Omics

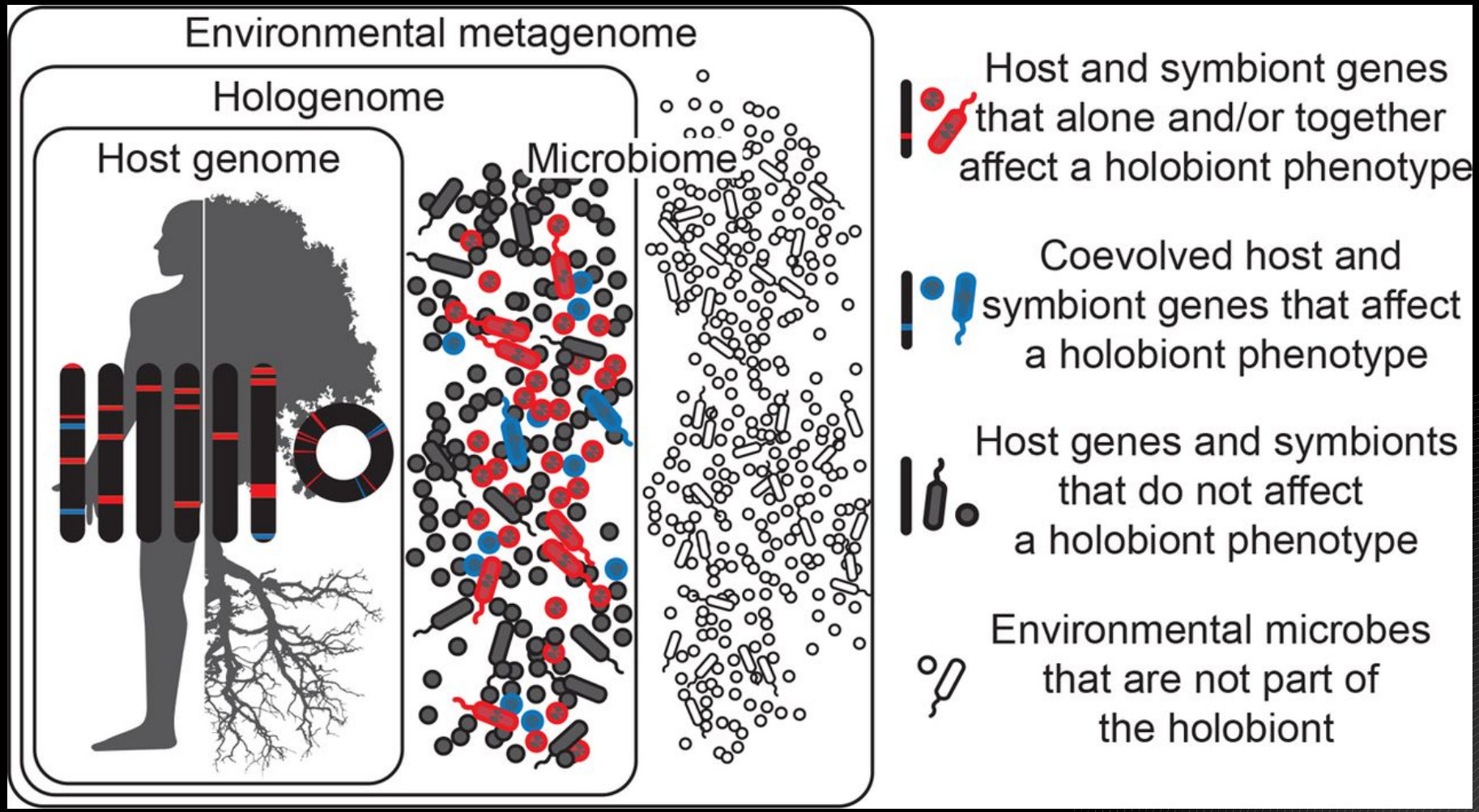
for the

Human holobiont studies

(and other environments)

The Holobiont concept

Holobiont = host + its symbiotic microbes



Kevin R. Theis et al. mSystems 2016

The Human Genome : February 2001

Lander, E., Linton, L., Birren, B., Nusbaum, C., Zody, M., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C. et al. (2001)

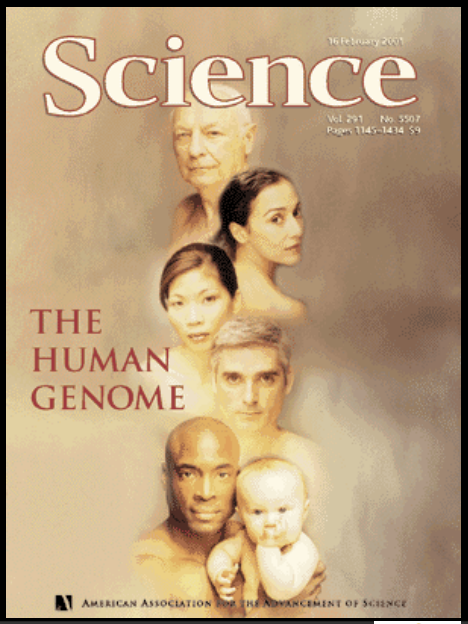
Initial sequencing and analysis of the human genome. Nature, 409:860-921.

[doi: 10.1038/35057062]

Venter, J., Adams, M., Myers, E., Li, P., Mural, R., Sutton, G., Smith, H., Yandell, M., Evans, C., Holt, R., Gocayne, J., Amanatides, P., Ballew, R., Huson, D., Wortman, J., Zhang, Q., Kodira, C., Zheng, X., Chen, L., Skupski, M., Subramanian, G., Thomas, P., Zhang, J., Gabor Miklos, G., Nelson, C., Broder, S., Clark, A., Nadeau, J., McKusick, V., Zinder, N. et al. (2001)

The sequence of the human genome. Science, 291:1304-1351.

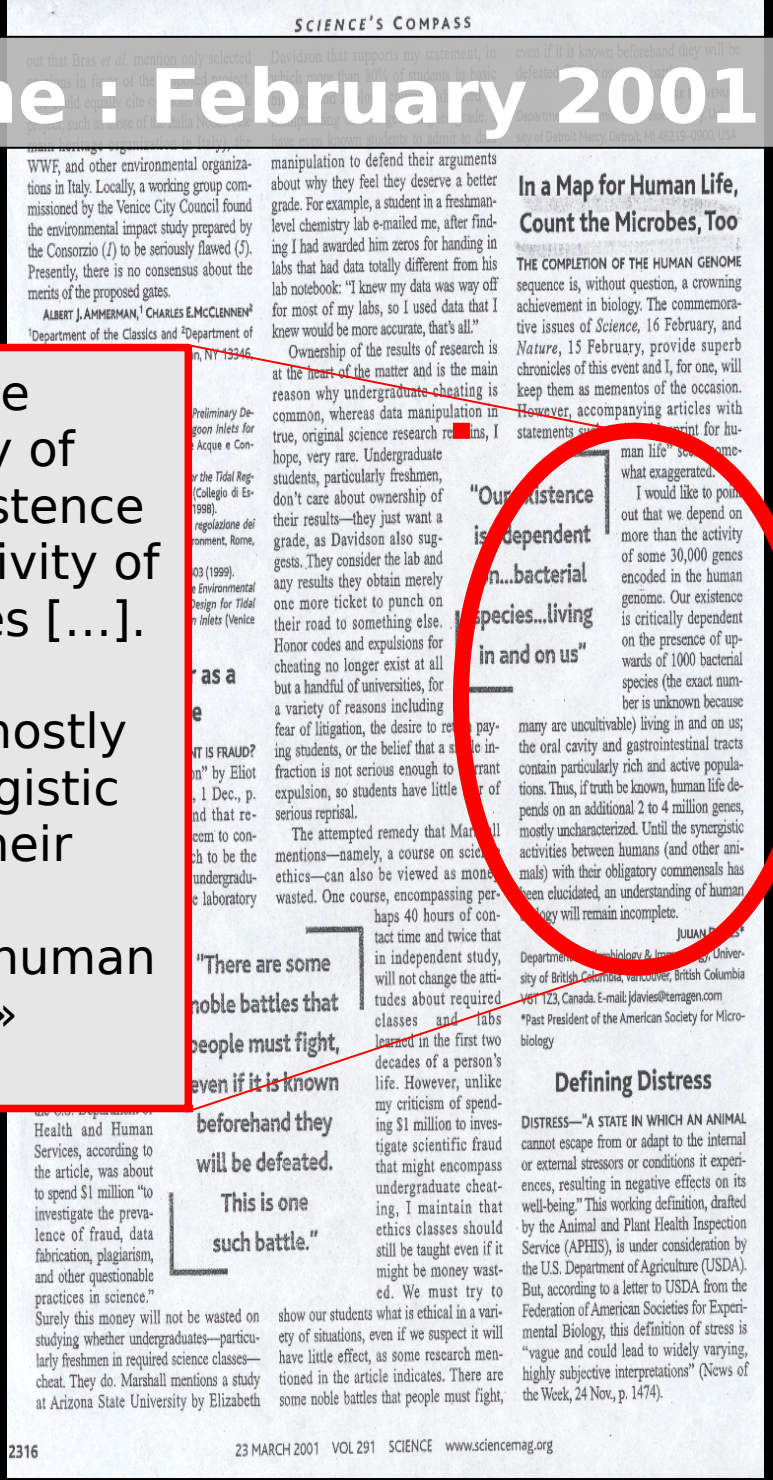
[doi: 10.1126/science.1058040]



The Human Genome : February 2001

« I would like to point out that we depend on more than the activity of some 30,000 genes [...]. Our existence is critically dependent on the activity of upmosts of 1000 bacterial species [...]. Thus [...] human life depends on additional 2 to 4 million genes, mostly uncharacterized. Untill the synergistic activities between human with their obligatory commensals has been elucidated, an understanding of human biology will remain incomplete. »

Julian Davies
Science 23 March 2001:
Vol. 291. no. 5512, p. 2316



SCIENCE'S COMPASS

WWF, and other environmental organizations in Italy. Locally, a working group commissioned by the Venice City Council found the environmental impact study prepared by the Consorzio (1) to be seriously flawed (5). Presently, there is no consensus about the merits of the proposed gates.

ALBERT J. AMMERMAN,[1] CHARLES E. McCLENNEN[2]

[1]Department of the Classics and [2]Department of

Health and Human Services, according to the article, was about to spend $1 million "to investigate the prevalence of fraud, data fabrication, plagiarism, and other questionable practices in science." Surely this money will not be wasted on studying whether undergraduates—particularly freshmen in required science classes—cheat. They do. Marshall mentions a study at Arizona State University by Elizabeth

manipulation to defend their arguments about why they feel they deserve a better grade. For example, a student in a freshman-level chemistry lab e-mailed me, after finding I had awarded him zeros for handing in labs that had data totally different from his lab notebook: "I knew my data was way off for most of my labs, so I used data that I knew would be more accurate, that's all."

Ownership of the results of research is at the heart of the matter and is the main reason why undergraduate cheating is common, whereas data manipulation in true, original science research remains, I hope, very rare. Undergraduate students, particularly freshmen, don't care about ownership of their results—they just want a grade, as Davidson also suggests. They consider the lab and any results they obtain merely one more ticket to punch on their road to something else. Honor codes and expulsions for cheating no longer exist at all but a handful of universities, for a variety of reasons including fear of litigation, the desire to retain paying students, or the belief that a single infraction is not serious enough to warrant expulsion, so students have little fear of serious reprisal.

The attempted remedy that Marshall mentions—namely, a course on science ethics—can also be viewed as money wasted. One course, encompassing perhaps 40 hours of contact time and twice that in independent study, will not change the attitudes about required classes and labs learned in the first two decades of a person's life. However, unlike my criticism of spending $1 million to investigate scientific fraud that might encompass undergraduate cheating, I maintain that ethics classes should still be taught even if it might be money wasted. We must try to show our students what is ethical in a variety of situations, even if we suspect it will have little effect, as some research mentioned in the article indicates. There are some noble battles that people must fight,

"There are some noble battles that people must fight, even if it is known beforehand they will be defeated. This is one such battle."

In a Map for Human Life, Count the Microbes, Too

THE COMPLETION OF THE HUMAN GENOME sequence is, without question, a crowning achievement in biology. The commemorative issues of *Science*, 16 February, and *Nature*, 15 February, provide superb chronicles of this event and I, for one, will keep them as mementos of the occasion. However, accompanying articles with statements such as "blueprint for human life" seem somewhat exaggerated.

I would like to point out that we depend on more than the activity of some 30,000 genes encoded in the human genome. Our existence is critically dependent on the presence of upwards of 1000 bacterial species (the exact number is unknown because many are uncultivable) living in and on us; the oral cavity and gastrointestinal tracts contain particularly rich and active populations. Thus, if truth be known, human life depends on an additional 2 to 4 million genes, mostly uncharacterized. Until the synergistic activities between humans (and other animals) with their obligatory commensals has been elucidated, an understanding of human biology will remain incomplete.

"Our existence is critically dependent on the activity of upmosts of 1000 bacterial species...living in and on us"

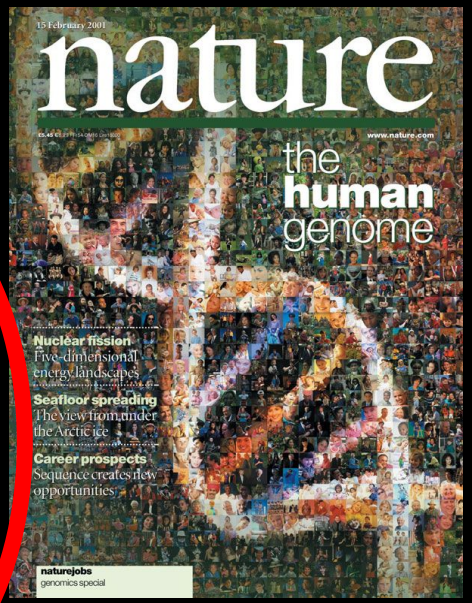
JULIAN DAVIES*

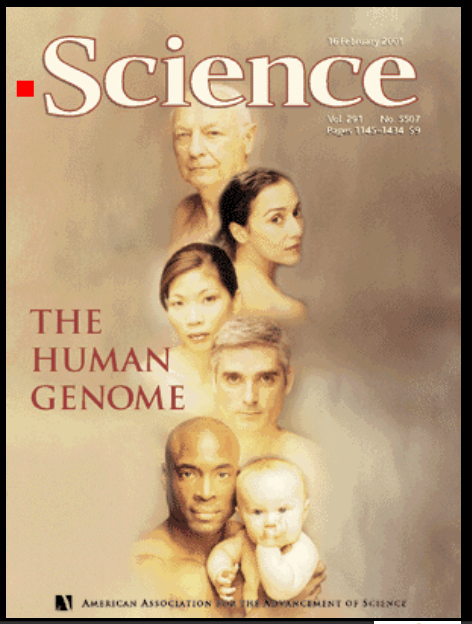
Department of Microbiology & Immunology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. E-mail: jdavies@terragen.com

*Past President of the American Society for Microbiology

Defining Distress

DISTRESS—"A STATE IN WHICH AN ANIMAL cannot escape from or adapt to the internal or external stressors or conditions it experiences, resulting in negative effects on its well-being." This working definition, drafted by the Animal and Plant Health Inspection Service (APHIS), is under consideration by the U.S. Department of Agriculture (USDA). But, according to a letter to USDA from the Federation of American Societies for Experimental Biology, this definition of stress is "vague and could lead to widely varying, highly subjective interpretations" (News of the Week, 24 Nov., p. 1474).

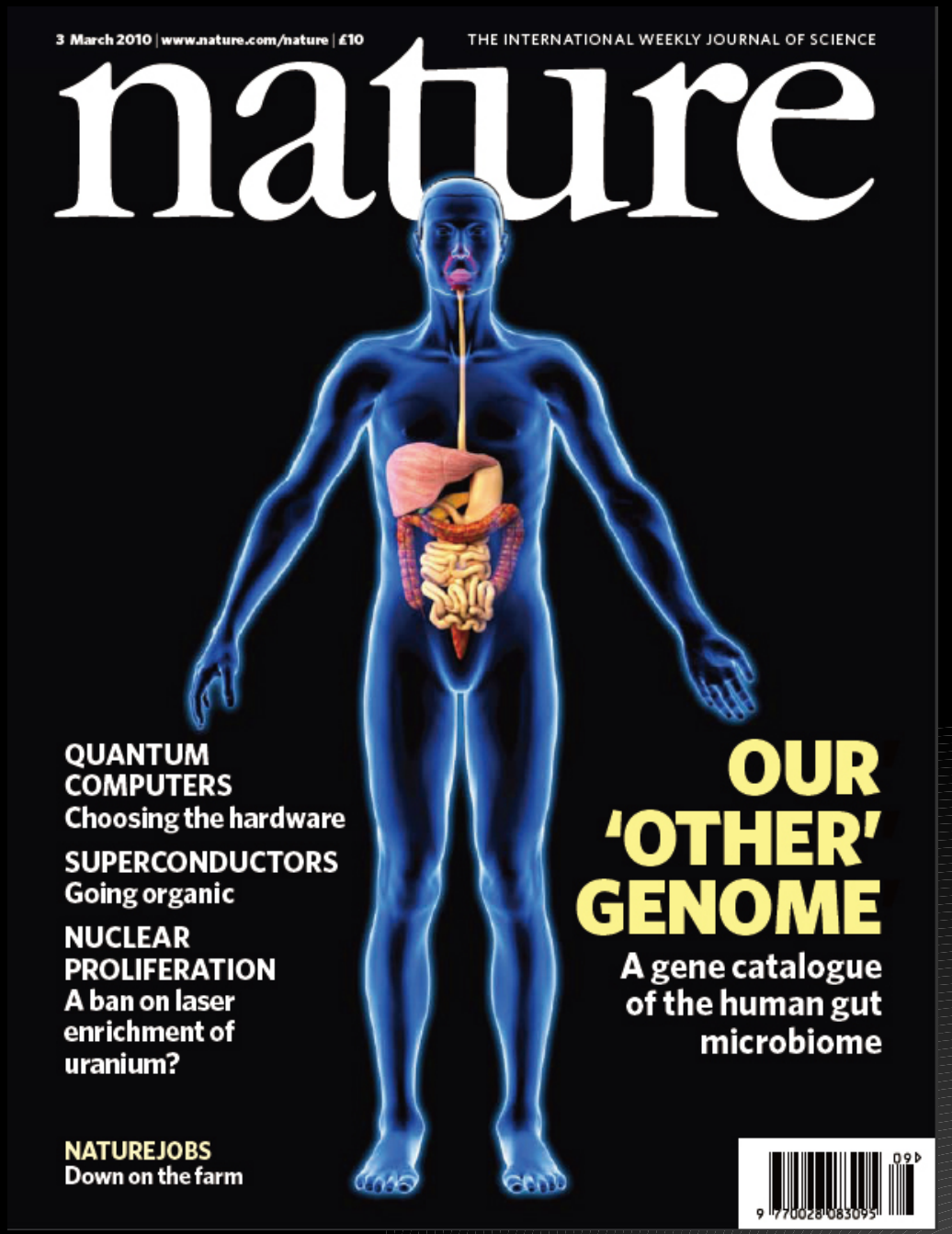




2010 : Our “second genome”

V1 :
124 subjects
3.3 millions genes

Last version (2014)
1267 subjects
9.9 millions genes



The Human companions :

BACTERIA

~ 8,000,000 genes

~ 99% beneficial
~ 1% pathogenic

FUNGI

~ 500,000 genes
?

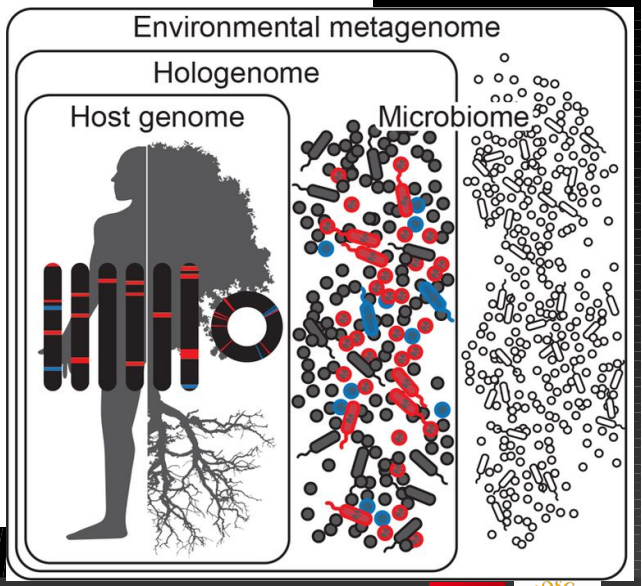
HUMAN

~ 22,000 genes

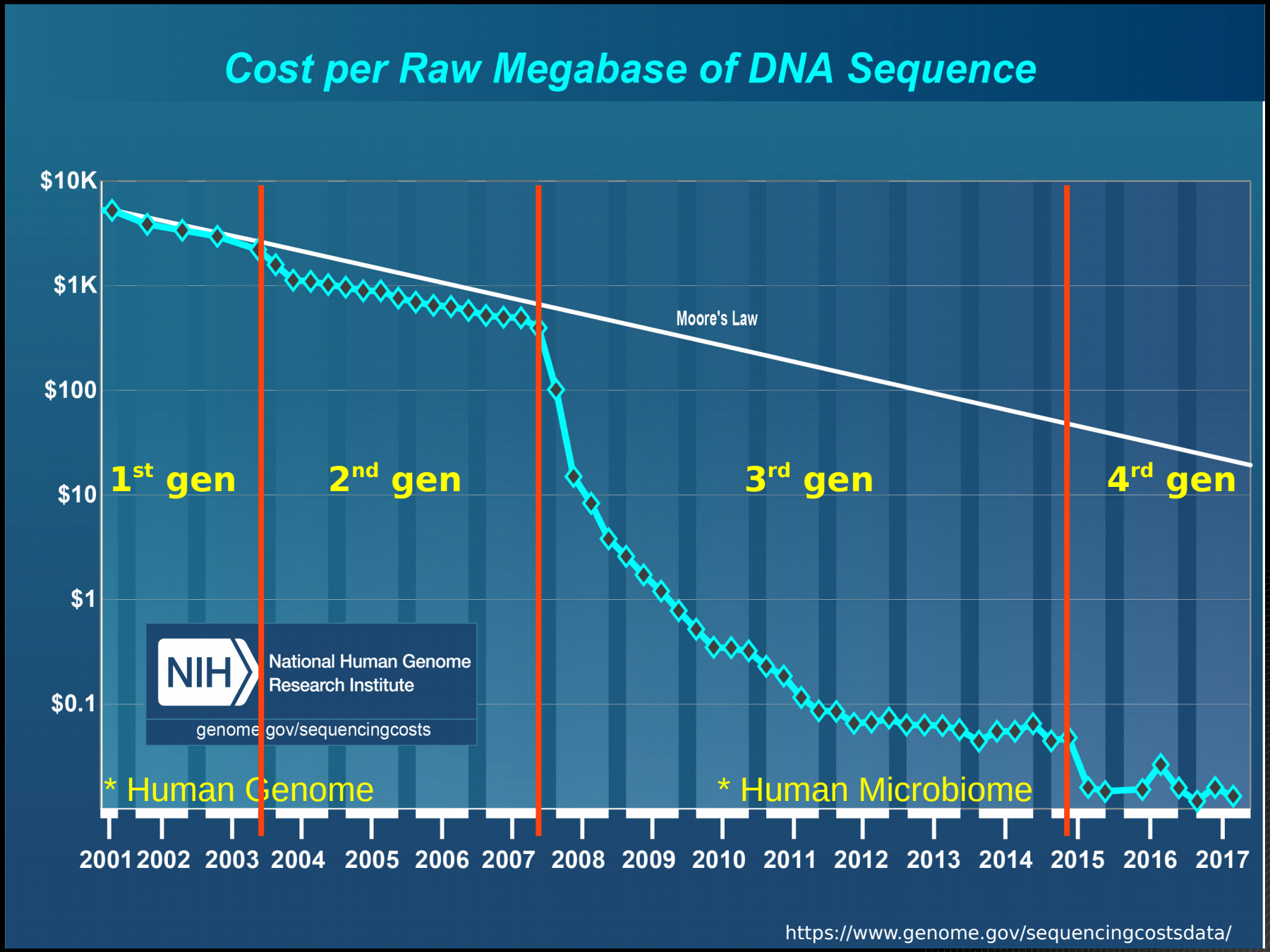
? VIRUSES

~ 80,000 genes
?

ARCHAEA



Sequencing technologies : a game changer

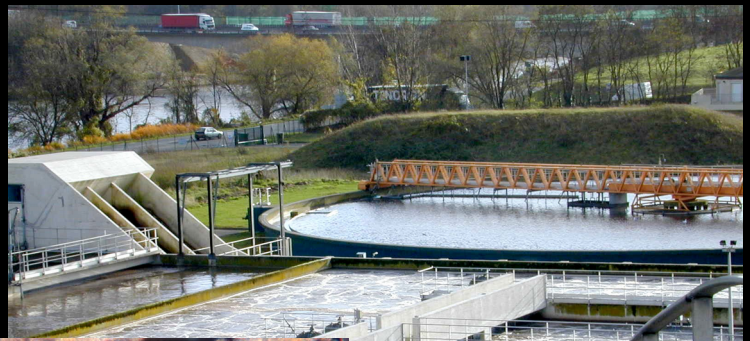


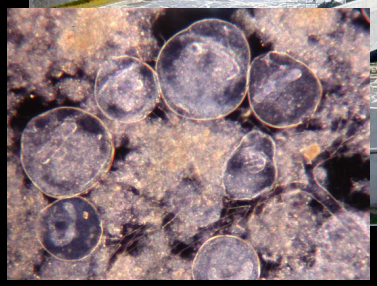
Almost all environments have been studied

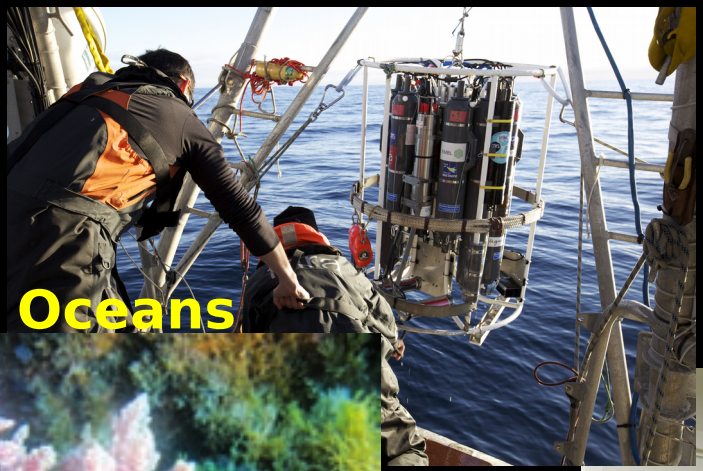
Soils

Air

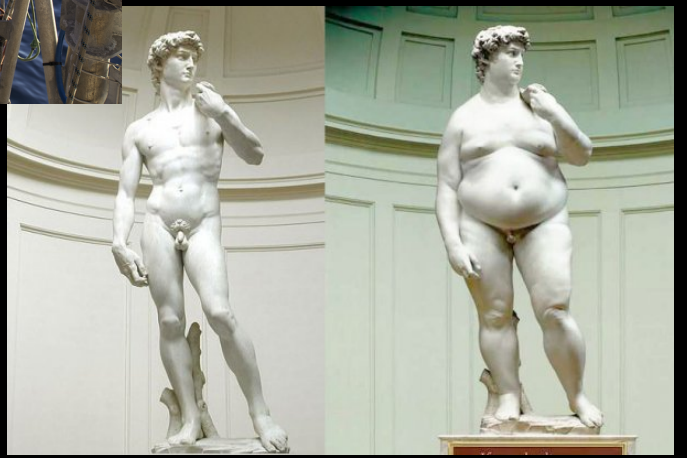
Wastewater plants







Oceans



Human microbiome

The environmental genomics Trinity

Who is there, with who ?

what are the organisms present in my sample
in which quantities

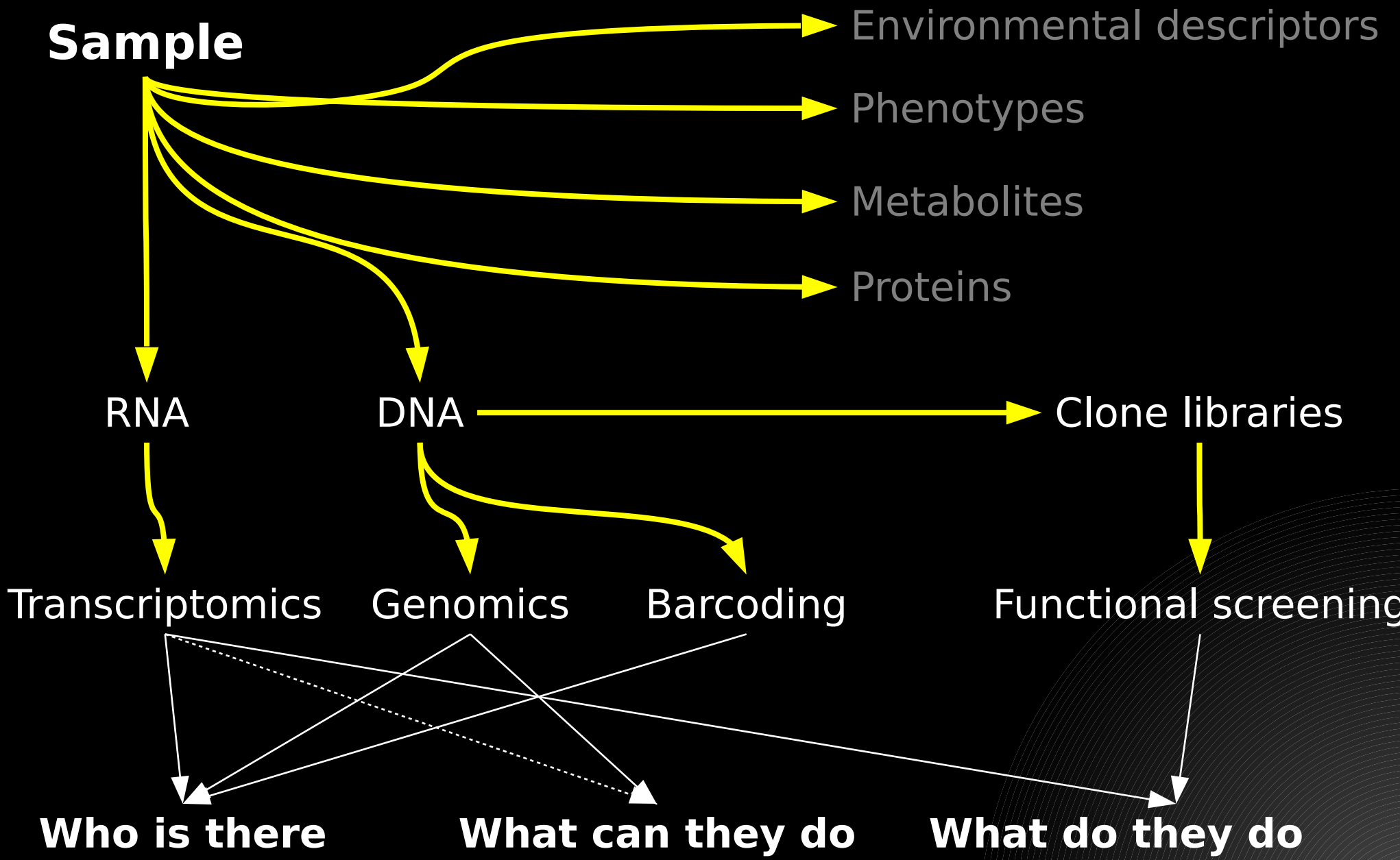
What can they do ?

what are their genes
what is their functional repertoire

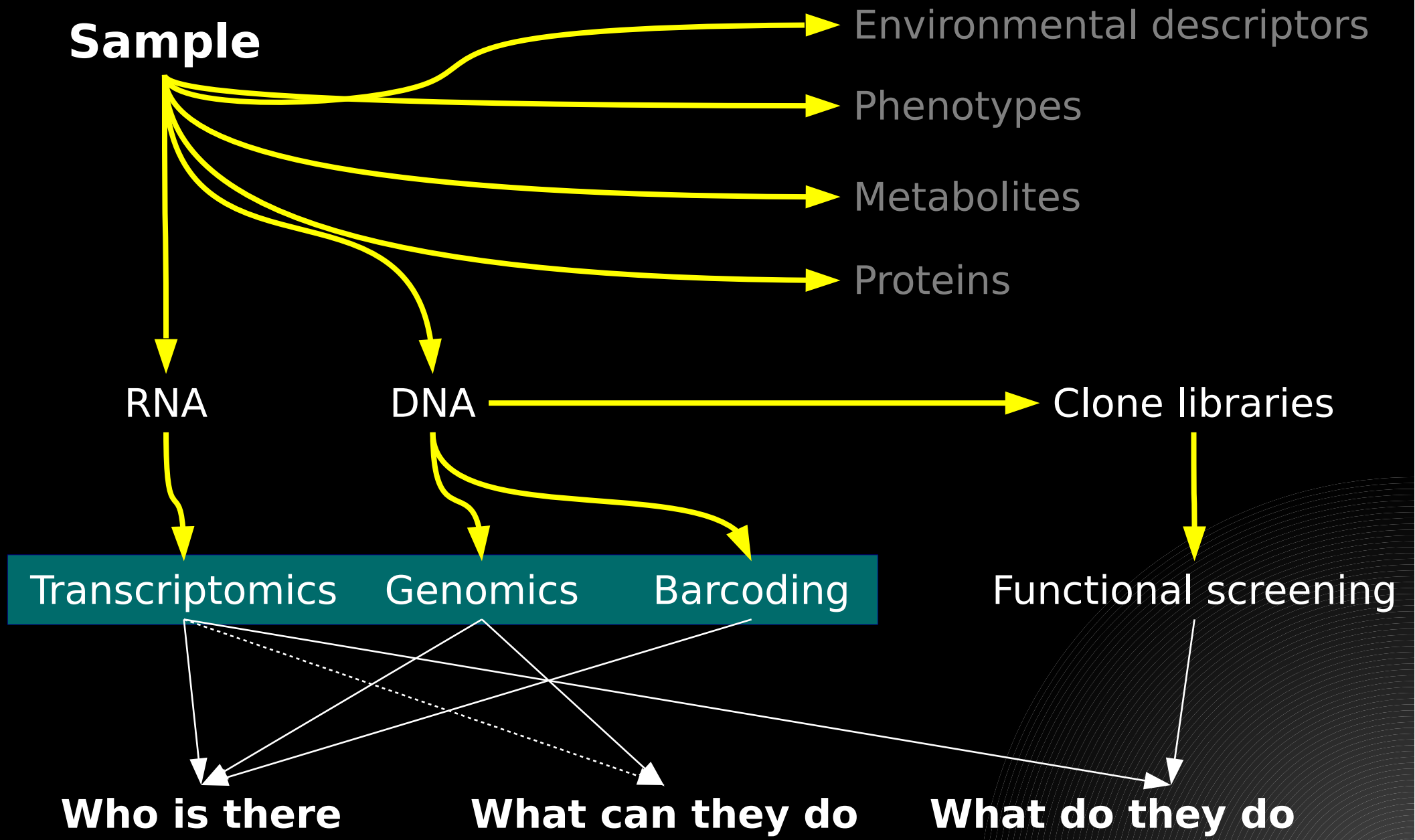
What do they do ?

what are the expressed functions
how do they interact (with the others and with the environment)

The multi meta'omics approach



The multi meta'omics approach



Usual strategies 1 - Barcoding

DNA

PCR amplification (marker gene)

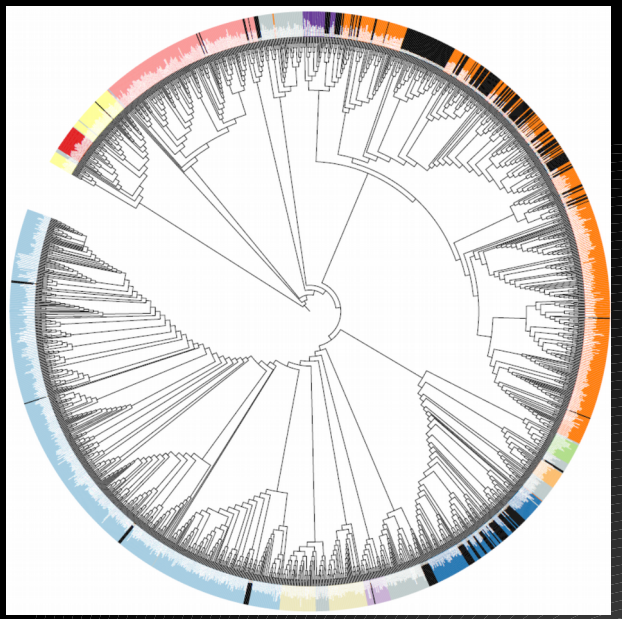
Reads

Clusters / swarms ★

Classification ★

Reference

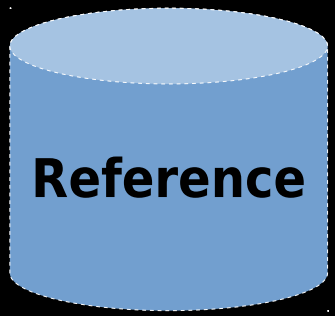
Taxonomical analysis



Usual strategies 2

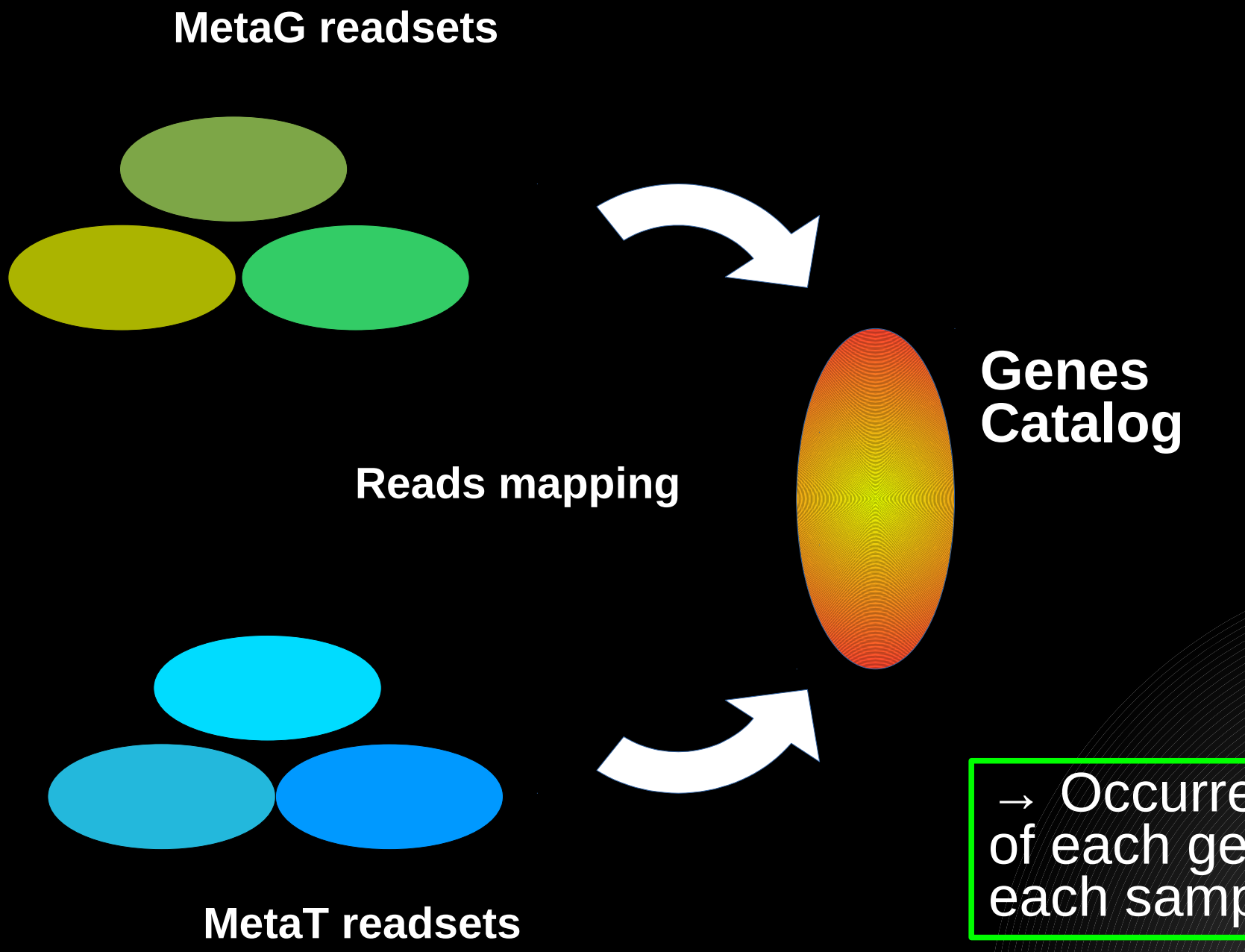
Genomics / Transcriptomics

DNA / RNA



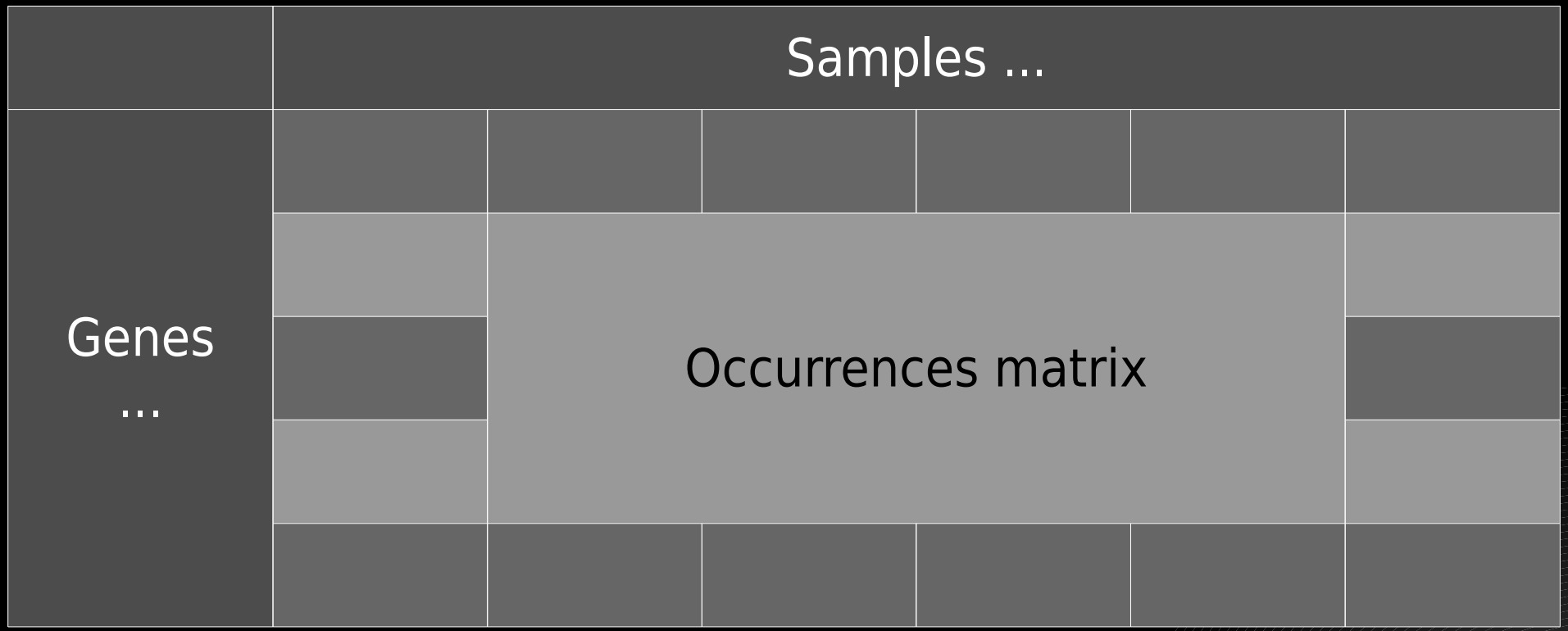
Taxonomic assignation + Functional classification

Getting quantitative



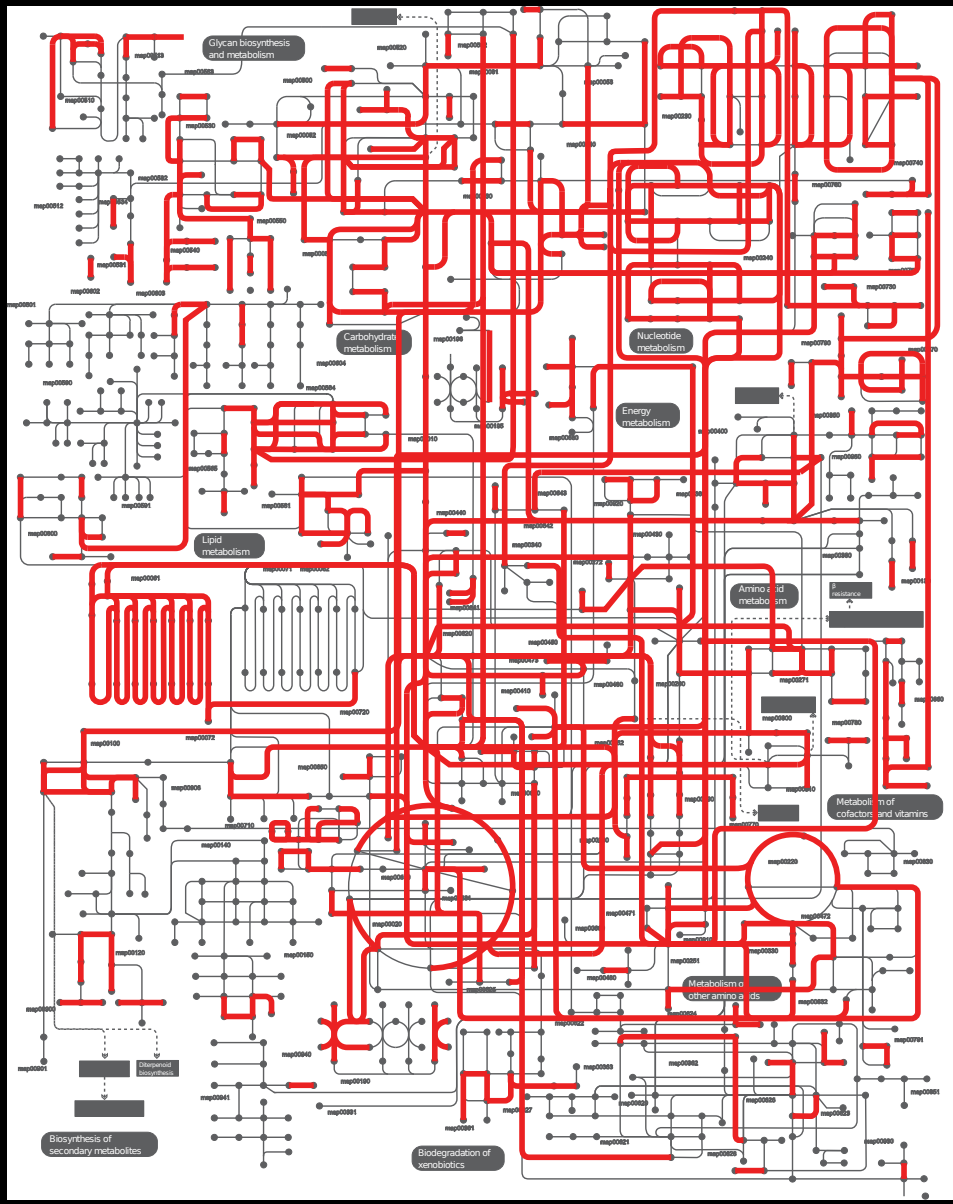
Uses of the occurrences data

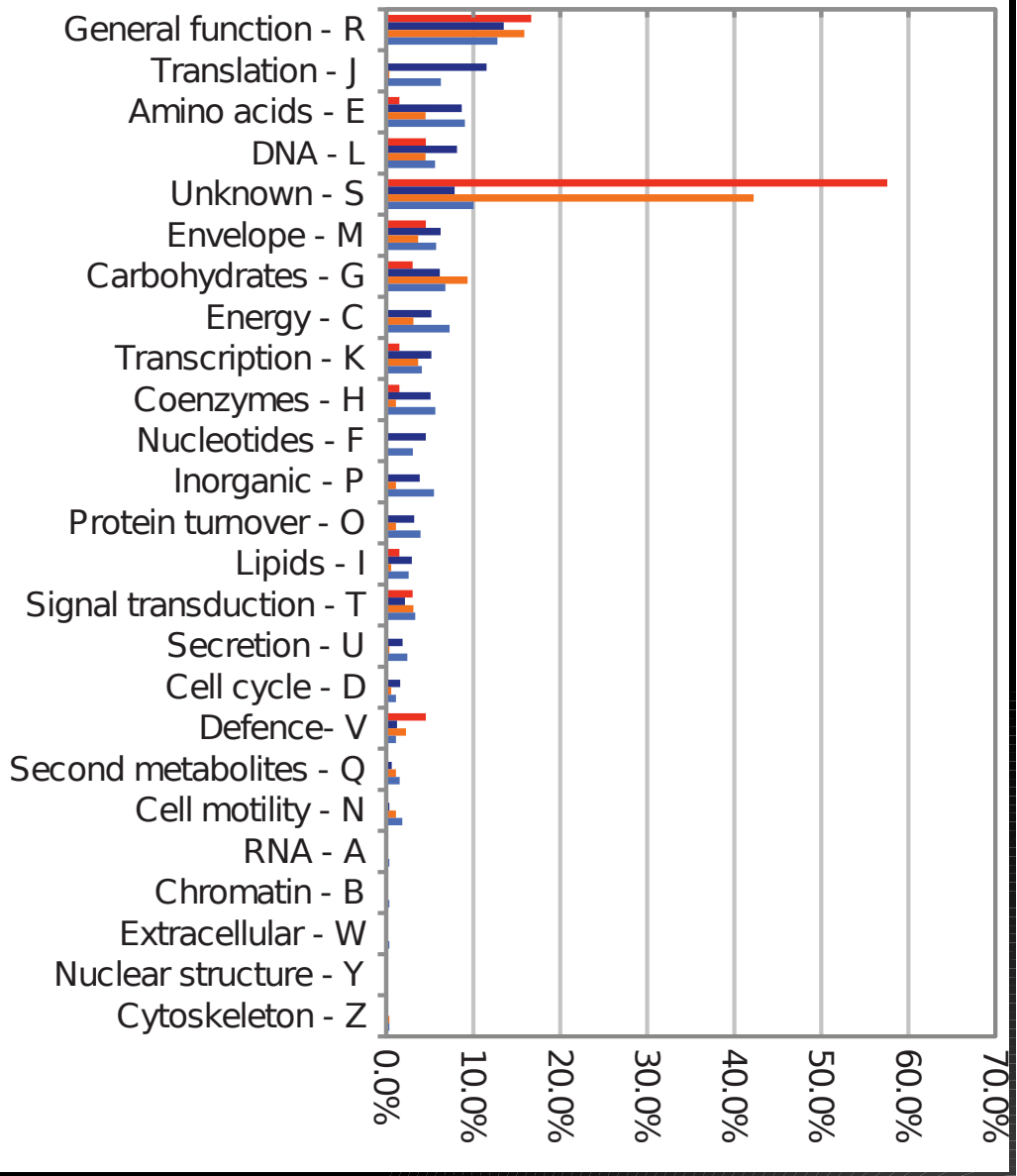
- Environmental parameters



- Functions
- Taxonomy

Functional analysis – metabolism reconstruction



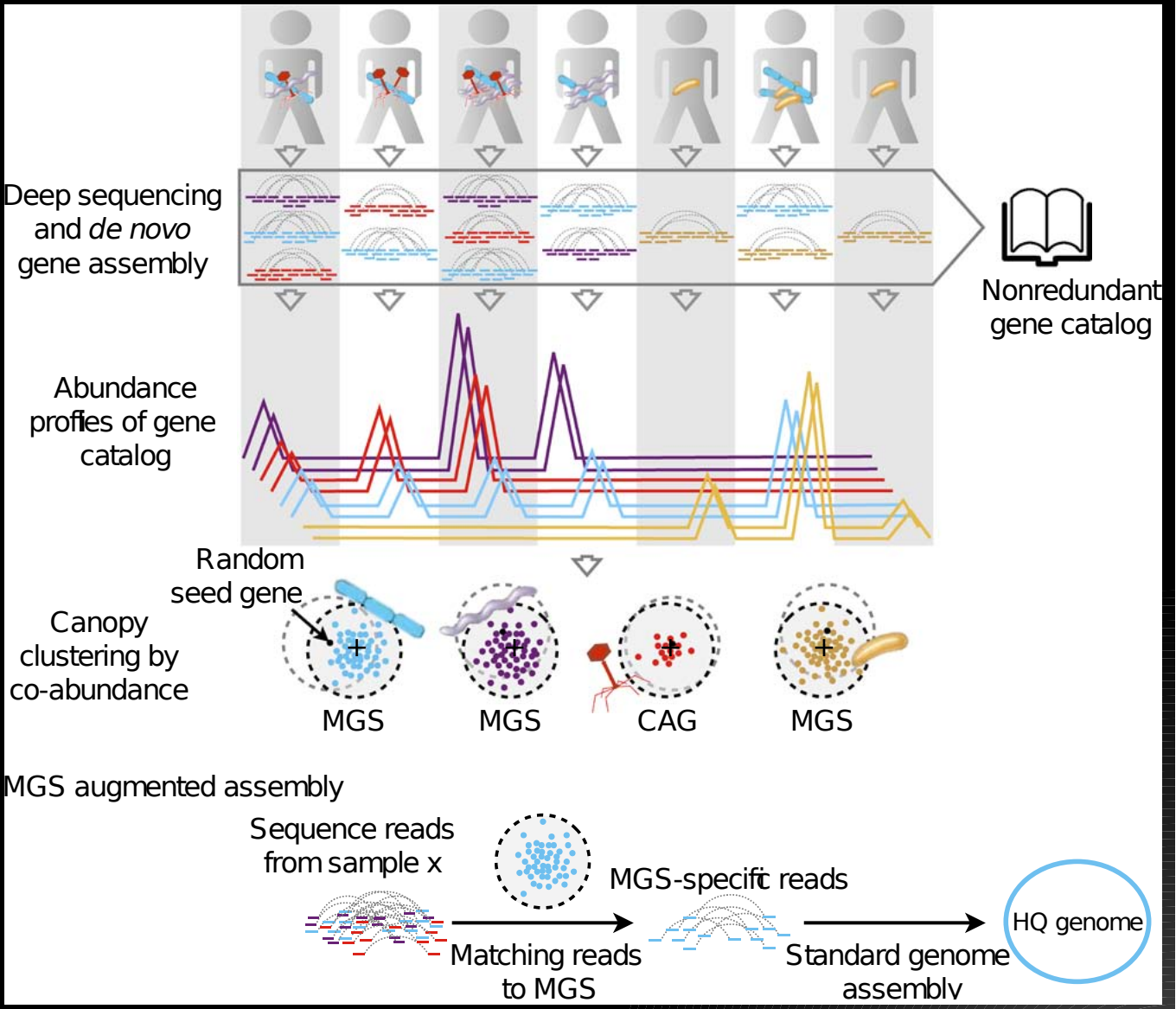


Qin et al. Nature 2010

From genes toward genomes : MetaGenomic Species

Gene clustering based on co-abundance variation profiles

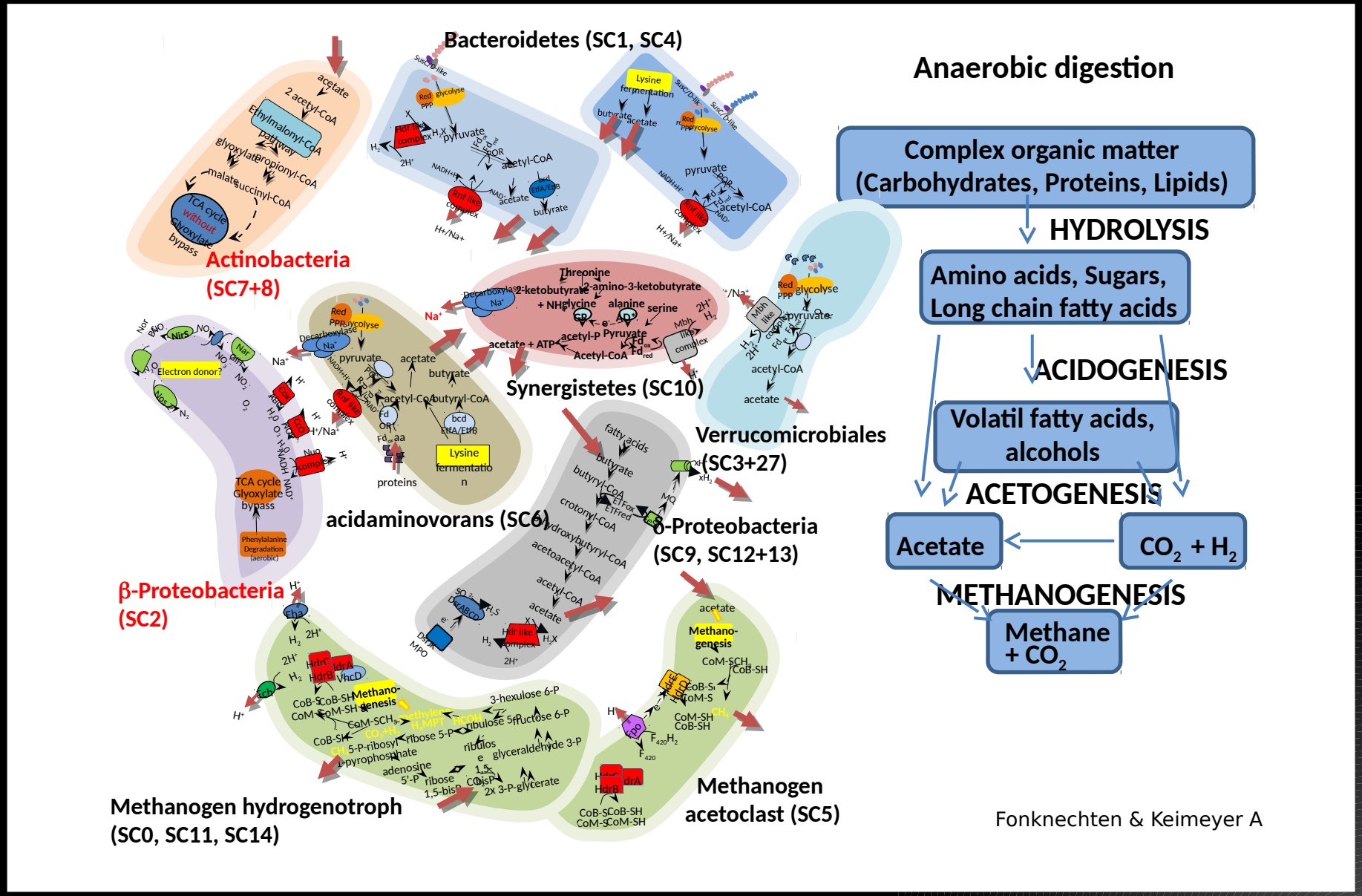
Allows reconstruction of high-quality complete microbial genomes from unknown organisms



Nielsen *et al.*, Nature Biotech (2014)

Global metabolic reconstruction and modelisation

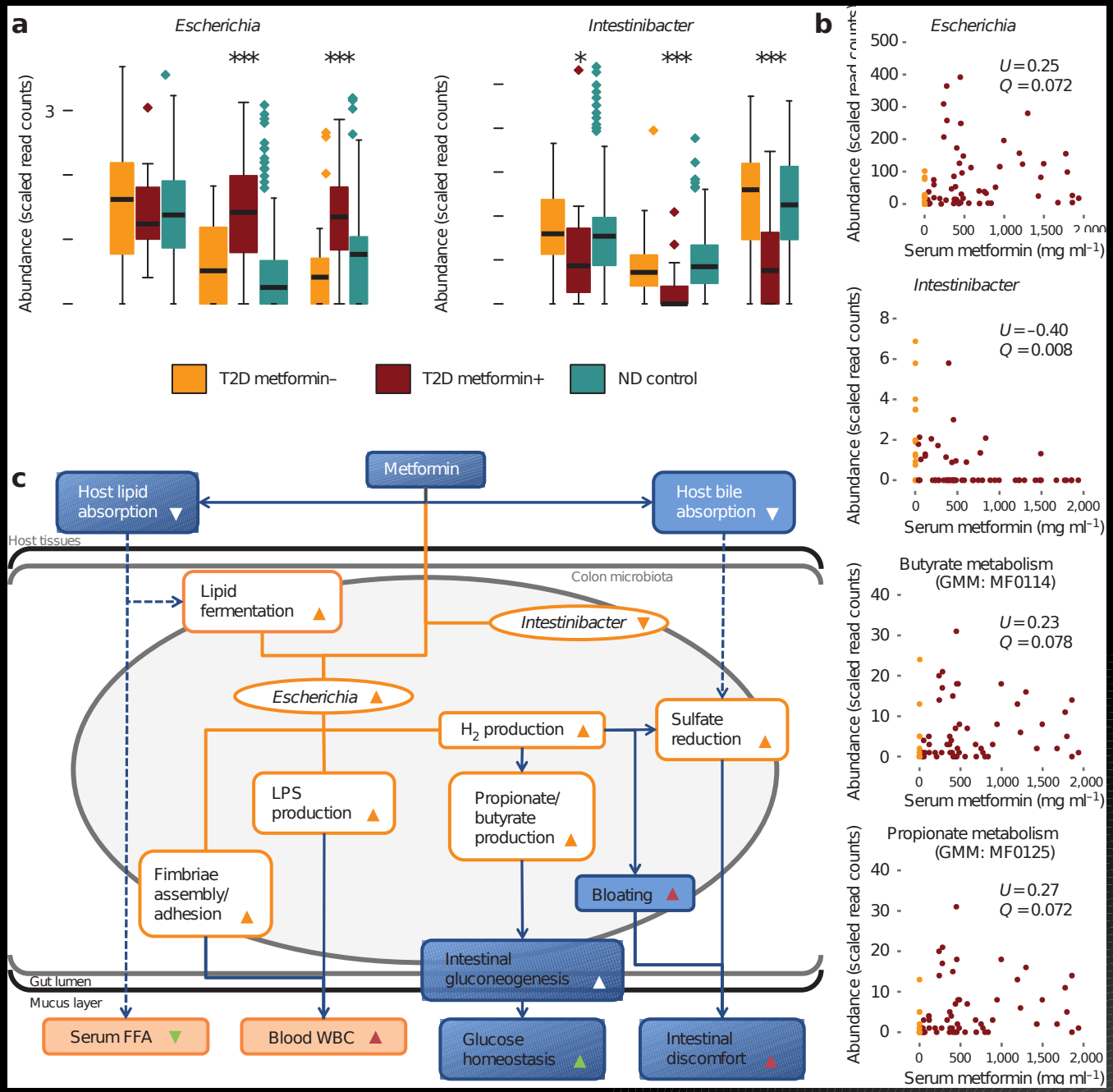
Example from a wastewater anaerobic digester



Integration with phenotype data

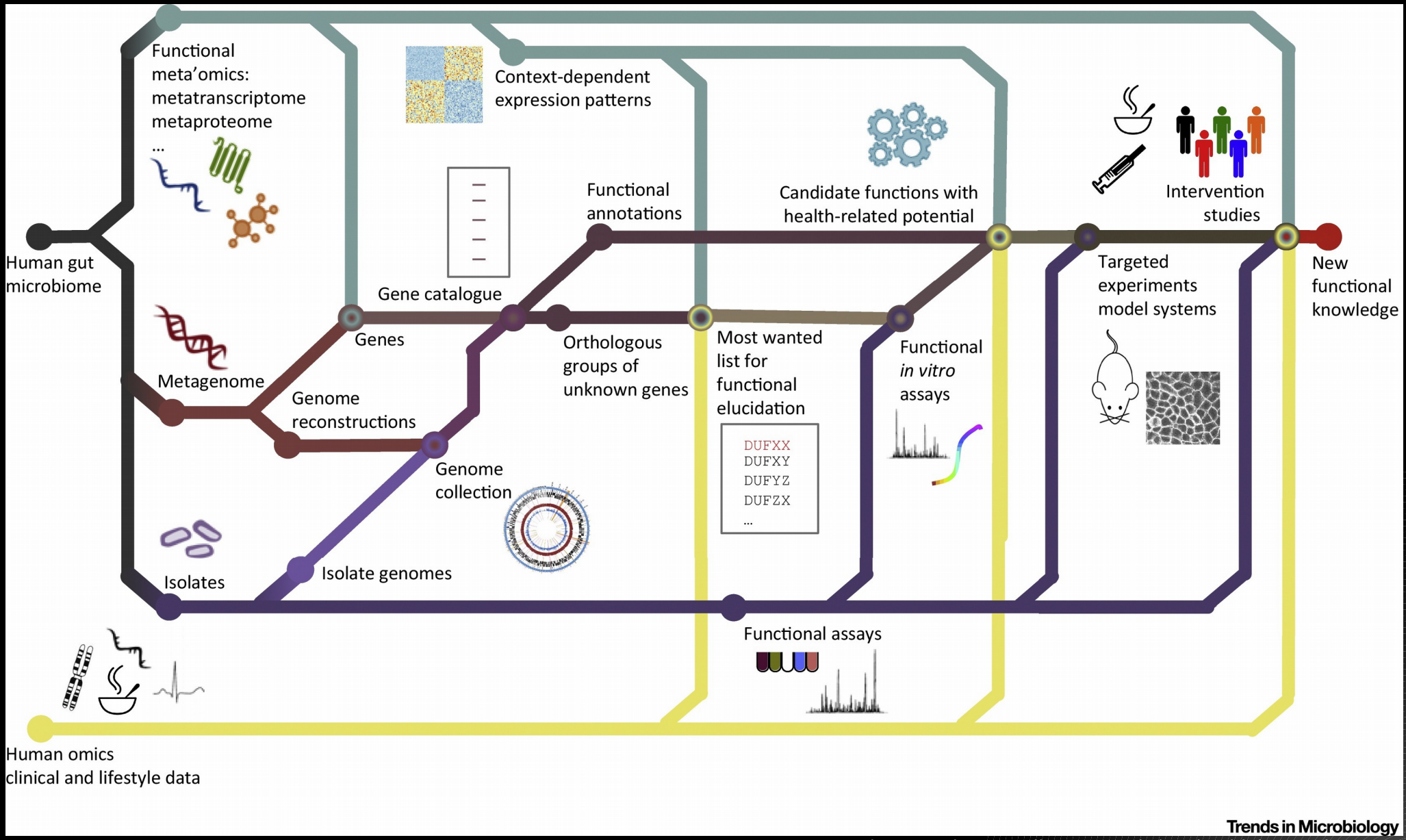
Type 2 diabetes and metformin treatment impact on human gut microbiota

Cross analysis of taxonomy distribution, functions, and phenotypic parameters



Forslund et al., Nature, 2015

Discovery by integration of heterogen



Trends in Microbiology

Heintz-Buschart & Wilmes, Trends in Microbiology, July 2018

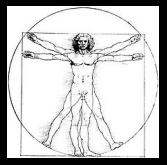
A (big) challenge for the HPC/HTC and

Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

Stephens *et al.*, PloS Biology, 2015

- Huge amount of raw data,
- Expanding during analysis
- Long life span
- Some analyses require huge in-memory space



Human Genome Project

Micro-Obes

Micro-Age

OCEANOMICS