ИНСТИТУТ БИОИНФОРМАТИКИ

Биоинформатические подходы анализа микробиоты человека

Alla L Lapidus, Ph.D.

International Human Genome Sequencing Consortium

Finishing of the human reference sequence beyond the publication of the draft human genome (Nature 409 861-921 (2001)):

The Wellcome Trust Sanger Institute (United Kingdom)

Washington University Genome Sequencing Center

Whitehead Institute for Biomedical Research, Center for Genome Research, Broad Institute, Cambridge

Joint Genome Institute (USA)

Baylor College Of Medicine Human Genome Sequencing Center

University of Washington Genome Center

etc

Scientific manag **NIH, USA**

Francis S. Colli tterstrand, Sandra Kamholz

Director, National Human Genome Research Institute, National Institutes of Health (NIH)

Director, NIH

Human genome project

Why: Human Health improvement

- identification of mutations linked to different forms of cancer
- genotyping of specific viruses to direct appropriate treatment
- design of medication and more accurate prediction of their effects

Impact on:

- advancement in forensic applied sciences, biofuels and other energy applications; agriculture, animal husbandry, bioprocessing; risk assessment; bioarcheology, anthropology and evolution.
- commercial development of genomics research related to DNA based products, a multibillion-dollar industry.

N50 size

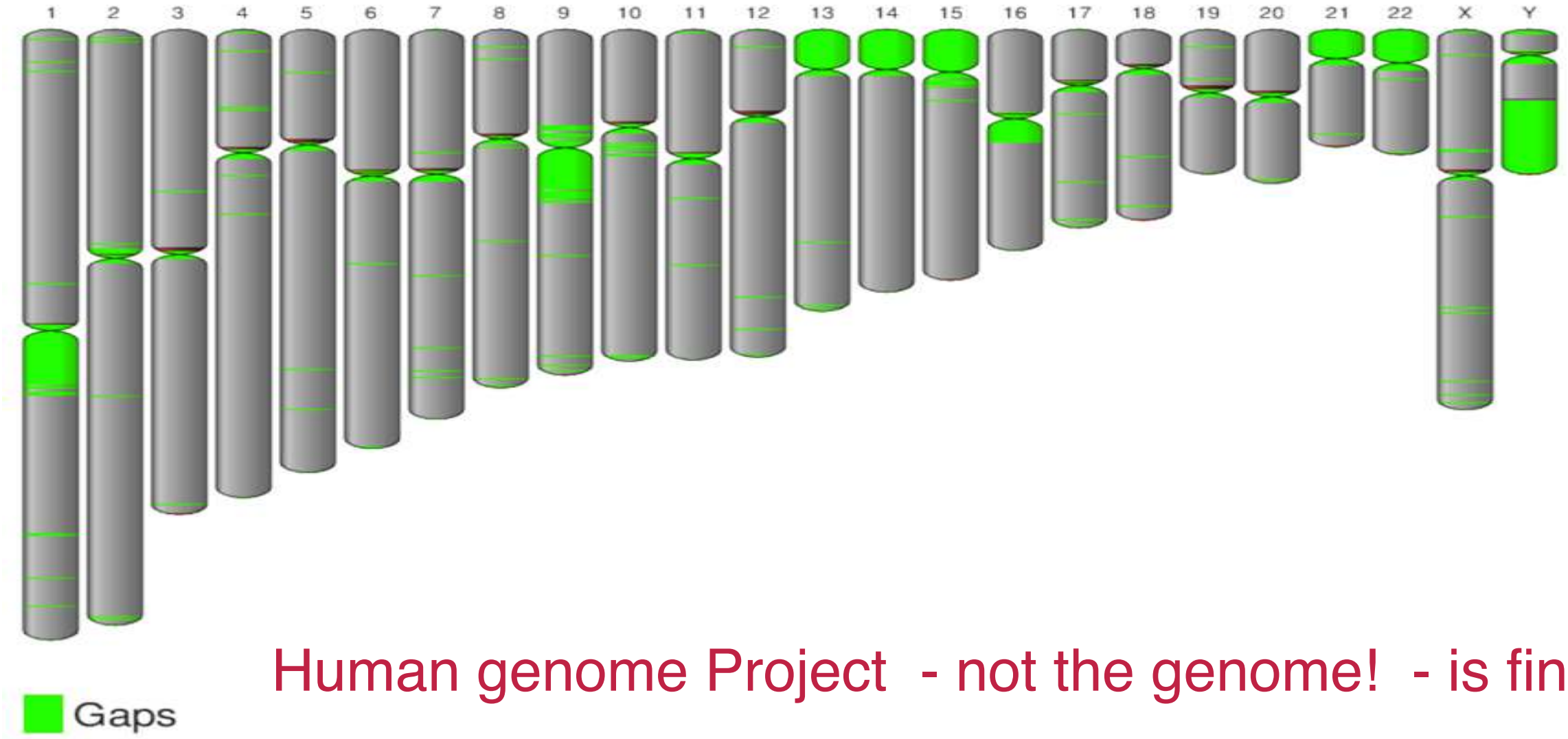
Def: 50% of the genome is in contigs as large as the N50 value

Example: 1 Mbp genome

N50 size = 30 kbp

(300k+100k+45k+45k+30k = 520k >= 500kbp)

Human genome

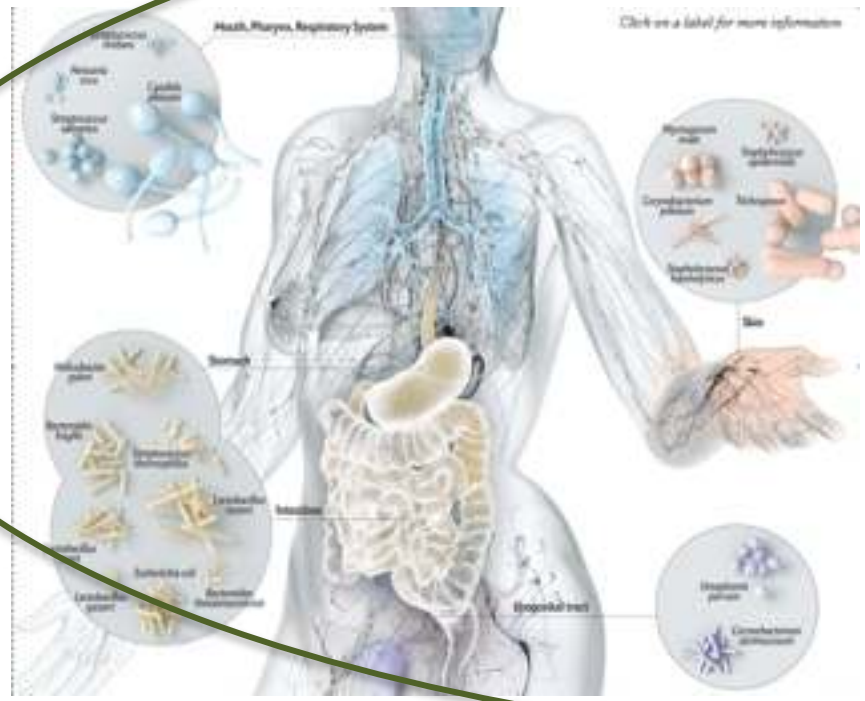


Human genome Project - not the genome! - is finishe

Microbes are everywhere…
Even humans are mostly microbes

Microbial cells outnumber human cells

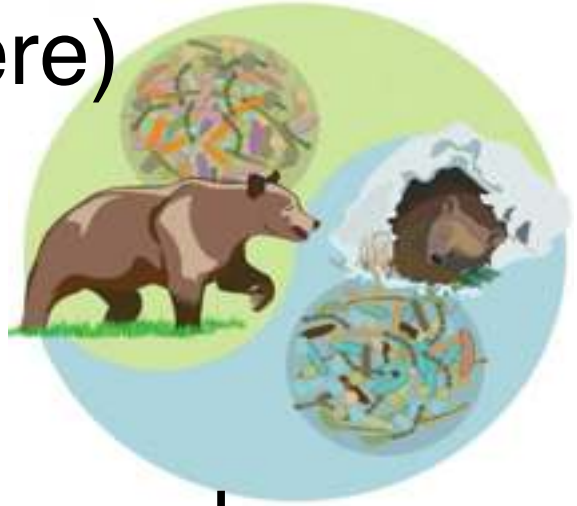
10 : 1



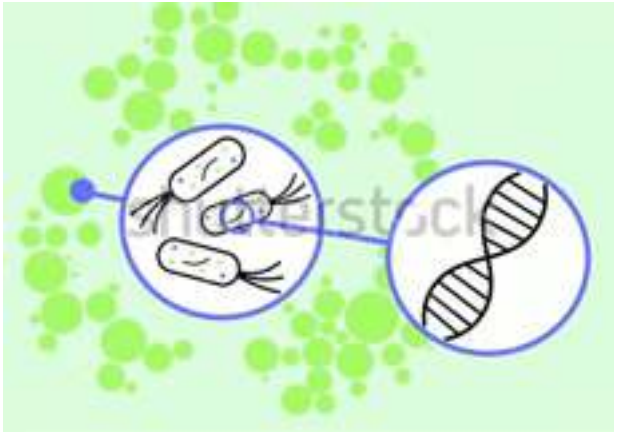
= super organism, made up of both human and bacterial DNA

Microbiota VS Microbiome

Microbiota: microorganisms in the environment (people, animals, plants, soil, oceans, atmosphere)

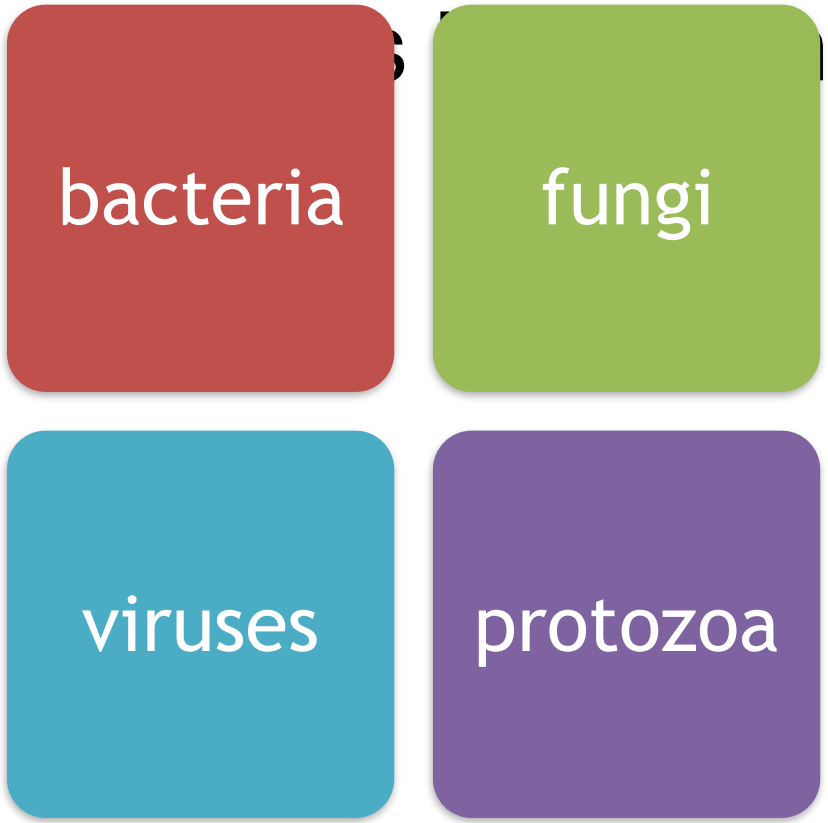


Microbiome: genetic information (genes and genomes) of a microbiota.



Human microbiome

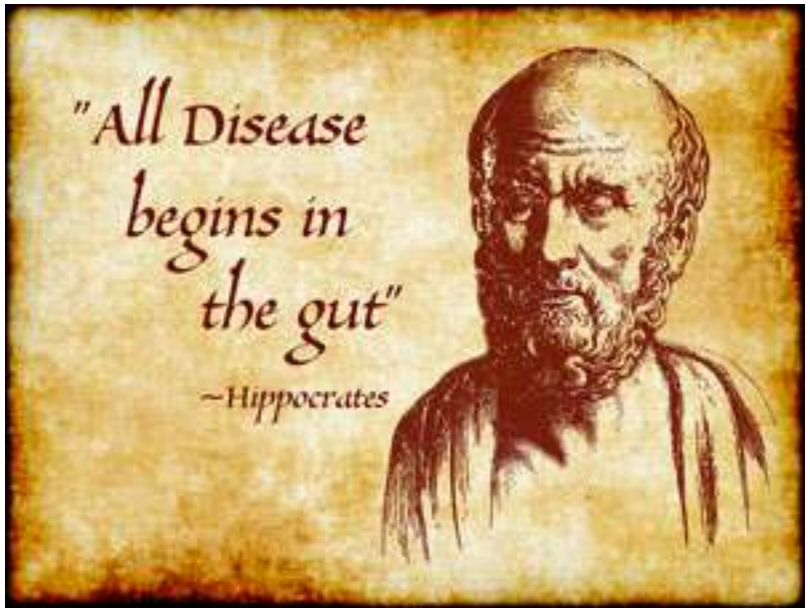
The number of genes in human microbiome



is 200 times the number of genes in the human genomes

200 : 1

Over 2000 years ago Hippocrates stated: “All disease begins in the gut”



Obviously, not all disease begins in the gut. For example, this does not apply to genetic diseases

BUT!

The gut hosts 80% of our immune system and helps to program it

Number of microorganism species found only in gut of healthy people exceeds 500

Human microbiota

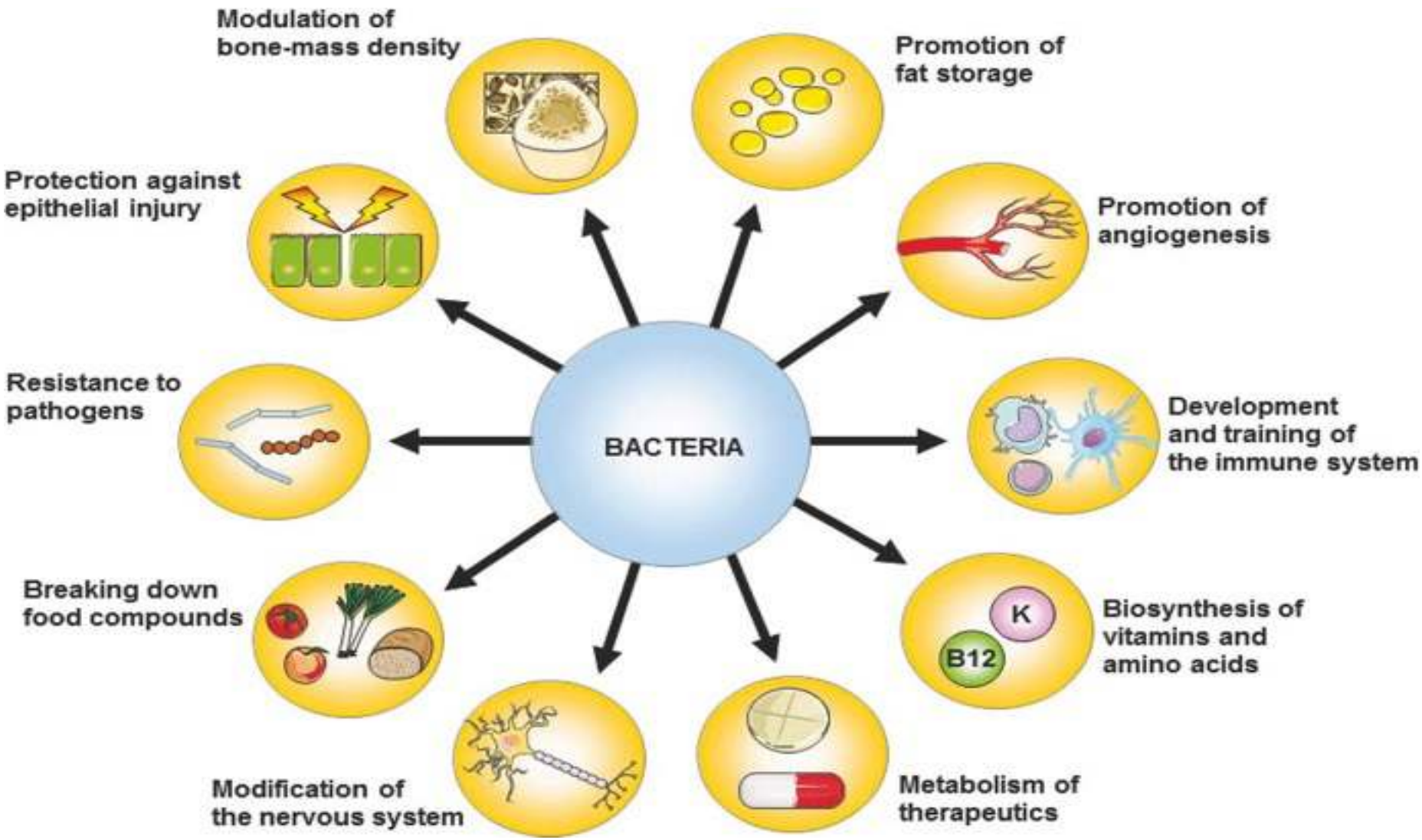
Studies started by *Antonie van Leewenhoek* as early as the 1680s

- compared his oral and fecal microbiota
- oral and fecal samples from individuals in states of health and disease

Good Bacteria

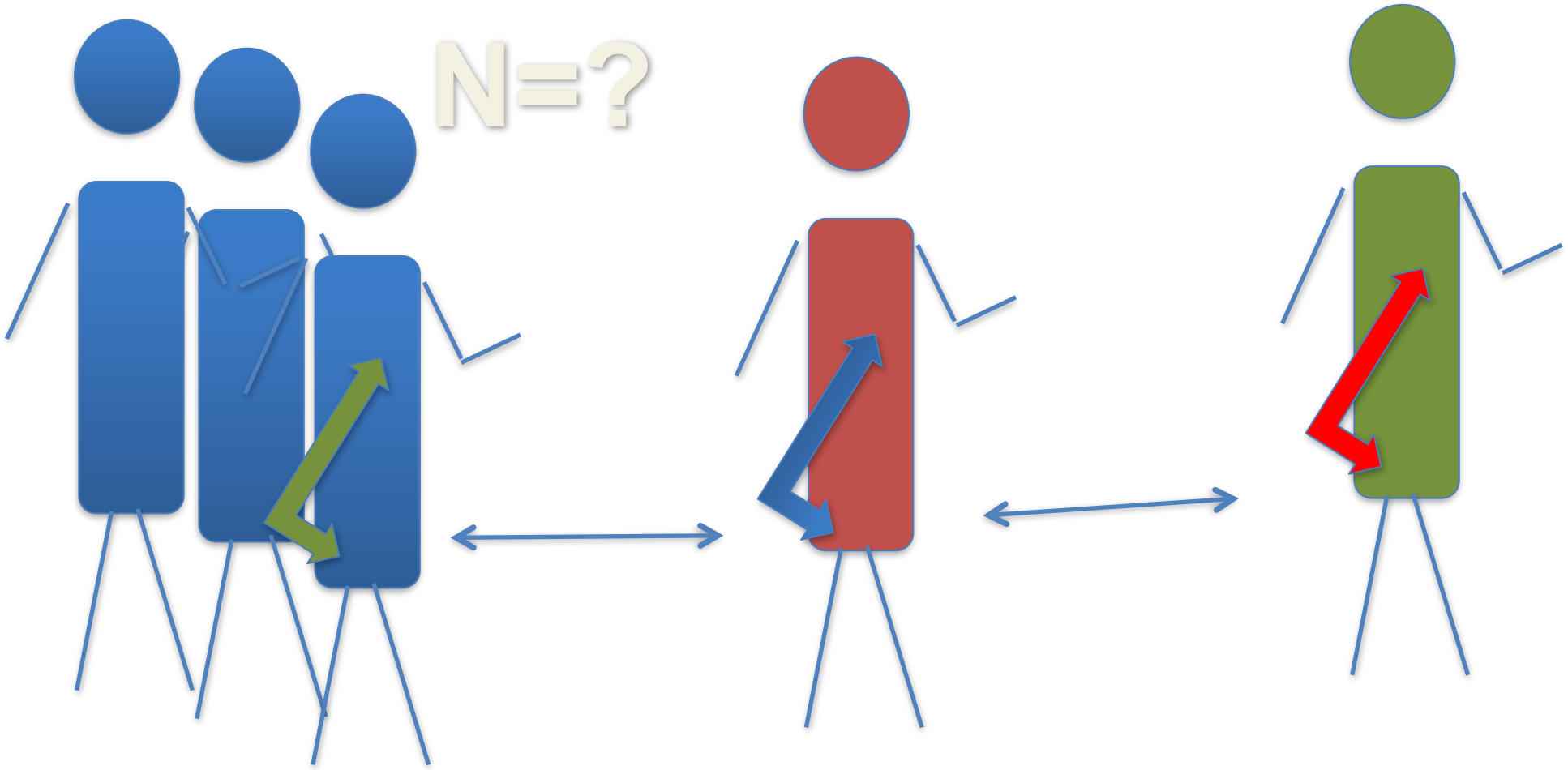
- help digest food
- regulate the immune system
- protect the organism against other disease causing bacteria
- produce vitamins B12, K, thiamine, riboflavin
- produce 100s of neurochemicals that the brain is able to use to regulate physiological and mental processes such as memory, mood and learning
- produce about 95% of the body's supply of serotonin (happy neurotransmitter that people experiencing mental health imbalances (for ex. depression) are lacking)

Main functions of bacteria in the gut



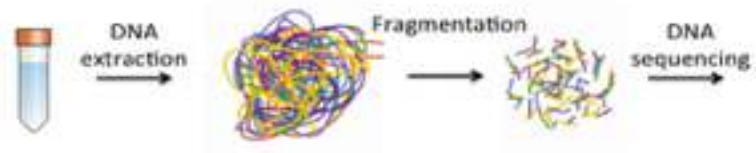
Debby Laukens et al. FEMS Microbiol Rev 2015;femsre.fuv036

Fundamental questions

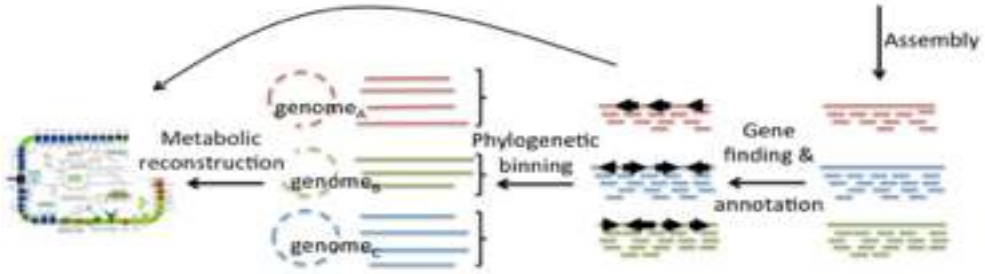


Metagenomic approaches

1. **16S rRNA**
2. **Shotgun sequencing**



3. **Metagenome analyses**
   - 3.1. Taxonomic diversity
   - 3.2. Phylogeny and community structure
   - 3.3. Assembly
   - 3.4. Gene calling and annotation
   - 3.5. Molecular function and biological roles



4. **Other analyses**
   - 4.1. Metatranscriptomes
   - 4.2. Metaproteomes
   - 4.3. Metametabolomes

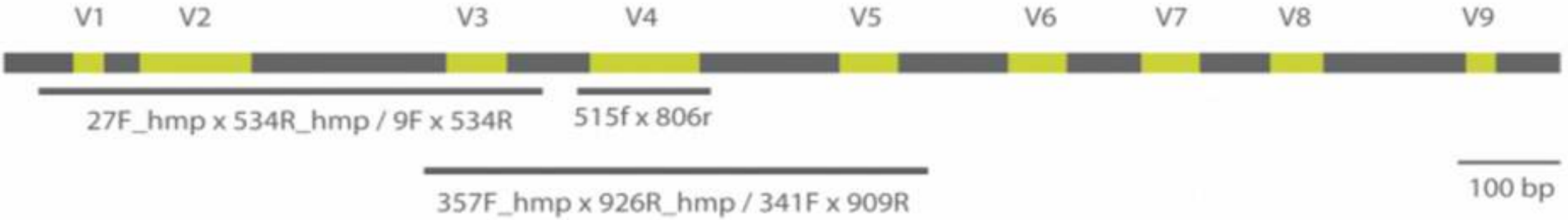
5. **Metagenomics Resources**
   - 5.1 Databases (16S rRNA – Ribosomal DB)
   - 5.2 Tools
   - 5.3 Raw data - Read Archives

NCBI

Metagenomic approaches: 16S

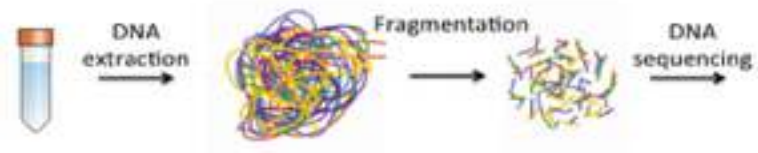
The majority of microbiome studies rely on 16S rRNA gene amplicon sequencing.

There are 9 different variable regions within the prokaryotes ubiquitous 16S rRNA gene (V1-V9), each flanked by highly conserved stretches of DNA suitable for primer binding

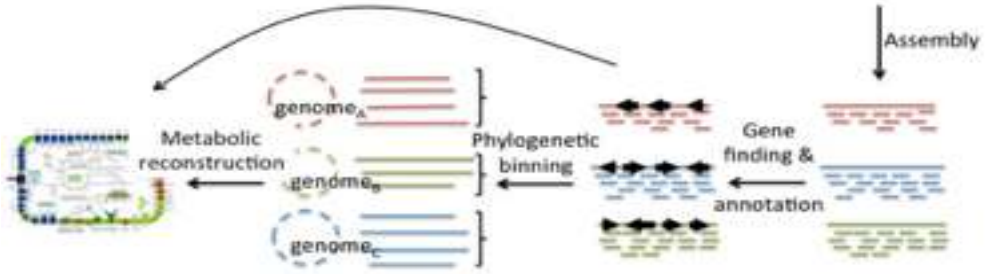


Metagenomic approaches

1. **16S rRNA**
2. **Shotgun sequencing**



3. **Metagenome analyses**
   - 3.1. Taxonomic diversity
   - 3.2. Phylogeny and community structure
   - 3.3. Assembly
   - 3.4. Gene calling and proxygenes
   - 3.5. Molecular function and biological roles

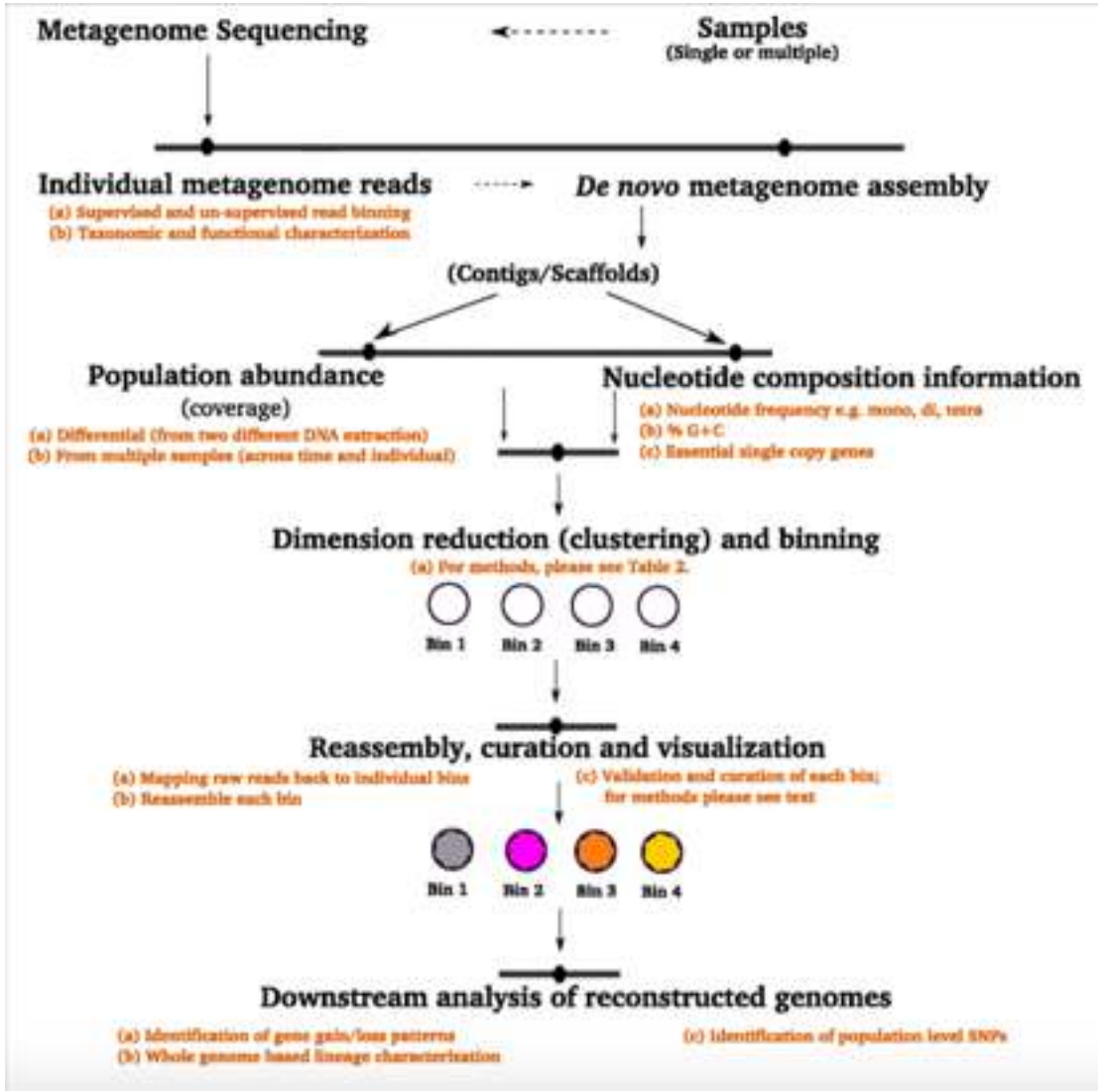


4. **Other analyses**
   - 4.1. Metatranscriptomes
   - 4.2. Metaproteomes
   - 4.3. Metametabolomes

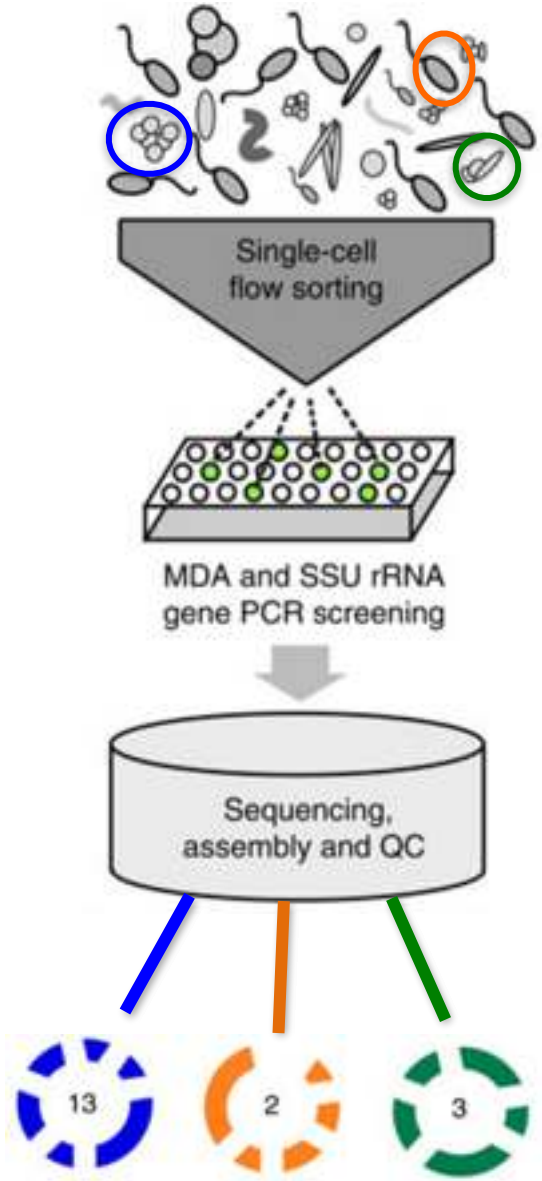
5. **Metagenomics Resources**
   - 5.1 Databases
   - 5.2 Tools
   - 5.3 16S rRNA - Ribosomal Database
   - 5.4 Raw data - Read Archives

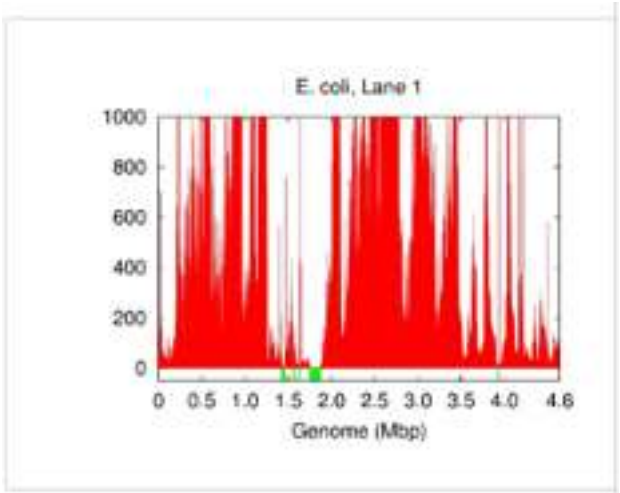
NCBI

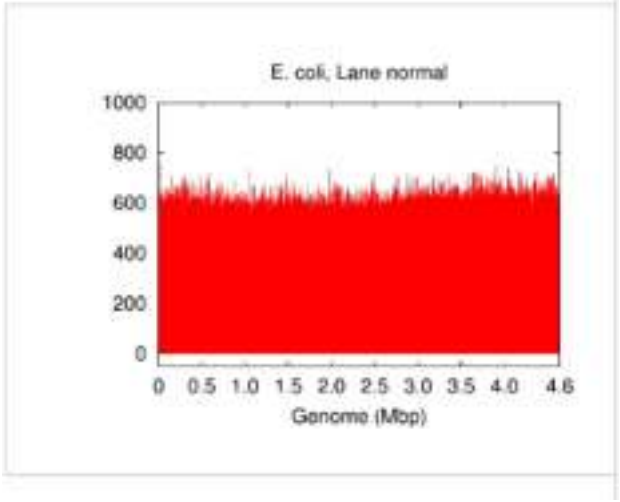
Recovering complete and draft population genomes from metagenome datasets



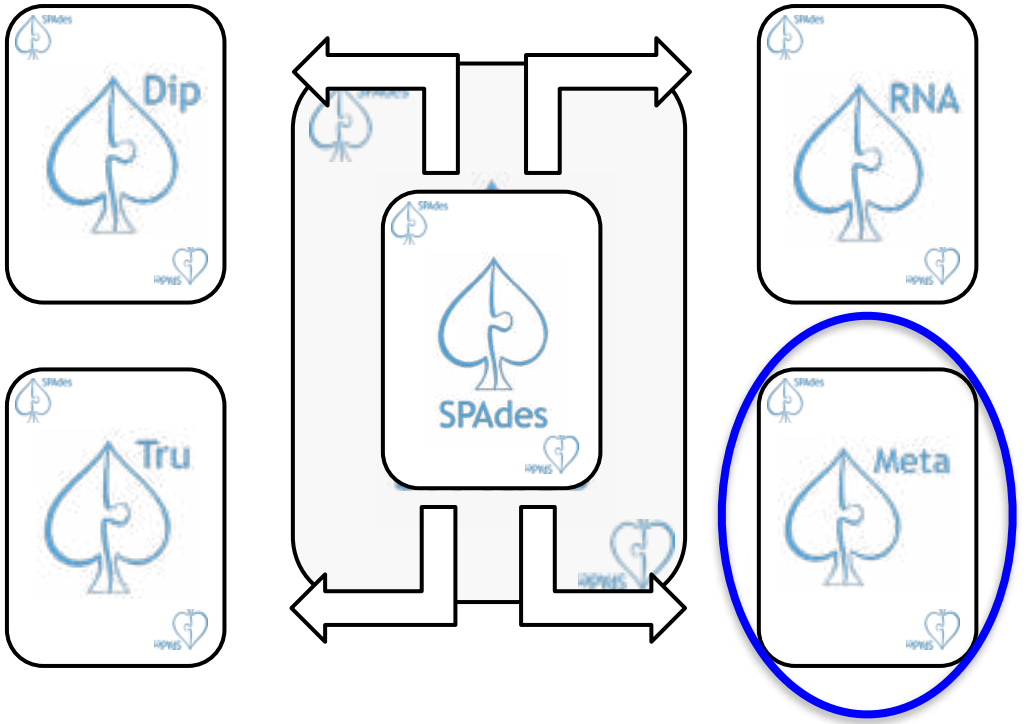
Single cell approach







Center for Algorithmic Biotechnology



http://cab.spbu.ru/

Plan your project in advance

- define the goal
- list questions to be answered
- select data analysis approaches
- publicly (commercially ?) available tools
- if none – consider analytical tool development
- IT infrustructure
- data storage

Diversity of the human microbiome

Individual humans are about 99.9% identical to one another in terms of their host genome

BUT

can be 80-90% different from one another in terms of

the microbiome

Personalized microbiota

Importance of the human-associated microbial communities has led to the continuing development of microbiome-based therapies such as fecal microbiota transplantation

BUT

microbial communities are very complex, dynamic and highly personalized ecosystems

So, fecal microbiota transplantation and other microbiome-based therapies

SHOULD BE PERSONALYSED

We need

Russian National Microbiome Initiative

THE WHITE HOUSE

OFFICE OF SCIENCE AND TECHNOLOGY POLICY

on May 13, 2016

Announced the National Microbiome Initiative

To support cross-ecosystem microbiome studies

National Institutes of Health (NIH) – Data bases and ioinformatics pipelines

Plant Soil Microbial Community Consortium - mathematical modeling, genomics and bioinformatics

The University of California Center for Pediatric Microbiome Research – brings metagenomics,

metabolomics, immunology, ecology,

genomics, bioinformatics,
to the project they start

biostatistics

Dirty Money Project

part of a larger project looking into New York City's "MetaGenome"
to examine the "microbes all around us"

Goals:

- identify potential health threats
- fight flu epidemics
- prevent disease outbreaks in urban environments
- even chart the environmental impact of major storms

Banknote is a medium of exchange for hundreds of different kinds of bacteria as banknotes pass from hand to hand. Each banknote carries about 3,000 types of bacteria on its surface as well as DNA from drug-resistant microbes.

New York University, http://cgsb.as.nyu.edu/page/home

http://www.qmul.ac.uk/media/news/items/se/57395.html

Take a look

1. Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies

Anastasis Oulas, Christina Pavloudi, Paraskevi Polymenakou, Georgios A Pavlopoulos, Nikolas Papanikolaou, Georgios Kotoulas, Christos Arvanitidis, and Ioannis Iliopoulos, Bioinform Biol Insights. 2015; 9: 75–88.

Published online 2015 May 5. doi: 10.4137/BBI.S12462 PMCID: PMC4426941

2. A bioinformatician's guide to metagenomics.

Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P. Microbiol Mol Biol Rev. 2008 Dec;72(4):557-78, Table of Contents. doi: 10.1128/MMBR.00009-08.

3. Metagenomic sequence assembly tools

https://omictools.com/metagenomic-assembly-category

Thank you ☺️


Bacteroides
Demodex
Malassezia
Endolimax
Methanobacter
Micrococcus
Brevibacterium
Lactobacillus
YOU ARE NEVER ALONE.


Feeling lonely?
Just remember, you're not alone.