

SHAMAN : a **SH**iny Application for Metagenomic ANalysis

Amine Ghozlane^{1,2*}, Stevenn Volant^{1*}, Hugo Varet^{1,2}, Christophe Malabat¹, Pierre Lechat¹, Sean Kennedy², Marie-Agnès Dillies^{1,2}

¹Institut Pasteur – Bioinformatics and Biostatistics Hub – C3BI, USR 3756 IP CNRS – Paris, France

²Institut Pasteur – Biomics – CITECH – Paris, France

^{*}Equally contributing authors

Background

- Quantitative metagenomics is an approach broadly employed to identify associations between a microbiome and an environmental / **individual** condition (disease, geographical condition, ...).
- To perform this type of approach, targeted sequencing of rDNA or shotgun sequencing is performed and quantitative measures are obtained by mapping the reads against the set of OTU identified or a gene catalog.
- These data can be analyzed by developing R scripts including statistical analysis (metagenomeseq, momr, edgeR, ...) or web interface dedicated to visualization (MEGAN, Shiny-phyloseq, Phinch).
- The lack of easy-access methods that ****** providing both relevant statistical analysis and specific visualization is a critical issue.
- Here we present SHAMAN, a Shiny-based application that offers an unified experience for the analysis of quantitative metagenomics data.



Contact: shaman@pasteur.fr

Start with SHAMAN



MBMA

SHAMAN requires as input of (1) a count table and (2) an annotation table (as csv or tsv file) or a BIOM file.

These data are provided by most pipelines like: MASQUE (docker: aghozlane/masque) for targeted metagenomics,

shotgun

- metagenomics (https://github.com/anitaannamale/MBMA).
- Imported dataset is analysed to identify which taxonomical level is



Quantitative Metagenomics (larger)	Heatmap	Sep 9th 2016
at OTU level Generalized Merge normalized counts at the user selected level		
Sufisical Analysis: DF See	Rarefaction	
Differential Analy ⁴ Low MJ, Haler W, Anders S. Mediestel estimation of Skil charge an		

SHAMAN is freely accessible through a web interface at http://shaman.c3bi.pasteur.fr/ and docker hub at aghozlane/shaman.



for

- the appropriate for the most analysis.
- Here 40% of the OTU are annotated at the Genus level and 38 different genera identified.

Experimental design / Statistical modeling

An experimental design table must be provided. The table is used to assign each sample to a condition, a time, an individual or an other metadata.

SHAMAN process is divided into two steps:

Normalization: The OTU/gene count is normalized using size factors defined as the median of the ratio between the count and the geometric mean of each OTU/gene (1) [Anders 2010].

$$s_j = median_i \frac{c_{ij}}{(\prod_{k \in S_i} c_{ik})^{1/n}}$$
(1)

Assume that $C = (c_{ij})_{1 \le i \le k; 1 \le j \le n}$ is a count table.

k and n correspond to the number of features (like OTU) and the number of samples, respectively. c_{ii} represents the count of feature *i* in sample *j*. s_i is

Experimental design	-	Options			-		
Select your target file Separator: Select ti level	he taxonomy		Linear moo	lel options			
Browse Target.tsv Tab Upload complete Genu	15 🔹	Type of transformation	Independent filtering	p-value adjustement	Level of significance		
Select the variables Add interactions			● True ○ False	 вн вү 	0.05	(3)	
condition time mice condition:mice con	ndition:time		Local function	Relationship			
			 Median 	 Parametric 			
▶ Run analysis		-	○ Shorth	🔿 Local			
Target file overview	-	🔿 Value	Options for the ne	ormalization step			
Show 10 ${}^{\vee}$ entries Search:							
label 🔶 condition 🔶 time 🗍	mice 🔶	Compute geometric mean without 0	Normalization by:	Define your own size factors	Separator:		
WT.m2 WT.m2 WT t0	m1	mean without 0		Browse No file	Tab 🔻		
WT2.38 WT2.38 WT t2	m1						
WT.m3 WT.m3 WT t0	m2					•	Sid
WT3.39 WT3.39 WT t2	m2	Contrasts (New)		-	Defined contrasts		Sig in
WT.m4 WT.m4 t0	m3	Compare To	For		Contrasts		in
WT4.40 WT4.40 WT t2	m3	Delta 🔻 DeltaCo	mpl All 🕶	+ Add	WT_0_24 ^ DeltaCompl_0_24		
WT5.41 WT5.41 WT t2	m4				Delta_0_24 WT_0_6		~~
WT.m5 WT.m5 WT t0	m4						cc fro
WT2.56 WT2.56 WT t3	m1	Contrasts (advanced use	er)	-	🗙 Remove 🏙 Export		ſ
WT3.57 WT3.57 WT t3	m2	Select a file of contrasts					Tro
Showing 1 to 10 of 36 entries Previous 1 2 3	4 Next	Browse Contrasts.txt					
		Upload complete					va
Count table (Genus)	-	Define contrasts by yourself					
Show 10 v entries Search:		Contrast name	+ Add contrast				
WT.m2 \Rightarrow WT2.38 \Rightarrow WT.m3 \Rightarrow W	/T3.39 🔷 WT.	Intercept					
[Eubacterium] 1054 521 394	228	0		÷			
Acetatifactor 0 4 1	3	conditionDelta					
Acetitomaculum 12 25 8	11	0		×			
Alistipes 2381 1533 1858	2073	conditionDeltaCompl					
Allobaculum 8 7 150	9	0		×			
Alloprevotella 8 0 8	4	conditionWT					
Anaerofustis 1 0 2	0	0		×			
Anaeroplasma 393 275 313	43	timet0					
Anaerotruncus 114 95 52	68	0		•			
Bacteroides 5309 906 3578	9625	timet2					
Showing 1 to 10 of 48 entries	,	0		÷			
Previous 1 2 3 4	5 Next	timet3					
		0		÷			
Letters Letter	2						
	2	micem1		÷			

The user defines a contrast vector to extract features that are significantly different in abundance according to the experimental design. A guided and expert mode are available in SHAMAN to perform this step.

nificant features are summarized in a table icating their **base mean** (mean normalized int), fold change (how much the count varies m one condition to the other) and adjusted pue.

how 25 $$					Search:
Id	$\stackrel{\wedge}{=}$	baseMean 🔶	FC 崇	log2FoldChange 🔷	padj 🍦
Turicibacter		995.24	0.004	-7.899	1.08466957119766e-9
Allobaculum		103.04	800.0	-6.955	0.000031191170815586
Roseburia		173.49	3.965	1.987	0.00152565569842405
Enterorhabdus		38.35	0.156	-2.676	0.00210484632042579
Ruminococcaceae		2369.48	0.324	-1.628	0.00364694347726387
Parasutterella		10.73	0.042	-4.56	0.00747095659486352
Mucispirillum		50.33	4.941	2.305	0.0186705242331407
Prevotellaceae		49.73	0.1	-3.321	0.0198884561087325
Alloprevotella		2.7	0.076	-3.727	0.0236316541392707
Streptococcus		8.73	0.221	-2.175	0.0264348153495083
Listeria		157.45	16.58	4.051	0.034455649855071

- the size factor of sample *j*.
- Modelization: DESeq2 local regression is used to get robust estimation of the OTU dispersion and a Generalized Linear Model is defined [Love 2014].

Visualizations

SHAMAN visualizations fall into three categories:

Diagnostic plots: These plots allow a quality check of the data.





- relevance of the statistical modeling.

Analysis plots: These plots are generated to highlight the differences in abundance identified by differential analysis.





Conclusion / Future work

SHAMAN:

 Combines strong statistical approach with a dynamic visualization interface.

features:

- ✓ Integrates most of the analysis required for publication.
- Functions in real time.
- ✓ Already used in a publication [Quereda et al. PNAS 2016].

