MissiON-T2D

Multiscale Immune System SImulator for the Onset of Type 2 Diabetes integrating genetic, metabolic and nutritional data

Work Package 2

Deliverable 2.4

Report on the integration of clinical data model to the overall workflow

Document Information

Grant Agreement	**Nº**	600803	**Acronym**	MISSION-T2D
Full Title	Multiscale Immune System SImulator for the Onset of Type 2 Diabetes integrating genetic, metabolic and nutritional data			
Project URL	http://www.mission-t2d.eu			
EU Project Officer	**Name**	Dr. Adina Ratoi		

Deliverable	**No**	2.4	**Title**	Report on the integration of clinical data model to the overall workflow
Work package	**No**	2	**Title**	Clinical data provision (genetics and aging) and gut microbiota modelling

Date of delivery	**Contractual**		**28.02.2015**		**Actual**		**28.02.2015**		
Status	**Version 2.1**				**Final**		**12.03.2015**		
Nature	**Prototype**		**Report**	**X**	**Dissemination**		**Other**		

Dissemination level	Consortium+EU	
	Public	X

Target Group (If Public)		Society (in general)	
Specialized research communities	X	Health care enterprises	
Health care professionals		Citizens and Public Authorities	

Responsible Author	**Name**	**Stefano Salvioli**	**Partner**	UNIBO
	Email	stefano.salvioli@unibo.it		

Version Log

Issue Date	Version	Author (Name)	Partner
10.03.2015	1.0	Stefano Salvioli, Gastone Castellani, Paolo Garagnani	UniBO
12.03.2015	2.0	Filippo Castiglione, Paolo Tieri	CNR

Executive Summary	In this deliverable we describe the work done in task Task 2.4 Validation and refinement of the model in the overall workflow.
Keywords	Metabolic flexibility; Insulin Resistance; Gut Microbiota composition analysis; Ecological modeling

Contents

1 Deliverable Description

In this document are described the hypotheses, the theoretical assumptions, the methodologies and the results of the Task 2.4: “Report on the integration of clinical data model to the overall workflow”. The main focus will be on the integration between clinical data analysis and modelling activities. We will describe the main steps that we have done for the creation of such models and for their exploitation on clinical data. The main goal is to develop a model for the metabolic flexibility switch in Health and Type 2 Diabetes. This will involve a skeletal muscle cell modelling and also a dynamical modelling of Gut Microbiota (GM), in order to characterize their response to external perturbations such as dietary changes, age or health state.

The metabolic flexibility model will be computed starting from the biochemical background described at the beginning of this document. Clinical datasets on gene expression of Healthy, Insulin Resistant and Type 2 Diabetes patients from the GEO database will be analyzed and exploited to assess which key points in our cell model could produce the metabolic flexibility impairment. Pathway analysis will also enable us to assess the effects of Insulin Resistance and Type 2 Diabetes on the gene expression of liver and adipose tissue.

For what concern the GM modelling, we will propose a new model based on a 2D chemostat, and whose stationary distribution will be obtain by a CME approach through numerical simulations in Python and, where possible, also analytical computations. Clinical 16S rRNA sequencing data from Claesson et al will be analysed and fitted with our model.

Finally, we will propose a table summarizing the main effects of diet on bacteria and the main products of such bacteria that could affect the host metabolic and immune systems, including its metabolic flexibility.

2 Deliverable Results

2.1 METABOLIC FLEXIBILITY AND INSULIN RESISTANCE

Metabolic flexibility is the capacity to switch from predominantly lipid oxidation and high rates of fatty acid uptake during fasting conditions to the suppression of lipid oxidation and increased glucose uptake, oxidation, and storage under insulin-stimulated

conditions or after a meal [**1**][**2**].

Skeletal muscle is the most important organ concerning the uptake and oxidation of fatty acids. [**2**]

High glucose concentrations have potential toxic effects, and skeletal muscle will prevent toxic glucose concentrations by increasing glucose uptake and oxidation at the cost of fatty acid uptake and oxidation. Also high FFA levels are toxic. However, after a meal, the body has – in contrast to glucose – the option to shuttle large amounts of excess FFA to the adipose tissue. [**2**]

2.1.1 Postprandial phase:

The postprandial phase is characterized by high blood concentration of nutrients (glucose and fatty acids). Pancreatic beta cells respond to this high glucose concentration releasing insulin in the bloodstream.

In myocytes, insulin receptor binds insulin and undergoes autophosphorylation on its carboxyl-terminal Tyr residues activating insulin signalling: insulin receptor phosphorylates IRS-1 on its Tyr residues; IRS-1, phosphorylated, activates PI-3K by binding to its SH2 domain; PI-3K converts PIP2 to PIP3; Akt/PKB bound to PIP3 is phosphorylated and activated by PDK1; PKB phosphorilates stimulates AS160 phosphorylation and in conclusion insulin-stimulated phosphorylation of AS160 (IR → IRS-1-PI3K → Akt2 → AS160) releases the inhibition on vesicular GLUT4 movement to the plasma membrane, as phosphorylated AS160 can no longer maintain rab proteins in a GDP-bound state. [**3**]

Similarly to insulin-induced GLUT4 translocation, the insulin-induced translocation of fatty acid transporters (CD36 and FATP1) requires the activation of the PI3K-Akt2 signaling pathway even if the pathway for GLUT4 and the one for fatty acid transporters can diverge at some point. [**3**]

Glucose is transported inside the cell through GLUT4 and enter the glycolysis pathway in which it is converted to pyruvate, that enters the TCA cycle in the mitochondria.

Fatty acids instead are converted by acyl-CoA synthetase to acyl-CoA, that enters the mitochondria through the carnitine shuttle and, converted to acetyl-CoA by b-oxidation, enters the TCA cycle.

Through the TCA cycle, glucose and fatty acid fluxes increase intra-mitochondrial

citrate, that through the citrate shuttle is transferred in the cytosol and is converted to acetyl-CoA.

Cytosolic citrate is able to stimulate ACC2 (acetyl-CoA carboxylase-2), that is the first enzyme of lipogenesis, that converts acetyl-CoA to malonyl-CoA (fat storage).

Malonyl-CoA is a potent allosteric inhibitor of carnitine-palmitoyl transferase-1 (CPT-1) (carnitine shuttle) and then blocks mitochondrial FA-CoA uptake and therefore limit b-oxidation, that is it inhibits FA oxidation. [**2**]

Meanwhile also insulin signalling (PIP3) promotes ACC2 activation through a protein phosphatase insulin dependent.[**6**]

2.1.2 Fasting phase:

During the fasting phase plasma glucose and insulin drop down, while plasma glucagon and FFA (coming from adipocytes) rise.

Consequently there will be less GLUT4 e CD36 in cells plasmatic membrane (since their insulin dependent positioning is reversible) and a lesser glucose and FFA uptake.

The energy demand will cause a drop in citrate levels, and consequently in cytosolic acetyl-coa levels. Then ACC2 won't be activated anymore, niether by citrate nor by insulin, but it will be instead phosphorylated and so inactivated by glucagon through PKA.

There will be a subsequent drop in malonyl-CoA levels (no fat storage), disinhibition of CPT-1 activity, increased mitochondrial fatty acyl-CoA uptake and an increase in fatty acid oxidative disposal. [**2**]

2.2 INSULIN RESISTANCE

Overfeeding causes an excess in plasma glucose and lipids concentration.

The **excess of plasma glucose** activates some processes that are strong sources of oxidative stress, like: glucose autooxidation, overproduction of ROS by mitochondria, non-enzymatic glycation, and the polyol pathway .

ROS may affect insulin signalling. This may occur through several mechanisms: ROS may induce serine phosphorylation of insulin receptor substrate, decreasing tyrosine phosphorylation thereby interfering with insulin signaling; similarly, ROS have been

shown to partially mediate the effect of Angiotensin II inhibition of insulin signalling. Methylglyoxal, a biologically active AGE precursor formed under conditions of hyperglycemia, inhibits phosphorylation of insulin receptor substrate and activation of the phosphatidylinositol 3-kinase (PI3K)/protein kinase B (PKB) pathway. [**4**]

Hyperglycemia-induced oxidative stress leads to the generation of intracellular signals that stimulate inflammation and cell death. They include protein kinase C (PKC), c-Jun-N-terminal kinase (JNK), and p38 mitogen-activated protein kinase (MAPK) and they inhibit insulin signalling .[**4**][**5**]

Inflammation in turn activates FOXOs mono-ubiquitinating them.

Hyperglycemia-induced FOXO plays an important role in the induction of proinflammatory cytokines and may be involved in a forward amplification loop.

Enhanced activation of FOXO1 affects the expression mitochondrial fusion and fission thereby affecting mitochondrial biogenesis and compromising their role against ROS and mitochondrial oxidation.[**4**]

Excess lipid accretion and/or lower triglyceride turnover can induce lipotoxicity, as reflected by the cellular accumulation of long chain fatty acyl-CoA (FA-CoA), diacylglycerol (DAG), triacylglycerol (IMTG) and ceramide. These lipid species ultimately impair insulin signaling through different mechanisms, either increased serine phosphorylation of the insulin receptor and insulin receptor substrate 1 and/or reduced serine phosphorylation of PKB/Akt . [**1**][**2**]

2.3 METABOLIC INFLEXIBILITY

2.3.1 Postprandial phase

Due to insulin resistance, that is an impaired insulin signalling, GLUT4 and CD36 won't be translocated in the plasmatic membrane and there will be a consequently lesser provision of glucose and FA inside the cell. It means a lesser citrate level coming from the TCA cycle. There will not be the ACC2 activation by insulin signalling and citrate and so ACC2 remains inactivated and FA storage will be unpaired. Meanwhile there won't be malonyl-CoA inhibition of carnitine shuttle, that is of FA oxidation.

Consequently there will be less glycolysis (due to the low glucose level inside the cell) and FA oxidation won't stop.

2.3.2 Fasting phase

During the fasting phase, due to insulin resistance, plasma glucose and FFA levels will remain high. So there won't be a relevant drop in mitochondrial citrate level and ACC2 won't be inactivated. Hyperglycemia will also impair the rising of plasma glucagon levels and there won't be even this factor to inactivate ACC2.

Fat lipogenesis and storage will keep on, as well as its inhibition of the carnitine shuttle that means of FA oxidation.

2.4 GENE EXPRESSION AND PATHWAY ANALYSIS

We studied the differences in gene expression between Healthy and Insulin Resistant or Type 2 Diabetes subjects in three tissues: skeletal muscle, adipose tissue and liver.

Since skeletal muscle is the most important organ concerning the uptake and oxidation of fatty acids, i.e. involved in metabolic flexibility, for such tissue we also assessed the effect of the impairment of the cell metabolic flexibility due to insulin resistance.

We identified 5 datasets from the GEO database:

- GSE 13070 [7]
- GSE 18732 [8]
- GSE 15773 [9]
- GSE 20950 [9]
- GSE 23343 [10]

In all five cases, data were generated with Affymetrix Human Genome U133 Plus 2.0. These arrays contain > 54000 probe sets, composed of 25 nucleotides each, corresponding to about 38500 genes overall.

The following table summarizes the number of samples of each type along the 5 datasets.

	MUSCLE NT	MUSCLE T	ADIPOSE TISSUE	LIVER
HEALTHY	15 (GSE13070) 47 (GSE18732)	16 (GSE13070)	6 (GSE13070) 5 (GSE15773) 10 (GSE20950)	7 (GSE23343)
INSULIN RESISTANT	46 (GSE13070)	46 (GSE13070)	25 (GSE13070) 4 (GSE15773) 9 (GSE20950)	X
TYPE 2 DIABETES	45 (GSE18732)	X	X	10 (GSE23343)

NT = Not Treated (fasting phase)

T = Treated with a 5-h hyperinsulinemic (80 mU m^{-2} min^{-1}) euglycemic clamp (simulates the postprandial phase)

Starting from the .CEL files, we extracted the data through the software Affymetrix Power Tool, with the Robust Multi-array Avarage (RMA) method.

We performed such reworking because data from different works had been processed in different ways.

In order to verify the datasets equivalence, we computed a Principal Component Analysis (PCA) considering the healthy not treated muscles samples form the 2 datasets GSE13070 and GSE18732.

As shown in Figure 1, the 2 datasets clearly separate, meaning that we can not consider their samples as equivalent. Thus, we chose to analyze the datasets separately.

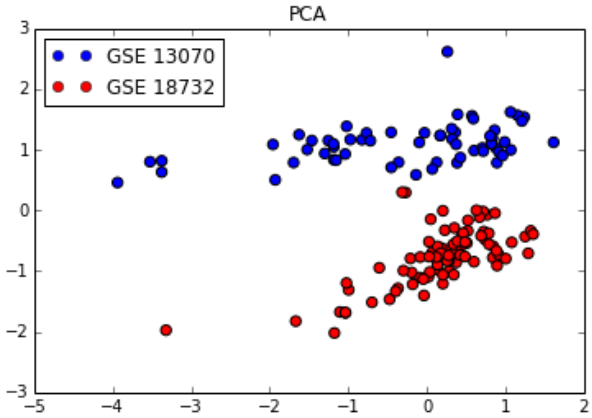


Figure 1. PCA considering the healthy not treated muscle samples from GSE13070 (blue) and GSE18732 (red).

In order to compare gene expression in healthy, insulin resistance and type 2 diabetes, we performed an ANalysis Of VAriance (ANOVA).

In order to perform the pathway analysis, we imported the KEGG classification of genes and pathways through the library keggrest.

We performed a Fisher's test to identify the significantly different pathways among those in which the significantly different genes are involved, through the library

multilinear.

2.5 Healthy vs Insulin Resistant

2.5.1 Muscle (GSE13070)

In order to assess which genes and pathways are significantly different in skeletal muscle cells between Healthy and Insulin resistant subjects, we considered data from GSE 13070. We computed the ANOVA considering fasting subjects, so as not to include the feeding variability. This comparison found that 3926 over 54675 probes had a significantly different expression in Healthy versus Insulin Resistant samples (p-value < 0.05). These corresponded to 2504 over 17526 genes according to the KEGG genes classification.

Among the 190 pathways that held these genes, pathway analysis identified 39 pathways that significantly changed between IR and Healthy samples (p-value < 0.05), that, excluding diseases associated pathways, are:

PATHWAY	P-VALUE
Proximal tubule bicarbonate reclamation	0,000001
Glyoxylate and dicarboxylate metabolism	0,00007
Metabolic pathways	0,00013
Valine, leucine and isoleucine degradation	0,00017
Arginine and proline metabolism	0,00085
Purine metabolism	0,00104
Carbon metabolism	0,00160
Alanine, aspartate and glutamate metabolism	0,00168
Long-term potentiation	0,00279
Morphine addiction	0,00433
Long-term depression	0,00529
D-Glutamine and D-glutamate metabolism	0,00586
GABAergic synapse	0,00641
Porphyrin and chlorophyll metabolism	0,00648
Oxidative phosphorylation	0,00776
Leukocyte transendothelial migration	0,00871
Retrograde endocannabinoid signaling	0,00970
Ribosome	0,01163
cGMP-PKG signaling pathway	0,01170
Rap1 signaling pathway	0,01170
Aminoacyl-tRNA biosynthesis	0,01519
Signaling pathways regulating pluripotency of stem cells	0,01688
Ras signaling pathway	0,02278
Vascular smooth muscle contraction	0,02424
Platelet activation	0,02666
Progesterone-mediated oocyte maturation	0,02990
Pyruvate metabolism	0,03043
Salivary secretion	0,03188
Synthesis and degradation of ketone bodies	0,03489
Dopaminergic synapse	0,03736
Proteoglycans in cancer	0,04025
Gap junction	0,04225
cAMP signaling pathway	0,04357
Adrenergic signaling in cardiomyocytes	0,04430
One carbon pool by folate	0,04846
Bile secretion	0,05181
Insulin signaling pathway	0,05205

2.5.2 Muscle – Metabolic flexibility (GSE13070)

In this part of the analysis we wanted to find out which genes and pathways significantly changed in the fasting versus postprandial comparison, because of the health state. In this way we assessed the genes and pathways involved in the metabolic flexibility impairment due to insulin resistance.

We computed a 2 way ANOVA and focused on the interaction term, which indeed describes the dependence of the differences between the fasting and postprandial phase on the health state.

This comparison found 663 over 54675 significant probes (p-value < 0.05), that corresponded to 343 over 17526 genes according to the KEGG genes classification.

Among the 135 pathways that held these genes, pathway analysis identified 5 significant pathways (p-value < 0.05):

PATHWAY	P-VALUE
Circadian rhythm	0,00008
Complement and coagulation cascades	0,0048
MicroRNAs in cancer	0,0333
Aldosterone-regulated sodium reabsorption	0,0444
Carbohydrate digestion and absorption	0,0444

2.5.3 1.3 Adipose tissue (GSE13070)

We computed a 1 way ANOVA for adipose tissue samples from GSE 13070, comparing Healthy versus Insulin Resistant subjects. This comparison found that 11327 over 54675 probes had a significantly different expression in the two groups (p-value < 0.05). These corresponded to 6407 over 17526 genes according to the KEGG genes classification.

The pathway analysis identified the following 95 over 238 pathways that significantly changed between Healthy and IR samples (p-value < 0.05), among which:

PATHWAY	P-VALUE
Lysosome	0,000000018
Metabolic pathways	0,000000056
Fatty acid metabolism	0,00000026
Valine, leucine and isoleucine degradation	0,00000031
Fc epsilon RI signaling pathway	0,0000013
Platelet activation	0,000021
Oxidative phosphorylation	0,00012
Leukocyte transendothelial migration	0,00018
Fatty acid degradation	0,00019
Fc gamma R-mediated phagocytosis	0,00019
Citrate cycle (TCA cycle)	0,00030
Mineral absorption	0,00033
HIF-1 signaling pathway	0,00036
Pathways in cancer	0,00054
AMPK signaling pathway	0,00073
Fatty acid elongation	0,00084
Apoptosis	0,00113
Calcium signaling pathway	0,00152
Other glycan degradation	0,00154
Glycosaminoglycan degradation	0,00178
Tight junction	0,00192
Phagosome	0,00192
B cell receptor signaling pathway	0,00195
VEGF signaling pathway	0,00203
Adipocytokine signaling pathway	0,00211
Inflammatory mediator regulation of TRP channels	0,00261
T cell receptor signaling pathway	0,00310
Insulin signaling pathway	0,00346
Focal adhesion	0,00357
Chagas disease (American trypanosomiasis)	0,00369
Biosynthesis of unsaturated fatty acids	0,00375
Toll-like receptor signaling pathway	0,00402
Regulation of actin cytoskeleton	0,00411
Glutamatergic synapse	0,00417
Propanoate metabolism	0,00433
Type II diabetes mellitus	0,00471
PPAR signaling pathway	0,00515
PI3K-Akt signaling pathway	0,00591
Natural killer cell mediated cytotoxicity	0,00632
Oxytocin signaling pathway	0,00703
MAPK signaling pathway	0,00718
Endocrine and other factor-regulated calcium reabsorption	0,00721
FoxO signaling pathway	0,00850
Carbon metabolism	0,00916
Pyruvate metabolism	0,00962
p53 signaling pathway	0,01098
Axon guidance	0,01563
Leishmaniasis	0,01639
Rap1 signaling pathway	0,01695
Thyroid hormone signaling pathway	0,01773
NF-kappa B signaling pathway	0,01857
Amino sugar and nucleotide sugar metabolism	0,02012
Melanogenesis	0,02476
Morphine addiction	0,02553
Glyoxylate and dicarboxylate metabolism	0,02589
Notch signaling pathway	0,02645
Amoebiasis	0,02722
GnRH signaling pathway	0,03048
Fatty acid biosynthesis	0,03085
Phosphatidylinositol signaling system	0,03423
TGF-beta signaling pathway	0,03423
Gastric acid secretion	0,03487
Hematopoietic cell lineage	0,03642
Tryptophan metabolism	0,03771
Neurotrophin signaling pathway	0,03879
Inositol phosphate metabolism	0,04219
Prolactin signaling pathway	0,04490
Dopaminergic synapse	0,04858

2.6 Healthy vs Type 2 Diabetes

2.6.1 Muscle (GSE18732)

We exploited data from GSE 18732 in order to assess the gene expression differences between Healthy and Type 2 Diabetes subjects in skeletal muscle cells. The ANOVA for this comparison found that 3546 over 54675 probes had a significantly different expression in the two groups (p-value < 0.05). These corresponded to 2168 over 17526 genes according to the KEGG genes classification.

The pathway analysis identified the following 28 over 171 pathways that significantly changed between T2D and healthy samples, among which:

PATHWAY	P-VALUE
Spliceosome	0,000006
Ubiquitin mediated proteolysis	0,00066
Regulation of actin cytoskeleton	0,00999
Protein processing in endoplasmic reticulum	0,01495
Axon guidance	0,01742
Glycosaminoglycan degradation	0,02810
MAPK signaling pathway	0,03314
Adherens junction	0,03362
B cell receptor signaling pathway	0,03892
AMPK signaling pathway	0,04092
FoxO signaling pathway	0,04292
Wnt signaling pathway	0,04474
Prolactin signaling pathway	0,04483

2.6.2 Liver (GSE23343)

The ANOVA for this comparison found that 5761 over 54675 probes had a significantly different expression in Healthy and Type 2 Diabetes samples (p-value < 0.05). These corresponded to 3804 over 17526 genes according to the KEGG genes classification.

In particular **TCF7** resulted significantly different (p-value = 0.047), being, on avarage, more expressed in Healthy than T2D.

Pathway analysis identified 49 over 192 pathways that significantly changed between T2D and Healthy samples (p-value < 0.05).

PATHWAY	P-VALUE
Protein processing in endoplasmic reticulum	0,00003
Tight junction	0,00009
Ubiquitin mediated proteolysis	0,00011
Ribosome	0,00026
Focal adhesion	0,00029
Hippo signaling pathway	0,00062
Ras signaling pathway	0,00062
Nucleotide excision repair	0,00105
One carbon pool by folate	0,00138
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	0,00183
Regulation of actin cytoskeleton	0,00194
RNA transport	0,00237
Platelet activation	0,00360
cGMP-PKG signaling pathway	0,00390
Circadian rhythm	0,00515
MAPK signaling pathway	0,00542
ErbB signaling pathway	0,00564
Spliceosome	0,00645
Glioma	0,00689
Non-alcoholic fatty liver disease (NAFLD)	0,00718
Oxytocin signaling pathway	0,00760
Metabolic pathways	0,01046
Endocytosis	0,01511
Adherens junction	0,01581
mRNA surveillance pathway	0,01597
TGF-beta signaling pathway	0,01831
Vascular smooth muscle contraction	0,02241
Protein export	0,02251
Oxidative phosphorylation	0,02349
Thyroid hormone signaling pathway	0,02672
Leukocyte transendothelial migration	0,02978
Serotonergic synapse	0,03392
NOD-like receptor signaling pathway	0,03938

2.7 METABOLIC FLEXIBILITY PATHWAY

We computed a metabolic flexibility pathway, summarizing the main involved steps described previously.

The principal KEGG implicated pathways:

hsa04930 - Type II diabetes mellitus
hsa04910 - Insulin signaling pathway
hsa00010 - Glycolysis / Gluconeogenesis
hsa00620 - Pyruvate metabolism
hsa00020 - Citrate cycle (TCA cycle)
hsa00061 - Fatty acid biosynthesis
hsa00062 - Fatty acid elongation
hsa00071 - Fatty acid metabolism
hsa03320 - PPAR signaling pathway
hsa00190 - Oxidative phosphorylation

hsa04920 - Adipocytokine signaling pathway
hsa04151 - PI3K-Akt signaling pathway
hsa04068 - FoxO signaling pathway

We propose a simple and a more complete version of the metabolic flexibility pathway.

The blue and red borders represent the two metabolic pathways among which the cell switches between the fasting (blue) and postprandial (red) phase. This is the switching that results impaired in insulin resistance and Type 2 Diabetes.

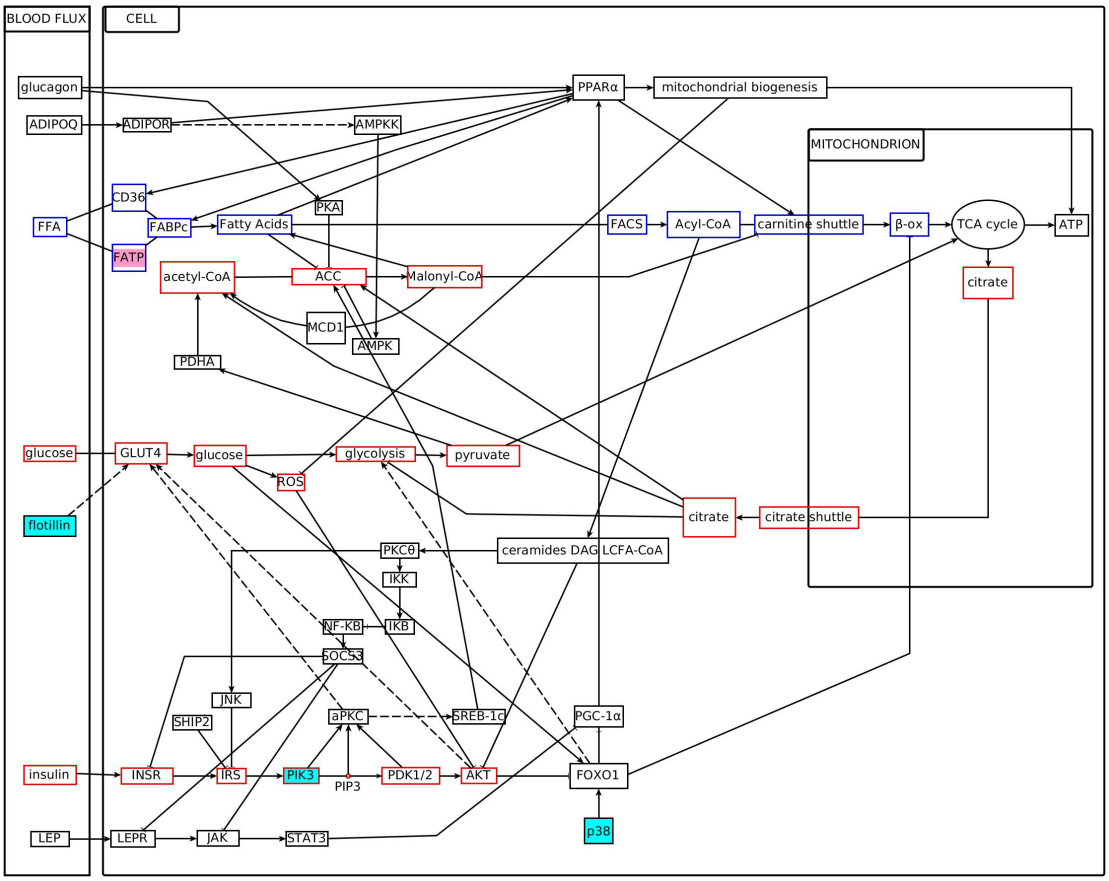
In the simple version we identify 3 core cellular steps that could summarize the whole process: Fatty Acids intake, Glucose intake and the Carnitine shuttle, through which FA enter the TCA cycle in the mitochondrion.

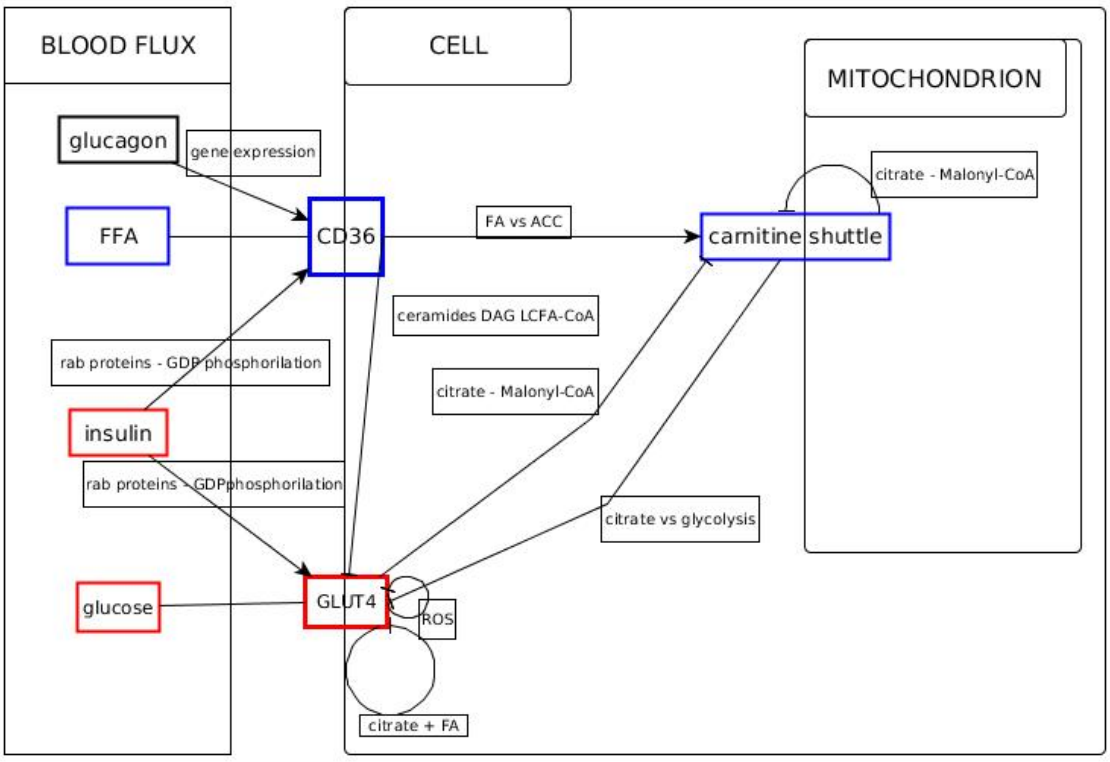
We plotted the results of the 2 way ANOVA used to determine metabolic flexibility differences between Healthy and Insulin Resistant in skeletal muscle on the complete pathway.

Filled cells represent significant genes, i.e. genes that have a different expression between the fasting and postprandial phase, that can be imputed to the different health state.

We considered the absolute difference of the average value of gene expression between the fasting and postprandial phase.

Cyan filled cells refer to genes that have a lower average difference in Insulin Resistant subjects than in Healthy, while pink filled cells correspond to genes for whom such difference is higher in IR.


BLOOD FLUX
CELL
glucagon
ADIPOQ
ADIPOR
AMPKK
PPARα
mitochondrial biogenesis
MITOCHONDRION
CD36
FFA
FABPc
Fatty Acids
PKA
FACS
Acyl-CoA
carnitine shuttle
β-ox
TCA cycle
ATP
FATP
acetyl-CoA
ACC
Malonyl-CoA
citrate
MCD1
AMPK
PDHA
glucose
GLUT4
glucose
glycolysis
pyruvate
ROS
flotillin
citrate
citrate shuttle
PKCθ
ceramides DAG LCFA-CoA
IKK
NF-KB
IKB
SOCS3
JNK
SHIP2
aPKC
SREB-1C
PGC-1α
insulin
INSR
IRS
PIK3
PIP3
PDK1/2
AKT
FOXO1
LEP
LEPR
JAK
STAT3
p38


BLOOD FLUX
CELL
MITOCHONDRION
glucagon
gene expression
citrate - Malonyl-CoA
FFA
CD36
FA vs ACC
carnitine shuttle
ceramides DAG LCFA-CoA
rab proteins - GDP phosphorilation
citrate - Malonyl-CoA
insulin
rab proteins - GDPphosphorilation
citrate vs glycolysis
glucose
GLUT4
ROS
citrate + FA

An interesting result is the lower variation of PIK3 in IR between the fasting and postprandial phase. This is a key gene of the insulin signaling pathway, and the result thus agrees with an impaired metabolic flexibility in IR subjects.

2.8 Ecological modeling of Gut Microbiota

2.8.1 PREPROCESSING

We analysed the data from [**11**], available on MG-RAST under the Project ID 154. These include 164 elderly subjects, non-antibiotic-treated, for whom we also had dietary information, and stratified by community residence setting: (1) community-dwelling, n=81; (2) attending an out-patient day hospital, n=20; (3) in short- term (<6 weeks) rehabilitation hospital care, n=12; (4) in long-term residential care (long-stay), n=51. The mean subject age was 78 (±8 s.d.) years, with a range of 64 to 102 years, and all were of Caucasian (Irish) ethnicity. The study also includes 13 young adults with a mean age of 36 (66 s.d.) years. For each subject, sequence reads from 16S rRNA gene V4 amplicons were generated with 454 Genome Sequencer FLX Titanium platform. These are suitable data for our analysis because starting from 16S rRNA gene (V4) data we can build OTUs, i.e. assess phylogenetic relationships between species.

Starting from the 16S rRNA sequences, we built the RSA distribution of the GUT ecological system computing the so called Operational Taxonomic Units (OTUs) through a clustering procedure based on sequence similarity. In particular the RSA was obtained representing the OTUs abundances in the form of Preston plot. This is the plot of how many species (y-axes) have a certain number of individuals (x-axis), with the x-axis transformed in logarithm to base 2 in order to compress the information of the otherwise very long tail of the distribution.

Before deepening the methodology, let us underline that in order to give an ecological description of the microbiota, it is highly recommended to base the analysis on de novo OTUs, rather then on some taxonomic classification. In fact, although phylotype-based methods are appealing approaches, since they enable investigators to place labels onto sequences, indicating their relationships to previously cultured and characterized

microbes, they are human-made methods and there are myriad examples of organisms that belong to the same species that have different phenotypes and organisms with the same phenotype belonging to different taxonomic lineages, without talking about unclassified organisms or about the fact that there are at least three different curated taxonomy outlines that contain significant conflicts with each other [**12**]. In order to estimate evolutionary relationships of organisms, genes, species, etc. the tool to be used is phylogenetic analysis, that is the analysis of OTUs [**13**]. Basing our analysis on OTUs, we will not be able to give names to species, but we will have the great advantage of avoiding the loss of information due to a taxonomic classification.

In order to build OTUs, we clustered the 16S rRNA sequences with UCLUST [**14**]. Sequences were previously sorted by quality, since UCLUST is sequence order sensitive and needs the highest quality sequences at the beginning of the input file. We applied the UCLUST algorithm using four different similarity thresholds: 90%, 93%, 95% and 97%. In this way, we obtained clusters (OTUs), that can be thought as groups of bacteria of the same taxon at a particular phylogenetic level [**12**]. Here, the similarity threshold is what defines the phylogenetic level at which we compute the RSA, the scale level at which we study the ecosystem.

Starting from UCLUST results, we estimated the OTUs abundances, that will correspond to species abundances. We then filtered for singletons (OTUs with just one sequence inside) in order to minimize the inclusion of sequencing artefacts [**15**]. Since the absence of singletons violates some fundamental assumptions of species richness analysis, we randomly removed one sequence per OTU, causing all OTUs with originally two reads (doubletons) to become singletons, those with three reads to become doubletons, and so on. The effect of an even more conservative interpretation of 454 reads on species richness analysis was evaluated by consecutively omitting doubletons and tripletons from the initial data sets. Thus, we computed the RSA in the form of Preston plot for all four data sets: singletons retained, singletons excluded, doubletons excluded and tripletons excluded. The greatest difference appears between retaining or removing singletons from the data set (see Fig.1). Thus, considering the less restrictive case, we chose to exclude just singletons, and to subtract 1 from all the other OTUs abundances.

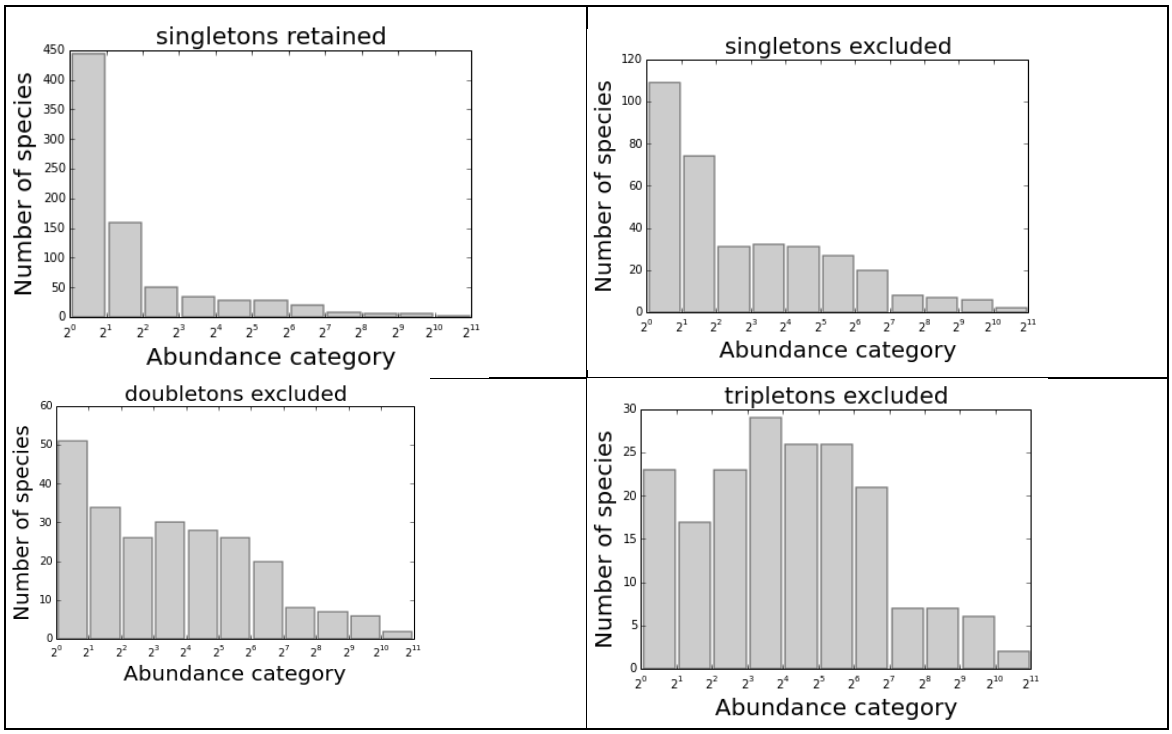


Figure 1. Preston plot of the RSA distribution of one GM sample at 97% similarity thresholds for the four data sets: singletons retained, singletons excluded, doubletons excluded and tripletons excluded.

2.8.2 MODELING: PURELY NEUTRAL MODEL

The GM RSA was modelled according to the neutral theory of ecosystems proposed by [**16**] and [**17**]. The dynamics of the population of a single species are governed by generalized birth and death events, that include speciation, immigration and emigration. If we neglect inter-species interactions after the community has reached a steady state, the number of individuals *n* of a given species evolves according to

$$\frac{dn}{dt} = b_n \cdot n - d_n \cdot n + S \quad (1)$$

where bn and dn denote the per-capita density-independent birth and death rates, with b−1 = 0 and d0 = 0, while the presence of the constant influx S produce a density dependence effect, which causes a rare species advantage and which can arise due to effective rates of immigration/emmigration/speciation/extinction in a local community, so as due to intraspecific interactions. Here, we do not incorporate speciation explicitly

into the model because it does not affect the functional form of the results (it can be incorporated into the immigration term at $n = 0$ by adding a constant).

Let us observe that, since we make the neutral equivalence of species assumption, this equation, that has been written for a single species, actually holds for the whole ecosystem, since the per-capita birth and death rates are the same for all the species.

In order to take into account stochasticity, that is fluctuations, we can write the corresponding Chemical Master Equation

$$\frac{\partial P_n(t)}{\partial t} = P_{n-1}(t)b_{n-1} + P_{n+1}(t)d_{n+1} - P_n(b_n + d_n) \quad (2)$$

where S has been included in the birth term, setting

$$b_n = b \cdot (n + \Upsilon) \quad (3)$$

that is

$$S = \Upsilon \cdot b \quad (4)$$

while the death rate simply results

$$d_n = d \cdot n \quad (5)$$

The stationary solution is easily obtained with the linear expansion method, in which the birth-death process null flux condition is exploited:

$$P_n = P_0 \prod_{i=0}^{n-1} \frac{b_i}{d_{i+1}} = P_0 \prod_{i=0}^{n-1} \frac{b \cdot (\frac{S}{b} + i)}{d \cdot (i+1)} \quad (6)$$

Here we consider $n > 0$ and we can deduce P_0 from the normalization condition $\sum_{n \geq 0} P_n = 1$.

The stationary solution results a Negative Binomial distribution

$$P_{RSA} = \frac{(1-\frac{b}{d})^{S/b}(\frac{b}{d})^{n}}{\Gamma(S/b)}\frac{}{n!}\Gamma(n+\frac{S}{b}) \quad (7)$$

The interesting thing about this distribution is that it is able to resemble both a Log-Series and a Log-Normal, the two RSA distribution widely observed and discussed in ecology [**18**].

2.8.3 MODELING: MODIFIED CHEMOSTAT 2D MODEL

A first insight on GM Preston plots showed that the RSA for this ecosystem is actually bimodal. This implied the need for a relaxation of the neutrality condition in favour of an hybrid niche-neutral model which allowed at least 2 niches, meaning 2 groups of bacteria with different dynamics parameters, but inside which neutrality still holds.

A similar solution had been proposed for the coral-reefs ecosystem in [**19**]. Here we propose a new model based on the chemostat, that is able to explain such situation with a competition for the same nutrient between two species, and that seems particularly suitable for bacteria in the Gut.

A chemostat is a bioreactor that consists of a vessel filled with culture medium and, in our case, bacteria. In the chemostat, we also supply a constant influx of medium (nutrient) and the vessel owns an outlet, so that nutrients and bacteria in the container can flow out.

This system can well model the microbiota: the vessel will be the host gut, the influx will be the ingested food (diet) and the efflux will be due to the host digestion and discard.

The deterministic equations for a one dimensional chemostat model, i.e. a chemostat with just one bacteria population *b* fed with nutrient *n*, are

$$\begin{aligned} \frac{db}{dt} &= b \cdot \frac{k_{max} n}{K_n + n} - D \cdot b \\ \frac{dn}{dt} &= n_0 - D_n n - \beta \cdot b \cdot \frac{k_{max} n}{K_n + n} \end{aligned} \quad (8)$$

where, *b* refers to the bacteria population measure, *n* to that of the nutrient, *D* corresponds to the bacteria death rate, D_n to the nutrient dilution rate, k_{max} is the Michaelis-Menten maximum growth rate, K_n is the half-saturation constant and $1/\beta$ is the yield.

So, the bacteria population abundance *b*, can change with time because some new individual is born (and this is related with some nutrient consumption), that is the first term of the sum, or because someone dies, as made explicit by the second term of the sum. For what concern the nutrient, instead, we have a constant influx n_0 (e.g. diet), a constant efflux $-D_n$ x *n* (e.g., food absorbed by our body or discard), and an efflux due to the bacteria consumption, that is the last term.

We can expand this chemostat model in two dimensions, considering two bacteria population b_1 and b_2, competing for the same nutrient *n*.

$$\begin{aligned}\frac{db_1}{dt} &= b_1\frac{k_{max1}n}{K_{n1}+n} - D_1 b_1 \\ \frac{db_2}{dt} &= b_2\frac{k_{max2}n}{K_{n2}+n} - D_2 b_2 \\ \frac{dn}{dt} &= n_0 - D_n n - \beta_1 b_1\frac{k_{max1}n}{K_{n1}+n} - \beta_2 b_2\frac{k_{max2}n}{K_{n2}+n}\end{aligned} \quad (9)$$

The competition for the same nutrient yields a coupling of the bacterias equations: for example, *n* partecipates to the growth term of b_1, but *n* depdens on both b_1 and b_2, thus the dyanamics of b_1 will depend also on b_2.

Moreover, following the idea in [**16**] and [**17**], we added to this model a further constant influx term, a density dependence effect that does not depend on the bacteria population abundance *b*.

$$\begin{aligned}\frac{db_1}{dt} &= b_1\frac{k_{max1}n}{K_{n1}+n} - D_1 b_1 + \mathcal{S}_1 \\ \frac{db_2}{dt} &= b_2\frac{k_{max2}n}{K_{n2}+n} - D_2 b_2 + \mathcal{S}_2 \\ \frac{dn}{dt} &= n_0 - D_n n - \beta_1 b_1\frac{k_{max1}n}{K_{n1}+n} - \beta_2 b_2\frac{k_{max2}n}{K_{n2}+n}\end{aligned} \quad (10)$$

The stochastic version of this model can be analytically solved under the nutrient saturation condition. In this case, the two differential equations for the bacteria populations b_1 and b_2 becomes uncoupled, being the nutrient n constant, and the solution will be a mixture of the distributions for b_1 and b_2, that will be two Negative Binomials, as for the purely neutral case of [**16**] and [**17**].

2.8.4 STOCHASTIC SIMULATIONS

The model was simulated stochastically with the Gillespie algorithm implemented in python. The simulation showed two main situations.

When K_{n1} and K_{n2} are little compare with n, the terms $\frac{k_{max1}n}{K_{n1}+n}$ and $\frac{k_{max2}n}{K_{n2}+n}$ approximate to k_{max1} and k_{max2}. In this situation the interaction due to nutrient competition has a little effect and simulations show two skewed distribution for the b_1 and b_2 populations (Fig. 2) that are well fitted by a Negative Binomial distribution, that is the continuous equivalent of the negative binomial distribution.

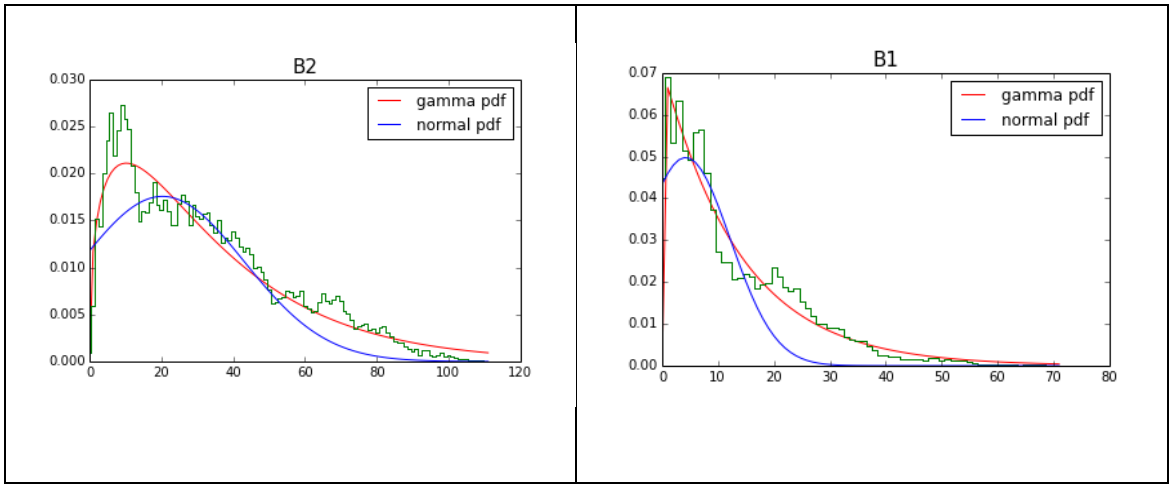


Figure 2. Simulation (green) of the chemostat 2D model with influx terms for two bacteria populations, with half-saturation constants K_{n1} = 0.001 and K_{n2} = 0.001 and an equilibrium value for the nutrient n ~ 40. The red and blue curves have been obtained with a least squares fitting respectively with a gamma and a normal distribution.

With higher K_{n1} and K_{n2} the 2 bacteria distributions becomes instead more normal and symmetric, but are again well fitted by the gamma (i.e. Negative Binomial) distribution (Fig. 3).

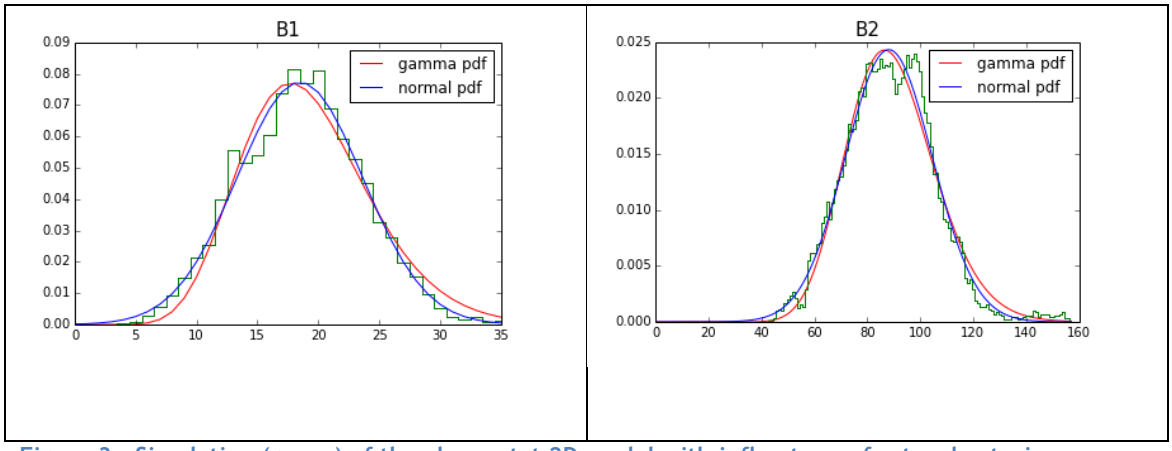


Figure 3. Simulation (green) of the chemostat 2D model with influx terms for two bacteria populations, with half-saturation constants K_{n1} = 10 and K_{n2} = 100 and an equilibrium value for the nutrient n ~ 60. The red and blue curves have been obtained with a least squares fitting respectively with a gamma and a normal distribution.

2.8.5 FIT WITH A MIXTURE OF 2 NEGATIVE BINOMIALS

The experimental RSA distributions have been fitted with a mixture of two Negative Binomials, one describing rarest species and the other describing the most abundant:

$$P_{RSA} = a \cdot \frac{(1-\theta_1)^{k_1}(\theta_1)^n}{\Gamma(k_1)}\frac{}{n!}\Gamma(n+k_1) + (1-a)\cdot\frac{(1-\theta_2)^{k_2}(\theta_2)^n}{\Gamma(k_2)}\frac{}{n!}\Gamma(n+k_2)$$

with $k_{1,2} = S_{1,2}/b_{1,2}$ and $\theta_{1,2} = b_{1,2}/d_{1,2}$.

As already pointed out, because of its very long tail, the RSA is represented in the form of Preston plot, i.e. with a logarithm to base 2 x-axis. In this way, each bin actually contains the sum of the number of species with abundance category between its minimum and maximum (e.g. bin $[2^2,2^3)$ represents the number of species having 4, 5, 6 or 7 individuals). Thus, the fit was computed considering the integral of the Negative Binomial distribution from the minimum to the maximum of each bin, that can be simply obtained subtracting the Negative Binomial cumulative in the minimum from the Negative Binomial cumulative in the maximum of the bin.

For what concern the Negative Binomial for the rarest species, we noted that the parameter $\theta = b/d \rightarrow 0$, making the Negative Binomial tend to a Poisson distribution. This limit resulted ambiguous for the fit computation, thus we used a different parametrization, fitting the rare Negative Binomials in term of $k = S/b$ and its mean $\mu = k\theta / (1 - \theta)$.

The fit worked well for all samples. In particular at 97% similarity threshold, that is the value we focused on for further analysis, the minimum R-squared was 0.891.

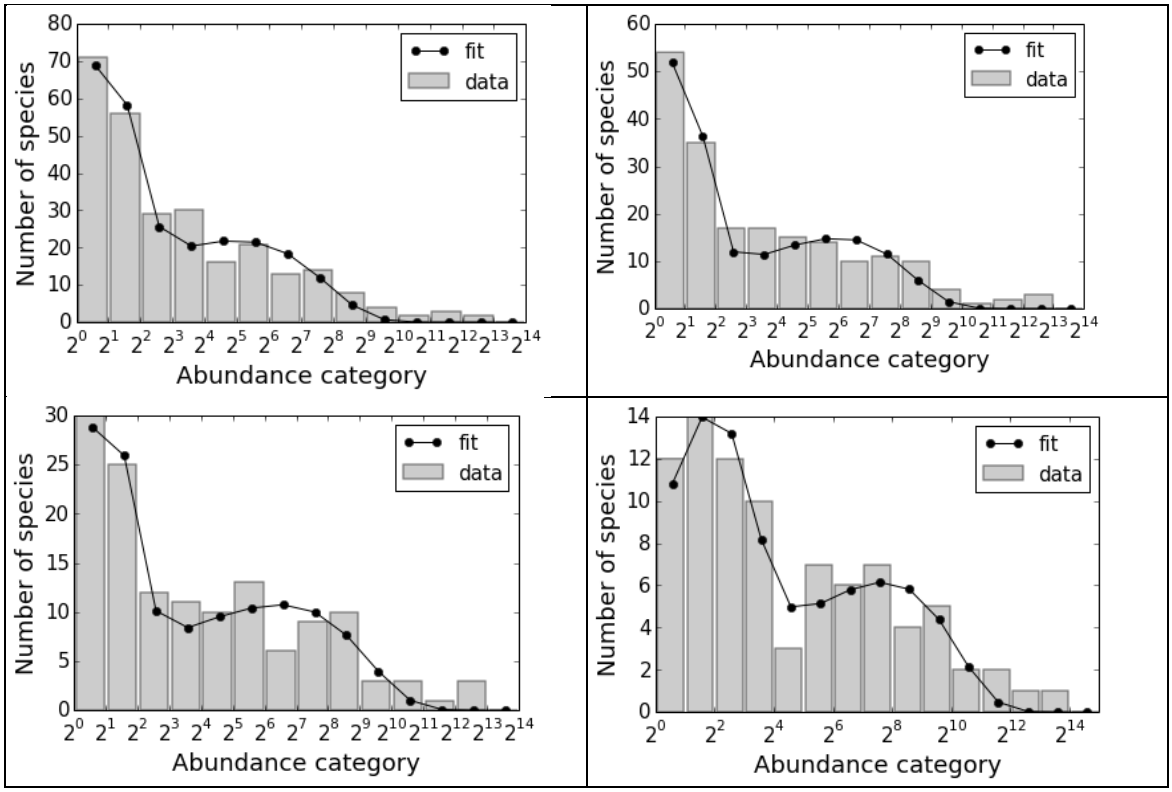


Figure 4. Fit with a mixture of 2 Negative Binomials of the RSA of one sample from [11] at different similarity thresholds: 97% (top-left), 95% (top-right), 93% (bottom-left) and 90% (bottom-right).

We focused on 97% similarity threshold, that is the common value used for taxonomy analysis, and we grouped the samples according to their age range and diet group. For both these comparisons we computed the mean and standard error of the fit parameters for all the groups. Results are shown in Figure 5 and Figure 6.

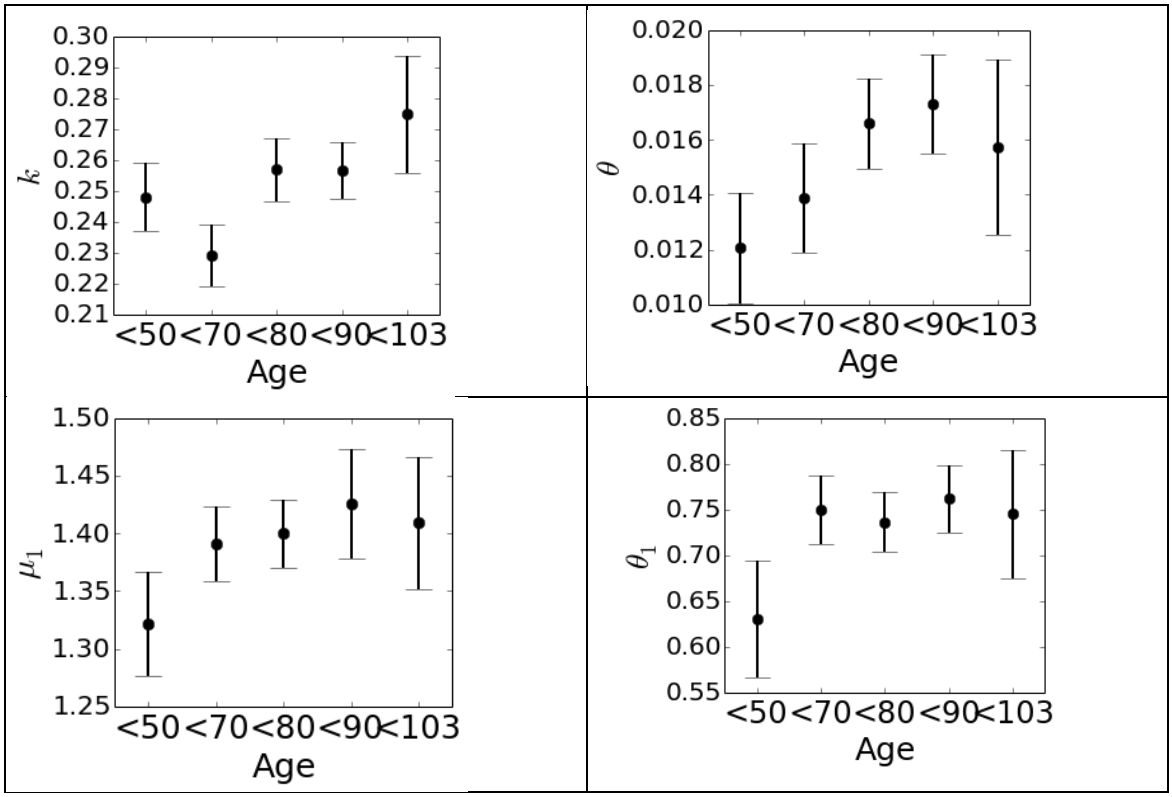


Figure 5. Mean fit parameters for 5 age groups. Top figures correspond to the parameters of the Negative Binomial describing abundant species, while bottom figures refer to the rarest species Negative Binomial.

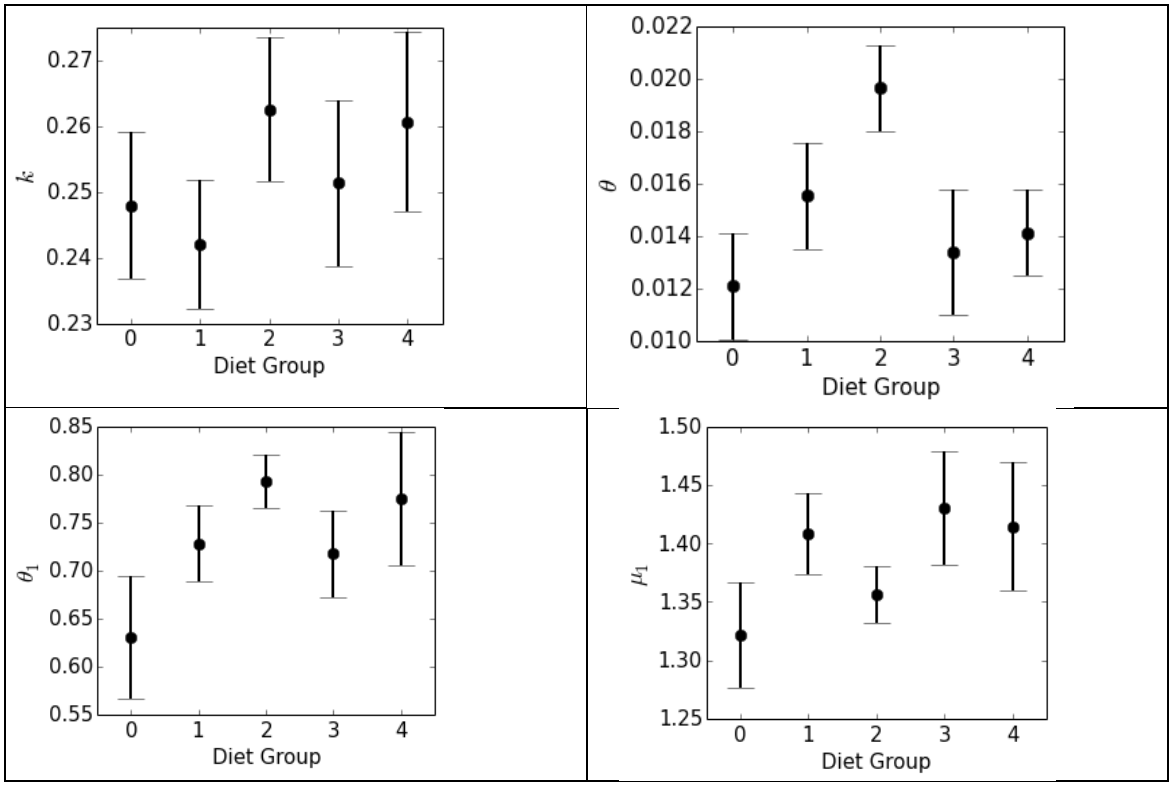


Figure 6. Figure 5. Mean fit parameters for the four diet groups (from 1 to 4); diet 0 means Young control samples, for which no diet information was available. Top figures correspond to the parameters of the Negative Binomial describing abundant species, while bottom figures refer to the rarest species Negative Binomial.

2.9 EFFECT OF DIET COMPONENTS ON GM BACTERIA

Basing mainly on the work of Wu et al. [20] and the review of Scott et al. [21], we defined a list of important GM bacteria and assessed how different diet component affect their abundance in the Gut.

In the following tables, symbol "+" means a positive correlation between the bacteria and the corresponding nutrient, while "-" stands for negative correlation. If no correlation have been observed in these works, no symbol is included.

BACTERIA	PUFA	SATURATED FATTY ACIDS	CLA FATTY ACIDS	TRANS FATTY ACIDS	VEGETABLE PROTEINS	ANIMAL PROTEIN	TOTAL PROTEINS	PREBIOTICS	SUGARS	TOTAL CARBOHYDRATES	FIBER
Bacteria.Actinobacteria											-
Bacteria.Actinobacteria.Actinobacteria.Actinomycetales.Actinomyces				+							-
Bacteria.Actinobacteria.Actinobacteria.Actinobacteridae.Bifidobacteriales.Bifidobacteriaceae.Bifidobacterium					+	-		+			- +
Bacteria.Actinobacteria.Actinobacteria.Actinobacteridae.Bifidobacteriales.Bifidobacteriaceae.Bifidobacterium.Bifidobacterium adolescentis						-					
Bacteria.Actinobacteria.Actinobacteria.Coriobacteriales							+				
Bacteria.Actinobacteria.Actinobacteria.Coriobacteriales.Coriobacteriaceae							+				
Bacteria.Bacteroidetes		+		+							
Bacteria.Bacteroidetes.Bacteroidia		+	+	+							
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales		+	+	+							
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Bacteroidaceae	-	+	+	+		+	+		-	-	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Bacteroidaceae.Bacteroides	-	+	+	+		+	+	-	-	-	
Bacteria.Bacteroidetes.Bacteroidetes.Bacteroidales.Bacteroidaceae.Bacteroides thetaiotamicron											+
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae	+	-				-	-				-
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae.Paraprevotella		-					-		+		
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae.Prevotella	+						-	-			
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae		+	+	+	-				-		-
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Barnesiella									-		
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Butyricimonas											
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Odoribacter		+					+				-
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Parabacteroides		+	+		-					-	-
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Parabacteroides diastonis											
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Rikenellaceae											-
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Rikenellaceae.Alistipes											-

BACTERIA	PUFA	SATURATED FATTY ACIDS	CLA FATTY ACIDS	TRANS FATTY ACIDS	VEGETABLE PROTEINS	ANIMAL PROTEIN	TOTAL PROTEINS	PREBIOTICS	SUGARS	TOTAL CARBOHYDRATES	FIBER
Bacteria.Firmicutes			-	-							
Bacteria.Firmicutes.Bacilli									+	+	
Bacteria.Firmicutes.Bacilli.Lactobacillales									+	+	
Bacteria.Firmicutes.Bacilli.Lactobacillales.Lactobacillaceae											
Bacteria.Firmicutes.Bacilli.Lactobacillales.Lactobacillaceae.Lactobacillus					+	-		+			
Bacteria.Firmicutes.Bacilli.Lactobacillales.Lactobacillaceae.Lactobacillus ruminis											+
Bacteria.Firmicutes.Bacilli.Lactobacillales.Streptococcaceae									+	+	
Bacteria.Firmicutes.Bacilli.Lactobacillales.Streptococcaceae.Streptococcus									+	+	
Bacteria.Firmicutes.Clostridia			-	-		+					
Bacteria.Firmicutes.Clostridia.Clostridiales			-	-							
Bacteria.Firmicutes.Clostridia.Clostridiales.Acidaminococcaceae.Phascolarctobacterium				-							+
Bacteria.Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Anaerotruncus					-				+		
Bacteria.Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Anaerotruncus.Faecalibacterium prausnitzii								+			+
Bacteria.Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Clostridium											
Bacteria.Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Clostridium.Clostridium histolyticum								-			
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIV									-		
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIV.Blautia									-		
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIII			-				-				
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIII.Anaerovorax					-						-
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae				-	+						+
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae.Coprococcus			-								
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae.Roseburia					+	-		+			+
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae.Dorea									-		
Bacteria.Firmicutes.Clostridia.Clostridiales.Eubacteriaceae											-
Bacteria.Firmicutes.Clostridia.Clostridiales.Eubacteriaceae.Eubacterium.Eubacterium rectale						-		+			+
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae			-				-				
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Butyricicoccus		-	-	-	+						+
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Oscillibacter							-		+	+	+
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Ruminococcus.Ruminococcus bromii											+
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Subdoligranulum			-								
Bacteria.Firmicutes.Clostridia.Clostridiales.Veillonellaceae.Megasphaera											-
Bacteria.Firmicutes.Erysipelotrichi		+								-	
Bacteria.Firmicutes.Erysipelotrichi.Erysipelotrichales		+								-	
Bacteria.Firmicutes.Erysipelotrichi.Erysipelotrichales.Erysipelotrichaceae		+								-	
Bacteria.Firmicutes.Erysipelotrichi.Erysipelotrichales.Erysipelotrichaceae.Coprobacillus		+								-	
Bacteria.Firmicutes.Erysipelotrichi.Erysipelotrichales.Erysipelotrichaceae.Holdemania					-		+				-
Bacteria.Firmicutes.Erysipelotrichi.Erysipelotrichales.Erysipelotrichaceae.Turicibacter									-		
Bacteria.Firmicutes.Negativicutes.Selenomonadales.Acidaminococcaceae.Acidaminococcus											-
Bacteria.Fusobacteria					-						
Bacteria.Fusobacteria.Fusobacteria					-						
Bacteria.Fusobacteria.Fusobacteria.Fusobacteriales				+	-						-
Bacteria.Proteobacteria		-		-	+						+
Bacteria.Proteobacteria.Betaproteobacteria		-			+						+
Bacteria.Proteobacteria.Betaproteobacteria.Burkholderiales		-									+
Bacteria.Proteobacteria.Betaproteobacteria.Burkholderiales.Alcaligenaceae.Sutterella									+	+	
Bacteria.Proteobacteria.Betaproteobacteria.Burkholderiales.Alcaligenaceae.Parasutterella									-		
Bacteria.Proteobacteria.Gammaproteobacteria							+				-
Bacteria.TM7			+								

2.10 GM BACTERIA MAIN PRODUCTS

Finally, the main products of the considered bacteria have been report in the following tables, with particular attention to short-chain fatty acids.

BACTERIA	PRODUCTS	OTHER INFO
Bacteria.Actinobacteria		
Bacteria.Actinobacteria.Actinobacteria.Actinomycetales.Actinomyces		pathogenic bacteria
Bacteria.Actinobacteria.Actinobacteria.Actinobacteridae.Bifidobacteriales.Bifidobacteriaceae.Bifidobacterium	acetate, butyrate, ammonia, molecular hydrogen	control intestinal pH and intestinal microbial homeostasis, inhibit pathogens, modulate the local and systemic immune responses, repress procarcinogenic enzymatic activities within the microbiota, produce SCFAs, produce vitamins, and bioconvert a number of dietary compounds into bioactive molecules.
Bacteria.Actinobacteria.Actinobacteria.Actinobacteridae.Bifidobacteriales.Bifidobacteriaceae.Bifidobacterium.Bifidobacterium adolescentis	SCFAs (propionate, butyrate, and acetate), lactate, vitamins, cyanocobalamin, nicotine, thiamin, folic acid, pyridoxine	
Bacteria.Actinobacteria.Actinobacteria.Coriobacteriales		
Bacteria.Actinobacteria.Actinobacteria.Coriobacteriales.Coriobacteriaceae		enriched in obese
Bacteria.Bacteroidetes		gram negative, decrease in obesity
Bacteria.Bacteroidetes.Bacteroidia		
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales		
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Bacteroidaceae		
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Bacteroidaceae.Bacteroides	acetic acid, iso valeric acid, and succinic acid	Gram-negative, pathogenic bacteira
Bacteria.Bacteroidetes.Bacteroidetes.Bacteroidales.Bacteroidaceae.Bacteroides thetaiotamicron	acetic acid, iso valeric acid, and succinic acid	Gram-negative, pathogenic bacteira
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae	SCFAs (Acetate, propionate), succinate	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae.Paraprevotella	succinate, acetate	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae.Prevotella	SCFAs	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae		
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Barnesiella	succinate, acetate	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Butyricimonas	butyrate, isobutyrate, succinate, acetate, propionate	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Odoribacter	Succinate, acetate, iso-valeric acid	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Parabacteroides	acetic acid,succinic acids,isovaleric acid, propionic acid, isobutyrate,formic acid, lactic acid, catalase, Mannose, raffinose, L-Rhamnose, Trehalose	
oides diastonis	catalase	
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Rikenellaceae		
Bacteria.Bacteroidetes.Bacteroidia.Bacteroidales.Rikenellaceae.Alistipes	succinate, acetate, mannose, raffinose, propionic acid	

BACTERIA	PRODUCTS	OTHER INFO
Bacteria.Firmicutes		increase with obesity
Bacteria.Firmicutes.Bacilli		
Bacteria.Firmicutes.Bacilli.Lactobacillales		
Bacteria.Firmicutes.Bacilli.Lactobacillales.Lactobacillaceae		
Bacteria.Firmicutes.Bacilli.Lactobacillales.Lactobacillaceae.Lactobacillus	lactate	help maintain the pH level
Bacteria.Firmicutes.Bacilli.Lactobacillales.Lactobacillaceae.Lactobacillus.Lactobacillus ruminis	lactate	help maintain the pH level
Bacteria.Firmicutes.Bacilli.Lactobacillales.Streptococcaceae		
Bacteria.Firmicutes.Bacilli.Lactobacillales.Streptococcaceae.Streptococcus	lactate	pathogenic bacteria
Bacteria.Firmicutes.Clostridia		
Bacteria.Firmicutes.Clostridia.Clostridiales		
erium	propionate	gram negative, succinate consumer
Bacteria.Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Anaerotruncus		
acterium prausnitzii	formate, D-lactate, butyrate	
Bacteria.Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Clostridium	butyrate	
histolyticum	organic solvents such as butanol	
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIV		
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIV.Blautia	acetate, ethanol, lactate, isobutyrate, isovalerate	
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIII		
Bacteria.Firmicutes.Clostridia.Clostridiales.Incertae_Sedis_XIII.Anaerovorax	acetate, butyrate, ammonia, molecular hydrogen	
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae	SCFA, butyrate, gas	
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae.Coprococcus	butyrate	
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae.Roseburia	butyrate	
Bacteria.Firmicutes.Clostridia.Clostridiales.Lachnospiraceae.Dorea		
Bacteria.Firmicutes.Clostridia.Clostridiales.Eubacteriaceae	SCFA	
um rectale	butyrate	
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae	SCFA	
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Butyricicoccus	butyrate	
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Oscillibacter	valerate	
inococcus bromii	ethanol, acetate, formic acids	
Bacteria.Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Subdoligranulum	butyrate, formic acid	
Bacteria.Firmicutes.Clostridia.Clostridiales.Veillonellaceae.Megasphaera	Butyrate, valerate, propionate, acetate, caproate	
Bacteria.Firmicutes.Erysipelotrichi		
Bacteria.Firmicutes.Erysipelotrichi.Erysipelotrichales		
Bacteria.Firmicutes.Erysipelotrichi.Erysipelotrichales.Erysipelotrichaceae		
cillus		
ania	acetate, succinate, lactate	
cter		
nococcus	propionate, acetate, butyrate	
Bacteria.Fusobacteria		
Bacteria.Fusobacteria.Fusobacteria		
Bacteria.Fusobacteria.Fusobacteria.Fusobacteriales	butyrate, acetate	
Bacteria.Proteobacteria		Gram-negative
Bacteria.Proteobacteria.Betaproteobacteria	nitrite, sulfate	gram negative
Bacteria.Proteobacteria.Betaproteobacteria.Burkholderiales		gram negative
la		Gram-negative, pathogenic bacteria
terella		gram negative
Bacteria.Proteobacteria.Gammaproteobacteria	sulfur	Gram-negative, pathogenic bacteria
Bacteria.TM7		related to inflammation
Bacteria.TM7.TM7_incertae_sedis		

3 Deliverable Conclusions

We created a metabolic flexibility pathway for skeletal muscle cell and we exploited it to assess the effects of Insulin Resistance on the impairment of such switch in skeletal muscle cells. This was performed basing on gene expression data from the GEO database, that also enabled us to evaluate the effects of IR and T2D on liver and adipose tissue. Several pathways resulted significantly different in all these 3 tissues

between healthy and unhealthy patients. Specifically, in skeletal muscle a significant gene was PIK3, that is a key gene in the insulin signaling pathway, so as in our metabolic flexibility pathway. In liver, a particularly interesting gene that changed significantly was TCF7.

We created a new stochastic model for the GM RSA, relaxing the neutral hypothesis previously proposed. The model was computed starting from a chemostat model for 2 populations competing for the same nutrient. A constant influx term corresponding to immigration, speciation or a density dependence effect was added. The model was solved analytically, when possible, and numerically in all other cases. For this purpose a Gillespie algorithm have been implemented in Python. The stationary solution resulted to be a mixture of two Negative Binomials and have been applied on data from Claesson et al. The results show that the model works well and that its parameters can be exploited to discriminate between different ages, diets or health states.

Finally a table summarizing how the main diet components affect GM have been proposed, also adding the information about the principal products arising from them. These products are the means by which diet, through its effects on GM, affects the host metabolism and immune system, contributing to the metabolic flexibility switch impairment that is observed in Insulin Resistant and Type 2 Diabetes subjects.

4 Perspective work

TCF7L2 genetics risk variants are the most robust and the most universal biomarkers associated with T2D. Since its first description in the study from Sladek et al. (Sladek R. et al., Nature, 2007) a number of studies confirmed the association between several TCF7L2 genetic variants and the occurrence of T2D. Recently, Garagnani et al. (Garagnani P. et al., Aging, 2012) performed an association study between TCF7L2 and T2D by using centenarians as super-controls. Centenarians can be considered as model of healthy disease-free longevity, thus representing the gold standard of genetic background. In the study the centenarians resulted significantly enriched of the non-risk allele in respect of general population, considered in the study as classical controls and in respect with patient with T2D, and in particular with the sub-group of patients that affected by T2D complications. In this study we suggested for the first time that, while the TCF7L2 rs7903146 T is a risk allele, the C allele can be considered a longevity one. This perspective was confirmed in a study reviewed in Corella et al. (Corella et al.,

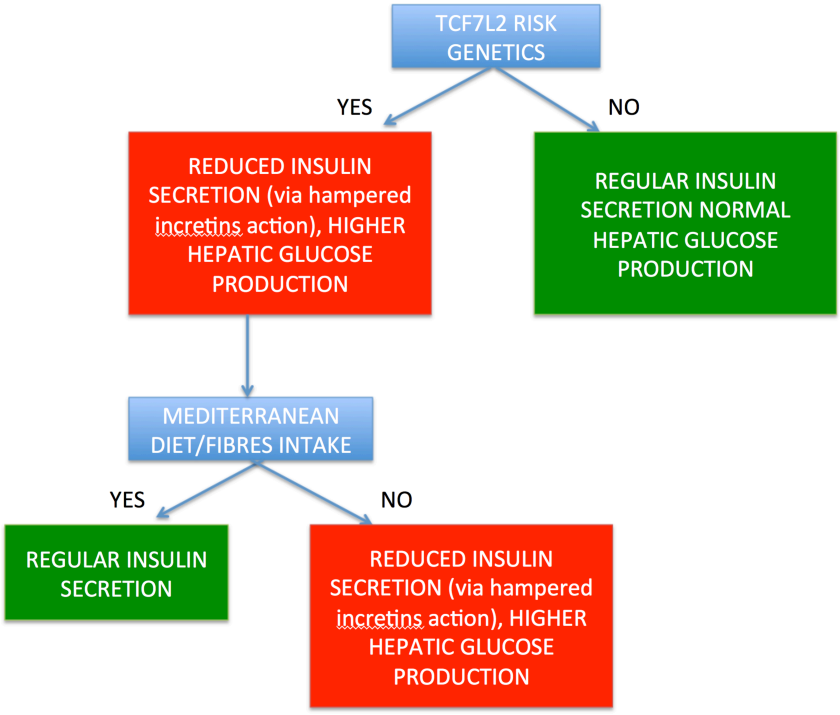
Ageing Research Review, 2014) were in a large cohort they showed that the three rs7903146 genotypes strongly correlate with stroke mortality. In this study they also demonstrate that the effect of the risk allele is completely counteracted by Mediterranean diet.

All these positive results have posed TCF7L2 in the middle of T2D research and a number of study have been conducted in order to elucidate the mechanism that leads TCF7L2 genetic variants to T2D.

Although the exact mechanism is still obscure, a number of studies on human model have elucidated specific effect of TCF7L2 variants on carrier phenotype. In particular a series of studies reviewed in Wagner et Al. (Wagner et al., Molecular Metabolism, 2014) demonstrate that the TCF7L2 rs7903146 reduce the insulin production by hampering the incretin efficacy (Incretins like GLP1 and GIP are secreted after eating and stimulate insulin secretion). This peculiar effect is counteracted by fibres intake (this strongly reminds the above cited study by the Corella et al. in which the effect of the T allele on stroke mortality is cancelled by Mediterranean Diet).

Another phenotypic effect of TCF7L2 gene variants is the increase of liver glucose production. In particular a study from Ling et al. (Ling et al., Journal of Hepatology, 2013) demonstrate that the occurrence of secondary T2D after liver transplantation is strongly associated with the TCF7L2 genetic background of the donor, indicating that the TCF7L2 risk genetics exerts a significant effect on liver physiology leading to a condition that favours the onset of T2D.

All these findings can be summarized in the picture below:



These results are certainly of interest to the MISSION-T2D consortium and indeed partner CNR in collaboration with UniBO is planning to implement the necessary modification to the project's model outcome pertaining the TCF7L2 genetics and to test whether the same conclusions of the above-mentioned study can be reproduced by the model.

More concretely, TCF7L2 genetics can be implemented (as a first approximation) in the MISSION-T2D simulator as follow:

- Represent by a Boolean variable the TCF7L2 genetic risk;
- The high-risk-allele carriers should present a reduced insulin production after calories intake and higher hepatic glucose production. The reduced insulin production can be cancelled by adding qualitative features to the calories intake such as Mediterranean Diet and/or fibres intake;
- Testing the dependency between a different macronutrients distribution intake and the effects on the risk of developing insulin sensitivity might therefore be calculated.

5 Appendix: List of abbreviations used

AT	Adipose Tissue
APC	Antigen-Presenting Cell (immunocompetent cell type)
CD	Cluster of Differentiation (surface markers of lymphocytes)
CTL	Cytotoxic T lymphocyte (immunocompetent cell type)
DC	Dendritic Cell (immunocompetent cell type)
HSP	Heath shock protein (signalling molecule)
IL	Interleukin (signalling molecule)
IFN-γ	Interferon gamma (signalling molecule)
LPS	Lipopolysaccharide (component of pathogen cell walls, signal of danger)
MCP-1	Monocyte Chemotactic Protein-1 (signalling molecule)
MHC	Major Histocompatibility Complex (surface protein of immunocompetent cell types)
MIP-1α	Macrophage Inflammatory Protein-1 alpha (signalling molecule)
T2D	Type 2 Diabetes
TNF-α	Tumor Necrosis Factor alpha (signalling molecule)

Table List of abbreviations

6 Bibliography

1. Jose E. Galgani, Cedric Moro and Eric Ravussin . Metabolic flexibility and insulin resistance . *Am J Physiol Endocrinol Metab 295*:E1009-E1017, 2008.

2. E. Corpeleijn, W. H. M. Saris and E. E. Blaak . Metabolic flexibility in the development of insulin resistance and type 2 diabetes: effects of lifestyle . *Etiology and Pathophysiology* . 2008

3. Jan F. C. Glatz, Joost J. F. P. Luiken and Arend Bonen . Membrane Fatty Acid Transporters as Regulators of Lipid Metabolism: Implications for Metabolic Disease .*Physiol Rev 90*: 367– 417, 2010.

4. Bhaskar Ponugoti, Guangyu Dong, and Dana T. Graves . Role of Forkhead Transcription Factors in Diabetes-Induced Oxidative Stress . *Hindawi Publishing Corporation Experimental Diabetes Research* . Volume 2012, Article ID 939751.

5. Alistair VW Nunn, Jimmy D Bell1 and Geoffrey W Guy. Lifestyle-induced metabolic inflexibility and accelerated ageing syndrome: insulin resistance, friend or foe? *Nutrition & Metabolism* 2009, 6:16

6. David L. Nelson, Micheal M. Cox, I principi di biochimica di Lehninger. *Ed. Zanichelli*. 2010

7. Sears, D. D., et al. "Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization." *Proceedings of the National Academy of Sciences* 106.44 (2009):

18745-18750.

8. Gallagher, Iain J., et al. "Integration of microRNA changes in vivo identifies novel molecular features of muscle insulin resistance in type 2 diabetes." *Genome Med* 2.9 (2010): b117.

9. Hardy, Olga T., et al. "Body mass index-independent inflammation in omental adipose tissue associated with insulin resistance in morbid obesity." *Surgery for Obesity and Related Diseases* 7.1 (2011): 60-67.

10. Misu, Hirofumi, et al. "A liver-derived secretory protein, selenoprotein P, causes insulin resistance." *Cell metabolism* 12.5 (2010): 483-495.

11. Claesson M.J. et al., Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-84 (2012).

12. Schloss P.D. and Westcott S.L. Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. *Applied and Environmental Microbiology* **77**, 3219–3226 (2011)

13. Lee S.H. et al., Embedding operational taxonomic units in three-dimensional space for evolutionary distance relationship in phylogenetic analysis. *WSEAS Transactions on Systems* **6**, 341 (2007).1

14. Edgar R.C., Search and clustering orders of magnitude faster than BLAST.. *Bioinformatics* **26**, 2460-1 (2010).

15. Unterseher M. et al., Species abundance distributions and richness estimations in fungal metagenomics - lessons learned from community ecology. *Molecular Ecology* **20**, 275–285 (2010).

16. Volkov, JR Banavar, SP Hubbell, A Maritan. Patterns of relative species abundance in rainforests and coral reefs.. *Nature* **450**, 45-9 (2007).

17. Sandro Azaele, Simone Pigolotti, Jayanth R. Banavar, Amos Maritan. Dynamical evolution of ecosystems. *Nature* **444**, 926–928 (2006)

18. Stephen P Hubbell. The unified neutral theory of biodiversity and biogeography (MPB-32). **32** (2001).

19. Tang, Junfeng, and Shurong Zhou. Hybrid niche-neutral models outperform an otherwise equivalent neutral model for fitting coral reef data. *Journal of theoretical biology* 317, 212-218 (2013).

20. Wu, Gary D., et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334.6052: 105-108 (2011).

21. Scott, Karen P., et al. The influence of diet on the gut microbiota. *Pharmacological research* 69.1: 52-60 (2013).