Microbiome is a term that describes a microbial community. It is used in relation to an ensemble of microorganisms associated with some host organisms. In the case of this talk, the host for the micrbiome is human. More specifically I will focus on the microbial community that resides in the large intestine of the human gut.

2 kilograms of bacteria!!

When thinking about the gut you should picture it, as a dynamic ecosystem in which microorganisms try to colonize it, they thrive there and are actively involved in proper functioning of the host organism. Microorganisms are also involved in different symbiotic interactions with each other, which is exemplified by their metabolic co-operation or competition for resources. Finally they are influenced by abiotic factors, like drugs, diet and other elements of that environment.

As an introduction, there are some general facts about the gut microbiome that I would like to bring up.

- Everyone of you, on average, carries 2 kg of microorganisms in the large intestine. The composition of your microbial community is very personal. And as I have already said, it is very dynamic. Antibiotics are one of the agents that can dramatically and permanently change the composition of you microbiome. Furthermore, many of you have travelled a long distance from your home country to participate in this conference and the change of diet that you experience will also affect you microbial community but rather in a temporary fashion.
- Another point, is that although each of us carries quite unique microbiome people can be grouped into a few enterotypes based on some similarities in microbial composition and its functional potential. These enterotypes are microbial constellations that do not correlate with host factors (e.g. age, nationality) but rather represent potential specialization of the gut community in how it generates energy from fermentable substrates available in the diet.
- Furthermore, a human on average has approximately 200 species in the gut microbiome. However, based on large cohort studies we know that the repertoire of possible gut microbiome species exceeds that number at least 3-fold. What is crucial, is that most of the microbial species that we find cannot be cultivated in a laboratory and we do not have their reference genome.
- Finally, another way of looking at the microbiome is through its gene repertoire. The latest study based on more than 1000 individuals from 3 continents, which included Spanish, Danish, Chinese and American individuals, report a microbial gene catalog containing 10 million genes, which exceeds the gene content of human genome 400-fold.

Enterotypes:
Bacteroides -
Prevotella -
Ruminococcus -

We are done with a short introduction and this is an overview of what I would like to talk to you about in more detail.

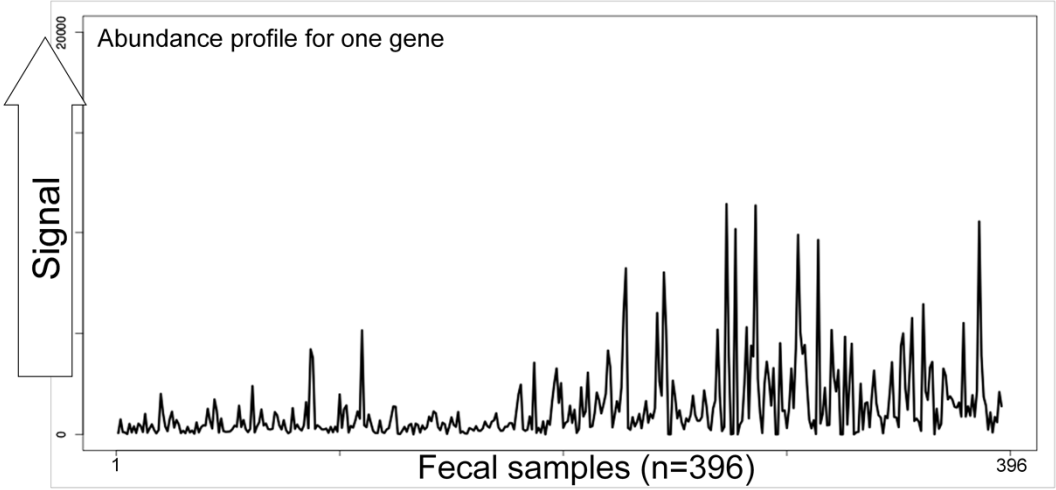
I will start with a method that we developed to identify species and other entitites in the gut microbiome. Then I will show you how we study and identify microbial persistence. I will finish my talk will with some of the on-going work on a symbiosis in microbial community.

The co-abundance principle

Why do we do it?

The way we identify species is based on the co-abundance principle.

Abundance profile for one gene

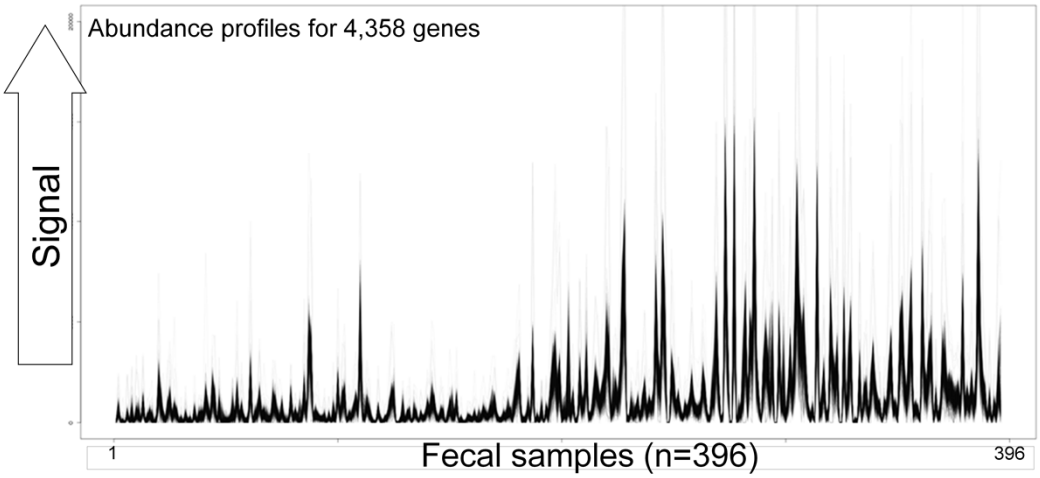


Before explaining this plot, I will just briefly mention that we took a stool sample from 396 individuals, sequenced them, identified genes and measured abundance of each gene in every sample.

Here I show abundance of a specific gene (y-axis) plotted against 396 fecal samples. As you can see the abundance of that gene varies between samples, and in some of them it is barely detectable. Of course this gene comes from an organism and other genes from its genome are also present in the sample.

And when I plot the abundance of all genes that organisms, using some semi-transparent lines, it looks like this.

a metagenomic species (MGS)



What you can notice is that genes from that organisms nicely follow each other in terms of abundance.

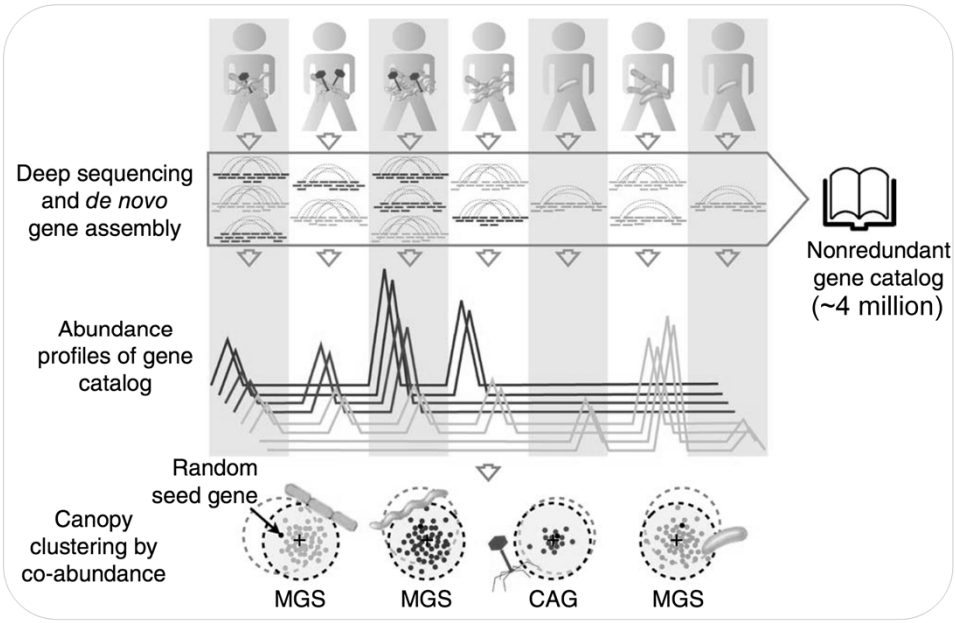
In more mathematical terms, what you see is that there is a fixed ratio between amount/abundance of genes in a single genome. Importantly, that ratio is not affected by the amount of that organism in a sample, as seen in the plot.

If the genes come from the same genetic entity (DNA) then they have the same abundance. Genes in physical linkage have the same abundance..

In other words if you were to correlate abundance of one of the genes in that organism to some other genes, across all these samples, you would see that the correlation would be very close to 1.

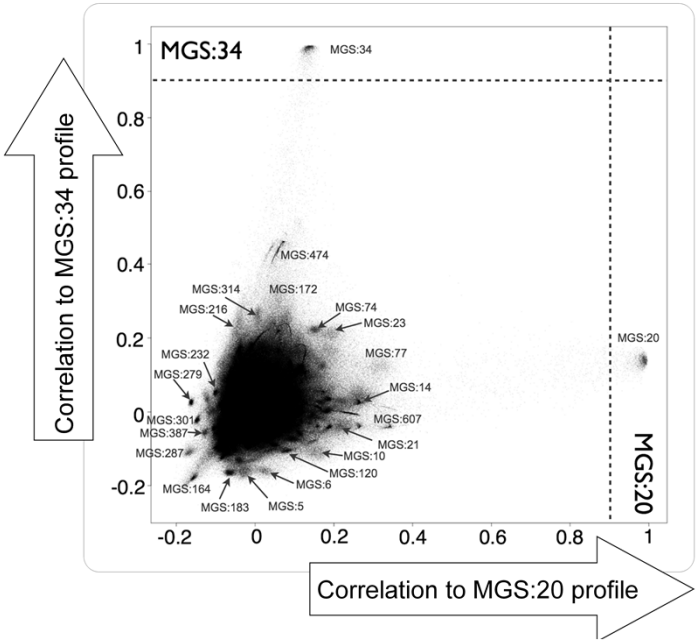
This is, in fact, the essence of the co-abundance principle that we use to identify species and other entities in the microbiome.

As you might have already noticed, I call this organisms for metagenomic species, MGS in short, which indicates that it has been derived from the metagenomic sample using co-abundance principle.



H. B. Nielsen et al., Nat. Biotechnol. (2014)

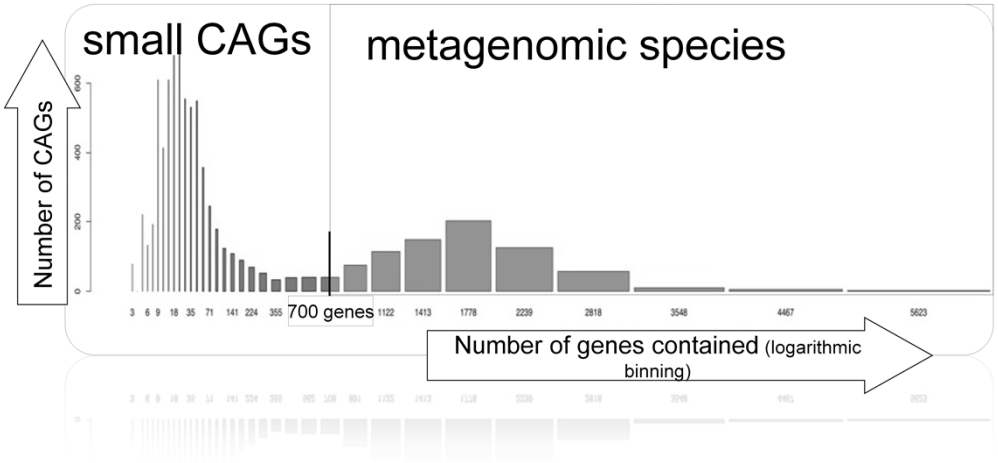
Genes are correlated



When we find a cluster of genes, hence an organisms, then we can calculate its median abundance. As I said, genes from the same organisms correlate in abundance. Here we look at 4 million microbial genes, where we plotted their correlation to two organisms, MGS:34 and MGS:20. In the right side of the plot you can see a cloud of genes that belong to MGS:20 and on top there are genes belonging to MGS:34.

What is also interesting is that in the center of the plot the cloud of points is not a uniform blop, but you can also distinguish clouds of gene highly correlated to each other. These are other metagenomic species.

7381 co-abundance gene groups (CAGs)



7381 CAGs

Size range: 3 - 6.319 genes

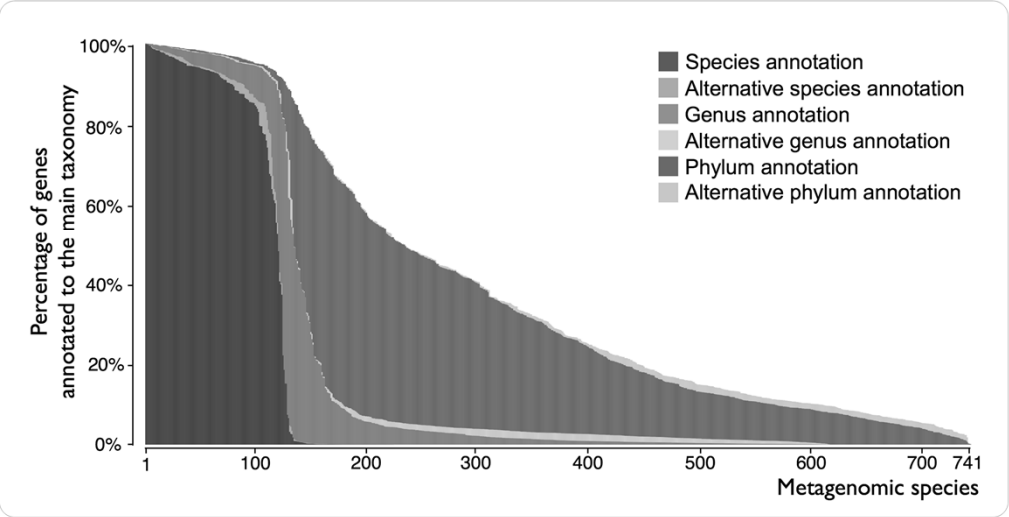
Total number of genes binned: 1,528,079

741 gene rich MGS (> 700 & median 1656 genes)

6640 Small MGS

1,253,770 genes were binned into gene rich MGS

Consistent taxonomy



> 700 genes

All the genes in gene catalog where annotated for the taxonomy by blasting them against reference databases and keeping the top annotation that passed our criteria.

It turned out that quite many metagenomic species have a very consistent taxonomical annotation. In this barplot, for each metagenomic species, I show on the y-axis percentage of genes annotated to the main taxonomy group, at species level (red), genus (green), or phylum (blue), given of course that the metagenomic species had any annotation at that taxonomy level.

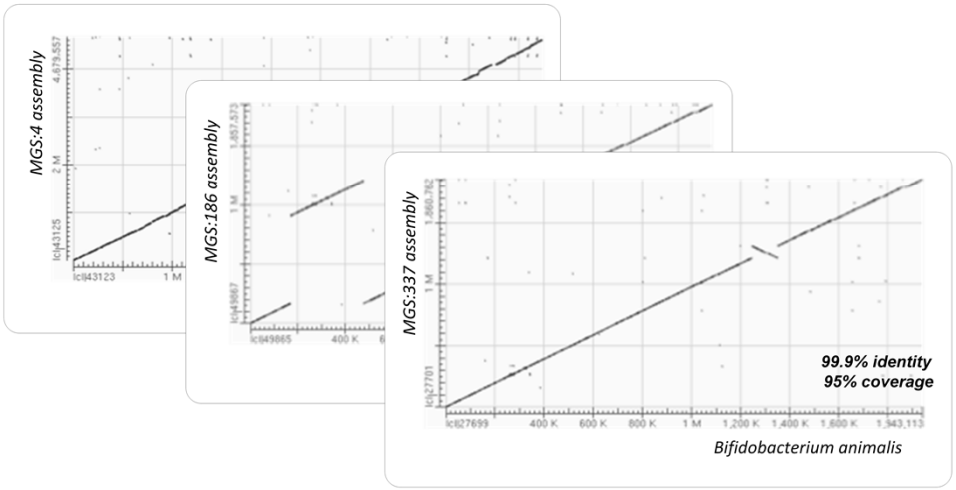
E.g. one of the red bars.. while one of blue bars…

At the point of writng 518 metagenomic species which we identified did not have species level similarity to any previously sequenced genome.

Species, genus and phylum level taxonomical annotation was defined as best sequence match with 95%, 85% and 75% identity over ≥ 100 bp

opaque vs transparent

Compared to known genomes

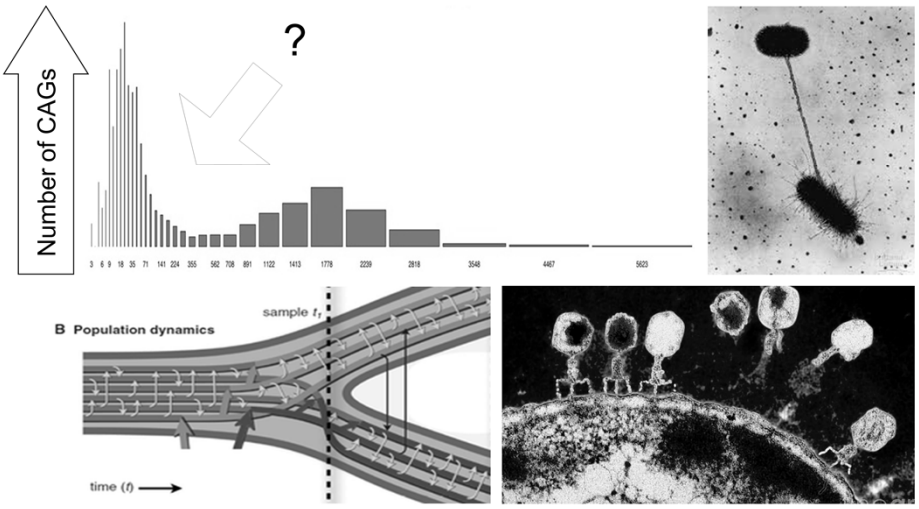


360 MGS assemblies passed **high quality draft assembly criteria**

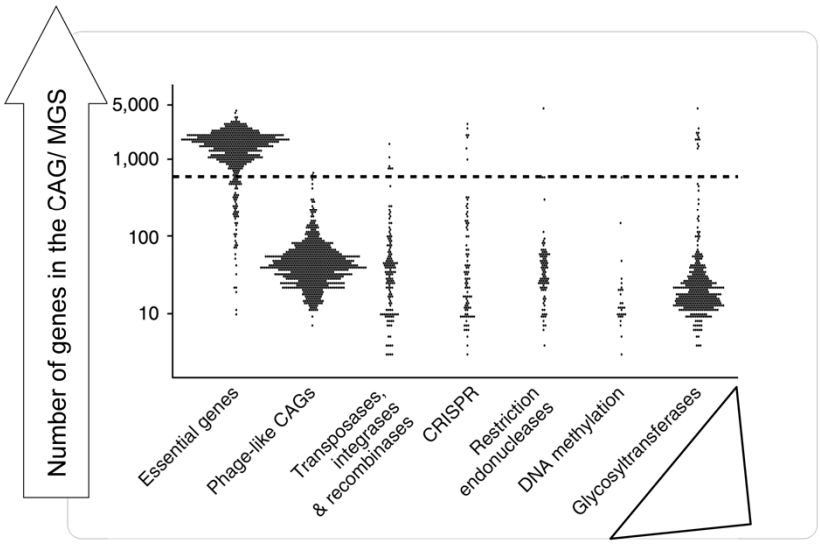
The fact that we can distinguish particular species from all the data, allowed us to go one step further and we could assemble 360 metagenomic species to a high quality drafts. Here I am showing examples of Escherichia coli, Methanobrevibacter smithi and Bifidobacterium animalis whose assemblies were plotted against their reference genomes.

!! maybe remove

Optional genomics

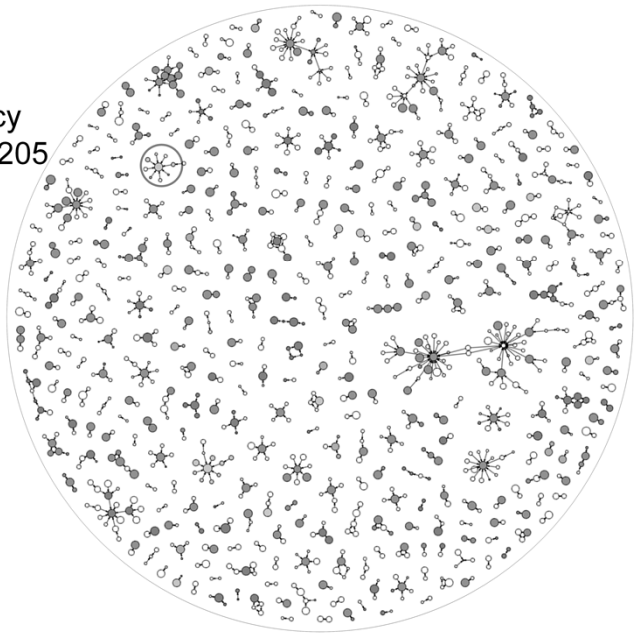


Biological functions



Dependency network

882 strong dependency associations among 1205 CAGs



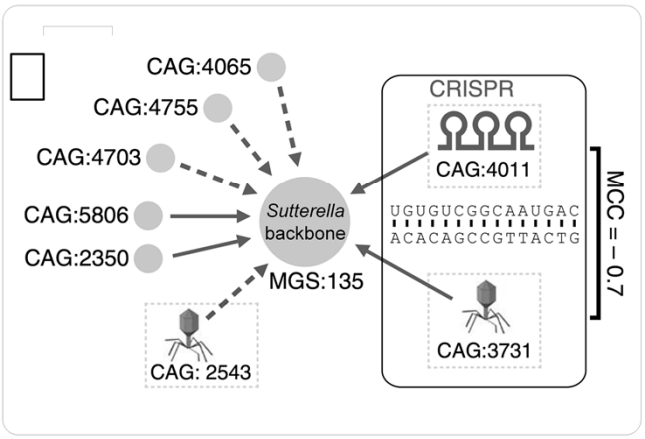
Mostly small CAGs to MGS dependency associations
(odd ratio 12.7)

!! maybe remove

Example: Sutterella

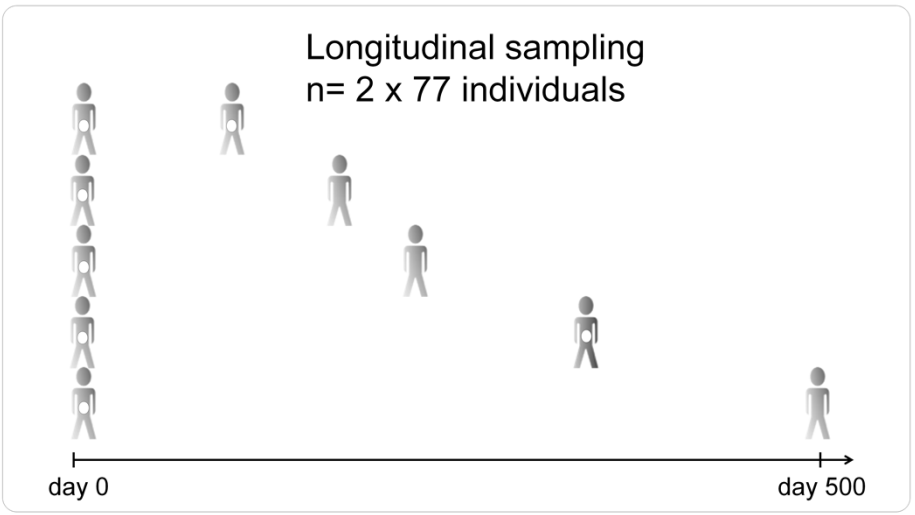
Several phages and a CRISPR containing CAG are dependency associated to the *Sutterella wadsworthensis* backbone MGS.

The associated CRISPR and phage-like CAGs appear to ‘avoid’ each other and CRISPR spacers match the phage DNA



Here is an example of Sutteralla cluster where the central node is the backbone or core of its genome, and the surronding nodes pointing towards it represent the optional genetic elements. Two of them represent bacteriophages and one contain CRISPR system, which is an adaptive immune system of bacteria and archaea that learns and protects against foreign DNA attacks. This CRISPR containing CAG:4011 and bacteriophage CAG:3731 were negatively correlated

Persistence

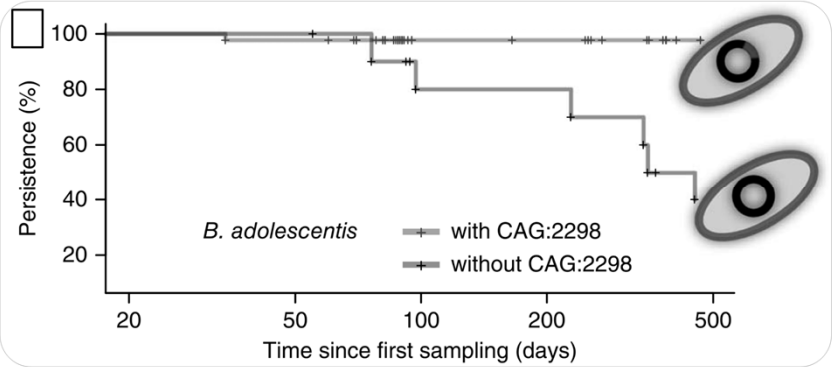


Persistence is the state of existing.
And

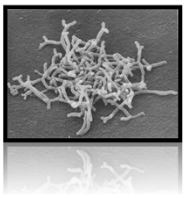
Conditional persistence.

Change colors

The little difference

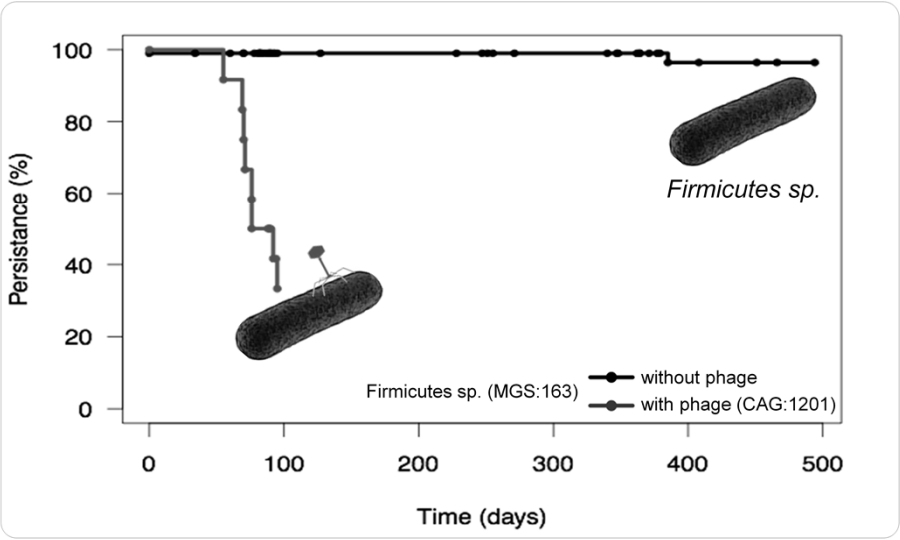


Across different species, the optional gene sets that enhance persistence were enriched for genes with important for resistance towards free-radicals

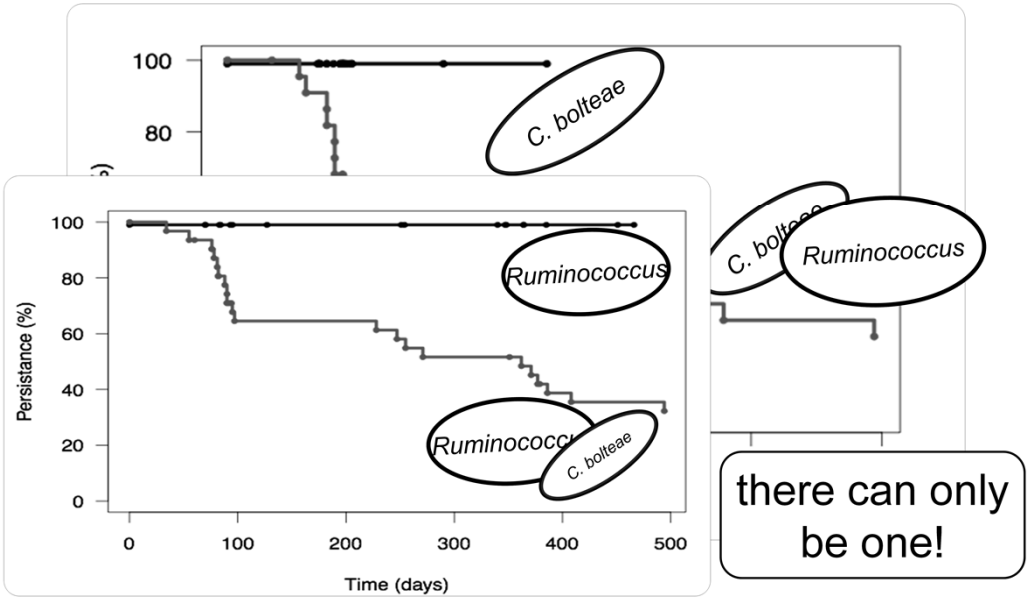


Bifidobacterium adolescentis

Co-existing with a phage – or not

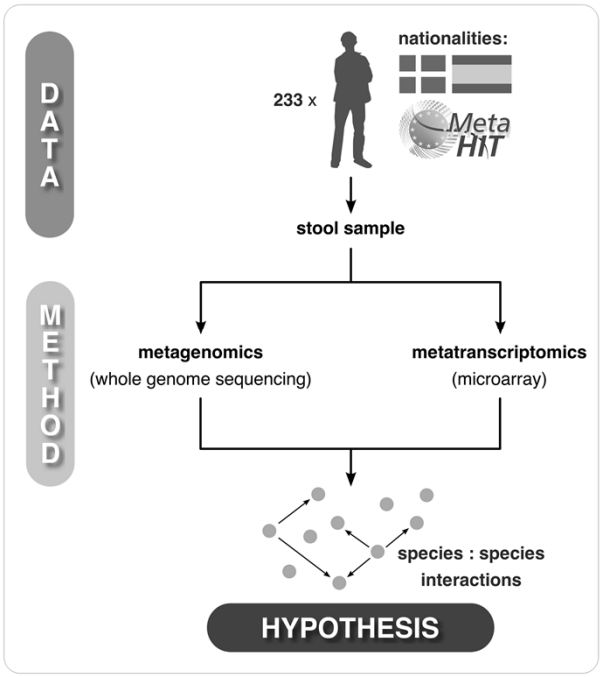


Competition may result in extinction

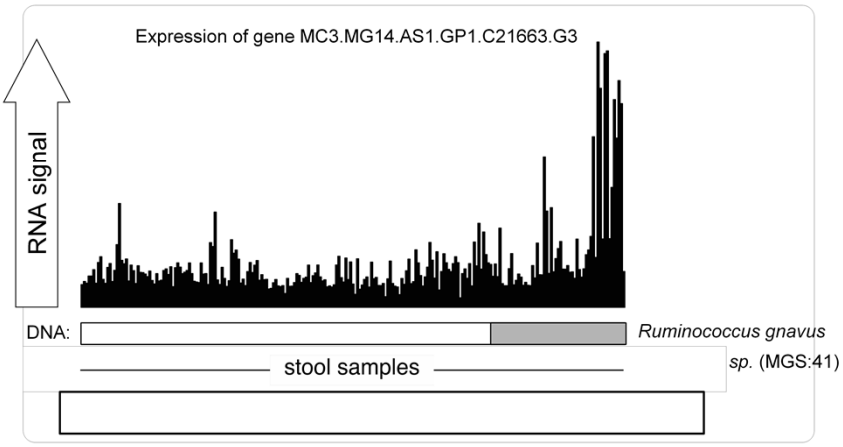


Genetic potential

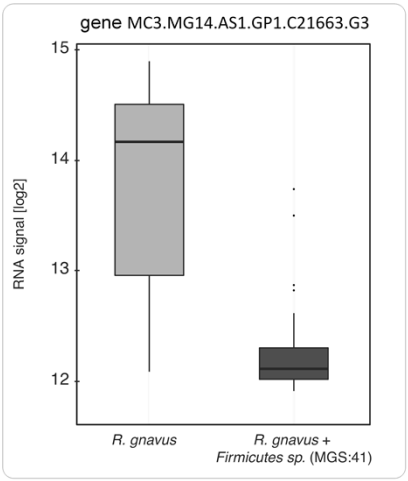
Transcriptional activity


DATA
METHOD
233 x
nationalities:
MetaHIT
stool sample
metagenomics
(whole genome sequencing)
metatranscriptomics
(microarray)
species : species
interactions
HYPOTHESIS

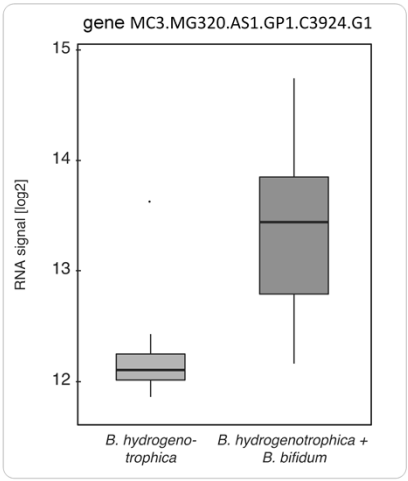
Gene expression is altered by co-existence



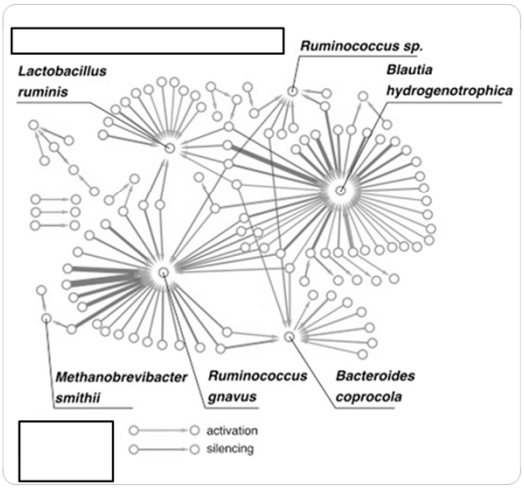
silencing

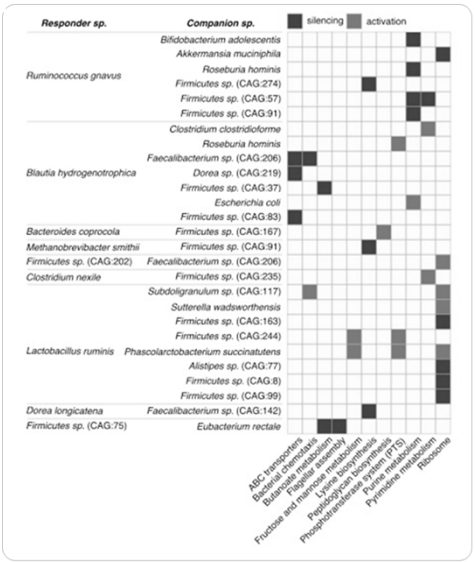


activation



Transcriptional adaptations





Conclusion

- Gene co-abundance principle allows for identification of **novel microorganisms**, **viruses** and **clonal variation**
- Smaller CAGs carry functions relevant for **inter-biotic interactions**
- Despite the complexity, patterns of **transcription** and **persistence** provide insights about **microbial symbiosis**

Perspectives:

- persistence / transcriptional interactions: learning about antagonistic relationships could be a way of identifying potentially therapeutic phages and bacteria.

The Team


Henrik Bjørn Nielsen


Mathieu Almeida


Agnieszka Juncker


Simon Rasmussen


Thomas Sicheritz Pontén


Søren Brunak


S. Dusko Ehrlich

MetaHIT

INRA
EMBL
BGI
Genoscope
Technical University of Denmark
Hospital Universitarii Vall d'Hebron
NNF Center for Basic Metabolic Research
Wageningen Universiteit

The Wellcome Trust Sanger Institute
Copenhagen University
Mérieux Alliance
SDU
UCB
Istituto Europeo di Oncologia - Italy
Danone