

On the use of interaction terms in GLMs to model complex experimental designs: the example of RNA-Seq data

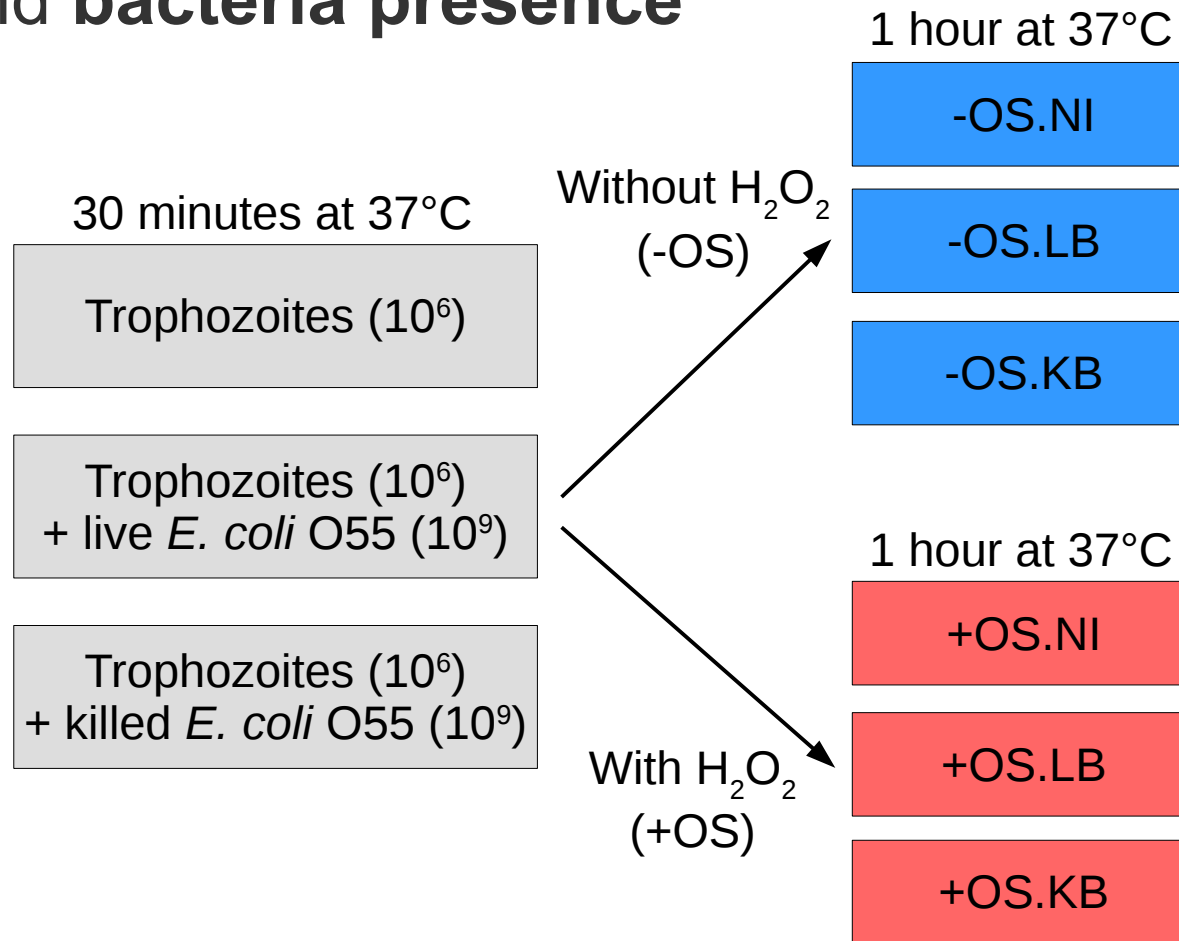
Journées du PEPI IBIS – June 6th & 7th, 2019 – Paris

Hugo Varet – hugo.varet@pasteur.fr

- Hub de Bioinformatique & Biostatistique – Département de Biologie Computationnelle
- Plate-forme Biomics – Centre de Ressources et Recherches Technologiques (C2RT)

RNA-Seq experimental design

Amoeba transcriptome study: impact of both **oxydative stress** and **bacteria presence**



Experiment replicated 3 times to obtain
3 biological replicates → RNA Sequencing of 18 samples

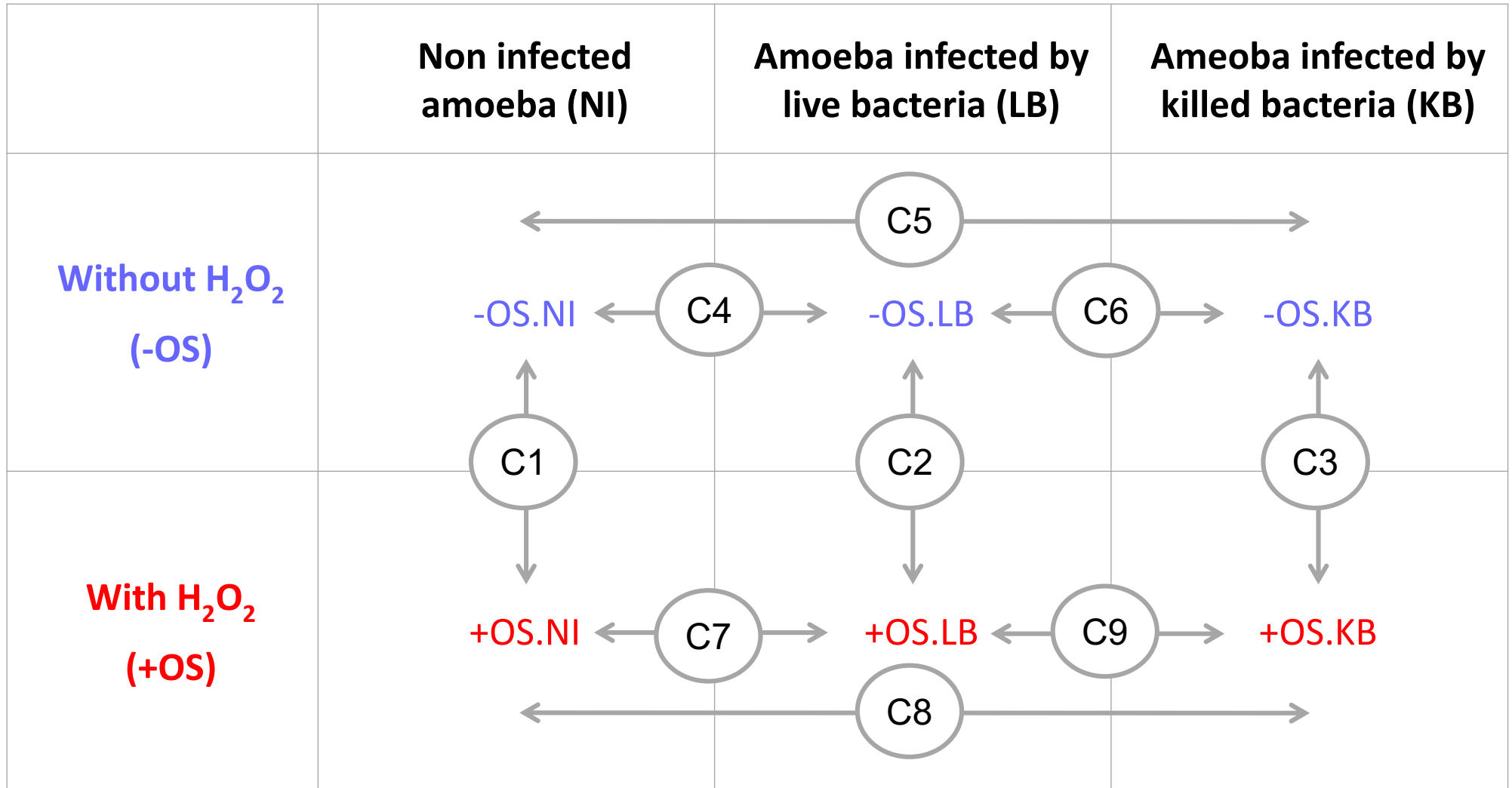
From raw reads to count data

Standard RNA-Seq pipeline for each sample:

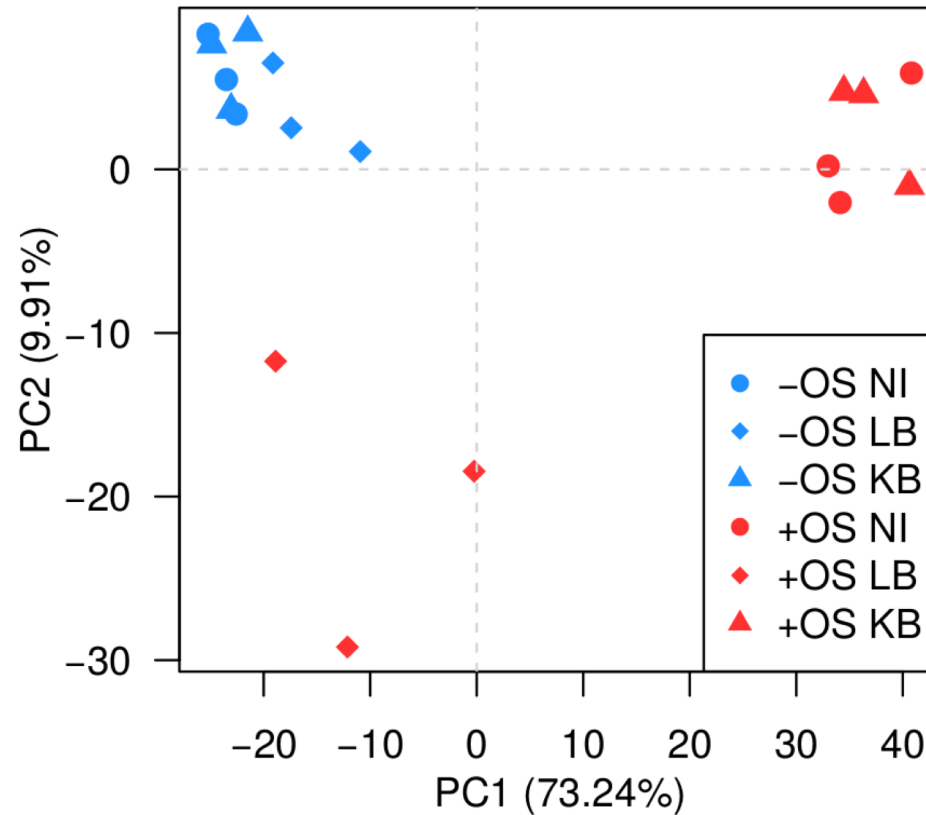
- Cleaning of the sequencing reads (in-house prog.)
- Mapping on the reference genome (tophat)
- Counting at the gene level (HTSeq-count)

	mOS_NI_A	mOS_NI_B	mOS_NI_C	...	pOS_LB_C
EHI_000010	919	992	842	...	738
EHI_000130	5037	5821	6874	...	9151
EHI_000140	270	324	312	...	283
...
EHI_000260	3342	3977	3942	...	6529

Possible comparisons

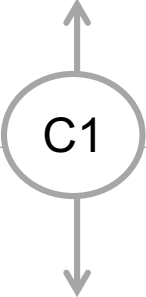
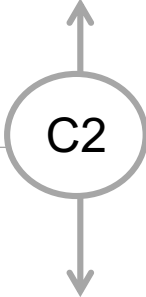


PCA plot



- PC1: stress effect for NI and KB samples
- PC2: stress effect for LB samples

Main biological question

	Non infected amoeba (NI)	Amoeba infected by live bacteria (LB)	Amoeba infected by killed bacteria (KB)
Without H ₂ O ₂ (-OS)	<p>-OS.NI</p>  <p>C1</p>	<p>-OS.LB</p>  <p>C2</p>	<p>-OS.KB</p>
With H ₂ O ₂ (+OS)	<p>+OS.NI</p>	<p>+OS.LB</p>	<p>+OS.KB</p>

Naive approach

Two **independent** statistical analyses using DESeq2 to test for the stress effect **+OS** vs **-OS** within NI and LB samples **separately**:

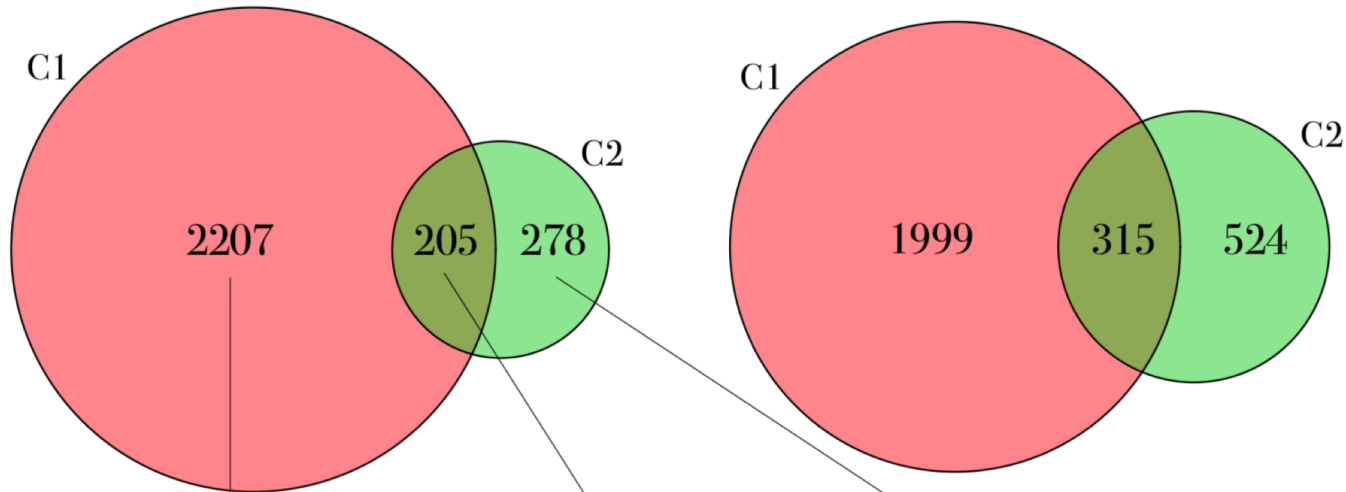
		# down	# up	# total	Interpretation
C1	+OS.NI vs -OS.NI	2412	2314	4726	Strong stress effect
C2	+OS.LB vs -OS.LB	483	839	1322	Moderate stress effect

→ Draw Venn diagrams to compare the results

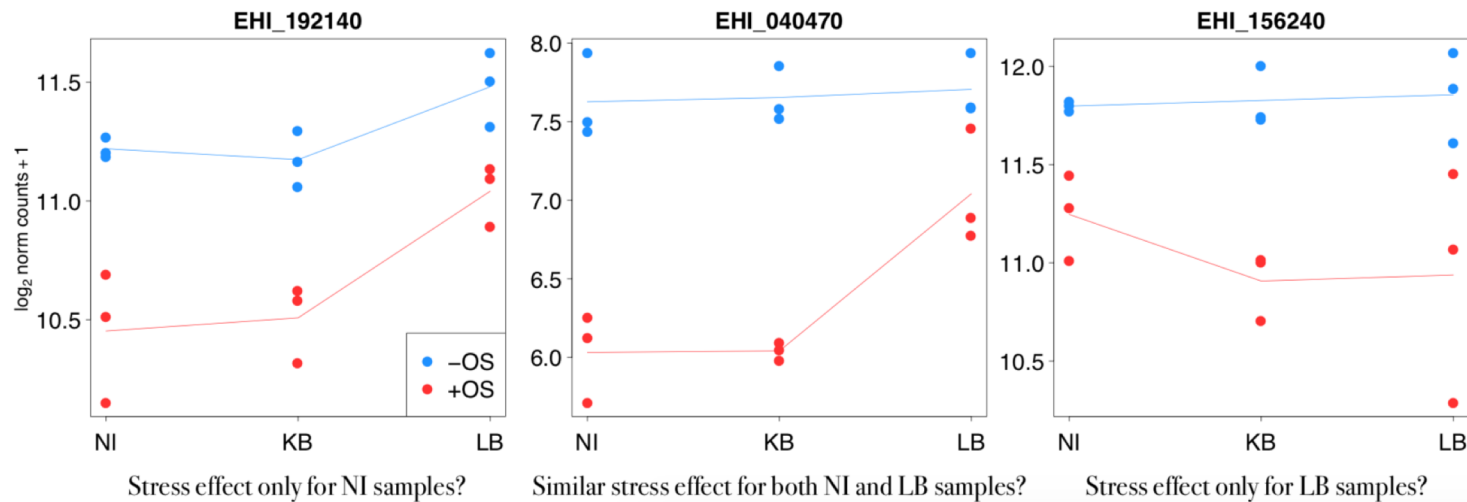
Naive approach

Down-regulated genes

Up-regulated genes



Examples:



Global analysis: design formula

counts ~ stress + bacteria + **stress:bacteria**

Use of specific contrasts to test for C1, C2, ..., C9:

		# down	# up	# total	Interpretation
C1	+OS.NI vs -OS.NI	2412	2314	4726	Strong stress effect
C2	+OS.LB vs -OS.LB	483	839	1322	Moderate stress effect
C3	+OS.KB vs -OS.KB	2346	2277	4623	
C4	-OS.LB vs -OS.NI	35	8	43	Low effect of live <i>E. coli</i>
C5	-OS.KB vs -OS.NI	0	0	0	No effect of killed <i>E. coli</i>
C6	-OS.KB vs -OS.LB	1	2	3	
C7	+OS.LB vs +OS.NI	1977	2201	4178	Strong effect of live <i>E. coli</i>
C8	+OS.KB vs +OS.NI	1	0	1	No effect of killed <i>E. coli</i>
C9	+OS.KB vs +OS.LB	2298	2032	4330	

stress : bacteria interaction

For each gene g , test for the null hypothesis:

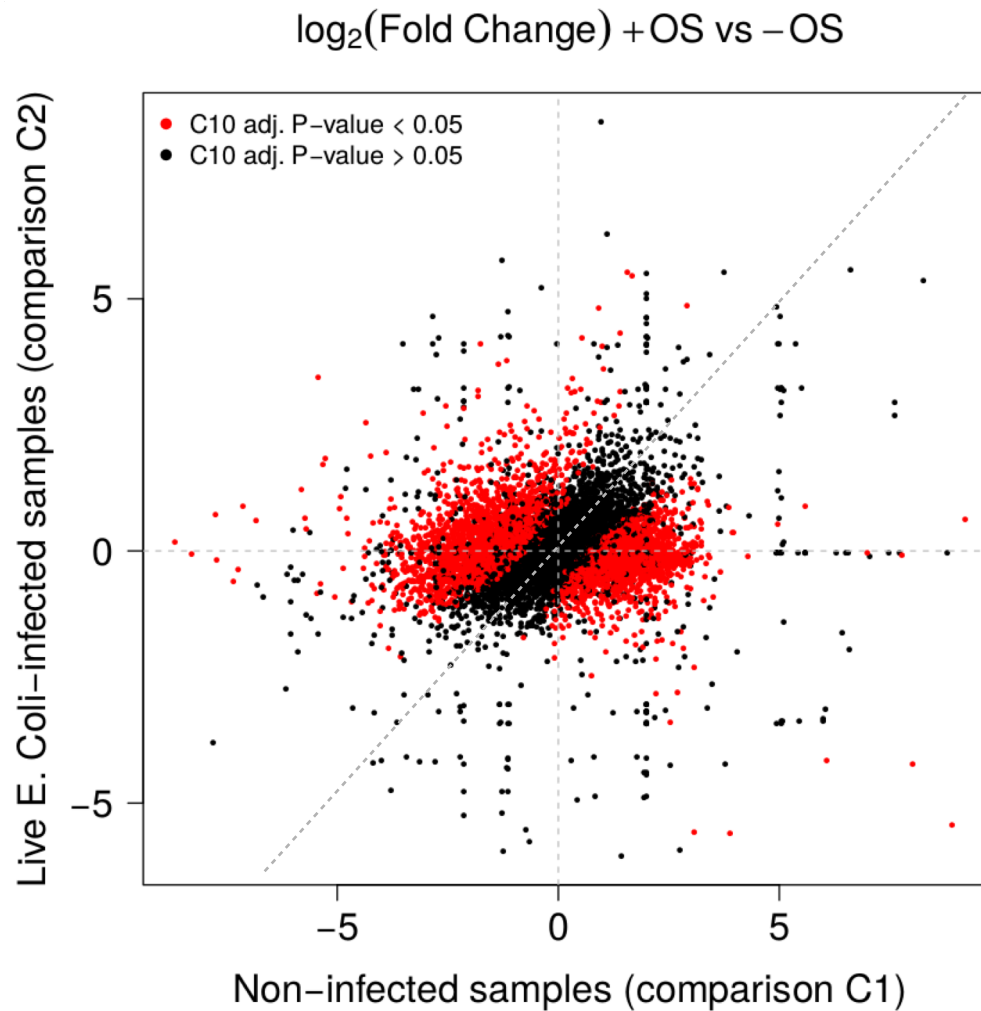
$$H_0: \log_2(\text{FC}_g)_{C1} - \log_2(\text{FC}_g)_{C2} = 0$$

On the \log_2 scale:

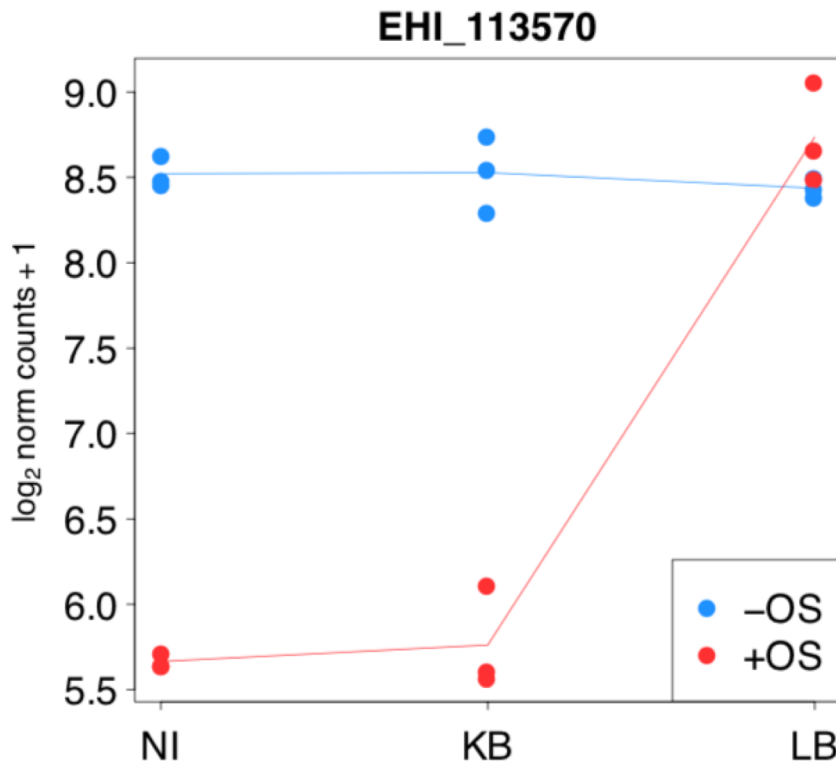
$$\begin{aligned} C2 - C1 &= (+\text{OS.LB vs } -\text{OS.LB}) \text{ vs } (+\text{OS.NI vs } -\text{OS.NI}) \\ &= +\text{OS.LB} - -\text{OS.LB} - +\text{OS.NI} + -\text{OS.NI} \\ &= +\text{OS.LB} - +\text{OS.NI} - -\text{OS.LB} + -\text{OS.NI} \\ &= (+\text{OS.LB vs } +\text{OS.NI}) \text{ vs } (-\text{OS.LB vs } -\text{OS.NI}) \\ &= \qquad\qquad C7 \qquad\qquad - \qquad\qquad C4 \end{aligned}$$

Interaction results

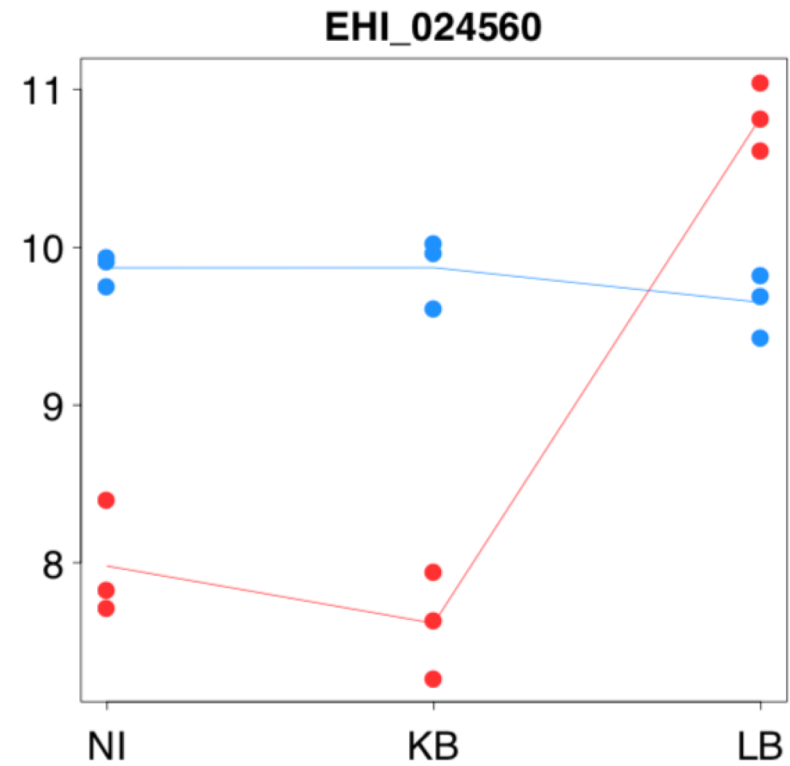
		# down	# up	# total
C10	(+OS.LB vs -OS.LB) vs (+OS.NI vs -OS.NI)	1757	2021	3778



Interaction results



Stress effect cancelled in presence of live *E. Coli*.



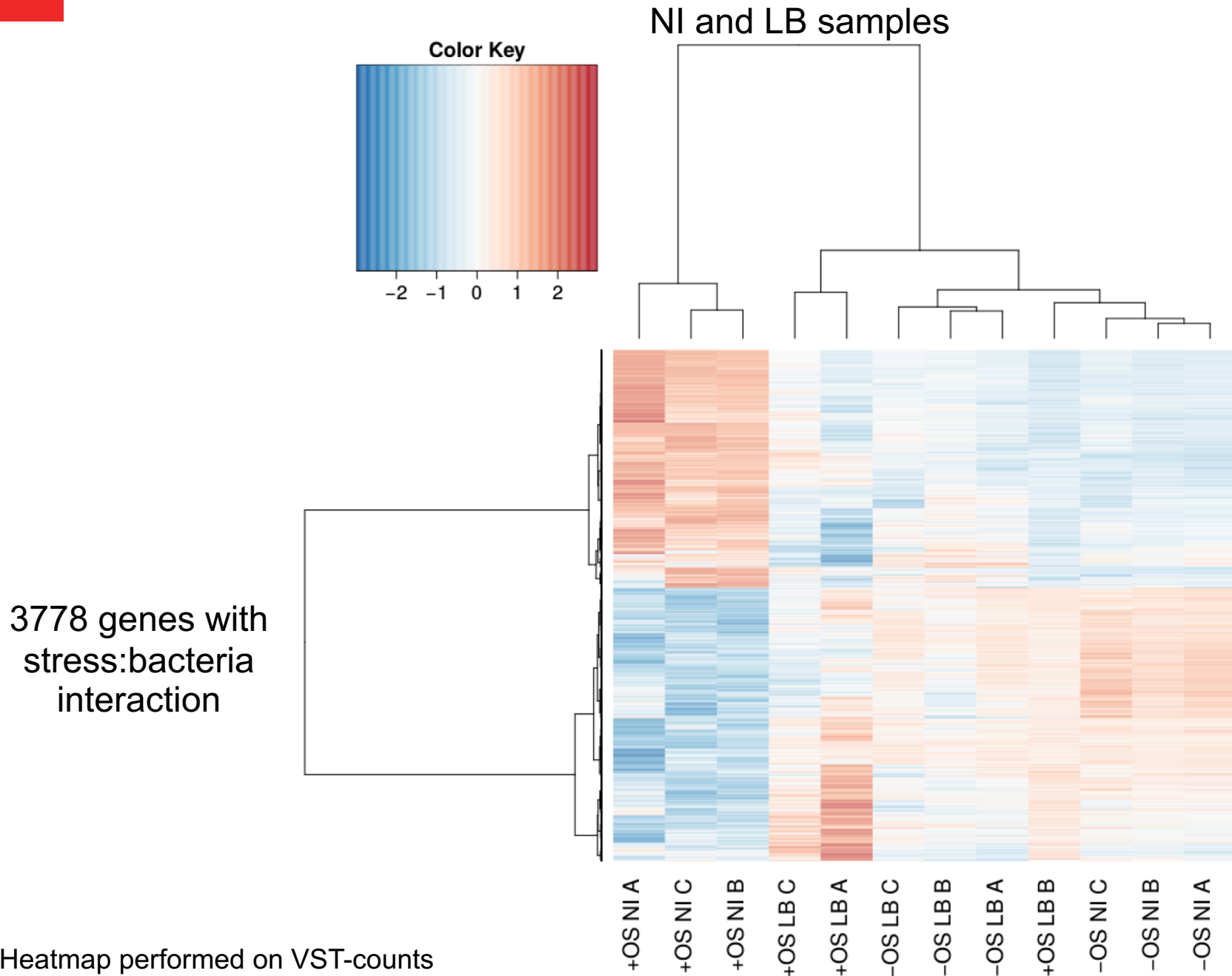
Inversion of the stress effect in presence of live *E. Coli*.



Threshold effects

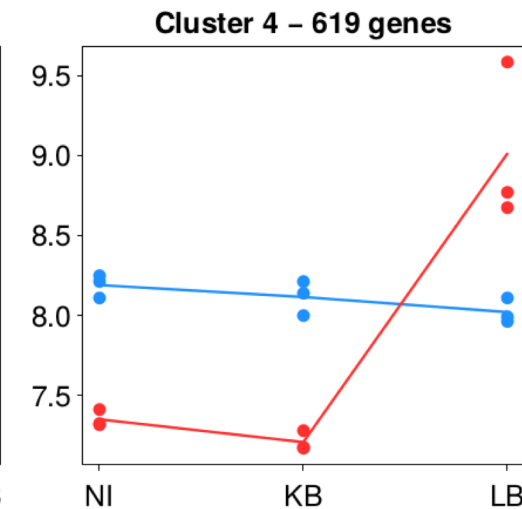
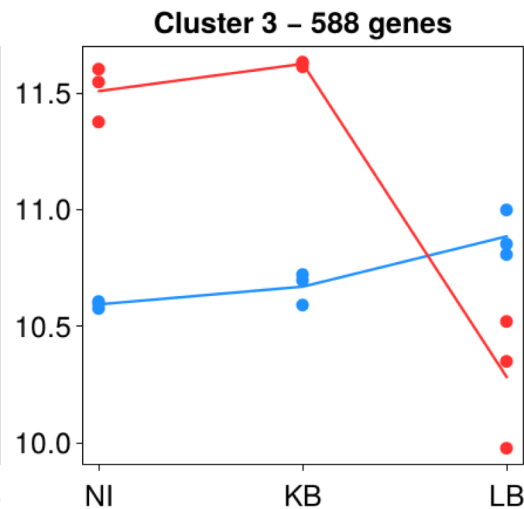
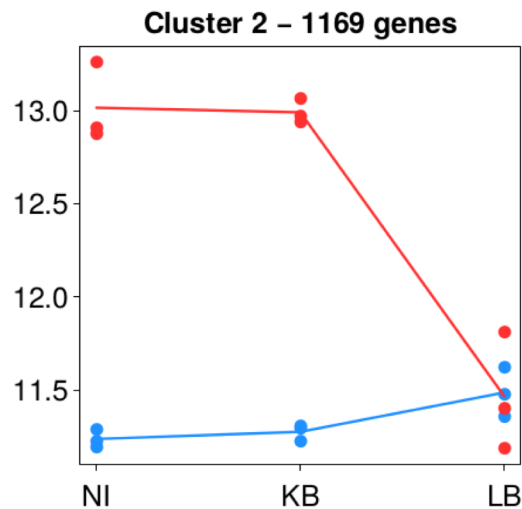
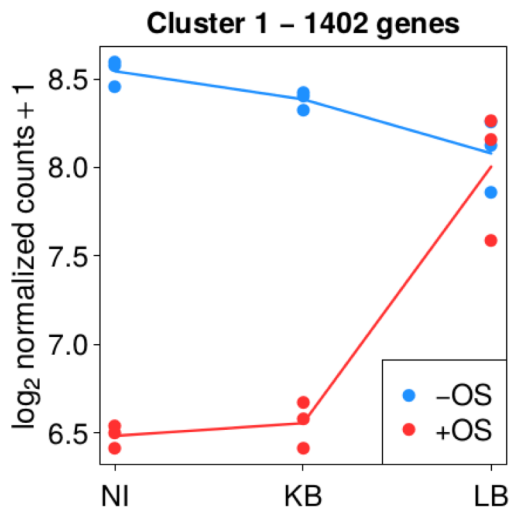
Gene	Comparison C1		Comparison C2		FDR interaction
	\log_2FC	FDR	\log_2FC	FDR	
EHI_192140	-0.755	0.001	-0.442	0.056	0.299
EHI_040470	-1.585	0.001	-0.649	0.049	0.030
EHI_156240	-0.541	0.052	-0.853	0.008	0.497
EHI_113570	-2.887	0.001	0.321	0.295	0.001
EHI_024560	-1.881	0.001	1.177	0.001	0.001

Focus on the interaction results



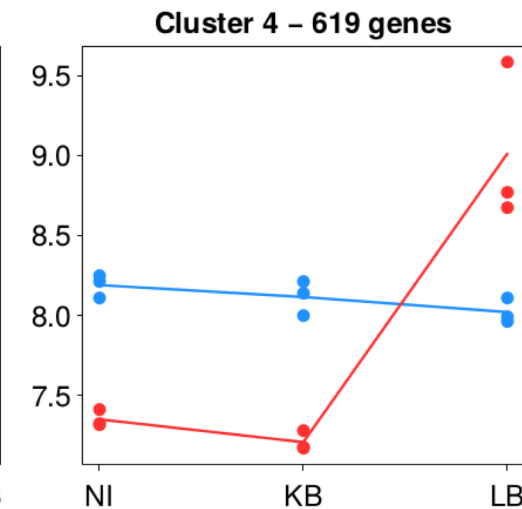
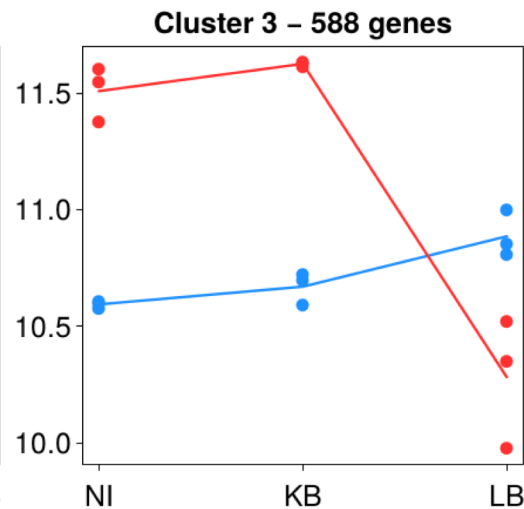
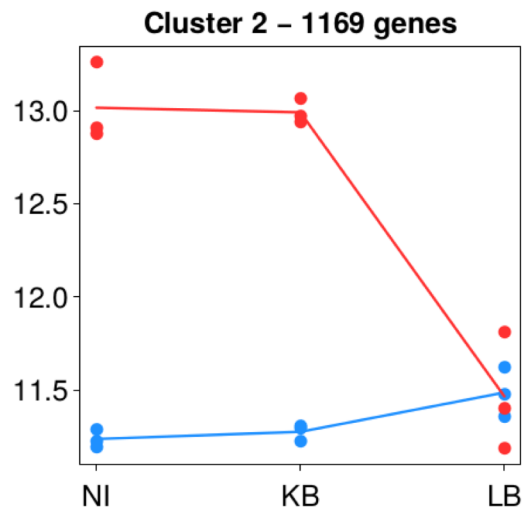
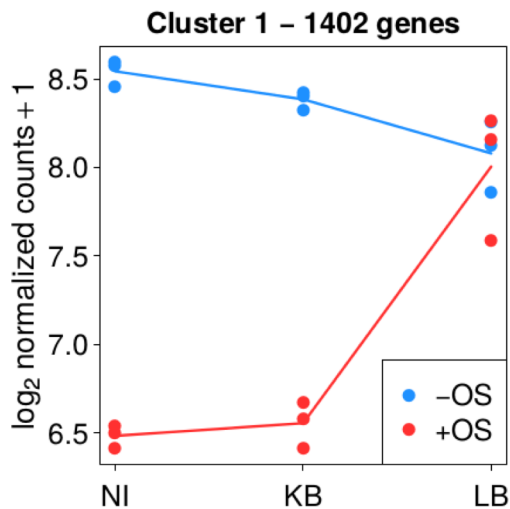
4 clusters of genes

	C1: +OS.NI vs -OS.NI			C2: +OS.LB vs -OS.LB		
	Down ↓	Not DE ↔	Up ↑	Down ↓	Not DE ↔	Up ↑
Cluster 1 1402 genes	1377 (98.2%)	25 (1.8%)	0	61 (4.3%)	1309 (93.4%)	32 (2.3%)
Cluster 2 1169 genes	0	0	1169 (100%)	12 (1.0%)	1121 (95.9%)	36 (3.1%)
Cluster 3 588 genes	1 (0.2%)	149 (25.3%)	438 (74.5%)	197 (33.5%)	390 (66.3%)	1 (0.2%)
Cluster 4 619 genes	329 (53.2%)	283 (45.7%)	7 (1.1%)	0	264 (42.6%)	355 (57.4%)



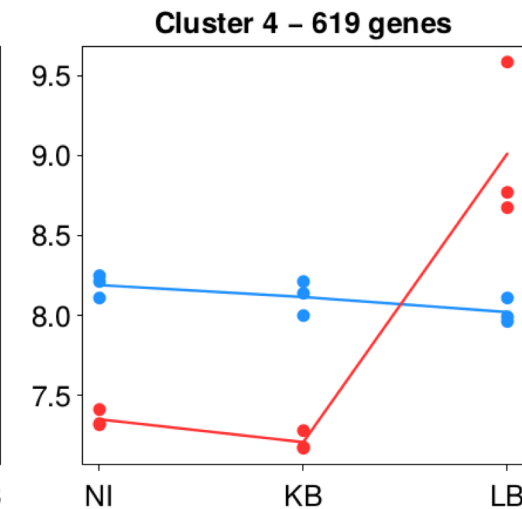
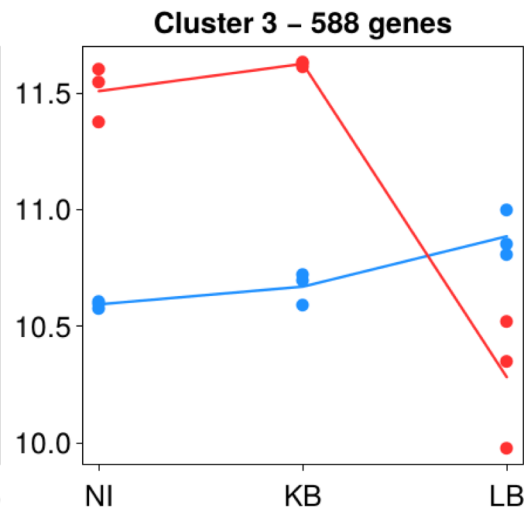
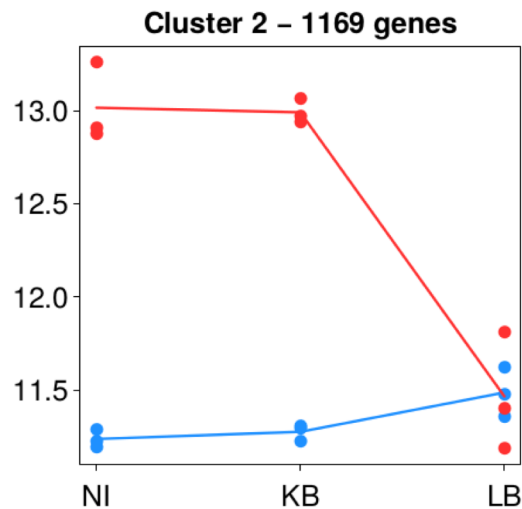
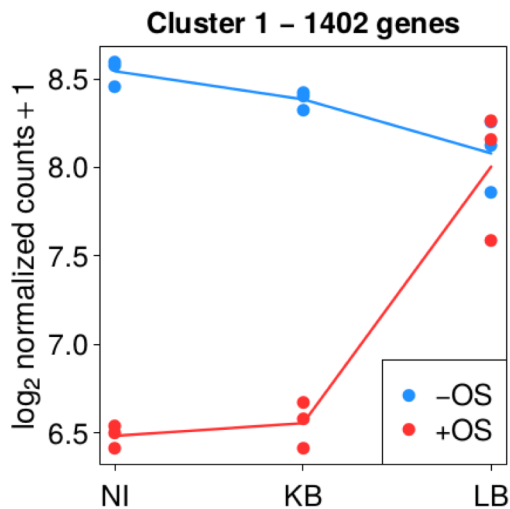
4 clusters of genes

	C1: +OS.NI vs -OS.NI			C2: +OS.LB vs -OS.LB		
	Down ↓	Not DE ↔	Up ↑	Down ↓	Not DE ↔	Up ↑
Cluster 1 1402 genes	1377 (98.2%)	25 (1.8%)	0	61 (4.3%)	1309 (93.4%)	32 (2.3%)
Cluster 2 1169 genes	0	0	1169 (100%)	12 (1.0%)	1121 (95.9%)	36 (3.1%)
Cluster 3 588 genes	1 (0.2%)	149 (25.3%)	438 (74.5%)	197 (33.5%)	390 (66.3%)	1 (0.2%)
Cluster 4 619 genes	329 (53.2%)	283 (45.7%)	7 (1.1%)	0	264 (42.6%)	355 (57.4%)



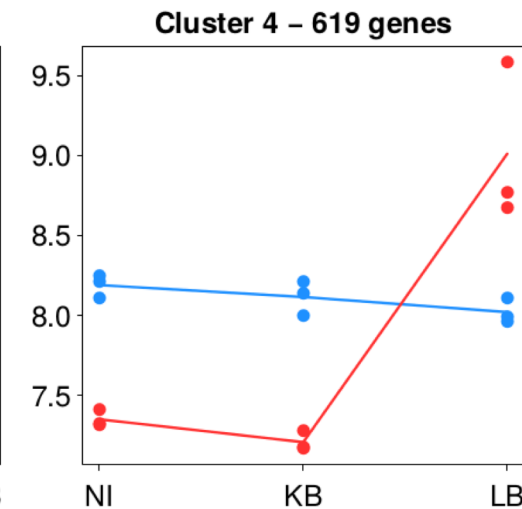
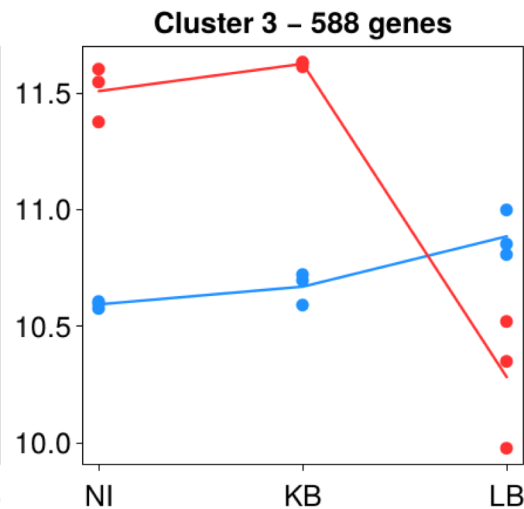
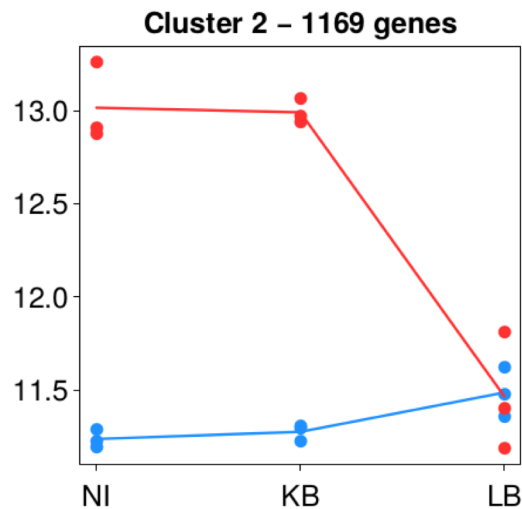
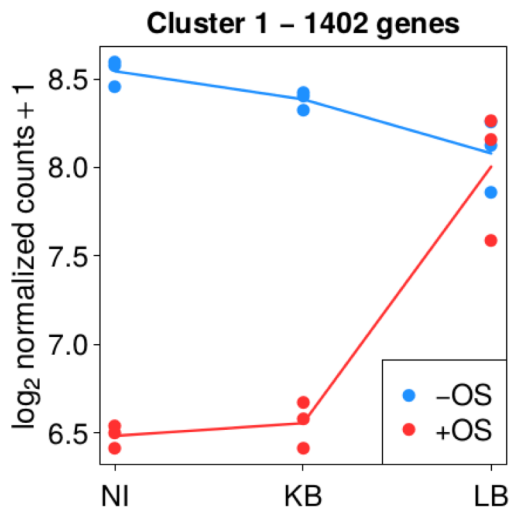
4 clusters of genes

	C1: +OS.NI vs -OS.NI			C2: +OS.LB vs -OS.LB		
	Down ↓	Not DE ↔	Up ↑	Down ↓	Not DE ↔	Up ↑
Cluster 1 1402 genes	1377 (98.2%)	25 (1.8%)	0	61 (4.3%)	1309 (93.4%)	32 (2.3%)
Cluster 2 1169 genes	0	0	1169 (100%)	12 (1.0%)	1121 (95.9%)	36 (3.1%)
Cluster 3 588 genes	1 (0.2%)	149 (25.3%)	438 (74.5%)	197 (33.5%)	390 (66.3%)	1 (0.2%)
Cluster 4 619 genes	329 (53.2%)	283 (45.7%)	7 (1.1%)	0	264 (42.6%)	355 (57.4%)



4 clusters of genes

	C1: +OS.NI vs -OS.NI			C2: +OS.LB vs -OS.LB		
	Down ↓	Not DE ↔	Up ↑	Down ↓	Not DE ↔	Up ↑
Cluster 1 1402 genes	1377 (98.2%)	25 (1.8%)	0	61 (4.3%)	1309 (93.4%)	32 (2.3%)
Cluster 2 1169 genes	0	0	1169 (100%)	12 (1.0%)	1121 (95.9%)	36 (3.1%)
Cluster 3 588 genes	1 (0.2%)	149 (25.3%)	438 (74.5%)	197 (33.5%)	390 (66.3%)	1 (0.2%)
Cluster 4 619 genes	329 (53.2%)	283 (45.7%)	7 (1.1%)	0	264 (42.6%)	355 (57.4%)



Biological interpretation

Strong stress effect on the *Amoeba* transcriptome


The presence of live *E. coli* can:


- cancel the stress effect (clusters 1 and 2)
- reverse the stress effect (clusters 3 and 4)

Many more details in:

Article | [OPEN](#) | Published: 13 June 2018

Enteric bacteria boost defences against oxidative stress in *Entamoeba histolytica*

Hugo Varet, Yana Shaulov, Odile Sismeiro, Meirav Trebicz-Geffen, Rachel Legendre, Jean-Yves Coppée, Serge Ankri  & Nancy Guillen 

Scientific Reports **8**, Article number: 9042 (2018) | [Download Citation](#) 

Aknowledgments



Serge Ankri and his team

Nancy Guillén



Odile Sismeiro

Rachel Legendre

Jean-Yves Coppée

Thanks for your attention!