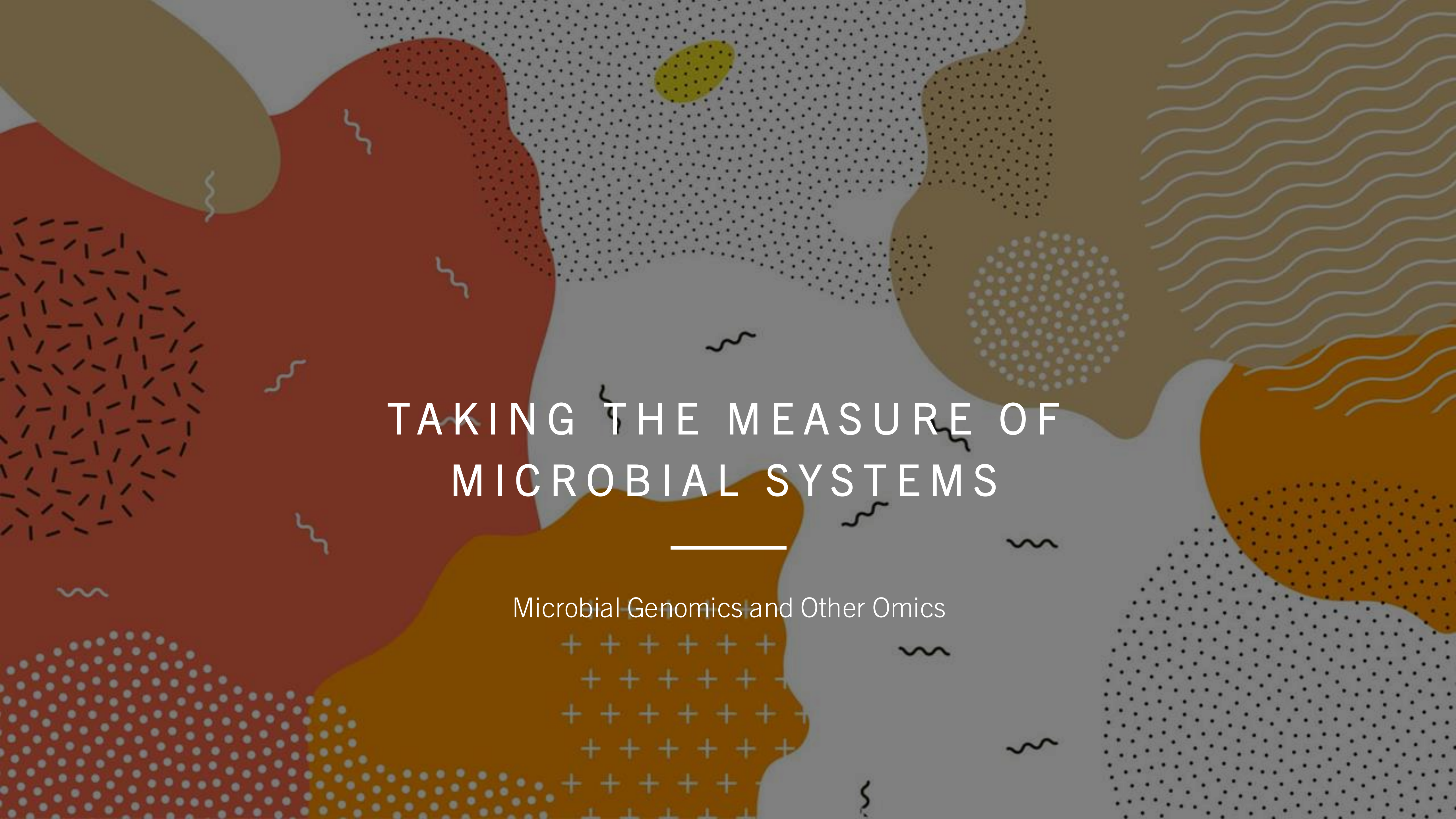
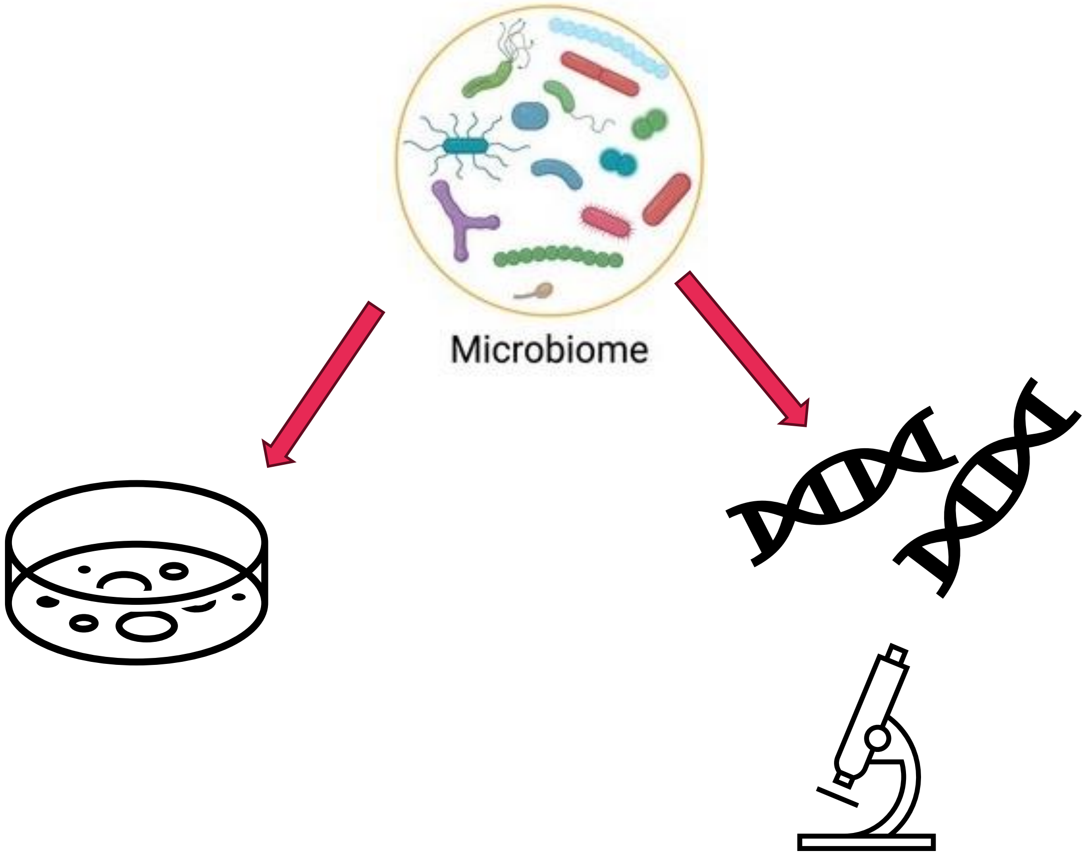


TAKING THE MEASURE OF MICROBIAL SYSTEMS

Microbial Genomics and Other Omics

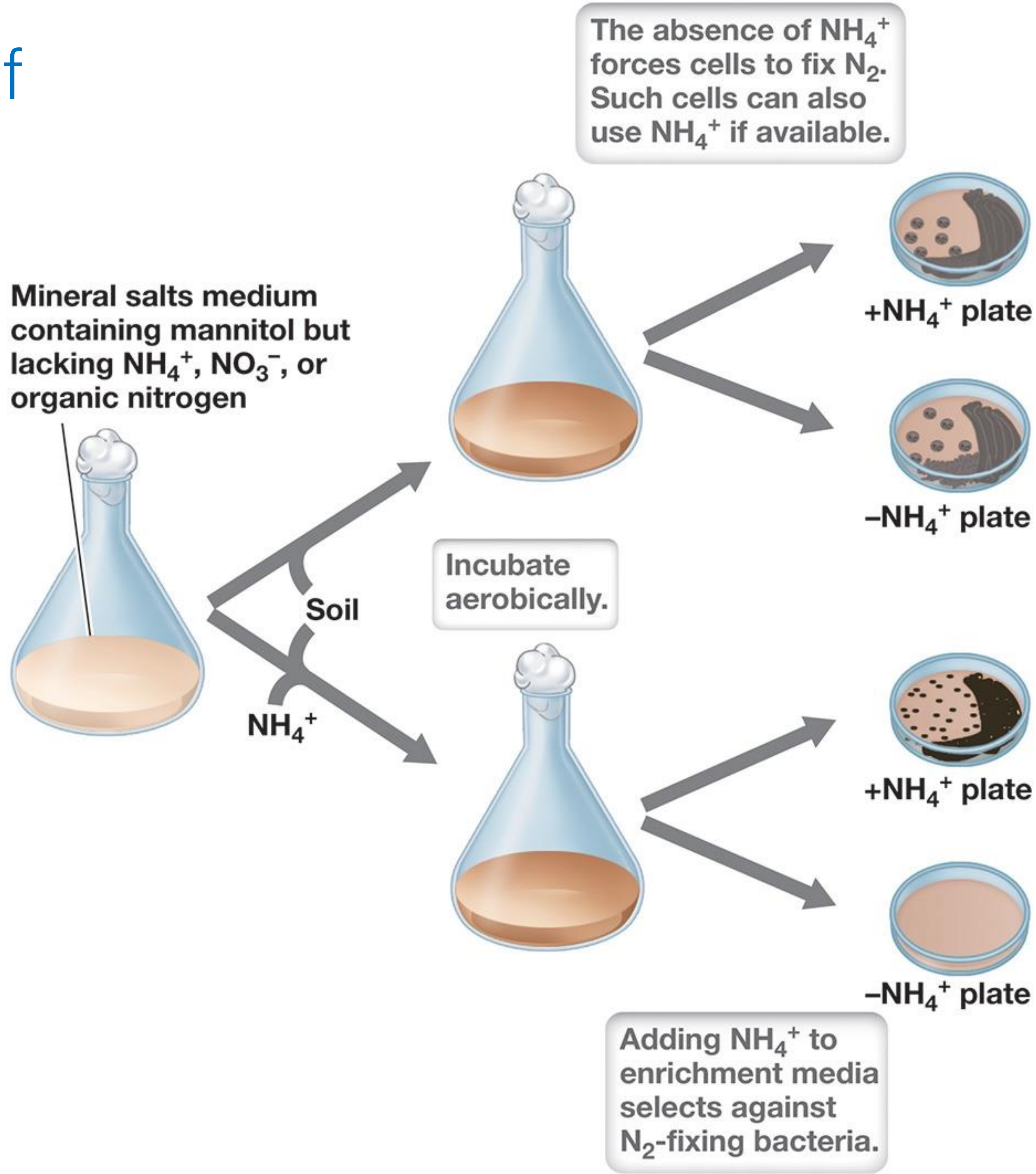
CULTURE DEPENDENT VS. CULTURE INDEPENDENT



ENRICHMENT CULTURE MICROBIOLOGY

- Inoculum
  - The sample from which microorganisms is isolated
  - Enrichment cultures are used to isolate bacteria from diverse and densely populated samples, such as feces or soil
- Isolation
  - The separation of individual populations from the mixed community
- Enrichment cultures
  - Select for desired organisms through manipulation of medium and incubation conditions. This means favoring the growth of target organisms while inhibiting the growth of non-target organisms.

The Isolation of *Azotobacter*



The absence of NH_4^+ forces cells to fix N_2. Such cells can also use NH_4^+ if available.

Mineral salts medium containing mannitol but lacking NH_4^+, NO_3^-, or organic nitrogen

Soil

NH_4^+

Incubate aerobically.

$+NH_4^+$ plate

$-NH_4^+$ plate

$+NH_4^+$ plate

$-NH_4^+$ plate

Adding NH_4^+ to enrichment media selects against N_2-fixing bacteria.

ENRICHMENT CULTURE MICROBIOLOGY

- Enrichment Culture Outcomes
  - Successful enrichment cultures have appropriate resources (nutrients) and conditions (temperature, p H, oxygen, osmotic considerations) that are needed for the target organisms to grow
  - Enrichment cultures can demonstrate the presence of an organism in a habitat
  - They cannot prove that an organism does not inhabit an environment

- Note: The ability to isolate an organism from an environment says nothing about its ecological importance or relative abundance in nature

Some Enrichment Culture Methods for Phototrophic Bacteria (Main C Source - CO_2)

Incubation in air

Incubation condition	Organisms enriched	Inoculum
N_2 as nitrogen source	Cyanobacteria	Pond or lake water; sulfide-rich muds; stagnant water; raw sewage; moist, decomposing leaf litter; moist soil exposed to light
NO_3^- as nitrogen source, 55°C	Thermophilic cyanobacteria	Hot spring microbial mat

Anoxic incubation

Incubation condition	Organisms enriched	Inoculum
H_2 or organic acids; N_2 as sole nitrogen source	Purple nonsulfur bacteria, heliobacteria	Same as above plus hypolimnetic lake water; pasteurized soil (heliobacteria); microbial mats for thermophilic species
H_2S as electron donor	Purple and green sulfur bacteria	
Fe^{2+}, NO_2^- as electron donor	Purple bacteria	

ENRICHMENT CULTURE MICROBIOLOGY

- Enrichment bias
  - Microorganisms cultured in the lab are frequently only minor components of the microbial ecosystem
    - Reason: quantity of nutrients available in the lab culture are typically much higher than in nature
    - Dilution of inoculum may be used to eliminate rapidly growing, but quantitatively insignificant, “weed” species

CLASSICAL PROCEDURES FOR ISOLATING MICROBES

Pure cultures contain a single kind of microorganism

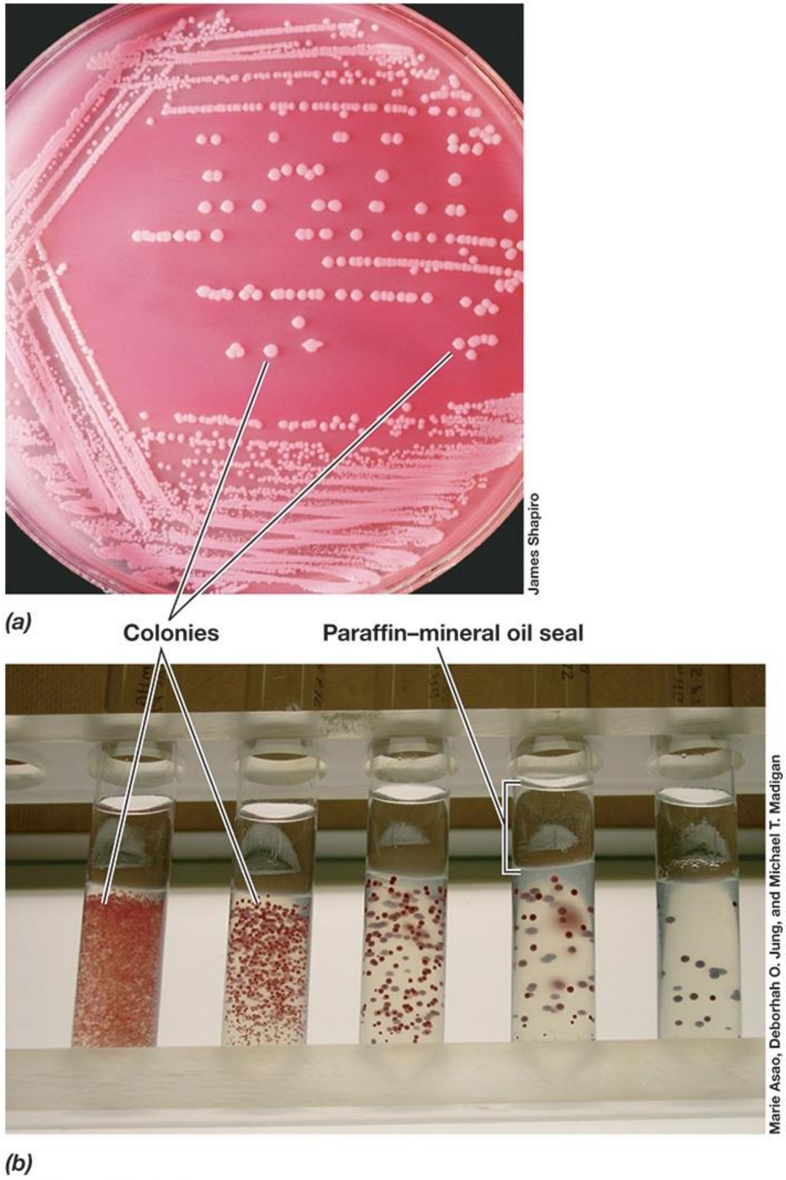
- Can then be used for molecular and physiological experiments

Streak plate

- A well-isolated colony is selected and restreaked several successive times in order to obtain a pure culture

Agar dilution tubes are mixed cultures diluted in molten agar

- Useful for purifying anaerobic organisms



James Shapiro

Marie Asao, Deborhah O. Jung, and Michael T. Madigan

CULTURE-INDEPENDENT ANALYSES OF MICROBIAL COMMUNITIES

Microscopic Analyses

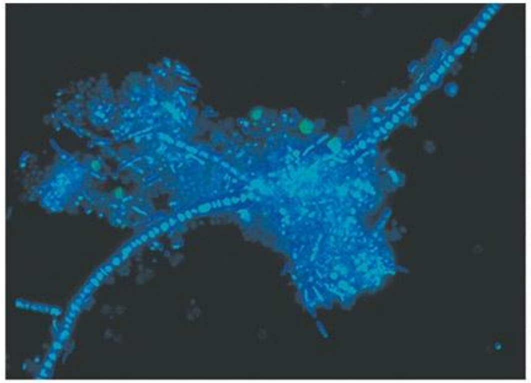
Genetic Analyses

MICROSCOPY ANALYSES OF MICROBIAL COMMUNITIES

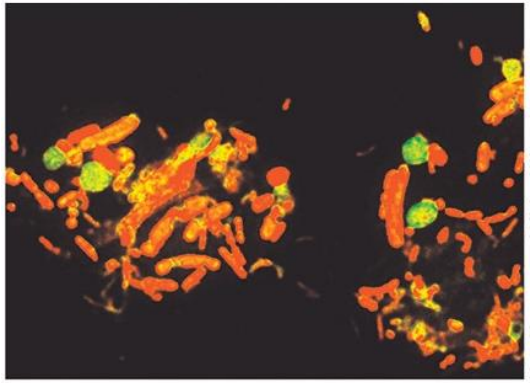
Fluorescent staining

- DAPI-stained cells fluoresce bright blue (DNA)
- AO-stained cells fluoresce orange (RNA) or green (DNA)
- SYBR-stained cells fluoresce green (DNA)

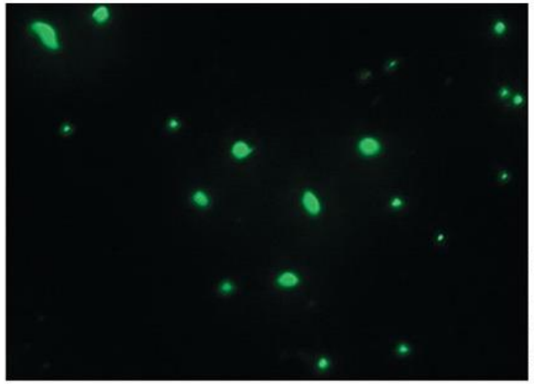
- Fluoresce under UV light
- Enumeration of microorganisms in samples
- Nonspecific and stain nucleic acids, such as DNA
- Stain nonspecifically
  - cannot differentiate between live and dead cells



Marc Mussman and Michael Wagner



Marc Mussman and Michael Wagner



Willm Martins-Habenna

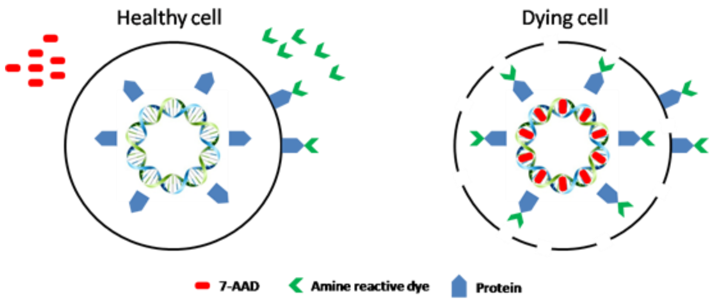
VIABILITY STAINS DIFFERENTIATE BETWEEN LIVE AND DEAD CELLS

Two dyes are used

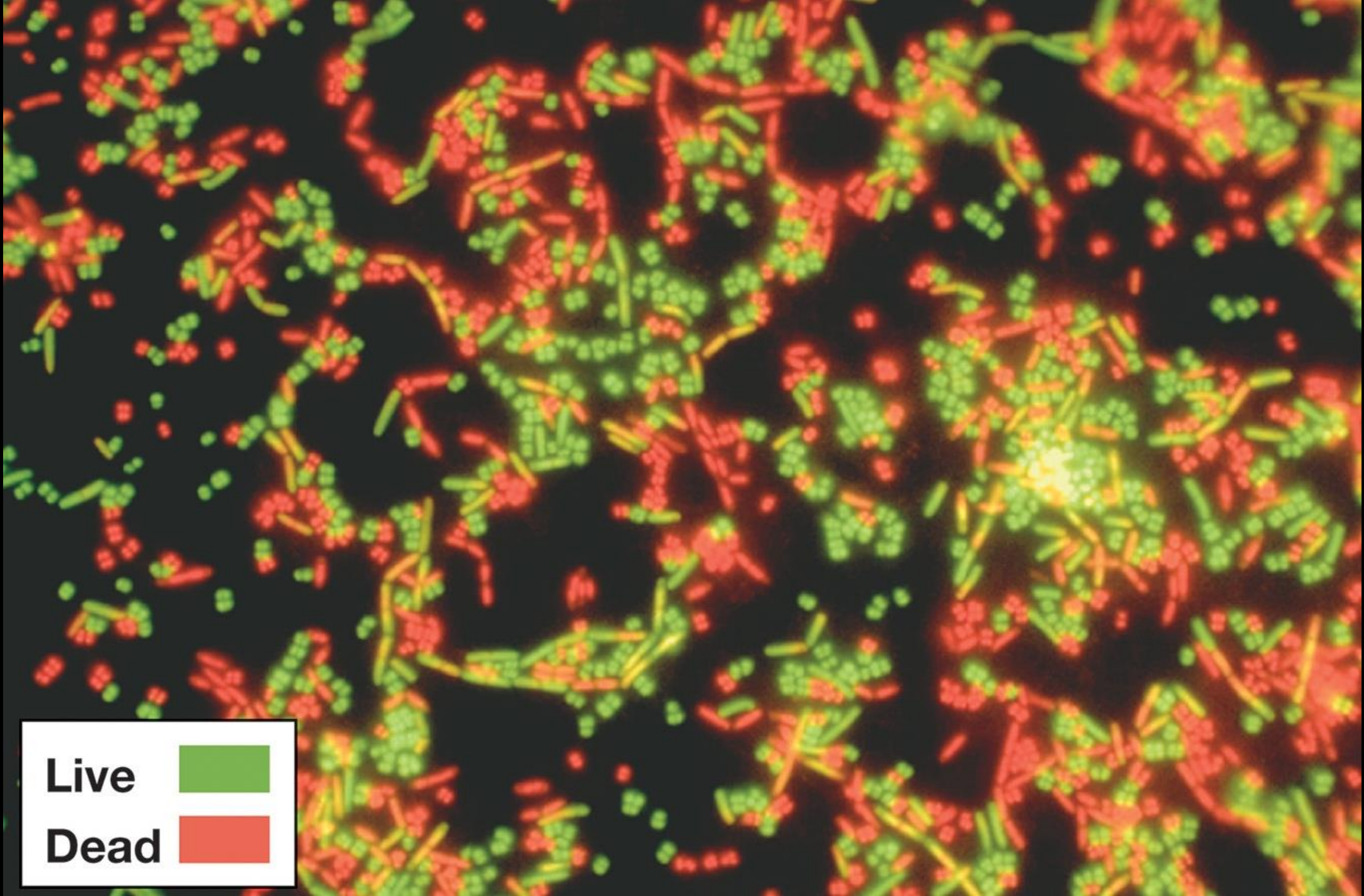
- one that stains live cells and the other that stains dead cells, based on integrity of cell membrane

Green cells are live

Red cells are dead



This method may stain background nonspecifically with environmental samples


Live 🟩
Dead 🟥

CULTURE-INDEPENDENT ANALYSES OF MICROBIAL COMMUNITIES

Microscopic Analyses

Genetic Analyses

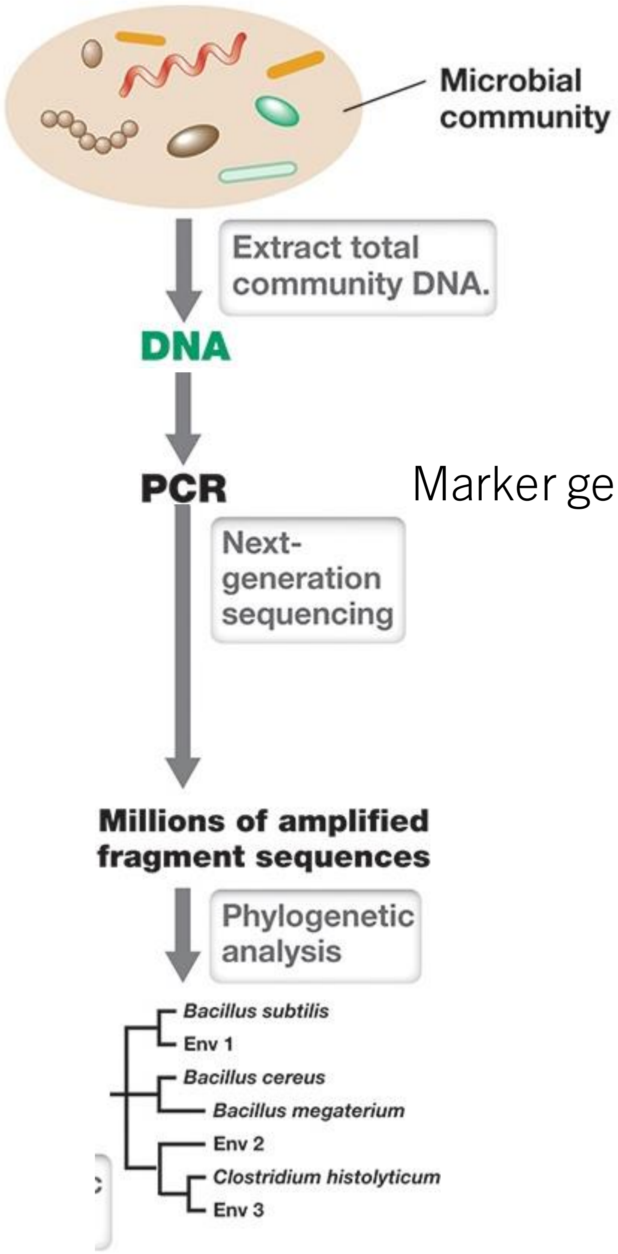
Culture-Independent Genetic Analyses of Microbial Communities

- PCR Methods of Microbial Community Analysis
- Environmental Multi-omics

PCR Methods of Microbial Community Analysis

Specific genes can be used as a measure of diversity: **marker genes**

- 16S – bacteria and archaea
- ITS – fungi
- Functional genes – antibiotic production, biogeochemical cycles, etc.

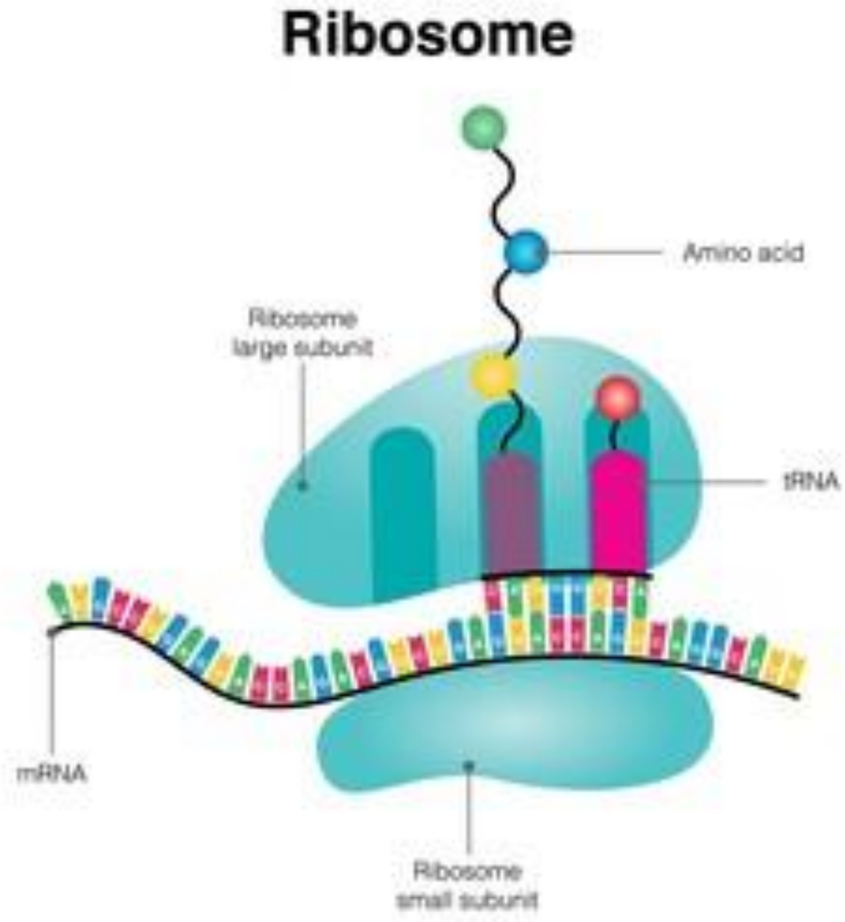


Marker gene

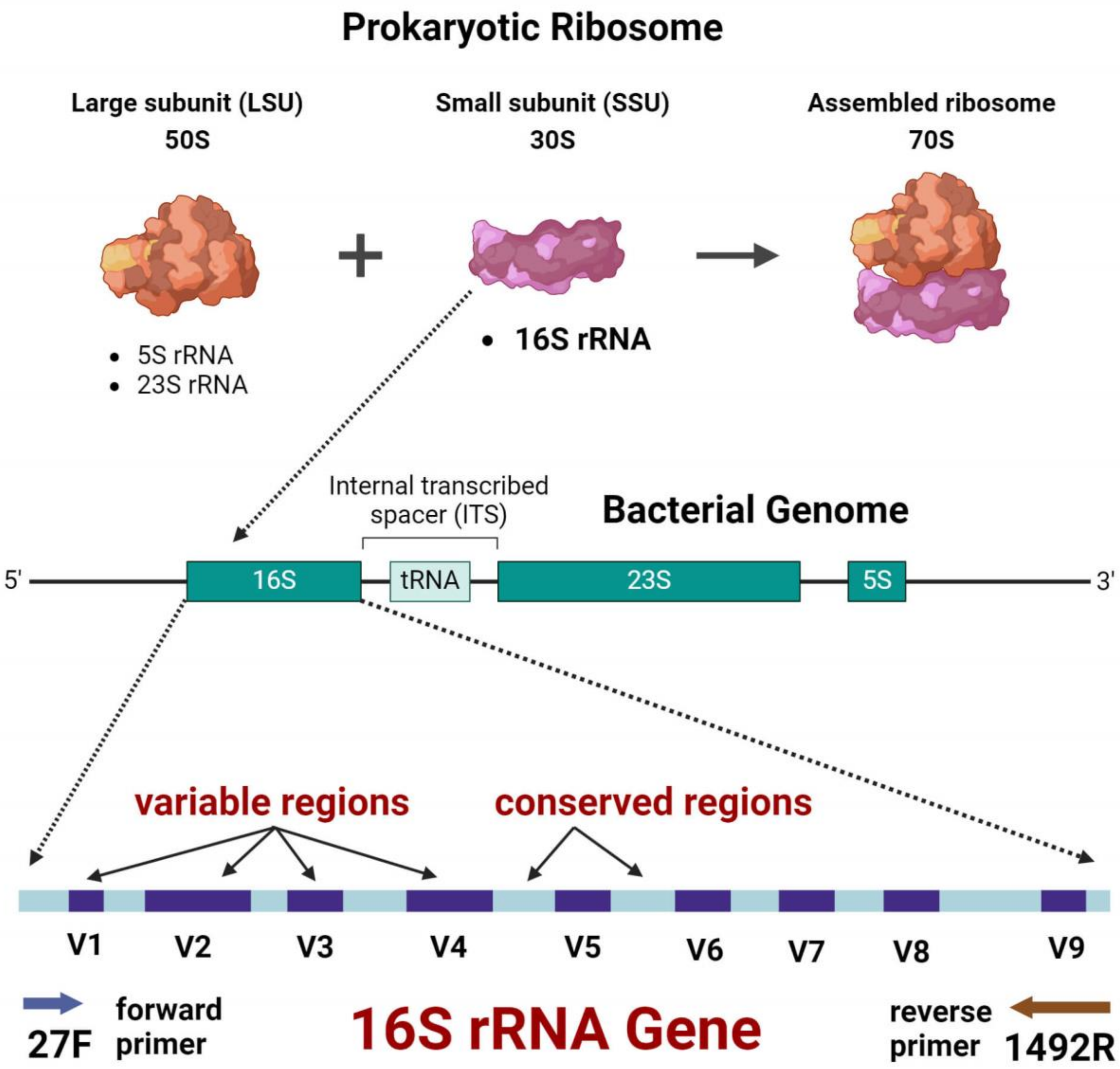
Ribosomal RNA

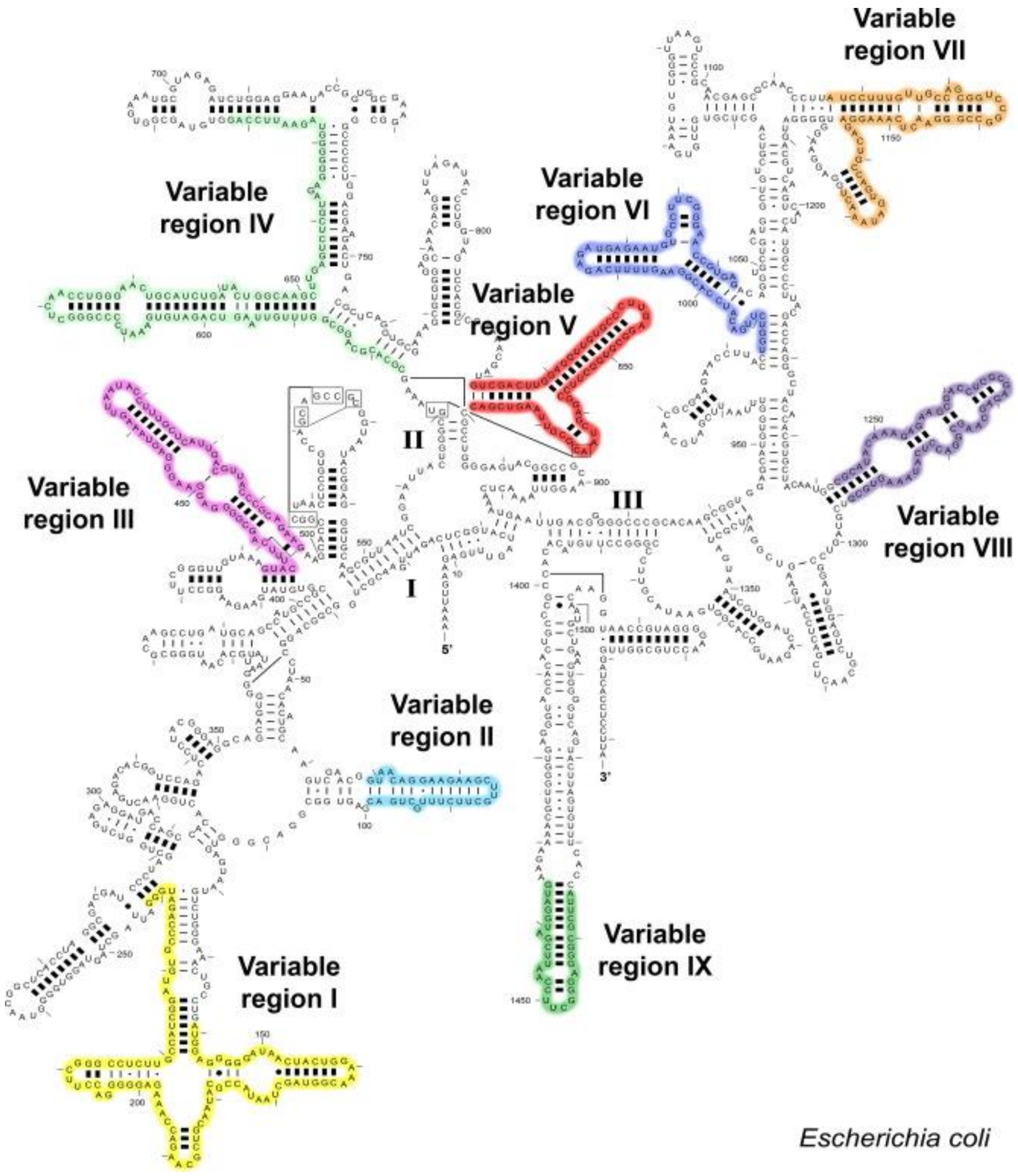
Ribosomal RNA (rRNA) – RNA molecules that associate with ribosomal proteins to form the ribosomes

Ribosomes are responsible for protein synthesis in the cell

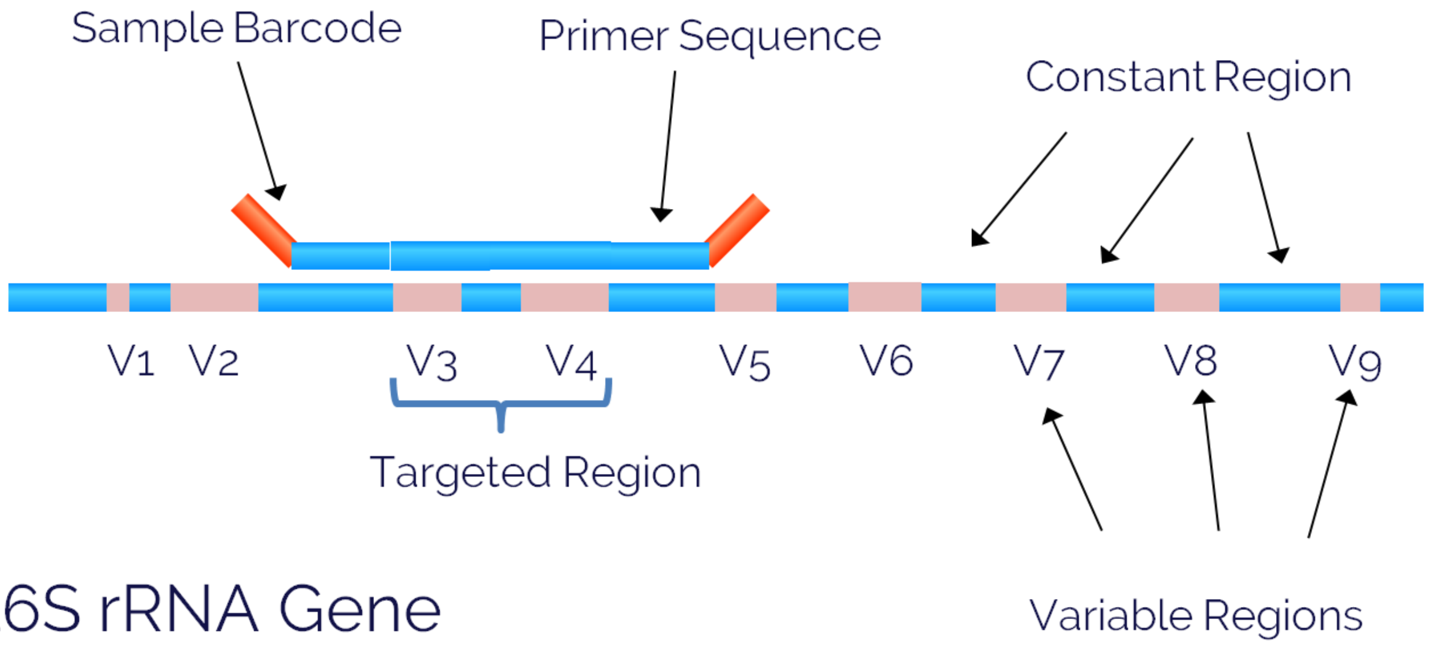


What is 16S rRNA gene?



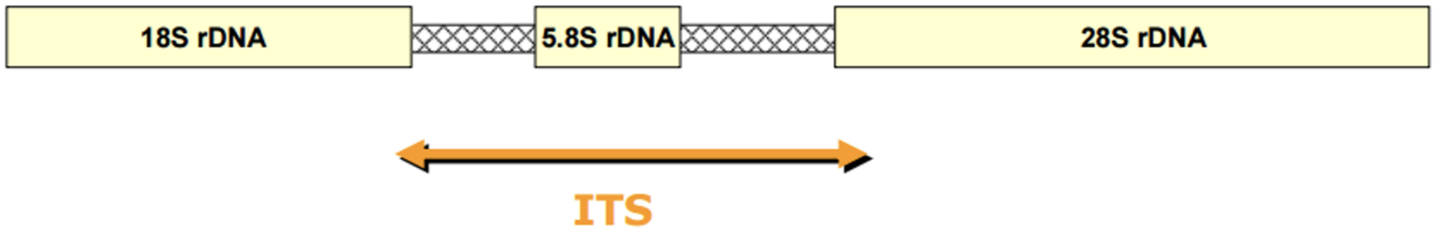


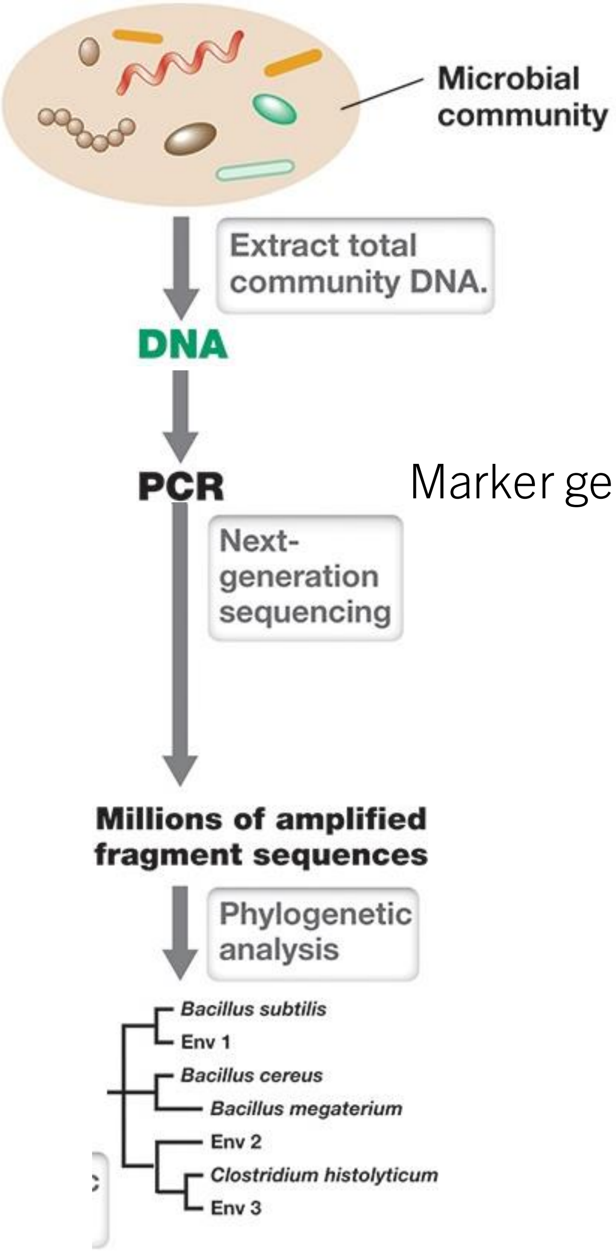
Marker gene for bacteria and archaea


Sample Barcode
Primer Sequence
Constant Region
V1
V2
V3
V4
V5
V6
V7
V8
V9
Targeted Region
16S rRNA Gene
Variable Regions

PCR Methods of Microbial Community Analysis

- ITS – fungi





Microbial community

Extract total community DNA.

DNA

PCR

Next-generation sequencing

Millions of amplified fragment sequences

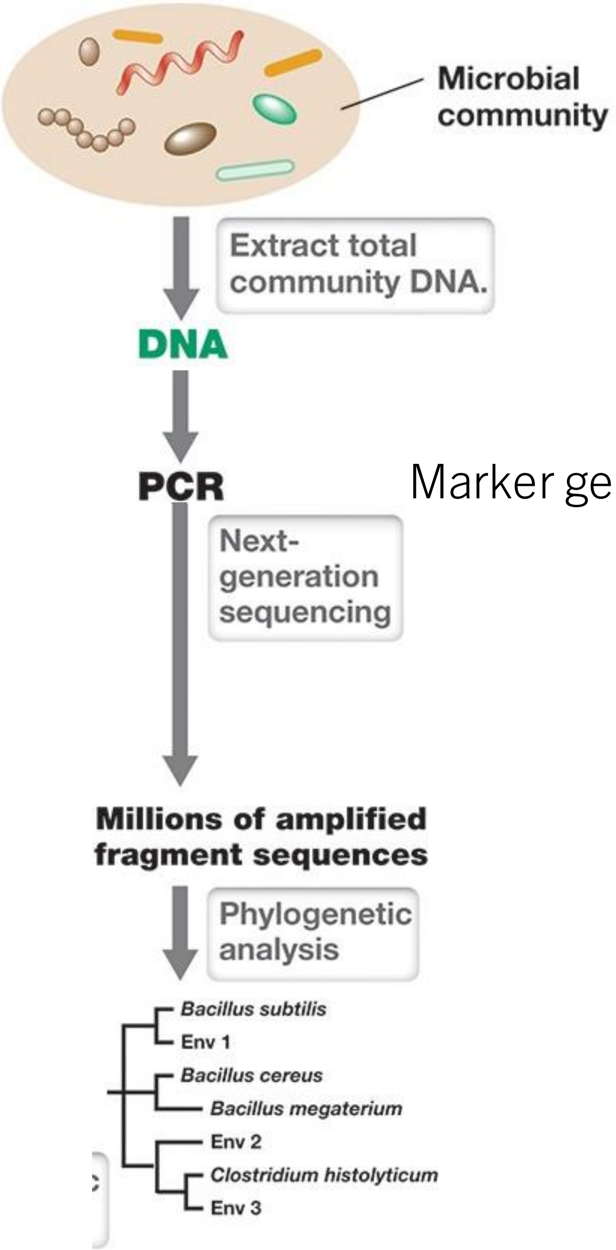
Phylogenetic analysis

Marker gene

PCR Methods of Microbial Community Analysis

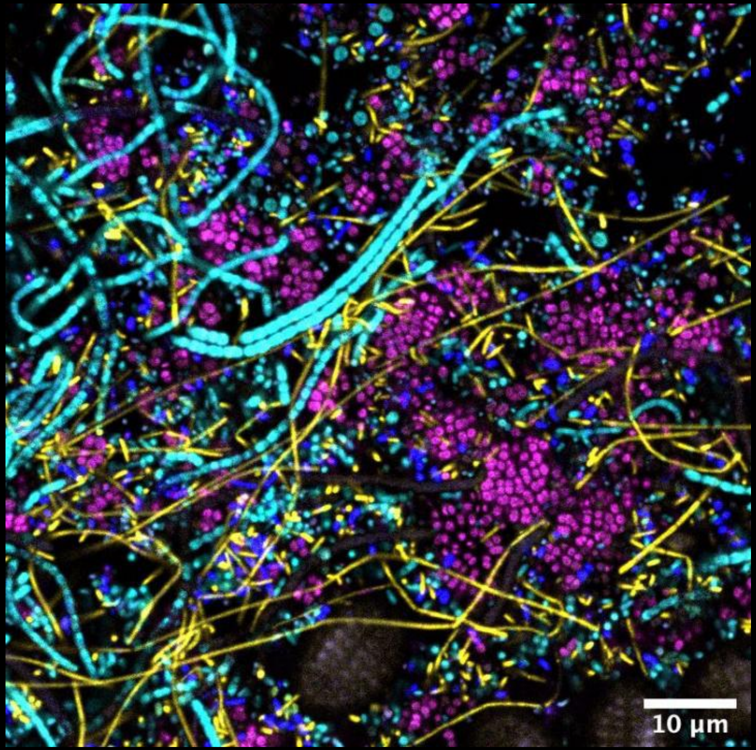
Specific genes can be used as a measure of diversity: **marker genes**

- 16S – bacteria and archaea
- ITS – fungi
- Functional genes – antibiotic production, biogeochemical cycles, etc.



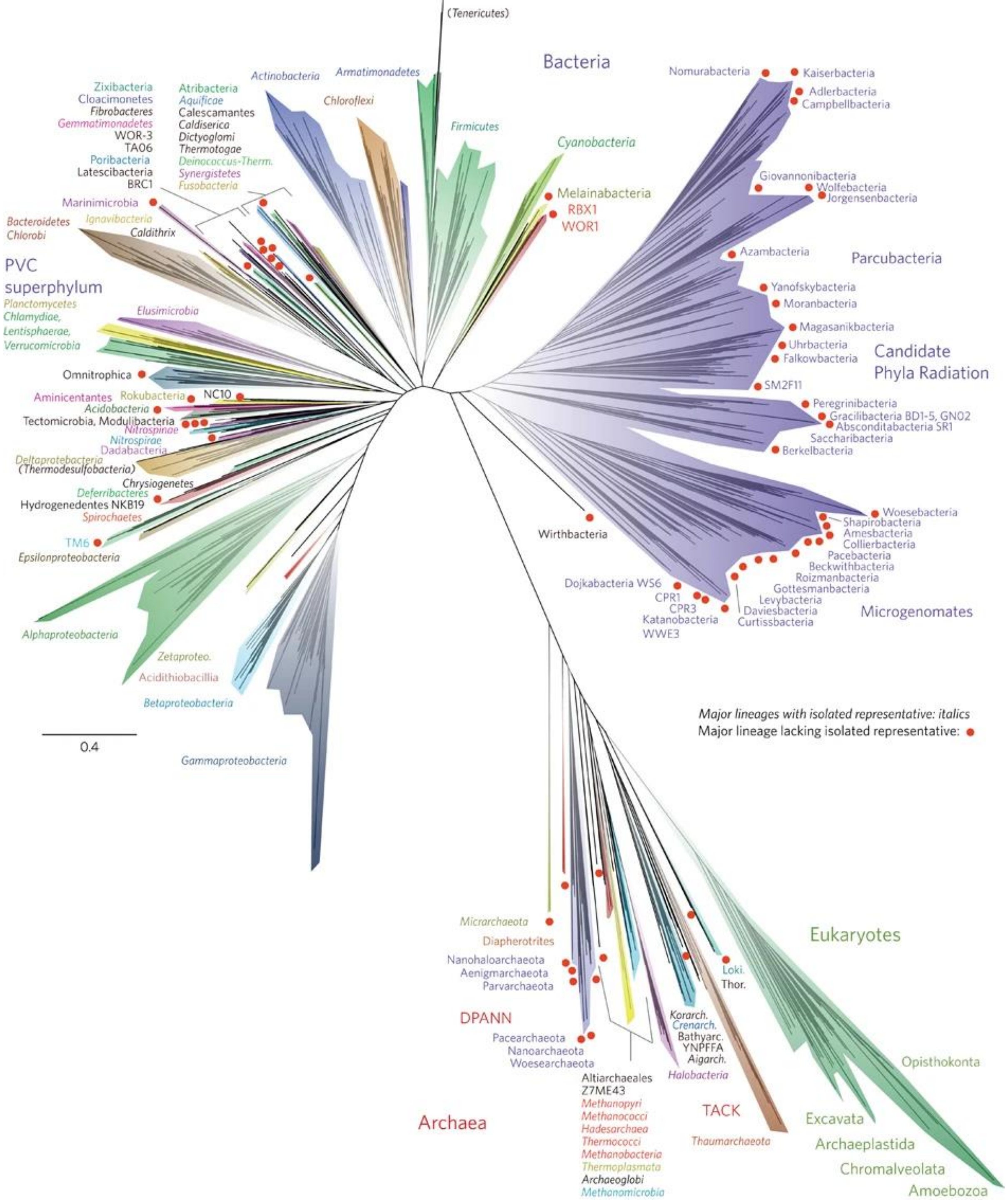
Marker gene

rRNA sequences in environmental samples differ from those of all known laboratory cultures



- fewer than 0.1% of bacteria have been cultured
- enrichment bias is a real problem to culture-based methods

A new view of the tree of life



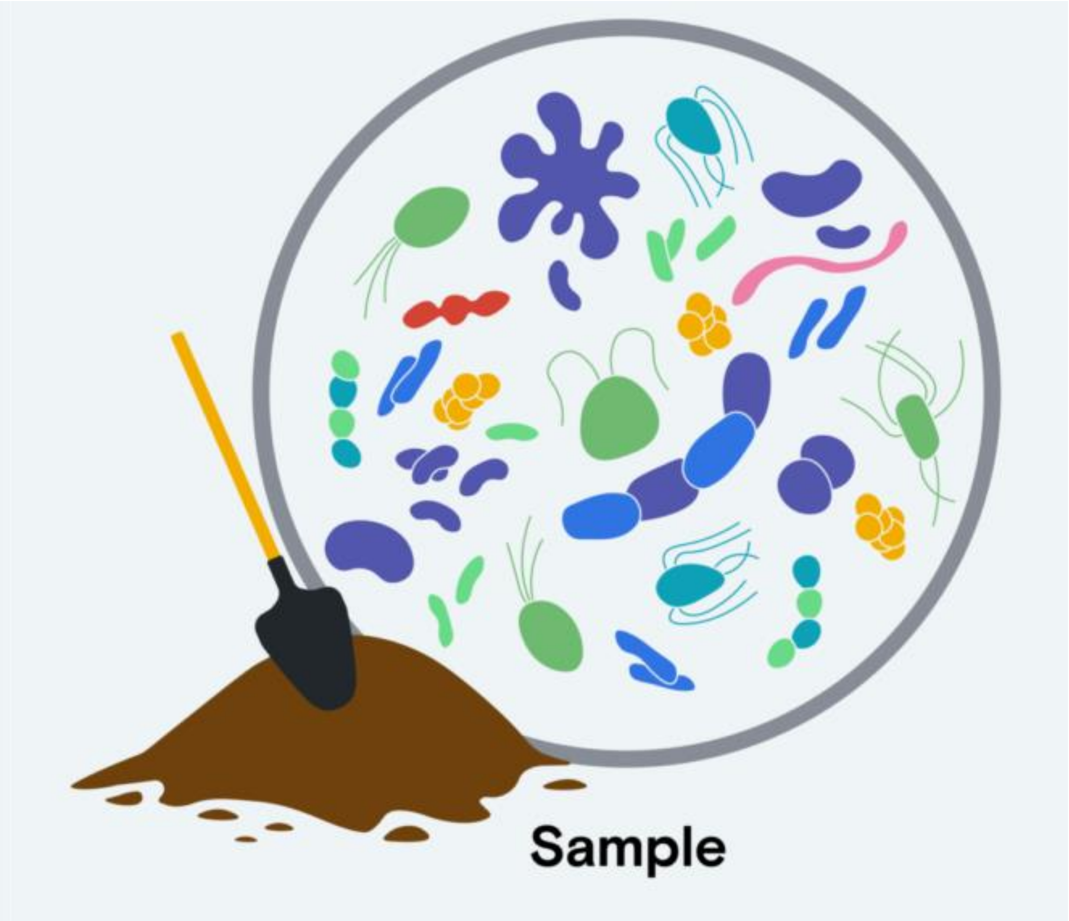
Hug et al. 2016

Metagenomics

- Metagenomics – Environmental genomics
- The study of the structure and function of entire nucleotide sequences isolated and analyzed from all the organisms (typically microbes) in a bulk sample.

  -structure of the microbial community (who is there)

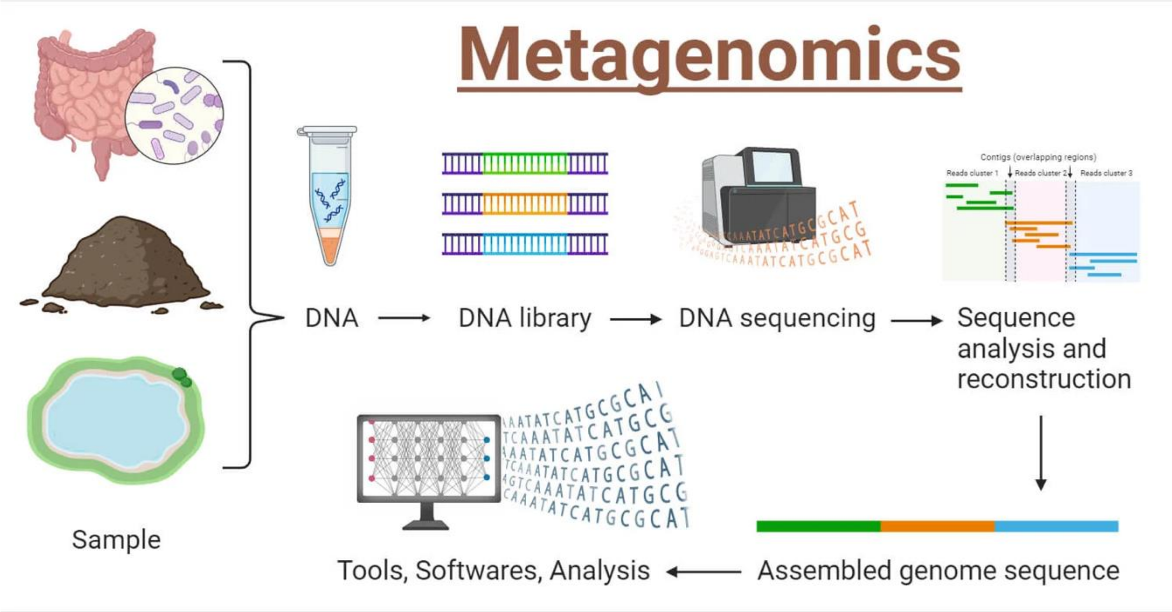
  -function of the microbial community (what can they do)



Metagenome: total gene content of microbial community

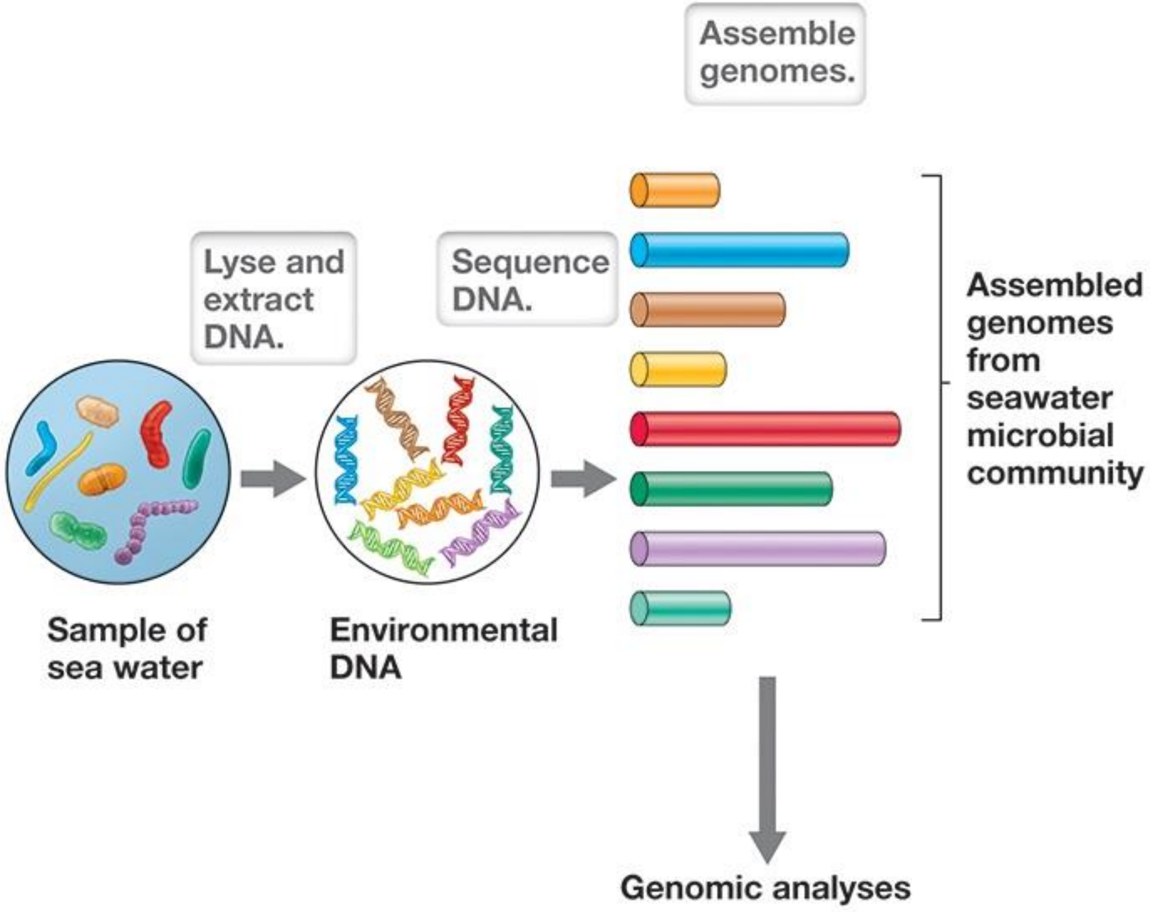
Metagenomics

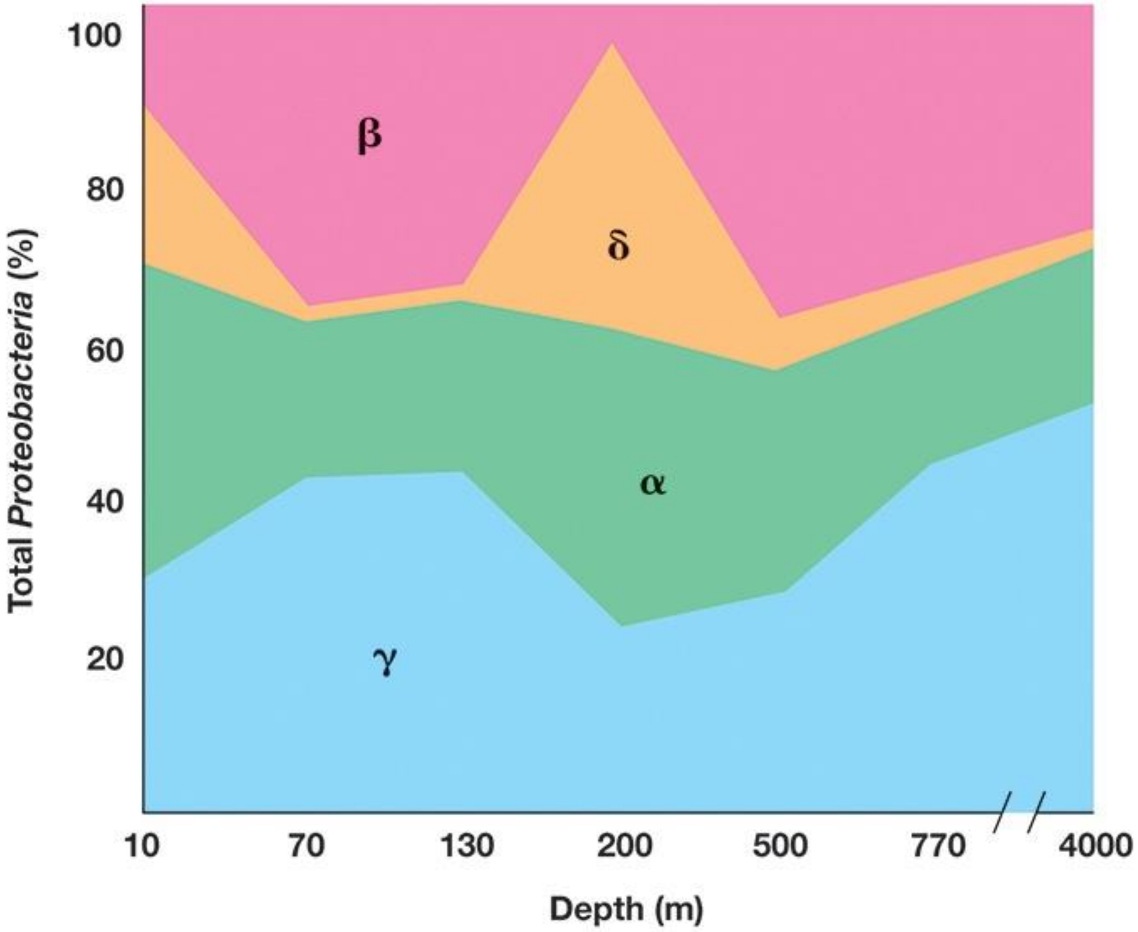
- Metagenomics – Environmental genomics
- DNA is isolated from the entire microbial community and sequenced
- Detects as many genes as possible
- Yields picture of gene pool in environment
- Can detect genes that are not amplified by current PCR primers
- Powerful tool for assessing the phylogenetic and metabolic diversity of an environment



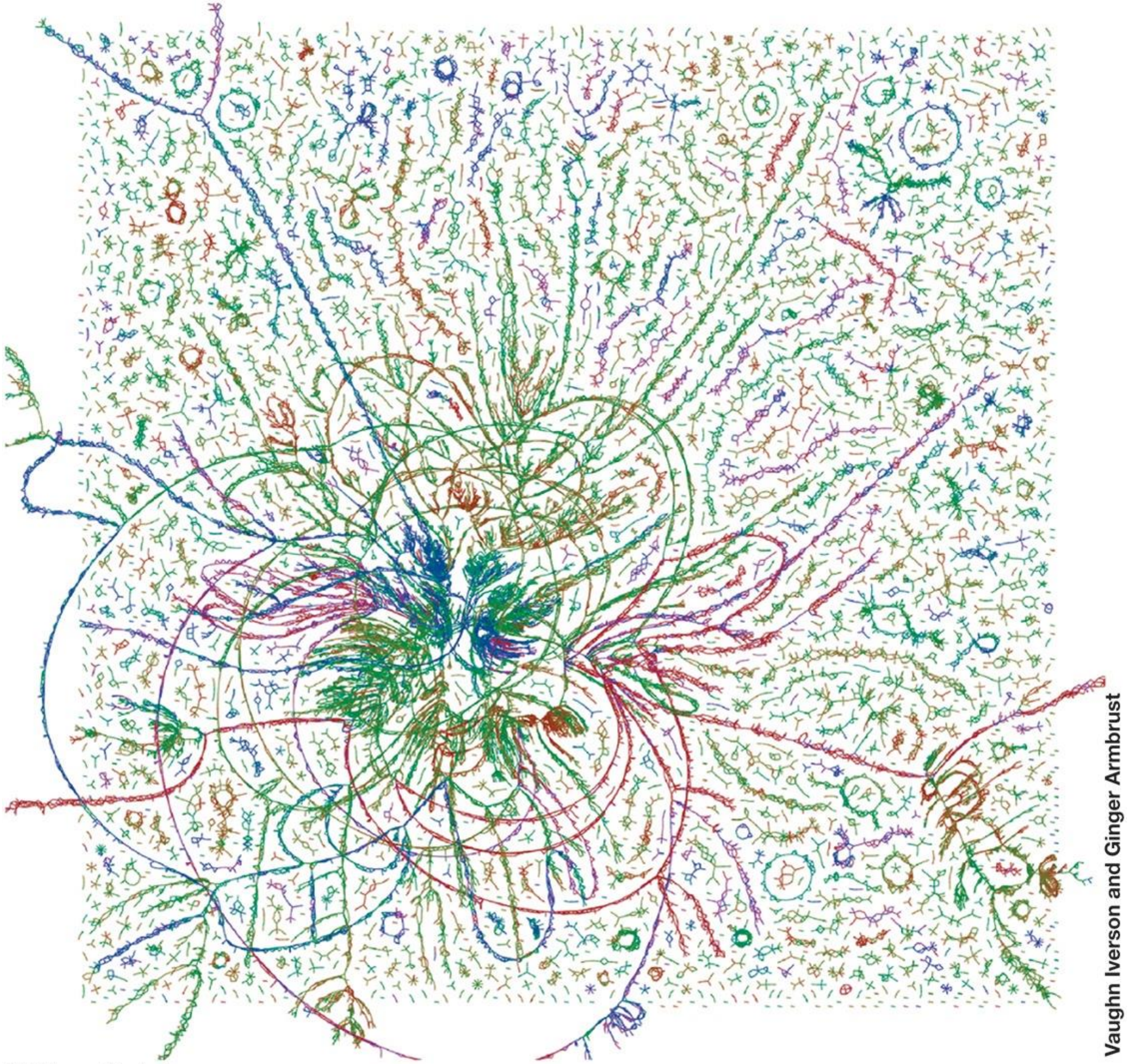
Metagenomics and the Microbiome

Metagenomics and the Microbiome



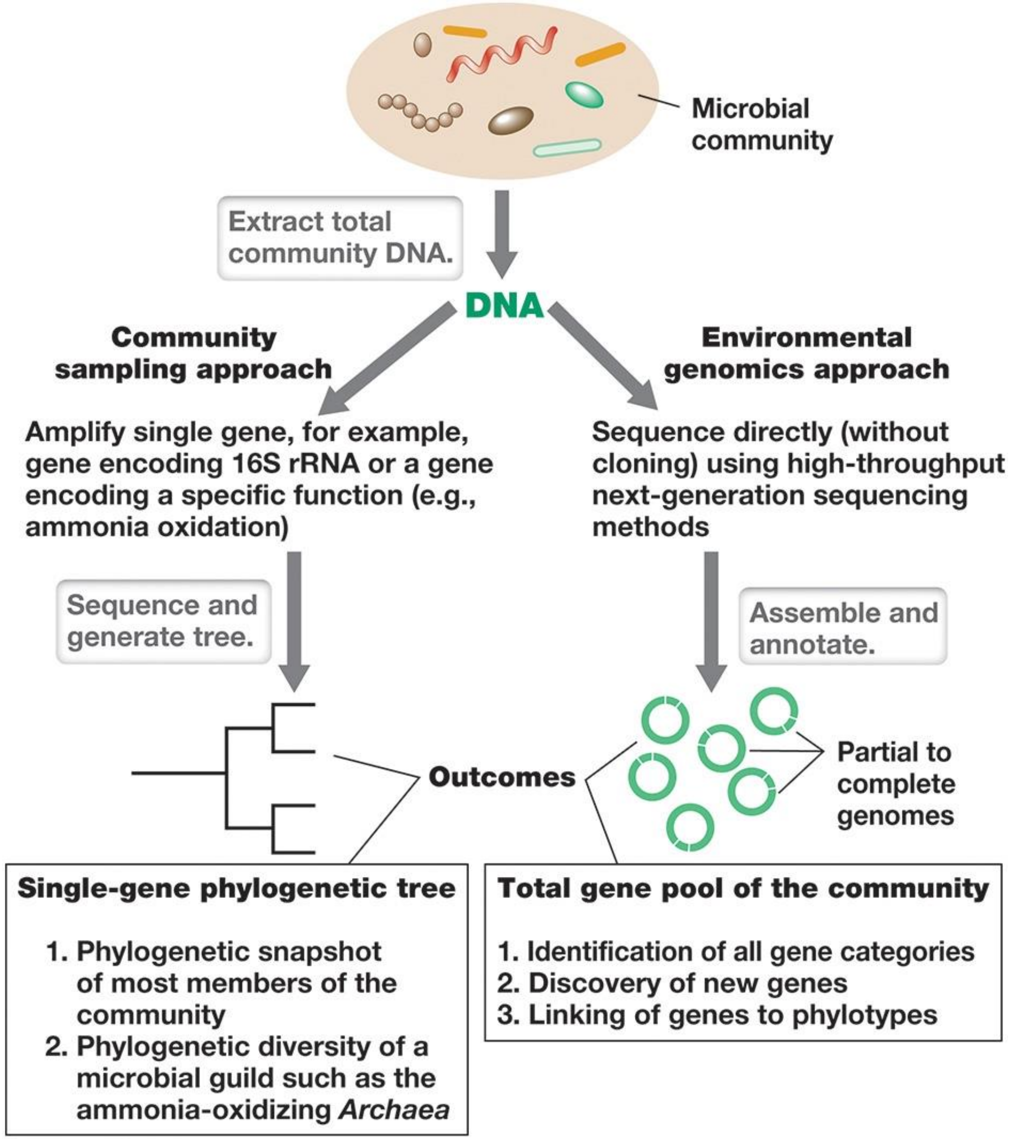


Genome Assembly From a Coastal Marine Metagenome Consisting of 58.5 Billion Nucleotides of Sequence



Vaughn Iverson and Ginger Armbrust

Single-Gene Versus Environmental Genomic Approaches to Microbial Community Analysis



Microbial community

Extract total community DNA.

DNA

Community sampling approach

Amplify single gene, for example, gene encoding 16S rRNA or a gene encoding a specific function (e.g., ammonia oxidation)

Sequence and generate tree.

Environmental genomics approach

Sequence directly (without cloning) using high-throughput next-generation sequencing methods

Assemble and annotate.

Partial to complete genomes

Outcomes

Single-gene phylogenetic tree

1. Phylogenetic snapshot of most members of the community
2. Phylogenetic diversity of a microbial guild such as the ammonia-oxidizing *Archaea*

Total gene pool of the community

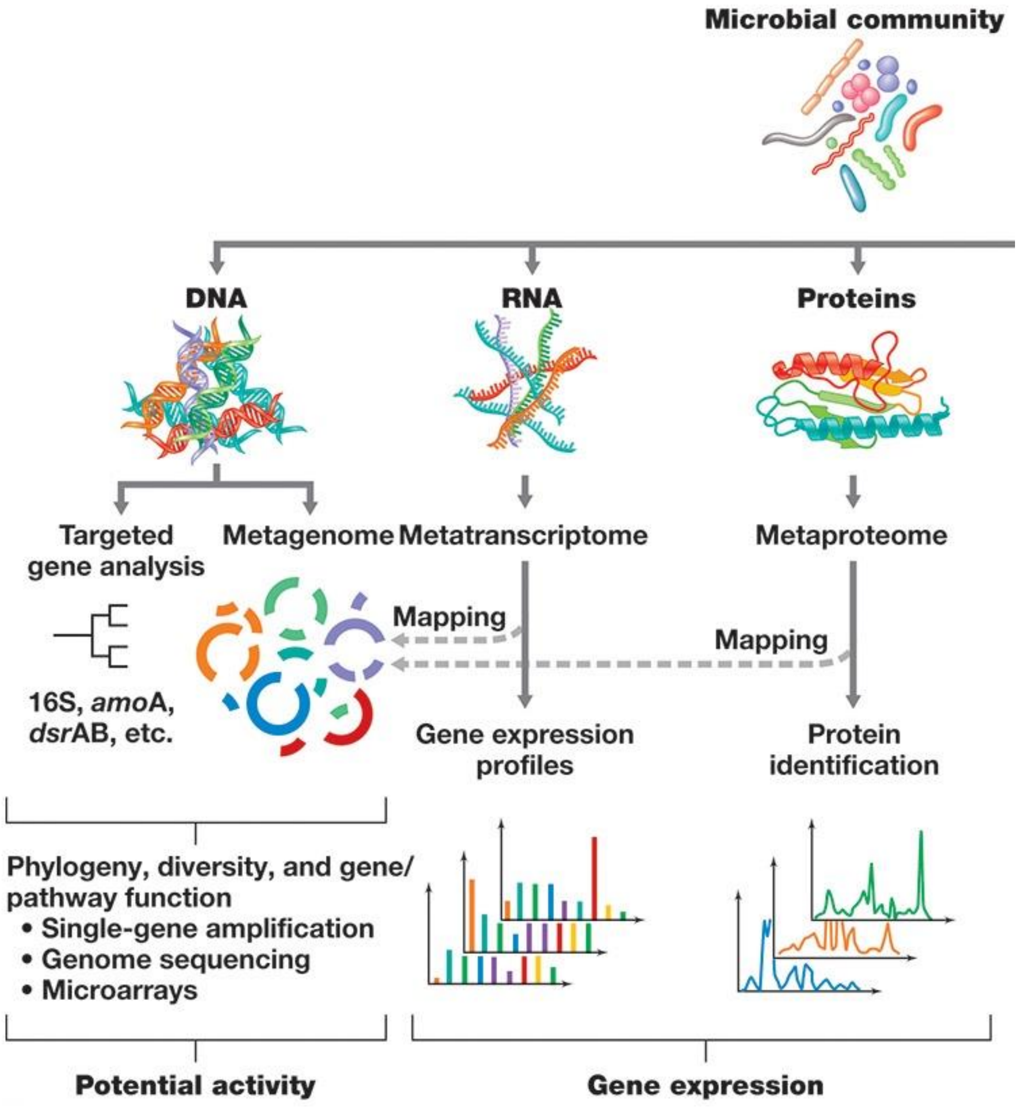
1. Identification of all gene categories
2. Discovery of new genes
3. Linking of genes to phylotypes

Metatranscriptomics and Metaproteomics

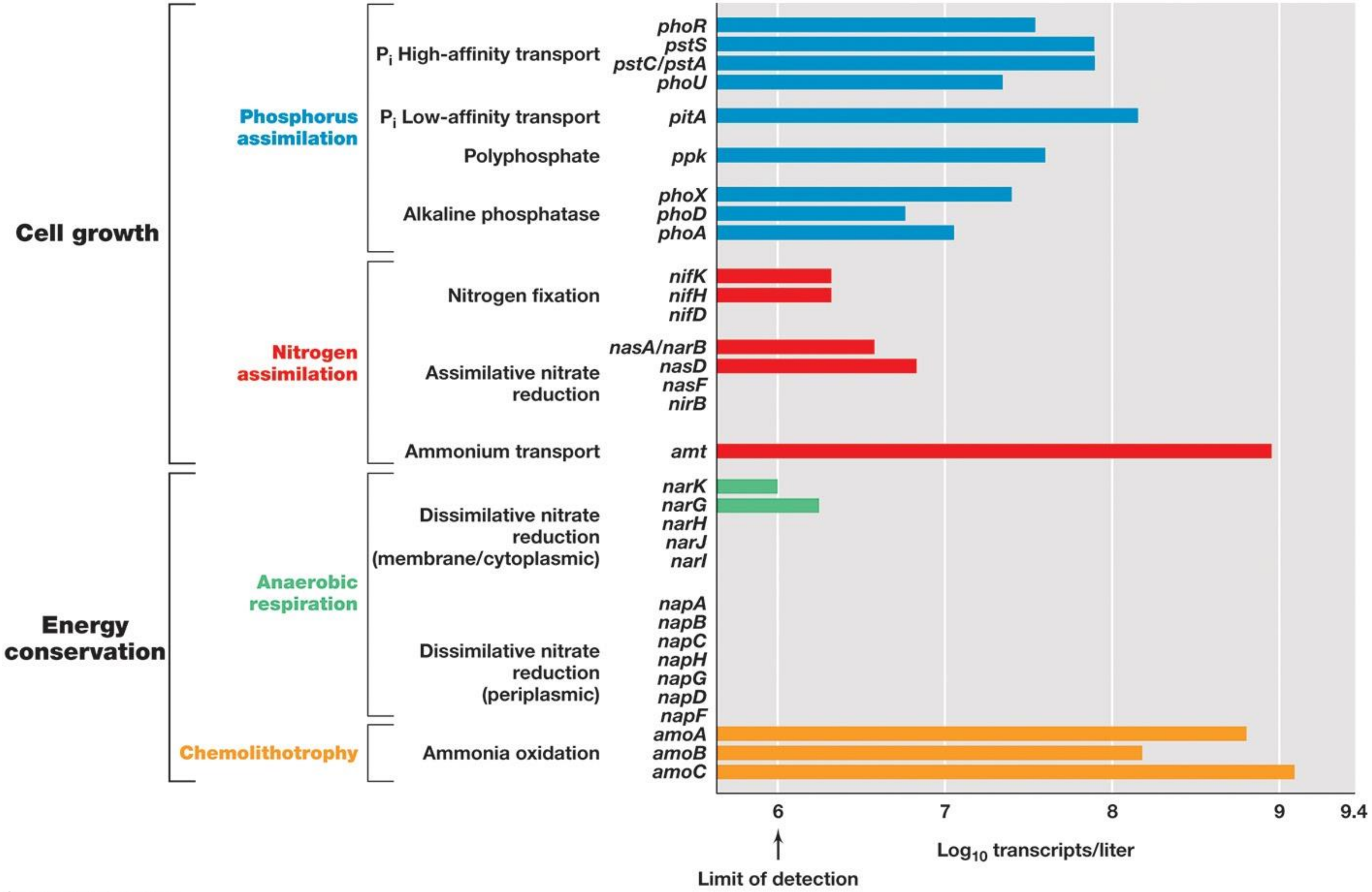
Functionaly active portion of the community

- Analyzes community RNA/proteins
  - (who are the active members)

- Reveals genes/proteins in a community that are actually expressed
  - (what they are actually doing)

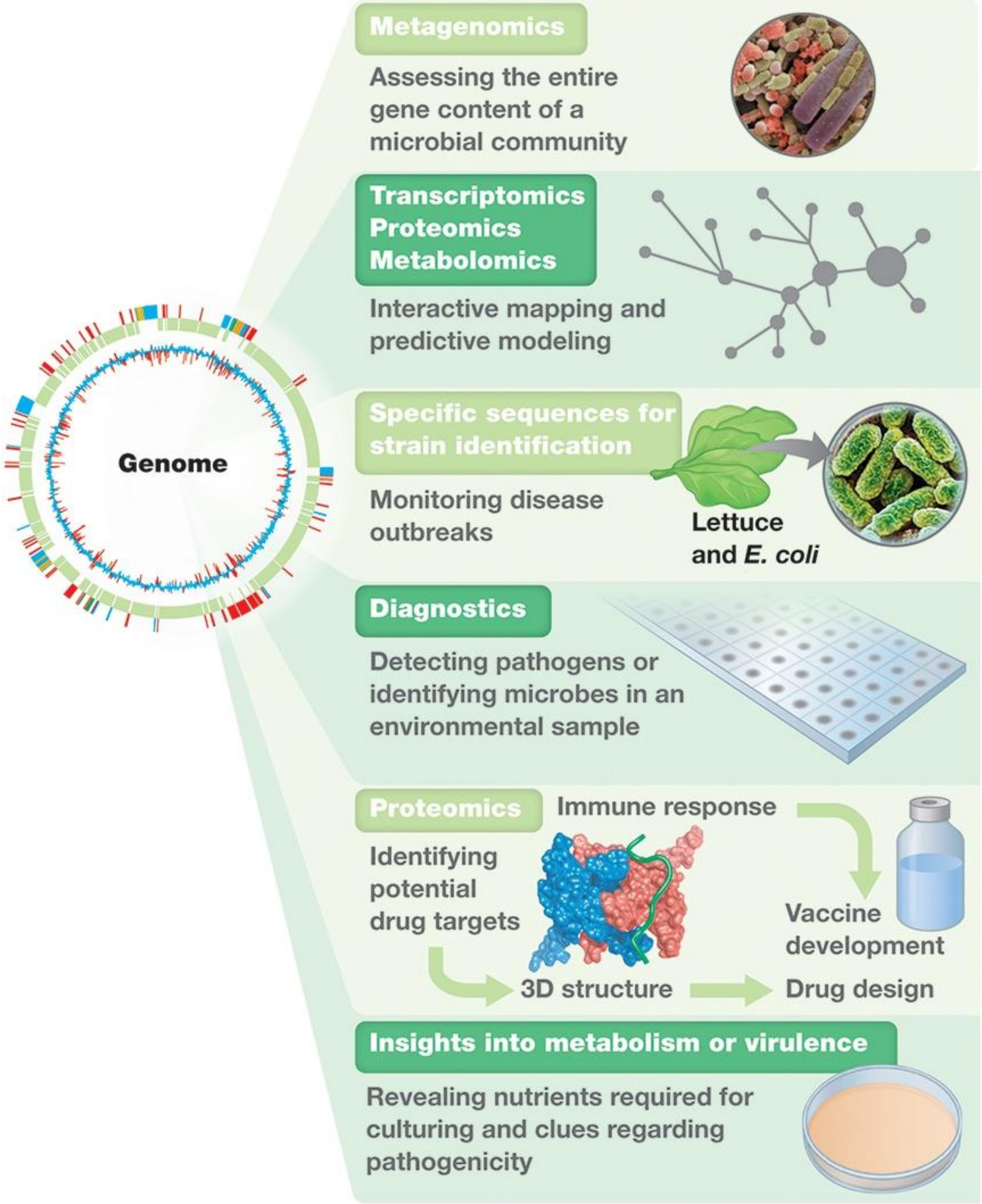


Metatranscriptomic Analysis of Coastal Marine Surface Waters



INTRODUCTION TO GENOMICS

- “Omics”: Broad discipline integrating different methodologies to characterize and quantify large pools of biomolecules
- **Genome:** entire complement of genetic information includes genes encoding proteins, RNA, and regulatory sequences, and noncoding DNA
- Major omic themes: genomics, transcriptomics, proteomics, metabolomics, metagenomics
- **Genomics:** discipline of mapping, sequencing, analyzing, and comparing genomes



Genome

Metagenomics

Assessing the entire gene content of a microbial community

Transcriptomics
Proteomics
Metabolomics

Interactive mapping and predictive modeling

Specific sequences for strain identification

Monitoring disease outbreaks

Lettuce and *E. coli*

Diagnostics

Detecting pathogens or identifying microbes in an environmental sample

Proteomics

Immune response

Identifying potential drug targets

3D structure

Vaccine development

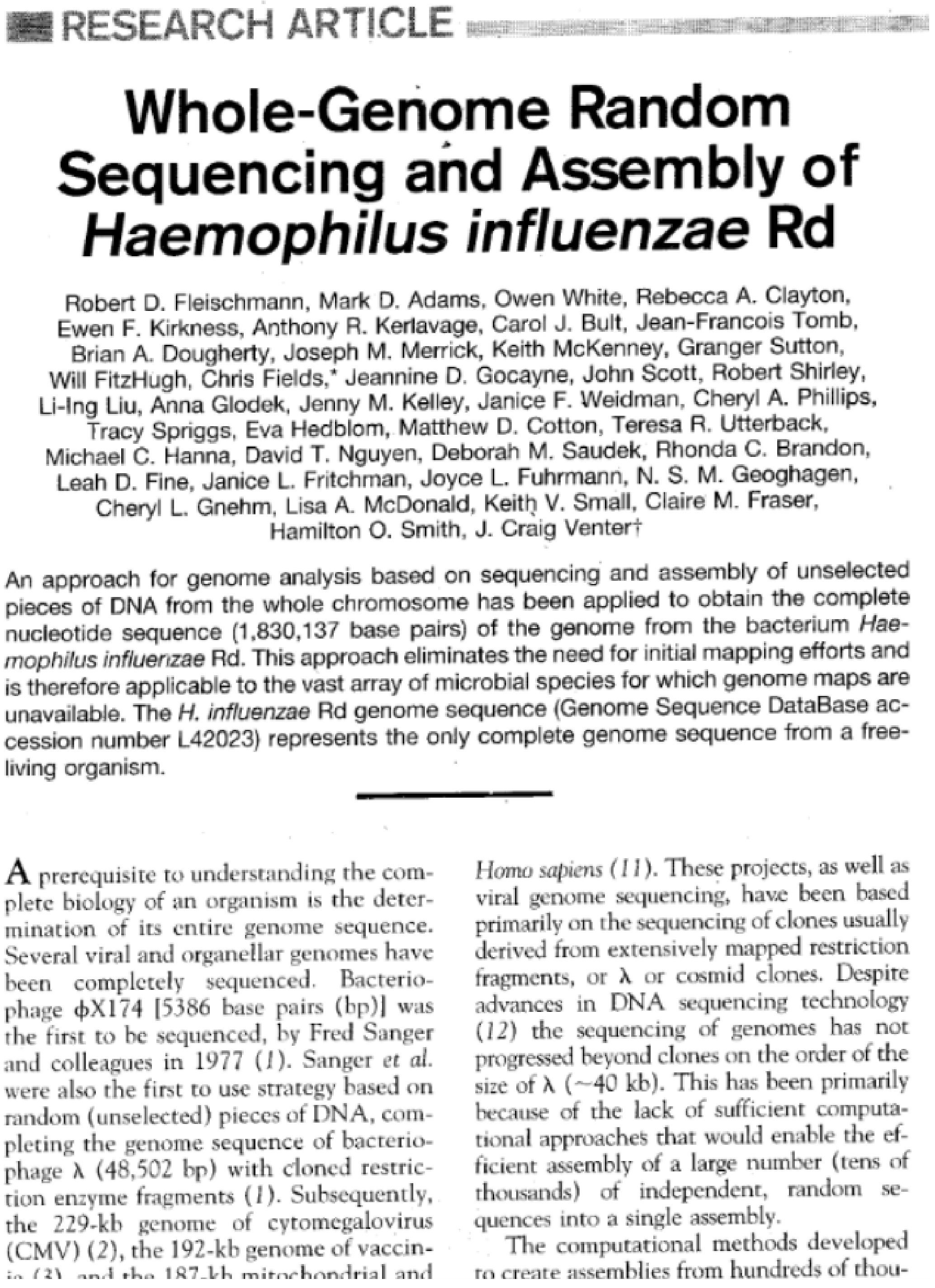
Drug design

Insights into metabolism or virulence

Revealing nutrients required for culturing and clues regarding pathogenicity

INTRODUCTION TO GENOMICS

- Advances rely on improvements in molecular technologies and computing power
- Number of sequenced genomes has grown rapidly
- Bottleneck is vast amounts of nucleic acid sequence data
- Genomics: Then and Now
  - First genomes sequenced were small viruses
  - First bacterial genome published in 1995
  - Today, DNA sequences from 125,000+ *Bacteria*, *Archaea*, and viruses are available publicly
  - Many eukaryotic genomes sequenced, including human genome



RESEARCH ARTICLE

Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

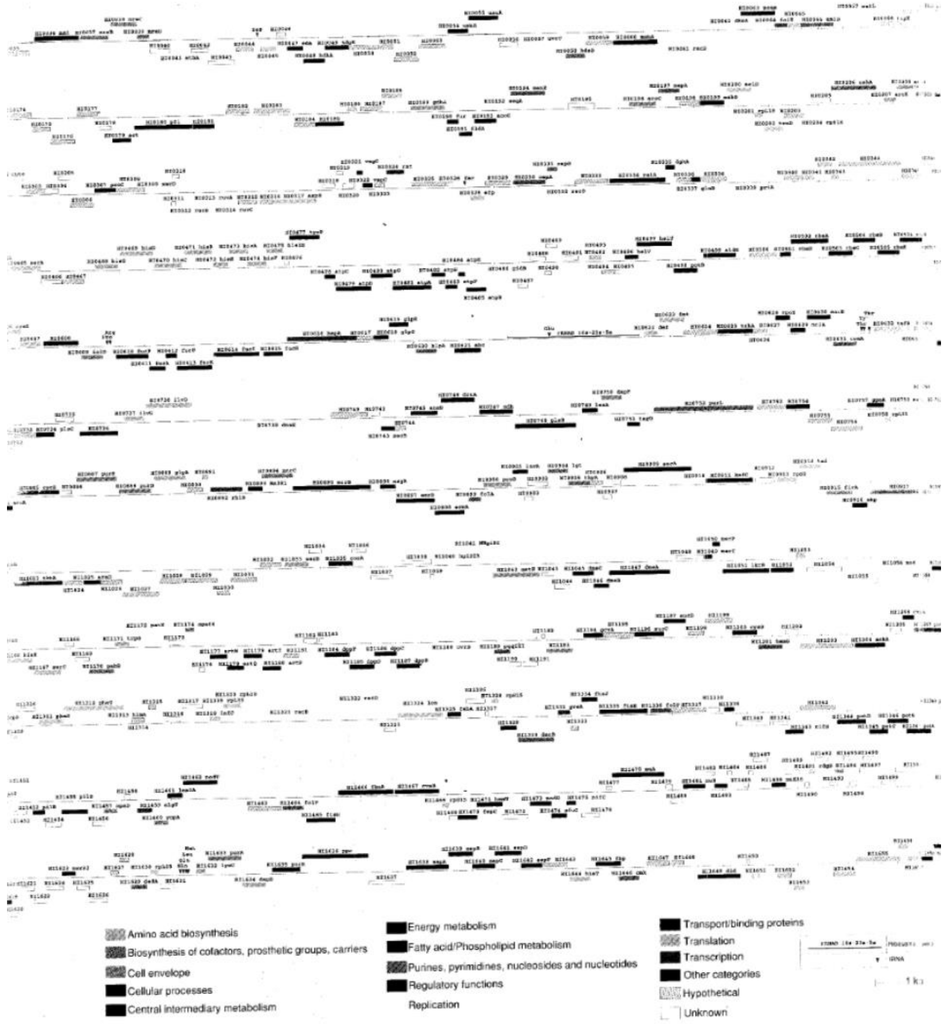
Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghagen, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium *Haemophilus influenzae* Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The *H. influenzae* Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.

A prerequisite to understanding the complete biology of an organism is the determination of its entire genome sequence. Several viral and organellar genomes have been completely sequenced. Bacteriophage ϕX174 [5386 base pairs (bp)] was the first to be sequenced, by Fred Sanger and colleagues in 1977 (*1*). Sanger *et al.* were also the first to use strategy based on random (unselected) pieces of DNA, completing the genome sequence of bacteriophage λ (48,502 bp) with cloned restriction enzyme fragments (*1*). Subsequently, the 229-kb genome of cytomegalovirus (CMV) (*2*), the 192-kb genome of vaccinia (*3*), and the 187-kb mitochondrial and

Homo sapiens (*11*). These projects, as well as viral genome sequencing, have been based primarily on the sequencing of clones usually derived from extensively mapped restriction fragments, or λ or cosmid clones. Despite advances in DNA sequencing technology (*12*) the sequencing of genomes has not progressed beyond clones on the order of the size of λ (~40 kb). This has been primarily because of the lack of sufficient computational approaches that would enable the efficient assembly of a large number (tens of thousands) of independent, random sequences into a single assembly.

The computational methods developed to create assemblies from hundreds of thou-



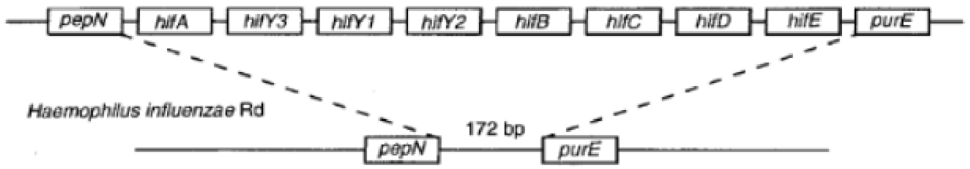


Fig. 3. A comparison of the region of the *H. influenzae* chromosome containing the eight genes of the fimbrial gene cluster present in *H. influenzae* type b and the same region in *H. influenzae* Rd. The region is flanked by *pepN* and *purE* in both organisms. However, in the noninfectious Rd strain the eight genes of the fimbrial gene cluster have been excised. A 172-bp spacer region is located in this region in the Rd strain and continues to be flanked by the *pepN* and *purE* genes.

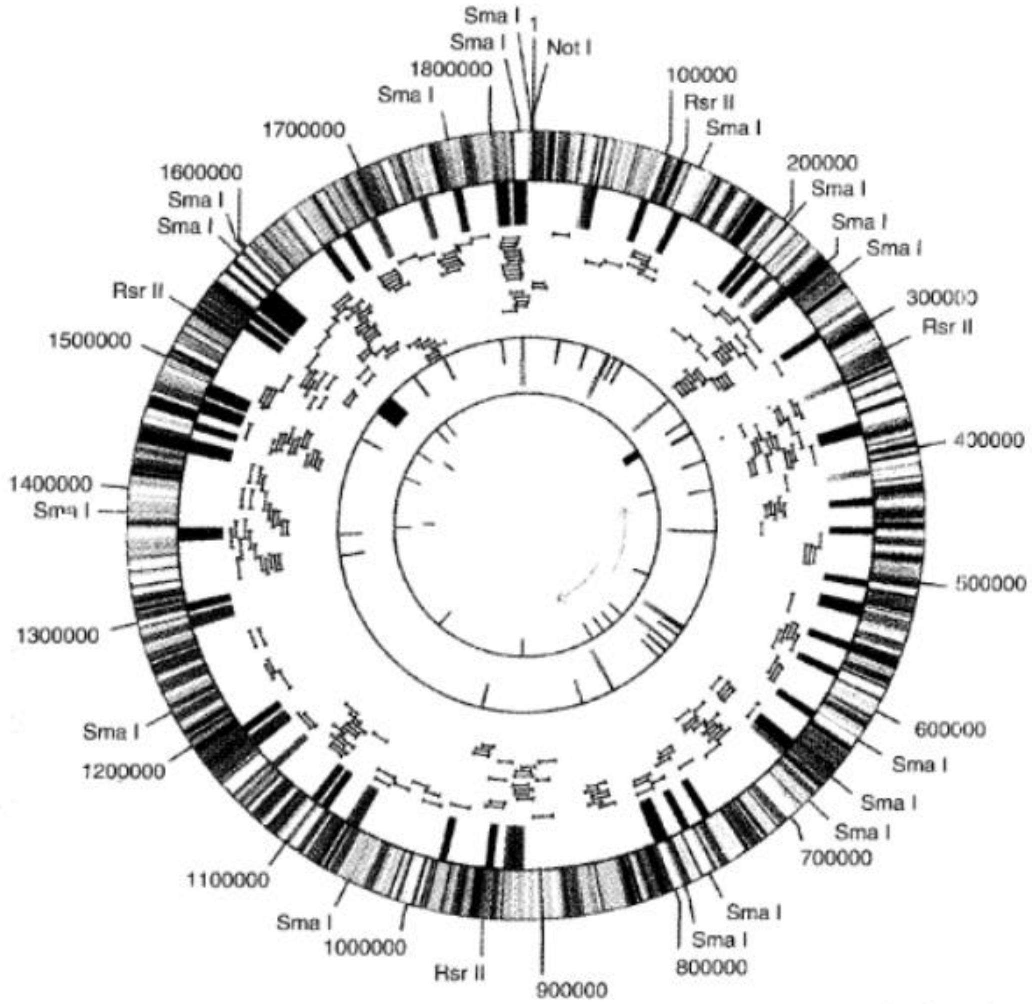


Fig. 1. A circular representation of the *H. influenzae* Rd chromosome illustrating the location of each predicted coding region containing a database match as well as selected global features of the genome. Outer perimeter: The location of the unique Not I restriction site (designated as nucleotide 1), the Rsr II sites, and the Sma I sites. Outer concentric circle: Coding regions for which a gene identification was made. Each coding region location is classified as to role according to the color code in Fig. 2. Second concentric circle: Regions of high G+C content (>42 percent, red; >40 percent, blue) and high A+T content (>66 percent, black; >64 percent, green). Third concentric circle: Coverage by λ clones (blue). More than 300 λ clones were sequenced from each end to confirm the overall structure of the genome and identify the six ribosomal operons. Fourth concentric circle: The locations of the six ribosomal operons (green), the tRNAs (black) and the cryptic mu-like prophage (blue). Fifth concentric circle: Simple tandem repeats. The locations of the following repeats are shown: CTGGCT, GTCT, ATT, AATGGC, TTGA, TTGG, TTTA, TTATC ,TGAC, TCGTC, AACC, TTGC, CAAT, CCAA. The putative origin of replication is illustrated by the outward pointing arrows (green) originating near base 603,000. Two potential termination sequences are shown near the opposite midpoint of the circle (red).

Genomes of Select Species of *Bacteria* and *Archaea*

Bacteria

Organism	Lifestyle[b]	Size (bp)	ORFs[c]	Features
Nasuia deltocephalinicola	E	112,091	137	Degenerate sap-feeding insect endosymbiont
Tremblaya princeps	E	138,931	121	Degenerate mealybug endosymbiont
Hodgkinia cicadicola	E	143,795	169	Degenerate cicada endosymbiont
Mycoplasma genitalium	P	580,070	525	Smallest nonsymbiotic bacterial genome
Rickettsia prowazekii	P	1,111,523	834	Obligate intracellular parasite, causes epidemic typhus
Treponema pallidum	P	1,138,006	1,041	Spirochete, causes syphilis
Methylophilaceae family, strain HTCC2181	FL	1,304,428	1,354	Marine methylotroph, smallest free-living genome
Thermotoga maritima	FL	1,860,725	1,877	Hyperthermophile

Genomes of Select Species of *Bacteria* and *Archaea*

Organism	Lifestyle[b]	Size (bp)	ORFs[c]	Features
Deinococcus radiodurans	FL	3,284,156	2,185	Radiation resistant, multiple chromosomes
Bacillus subtilis	FL	4,214,810	4,100	Gram-positive genetic model
Mycobacterium tuberculosis	P	4,411,529	3,924	Causes tuberculosis
Escherichia coli K-12	FL	4,639,221	4,288	Gram-negative genetic model
Escherichia coli O157:H7	FL	5,594,477	5,361	Enteropathogenic strain of *E. coli*
Bradyrhizobium japonicum	FL	9,105,828	8,317	Nitrogen fixation, nodulates soybeans
Sorangium cellulosum	FL	14,782,125	11,559	Forms multicellular fruiting bodies
Minicystis rosea	FL	16,040,666	14,018	Forms multicellular fruiting bodies

Genomes of Select Species of *Bacteria* and *Archaea*

Archaea

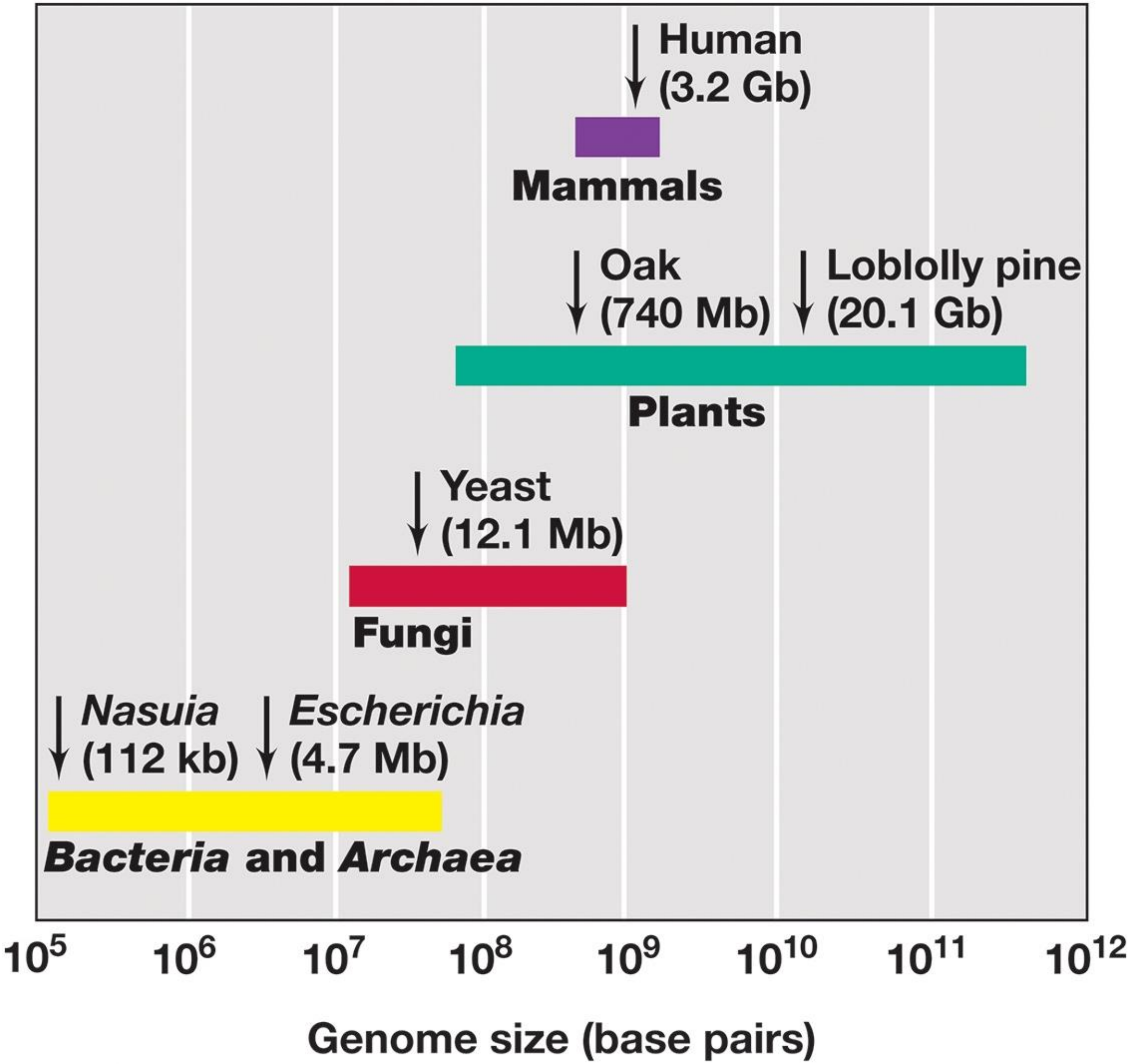
Organism	Lifestyle [b]	Size (bp)	ORFs[c]	Features
Nanoarchaeum equitans	P	490,885	552	Smallest nonsymbiotic cellular genome
Methanocaldococcus jannaschii	FL	1,664,976	1,738	Methanogen, hyperthermophile
Pyrococcus horikoshii	FL	1,738,505	2,061	Hyperthermophile
Sulfolobus solfataricus	FL	2,992,245	2,977	Hyperthermophile, sulphur chemolithotroph
Haloarcula marismortui	FL	4,274,642	4,242	Extreme halophile, bacteriorhodopsin
Methanosarcina acetivorans	FL	5,751,000	4,252	Acetate using methanogen

[a] Information on prokaryotic genomes can be found at

[b] E, endosymbiont; P, parasite; FL, free-living.

[c] Open reading frames. Genes encoding known proteins are included, as well as ORFs that could encode a protein greater than 100 amino acid residues.

Genome Sizes of Microbial Cells and Higher Organisms



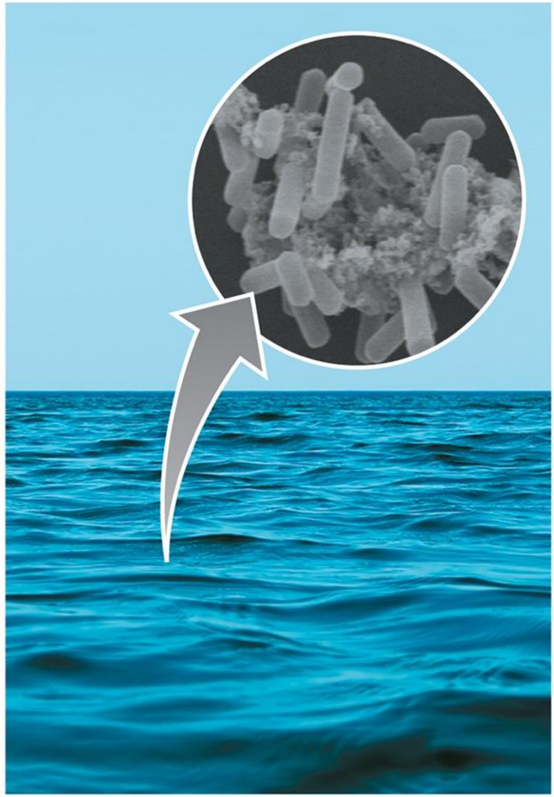
Introduction to Genomics

What Can Genomes Tell Us?

- Modern microbiology thrives on sequences
- Heat-stable enzymes to virulence factors
- Studying gene expression, detecting horizontal transfer, monitoring and diagnosing disease outbreaks, anti-bacteriophage systems, understanding metabolism, and determining growth requirements…….

- Solving medical mysteries (**e.g.,** cause of "Black Death")
- Identifying new microbial phyla

Yersinia pestis

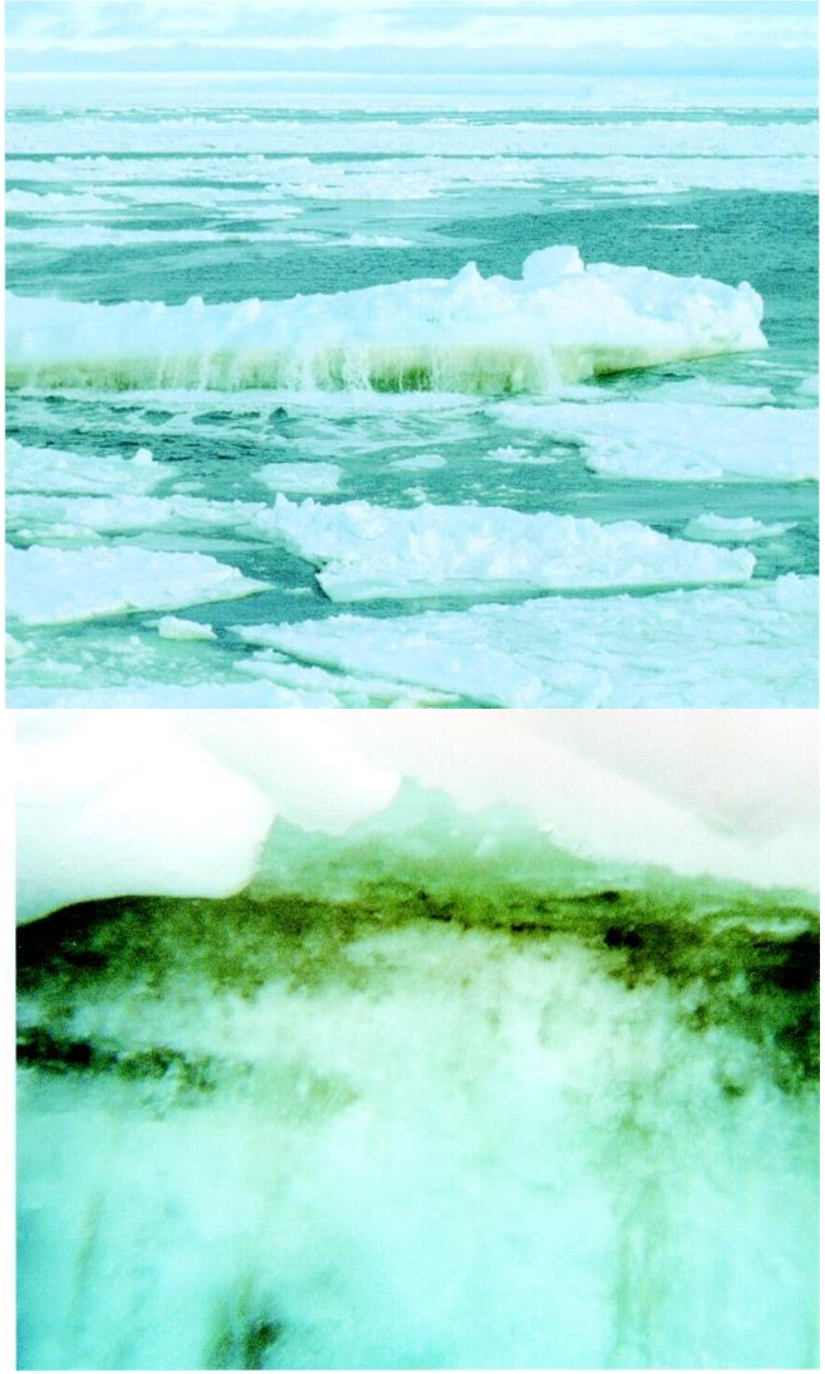


marine ammonia-oxidizing archaeon Nitrosopumilus assigned to a new phylum: *Thaumarchaeota*

Introduction to Genomics

What Can Genomes Tell Us?

- How do certain organism are able to exist in their environments
- Determine adaptation strategies
- How do certain organism contribute to biogeochemical cycles



Sequencing and Annotating Genomes

- **Sequencing:** determining the precise order of nucleotides in a DNA or RNA molecule
- **Genome annotation:** converting raw sequence data into a list of genes and other functional sequences present in the genome
- **Bioinformatics:** storing and analyzing sequences and structures of nucleic acids and proteins
- Annotation is “bottleneck” in genomics

Sequencing and Annotating Genomes

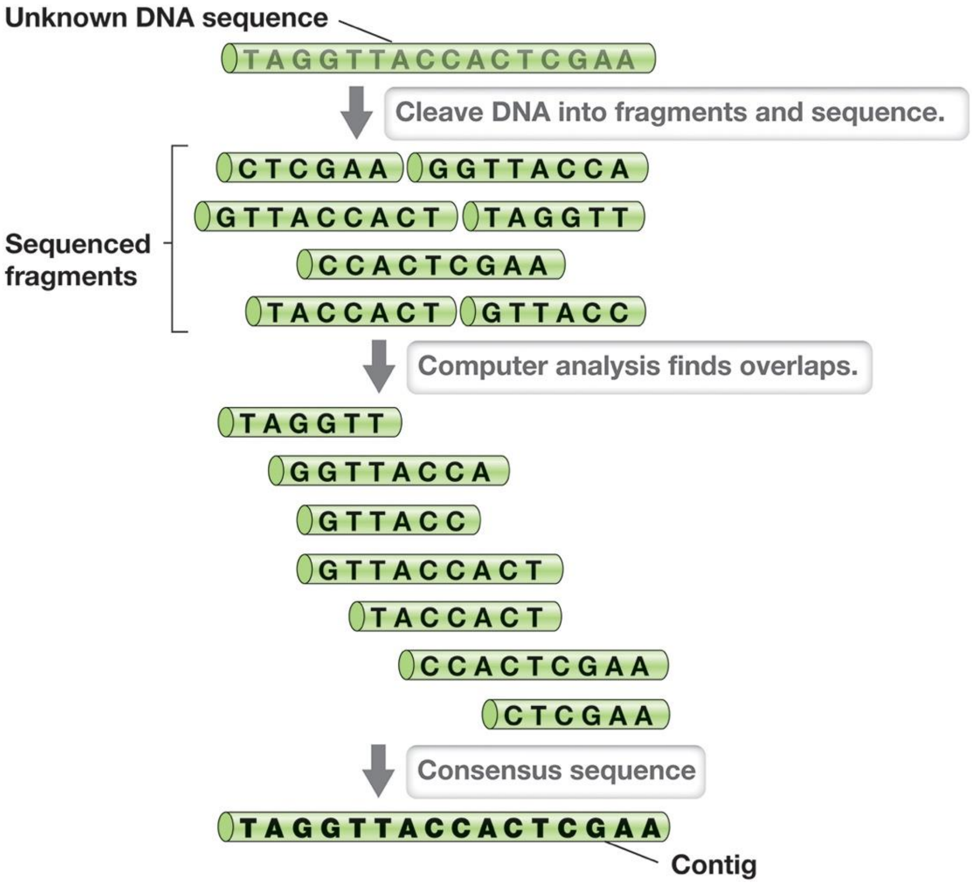
Genome Assembly and Annotation

Genome assembly consists of putting fragments in the correct order and eliminating overlaps

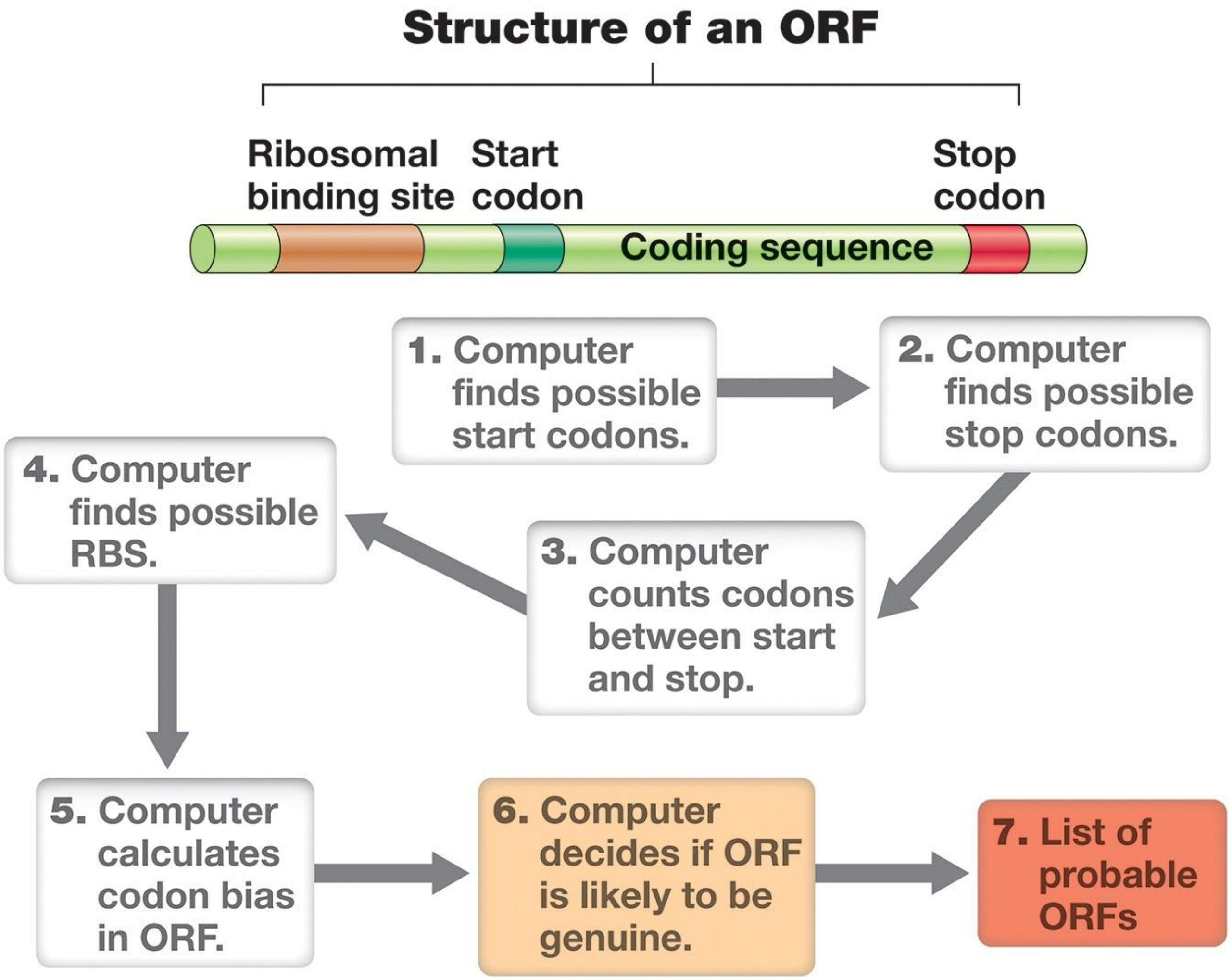
Annotation identifies genes

Highly computational

- Map of complete genome is generated
- Bacterial and archaeal genomes are a series of open reading frames (ORFs)
- Functional ORF: encodes a protein



Computer Identification of Possible ORFs



Structure of an ORF

1. Computer finds possible start codons.
2. Computer finds possible stop codons.
3. Computer counts codons between start and stop.
4. Computer finds possible RBS.
5. Computer calculates codon bias in ORF.
6. Computer decides if ORF is likely to be genuine.
7. List of probable ORFs

Sequencing and Annotating Genomes

Never 100 percent identification; many genes encode proteins of unknown function

Number of genes with role that can be clearly identified in a given genome is ~70% or less of total ORFs detected

Hypothetical proteins: uncharacterized ORFs; proteins that likely exist but whose function is currently unknown

- lack sufficient amino acid sequence homology with known proteins for identification
- assign to family or general function

Sequencing and Annotating Genomes

Some genes encode RNA that is not translated (noncoding RNA)

- lack start codons and may have multiple stop codons
    - transfer RNA (tRNA)
    - ribosomal RNA (rRNA)
    - noncoding regulatory RNA molecules

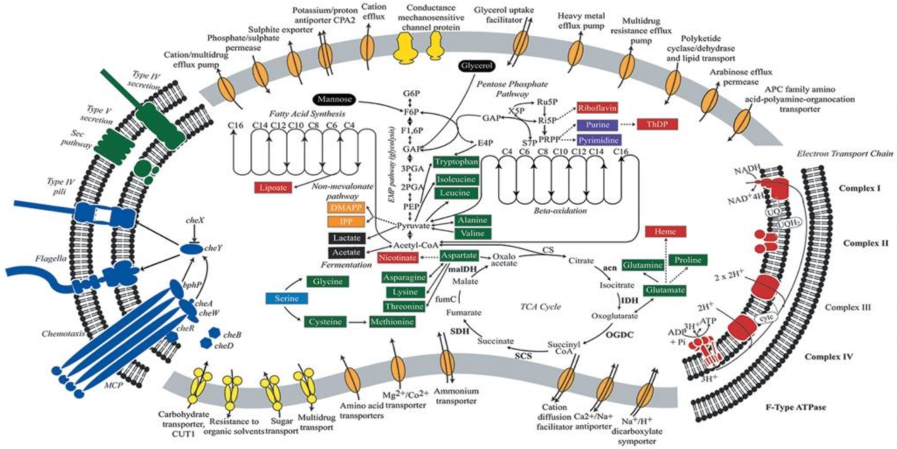
Gene Content of Bacterial Genomes

Complement of genes in a particular organism defines its capabilities

Genomes are also influenced by an organism’s lifestyle (or vice-versa?)

Enzymes in the genome can be related to lifestyle of the microorganism

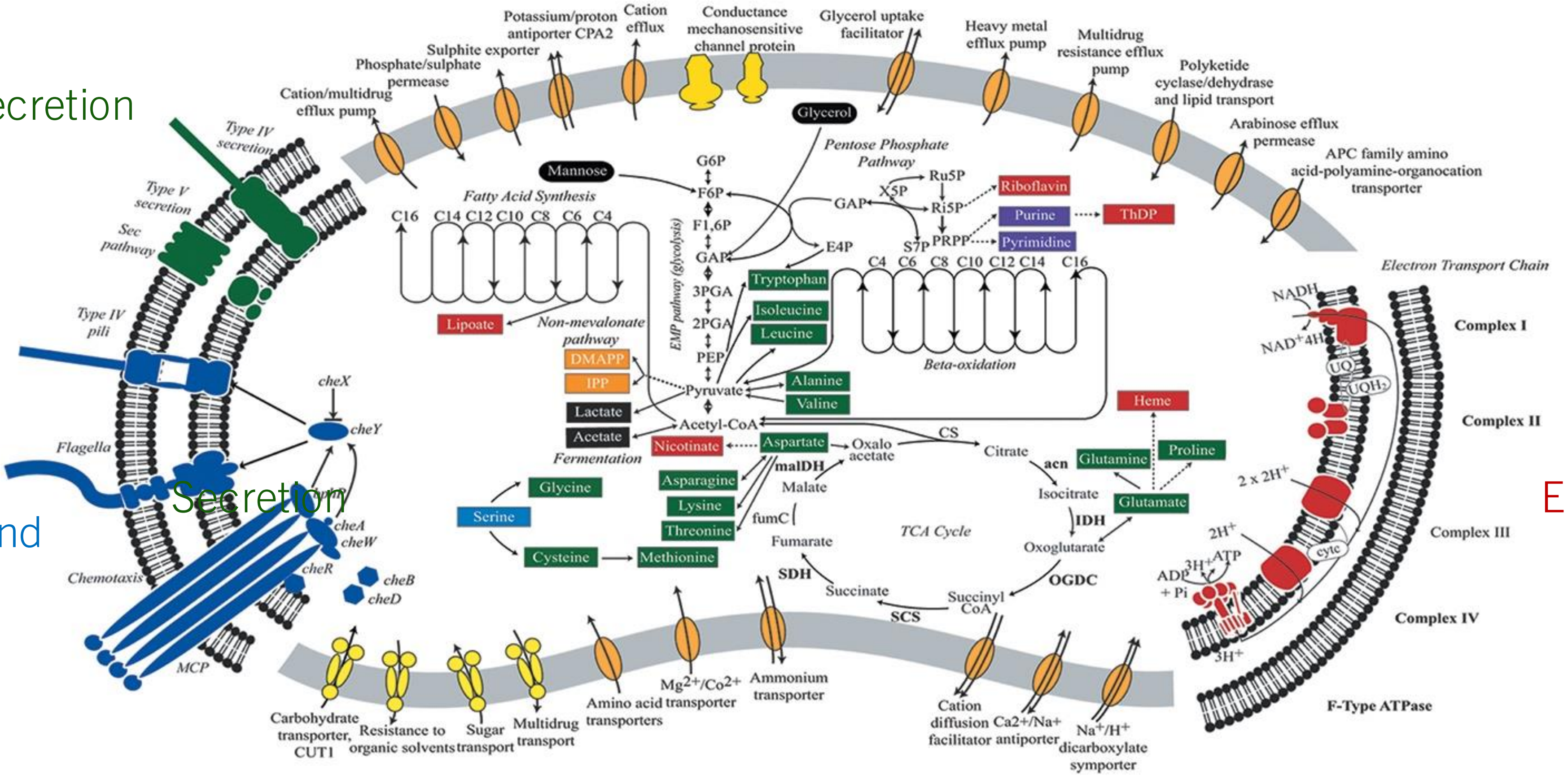
Metabolic genes typically most abundant in bacterial genomes



Soo, R.M., Woodcroft, B.J., Parks, D.H., Tyson, G.W., and P. Hugenholtz. 2015, *PeerJ* 3: e968

Functional and Metabolic Predictions for *Vampirovibrio chlorellavorus* Based on Genomic Annotation

Secretion



Secretion

Chemotaxis and movement

Electron transport

Pumps and transporters

Gene Function in Some Genomes of *Bacteria*

Functional categories	Percentage of genes *Escherichia coli* (4.64 Mbp)[a]	Percentage of genes *Haemophilus influenzae* (1.83 Mbp)[a]	Percentage of genes *Mycoplasma genitalium* (0.58 Mbp)[a]
Metabolism	21.0	19.0	14.6
Structure	5.5	4.7	3.6
Transport	10.0	7.0	7.3
Regulation	8.5	6.6	6.0
Translation	4.5	8.0	21.6
Transcription	1.3	1.5	2.6
Replication	2.7	4.9	6.8
Other, known	8.5	5.2	5.8
Unknown	38.1	43.0	32.0

[a] Chromosome size, in megabase pairs. Each organism listed contains only a single circular chromosome.

- 25 % of the grade
- In groups or individual

PROJECT REPORT

- Research paper

Based on data collected during the course

Intro, Methods, Results, Discussion, Conclusion, References

PROJECT REPORT

- 25 % of the grade
- In groups or individual

Isolated microorganism

What is its taxonomy?

What kind of metabolism does it have?

How does it acquire Nitrogen and Carbon?

Can it produce greenhouse gases? Which?

What kind of metabolites does it make/excrete?

Can it detoxify metal contamination?

Does it have antibiotic resistance genes?

How can it potentially interact with other organisms?

Field Data

Use the data gathered in the field to contextualize the analysis of your isolate.

- Does the chemistry/temperature/gas flux of the environment where you isolated your organism from make sense with what you know about four organism microorganism (from your results and literature)?

Use published literature and your results.
Compare results to published data in the Discussion