

This presentation is licensed under the Creative Commons Attribution 3.0 Unported License available at <http://creativecommons.org/licenses/by/3.0/>



HMP 2010, St Louis, MO
S. Dusko Ehrlich

Metagenomics of the **H**uman **I**ntestinal **T**ract

<http://www.metahit.eu>

The MetaHIT approach to relation between microbes & us

Establish a **reference gene set** by metagenomic & genomic sequencing of the Human GI tract microbes

Use the gene set to develop **tools for profiling** the GI tract microbiota genes : arrays and high throughput DNA sequencing

Use the profiling tools to search **association of microbial genes and disease** in Inflammatory Bowel Diseases and Obesity

Carry out function analysis to go **from associations to mechanisms**

Bioinformatics overlays all activities



The partners

- Thirteen European, one Chinese Institutions
- Eight countries, two continents
- Nine public and four private Institutions



Beyond & above all – the people, a stellar team!

MetaHIT budget : 21.2 M €

EC Contribution: 11.4 M €

100 scientists

Starting date: March, 2008

Duration: 4 years

Illumina-based intestinal bacteria reference gene set



Qin et al., 2010

Illumina sequencing

Samples	124 individuals (85 Danes, 39 Spaniards)	
Library type	15 samples	200bp
	109 samples	140bp
		350bp
Sequencing type	Paired-end (PE) sequencing	
Read length (bp)	45 b (15 samples) 75 b (109 samples)	
Sequences per sample	31million \pm 0.5 million	

In total, **~0.58 Terabase** of sequence

A high-quality contig set

- SOAPdenovo (de Bruijn graph-based tool)
- Removal of short contigs (<500bp)
- Removal of redundancy

Total Size	Number	N50 Size	N90 Size	Max. Length
10.3 Gb	6.6 Million	2.2 kb	0.7 kb	237.6 kb

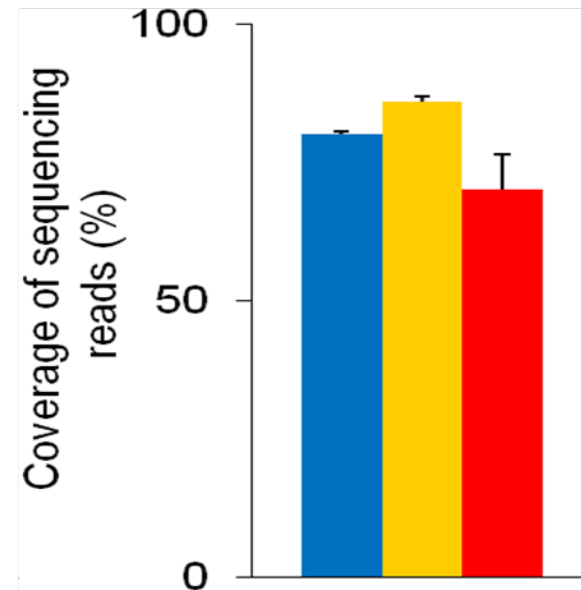
BGI, Wang Jun et al.

Representation of the human gut microbiome in the contig set

Sequences from three studies were mapped on the contig set

- **124 Europeans** (0.58Tb Illumina)
- **18 US** (1.83 Gb 454 Roche)
- **13 Japanese** (0.79 Gb Sanger)

The contig set represents well the whole human metagenome



The gene set

Metagene prediction on the contigs:

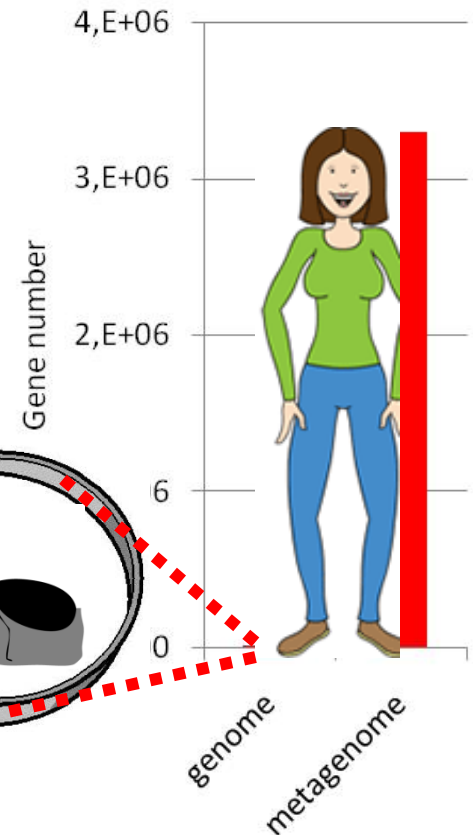
- 14 million **O**pen **R**eadin**G** **F**rames

Removal of redundancy :

≥ 95 % nucleotide identity over at least 90 % of the length of the shorter ORF

- 3.3 million ORFs

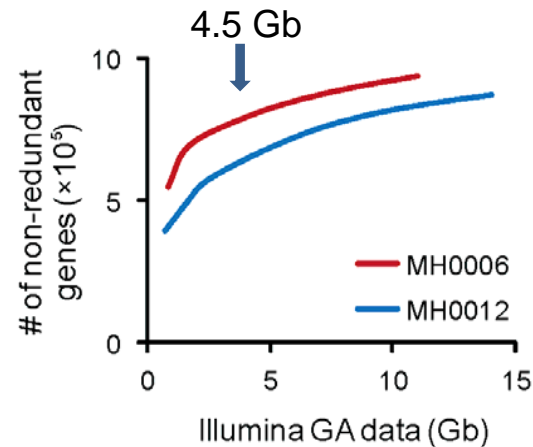
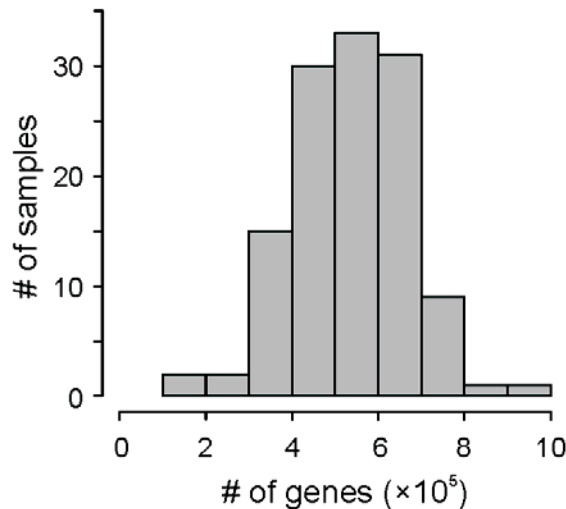
150 times human gene com



Human intestinal microbial genes are largely shared in the cohort

Each individual has ~540 000 prevalent genes, on average

Deeper sequencing reveals more genes



40 % of an individual's genes are shared by individuals of the cohort

The half'n half rule!

Bacterial species are also largely shared

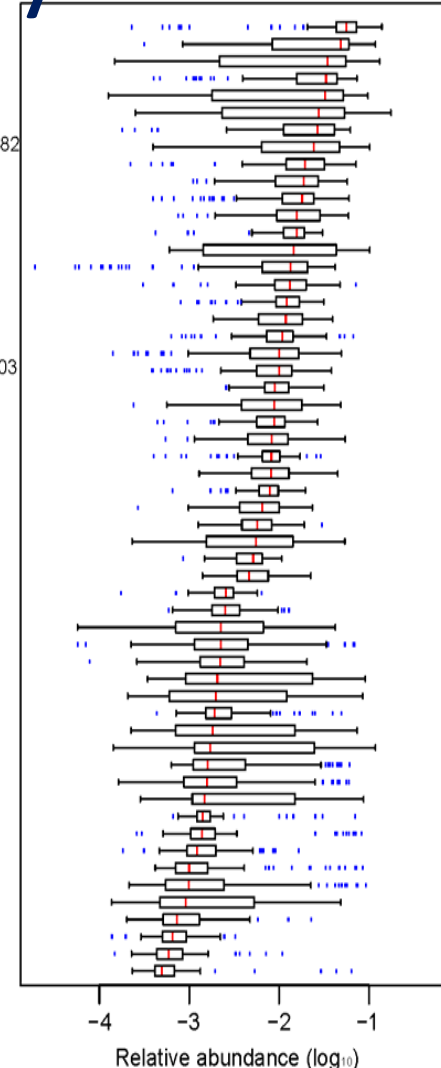
Illumina reads used to **identify** bacterial species **and measure their abundance** in different individuals of the cohort

Individuals	Species
All 124	18
>90%	57
>50%	75

Abundance of a bacterial species varies 12-2200 fold in individuals

We are all similar!

Bacteroides uniformis
 Alistipes putredinis
 Parabacteroides merdae
 Dorea longicatena
 Ruminococcus bromii L2-63
 Bacteroides caccae
 Clostridium sp SS2-1
 Bacteroides thetaiotaomicron VPI-5482
 Eubacterium hallii
 Ruminococcus torques L2-14
 unknown sp SS3 4
 Ruminococcus sp SR1 5
 Faecalibacterium prausnitzii SL3 3
 Ruminococcus lactaris
 Collinsella aerofaciens
 Dorea formicigenerans
 Bacteroides vulgatus ATCC 8482
 Roseburia intestinalis M50 1
 Bacteroides sp. 2_1 7
 Eubacterium siraeum 70 3
 Parabacteroides distasonis ATCC 8503
 Bacteroides sp. 9_1_42FAA
 Bacteroides ovatus
 Bacteroides sp. 4_3_47FAA
 Bacteroides sp. 2_2 4
 Eubacterium rectale M104 1
 Bacteroides xylanisolvens XB1A
 Coprococcus comes SL7 1
 Bacteroides sp. D1
 Bacteroides sp. D4
 Eubacterium ventriosum
 Bacteroides dorei
 Ruminococcus obeum A2-162
 Subdoligranulum variabile
 Bacteroides capillosus
 Streptococcus thermophilus LMD-9
 Clostridium leptum
 Holdemania filiformis
 Bacteroides stercoris
 Coprococcus eutactus
 Clostridium sp M62 1
 Bacteroides eggerthii
 Butyrivibrio crossotus
 Bacteroides finegoldii
 Parabacteroides johnsonii
 Clostridium sp L2-50
 Clostridium nexile
 Bacteroides pectinophilus
 Anaerotruncus colihominis
 Ruminococcus gnavus
 Bacteroides intestinalis
 Bacteroides fragilis 3_1 12
 Clostridium asparagiforme
 Enterococcus faecalis TX0104
 Clostridium scindens
 Blautia hansenii



Bacterial genes/species/communities associated to a disease ?

Take-home messages

- 3.3 million prevalent human intestinal bacterial genes were identified in a cohort of 124 individuals, 150 times more than the human gene complement
- The gene catalog includes most of the genes identified in the studies over three continents
- Combinations of species (i.e. bacterial communities!) are associated to chronic diseases

After the human genome

the human metagenome!!!

Where do these studies lead to and when?

Diagnostic & prognostic tests - soon

- arrays, sequencing, Q-PCR; immunomarkers (?)

Better treatments - next

- personalized medicine

Responders/non-responders

Novel treatments – last: target our other genome!

- modulation of microbiota

Promoters (prebiotics, probiotics)

Inhibitors (“soft” AB-like?? “Re-purposing”?)

- transplantation of microbiota; biobanking

Acknowledgments

MetaHIT Consortium

P. Bork

J. Raes

M. Arumugam

Wang Jun

Qin Junjie

Li Ruiquiang

J. Doré

F. Guarner

J. Parkhill

O. Pedersen

J. Weissenbach



INRA

N. Pons

J.M. Batto

E. Le Chatelier

M. Almeida

P. Renault



Thank you for
your attention

Coda

You can follow us on:

 www.twitter.com/metahit

You can follow human microbiome news at:

 http://www.netvibes.com/metahit#Live_News

Both can be reached from the MetaHIT web page:

<http://www.metahit.eu/>

Yoni Winogradski
MetaHIT manager