-Curtis distance between samples given presence/absence of 81 PPOs present in at least one sample. Stress=0.1542.

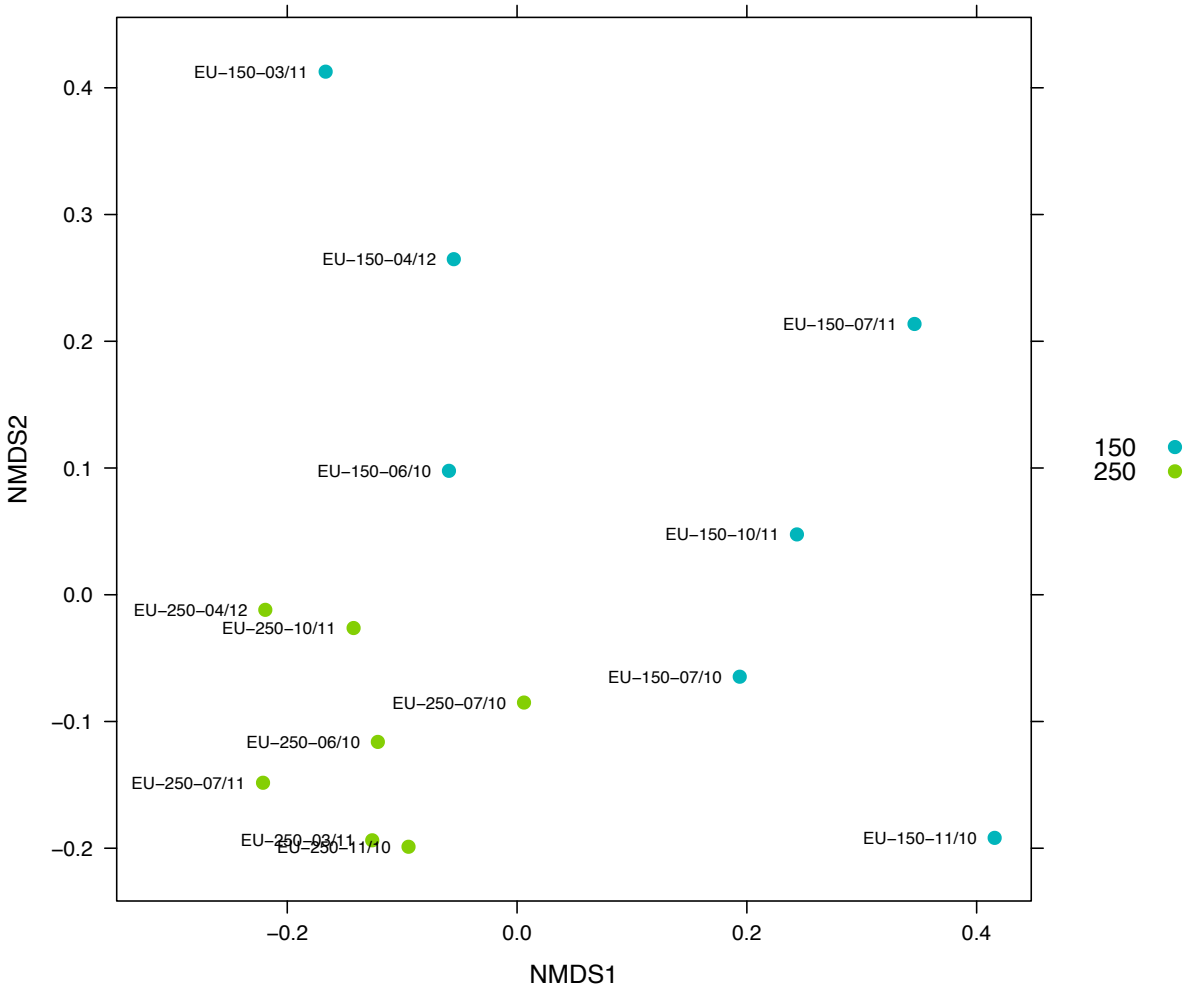


FIGURE 8-5

OBSERVATIONS

1. Two separate groups of samples are observed based on module category.
2. Samples from module 250 exhibit lower intra-group dissimilarities than samples from module 150.

FIGURE 8-6

(bt1.bray.HCAN) Hierarchical Clustering (average linkage) based on Bray-Curtis distance between samples given presence/absence of 81 PPOs present in at least one sample.

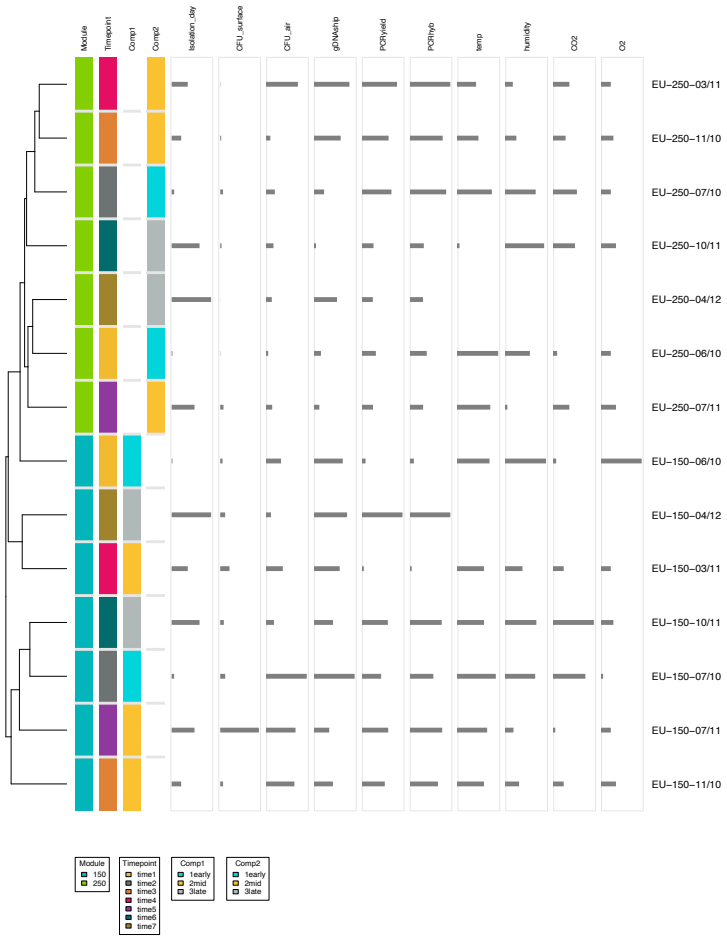


FIGURE 8-6

OBSERVATIONS

1. No entire separation of the microbiomes is observed in HC-AN analysis considering the incidence values of the PPOs.
2. One cluster is consisted only of module 250 samples and sample EU-150-06/1.

FIGURE 8-7

(PPO_spearman_correlation_time.pdf) Heatmap of PPOs that showed a significant correlation (p<0.05) with the factor time in module 150 and 250. White horizontal line separates module 150 PPOs from module 250 PPOs.

FIGURE 8-7

OBSERVATIONS

1. Six PPOs are identified to have a significant correlation with time, 4 in module 150, 2 in module 250.
2. Based on relative abundance, 4 of 6 of these six PPOs correlate positively with time.

APPENDIX

Glossary of Terms

Adonis test	The Adonis test is utilized for finding significant differences in the whole microbiome among discrete categorical or continuous variables. The samples are randomly reassigned to the various sample categories, and the fraction of permutations with larger cross-category differences relative to within-category differences is reported as the p-value for the Adonis test.
Bray-Curtis Distance	Bray-Curtis Distance is a statistic used to quantify the compositional dissimilarity between two different communities. The Bray–Curtis dissimilarity considers differences in the abundance of a species or OTU across two communities.
Diversity	The degree or amount of variation of microbial community structure within a sample, referred to an alpha-diversity, or between samples referred to beta-diversity. Many metrics of diversity have been proposed and can be derived from tables of relative abundance and/or incidence
Filter	A PhyCA-Stats filter selects important taxa for further consideration. Many employ an appropriate statistical test (i.e. Welch test, paired t-test, etc.) to highlight those taxa with changes in abundance or incidence across experimental groups.
HC-AN	HC-AN is a hierarchical clustering technique using the Average-Neighbor method; it graphically summarizes the inter-sample relationships in the form of a dendrogram. Biologically similar communities have a shorter branch length between them.
HybScore	The Hybridization Score (HybScore) is the abundance metric used to compare relative changes in a taxon's population across samples. It is derived from the background-subtracted fluorescence intensity of the multiple probes within a probe set. The details of the calculation are described in the Methods section.
Incidence	In ecology, incidence is the frequency at which a taxon is found in samples. The incidence table contains binary values of 1's and 0's, to denote presence and absence, respectively, of each taxon in each sample. It is sometimes referred to as "species richness". (Also see Relative Abundance.)
iTOL	The Interactive Tree of Life visualizes the changes in each OTU's relative abundance while displaying the phylogenetic relationships among those OTUs via a phylogenetic tree.

APPENDIX

Relative abundance Relative abundance in alpha-diversity analysis refers to how common or rare an OTU is relative to other OTUs in the same community. In beta-diversity analysis, relative abundance is useful for monitoring changes in a taxon's population between samples. For PhyloChip analysis, the abundance scores, or "HybScores", are used for beta-diversity analysis. It is sometimes referred to as "species evenness". (Also see Incidence.)

Richness Richness refers to the number of OTUs present in a community, and is one of the common components used to measure the biodiversity of an ecosystem. A larger number of samples within an experiment or a larger number of molecules sequenced or hybridized from any one sample may reveal higher richness because the rare species are more likely to be observed. However, richness should plateau as sampling increases.

APPENDIX

References

1. Deep-sea oil plume enriches indigenous oil-degrading bacteria. Hazen T.C. et al., Science. 2010 Oct 8;330(6001): 204-8. (First publication using the G3 PhyloChip assay)
2. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. DeSantis, T.Z., et al., Microb Ecol, 2007 53(3):371-83.
3. High-density microarray of small-subunit ribosomal DNA probes. Wilson, K.H., et al., Appl Environ Microbiol, 2002 68(5):2535-41.
4. Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. Letunic I. and Bork P., Bioinformatics, 2007 23(1):127-8.
5. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. DeSantis, T.Z., et al., FEMS Microbiol Lett, 2005 245(2): 271-8.
6. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Benjamini and Hochberg (1995) *J Roy Statist Soc Ser B (Methodological) 57:289-300.*