Linking diet and microbiome data

Supervisors:

M. LUZ CALLE ROSINGANA (Uvic – UCC)
MARC NOGUERA JULIÁN (IrsiCaixa)

JAVIER RIVERA PINTO

IrsiCaixa
Institut de Recerca de la Sida

NOEL

UVIC
UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

Index

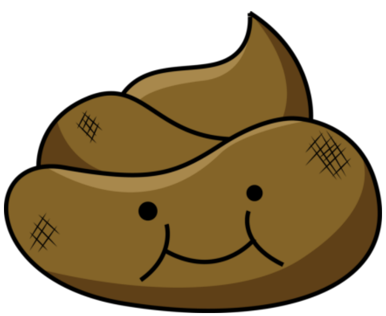
- Introduction
- Previous results
- New hypothesis
- Methodology
- Diet
- Results

Introduction

IrsiCaixa

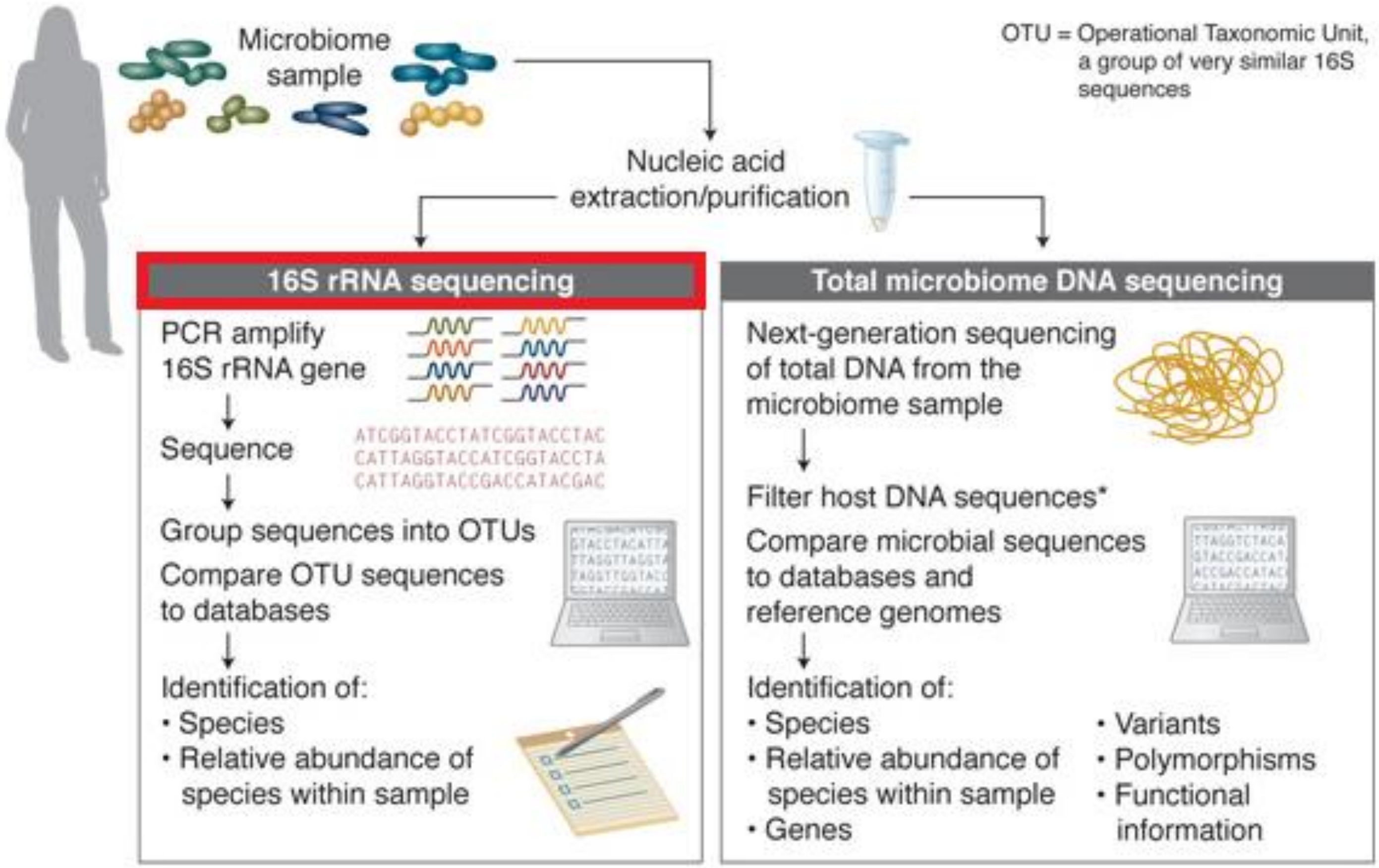
Institut de Recerca de la Sida

IrsiCaixa works with HIV – infected patients to improve their health and the knowdlege of this disease.



In GEM group the gut microbiome is analyzed by means of fecal samples of the patients.

Extracting information



Data transformation

- Fix a **percentage of similarity** between sequences (*97%*)

- Select a **taxonomic level** (*genus*)

- Apply a **filter** (to delete artificial OTUs)

- Data **normalization** (*proportion*, *rarefaction*, . . .) if necessary.

Data appearance

Microbiome information

	Bacteria 1	Bacteria 2	Bacteria 3	. . .	Bacteria k
	25	1	0	…	2151
	0	11	183	…	3
. . .	…	…	…	…	…
	130	0	52	…	1021

	Age	Gender	HIVStatus	. . .	BMI
	31	M	Negative	…	21.7
	24	F	Positive	…	23.5
. . .	…	…	…	…	…
	19	M	Positive	…	19.2

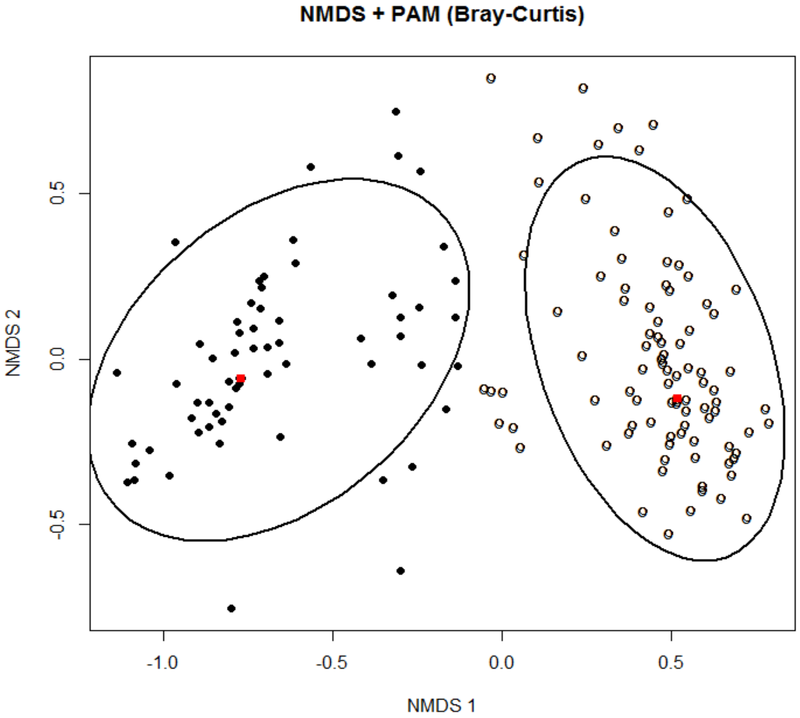
Metadata

Previous results

Main results

- **Two cluster** where found according to microbiome composition.

- These **differences** were **described** mainly by *SexualPractice* variable (unexpected result).



New hypothesis

New hypothesis

- The **diet** has an **influence** over the **gut microbiome** composition

HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2014 July 23.

Published in final edited form as:
Nature. 2014 January 23; 505(7484): 559–563. doi:10.1038/nature12820.

Diet rapidly and reproducibly alters the human gut microbiome

Lawrence A. David[1,2,#], Corinne F. Maurice[1], Rachel N. Carmody[1], David B. Gootenberg[1], Julie E. Button[1], Benjamin E. Wolfe[1], Alisha V. Ling[3], A. Sloan Devlin[4], Yug Varma[4], Michael A. Fischbach[4], Sudha B. Biddinger[3], Rachel J. Dutton[1], and Peter J. Turnbaugh[1,*]

NIH Public Access

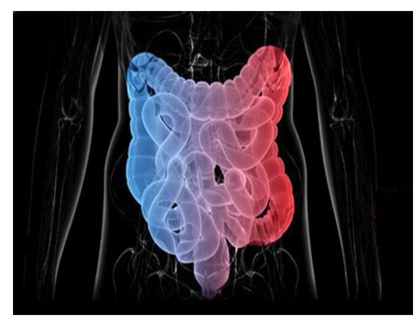
Author Manuscript

Science. Author manuscript; available in PMC 2012 June 06.

Published in final edited form as:
Science. 2011 October 7; 334(6052): 105–108. doi:10.1126/science.1208344.

Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes

Gary D. Wu[1,*], Jun Chen[2,3], Christian Hoffmann[4,5], Kyle Bittinger[4], Ying-Yu Chen[1], Sue A. Keilbaugh[1], Meenakshi Bewtra[1,2], Dan Knights[6], William A. Walters[7], Rob Knight[8,9], Rohini Sinha[4], Erin Gilroy[2], Kernika Gupta[10], Robert Baldassano[10], Lisa Nessel[2], Hongzhe Li[2,3], Frederic D. Bushman[4,*], and James D. Lewis[1,2,3,*]



Methodology

Methodology

- The link between diet and microbiome is estimated following these steps:

- Assume a **Dirichlet – Multinomial** distribution over the gut microbiome count expression.

- **Estimate** the **parameters** of the distribution

- Fit a **log - linear regression** model over these parameters using **LASSO** techniques.

NIH Public Access
Author Manuscript
Ann Appl Stat. Author manuscript; available in PMC 2014 March 01.
Published in final edited form as:
Ann Appl Stat. 2013 March 1; 7(1): . doi:10.1214/12-AOAS592.

VARIABLE SELECTION FOR SPARSE DIRICHLET-MULTINOMIAL REGRESSION WITH AN APPLICATION TO MICROBIOME DATA ANALYSIS*

Jun Chen and
Department of Biostatistics and Epidemiology, University of Pennsylvania
chenjun@mail.med.upenn.edu
Hongzhe Li
Department of Biostatistics and Epidemiology, University of Pennsylvania hongzhe@upenn.edu

Dirichlet – Multinomial distribution

1) Multinomial distribution

Multinomial distribution

Probability mass function

$$f_M(y_1, y_2, \cdots, y_q, \pi) = \binom{y_+}{y} \prod_{j=1}^{q} \pi_j^{y_j} = \frac{y_+!}{y_1! \cdots y_q!} \prod_{j=1}^{q} \pi_j^{y_j}$$

Mean

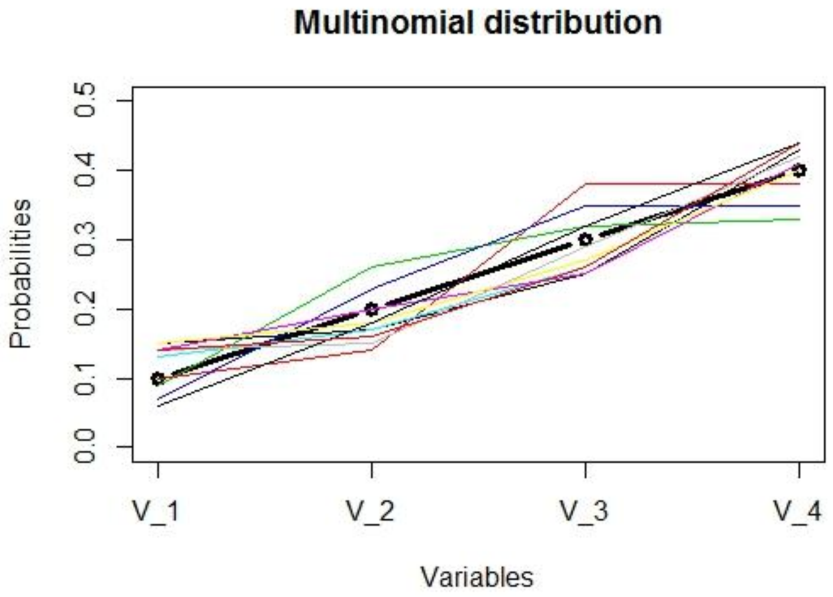
$$\mathbf{E}[Y_j] = y_+ \pi_j$$

Variance

$$\boxed{\mathbf{Var}[Y_j] = y_+ \pi_j (1 - \pi_j)}$$

High variability of real data

- Proportions of each taxa along individuals have **higher variability** than the defined using a Multinomial distribution



Dirichlet – Multinomial distribution

2) Dirichlet distribution

Dirichlet distribution

$$f_D(\pi_1, \pi_2, \cdots, \pi_q; \gamma) = \frac{\Gamma(\gamma_+)}{\prod_{j=1}^{q} \Gamma(\gamma_j)} \prod_{j=1}^{q} \pi_j^{\gamma_j - 1}$$

Mean

$$\mathbf{E}[\pi_j] = \frac{\gamma_j}{\gamma_+}$$

Variance

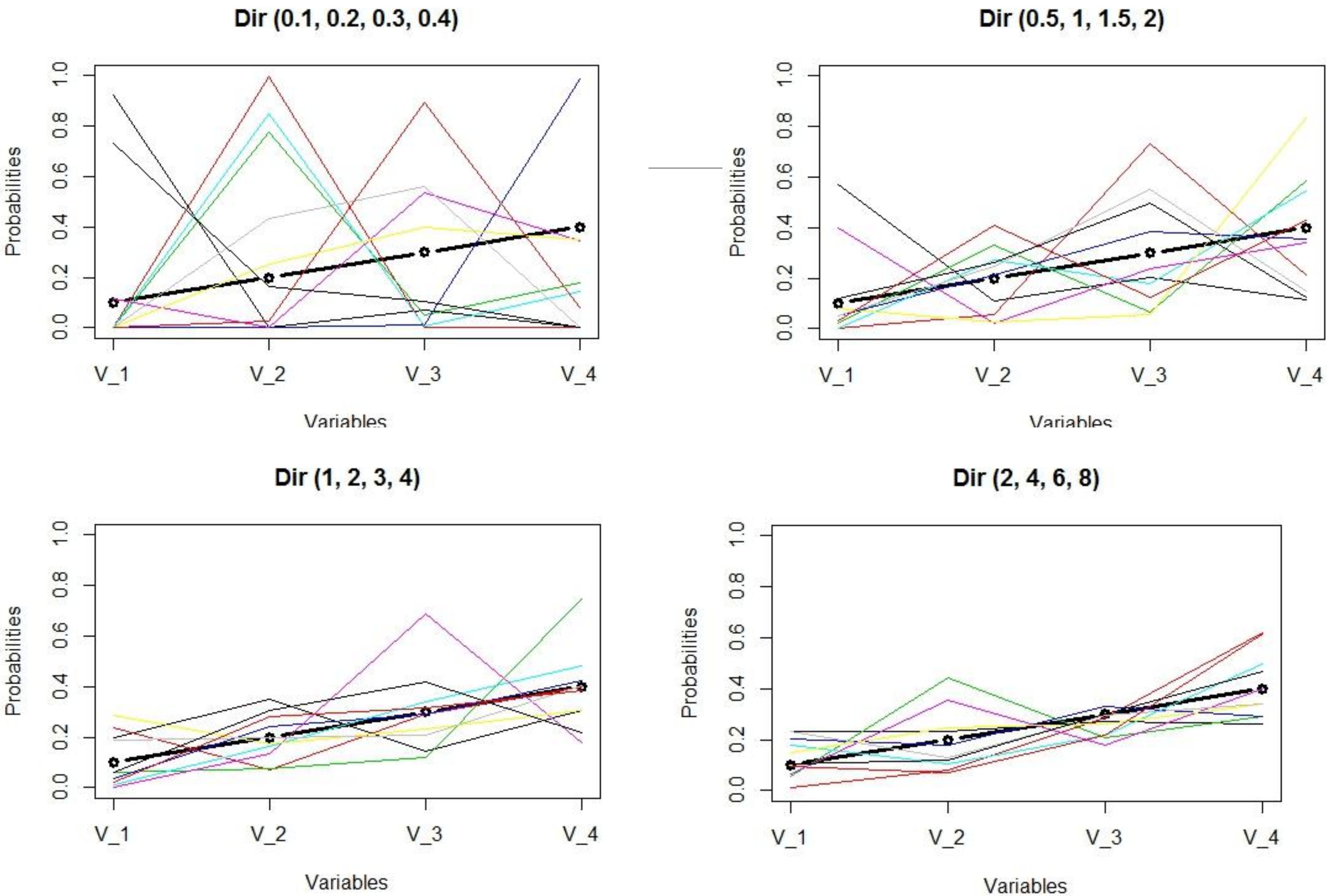
$$\mathbf{Var}[\pi_j] = \frac{\gamma_j(\gamma_+ - \gamma_j)}{(1 + \gamma_+)\gamma_+^2}$$

Some details

- In a Dirichlet distribution, **underlaying proportions** are themselves **positive random variables** subjects to the constraint that they sum 1.

- $\gamma = \{\gamma_1, \cdots, \gamma_q\}$ parameters control the **variability** and the **mean** of each component.

- $\gamma = \{\gamma_1, \cdots, \gamma_q\} = s\{\pi_1, \cdots, \pi_q\}$ with $s \in \mathbb{R}$ The higher ***s*** is, the lower variance.


Dir (0.1, 0.2, 0.3, 0.4)
Dir (0.5, 1, 1.5, 2)
Dir (1, 2, 3, 4)
Dir (2, 4, 6, 8)
Probabilities
0.0
0.2
0.4
0.6
0.8
1.0
V_1
V_2
V_3
V_4
Variables

Dirichlet – Multinomial distribution

3) Dirichlet - Multinomial distribution

Dirichlet – Multinomial distribution

$$f_{DM}(y_1, y_2, \cdots, y_q; \gamma) = \binom{y_+}{y} \frac{\Gamma(y_+ + 1)\Gamma(y_+)}{\Gamma(y_+ + \gamma_+)} \prod_{j=1}^{q} \frac{\Gamma(y_j + \gamma_j)}{\Gamma(\gamma_j)\Gamma(y_j + 1)}$$

Mean

$$\mathbf{E}[Y_j] = y_+ \pi_j$$

Variance

$$\mathbf{Var}[Y_j] = y_+ \pi_j (1 - \pi_j) \left(\frac{y_+ + \gamma_+}{1 + \gamma_+} \right)$$

Regression model over parameters

- After the estimation of γ vector, the following regression model is fitted:

$$\gamma_j(x^i) = exp(\sum_{k=0}^{p} \beta_{jk} x_{ik})$$

being x_{ik} the *k*-th nutrient value for the *i*-th individual in the considered example.

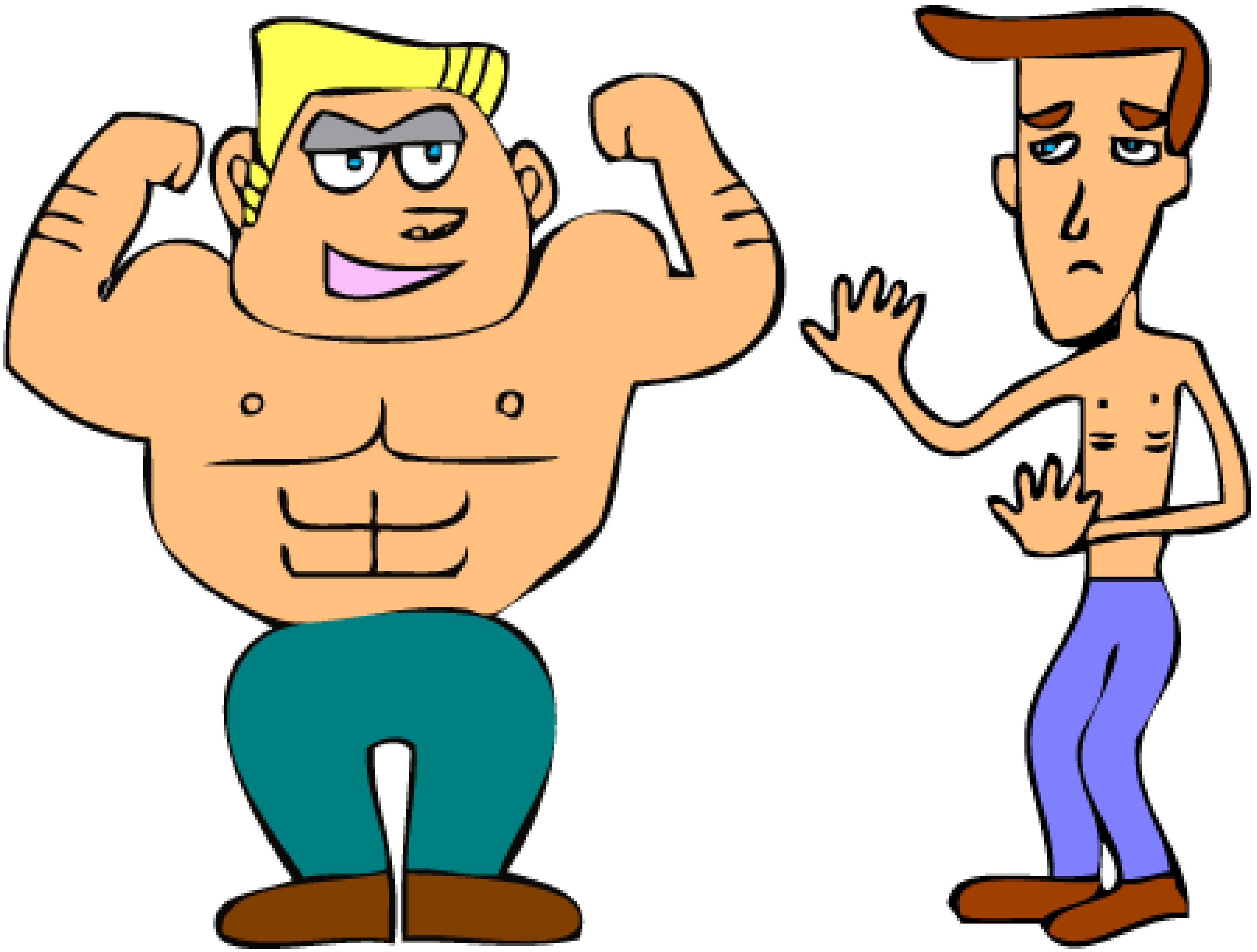
NOTE: LASSO techniques applied to select only the strongest associations.

Diet

Available information

- **Nutrient questionnaire** : consisted of a prospective dietary nutrient survey aimed at recording, as precisely as possible, any food, supplement or liquid intake during 3 to 5 consecutive days, including at least one weekend day.

- **Food portion questionnaire** : based on recall of food portions intake per week, on average, during the last year.



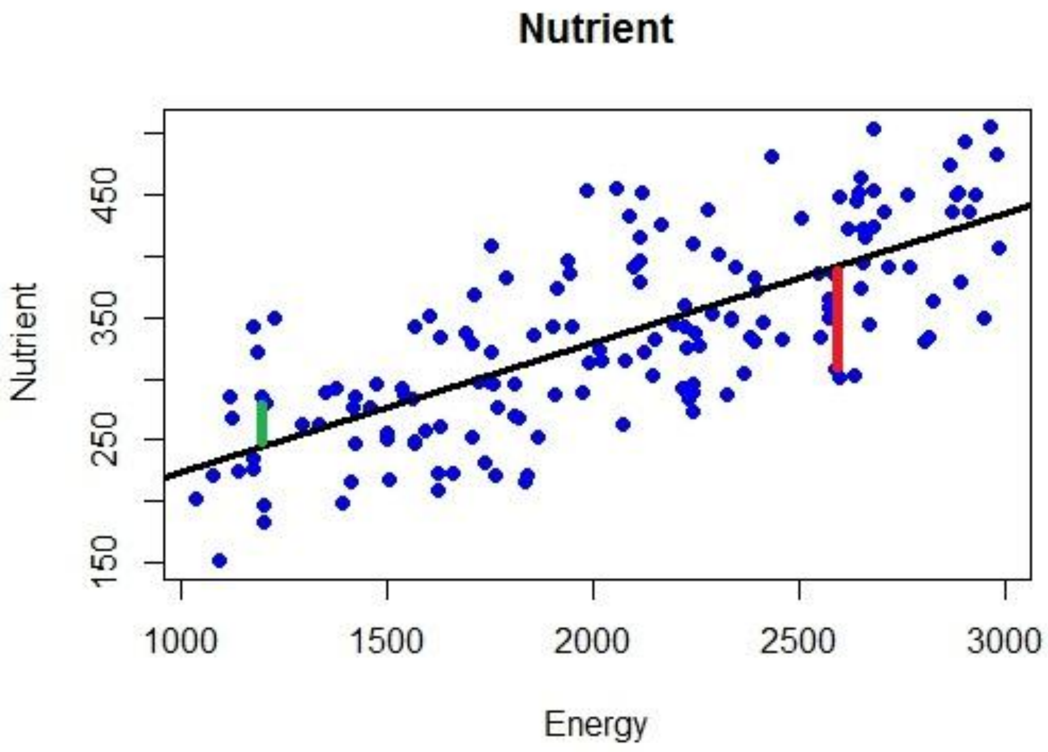
They both eat 100gr. of vegetables every day

Diet transformation I

- **Idea:** Transform data taking into account *energy* variable
- **How?:** For each variable fit a linear model and take residuals as new nutritional values
- **Why?:** Is not the same to consume 100 gr of proteins if your total energy intake is 1500 Kcal as it would be 2500 Kcal.

NOTE: after transforming, scale data by nutrient (mean 0 and standard deviation 1).

Diet transformation II



Results

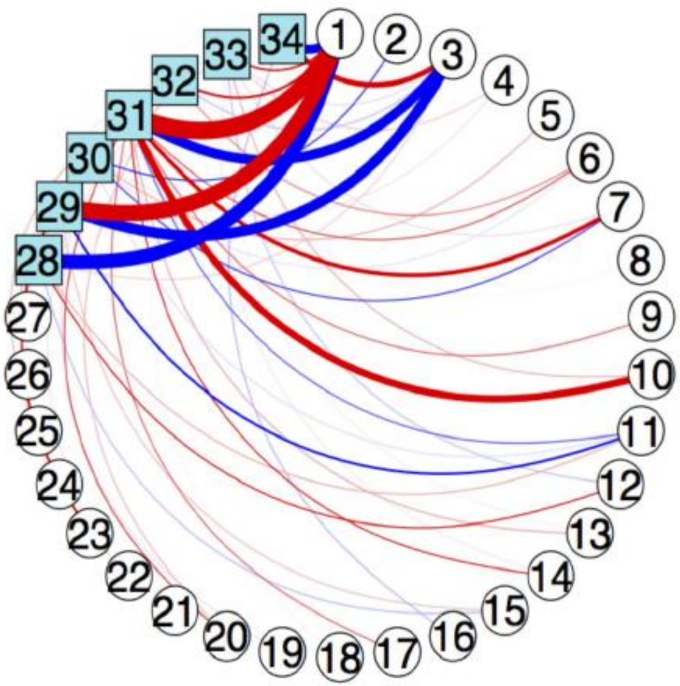
Results

$$\begin{array}{cc} & \textbf{Nutrients} \\ \begin{array}{c}\text{T}\\ \text{A}\\ \text{X}\\ \text{A}\\ \text{S}\end{array} & \begin{bmatrix} \beta_{1,1} & \beta 1, 2 & \cdots & \beta 1, p \\ \beta_{2,1} & \beta 2, 2 & \cdots & \beta 2, p \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{q,1} & \beta q, 2 & \cdots & \beta q, p \end{bmatrix} \end{array}$$

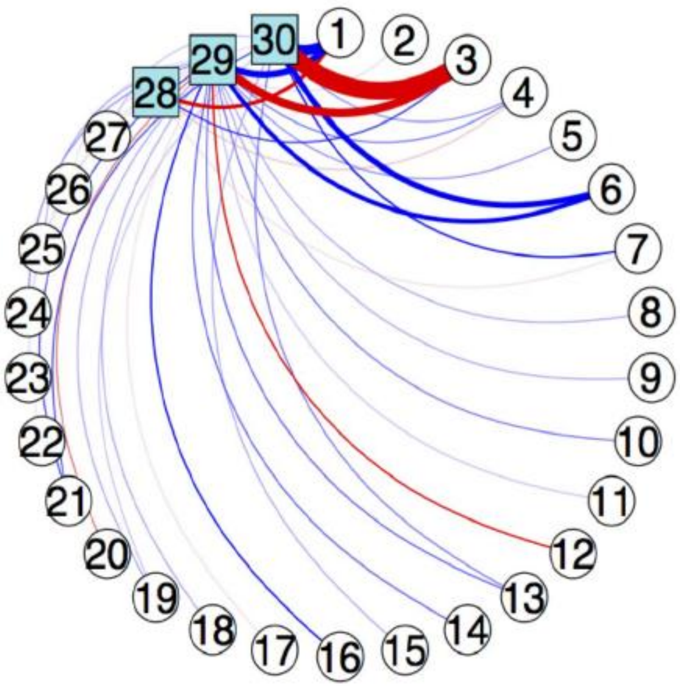
A **matrix** with the **regression coefficients** linking taxas abundance and nutrients standardized consumption.

Plots

Nutrients



Portions



Interpretation

- The **thickness** of the line is linked with strongest associations (fatter lines).

- The **color** represents positive (*red*) or negative (*blue*) associations.

Further interpretations

- There were **differences in genus** between the clusters

- DM – regression pointed some **nutrients associated** to these genus.

- If diet had an impact over the microbiome structure, it would be **expected** to find **differences** between the nutrients consumption for individuals in **Cluster 1** and **2**.

- Only one difference was found (*water*)

Javier Rivera | Mariona Parera | María Casadellà | Chiara Mancuso | Marc Noguera
Rocío Bellido | Roger Paredes | Muntsa Rocafort | Cristina Rodríguez | Yolanda Guillén