



Questions and (partially) answers on The American Gut project

Brusati Alberto 766454

Matamoros Ricardo 807450

Villa Giacomo 807462

Table of Contents



01

The American Gut Project

02

The pipeline

03

Mens sana in corpore sano

04

Life Style

05

A brief focus on diseases

06

Future development

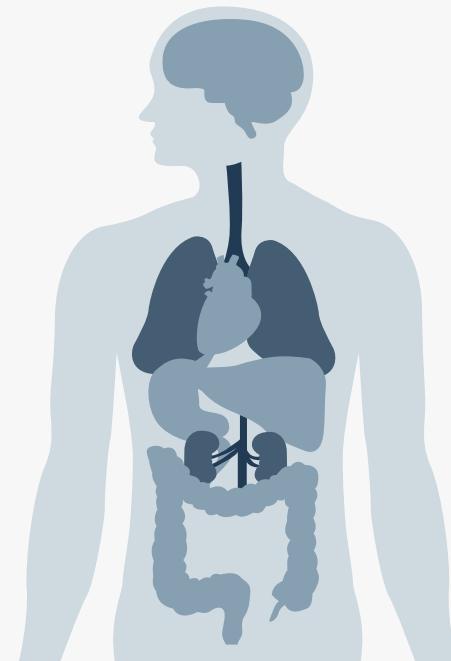
01

The American Gut Project

The American Gut Project

The American Gut Project was co-founded in November 2012 by Rob Knight, Jeff Leach, and Jack Gilbert. The project's goal is to better understand human microbiomes — which types of bacteria live where, how many of each, and how they are influenced by diet, lifestyle and disease.

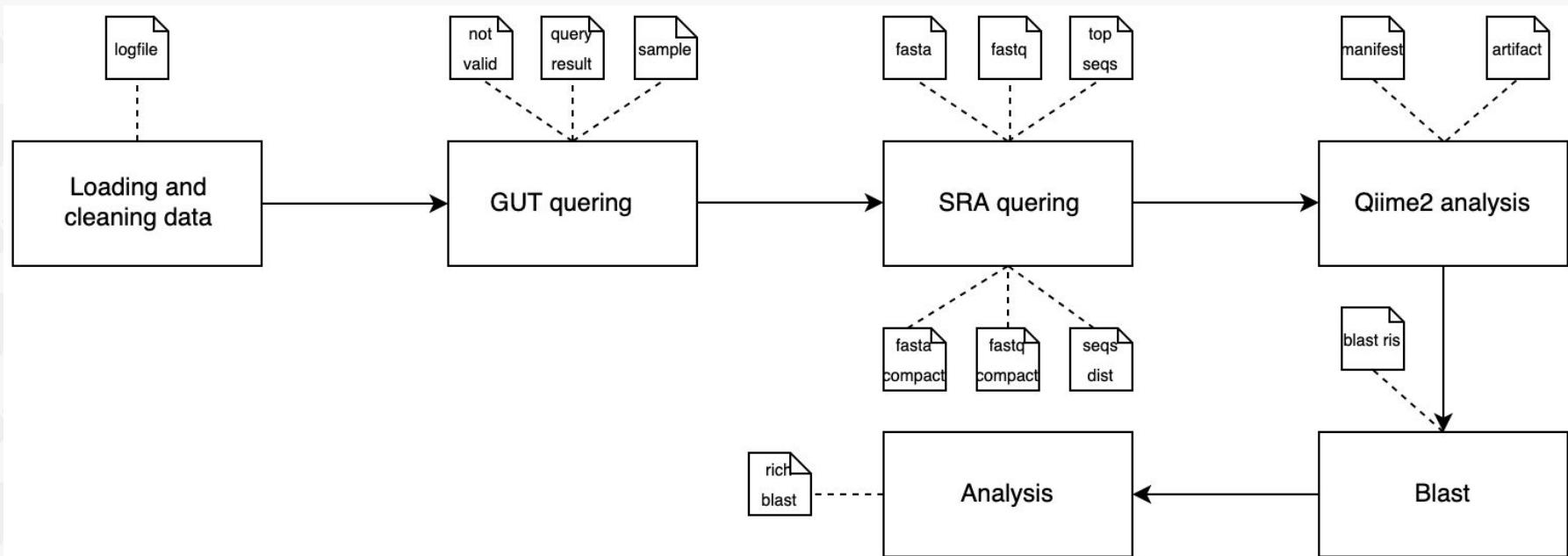
Sample types aren't restricted to fecal, which provides a window into a person's gut microbiome. People can also contribute from skin, oral or other samples.



02

The pipeline

The pipeline



It is drawn in order to enable future and different analysis from gut dataset. The pipeline is split in 3 different notebooks for reasons of readable, they are connected by output file format.

The pipeline - Loading and cleaning data

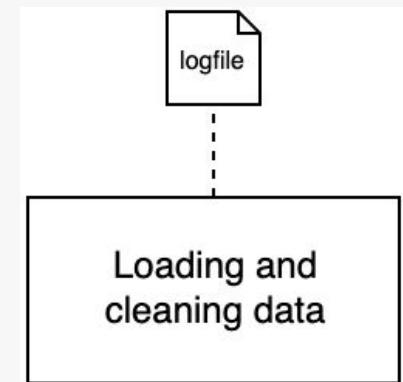
In this phase:

- loading and cleaning the gut data
- Initialize a logfile

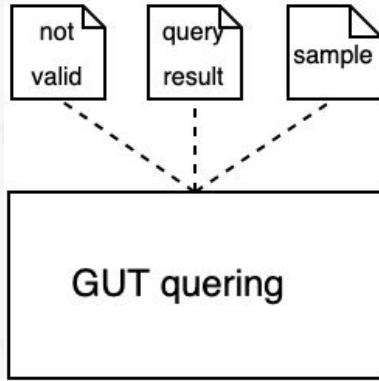
```
RUN TIME: 01/06/2020 10:53:47
-----
-----HEALTHY-----
Total number of healthy people: 615
Total number of healthy man: 244
Total number of healthy woman: 371
Total number of non valid sex 0
Mean Age for total: 36.9545
Mean Age for man: 37.0984
Mean Age for woman: 36.8598

Total rows: 615
Valid rows: 439
Total number of sample people: 40
Total number of sample man: 18
Total number of sample woman: 22
Total number of non valid sex 0
Mean Age for total: 36.8
Mean Age for man: 34.0556
Mean Age for woman: 39.0455
Number of sequences for healthy: 4658723
Number of grouped sequences: 3296480
Number of taken sequences: 731
```

Example of logfile compiling



The pipeline - Gut querying



```
sample_name,typology
10317.000093400,not_healthy_old
10317.000059040,not_healthy_old
10317.000070708,not_healthy_old
10317.000047151,not_healthy_old
10317.000097663,not_healthy_old
10317.000033063,not_healthy_old
10317.000050274,not_healthy_old
```

First line of sample

```
not_valid_sample_name,typology
10317.00009431,mental_illness
10317.000011356,mental_illness
10317.000022548,mental_illness
10317.000027676,mental_illness
10317.000031270,mental_illness
10317.000033765,mental_illness
10317.000039515,mental_illness
```

First line of not valid file

GUT querying and sampling are the core of this phase.

Before sampling operation we validate each sample-names and create a not valid list of them.

The pipeline - SRA querying

Core of the entire pipeline, in this phase the most important file has been generated and others have been updated.

For each experiment typology are generated fast and fastq file for each SRA record on NCBI.

We compact this file, based on typology, and count for take in account only the most recurring sequences.

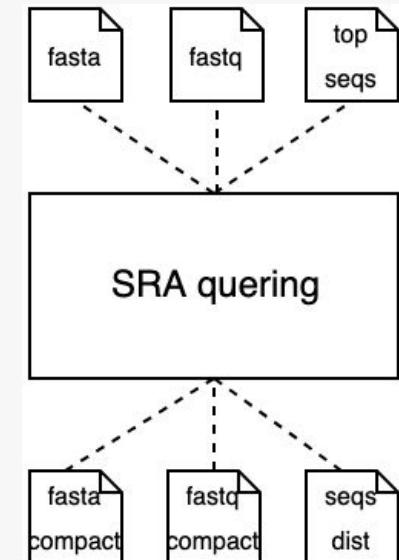
After we create a sequence distribution over the samples.

sample_name	seq 1	seq 2	seq 3
10317.000069001	0	0	0
10317.000054289	18	375	0
10317.000069002	94	5369	1
10317.000068657	311	14	0
10317.000093103	166	1028	0

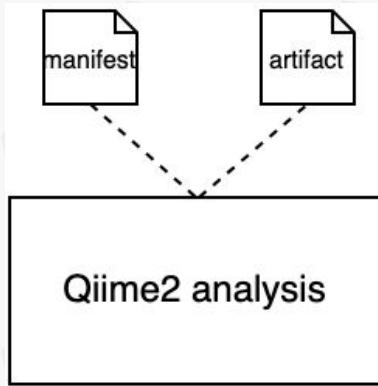
Top seqs distribution example

```
sample_name,typology,bio_sample_id,runId
10317.000066612,healthy,7496859,['ERR2092040']
10317.000051168,healthy,6365139,['ERR1841778', 'ERR1843783']
10317.000065770,healthy,7354089,['ERR2032865']
10317.000073253,healthy,8577551,['ERR2314357']
```

Sample file update example



The pipeline - Qiime2 analysis



```
sample-id      absolute-filepath  
sample-1       $PWD/some/filepath/sample1_R1.fastq  
sample-2       $PWD/some/filepath/sample2_R1.fastq
```

Manifest file format example

We implement a Quiime2 analysis about quality score for fastq file retrieved by NCBI.

First of all, for each sample, we compact all sequences in possibly different files, in only one.

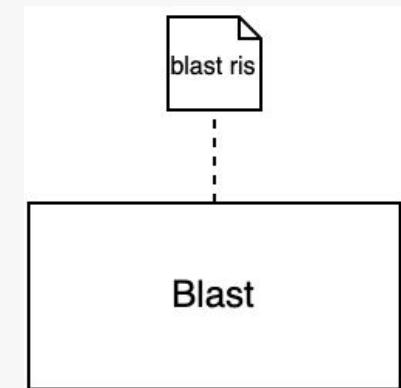
After creating a manifest file in order to link the samples with their sequences and create artifact for Quiime2 pipeline.

The pipeline - Blast

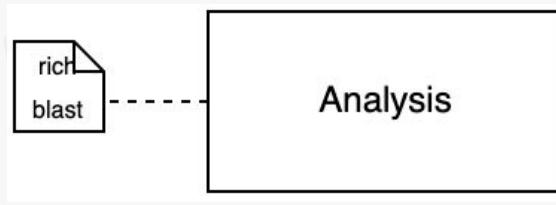
After all we launch a blast algorithm in order to retrieve, for each sequence, the taxonomy associated.

```
seqid    accession   identity    evalue   sciname  taxids   taxonomy
```

Blast file columns result



The pipeline - Analysis



All major analysis has been done in this last phase.

A fundamental file has been generated here: *rich blast*; that contains informations about bacteria frequency in the sample.

`sciname, taxonomy, seqnumber, total_sample, sample_with, len_sample_with`

Rich blast file columns

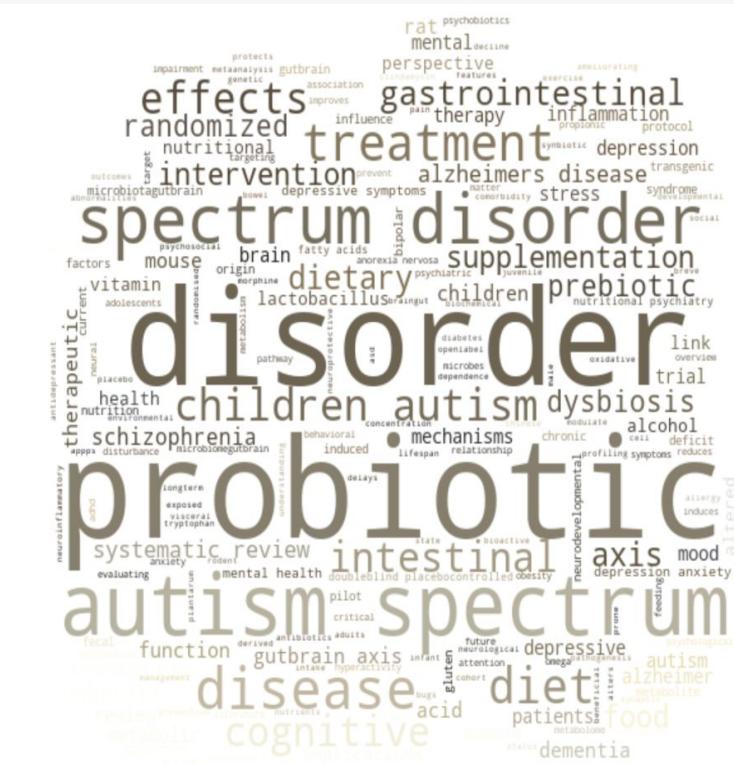
03

Mens sana in corpore sano

Mens sana in corpore sano

Few people are aware of the connection between nutrition and mental illness.

Many of the easily visible dietary patterns that precede depression are the same as those that occur in people diagnosed with mental illness¹.



1: Food, Mood, and Brain Health: Implications for the Modern Clinician

Mens sana in corpore sano

Considering the data in the Gut dataset it was easy to highlight people with mental illnesses considering the attribute¹:

- mental illness

The main challenge was the selection of attributes to classify people with eating disorders²:

- mental illness
- bmi
- fruit frequency
- exercise frequency



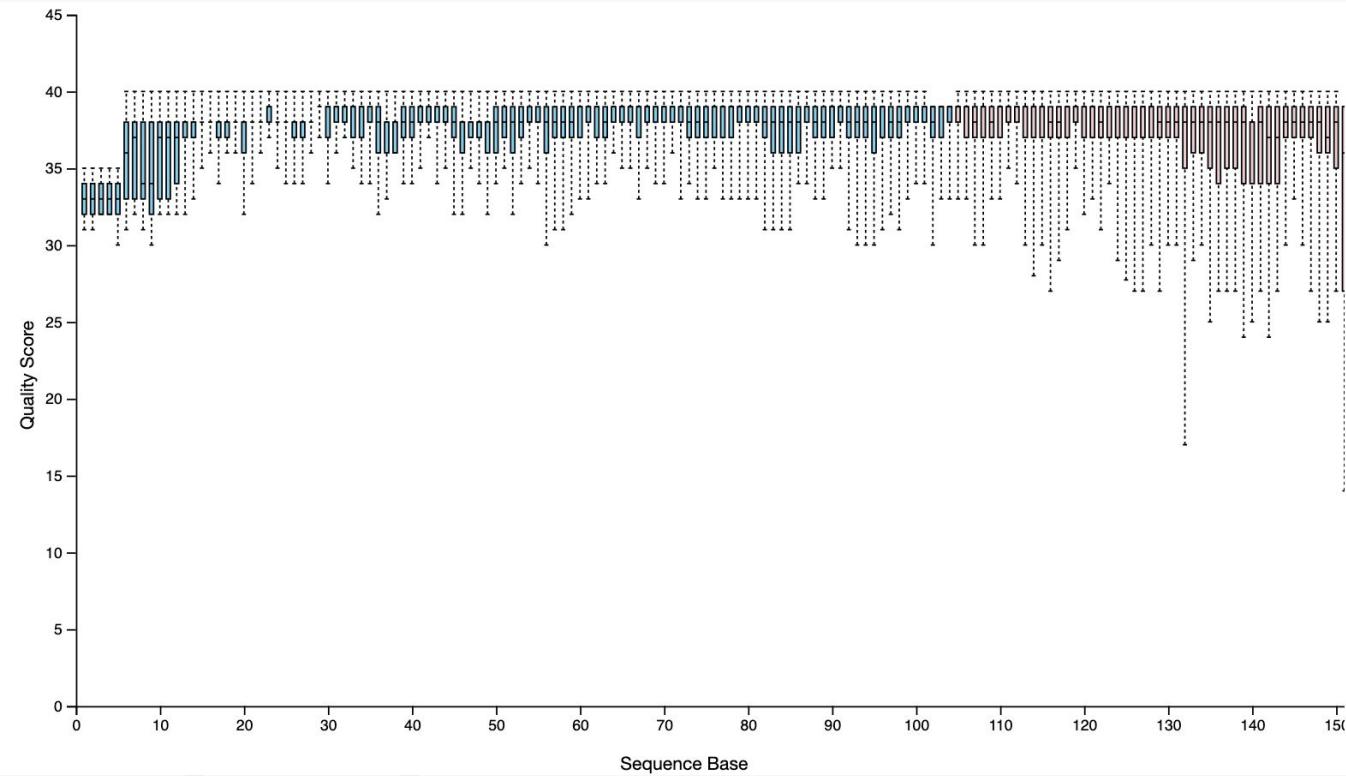
1: [The microbiome and mental health: Hope or hype?](#)

2: [Understanding nutrition, depression and mental illnesses](#)

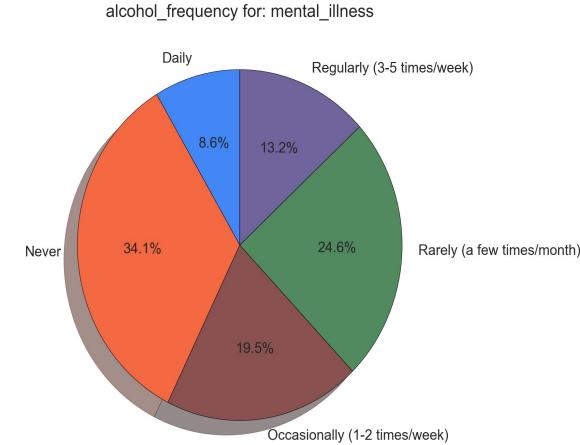
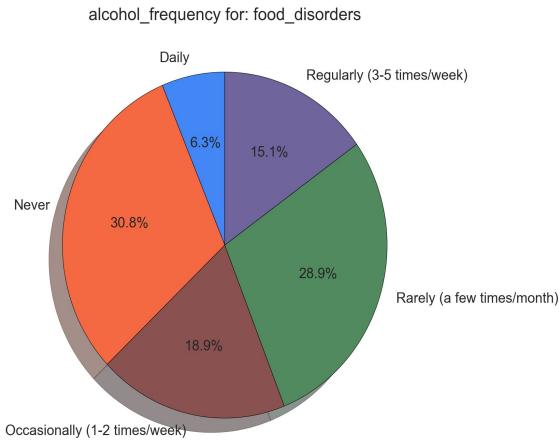
Mens sana in corpore sano - Qiime2 analysis

The analysis with Quiime2 shows that the quality of each base is high.

Less than 1% of the sequences are not 150 characters long, such are not considered since they do not appear in the higher frequencies on blast.



Mens sana in corpore sano - GUT analysis

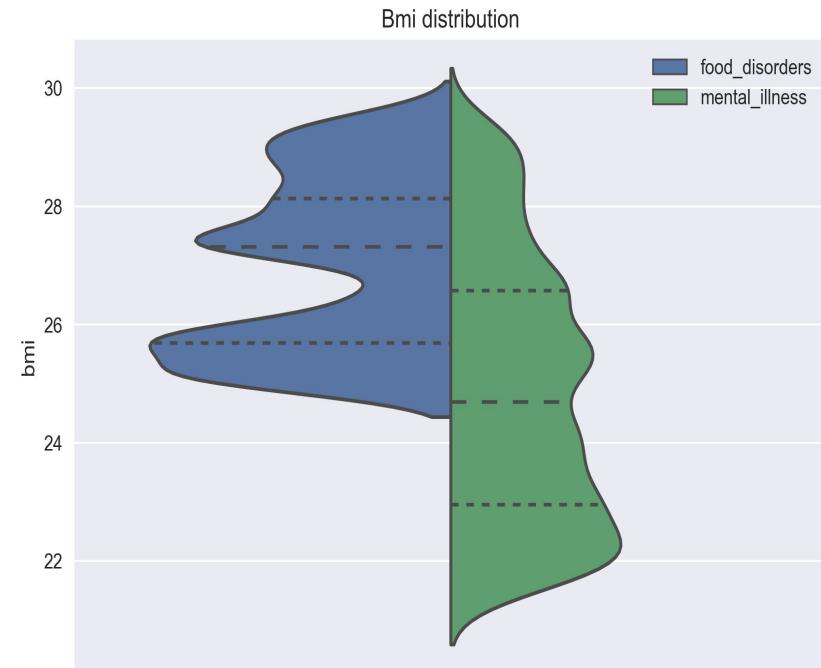


Alcohol can temporarily relieve feelings of anxiety and depression so it is often used as a form of "self-medication". Self-medication can lead to eating disorders such as obesity.

Mens sana in corpore sano - GUT analysis

BMI distribution

Analyzing individuals sampled in Gut with $bmi < 18.5$ or $bmi > 24.99$, it can be seen that all individuals with eating disorders have obesity problems and even more than 50% of individuals with mental illnesses have weight problems.

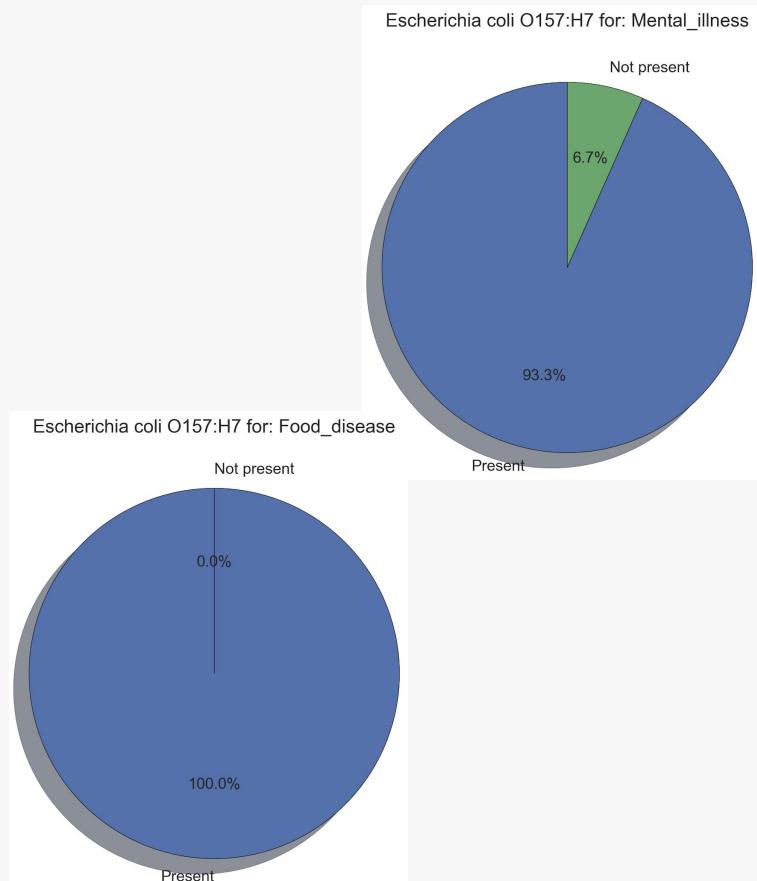


Mens sana in corpore sano - Bacteria analysis

Escherichia coli O157: H7

Defined as a Gram negative bacterial species present in the intestine, most strains of E. coli harmlessly colonize the gastrointestinal tract of humans and animals as normal flora¹.

Strains such as o157 produce toxins and cause food-borne diseases in general.



1: A Brief Overview of Escherichia coli O157:H7 and Its Plasmid O157

Mens sana in corpore sano - Bacteria analysis

Escherichia coli O157:H7

Infection with this type of pathogenic bacteria can lead to hemorrhagic diarrhea, and it is also reported that they cause death.

In the United States most cases of hemolytic uremic syndrome caused by this bacterium lead to kidney failure in children¹.



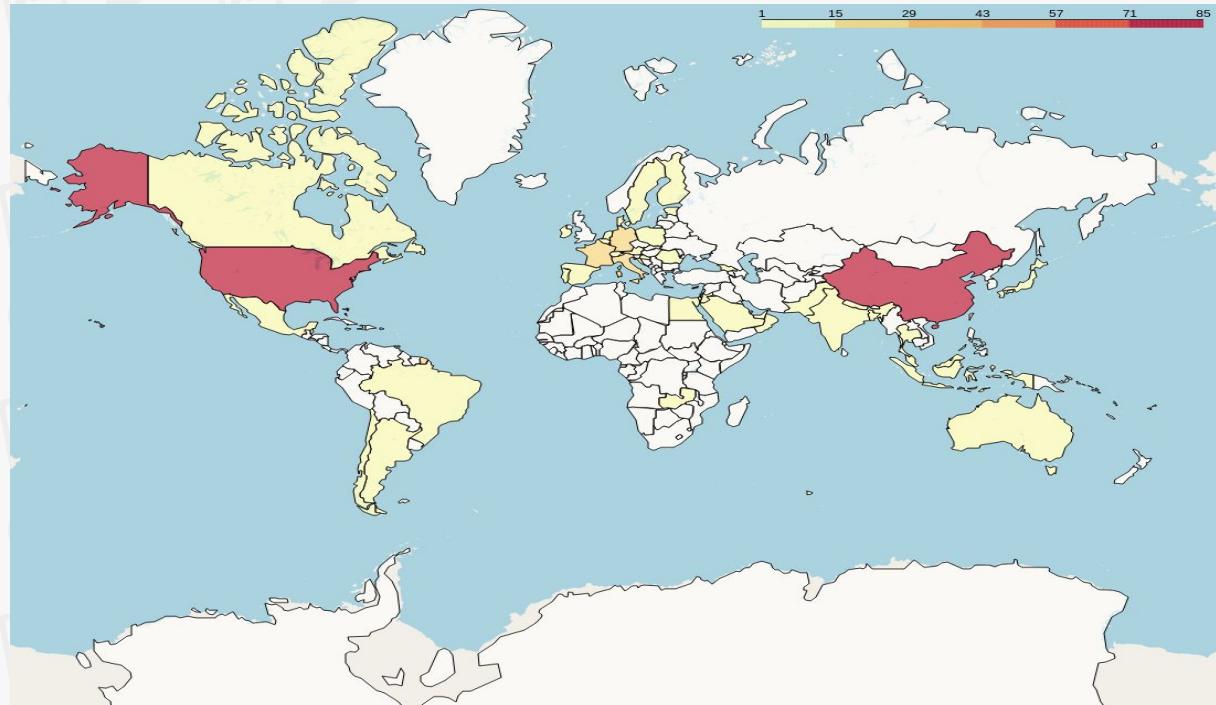
1: Infection by *Escherichia coli* O157:H7 and Other Enterohemorrhagic *E. coli*

Mens sana in corpore sano - Bacteria analysis

Escherichia c. is widely studied in the United States, Great Britain, Germany and many states in Australia.

The presence of this bacterium is associated with the diarrhea epidemic that occurred in students and staff of a school for adults and children with mental health problems in the USA in 1981¹.

Escherichia coli O157: H7 in the world



Mens sana in corpore sano - Bacteria analysis

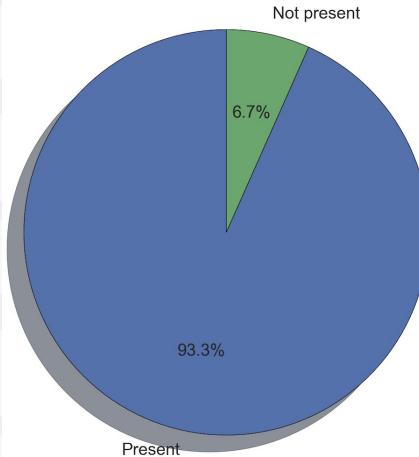
Bacteroides Ovatus

Gram negative bacteria, it's one of the predominant bacteria in the intestine.

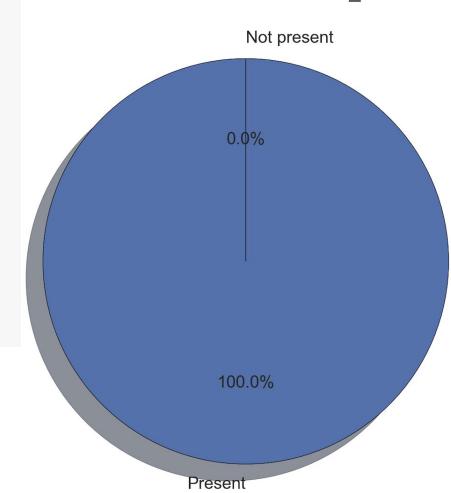
A study of bacterial species in intestinal microbes has shown that *B. ovatus* causes antibody responses in serum in patients with inflammatory bowel disease (IBD).

B. ovatus were more abundant in controls than patients with celiac¹ disease that causes chronic fatigue related to severe obesity and psychiatric illness².

Bacteroides ovatus for: Food_disease



Bacteroides ovatus for: Mental_illness



1: [Chronic Fatigue Syndrome](#)

2: [Intestinal Bacteroides Species Associated With Coeliac Disease](#)

Mens sana in corpore sano - Bacteria analysis

Bacteroides Ovatus

The high presence of the bacteria is linked to type 1 diabetics¹.

Obesity has recently been shown to be a risk factor for people with type 1 and type 2 diabetes².

Conditions such as stress, anxiety and even depression are linked to diabetes, it is estimated that 50% of people with diabetes will suffer from a mental disorder sooner or later³.



1: Verso la definizione del microbioma autoimmune per il diabete di tipo 1

2: IL DIABETE DI TIPO 1

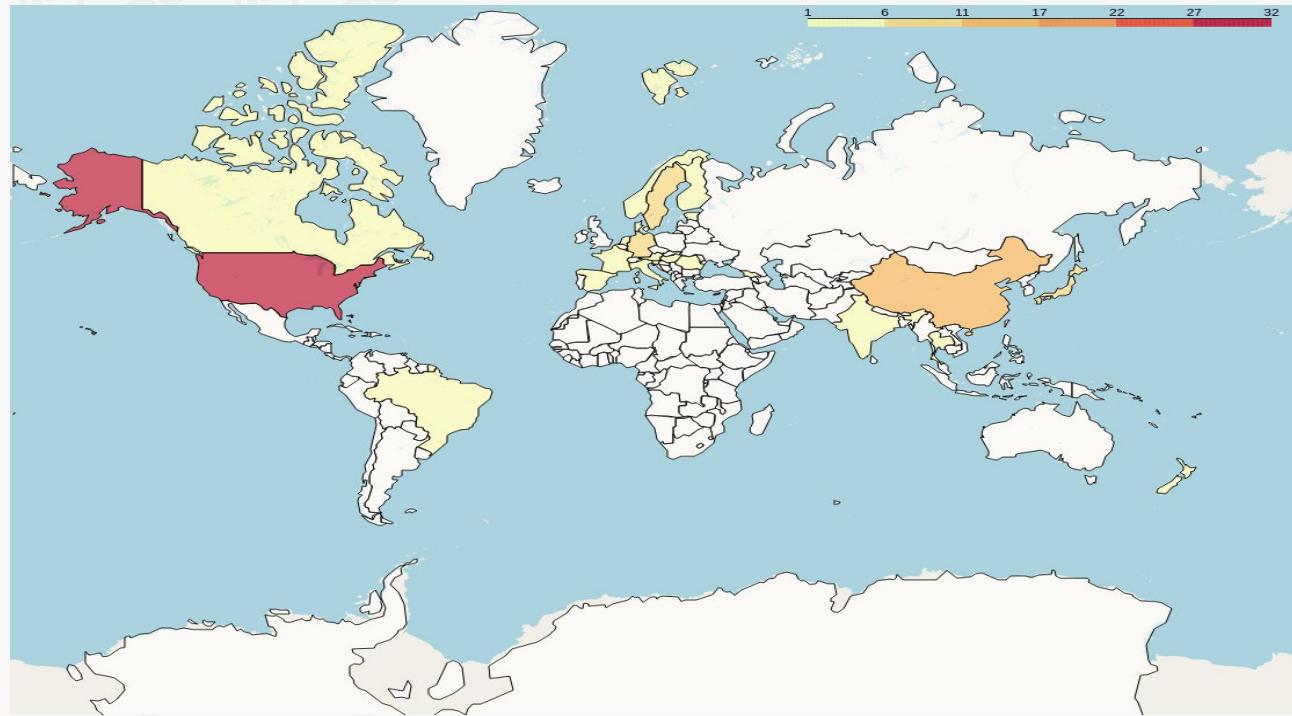
3: Aspetti psicologici del diabete

Mens sana in corpore sano - Bacteria analysis

For the ovarian strain ELH-B2 is considered a next generation probiotic potential because of its preventive effects on intestinal inflammation.

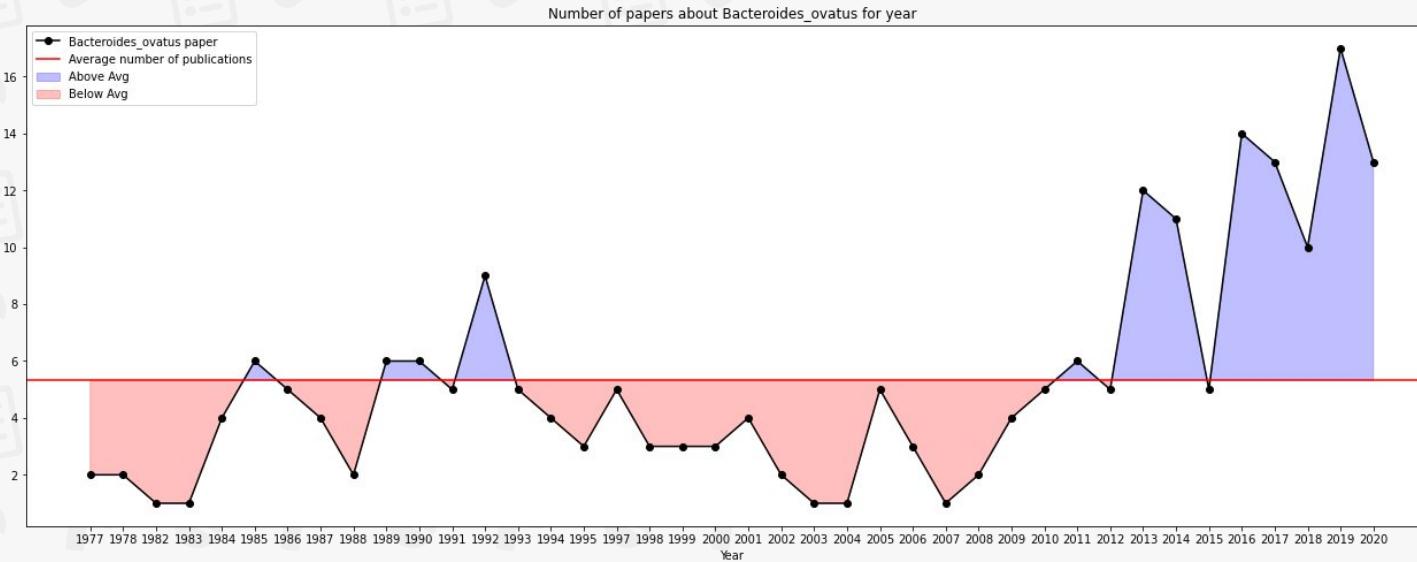
It is estimated that probiotics will have an expected global turnover value of \$46.55 billion by 2020¹.

Bacteroides Ovatus in the world



Mens sana in corpore sano - Bacteria analysis

Bacteroides Ovatus over the years



In recent years this bacterium has seen an increase in related research both in terms of its benefits (use in probiotics) and the diseases it can cause (celiac disease, diabetes). Scientists are looking for connections that allow them to understand the impact of the intestine on the brain and mental illnesses.

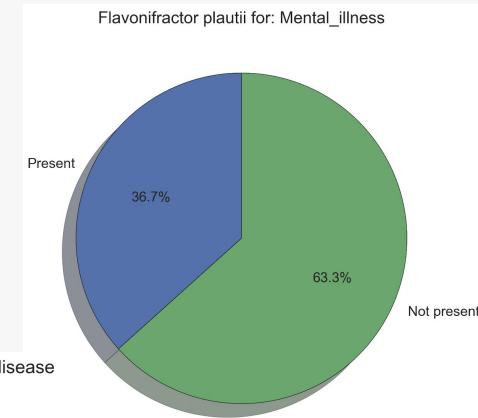
Mens sana in corpore sano - Bacteria analysis

Flavonifractor Plautii

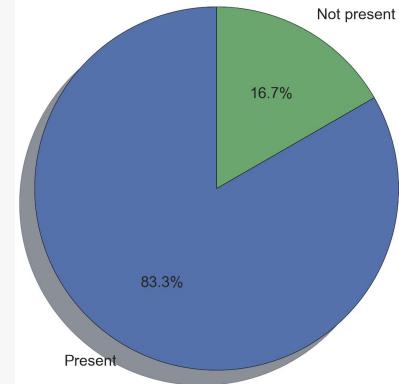
It is a gram positive bacterium that often has gram negative variants, it is a common organism in the human intestinal microbiome and are anaerobic.

There are only two infections related to this bacterium in the world, it is considered a potential pathogen for cholecystitis causing blood flow infection¹.

Flavonifractor plautii for: Mental_illness



Flavonifractor plautii for: Food_disease



1: [Flavonifractor \(Eubacterium\) plautii bloodstream infection following acute cholecystitis](#)

Mens sana in corpore sano - Bacteria analysis

Flavonifractor Plautii

The association of colorectal cancer (CRC) with Flavonifractor plautii in Indian patients has emerged as a new discovery¹.

Many studies show that a diet rich in animal fat and protein and low in fiber is associated with an increase in intestinal cancer.

Obesity and sedentary lifestyle are additional risk factors².

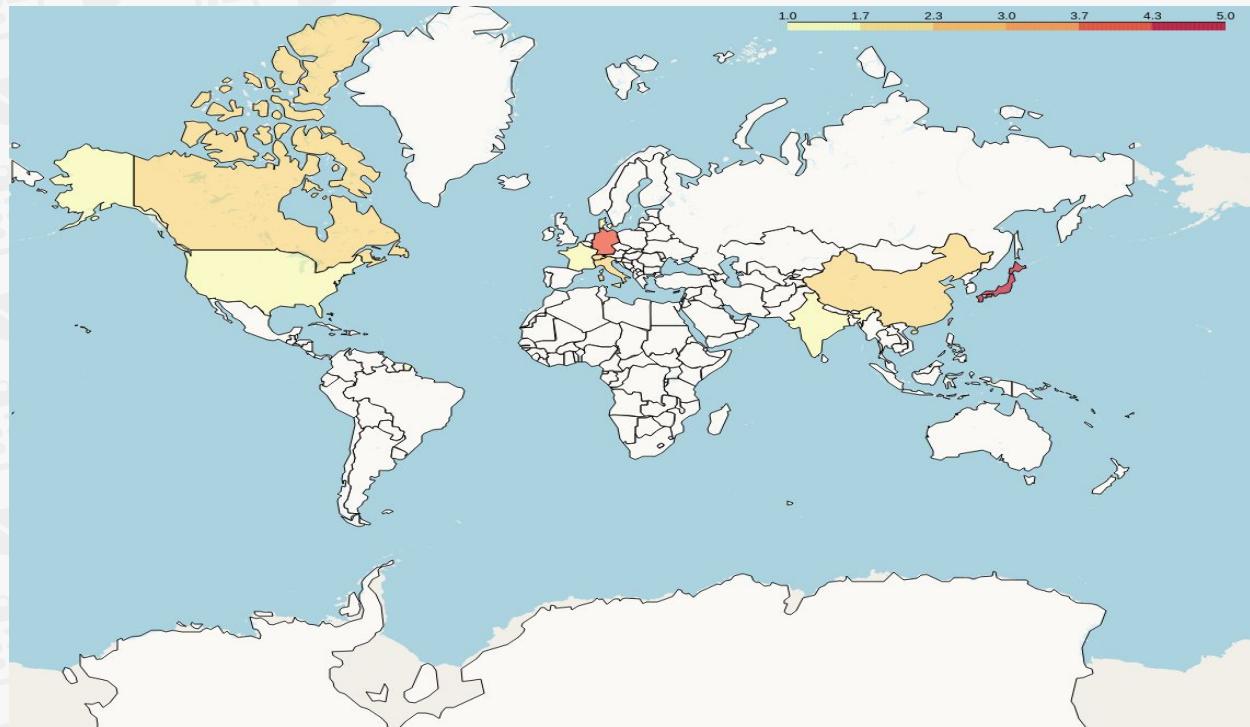


1: Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India

2: Tumore del colon-retto

Mens sana in corpore sano - Bacteria analysis

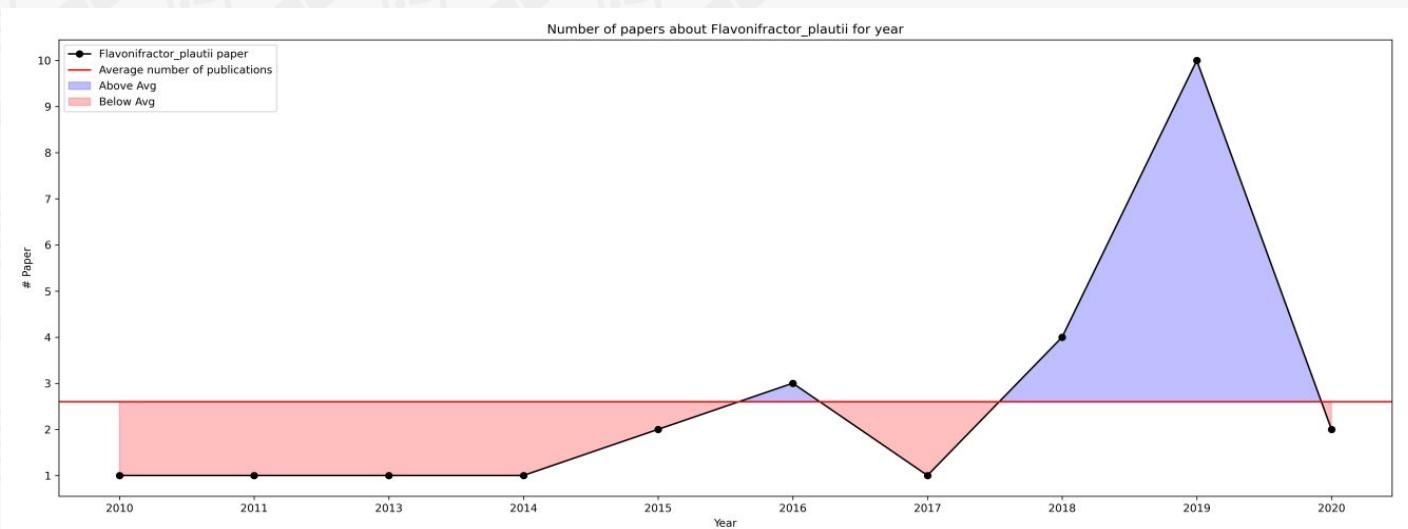
Flavonifractor Plautii in the world



Analyses of this bacterium are few and far between, as it is only in recent years that the link of Flavonifractor plautii with bipolar disorders and rectum cancer has been discovered.

Mens sana in corpore sano - Bacteria analysis

Flavonifractor Plautii over the years



One of the reasons for the increase in studies on this bacterium was the discovery in 2019 proposed by Klara Coello on the presence of this bacterium in the composition of the intestinal microbiome in patients with bipolar disorder, Flavonifractor was present in 61% of patients with BD, moreover a greater correlation was found with smokers and the female sex¹.

1: [Gut Microbiota Composition in Patients With Newly Diagnosed Bipolar Disorder and Their Unaffected First-Degree Relatives](#)

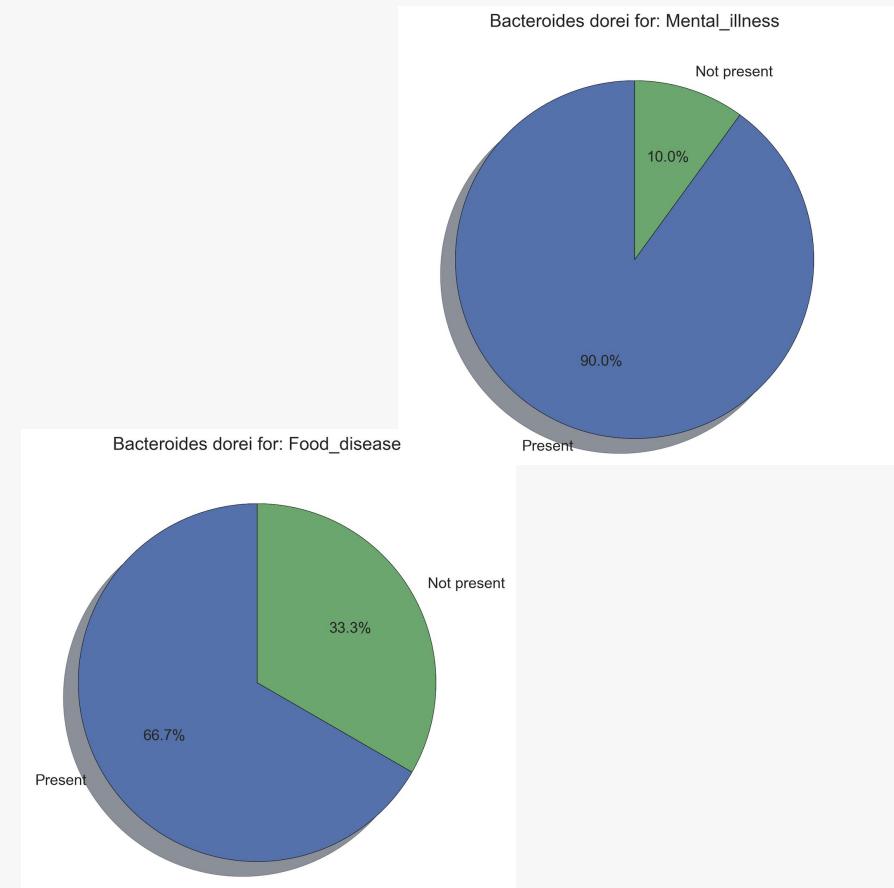
Mens sana in corpore sano - Bacteria analysis

Bacteroides Dorei

They are Gram-negative bacteria, which are found in the gastrointestinal tract of humans and play a role in normal intestinal functioning.

They are also known as opportunistic pathogenic bacteria present in anaerobic infections.

When an alteration of the mucous barrier occurs the bacterium is able to enter creating a pathogen.



Mens sana in corpore sano - Bacteria analysis

Bacteroides Dorei

The difference in the bacterial composition of the intestinal system is considered important for the understanding of autoimmune diseases such as type 1 diabetes.

The results of metagenomic sequencing in 2014 showed significantly higher composition of *B. dorei* and *Bacteroides vulgatus* in some children.

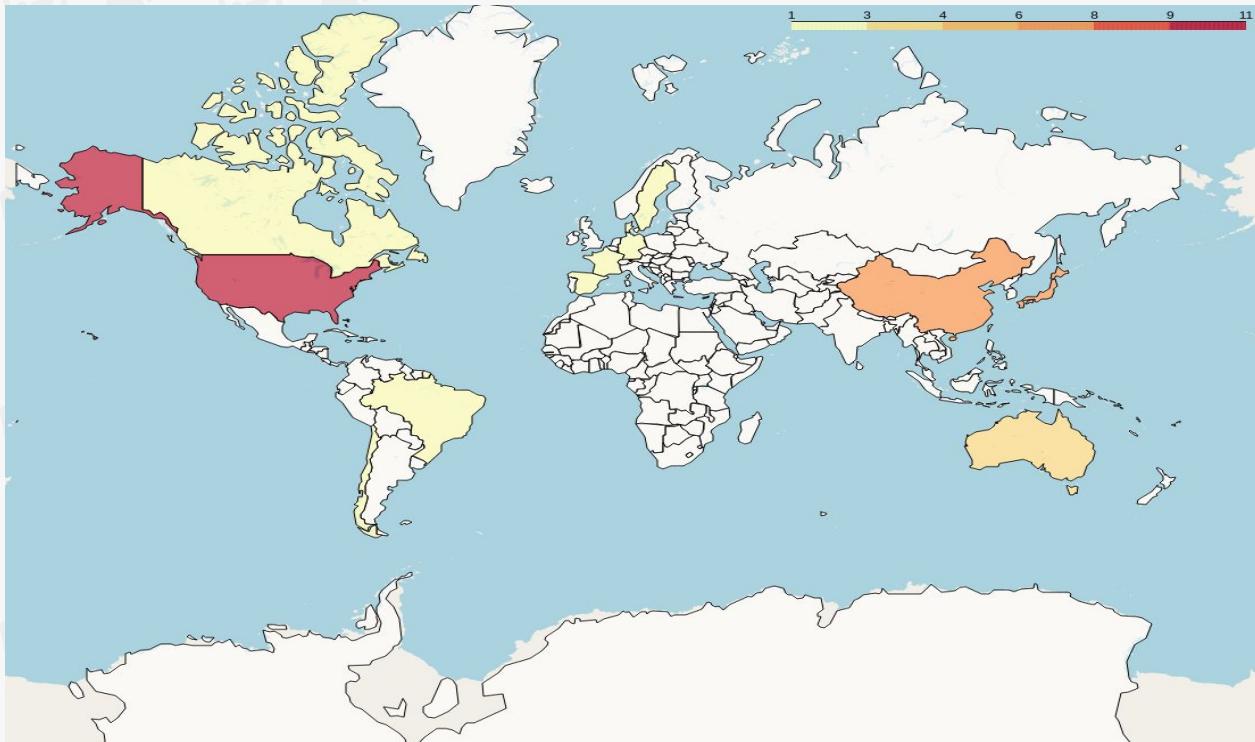
The increase in the amount of *B. dorei* is a potential indicator of T1D1 development¹.



1: *Bacteroides dorei* dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes

Mens sana in corpore sano - Bacteria analysis

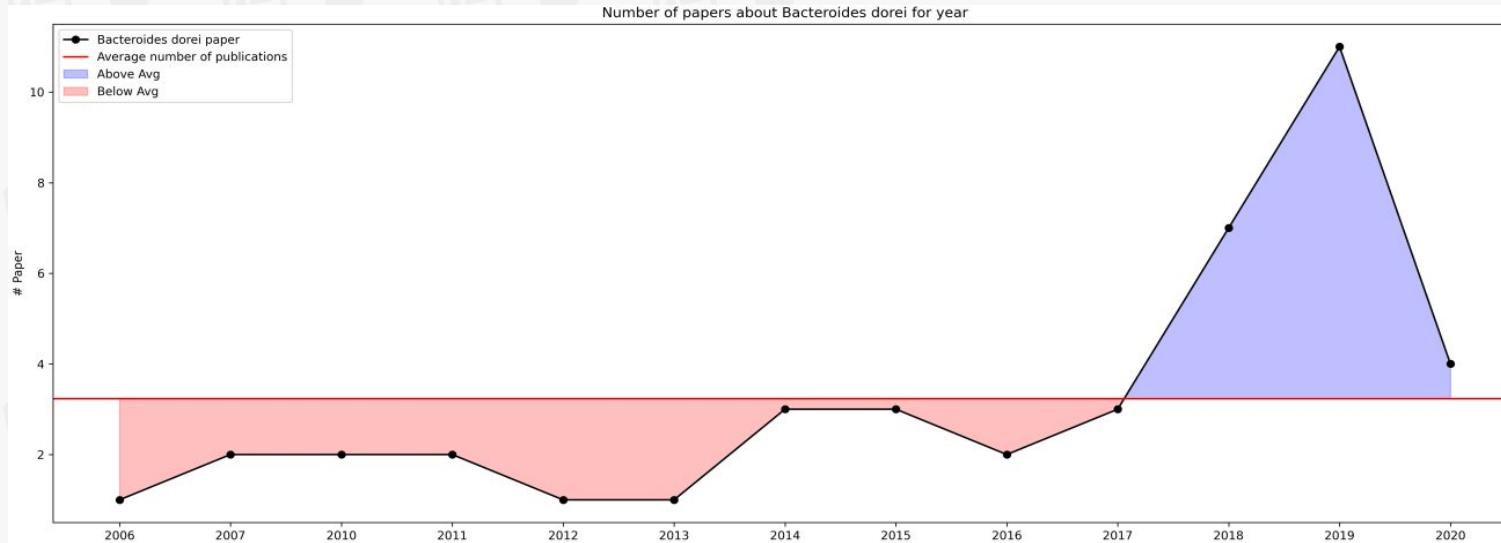
Bacteroides Dorei in the world



Most universities and research centers that have studied this bacterium belong to the United States, China, Japan, or European countries such as the authors of the article on type 1 diabetes.

Mens sana in corpore sano - Bacteria analysis

Bacteroides Dorei over the years



The results of some research in 2018 identified a previously unknown link between specific intestinal bacteria and atherosclerosis. *B. dorei* treatment may help prevent coronary heart disease¹.

The growth of research in 2019 is linked to *B.dorei*'s link to type 1 diabetes how to control it².

1: [Bacteroides vulgatus and Bacteroides dorei Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis](#)

2: [Non-lethal Growth Inhibition by Arresting the Starch Utilization System of Clinically Relevant Human Isolates of Bacteroides dorei](#)

04

Life Style

Life Style

The people life style is important in order to define a good healthy state.

A healthy microbioma could help better respond to infections and diseases that can affect organism.

In this analysis we focused on define a good life style and a bad life style taking in account:

- Smoking frequency
- Drinking frequency
- BMI index

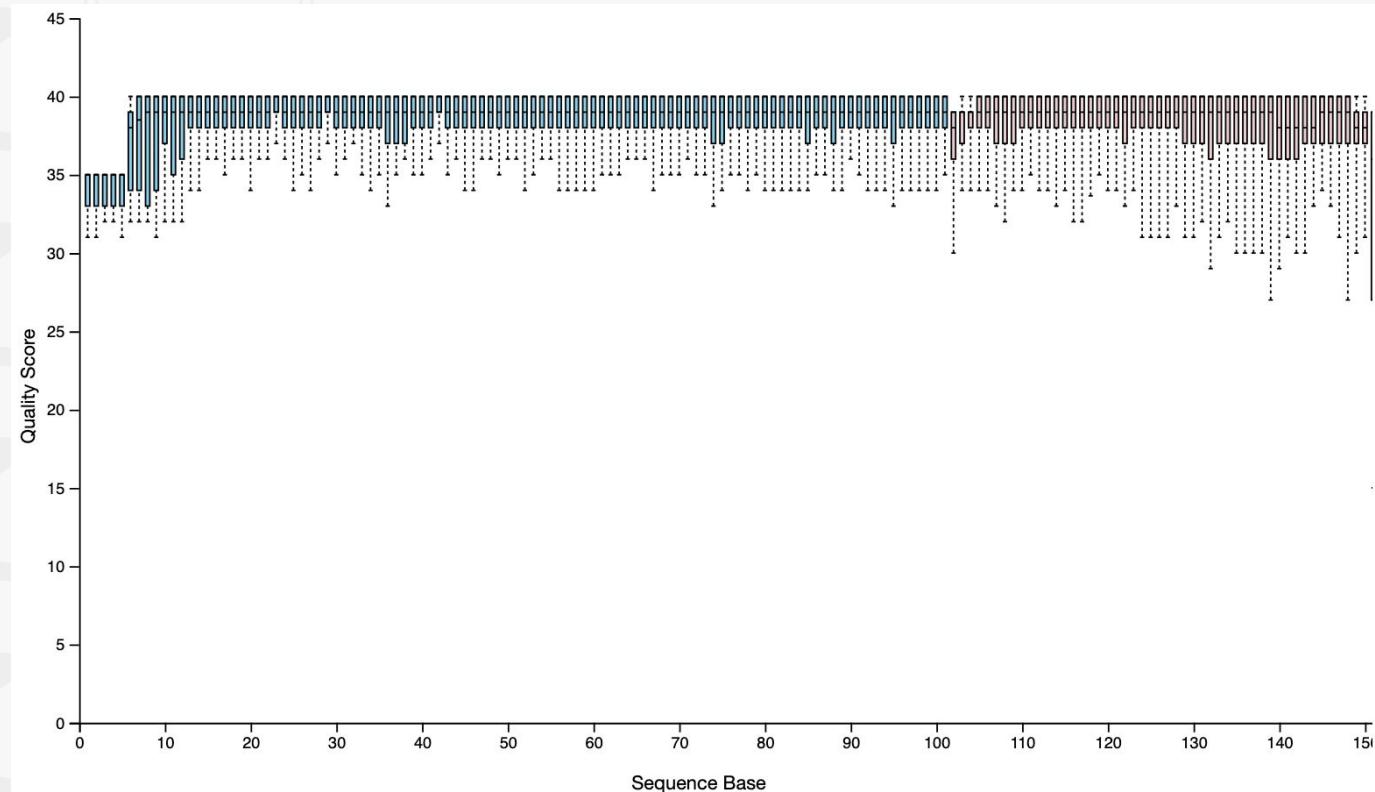
Final sample characteristics:

	Mean man age	Mean woman age	# Woman	# Man
Healthy	34	39	22	18
Not healthy	35	38	13	25

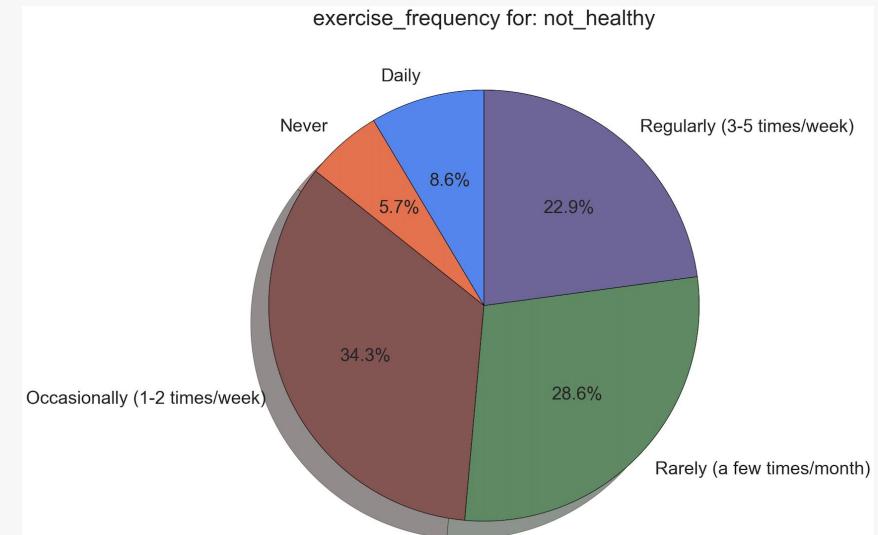
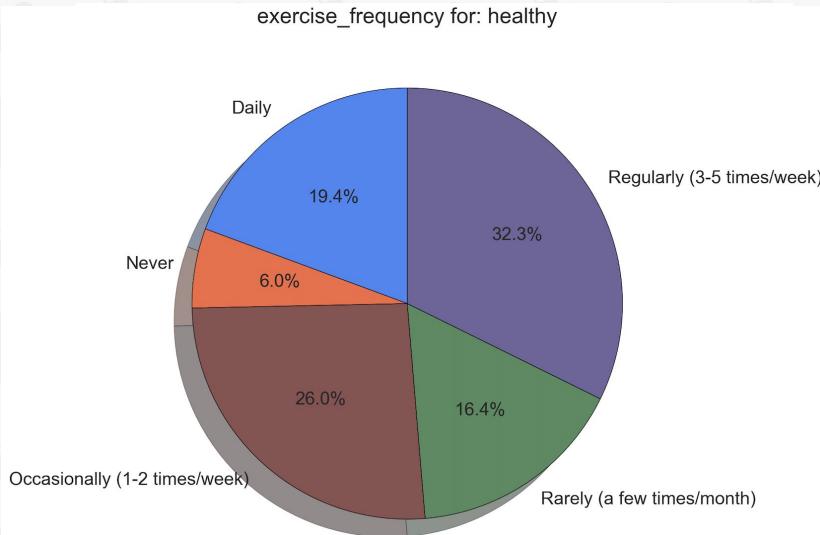
Life Style - Qiime2 analysis

We can observe a relative high quality score for each base.

Pink color is due by short length of a sequences set.



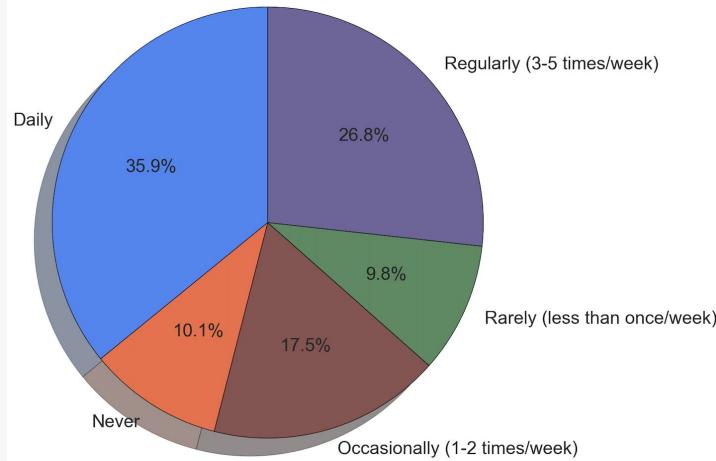
Life Style - GUT analysis



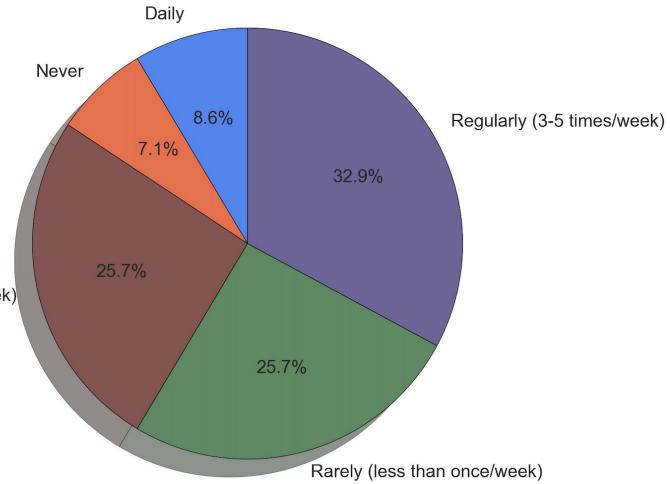
A first GUT analysis shows a very important difference in exercise frequency between two groups; regular physical activity helps to control the risk of pathologies.

Life Style - GUT analysis

fruit_frequency for: healthy



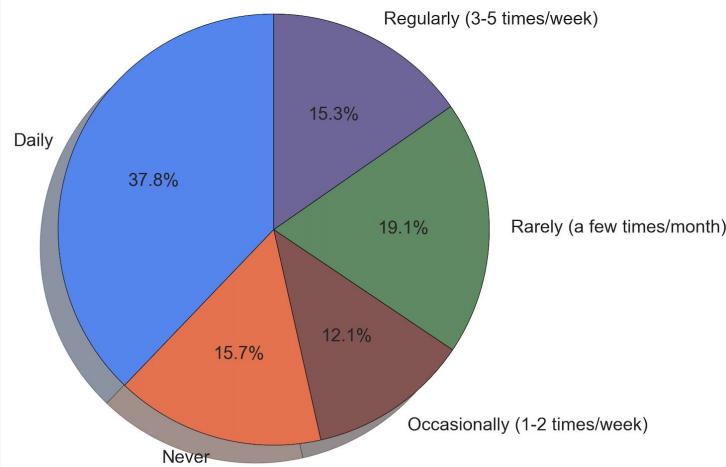
fruit_frequency for: not_healthy



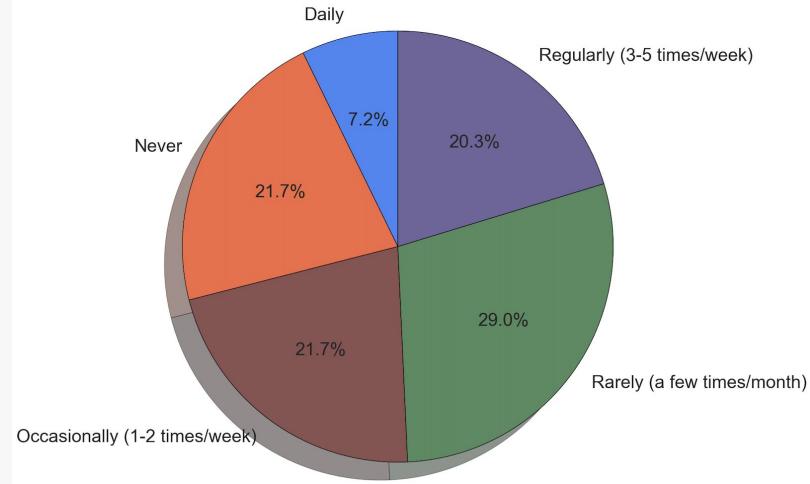
Combine a fruit-rich diet, with a regular physical activity improve immune response, these analysis shows a marked difference between the two sets of people.

Life Style - GUT analysis

flossing_frequency for: healthy



flossing_frequency for: not_healthy



Last analysis on GUT, focused on self-care, may not help healthy life style identification but increase our knowledges about this experiment.

Life Style - Bacteria analysis

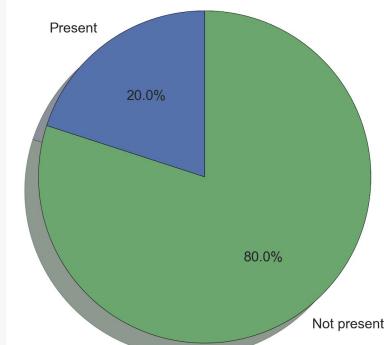
Bifidobacterium pseudocatenulatum

Bifidobacterium species are majority bacteria among gastrointestinal tract (GIT) of humans.

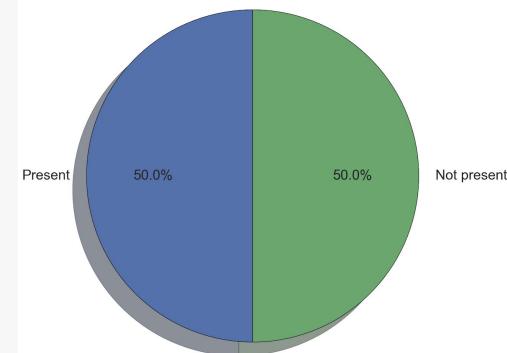
They play important roles in maintaining human health via the digestion of foods, production of essential vitamins, as well as by preventing the colonization or overgrowth of pathogens in the GIT¹.

Different studies show as the **Bifidobacterium pseudocatenulatum** decreased significantly in patients with liver diseases².

Bifidobacterium pseudocatenulatum for: not_healthy



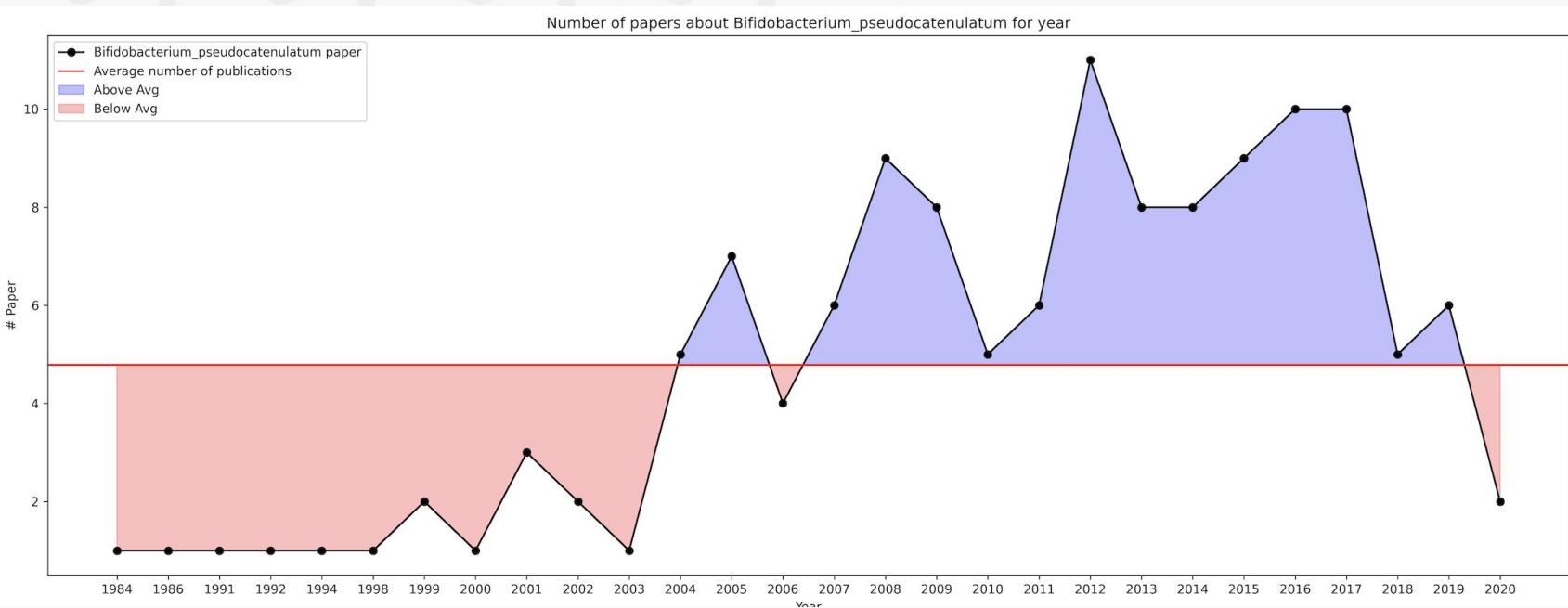
Bifidobacterium pseudocatenulatum for: healthy



1: [The genome of Bifidobacterium pseudocatenulatum](#)

2: [Bifidobacterium pseudocatenulatum attenuate D-galactosamine-induced liver injury by modifying the gut microbiota](#)

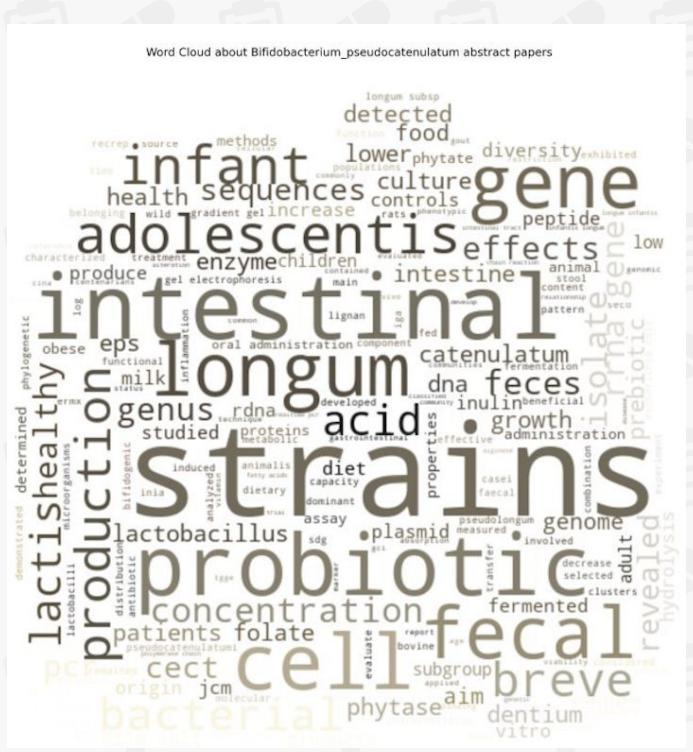
Life Style - Bacteria analysis



The number of papers associated with this bacteria suggest an increase interested in the last years, the American Gut project was founded in the end of 2012¹, after that the number of publications is higher than before.

1: First Major Results of "American Gut Project" Published

Life Style - Bacteria analysis



On the paper abstract wordcloud we can notice, of course, a word that recall a search object.

Other words, as healthy and probiotic, suggest that this bacteria could be inject in order to obtain a healthy increase in end host.

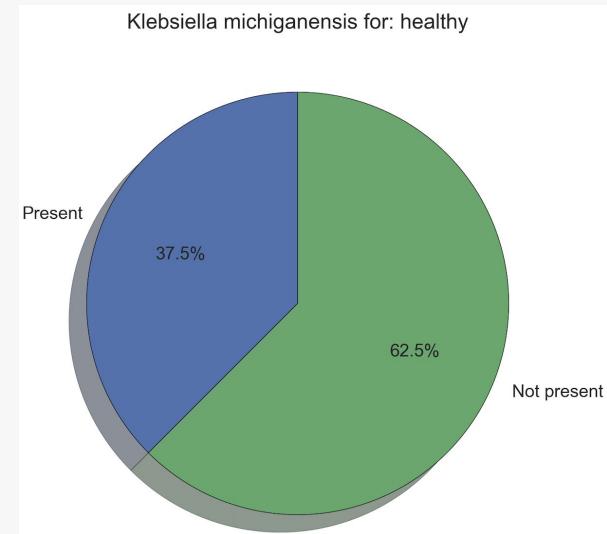
Life Style - Bacteria analysis

Klebsiella michiganensis

Gram-negative bacteria.

Different studies of klebsiella michiganensis showed that it could help in order protect against gut pathogen colonization.

K. michiganensis also hampered colonization of the pathogen Salmonella, prolonging host survival¹.



1: [Klebsiella michiganensis transmission enhances resistance to Enterobacteriaceae gut invasion by nutrition competition](#)

Life Style - Bacteria analysis

Akkermansia muciniphila

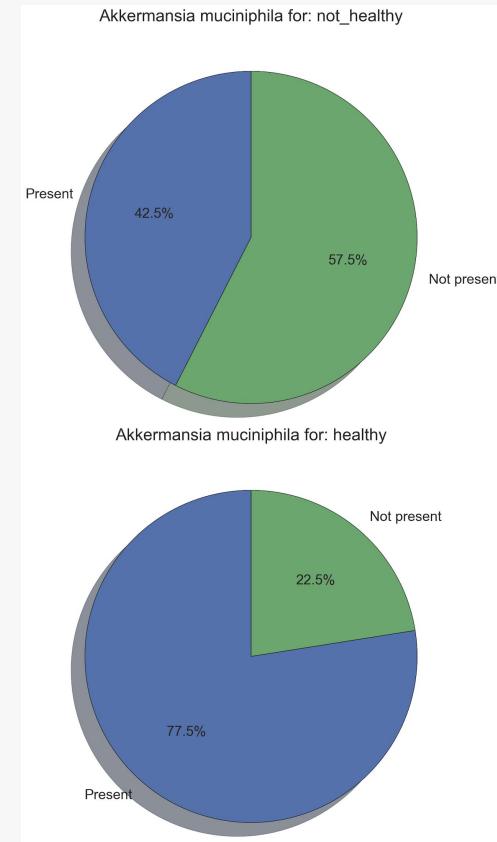
Gram-negative bacteria.

Researchers have discovered that *A. muciniphila* could be used to combat obesity and type 2 diabetes¹.

The bacterium is naturally present in the human digestive tract, but has been seen to fall with obesity² and irritable bowel disease.

1: [Akkermansia muciniphila](#)

2: [What is Akkermansia muciniphila and how does it protect gut health?](#)



Life Style - Bacteria analysis

Words as obese, probiotic, diet, healthy recall an already spoken *A. muciniphila* behaviour in order to improve interaction with lipid metabolism¹.

Inflammatory and inflammation words are referred to A. muciniphila anti-inflammatory effects in humans².



1: Function of *Akkermansia muciniphila* in Obesity: Interactions With Lipid Metabolism, Immune Response and Gut Systems

2: Akkermansia muciniphila, a bacterium that may help reduce metabolic syndrome

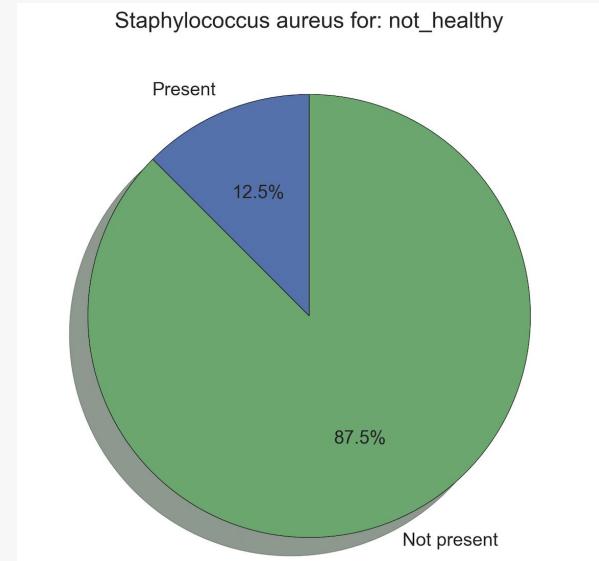
Life Style - Bacteria analysis

Staphylococcus aureus

Gram-positive bacteria.

In humans, *S. aureus* is not part of the normal microbiota present in the upper respiratory tract or gut mucosa or on the skin.

S. aureus is one of the most common causes of bacteremia and infective endocarditis¹.



1: [Staphylococcus aureus](#)

Life Style - age focus

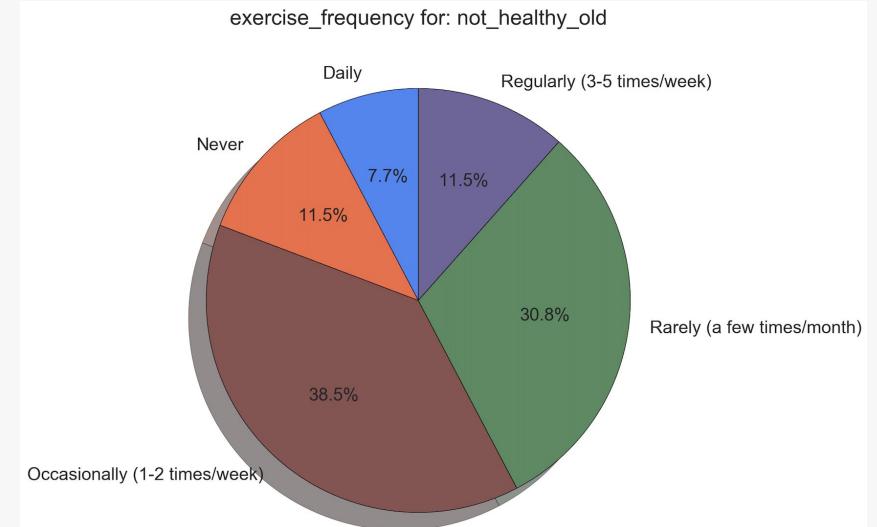
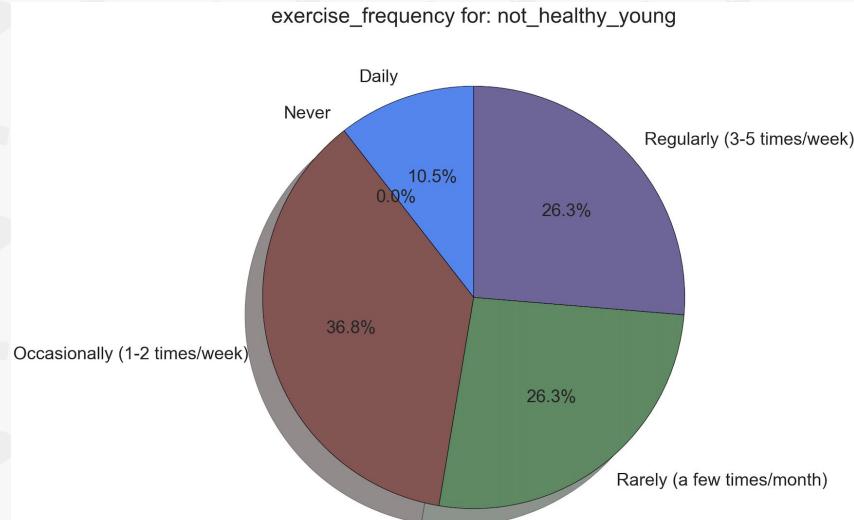
In order to weight better the age variable, it has been decided to split the not healthy people focusing on the age year.

We defined two groups: young/old not healthy people.

Final sample characteristics:

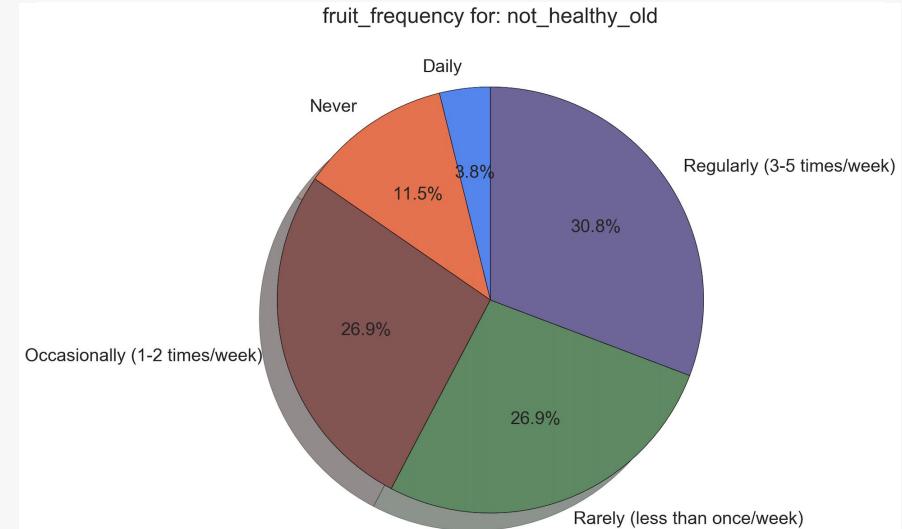
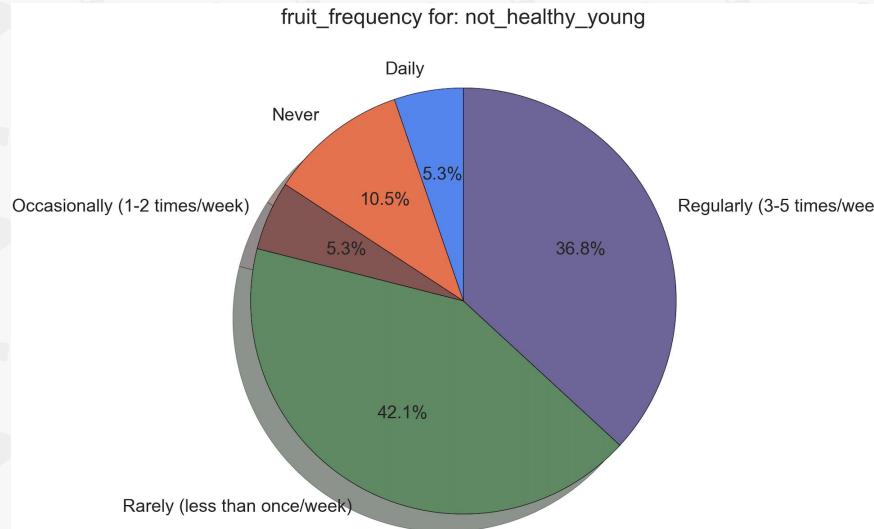
	Mean man age	Mean woman age	# Woman	# Man
Not Healthy Young	26	23	8	11
Not healthy Old	46	46	5	12

Life Style - age focus - GUT analysis



A first similitudine between healthy and not healthy young is represented by the exercise frequency.

Life Style - age focus - GUT analysis

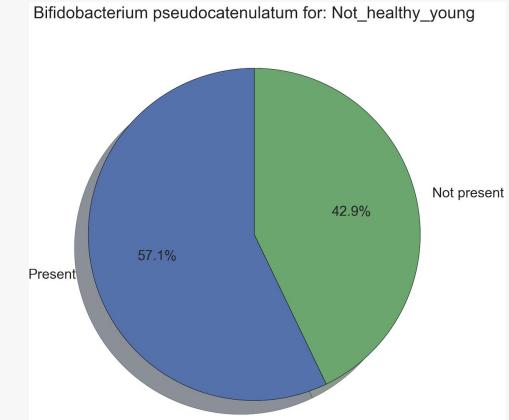
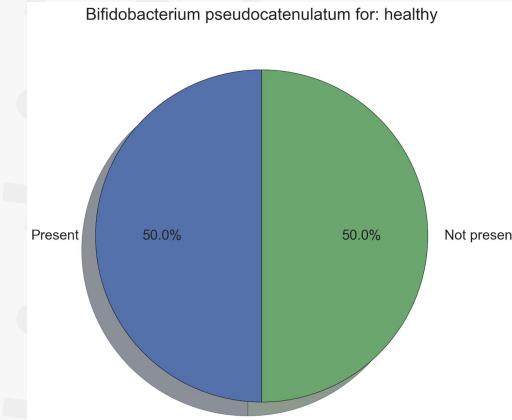
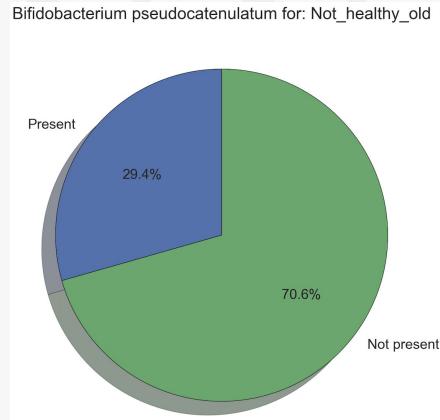


A fruit intake frequency does not show a big similarity number to healthy people; this could be caused by young people tendency of not eating fruit.

Life Style - age focus - Bacteria analysis

Bifidobacterium pseudocatenulatum

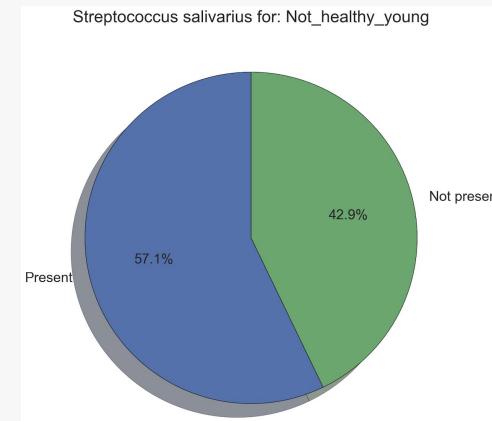
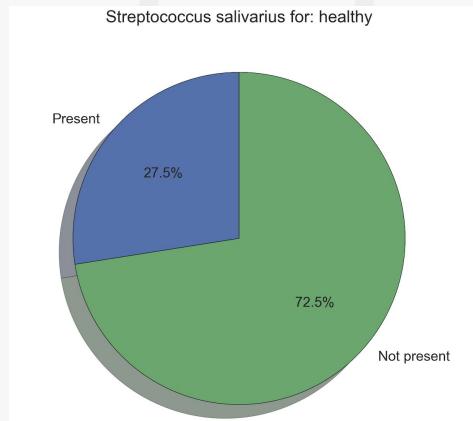
An already seen bacteria, highest values of this in the not healthy young group put this category closest to healthy people.



Life Style - age focus - Bacteria analysis

Streptococcus salivarius

Another interesting similitudine is about Streptococcus salivarius presence; this bacteria may contribute to the establishment of immune homeostasis and regulation of host inflammatory responses¹. Not healthy old doesn't observe the presence of this bacteria.



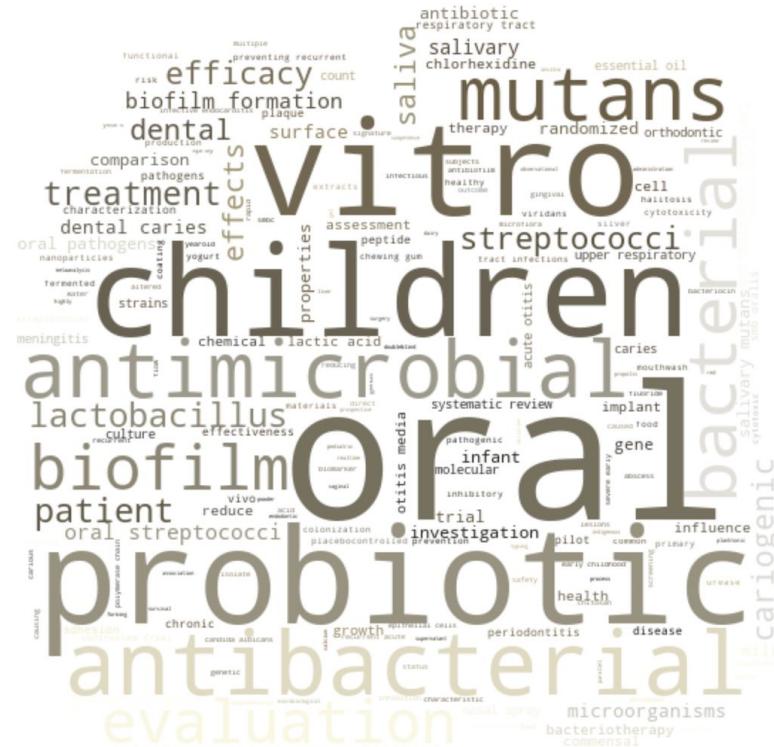
1: Anti-Inflammatory Properties of Streptococcus salivarius

Life Style - age focus - Bacteria analysis

The word children could be justified by the fact that *Streptococcus salivarius* is one of the first colonizers of the human oral cavity and gut after birth.

Other words, as probiotic and antibacterial, could recall the anti-inflammatory bacteria behaviour. Moreover *S. salivarius* are being trialed for their use as a probiotic in the prevention of oral infections¹.

Word Cloud about *Streptococcus_salivarius* title papers ever



1: *Streptococcus salivarius*

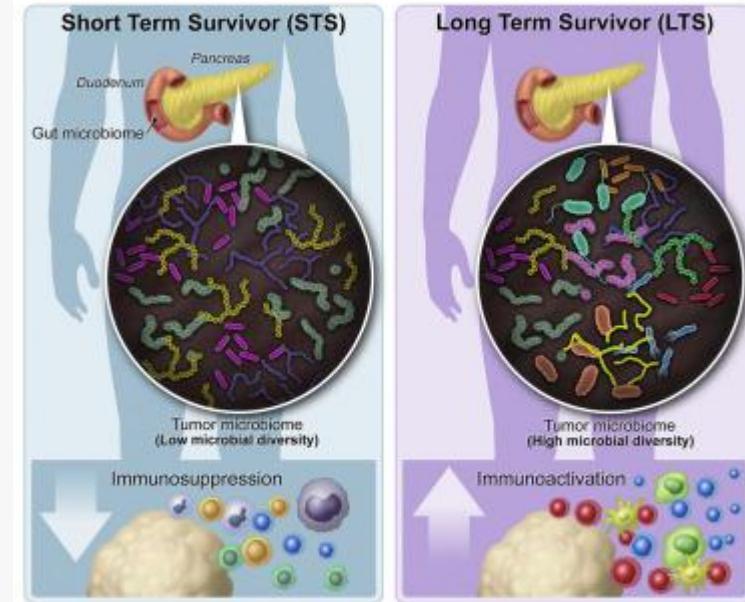
Life Style - effects of the loss of microbial diversity.

Most of the Human diseases affecting westernized countries are associated with dysbiosis and loss of microbial diversity in the gut microbiota.

It also seem that for some pathologies, patients with an higher microbiome biodiversity could survive longer than patients with low microbiome biodiversity.

<https://www.sciencedirect.com/science/article/pii/S0092867419307731>

<https://www.frontiersin.org/articles/10.3389/fmicb.2016.00455/full>

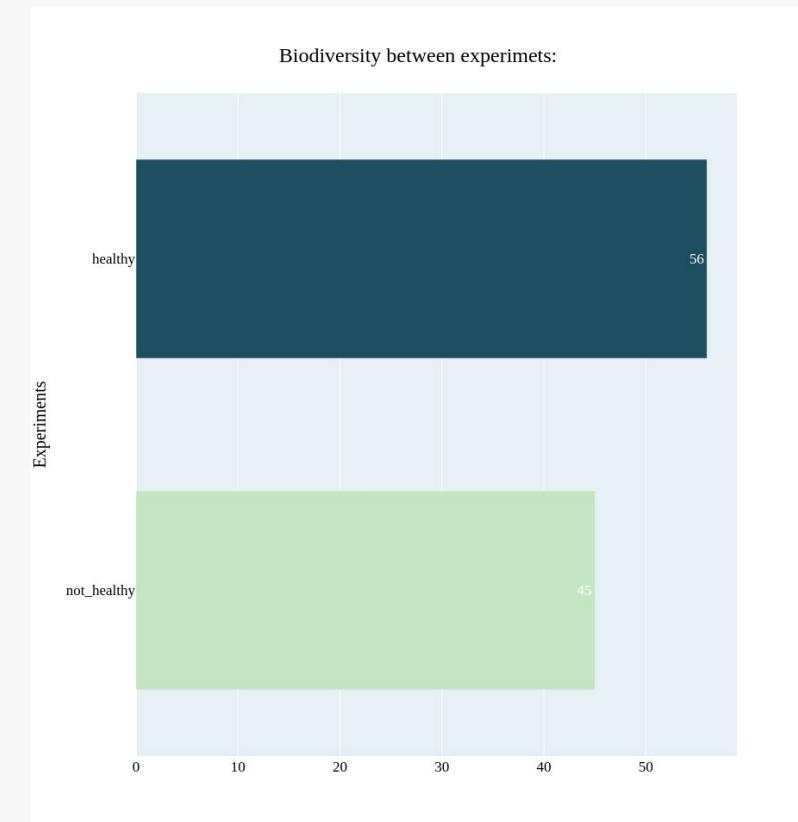


Life Style - effects of the loss of microbial diversity.

We evaluated for this purpose the microbiome biodiversity among our experiments.

We achieved our goal by counting the number of species for each experiments, we also excluded those that occurred less than 5 times among samples....

...as we can expect, the number of species is higher into the healthy experiment!!!!



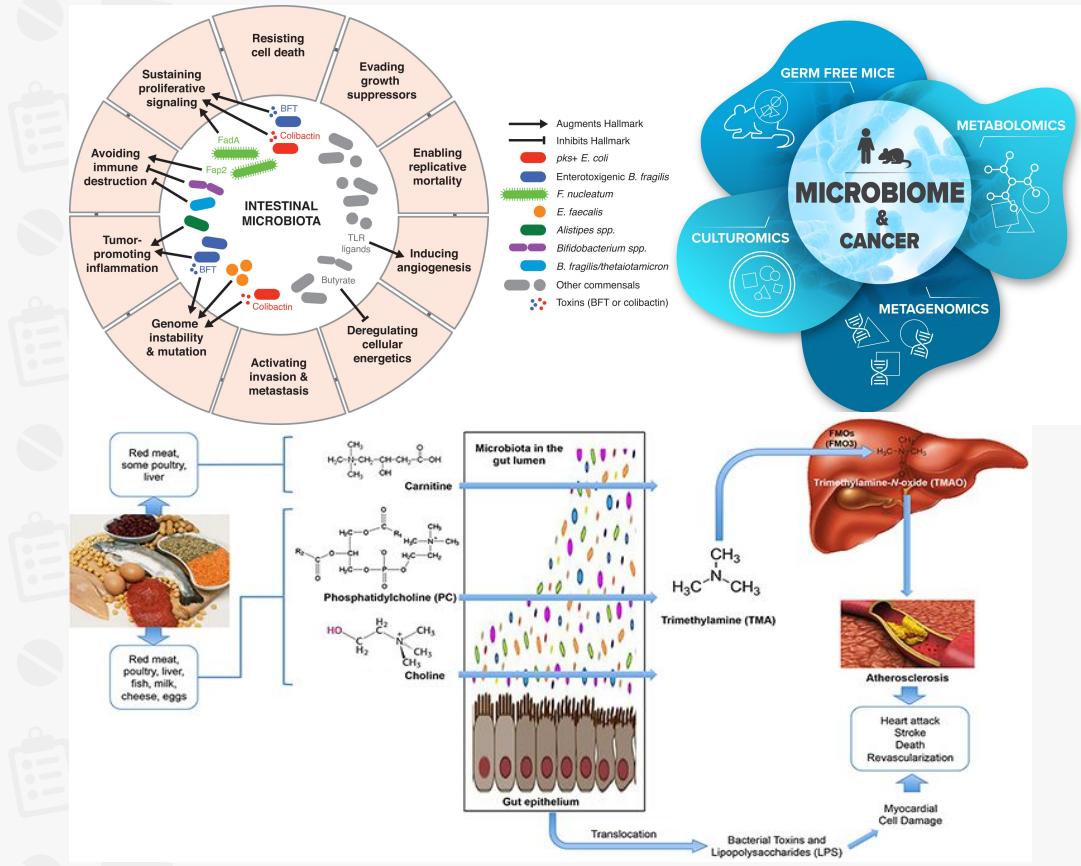
05

A brief focus on diseases

A brief focus on diseases

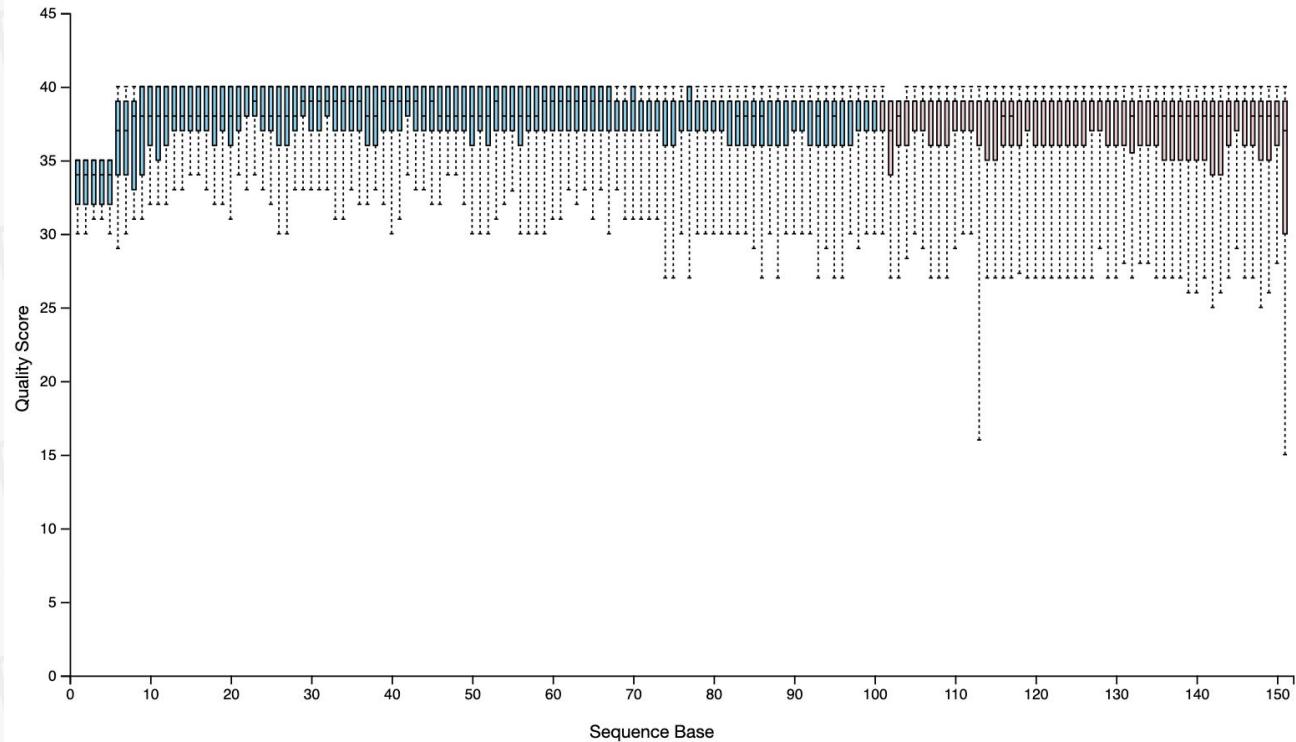
Perturbations to the microbiota also disrupt these homeostatic processes, promoting the development of numerous diseases such as different type of cancer and cardiovascular disease.

It's also clear that some specific bacteria genus partecipates in the onset of cardiovascular disease. In this final part we evaluated how much a borderline lifestyle is linked to these kind of pathologies.



A brief focus on diseases - Qiime2 analysis

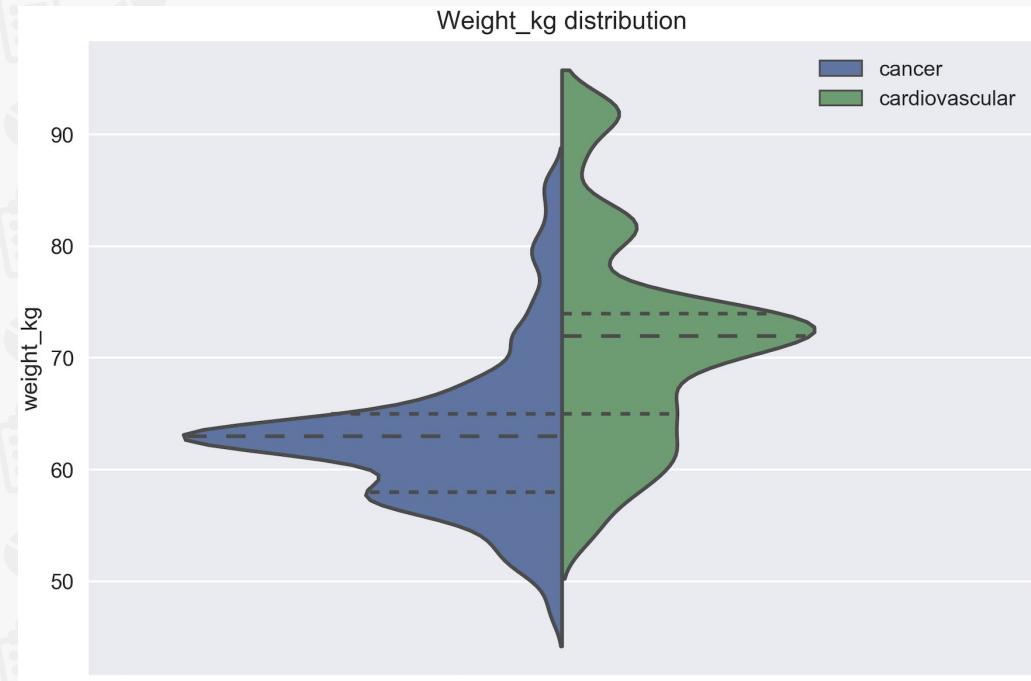
The quality of
the sequences
used for this part
of the analysis
also seems to be
enough high.



A brief focus on diseases - GUT analysis

The weight distribution reflects the loss in cancer patients due to the effects of the pathology and the treatments used.

It's not surprising to see weight gain in cardiovascular patients.



<https://www.nature.com/articles/0802243>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3250069/>

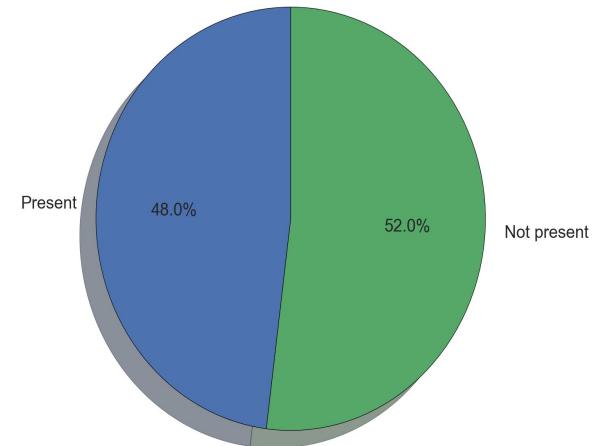
A brief focus on diseases- Bacteria analysis cardiovascular

Akkermansia muciniphila

The presence of Akkermansia muciniphila has significantly decreased among the samples of cardiovascular patients. It seems that this bacterium also plays an important role in reducing the size of atherosclerotic plaques.

<https://openheart.bmj.com/content/6/1/e000993>

Akkermansia muciniphila for: Cardiovascular



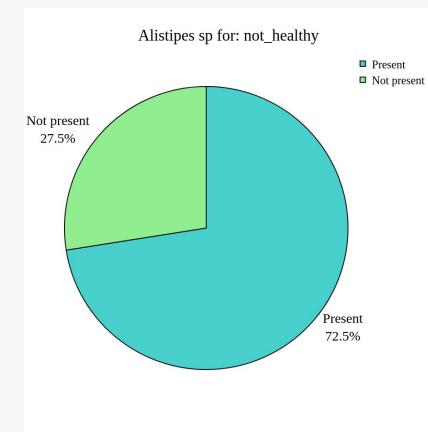
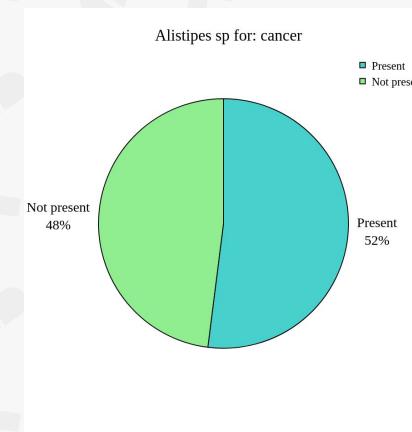
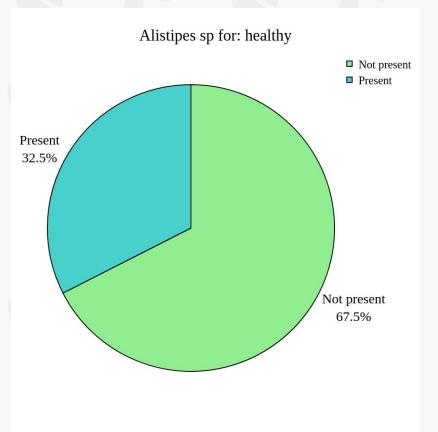
A brief focus on diseases - Bacteria analysis cancer

Alistipes spp.

The *Alistipes* genus includes obligate anaerobic gram negative bacteria. Its presence seems to increase among the cancer samples and not-healthy samples compared to healthy. This genus of bacteria may contribute to inflammation and tumorigenesis with enhanced IL-6 production, STAT3 activation, epithelial hyperplasia, and epithelial barrier dysfunction.

<https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1006480>

<https://www.sciencedirect.com/science/article/pii/B9780128023044000128>



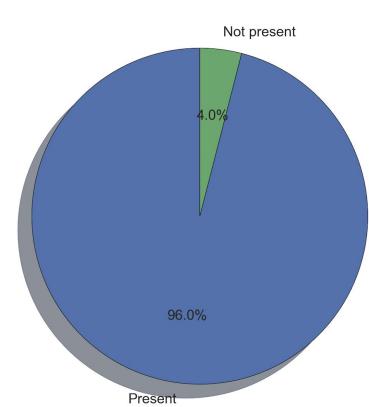
The situation is not always totally compromised!!!

Some specific gut microbes can modulate cancer immunotherapy (ICT). Several reports in the past 5 years have shown that alterations in the gut microbiota influence immunotherapy efficacy in mice, such as *Bacteroides dorei*, *Akkermansia muciniphila*, *Bacteroides uniformis* and *Bacteroides fragilis*.

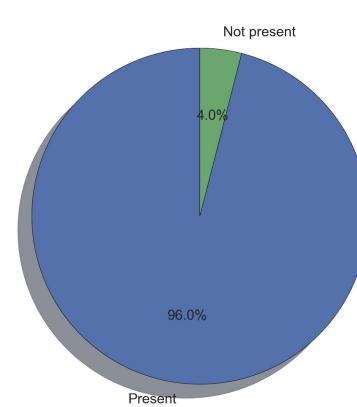
Many of our samples shares the presence of these kind of bacteria, maybe reflecting the hopes in these therapies.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6482659/>

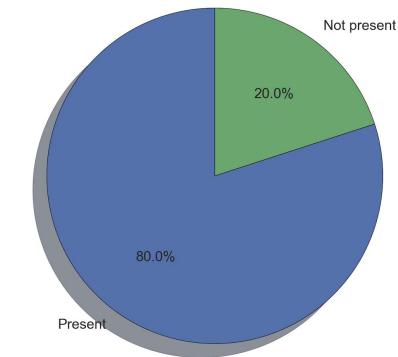
Bacteroides uniformis for: Cancer



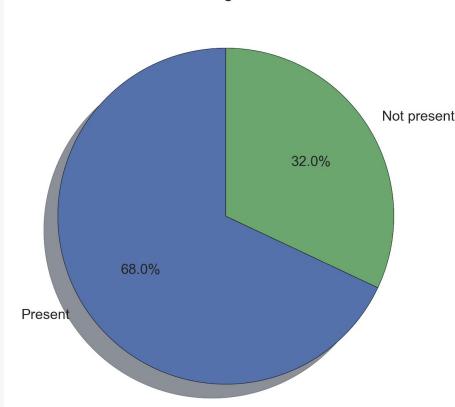
Bacteroides dorei for: Cancer



Akkermansia muciniphila for: Cancer



Bacteroides fragilis for: Cancer



06

Future development

Future development

In order to improve analysis result data, could be useful (and feasible), counting bacteria frequency inside the single sample.

Other Quiime2 analysis could be implemented in order to increase the strength of our statements (a good starting point could be the comparison with some specific bacteria databases , such as Silva or Greengenes).

Improving sampling techniques.

...and a degree in biological sciences with specialization in microbiology



Thank you for the attention!

