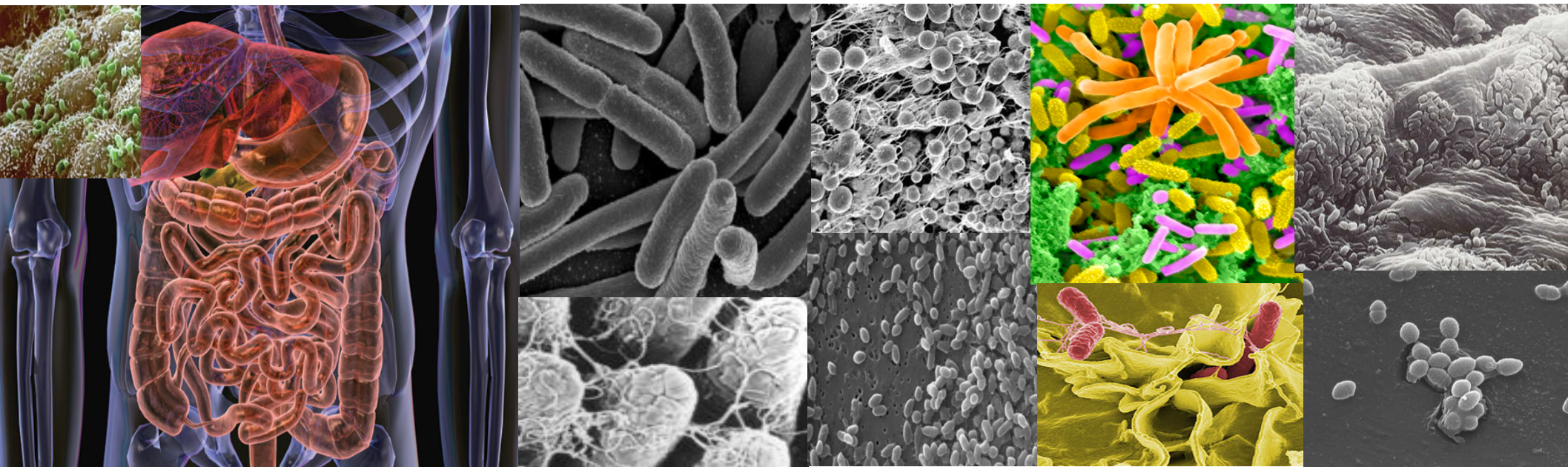# Introduction to metagenomics

Agnieszka S. Juncker

Center for Biological Sequence Analysis
Technical University of Denmark
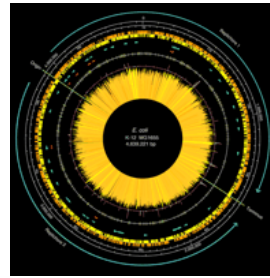
Outline

- Metagenomics
- The human gut

From genomics to metagenomics
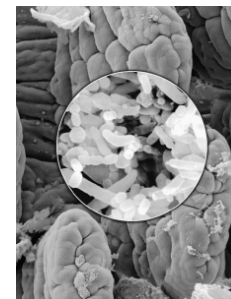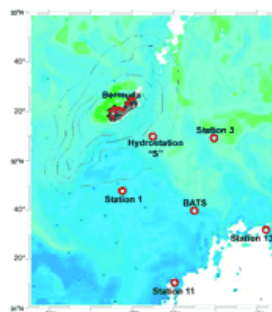
Genomics

E. coli, Science, 1997
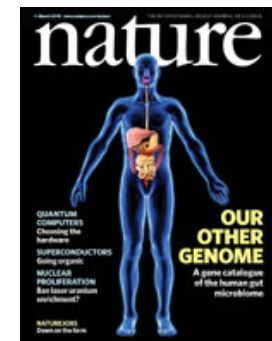

Human, Nature/Science, 2001

Metagenomics

Saragasso sea, Science, 2004

Human gut, Nature, 2010

What is Metagenomics?

Metagenomics (Environmental Genomics, Ecogenomics or Community Genomics) is the study of genetic material recovered directly from environmental samples.

Chen & Pachter, 2005
Metagenomics is application of modern genomic techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species

About Metagenomics

A) Most microbial activities are carried out by complex communities of microorganisms ...

B) 99% of microbial species cannot currently be cultivated

A hand full of soil ...

- Culturing: a few hundreds species per gram
- 16S sequencing: few thousands per gram

Why Metagenomics?

Discovery of:
- novel natural products
- new antibiotica
- new molecules with new functions
- new enzymes and bioactive molecules

- what is a genome/species
- diversity of life
- interplay between human and microbes
- how do microbial communities work and how stable are they
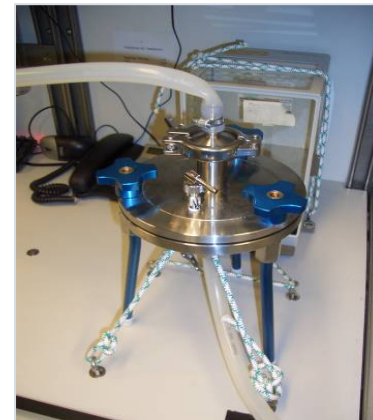- holistic view on biology

Environments

Sample preparation

Design of study and sampling (sample size, timing, replicates)
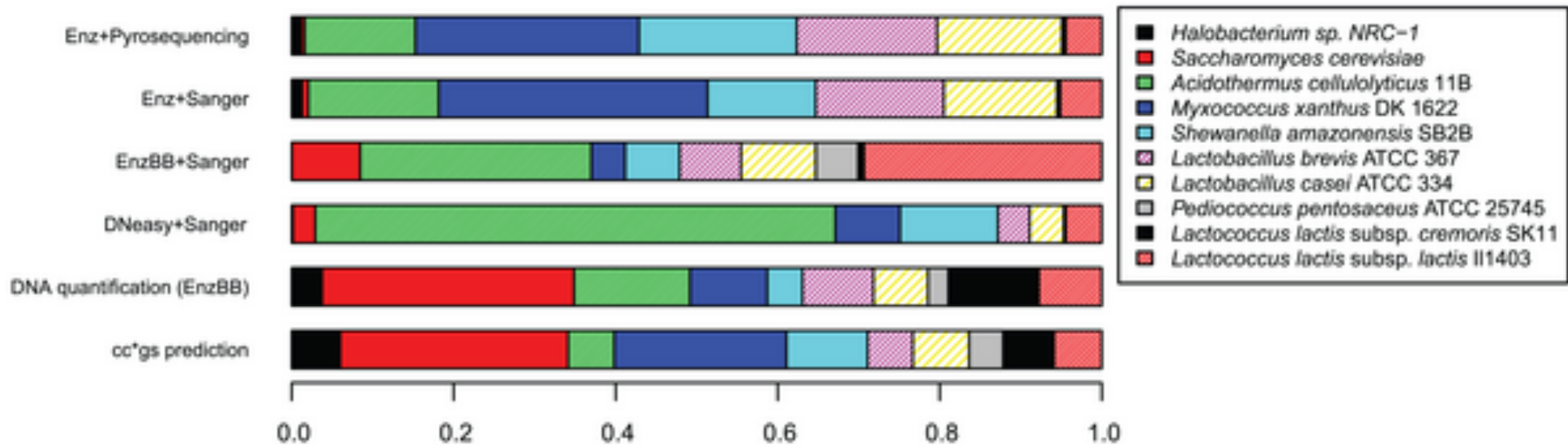
Avoid contamination

Pre-treatment, e.g. filtering
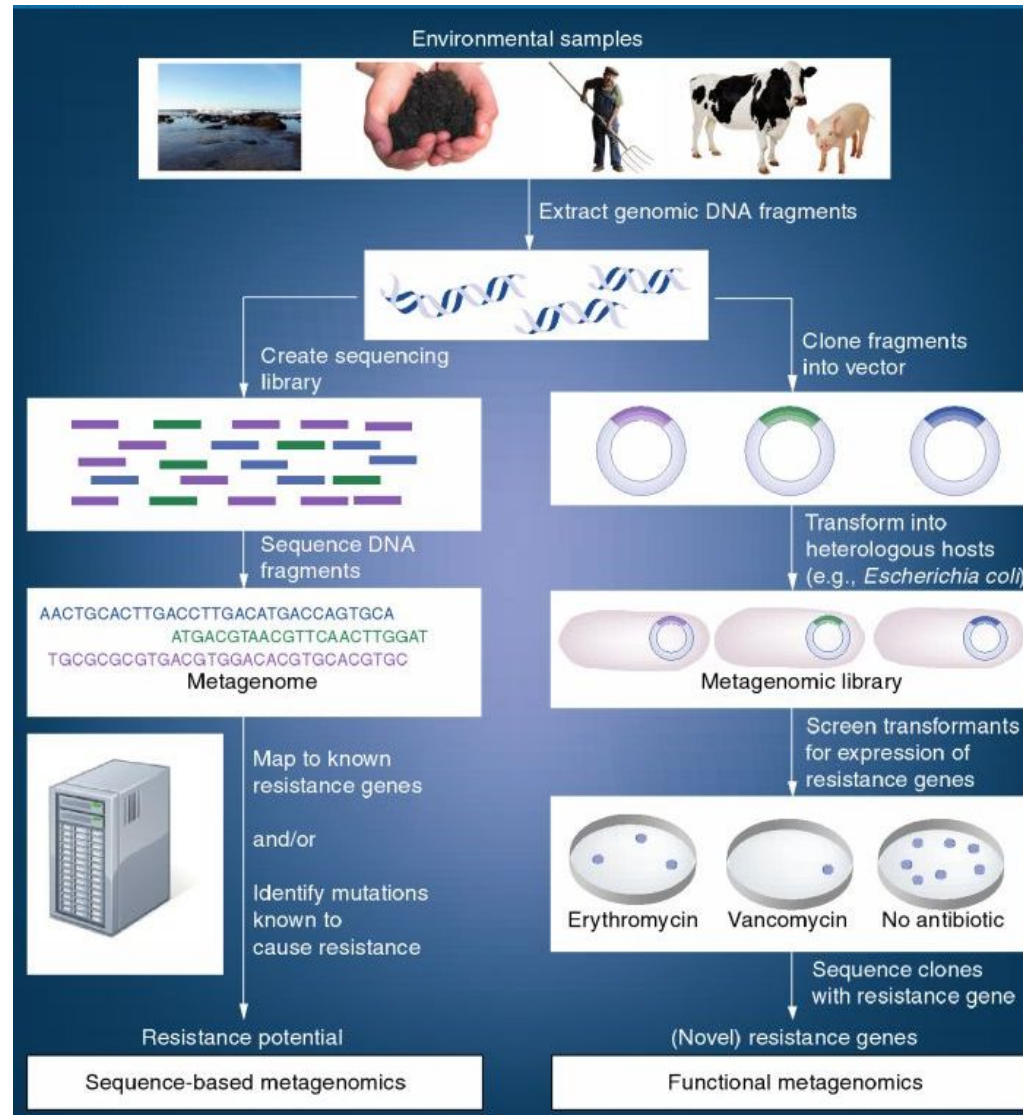
DNA extraction from sample

Lysation and DNA extraction, many methods availble, different biases

Metagenomics approaches

Sequence-based
(computational)

Functional
(experimental)

Sequence-based metagenomics

16S rDNA sequencing   **Whole-genome sequencing**

Sampling
DNA extraction

PCR+sequencing of 16S rDNA
Phylogeny analysis

Sequencing

Assembly

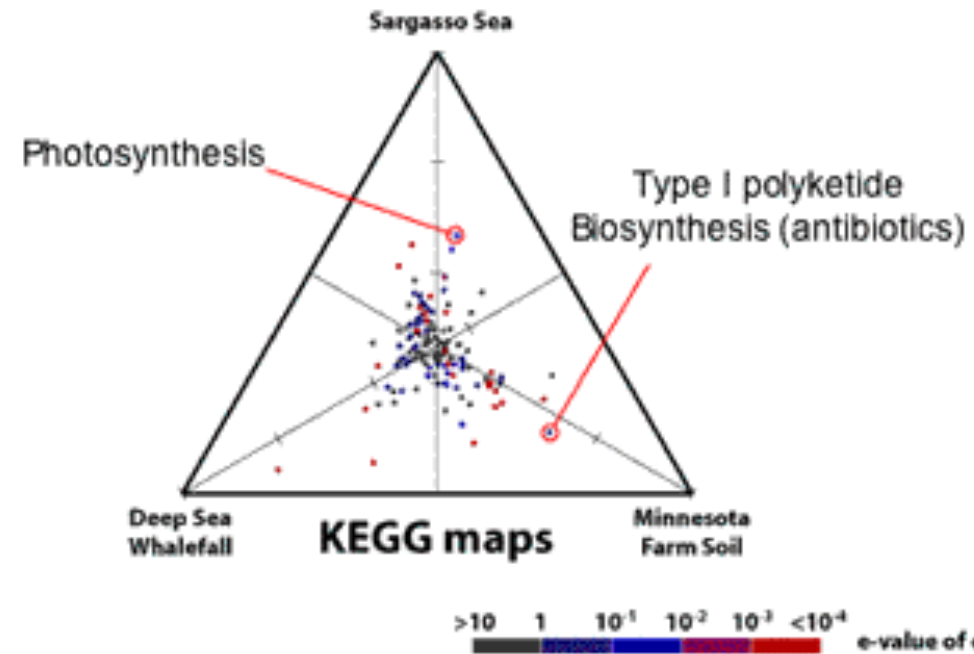Comparison

Gene finding
and annotation

Comparison

Metagenomics data analysis
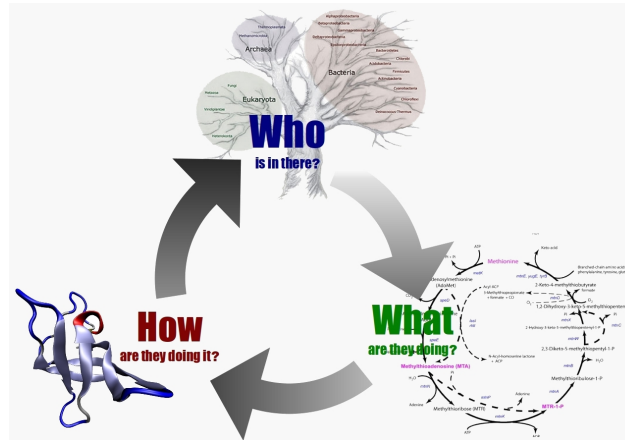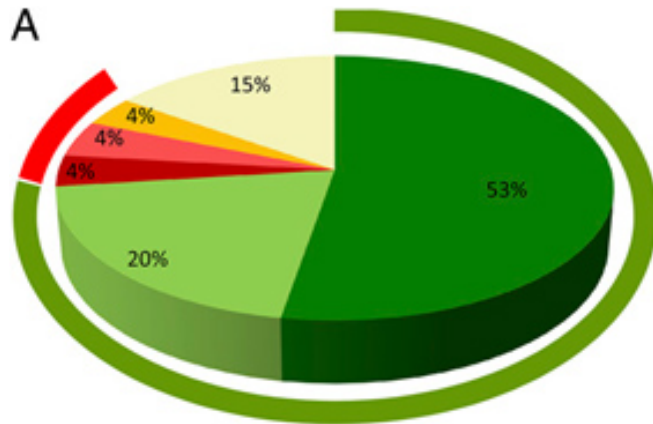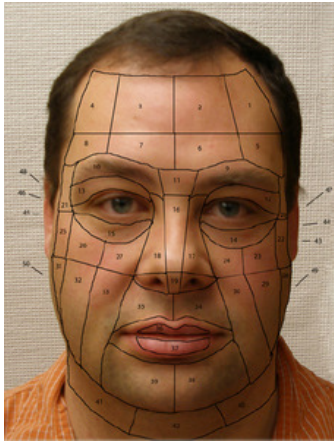

Taxonomy annotation

Functional annotation
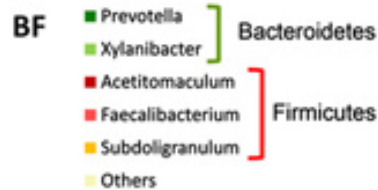
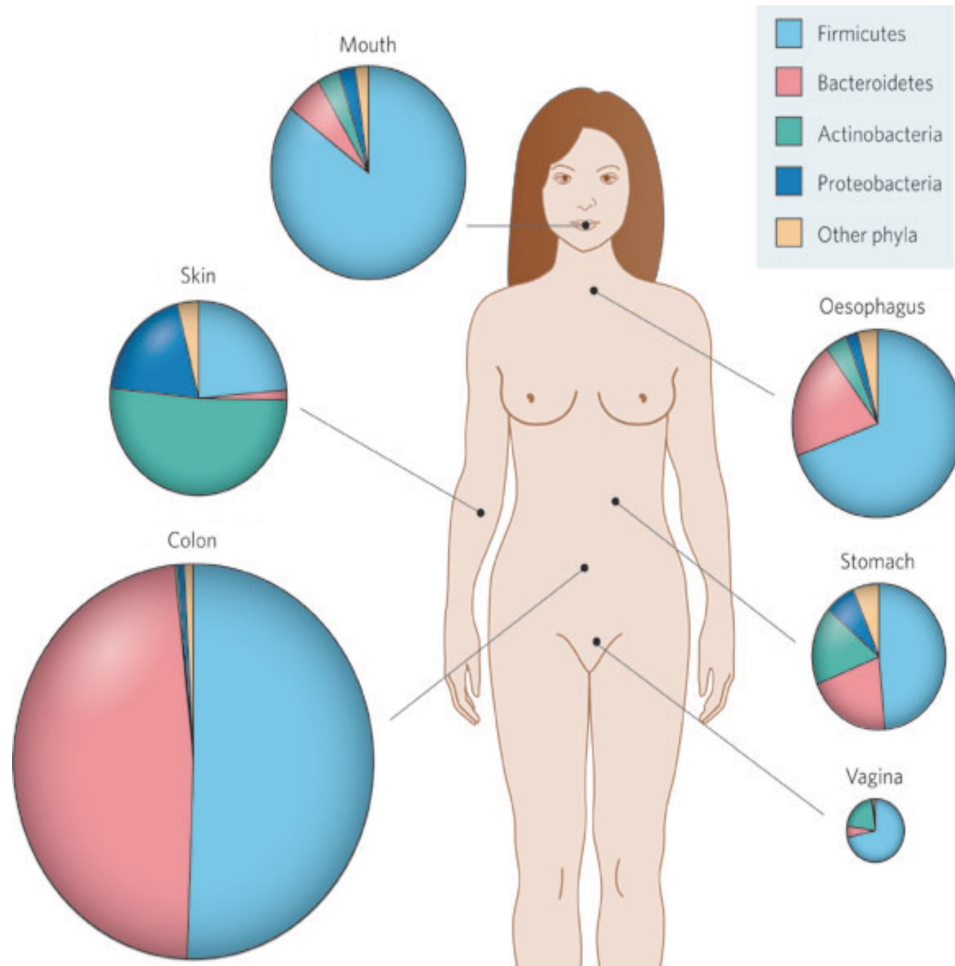# Metagenomics data analysis – wrap up

- Sequence reads

- Assembly (contigs)

- Gene prediction

- Count matrix calculation (in case of many samples)

- Taxonomy annotation (BLAST, LCA)

- Functional annotation (COG, KEGG, GO)

- Main statistical analysis
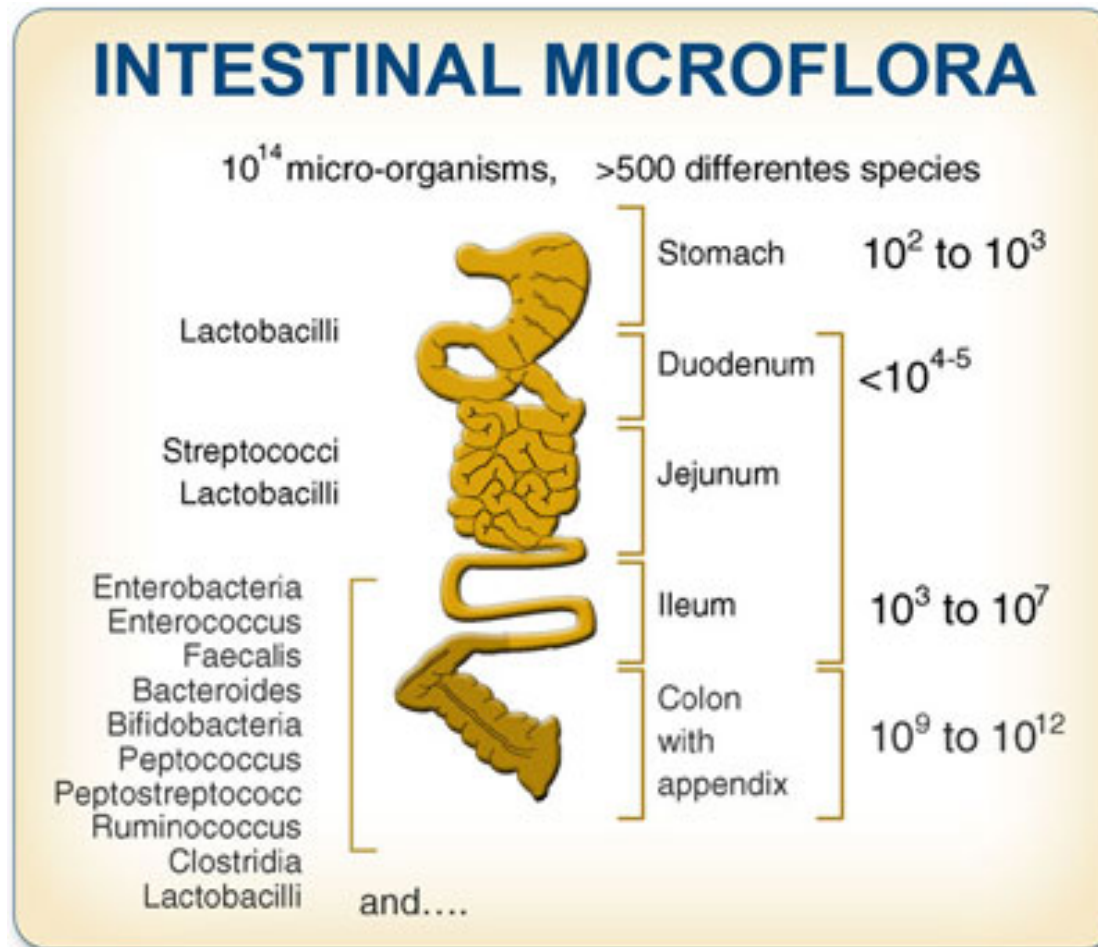
Project examples

The human microbiome

Human Microbiota

Human intestines

Metagenomics of the human gut

ARTICLES

# A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin[1]*, Ruiqiang Li[1]*, Jeroen Raes[2,3], Manimozhiyan Arumugam[2], Kristoffer Solvsten Burgdorf[4], Chaysavanh Manichanh[5], Trine Nielsen[4], Nicolas Pons[6], Florence Levenez[6], Takuji Yamada[2], Daniel R. Mende[2], Junhua Li[1,7], Junming Xu[1], Shaochuan Li[1], Dongfang Li[1,8], Jianjun Cao[1], Bo Wang[1], Huiqing Liang[1], Huisong Zheng[1], Yinlong Xie[1,7], Julien Tap[6], Patricia Lepage[6], Marcelo Bertalan[9], Jean-Michel Batto[6], Torben Hansen[4], Denis Le Paslier[10], Allan Linneberg[11], H. Bjørn Nielsen[9], Eric Pelletier[10], Pierre Renault[6], Thomas Sicheritz-Ponten[9], Keith Turner[12], Hongmei Zhu[1], Chang Yu[1], Shengting Li[1], Min Jian[1], Yan Zhou[1], Yingrui Li[1], Xiuqing Zhang[1], Songgang Li[1], Nan Qin[1], Huanming Yang[1], Jian Wang[1], Søren Brunak[9], Joel Doré[6], Francisco Guarner[5], Karsten Kristiansen[13], Oluf Pedersen[4,14], Julian Parkhill[12], Jean Weissenbach[10], MetaHIT Consortium†, Peer Bork[2], S. Dusko Ehrlich[6] & Jun Wang[1,13]

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.

|	Human (isolated)	Microbiota
weight	~ 50-100 kg	~ 2 kg
species	1	1000-5000
cells	~ 10^{12}	10^{13} - 10^{14}
genes	25.000	>4.000.000

Bacteria of the gut
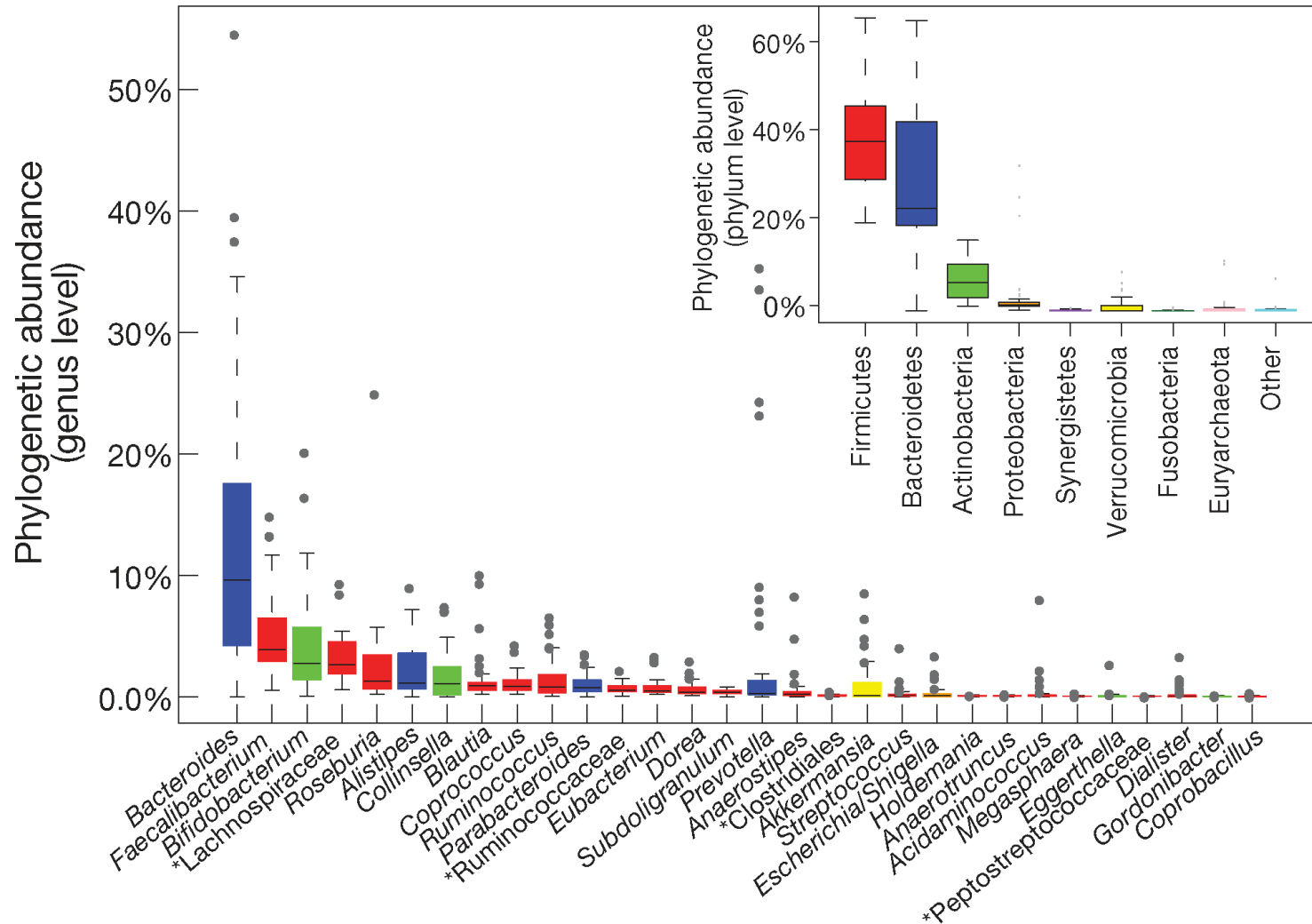
1 *Citrobacter sp. 30_2*
2 *Actinomyces odontolyticus*
3 *Enterobacter sp. 638*
4 *Streptococcus mutans*
5 *Lactobacillus gasseri*
6 *Pasteurella multocida*
7 *Pediococcus pentosaceus*
8 *Lactobacillus helveticus*
9 *Campylobacter concisus*
10 *Enterobacter cancerogenus*
11 *Lactobacillus reuteri*
12 *Lactobacillus sakei*
13 *Streptococcus suis*
14 *Streptococcus thermophilus*
15 *Actinobacillus pleuropneumoniae*
16 *Haemophilus parasuis*
17 *Cronobacter sakazakii*
18 *Haemophilus influenzae*
19 *Lactobacillus salivarius*

20 *Streptococcus gordonii*
21 *Lactobacillus johnsonii*
22 *Lactobacillus fermentum*
23 *Citrobacter sp.*
24 *Escherichia fergusonii*
25 *Streptococcus pneumoniae*
26 *Streptococcus pyogenes*
27 *Salmonella enterica*
28 *Citrobacter koseri*
29 *Escherichia coli*
30 *Lactobacillus hilgardii*
31 *Blautia hansenii*
32 *Clostridium scindens*
33 *Streptococcus sanguinis*
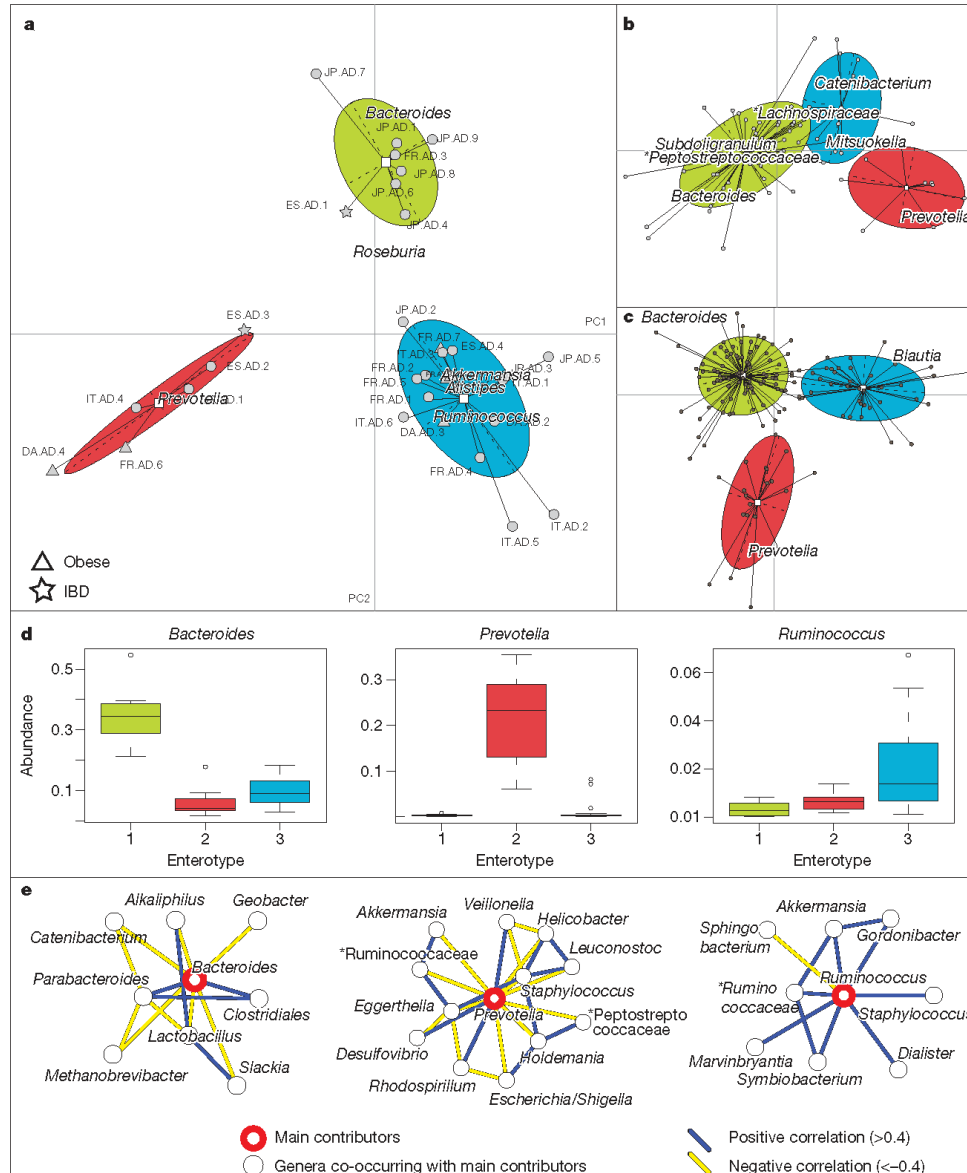34 *Clostridium perfringens*
35 *Lactobacillus ultunensis*

Actinobacteria
Bacteroidetes
Firmicutes
Fusobacteria
Proteobacter
Tenericutes

Positive correlation
Negative correlation

Qin *et al.* (Nature 2010)
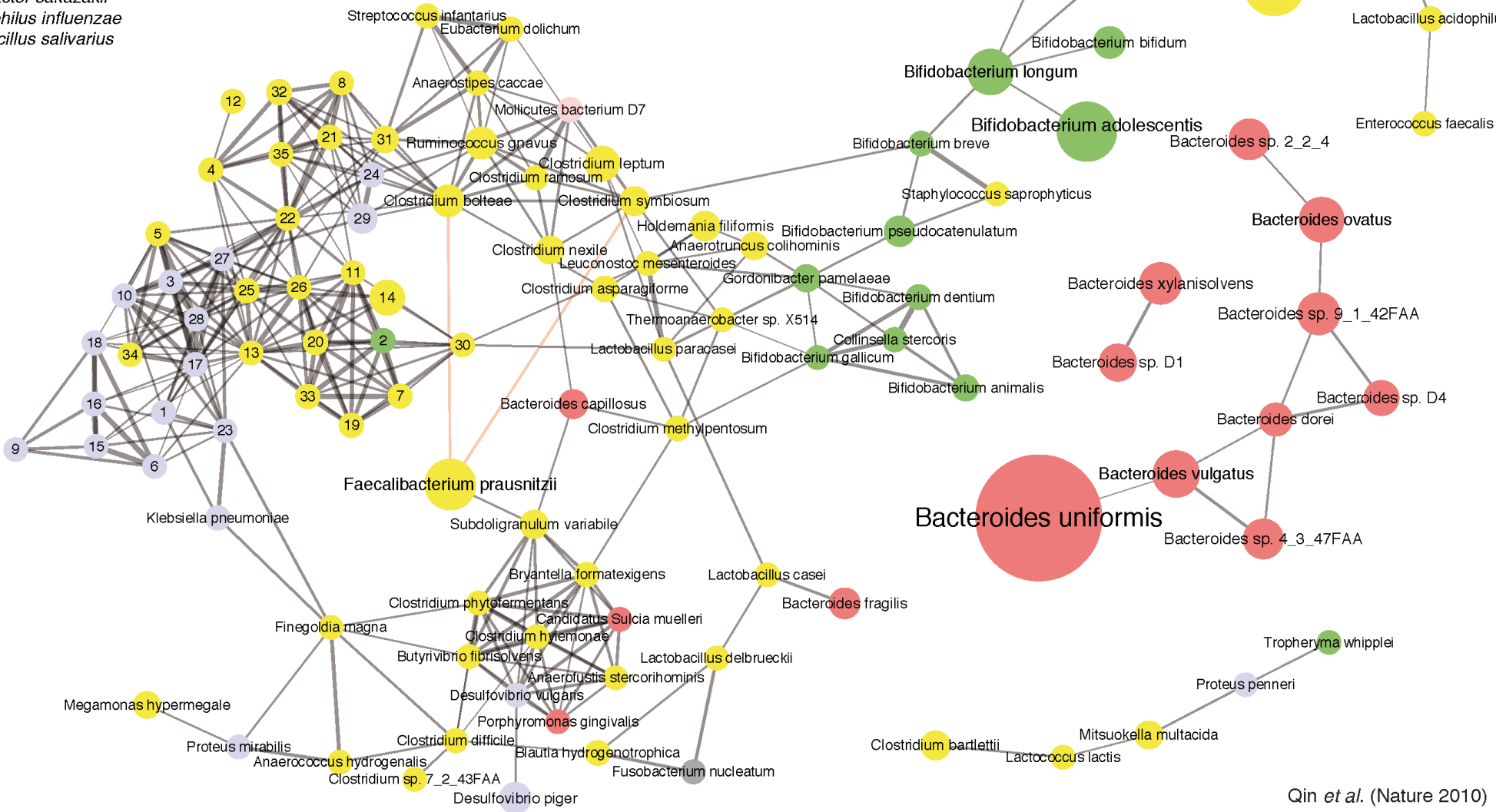
Functions of the human gut microbiome

Enterotypes

Arumugam *et al.* (Nature, 2011)
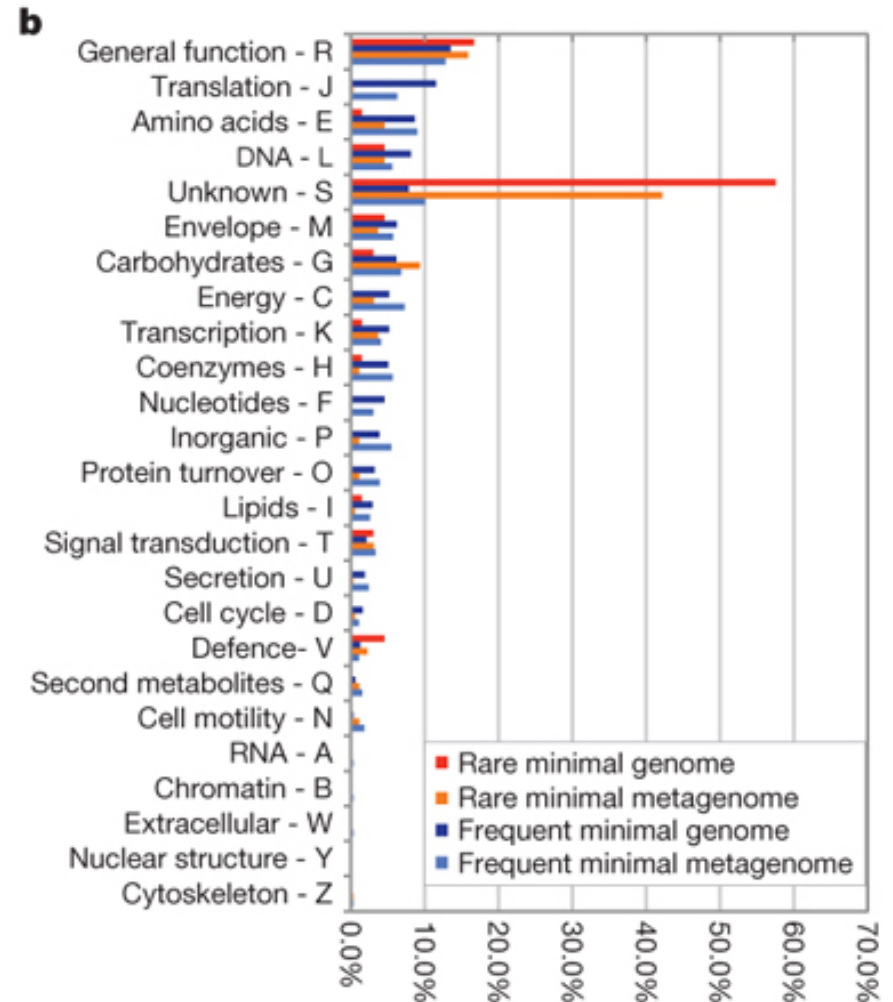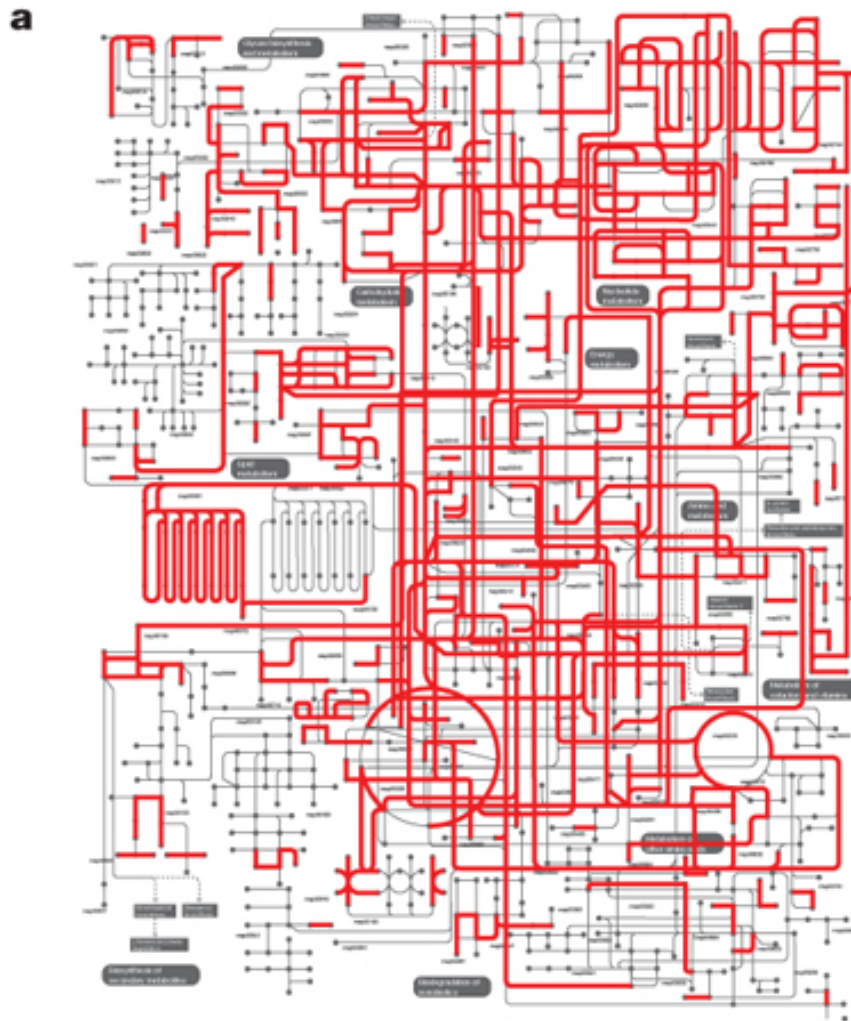
Acknowledgements

Damian Plichta

Marcelo Bertalan

H. Bjørn Nielsen

Falk Hildebrand
(from Jeroen's group)

Pia Friis

Lene Blicher

Laurent Gautier

Thomas Sicheritz Pontén

Søren Brunak