Metagenomics

Introduction

- **Metagenomics was first described in 1998 by Handelsman and Rodon.**
- **Metagenomics** is the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics.
- **Metagenomics** is a molecular tool used to analyse DNA acquired from environmental samples, in order to study the community of microorganisms present, without the necessity of obtaining pure cultures.
- There are two common methods **used** in **metagenomics**: shotgun sequencing and directed sequencing. In shotgun sequencing, **scientists** sequence many small sections of the genome and reconstruct the entire genome by figuring out how these small sections fit together.

Metagenomics enables the study of all microorganisms, regardless of whether they can be cultured or not, through the analysis of genomic data obtained directly from an environmental sample, providing knowledge of the species present, and allowing the extraction of information regarding the functionality of microbial communities in their natural habitat.

Metagenomics has been applied to explore pharmaceutical and industrial products from environmental samples. The most popular genes that have been isolated by metagenomics were polyketide synthases (PKS) genes, which are key enzymes for synthesizing polyketide antibiotics.

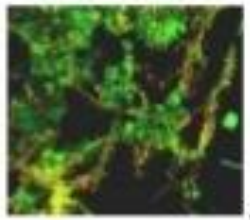
Metagenomic approaches are now commonly used in microbial ecology to study microbial communities in more detail, including many strains that cannot be cultivated in the laboratory.
Bioinformatic analyses make it possible to mine huge metagenomic datasets and discover general patterns that govern microbial ecosystems.

Each organism's specific DNA sequence frames the story of its life: how it grows and develops, how it uses energy, and how it reproduces. ... Thus, we use DNA sequence similarity as a way of telling how similar or different two organisms are from one another.

Metagenomics will be the systems biology of the biosphere.

Metagenomics provides a means for studying microbial communities on their own "turf."
Complex ecological interactions—including lateral gene transfer, phage-host dynamics, and metabolic complementation—can now be studied with the lens of metagenomics.

The study of uncultured Microorganisms has expanded beyond asking "WHO IS THERE" to include the different question "WHAT ARE THEY DOING"



DNA extraction

100++ Abundant species (≈3 Mbp each)

Sequencing

Reads

100-150 bp

Assembly

Contigs

1000+ bp

Search against database

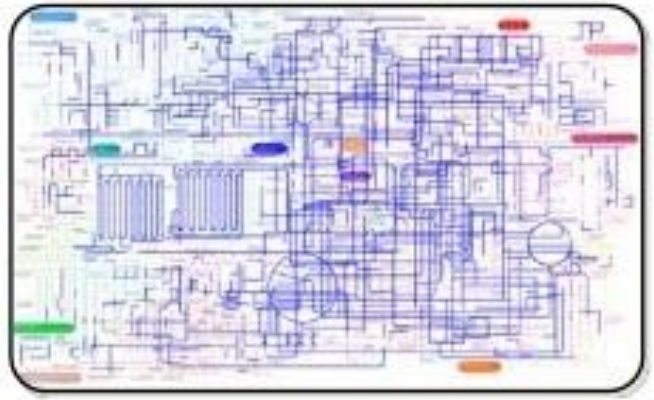
Phylogenetic classification

Who is there?

Bacterium A
Bacterium B
...
Bacterium X

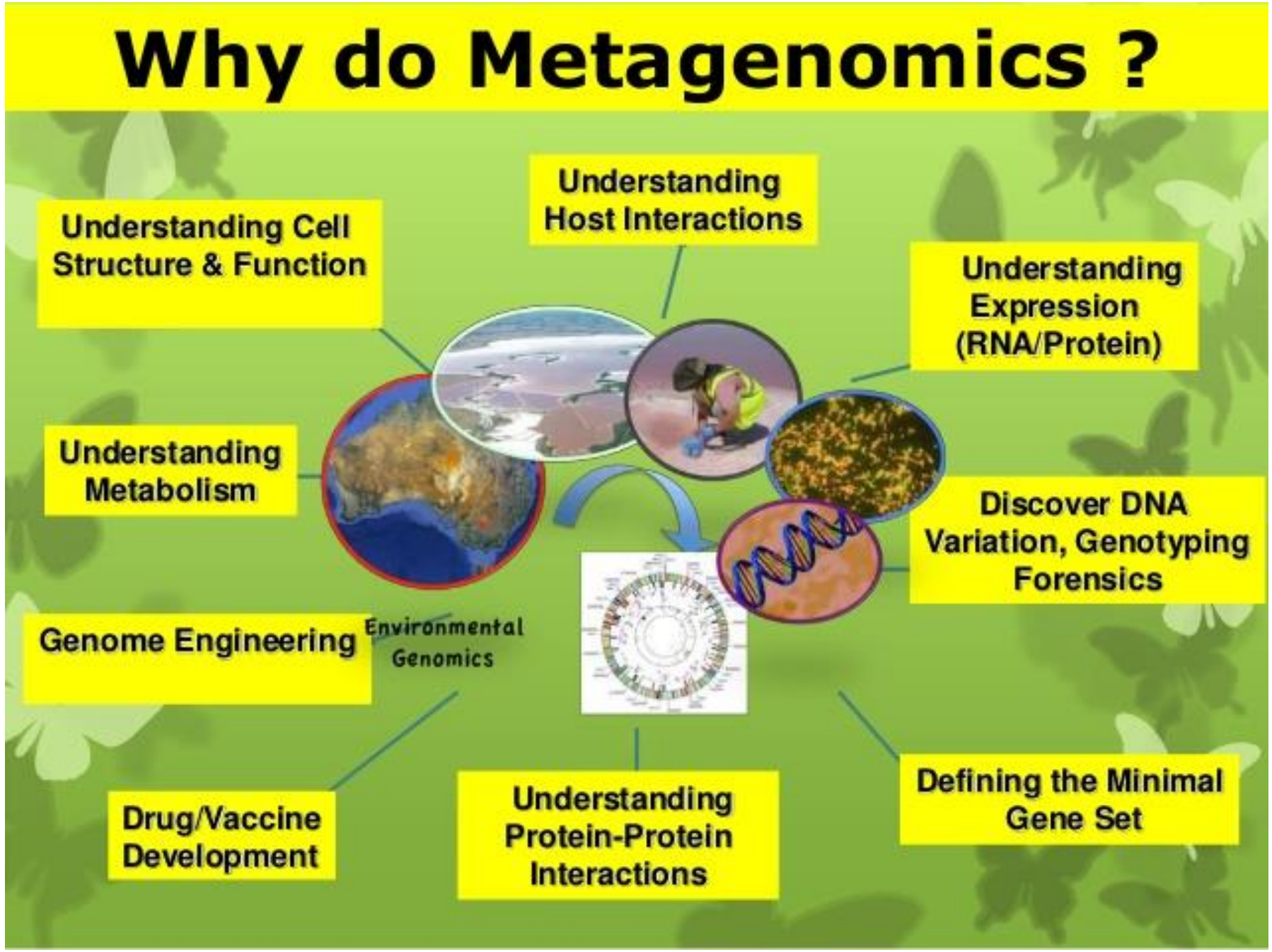
Functional classification

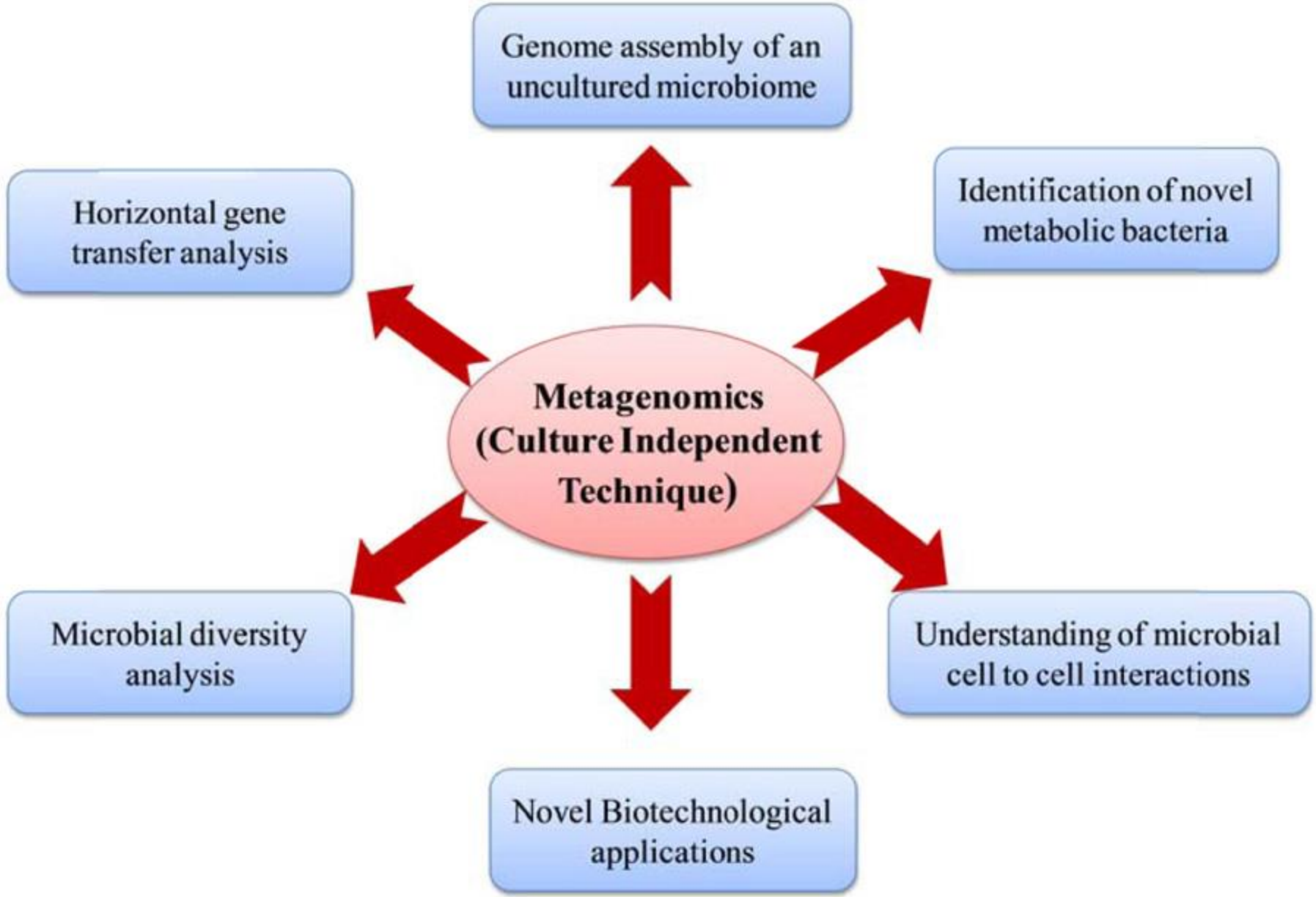
What can they do?



Gene A
Gene B
...
Gene X

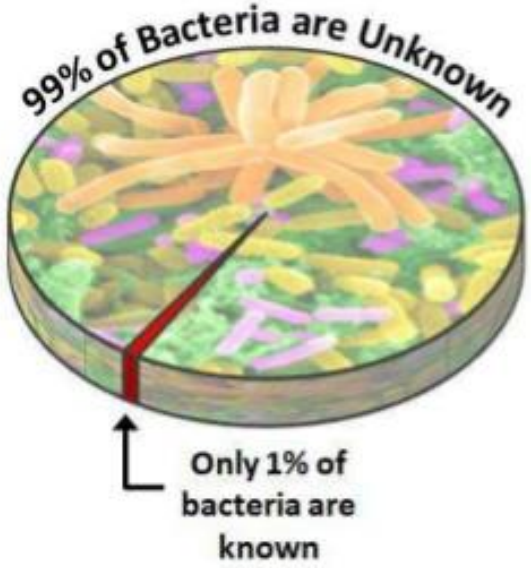
Why do Metagenomics ?




Genome assembly of an uncultured microbiome
Horizontal gene transfer analysis
Identification of novel metabolic bacteria
Metagenomics (Culture Independent Technique)
Microbial diversity analysis
Understanding of microbial cell to cell interactions
Novel Biotechnological applications

Introduction

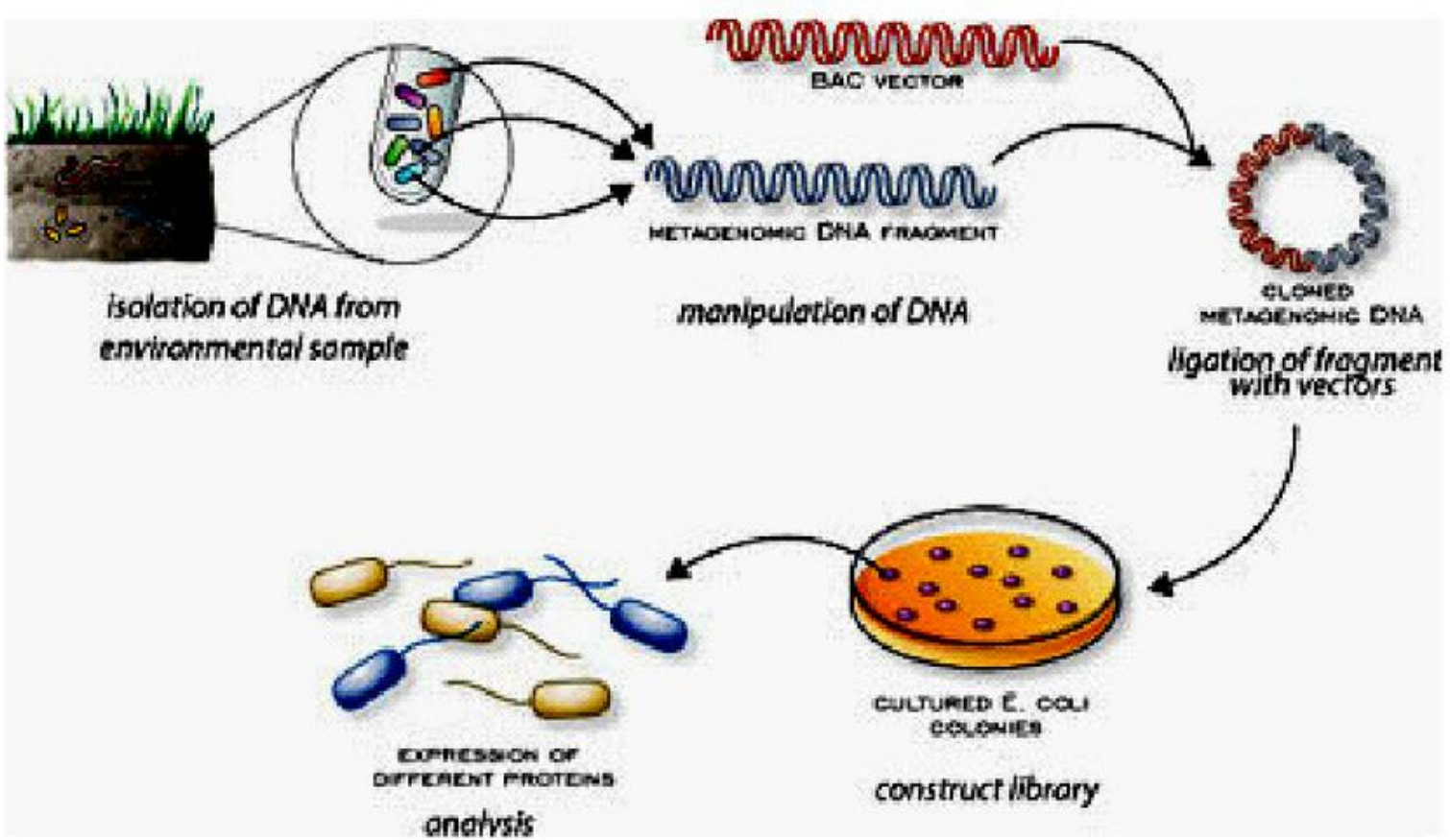
Microbiology Till Metagenomics

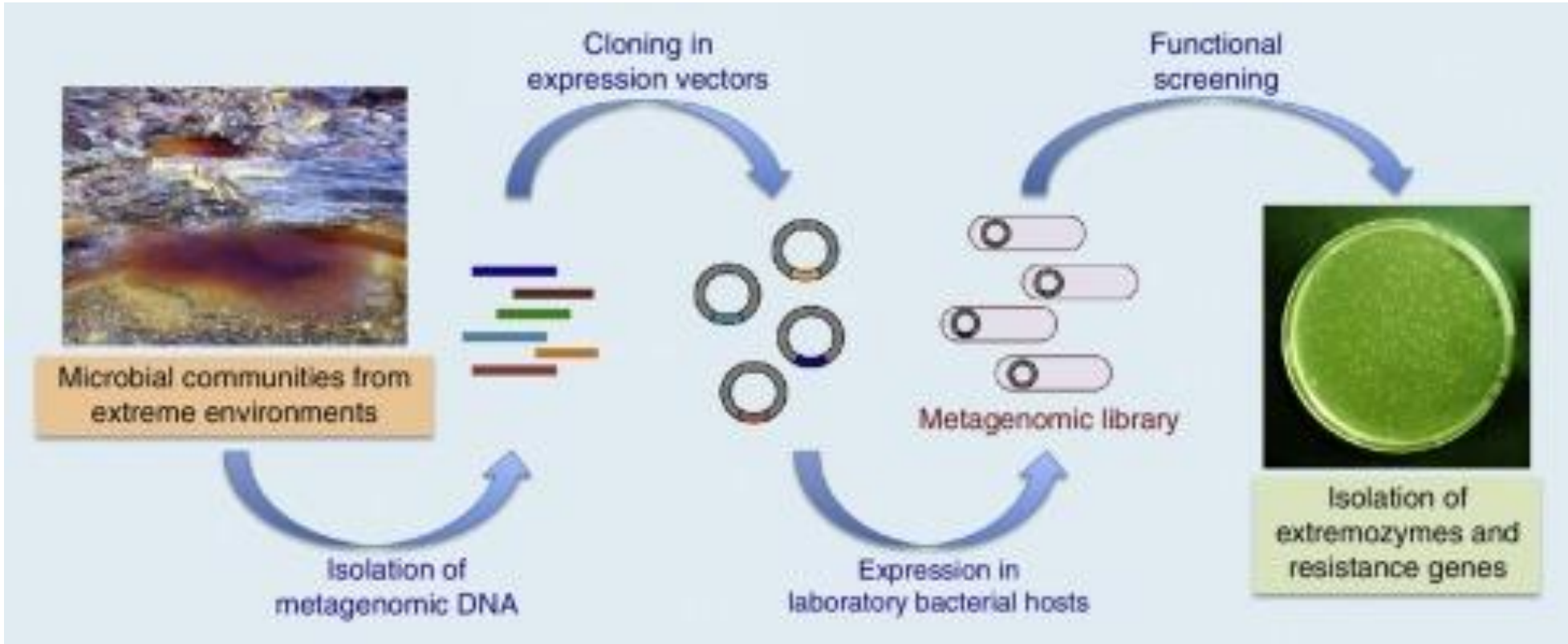


- Until recently, microbiology research required culture-based techniques.
- <1% of all microbes can be cultured (Sleator *et al*. 2008)
- Traditional techniques have left us with a biased and incredibly incomplete understanding of microbes.
  - Sequence the genome of one organism at a time
  - Use cultures to isolate microbe of interest

Steps and process of metagenomics:

Sample collection.
DNA extraction.
Sample pre-preparation.
Sample analysis. PCR. DNA sequencing. DNA microarray.
Bioinformatics studies.
Results and interpretation.




Cloning in
expression vectors
Functional
screening
Microbial communities from
extreme environments
Metagenomic library
Isolation of
metagenomic DNA
Expression in
laboratory bacterial hosts
Isolation of
extremozymes and
resistance genes

What is Metagenomics?

- Traditional microbial genomics
  - Sequence the genome of one organism at a time
  - Use cultures to isolate microbe of interest
- Metagenomics
  - Extract sequence data from microbial communities as they exist in nature
  - Bypass the need for culture techniques
    - Sequence all DNA in sample
    - Select DNA based on universal sequences

direct isolation of DNA from the environment

metagenomics

DNA knowledge application

Metagenomics can in principle access 100% of the genetic resources of an environment.

cultivation

DNA knowledge application

isolation of DNA

cultivatable species

genomics

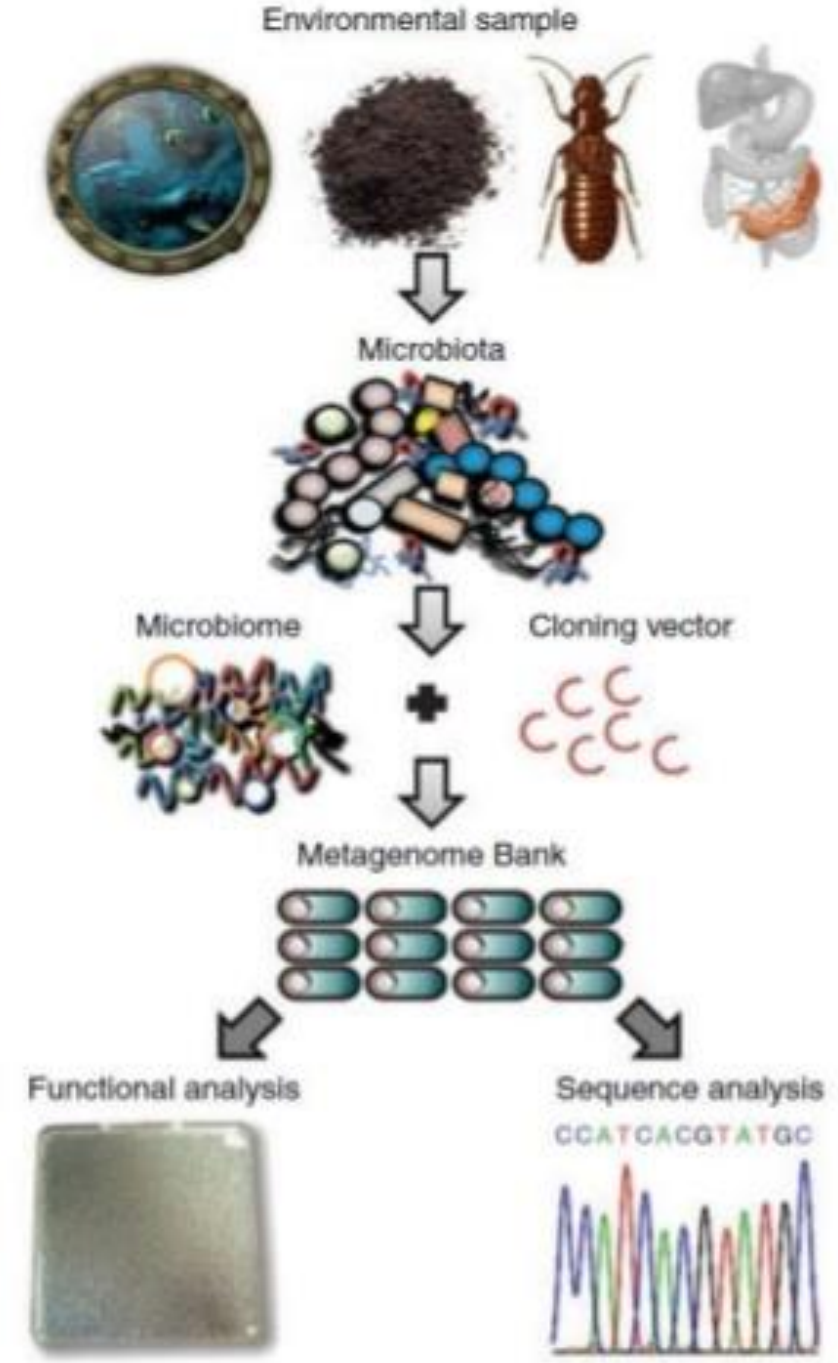
Traditional cultivation methods and traditional genomics can at best access 1%.

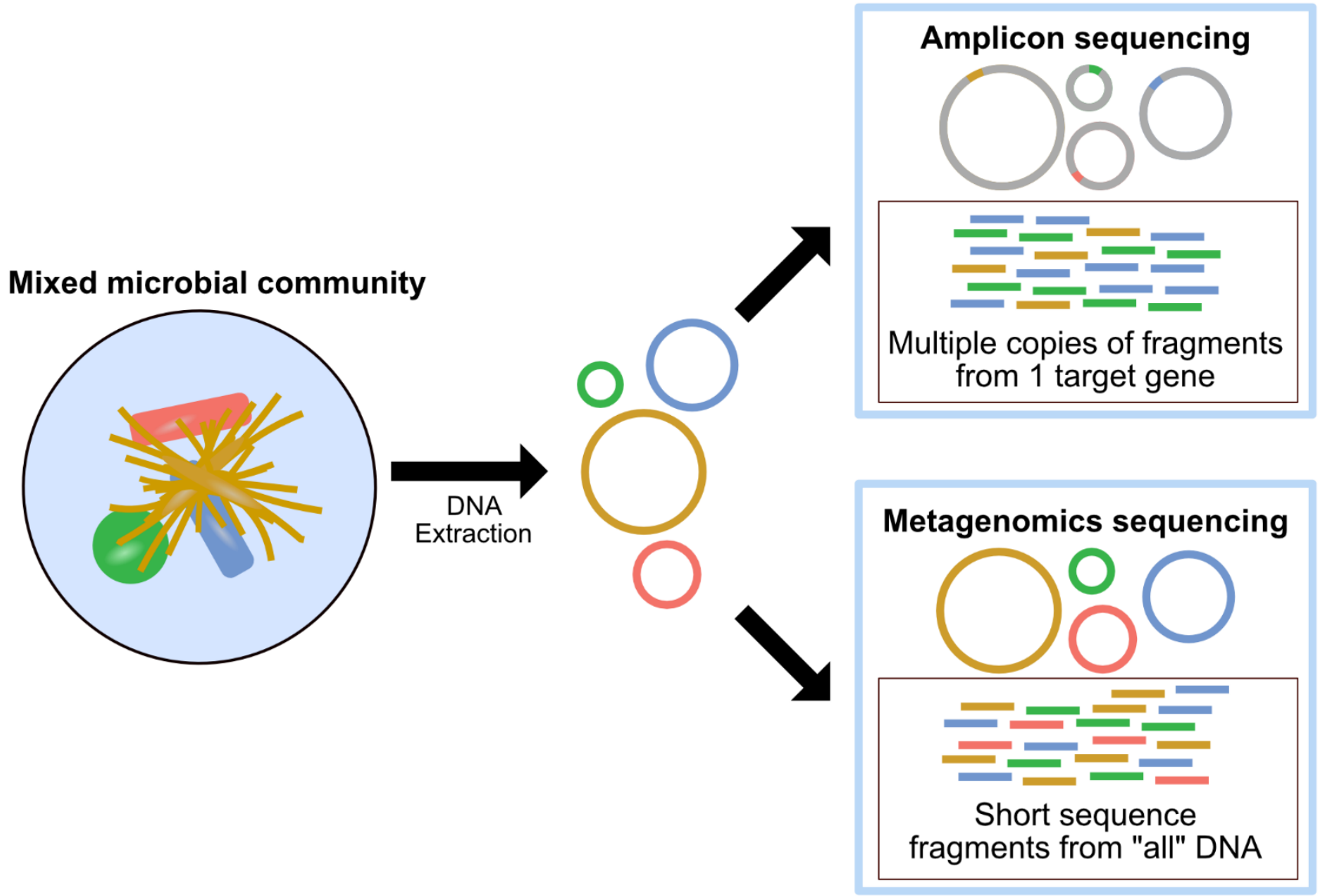
Type of metagenomics

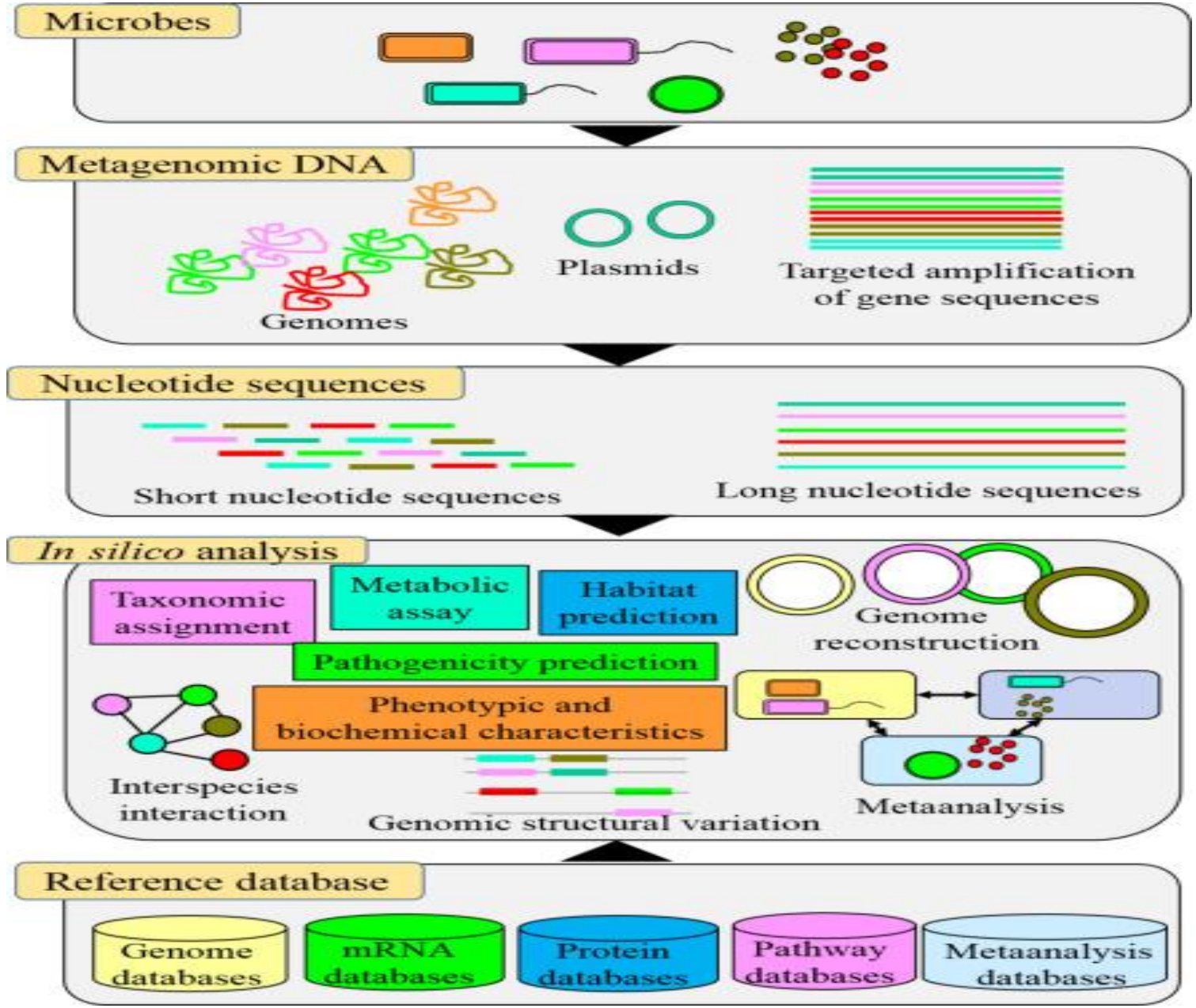
- There are two basic types of Metagenomics studies

I. **Sequence-based Metagenomics-** involves sequencing and analysis of DNA from environmental samples

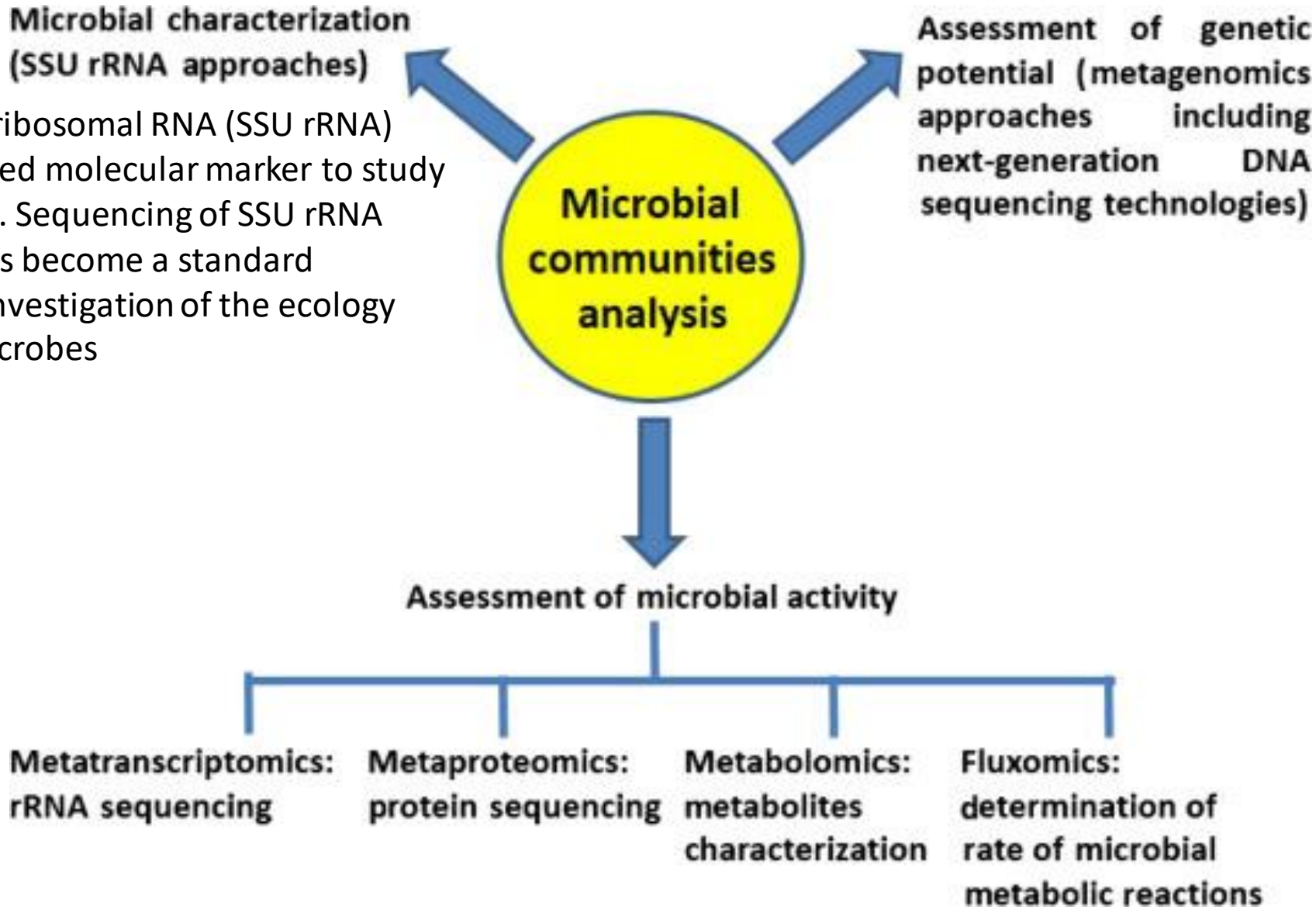
II. **Function-based Metagenomics** involves screening for a particular function or activity

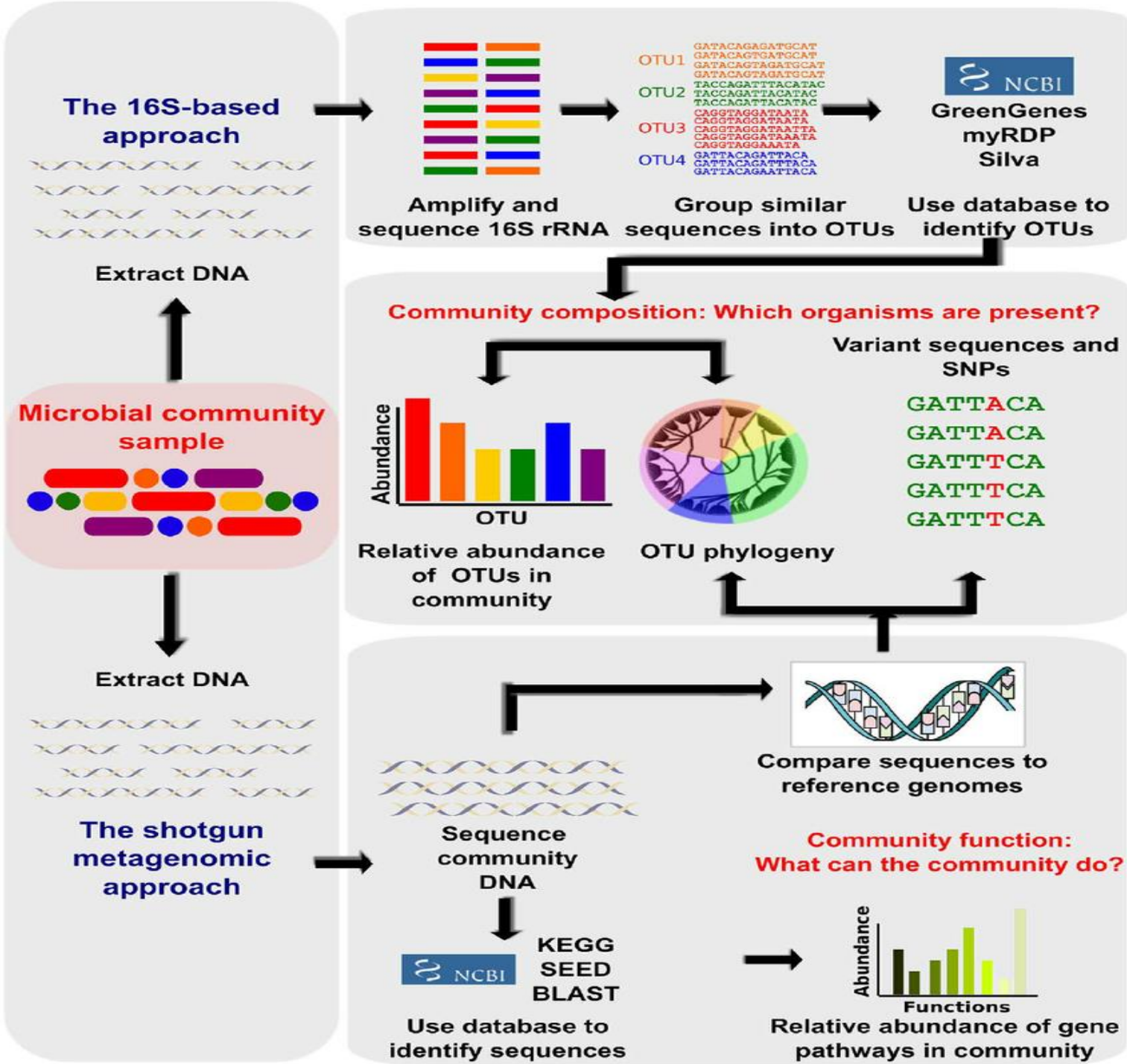



Mixed microbial community
DNA
Extraction
Amplicon sequencing
Multiple copies of fragments
from 1 target gene
Metagenomics sequencing
Short sequence
fragments from "all" DNA


Microbes
Metagenomic DNA
Genomes
Plasmids
Targeted amplification of gene sequences
Nucleotide sequences
Short nucleotide sequences
Long nucleotide sequences
In silico analysis
Taxonomic assignment
Metabolic assay
Habitat prediction
Genome reconstruction
Pathogenicity prediction
Phenotypic and biochemical characteristics
Interspecies interaction
Genomic structural variation
Metaanalysis
Reference database
Genome databases
mRNA databases
Protein databases
Pathway databases
Metaanalysis databases

The small subunit ribosomal RNA (SSU rRNA) gene is a widely used molecular marker to study the diversity of life. Sequencing of SSU rRNA gene amplicons has become a standard approach for the investigation of the ecology and diversity of microbes





Operational Taxonomic Unit is an operational definition used to classify groups of closely related individuals

Genomic DNA extraction

Heterologous genomic DNA

+

Restriction-digested vector

Ligation

Vector DNA

+

E. coli DH10B

Transformation

Function-driven analysis

Heterologous gene expression

Transcription

mRNA

Translation

Protein

Secretion

Metagenomic library

Sequence-driven analysis

Cloned DNA preparation

atgacgac...gatttaca
tgggctcccatcgctag

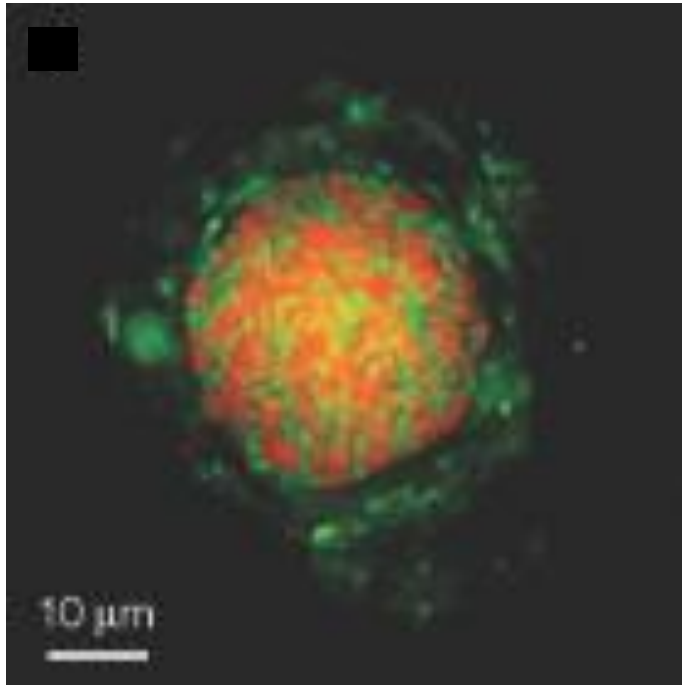
Genomic sequence analysis

Analysis of Metagenomics Data

- Metagenomes are big
  - Soil has as many as 40,000 individual microbial species
  - Soil metagenome orders of magnitude bigger than human genome
- Analyzing the metagenome
  - Screens
  - Phylogenetic studies
  - Sequencing uncultivated organisms
  - Studying metagenome under different conditions

Understanding Microbial Communities

- Some questions metagenomics may answer
  - Are certain adaptations observable across environmental gradients?
  - How do different species interact?
  - Can lateral gene transfer be detected?



From Figure 2 in Schleper, C., et al. (2005) "Genomics studies of uncultivated Archaea" Nature Reviews Microbiology 3: 479-488.

Permission for figure 2 granted by K.Knittel and T.Loesekann,MPI Bremen, www.mumm-research.de.

Applications

- $2.3 billion in sales of industrial enzymes in 2003
- Discovery of novel enzymes and catalysts with industrial uses by screening thousands of microbial species simultaneously
- Look for pharmacologically interesting genes (e.g. antibiotics) that exist in organisms that cannot be cultured

Caveats (Limitations)

- Problems with DNA purification
- Sample contamination
- Issues with sequencing
  - Immensity of metagenome (gigabases)
  - Errors in assembly due to inter-species similarities
  - Difficulties in sequencing less well-represented genomes

Metagenomics projects

- The Acid Mine Drainage (AMD) Project

Biofilms growing on the surface of flowing AMD in the five-way region of the Richmond mine at Iron Mountain, California, were sampled in March 2000

- Acid is produced by oxidation of sulfide minerals that are exposed to air as a result of mining activity.
- **Why AMD Biofilm?**
- Extreme acidic environment (self-contained ecosystem)
- Scientists are interested in the metabolic potential of such an environment: nitrogen fixation, sulfur oxidation, and iron oxidation
- To understand how the microbes tolerate the extremely acidic environment
- And it is a good pick -- low species complexity

Metagenomics projects

- Preliminaries
- Group-specific fluorescence in situ hybridization (FISH)
- – Results indicated the presence of mixtures of bacteria (*Leptospirillum*, *Sulfobacillus* and, in a few cases, *Acidimicrobium*) and archaea (*Ferroplasma* and other members of the *Thermoplasmatales*)
- 16S ribosomal RNA gene clone library – 384 clones were end-sequenced
- Results indicated the presence of three bacterial and three archaeal lineages. The most abundant clones are close relatives of *L.ferriphilum* and belong to *Leptospirillum* group II

Metagenomics projects

- Whole-genome Assembly (for Metagenomic Data)
- Assembled using JAZZ (whole-genome shotgun assembler)
- Over 85% of the reads were assembled into 1,183 scaffolds longer than 2 kb; with 92.7% of end pairs from the same clone assembled with the appropriate orientation and separation.
- Assignment of scaffolds to organism type: separate scaffolds by ***G+C content*** (low G+C content, and G+C content bins); then subdivide them using ***read depth*** (coverage)

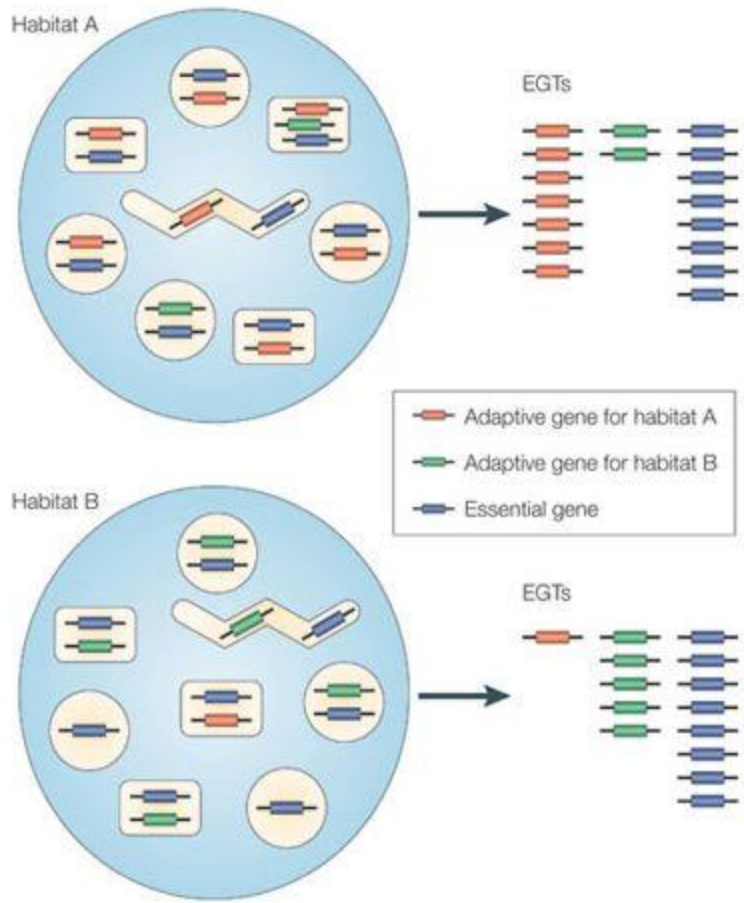
Metagenomics projects

- The Sargasso Sea Project
- Science 2004, 304:66-74
- A pilot project of Venter's global ocean voyage (with the ultimate goal of finding solutions to energy problems)
- Venter is not the first to sequence the genes of microbes from the ocean; but he is the first to do it on a large and ambitious scale
- Sargasso Sea takes its name from the sargassum seaweed that floats on its surface
- It is in the middle of the North Atlantic Ocean near Bermuda.

Environmental Gene Tag (EGT)

- **Environmental gene tags (EGTs) are short DNA sequences used to characterize and distinguish microbial environments.**
- EGTs include fragments or complete sequences of genes that are important to survive in specific environmental conditions and hence are present in the genomes of many microbes of one environment and less frequent or even completely absent in another environment.
- The frequencies of all EGT sequences in the whole genomic content of a specific environment is used as a fingerprint that characterizes the environmental adaptation of microbes.
- Genes that are more frequently found in one type of microbial community compared to another are supposed to provide beneficial functions for the community living in a specific environment.

Environmental Gene Tags



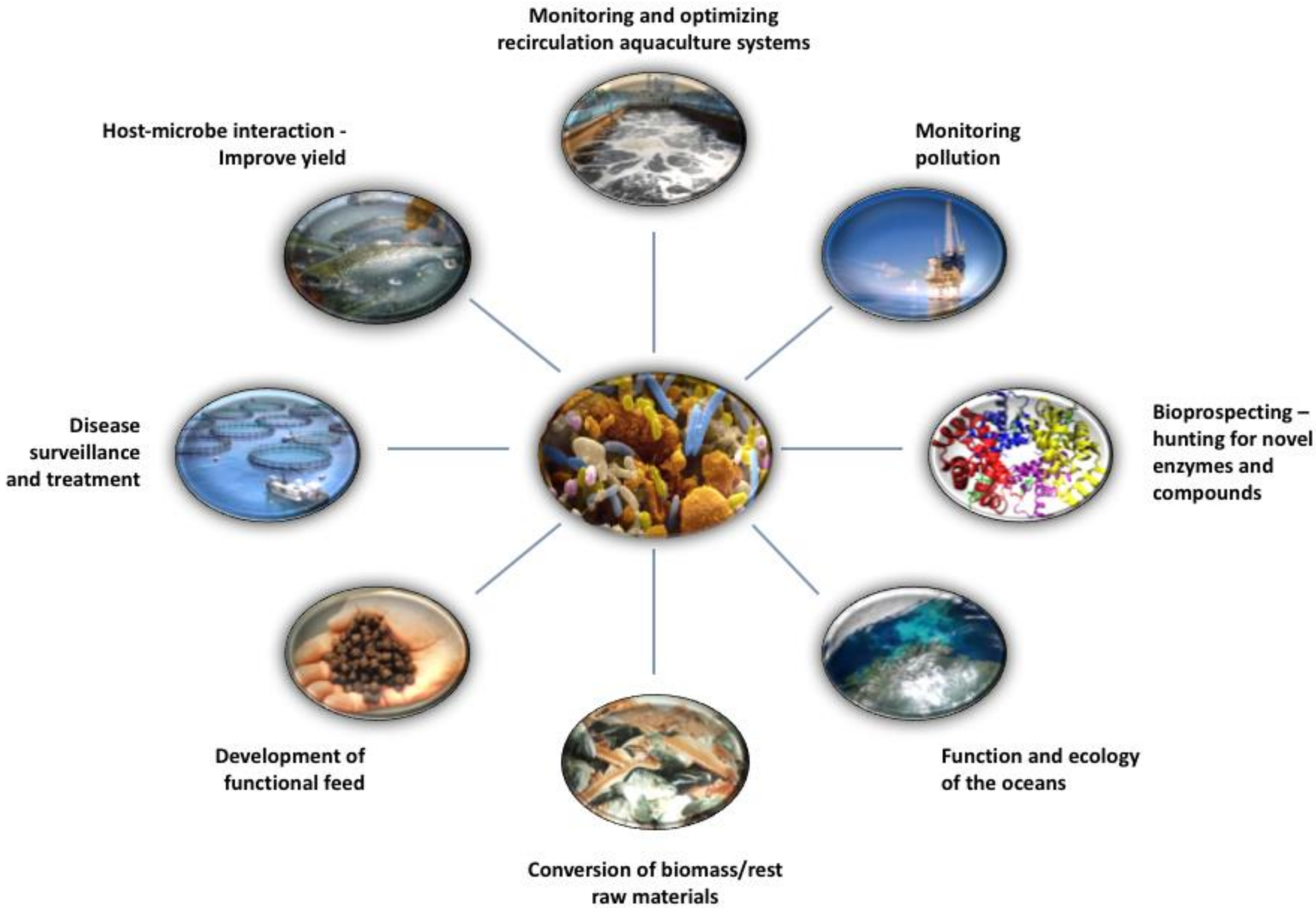
Environmental gene tags (EGTs) are short sequences from the DNA of microbial communities that contain fragments of functional genes.

Essential genes will occur frequently regardless of enviroment.

Genes adaptive for a particular environment will be occur frequently in that environment but not in others.

FUTURE OF METAGENOMICS

- To identify new enzymes & antibiotics
- To assess the effects of **age, diet, and pathologic states (e.g., inflammatory bowel diseases, obesity, and cancer)** on the distal gut microbiome of humans living in different environments
- Study of more exotic habitats
- Study antibiotic resistance in soil microbes
- Improved bioinformatics will quicken analysis for library profiling
- Investigating ancient DNA remnants
- Discoveries such as phylogenic tags (rRNA genes, etc) will give momentum to the growing field
- Learning novel pathways will lead to knowledge about the current nonculturable bacteria to then culture these systems


Monitoring and optimizing recirculation aquaculture systems
Host-microbe interaction - Improve yield
Monitoring pollution
Disease surveillance and treatment
Bioprospecting – hunting for novel enzymes and compounds
Development of functional feed
Function and ecology of the oceans
Conversion of biomass/rest raw materials

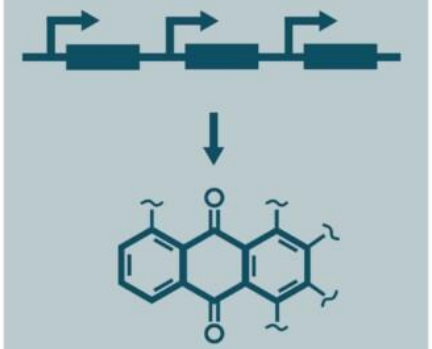
Discoveries from the field of metagenomics

Only a small fraction of the world's microorganisms have been cultured. Yet, we do have large databases of DNA sequences from a wide array of organisms. In "metagenomics," researchers use computational tools to search through these databases and find candidate DNA sequences that encode new kinds of proteins, groups of proteins, or even new organisms. Using metagenomics, we can find many new tools with biotechnological applications.

New CRISPR proteins

Uses: Expand the CRISPR toolkit and improve genome editing

Biosynthesis pathways



Uses: Create chemicals with bioactivity

New viruses

Uses: New kinds of antibiotics, sources of CRISPR proteins

MammothBiosciences