Metagenomics

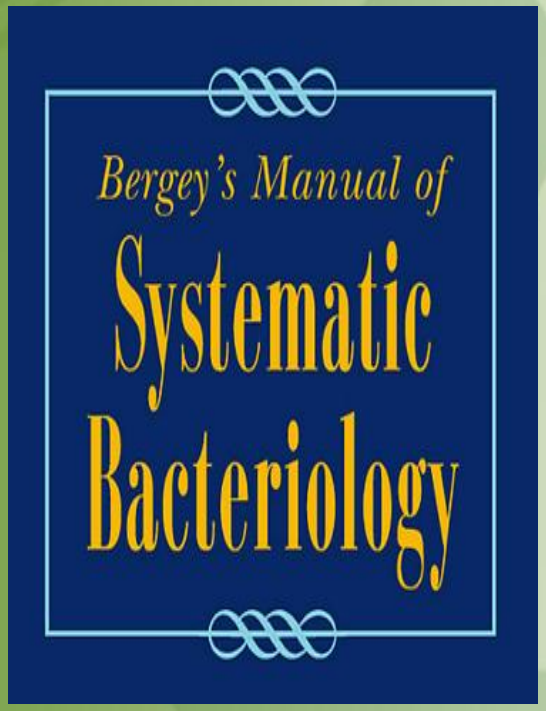
Lapidus, A.L.

Microbes run the world

Microbes are the foundation of the biosphere and have dominated life on Earth for most of its 4.5 billion year history.

Life is completely dependent upon these microscopic life forms

Only ~ 2% of all microorganisms can be grown in the lab environment



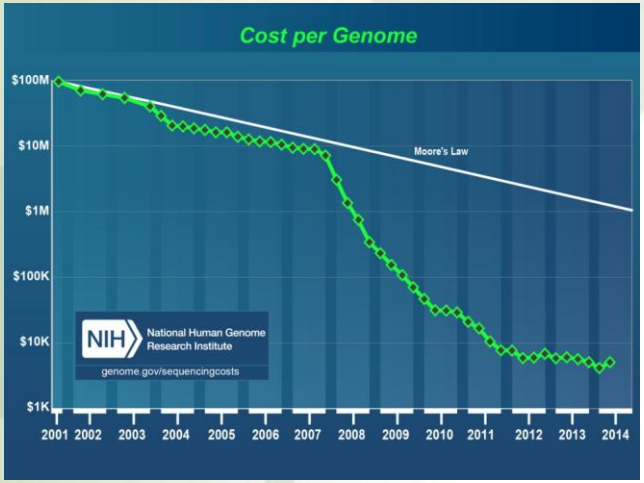
The first edition of *Bergey's Manual of Determinative Bacteriology* was print in 1923

One of the most authoritative works in bacterial taxonomy

What is metagenomics ?

Metagenomics is the study of genomes from whole communities rather than individual species

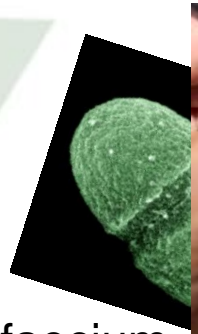
Recent advances and decreases in cost have allowed biologists to study genomes of organisms that cannot survive on their own



Where do they live?

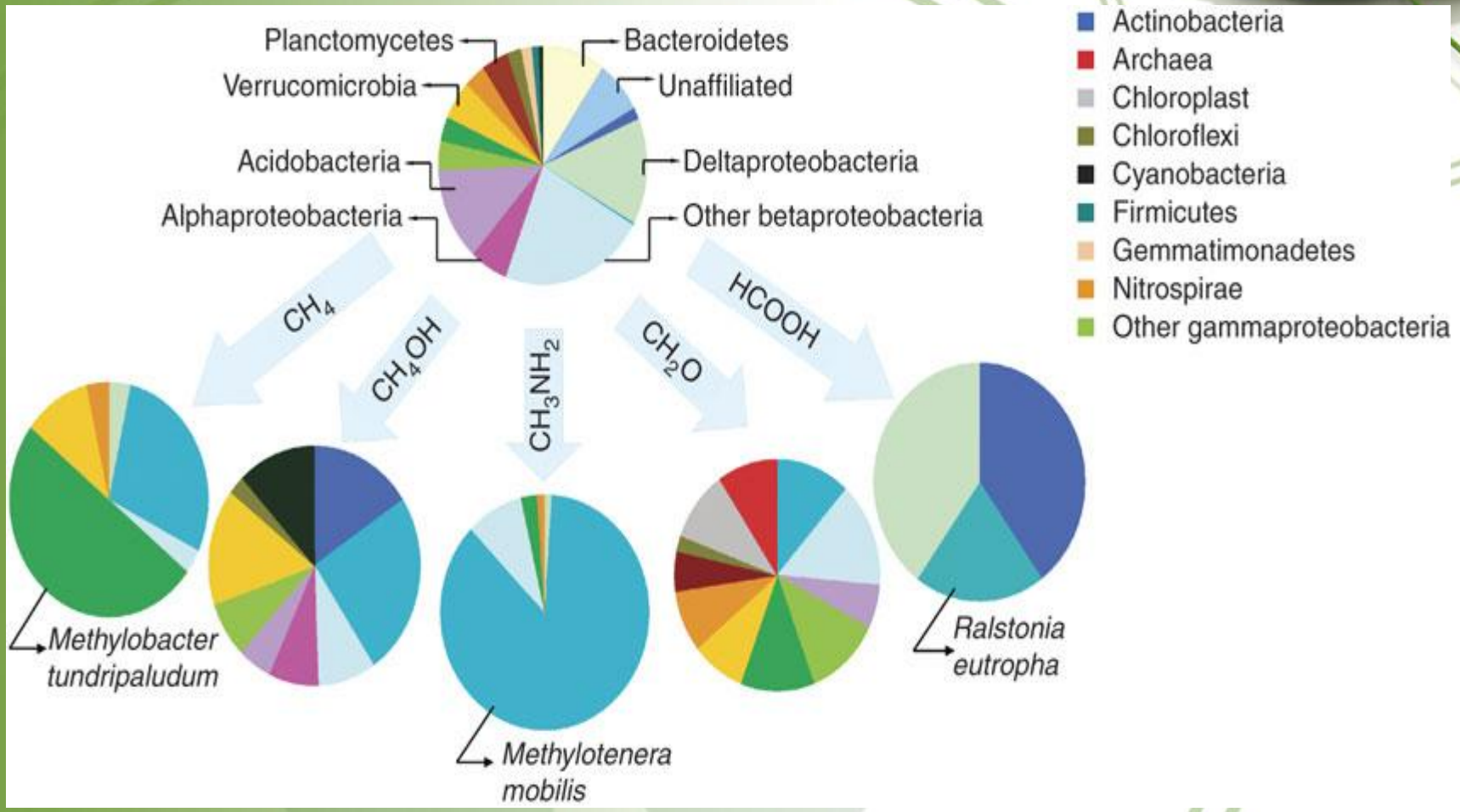
Everywhere: earth, air, water, plants, with us (skin), within us (gut, nose, etc).

MRSA



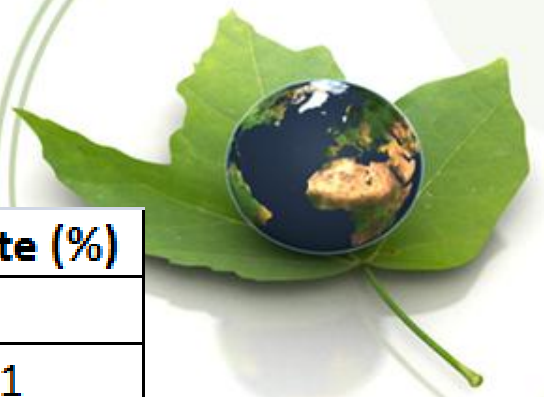
Enterococcus faecium

Simple Communities are Very Complex



High-resolution metagenomics targets specific functional types in complex microbial communities
M. Kalyuzhnaya, A.Lapidus, N. Ivanova, A.Copeland, A. McHardy, E. Szeto, A.Salamov, I. Grigoriev, D. uciu, S. Levine, V.M. Markowitz, I.Rigoutsos, S.Tringe, D. Bruce, P. Richardson, M.Lidstrom & L.Chistoserdova
Nature Biotechnology **26**, 1029 - 1034 (2008) Published online: 17 August 2008

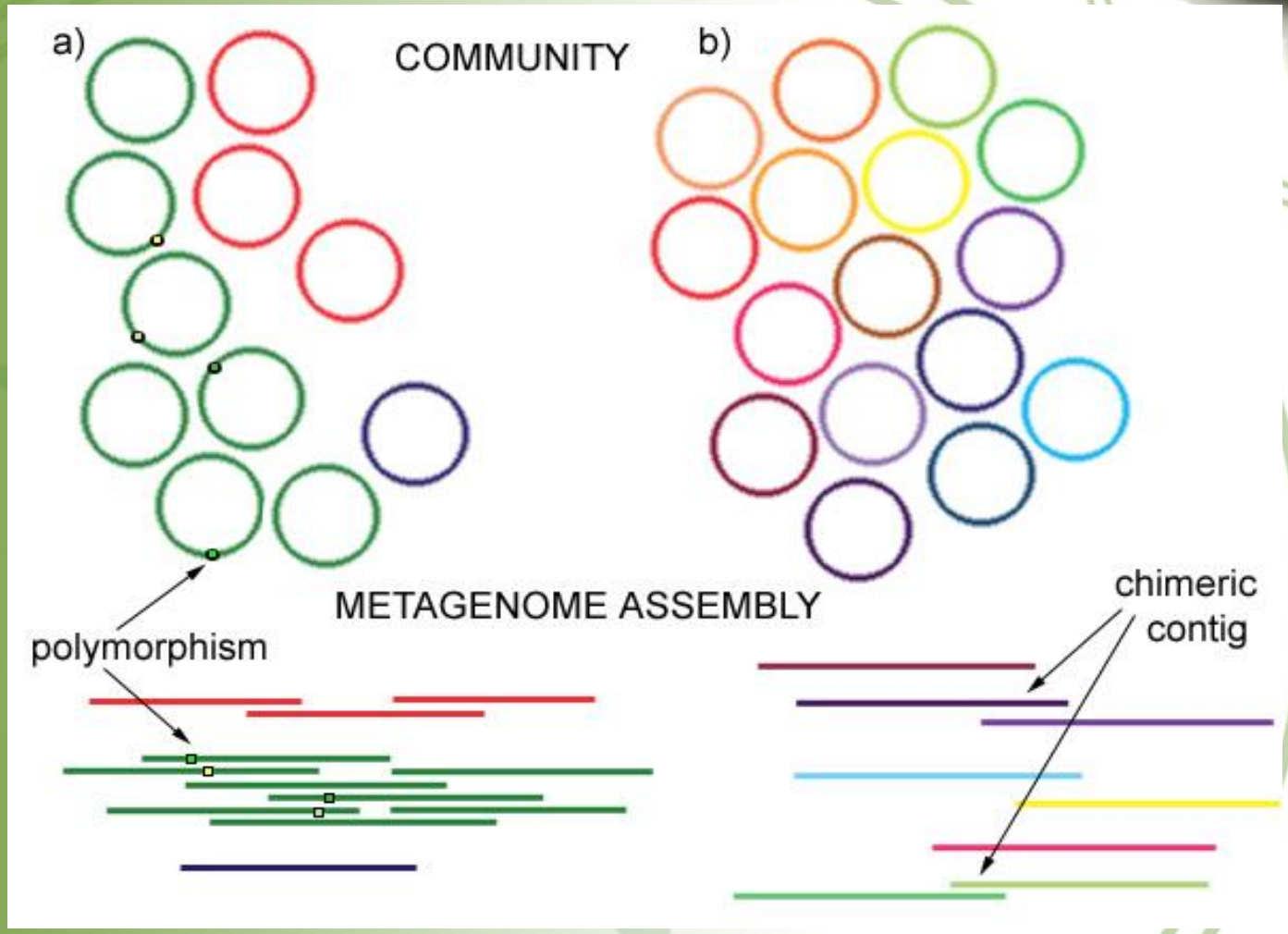
Shotgun sequencing



Platform	Read length (bp)	error type	error rate (%)
454 FLX	700	indels	1
Illumina	100-300	substitution	>0.1
Life technologies (SOLiD)	33-75	A-T Bias	>0.06
Ion torrent	100-200	indels	1
PacBio	>1500	Insertions	13-15

- Sargasso Sea-low nutrient level
  - 7 libraries; 1.6 Gb DNA from this study
  - only 3% of this was accounted for
  - 1.2 million genes found similar genes in database
  - less than 1/3 could be assigned to a cellular role
  - most genes couldn't be assigned to phylogenetic group
  - encode for phosphorus uptake, proteorhodopsin, and many others
- Nutrient rich environments
  - agricultural soil and whale falls
  - couldn't complete genomes because of high number of organisms, but identified gene families of importance
    - environmental gene tag (EGT)

Assembly



http://wiki.biomine.skelleftea.se/wiki/index.php/Metagenomics

Assembly Challenges

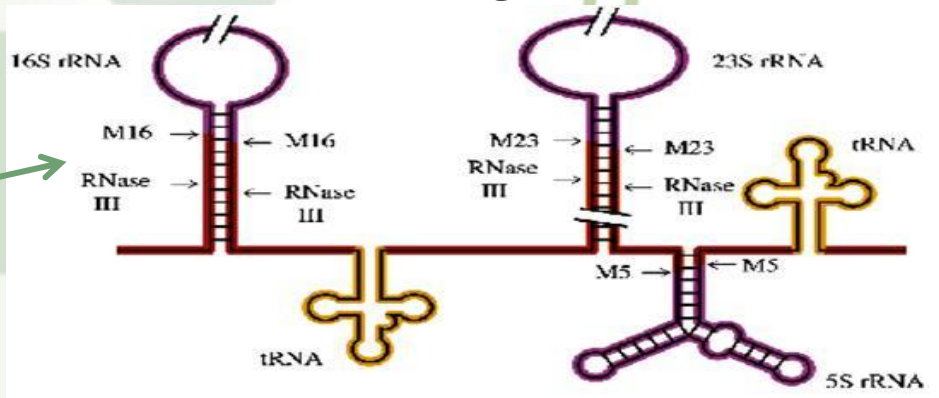
- Typically size of metagenomic sequencing project is very large
- Different organisms have different coverage. Non-uniform sequence coverage results in significant under- and over-representation of certain community members
- Low coverage for the majority of organisms in highly complex communities leads to poor (if any) assemblies
- Chimerical contigs produced by co-assembly of sequencing reads originating from different species.
- Genome rearrangements and the presence of mobile genetic elements (phages, transposons) in closely related organisms further complicate assembly.
- No assemblers developed for metagenomic data sets
- Quality of the sequence you work with is even more important

Binning

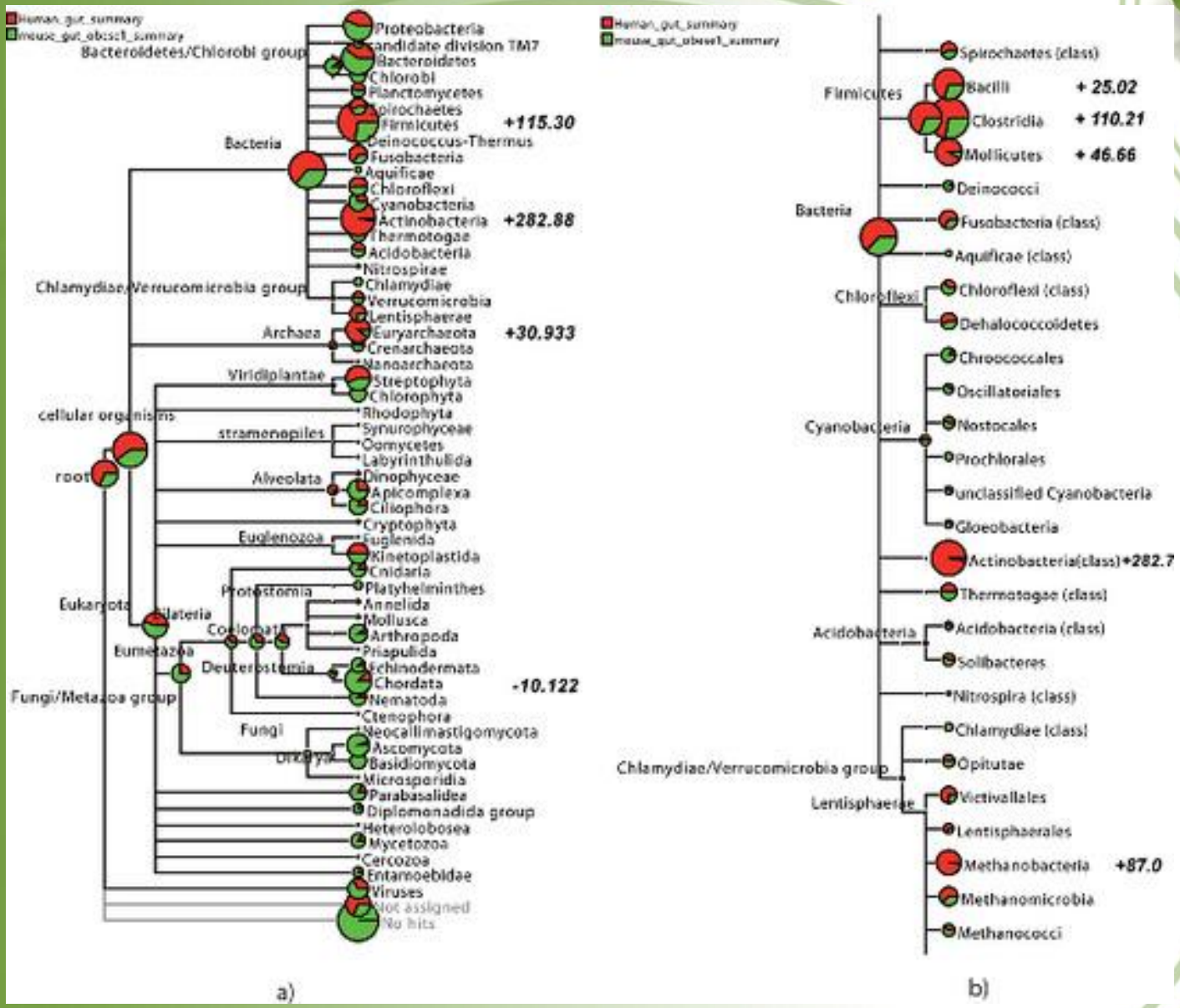
One of the essential steps in metagenome analysis is reconstruction of draft (whole??) genomes for population of a community

'Taxonomic binning' corresponds to the process of assigning a taxonomic identifier to sequence fragments, based on information such as sequence similarity, sequence composition or read coverage.

(16S ribosomal RNA)



16S survey (MEGAN)

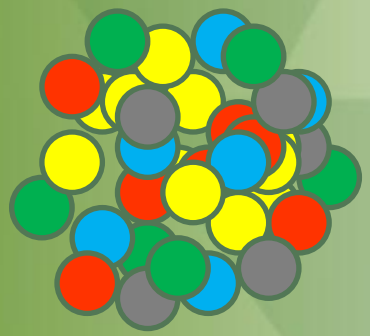


Human gut microbioita

Binning (2)

Tetranuclotide frequency usage patterns in genomic fragments

$4^4 = 256$



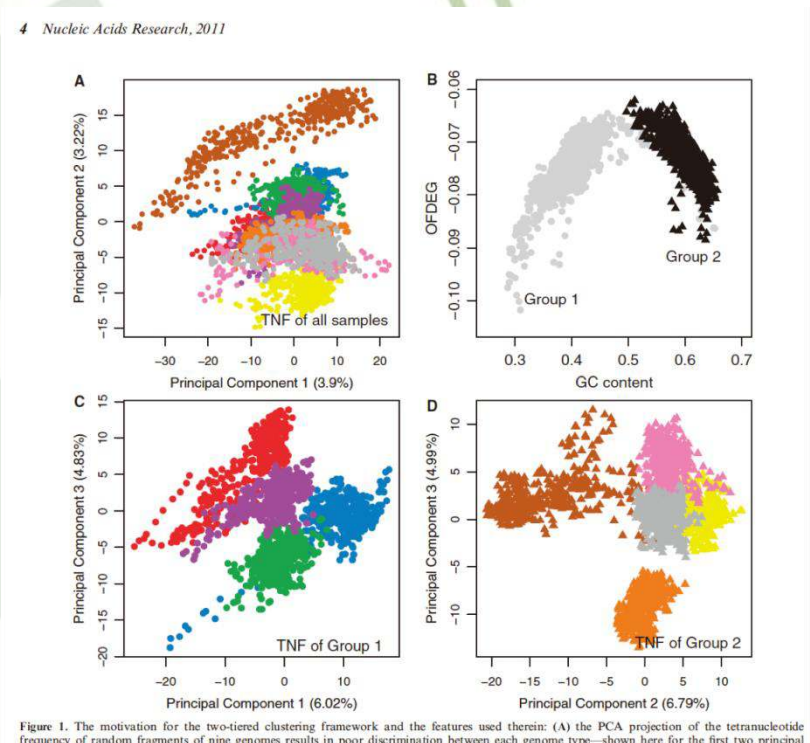
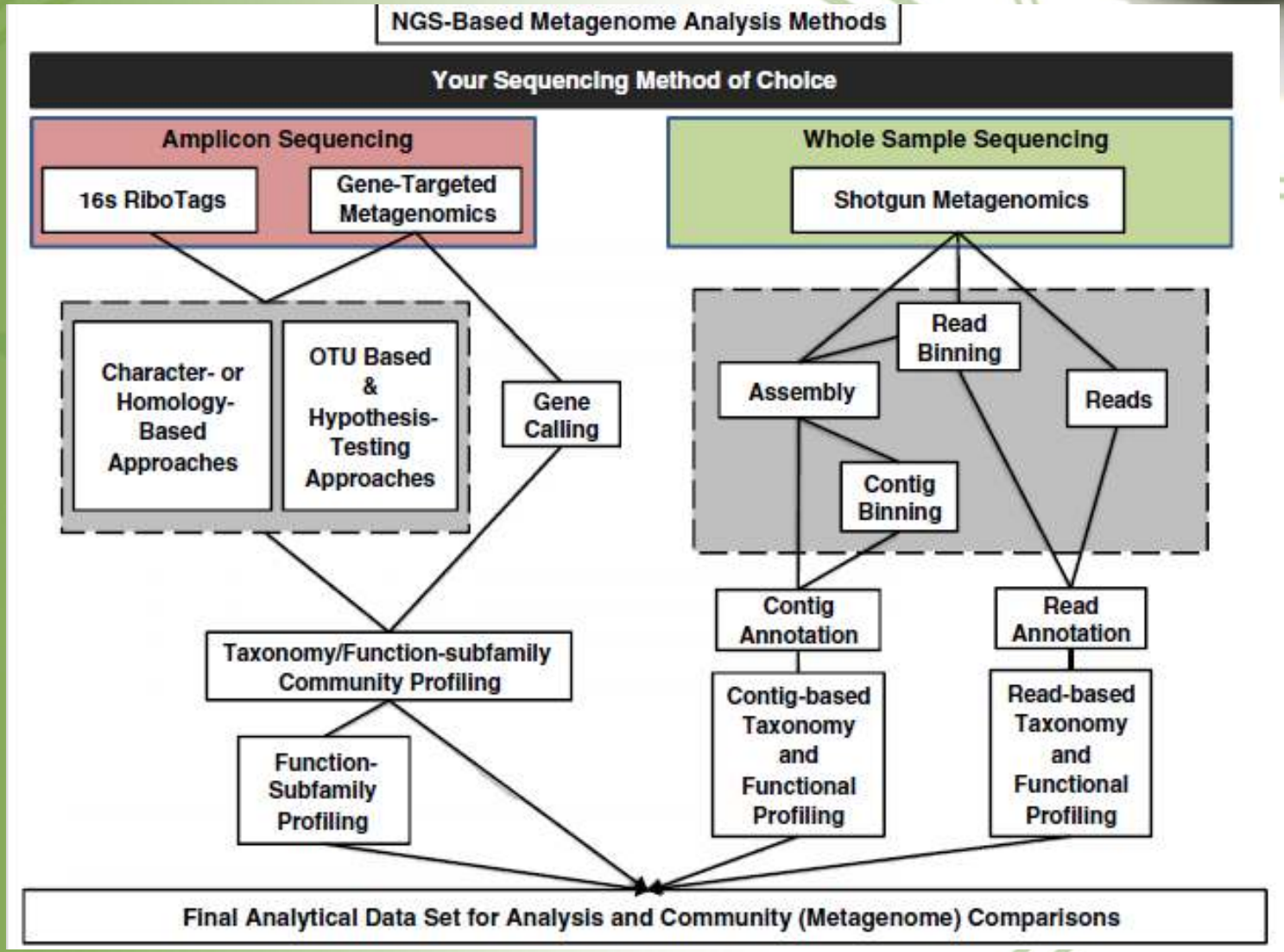


Figure 1. The motivation for the two-tiered clustering framework and the features used therein: (A) the PCA projection of the tetranucleotide

Analysis methods for metagenomics



Metagenome analysis

Most metagenomes = microbes+ viral and eukaryotic components

No centralized method to annotate such diverse sequences

Microbial online metagenome annotation services
IMG-M, MG-RAST

Virus specific annotation pipeline – VMGAP

Workflow managements systems: CAMERA and Galaxy

Comparative metagenomics

IMG-M, CAMERA and MG-RAST allow

- taxonomic and functional assignments,
- taxonomic comparison
- pathway reconstruction
- analyze against finished references

Terragenome: International Soil Metagenome Sequencing Consortium

"We know more about the movement of celestial bodies than about the soil underfoot." - *Leonardo da Vinci*

Terragenome is an international group of scientists interested in soil metagenomics all over the world.

The **goal** of the project is providing the first complete sequence of a soil metagenome over a ten-year period.

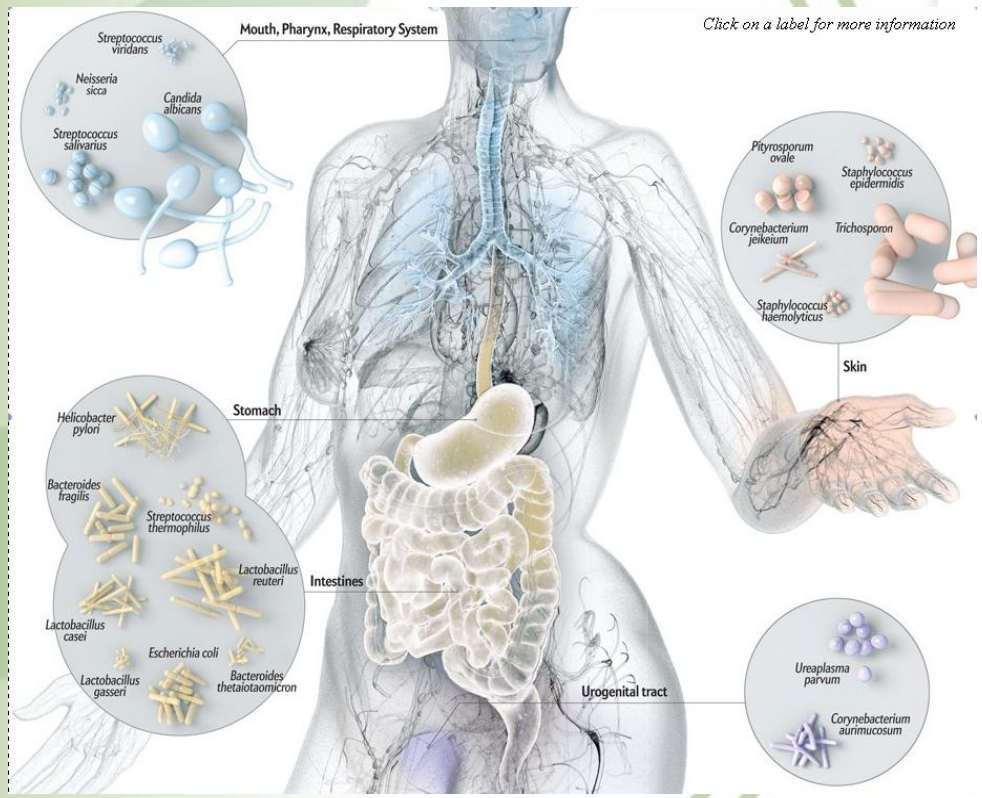
These metagenome sequence data will constitute the "reference" sequence to which other soils around the world could be compared.

Metagenomics of the Human Intestinal Tract

The Human Microbiome Project (HMP)

NIH HUMAN MICROBIOME PROJECT

You are what you eat!



Gut bacteria have a role in obesity

Input: four sets of human twins in which one of each pair was lean and one was obese,

Mice given bacteria from a lean twin stayed slim, whereas those given bacteria from an obese twin quickly gained weight, even though all the mice ate about the same amount of food.

Jeffrey Gordon, Science, 2013

Forensic genomics

use of NGS for crime investigations and missing person identification, kinship testing and ancestry investigation

Task: reveal Forensic DNA evidences from tiny and highly mixed samples.

Study:

-short tandem repeat (STR) typing (repeating units of 2–6 nucleotides

-mitochondrial DNA analysis,

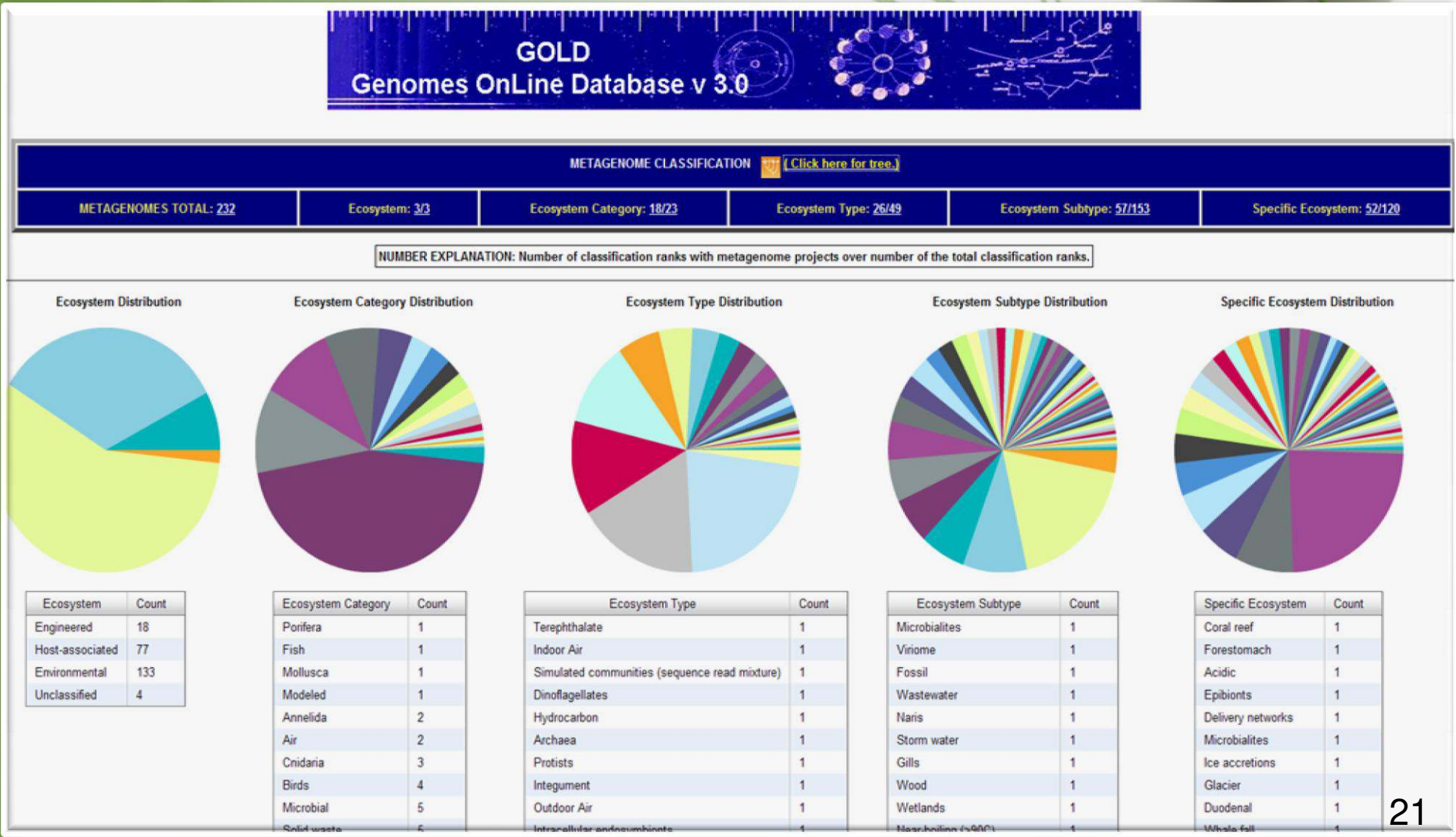
-dense panels of single nucleotide polymorphisms (SNPs) offering

And more….


Lignocellulose & humic acids
Mono- & oligomers
N_2
Organic acids
Amino acids
H_2 & CO_2
Acetate
CH_4
CO_2


Nata-Lia/Shutterstock

Metagenome Classification



Ecosystem	Count
Engineered	18
Host-associated	77
Environmental	133
Unclassified	4

Ecosystem Category	Count
Porifera	1
Fish	1
Mollusca	1
Modeled	1
Annelida	2
Air	2
Cnidaria	3
Birds	4
Microbial	5
Solid waste	5

Ecosystem Type	Count
Terephthalate	1
Indoor Air	1
Simulated communities (sequence read mixture)	1
Dinoflagellates	1
Hydrocarbon	1
Archaea	1
Protists	1
Integument	1
Outdoor Air	1
Intracellular endosymbionts	1

Ecosystem Subtype	Count
Microbialites	1
Viriome	1
Fossil	1
Wastewater	1
Naris	1
Storm water	1
Gills	1
Wood	1
Wetlands	1
Near boiling (>90C)	1

Specific Ecosystem	Count
Coral reef	1
Forestomach	1
Acidic	1
Epibionts	1
Delivery networks	1
Microbialites	1
Ice accretions	1
Glacier	1
Duodenal	1
Whale fall	1