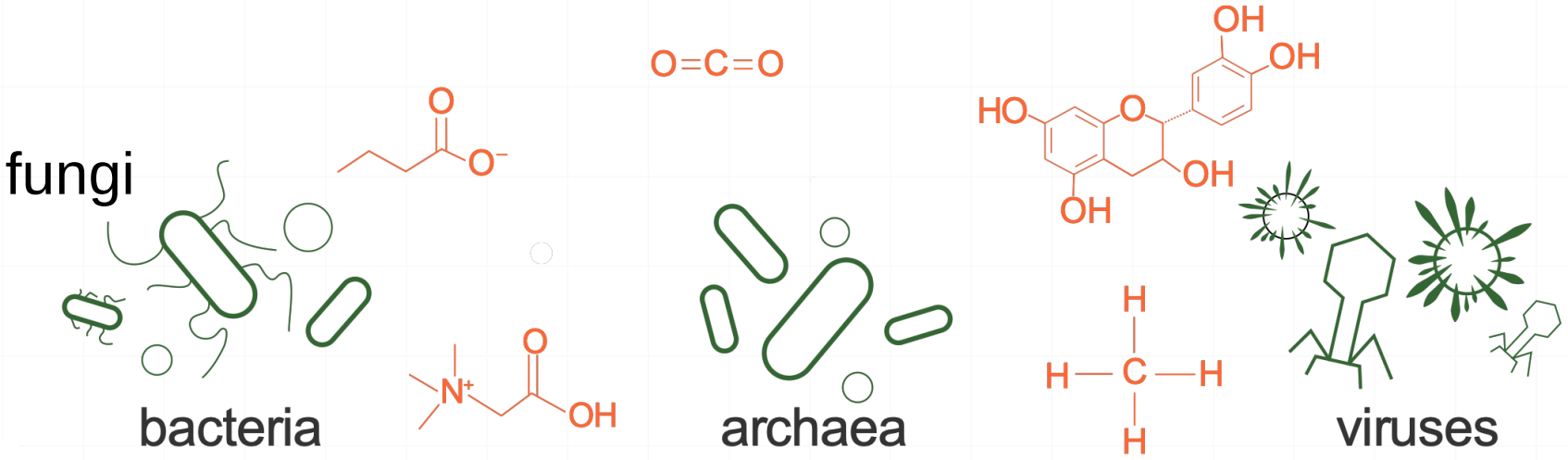
MICROBIOME- scaling genomics to communities

Kelly Wrighton
@kwrighton
Wrightonlab.com

What is the microbiome?

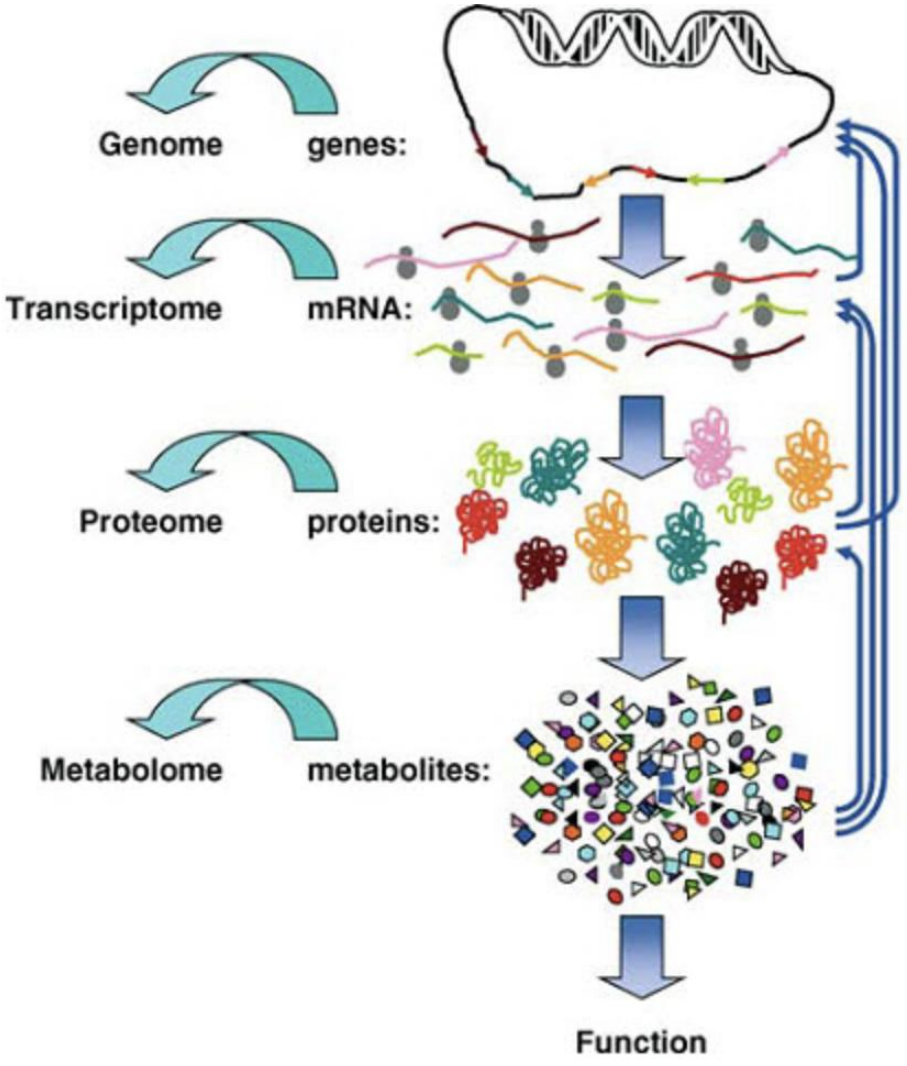
The study of microorganisms, their genomes, and their surrounding environmental conditions



Today we are going to discuss many of these multi-omics tools

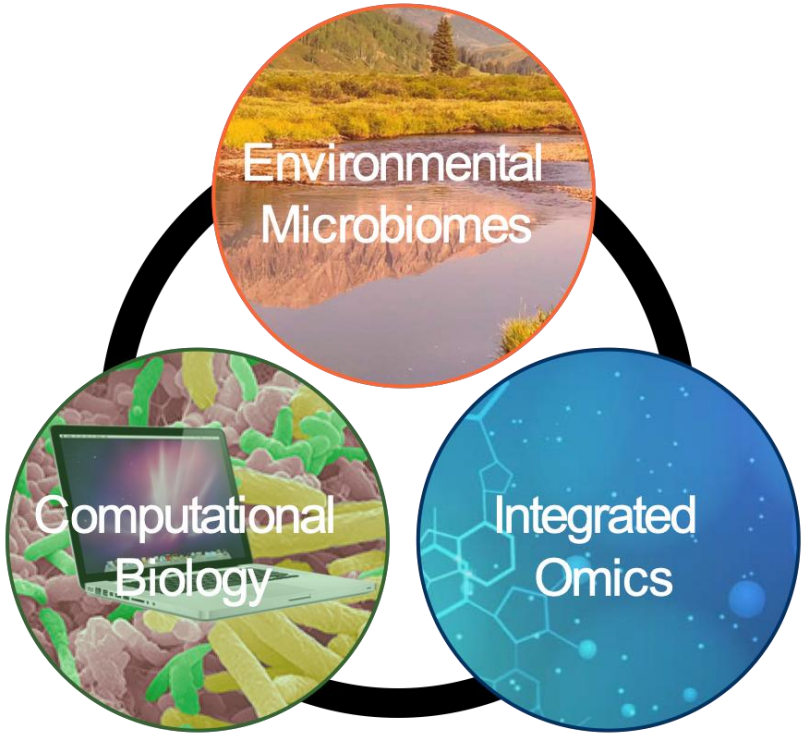
&

provide case studies to help contextualize



Genomics meets the Microbiome

- **About Wrighton Microbiome Lab**
- Overview of microbiome relevant methods
- Session 1: Metagenomics: where have we been, where are we today.
- Session 2: Case study to highlight how we use these tools to gain new insights into ecosystems



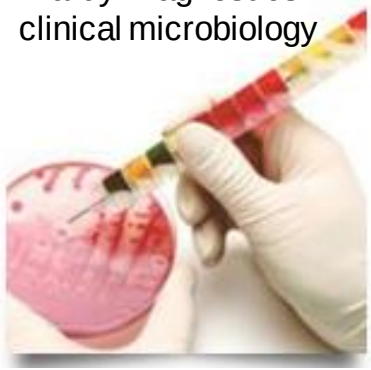
My Science Trajectory

Hardy Diagnostics-
clinical microbiology

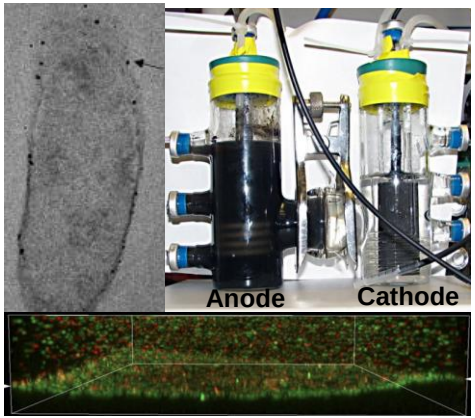
Maters-molecular tools
bioremediation

My Science Trajectory

Hardy Diagnostics-
clinical microbiology



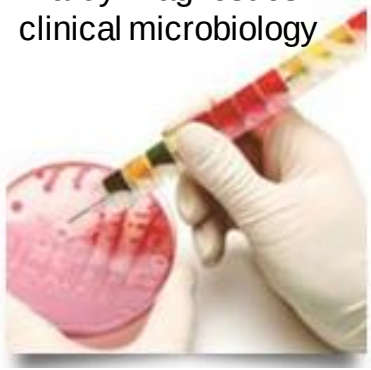
PhD UC Berkeley- Coates Lab
mechanisms extracellular respiration



Maters-molecular tools
bioremediation

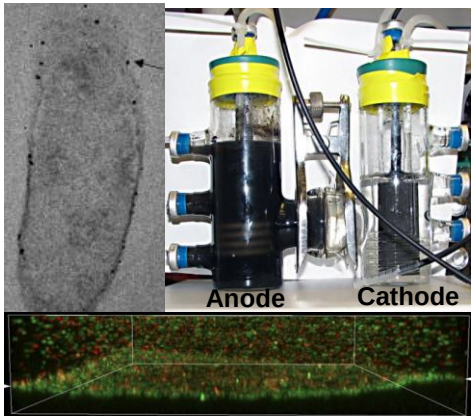
My Science Trajectory

Hardy Diagnostics- clinical microbiology

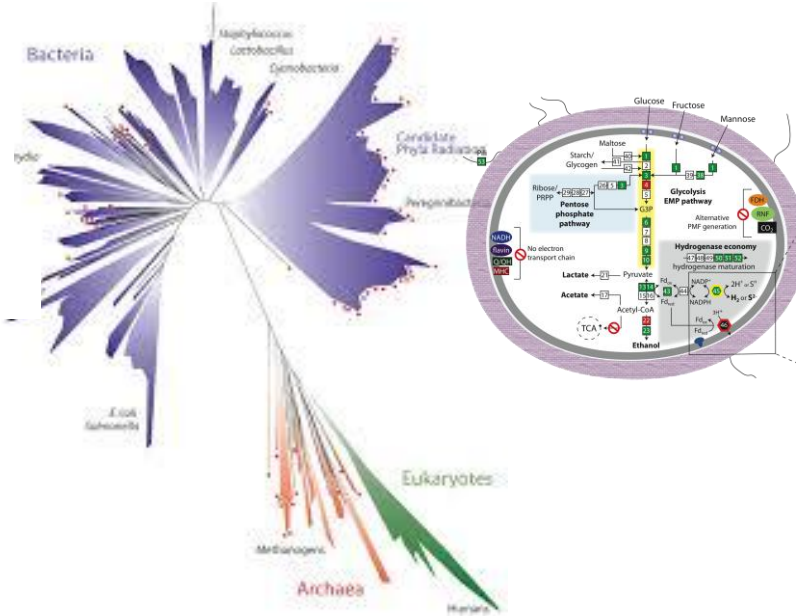


Maters-molecular tools bioremediation

PhD UC Berkeley- Coates Lab
mechanisms extracellular respiration



Post-doc- UC Berkeley Banfield Lab
Omics to discover new branches of tree of life



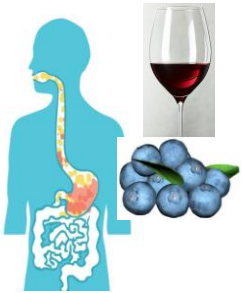
Wrighton Lab Started in 2013

2002 MS in Ecology
2005 Phd UC Berkeley- Microbiology
2010 Post-Doc UC Berkeley- Computational Biology
2013 Assistant Professor, Ohio State University
Microbiology
2018 Assistant Professor, CSU
Microbiome Data Sciences, Soil and Crop Science

@kcwrighton
Wrighton@colosate.edu
Wrightonlab.com

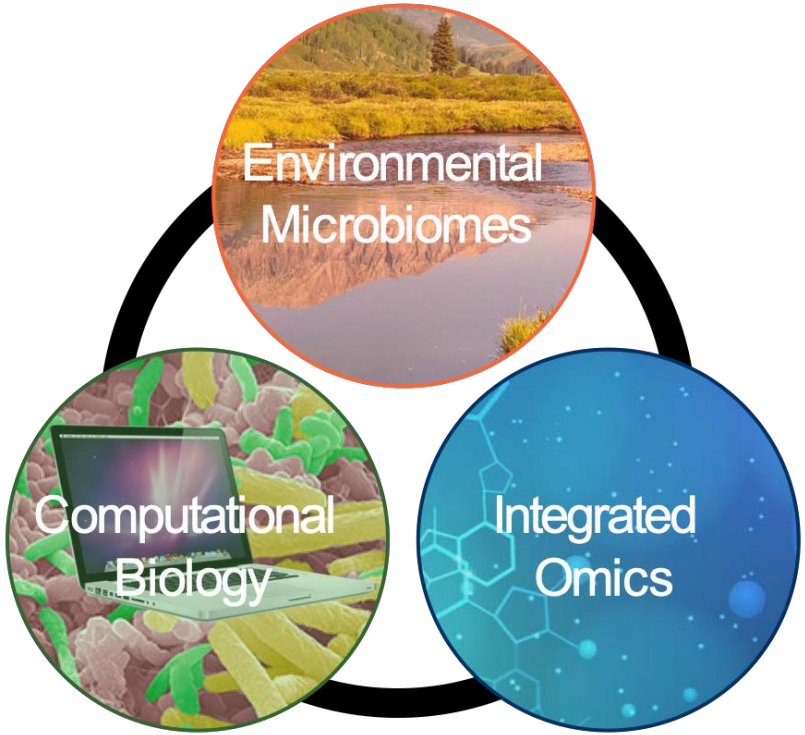
A little about the Wrighton Microbiome Lab Group

Ecosystem agnostic- but with a consistent focus on microbial metabolism and metabolic connectivity that contribute to ecosystem properties



Genomics meets the Microbiome

- About Wrighton Microbiome Lab
- **Overview of microbiome relevant methods**
- Session 1: Metagenomics: where have we been, where are we today.
- Session 2: Case study to highlight how we use these tools to gain new insights into natural systems

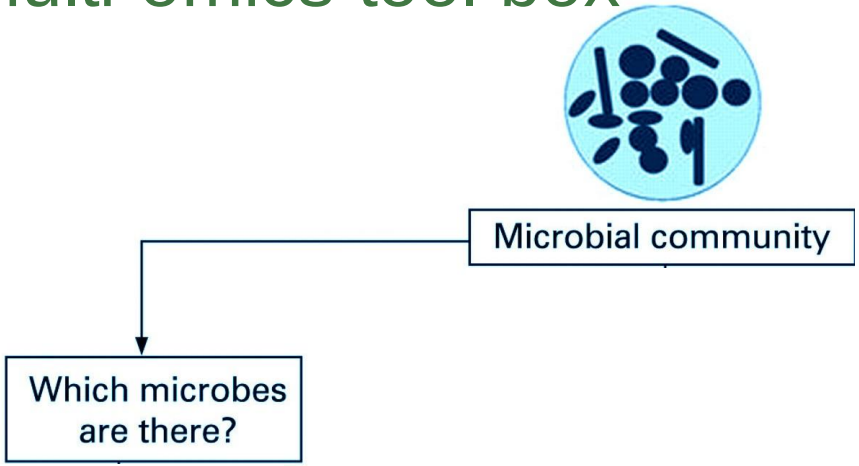


Microbiome in the genomics era :

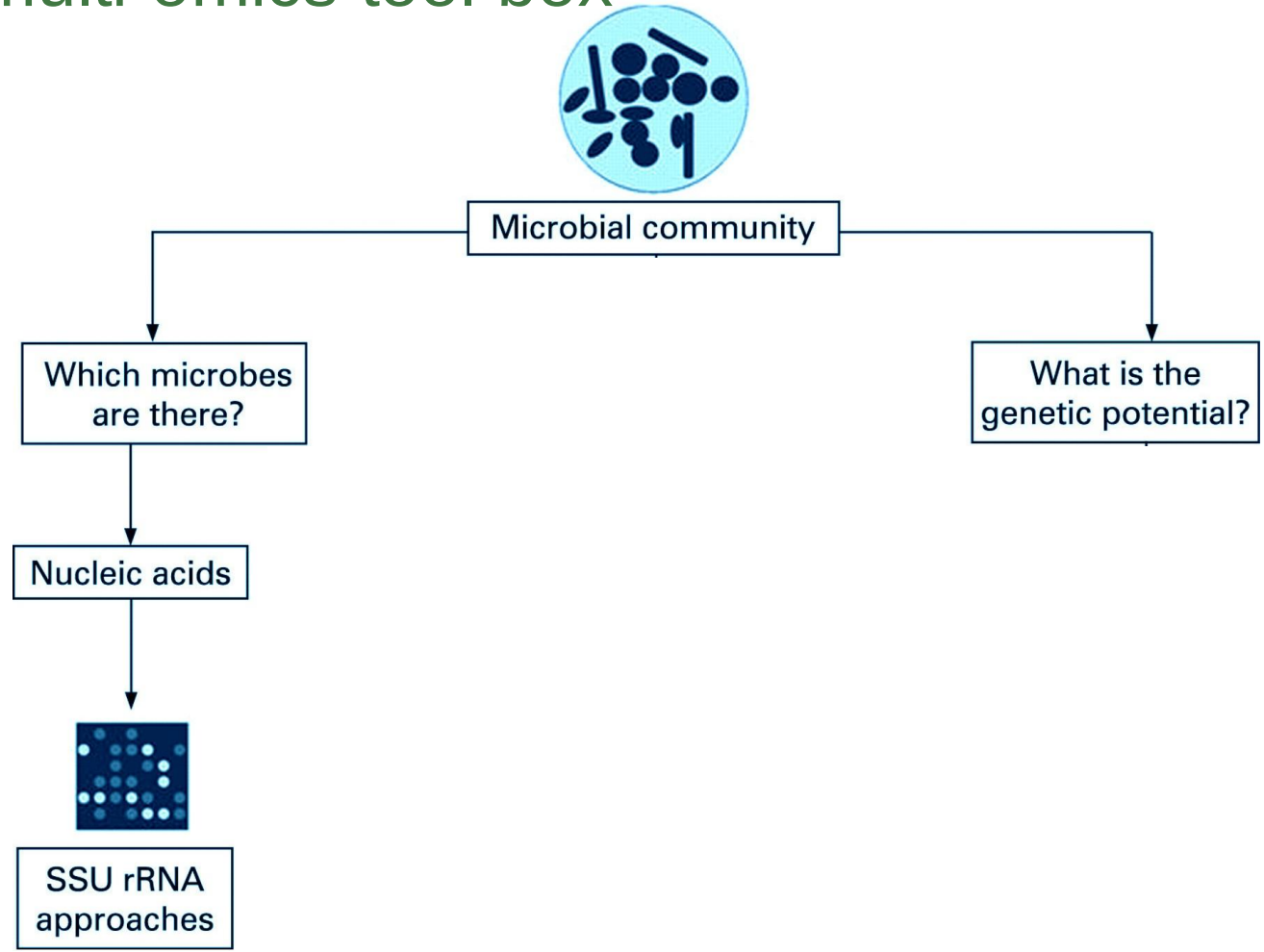
You need to use the right tool for the job,

but this depends on your question

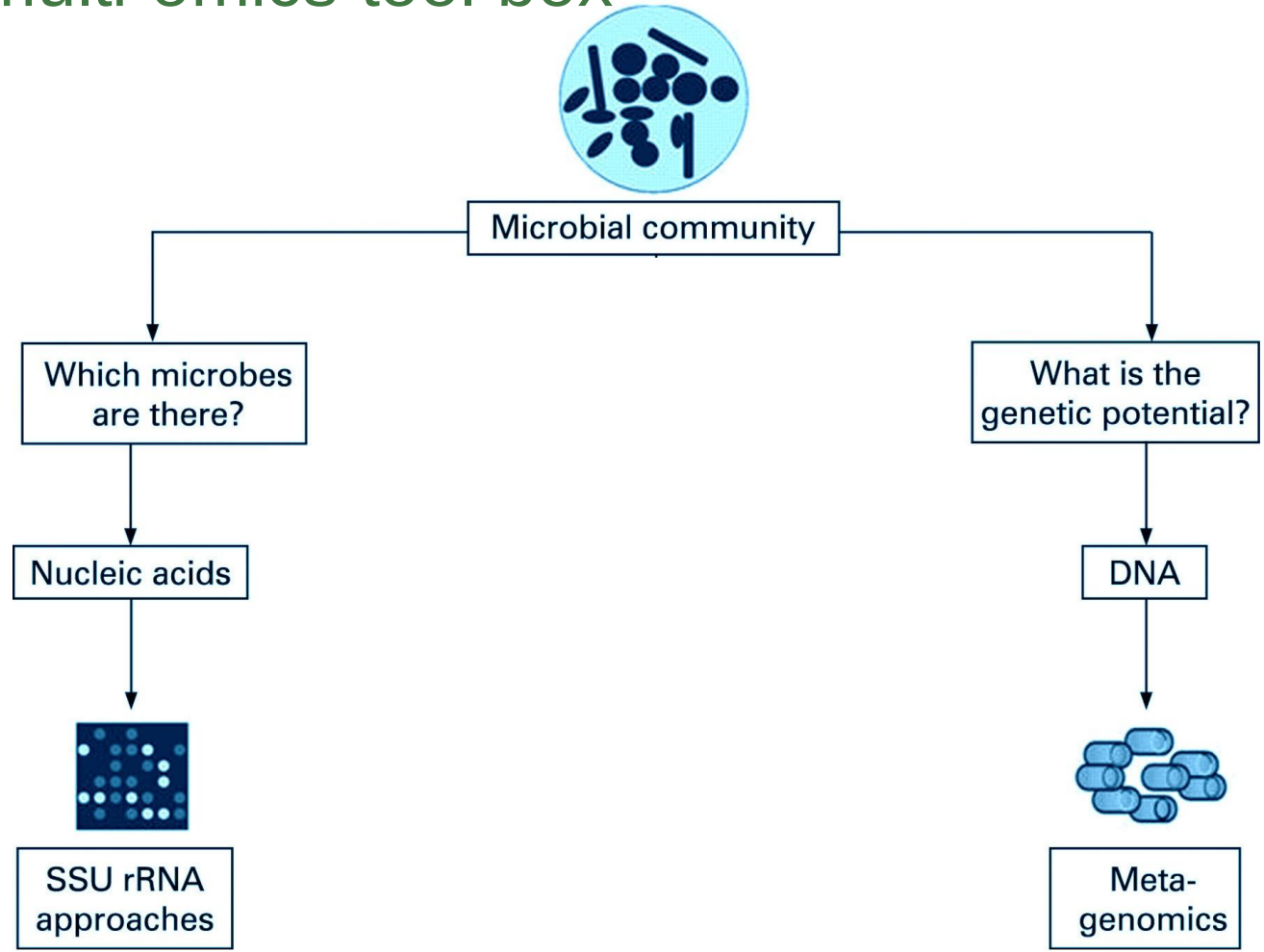
The multi-omics tool box



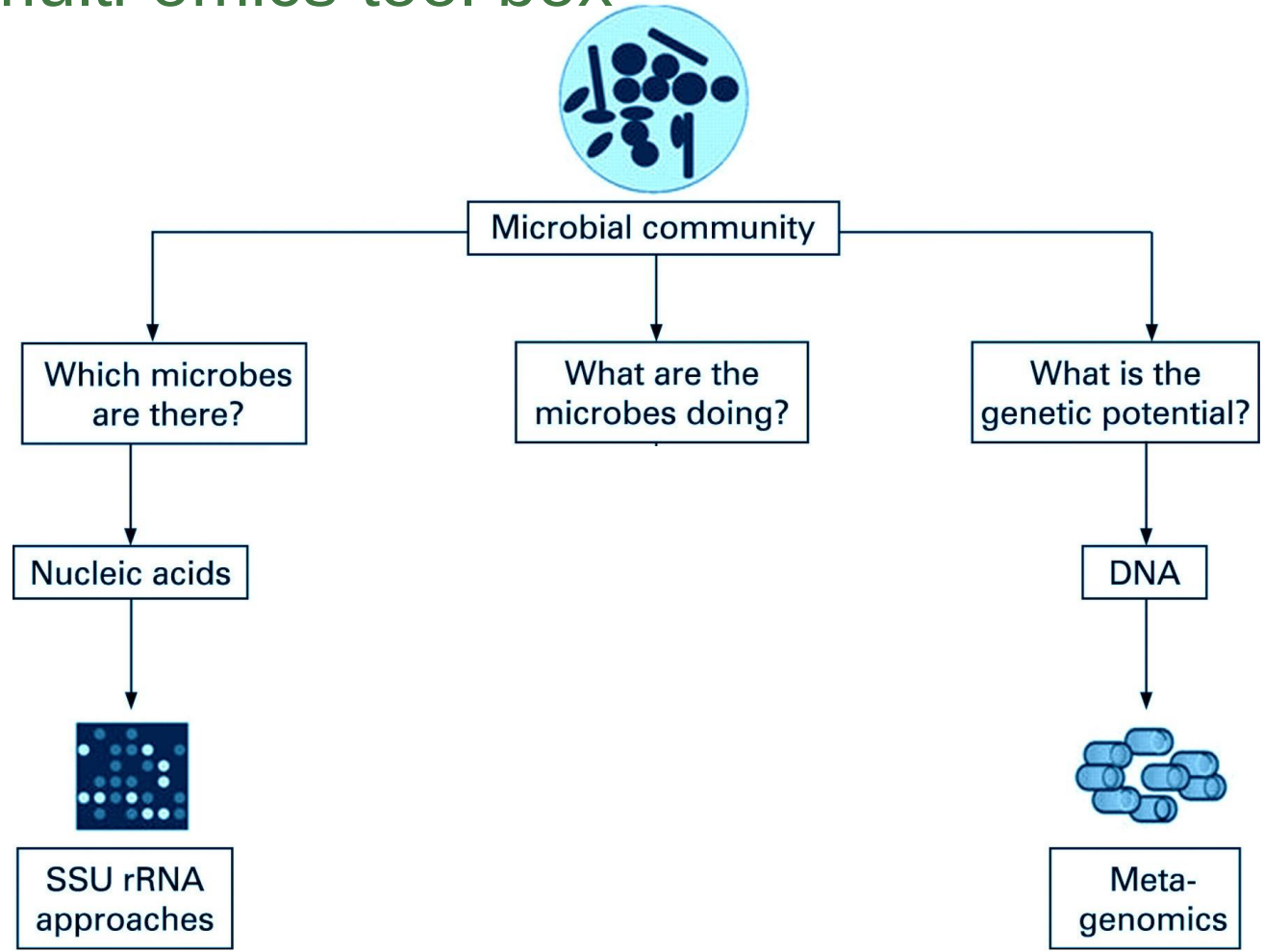
The multi-omics tool box



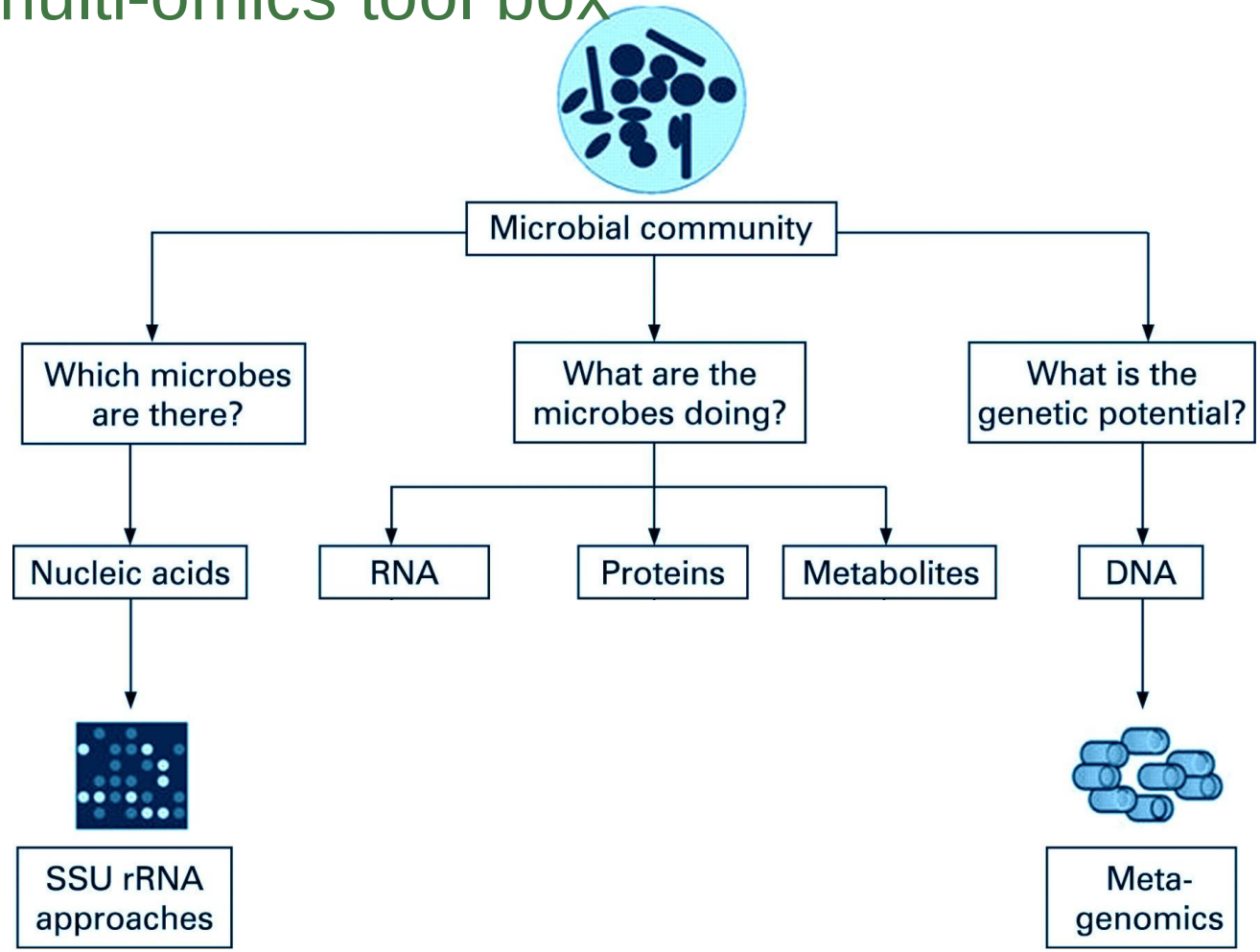
The multi-omics tool box

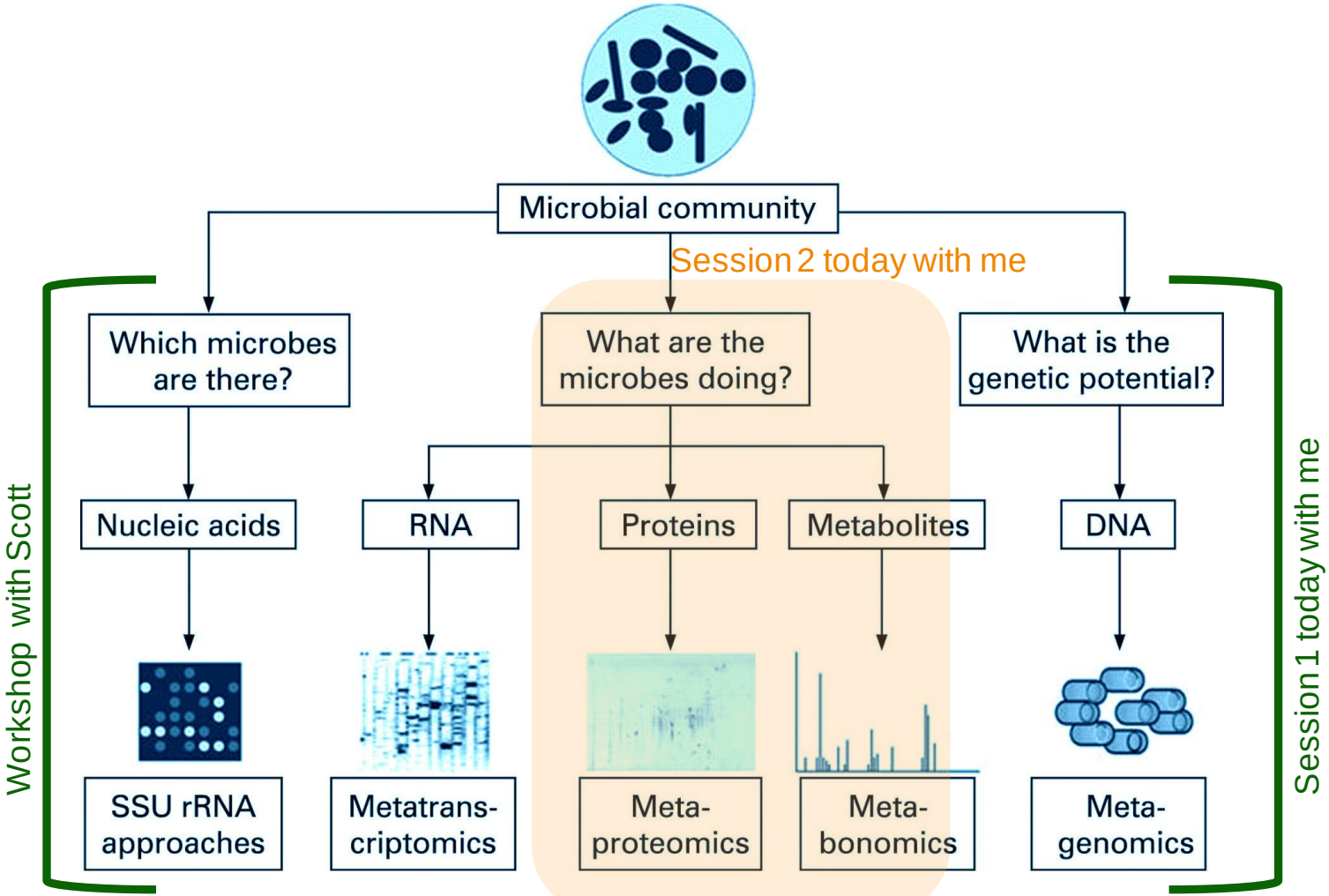


The multi-omics tool box



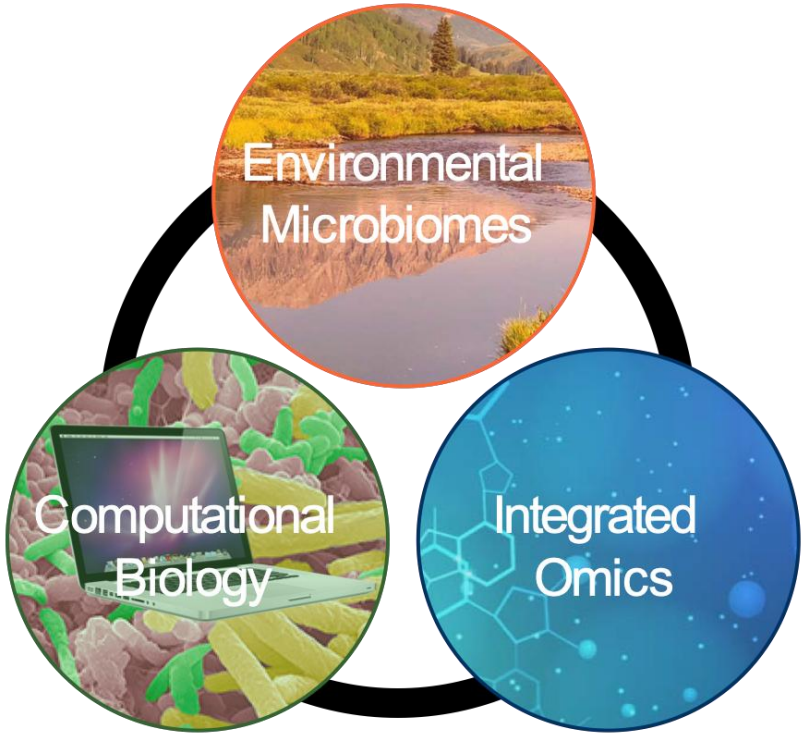
The multi-omics tool box




Microbial community
Session 2 today with me
Which microbes are there?
What are the microbes doing?
What is the genetic potential?
Nucleic acids
RNA
Proteins
Metabolites
DNA
SSU rRNA approaches
Metatrans-criptomics
Meta-proteomics
Meta-bonomics
Meta-genomics
Workshop with Scott
Session 1 today with me

Genomics meets the Microbiome

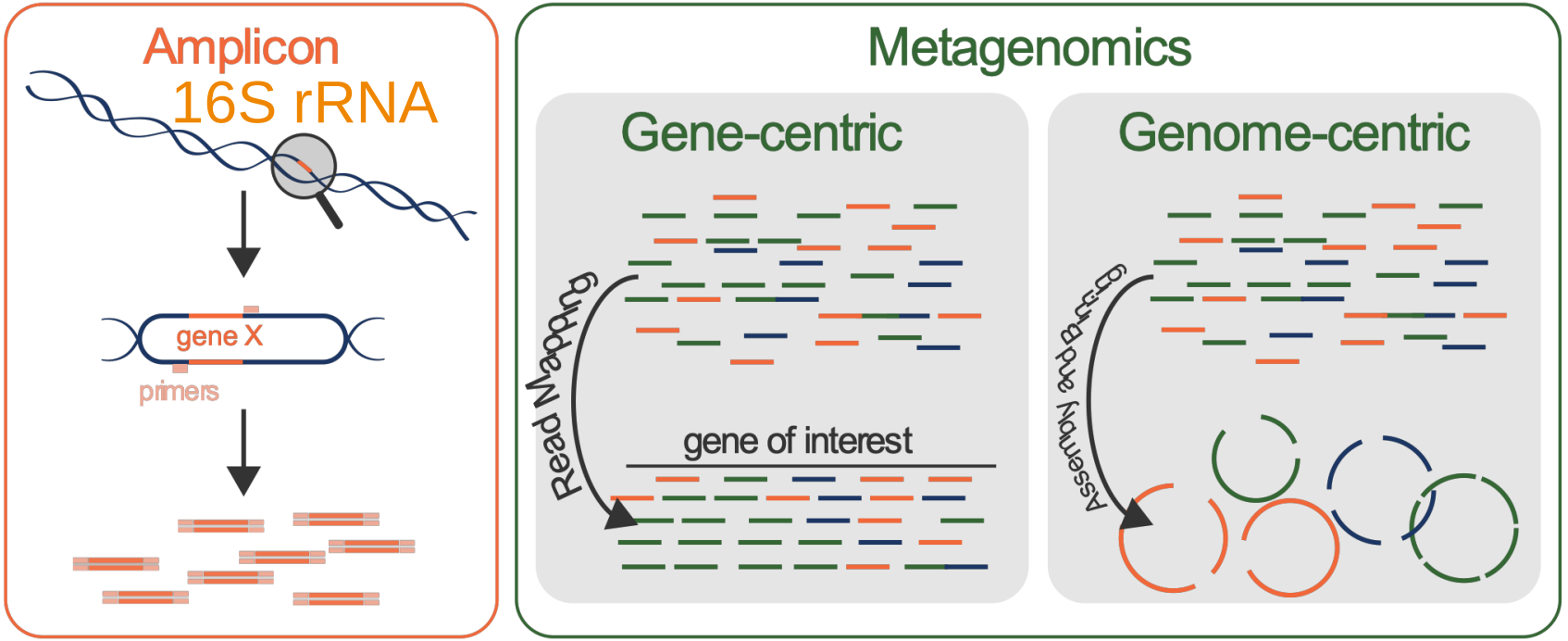
- About Wrighton Microbiome Lab
- Overview of microbiome relevant methods
- **Session 1: Metagenomics: where have we been, where are we today.**
- Session 2: Case study to highlight how we use these tools to gain new insights into natural systems



Session 1 Learning objectives

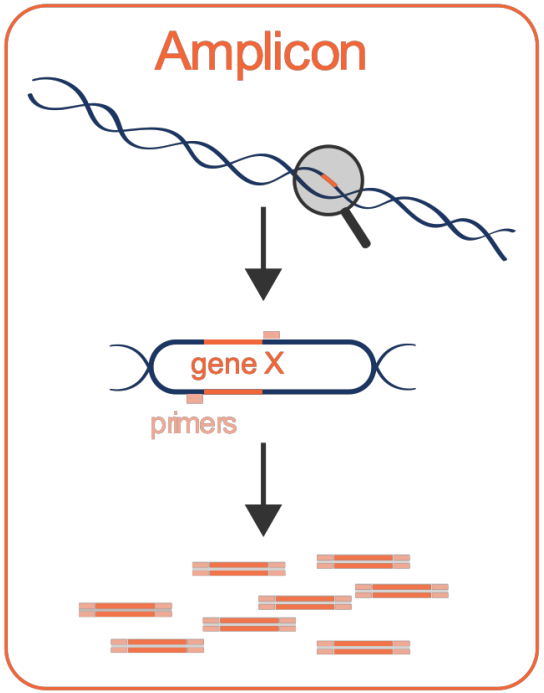
- What is microbial dark matter, why does it exist?
- What is metagenomics and how has it changed our understanding of microbial diversity (to uncover dark matter_
- What is the difference between 16S and metaG when to use each method
- What is the value of the unbinned metagenome approach- know example
- What is the value of the binned metagenome approach- know example
- Since these original papers, where is the field today

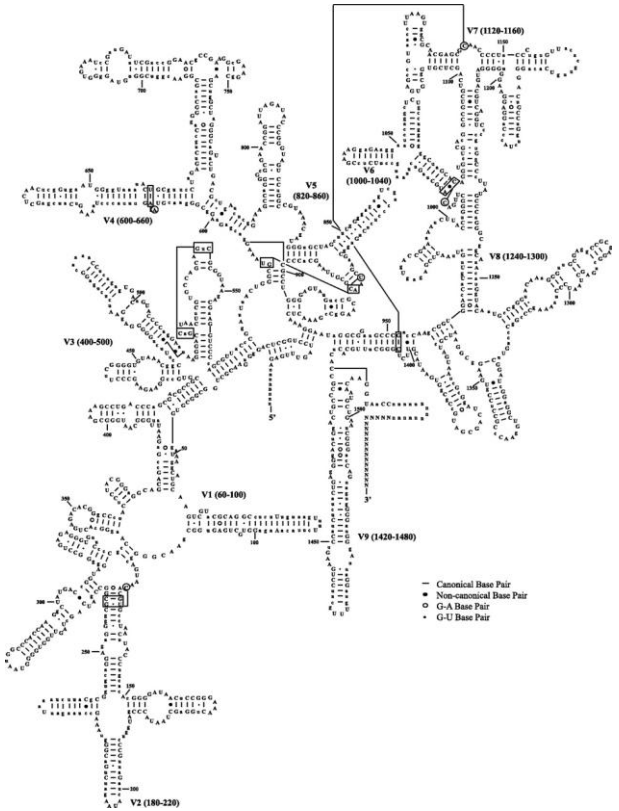
DNA tools in the microbiome field



16S rRNA versus Whole Genome Sequencing

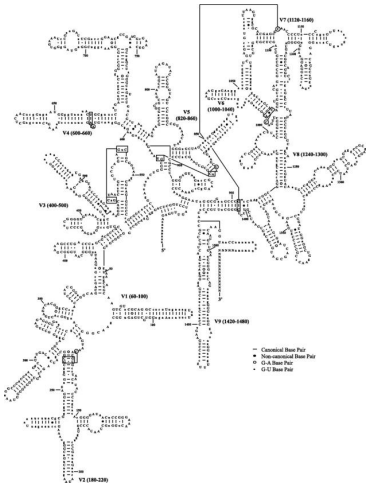
16S rRNA gene is a commonly used biomarker





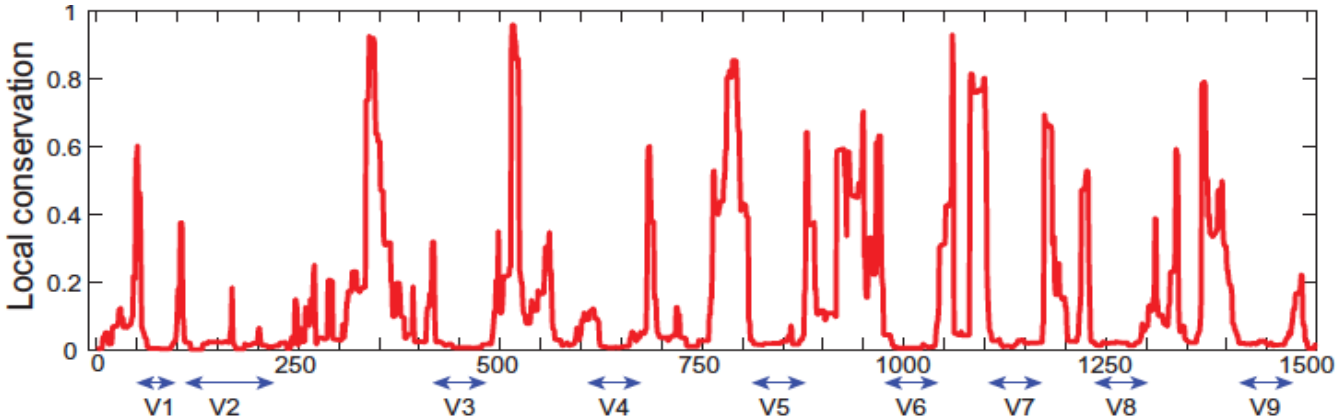
track taxonomy and ‘relative abundance’

16S rRNA gene is a good biomarker
ubiquity, variability, conservation



Series of 9 variable regions (denoted by V) represented by valleys

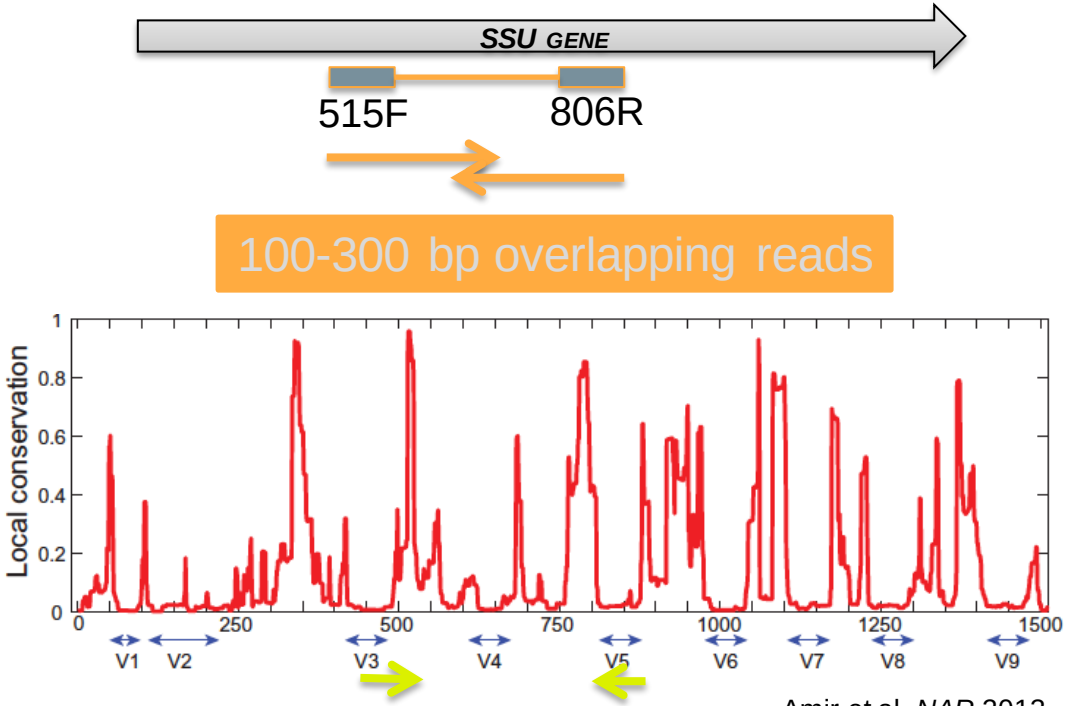
Conserved regions are represented by peaks



Amir et al. *NAR* 2013

From Woese to PCR :16S rRNA sequencing today

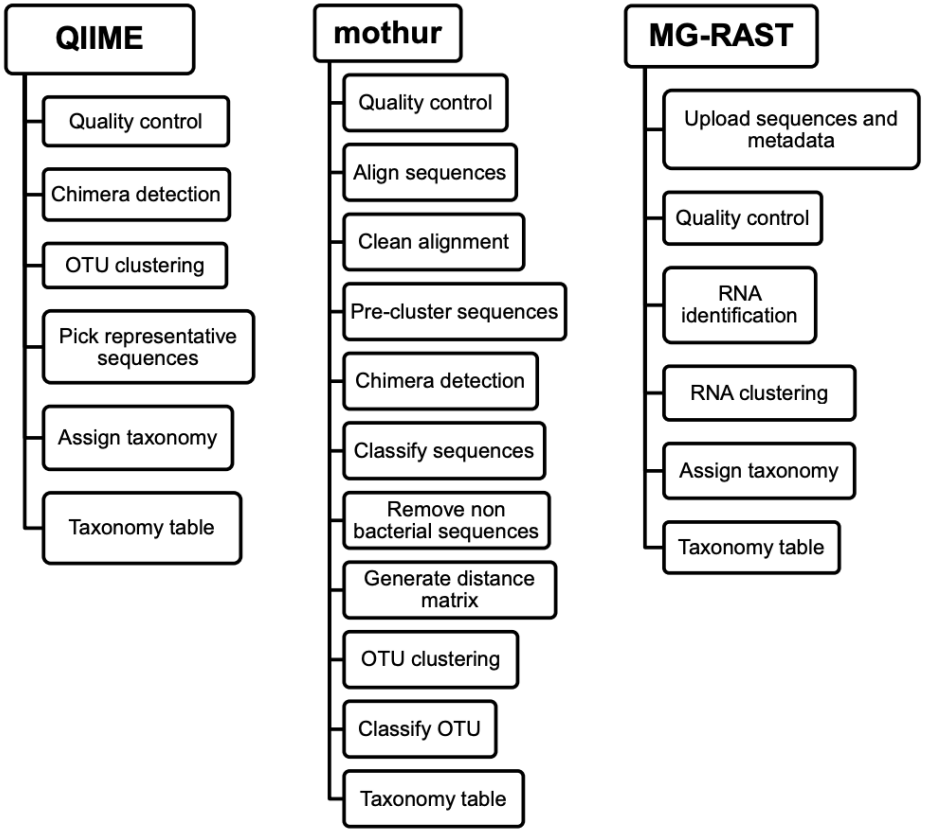
High-throughput sequencing of hypervariable fragment of 16S rRNA



Amir et al. *NAR* 2013

e.g. Kozich et al *AEM* 2013; Fadrosh et al *Microbiome* 2014

16S rRNA gene is a good biomarker
sequencing affordable and computationally accessible



OPEN
doi:10.1038/nature24621

A communal catalogue reveals Earth's multiscale microbial diversity

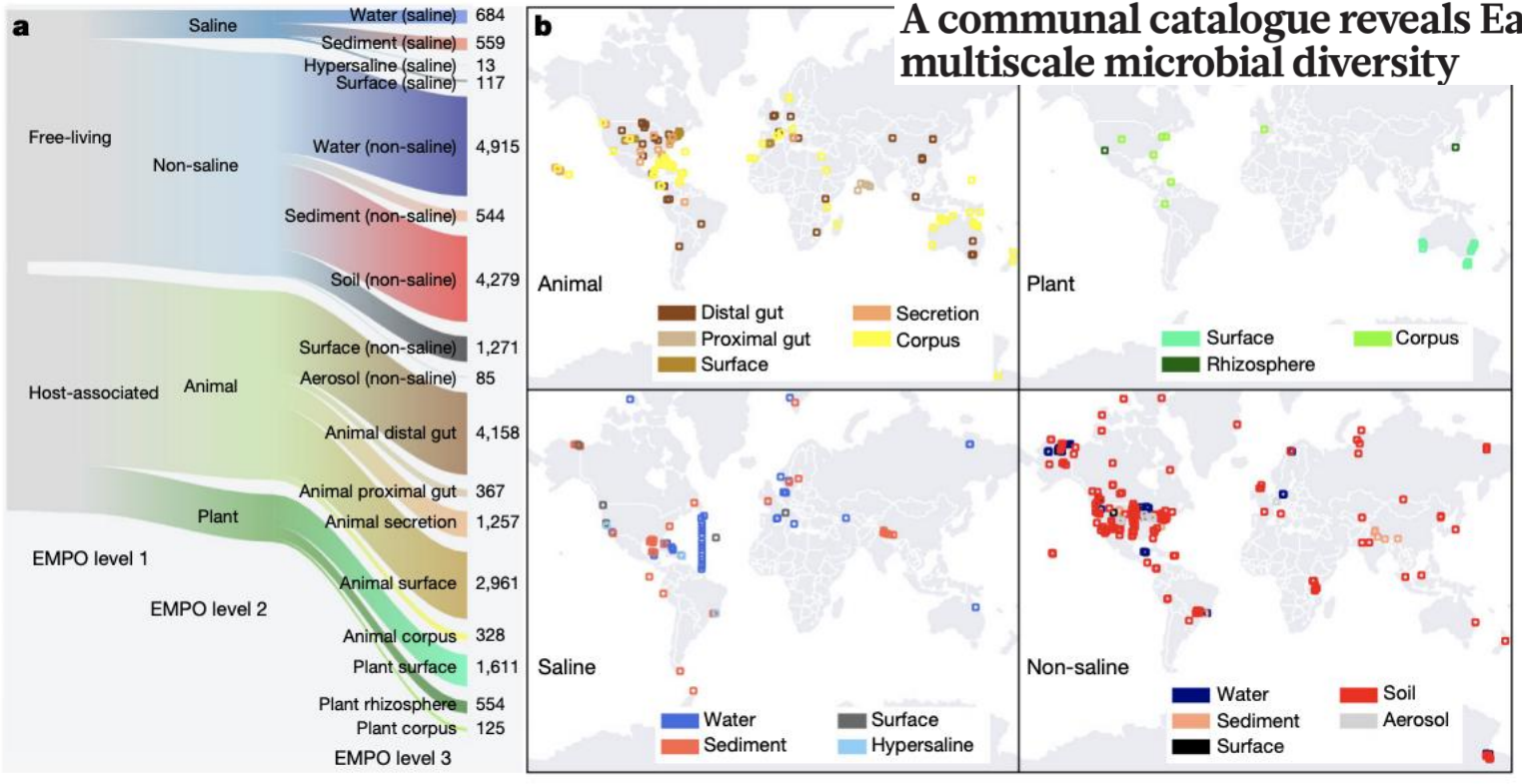
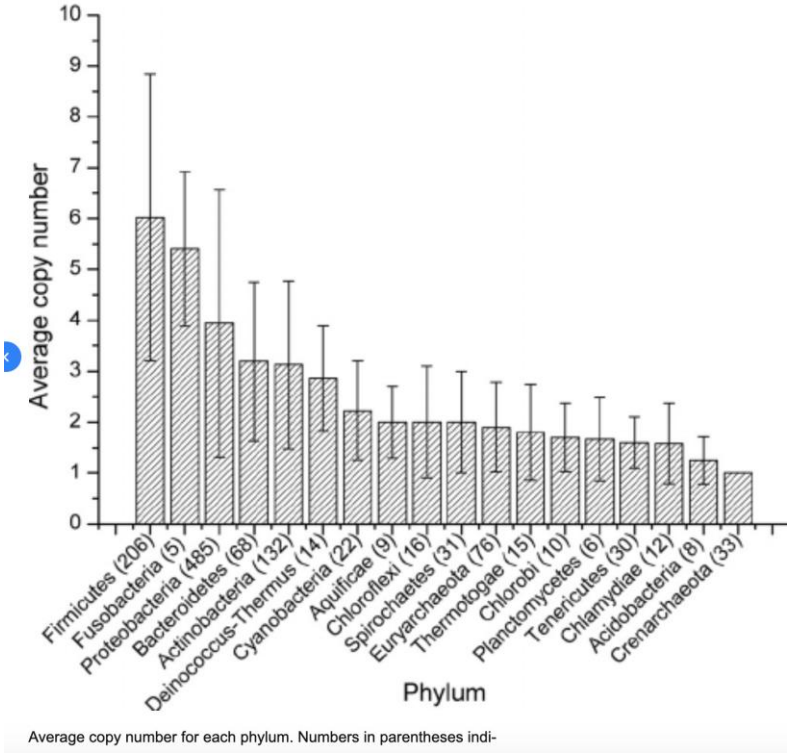
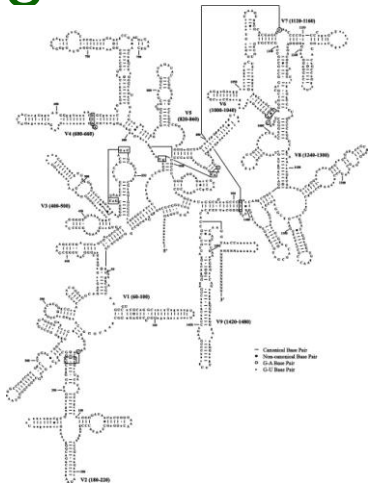


Figure 1 | Environment type and provenance of samples. a, The EMP ontology (EMPO) classifies microbial environments (level 3) as free-living or host-associated (level 1) and saline or non-saline (if free-living) or animal or plant (if host-associated) (level 2). The number out of 23,828 samples in the QC-filtered subset in each environment is provided. EMPO is described with examples at http://www.earthmicrobiome.org/protocols-and-standards/empo. **b**, Global scope of sample provenance: samples come from 7 continents, 43 countries, 21 biomes (ENVO), 92 environmental features (ENVO), and 17 environments (EMPO).

16S rRNA gene has copy number differences can skew interpretation of relative abundance

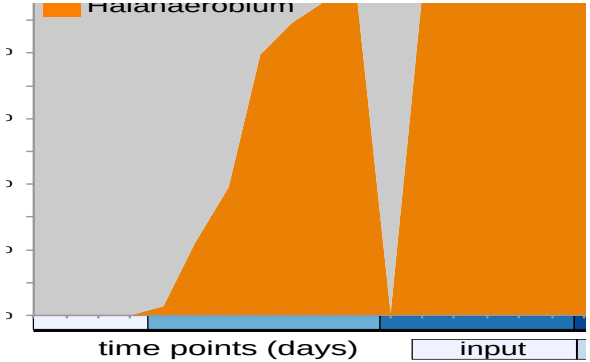


Average copy number for each phylum. Numbers in parentheses indi-

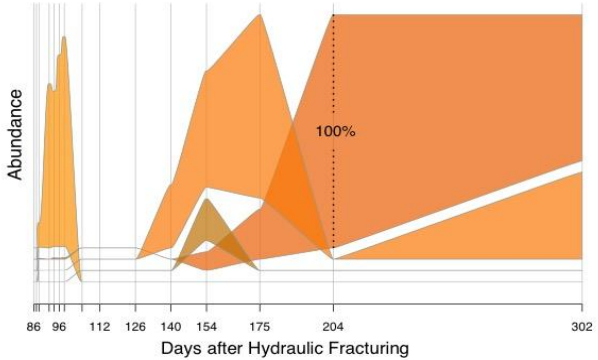


2013 completely sequenced genomes of bacteria and archaea were analyzed

For many organisms, 16S might not provide enough resolution for ecological significance

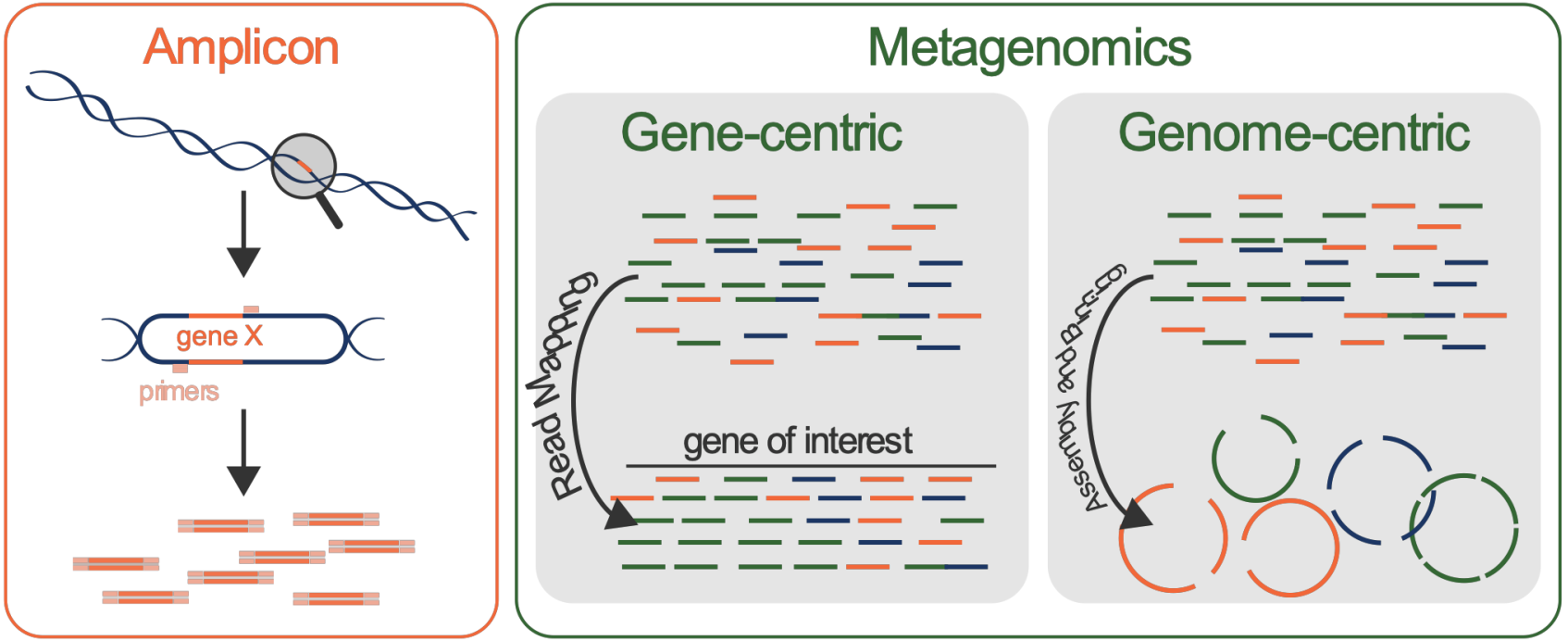


FROM 16S rRNA-full length only 1 ASV

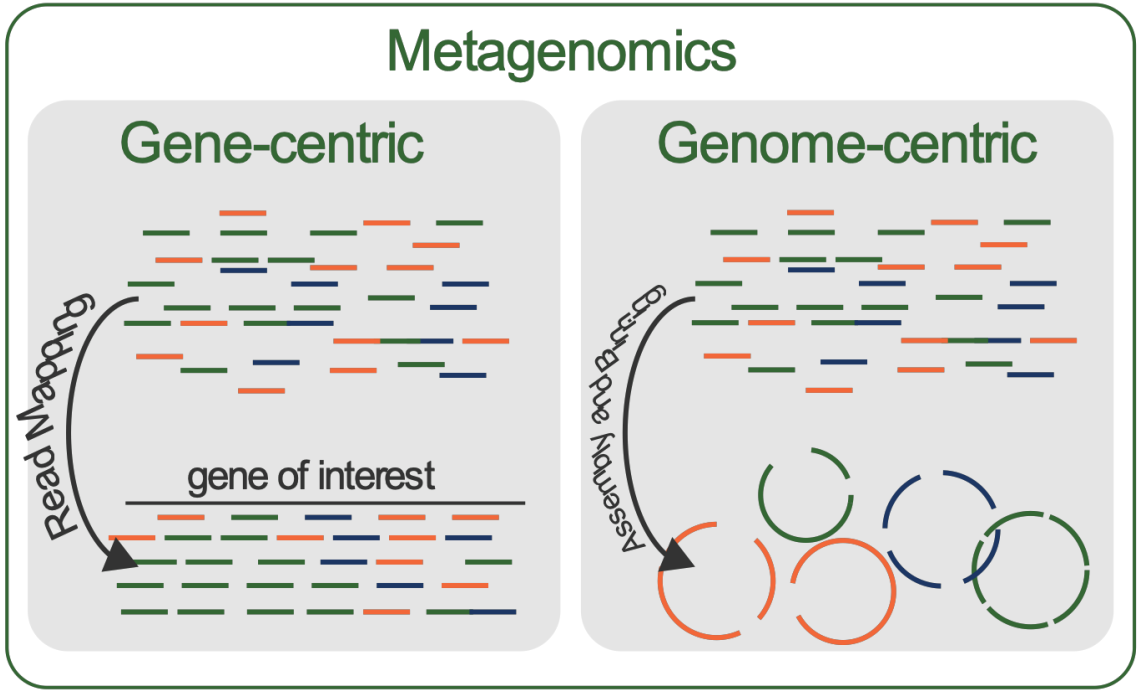


FROM genomes resolved from metagenomes, 5 strains with >99.9% identical 16S rRNA gene

DNA based microbiome tools

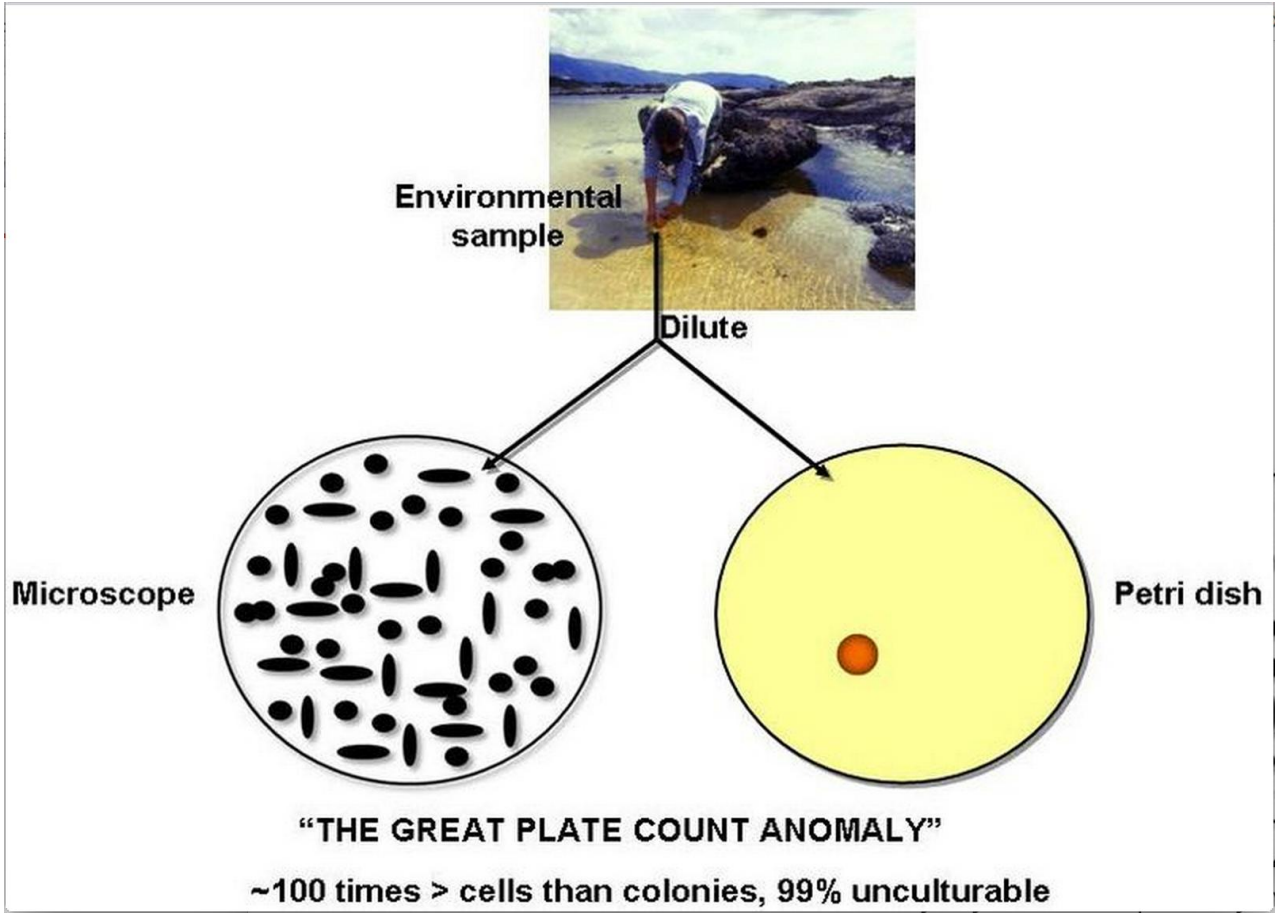


Metagenome- or community genomics- provides organisms in context of their environment

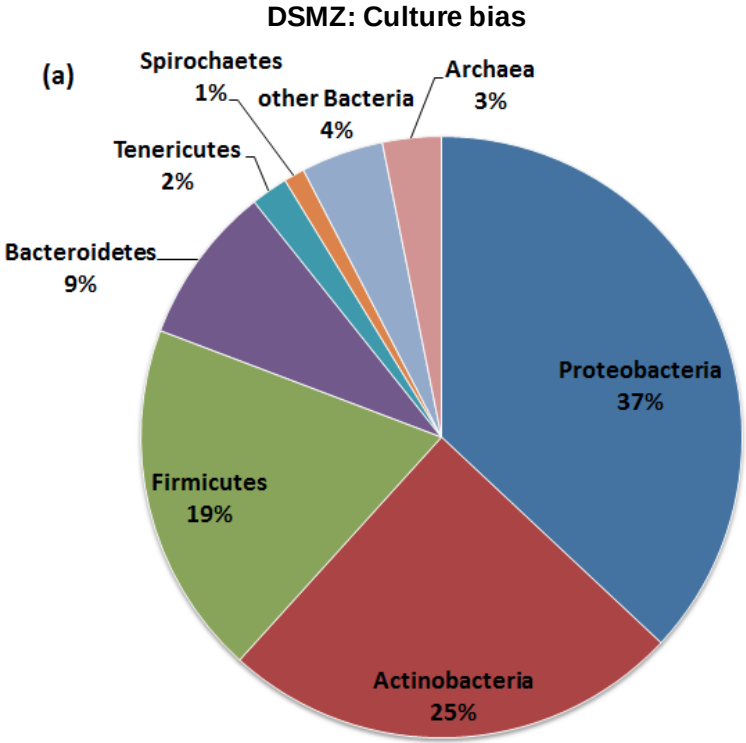


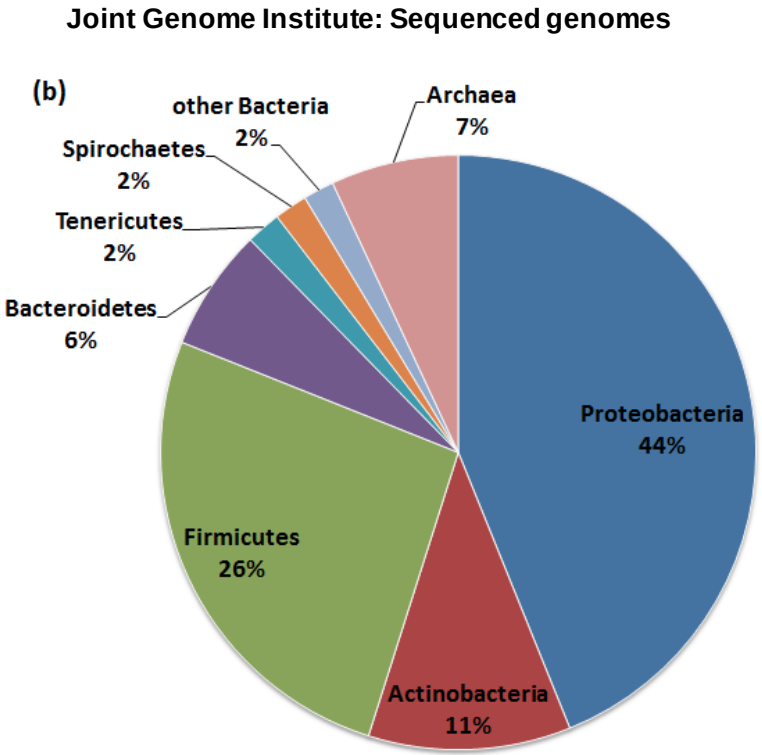
Two methods worth discussing

the microbial world before metagenomics



Genomes used to be only from cultured- bias!





- Four phyla comprise more than 88% of all isolates maintained at DSMZ.

Rinke et al (Nature, 2012)

- Four phyla comprise more than 87.8% of all isolates maintained at DSMZ.

Microbial dark matter: The uncultivated majority

Dark matter is a type of matter hypothesized in astronomy to account for a large part of the mass that appears missing from the universe.

Dark matter cannot be seen directly.

Instead, dark matter existence and properties are inferred.

Microbial dark matter: The uncultivated majority

Because of its ability to reveal the previously hidden diversity of microscopic life, metagenomics offers a powerful lens for viewing the microbial world that has revolutionize understanding of the entire living world

Name that noun and verb pair

DNA is sequenced into ?

are Assembled into **?**

are Binned into **?**

Bins are curated into **Metagenome Assembled Genomes (MAGs)**

Name that noun and verb pair

unbinned

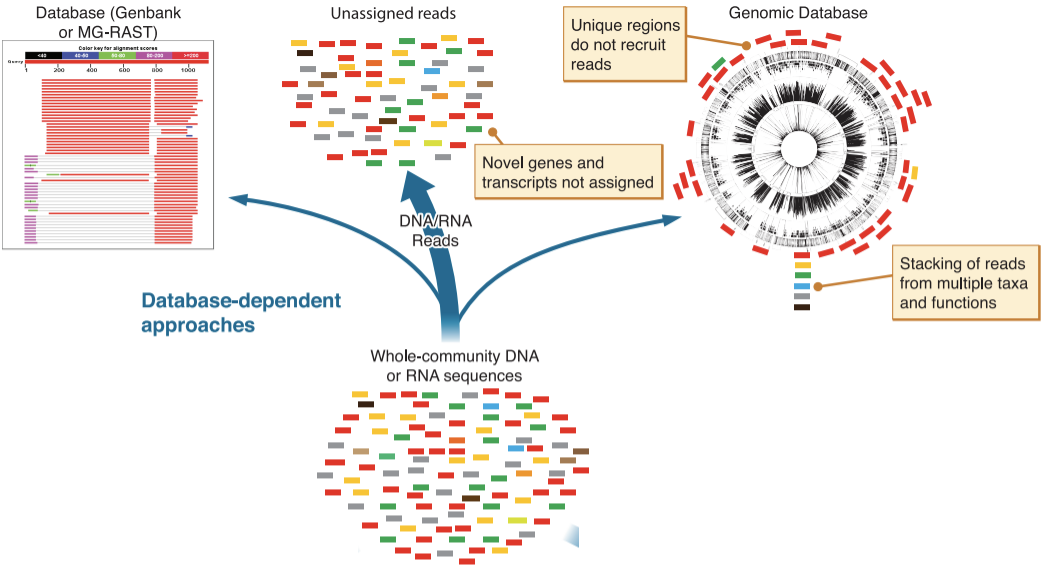
binned

DNA is sequenced into **Reads**

Reads are Assembled into **Contigs**

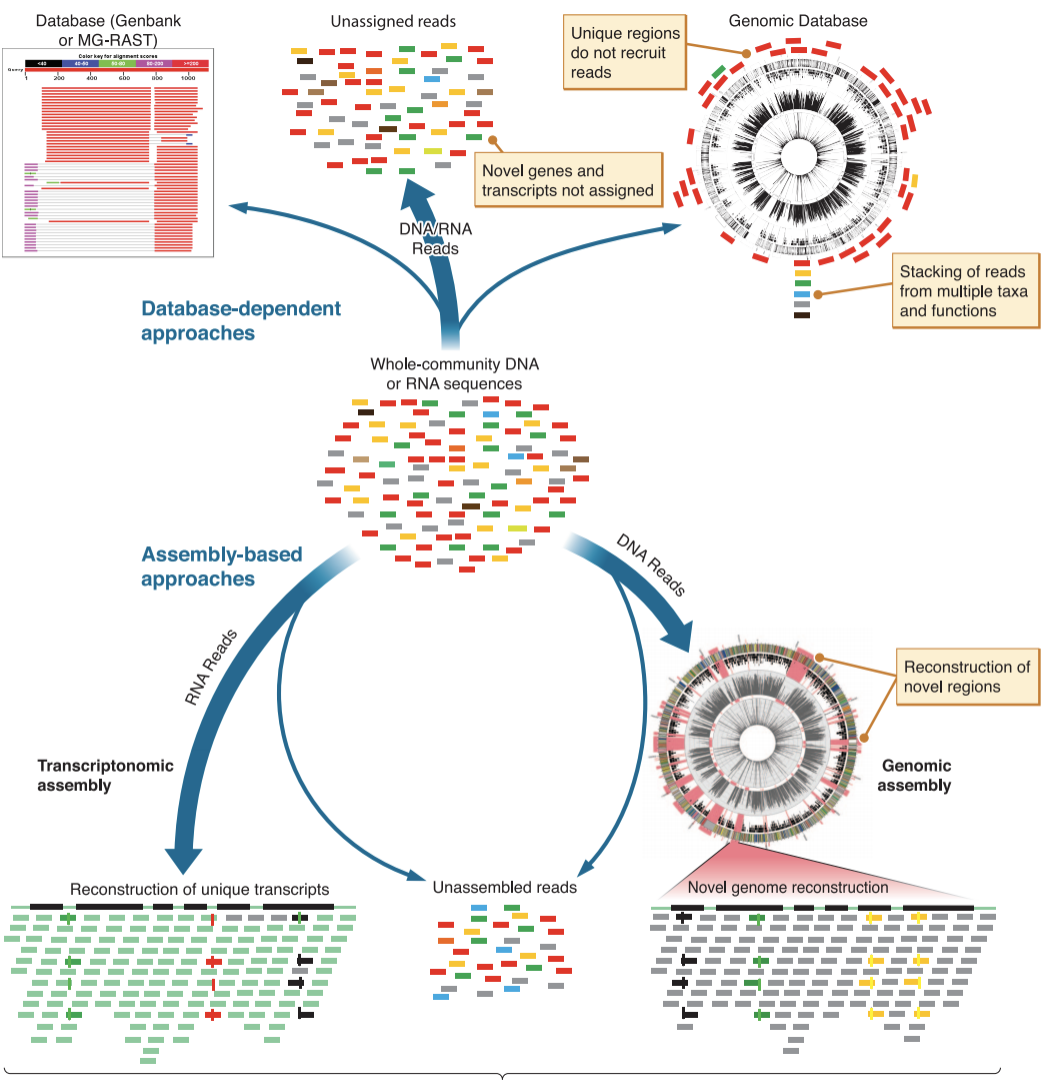
Contigs are Binned into **Bins**

Bins are curated into **Metagenome Assembled Genomes (MAGs)**



1 Unbinned

To bin or not to bin that is the question



1 Unbinned

To bin or not to bin that is the question

2 Binned

❶ Unbinned Venter paper work flow

unbinned

DNA is sequenced into **Reads**

Reads are Assembled into **Contigs**

Contigs are Annotated to **Genes**

Reads are Mapped to contigs for **Coverage**

Unbinned metagenome can uncover new functional diversity in the world around us

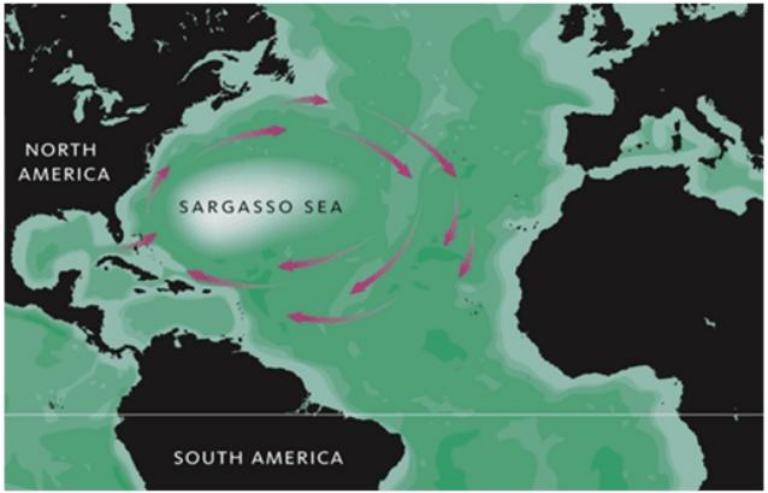
Global ocean sampling expedition (GOS)

RESEARCH ARTICLE

APRIL2004

Environmental Genome Shotgun Sequencing of the Sargasso Sea

Sargasso Sea (from Wikipedia)

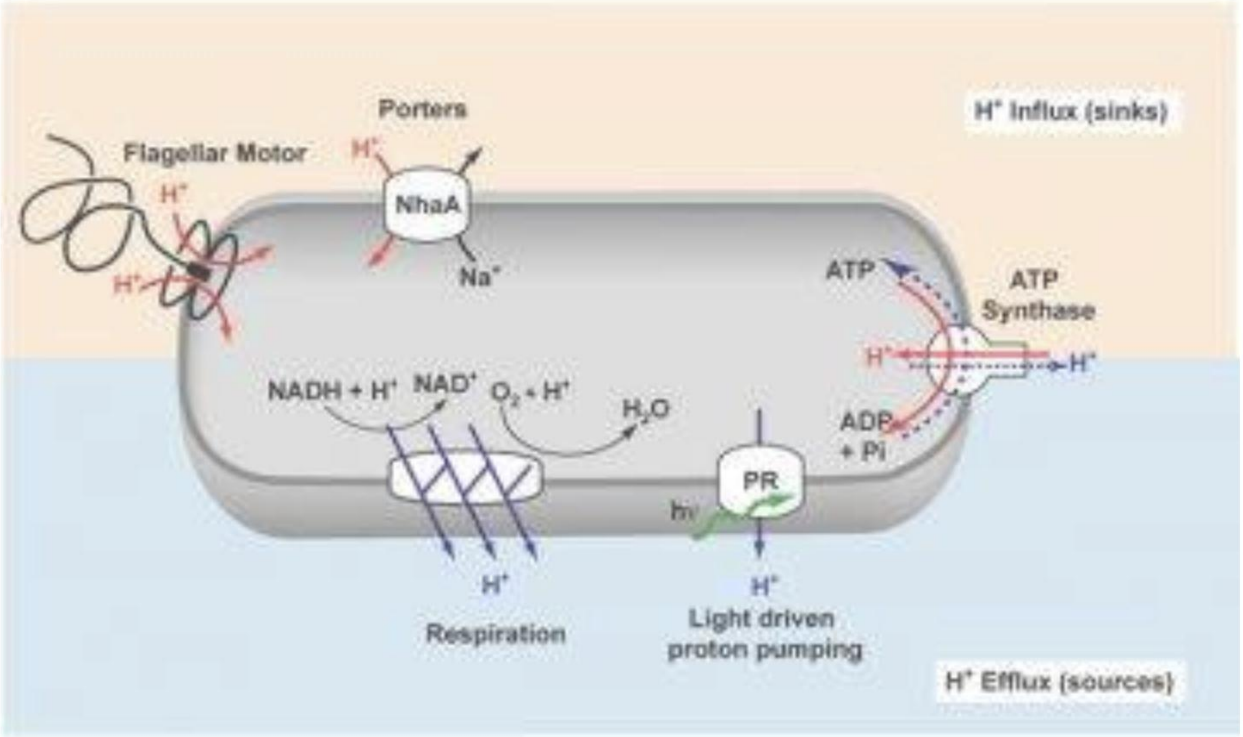


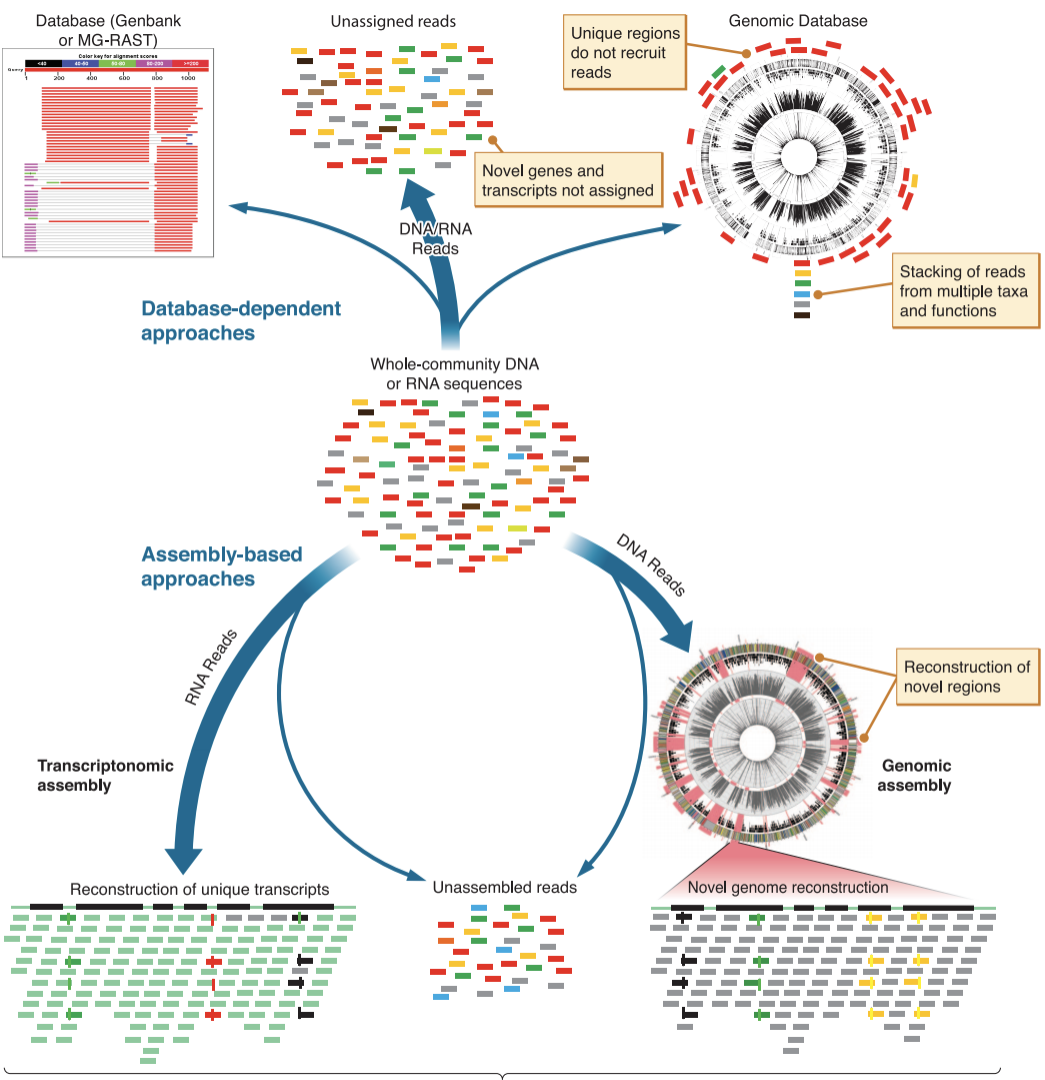
We have applied “whole-genome shotgun sequencing” to microbial populations collected en masse on tangential flow and impact filters from seawater samples collected from the Sargasso Sea near Bermuda. A total of 1.045 billion base pairs of nonredundant sequence was generated, annotated, and analyzed to elucidate the gene content, diversity, and relative abundance of the organisms within these environmental samples. These data are estimated to derive from at least 1800 genomic species based on sequence relatedness, including 148 previously unknown bacterial phylotypes. We have identified over 1.2 million previously unknown genes represented in these samples, including more than 782 new rhodopsin-like photoreceptors. Variation in species present and stoichiometry suggests substantial oceanic microbial diversity.

New genes recovered- expand gene diversity by sampling the uncultivated

- Mind blowing compared to single gene approaches!!!
- A total of 69,901 novel genes belonging to 15,601 clusters were identified
- 782 Proteorhodopsin genes
  - Increase number order magnitude (blue previously know)
  - Increase diversity into uncultivated lineages (from 4 to 13 gene families)

Challenge- assign these genes to organisms or provide overall organismal context





1 Unbinned

To bin or not to bin that is the question

2 Binned

2 **Binned**

Banfield lab work flow

binned

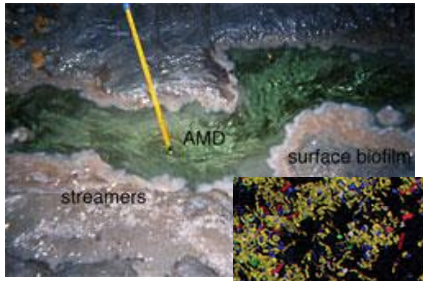
DNA is sequenced into **Reads**

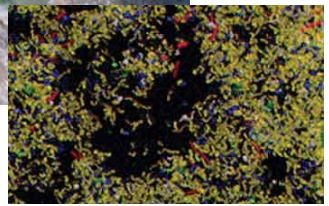
Reads are Assembled into **Contigs**

Contigs are Binned into **Bins**

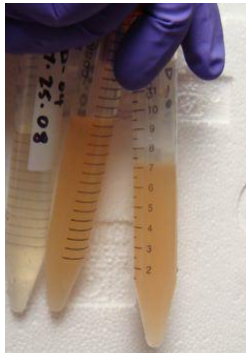
Bins are curated into **Metagenome Assembled Genomes (MAGs)**

1. Environmental sample





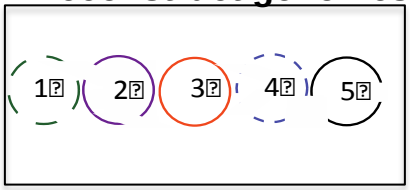
2. Get microbial biomass



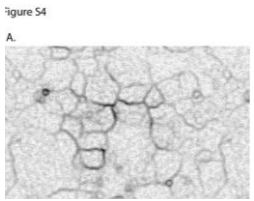
3. Extract genomic DNA

4. Sequence

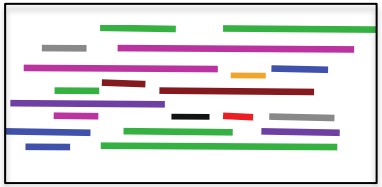
7. Reconstruct genomes



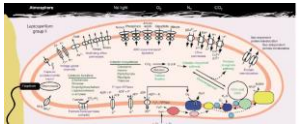
6. Bin

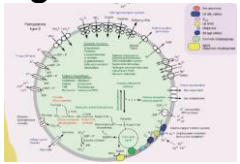


5. Assemble



8. Phylogeny and Infer organism metabolism





Genomics meet biogeochemistry

Pyrite

Fe(II)
Ferrous sulfate

Sulfuric acid
Drops pH

1. $2FeS_2 + 7O_2 + 2H_2O \rightarrow 2FeSO_4 + 2H_2SO_4$

Abiotic reaction

2. $2Fe^{2+} + 1/2\ O_2 + 2H^+ \rightarrow 2Fe^{3+} + H_2O$

Microbial aerobic ferrous iron oxidation

3. $Fe^{3+} + 3H_2O \rightarrow Fe(OH)_{3(s)} + 3H^+$

Abiotic precipitation reaction

Iron hydroxide waste stream- causes additional leaching of Cu, Ca, radioactive metals

MAGs from AMD biofilms

Use 16S FISH and metagenomics to reconstruct genomes to answer these questions in a new ecosystem:

- Who is there?
- How do they grow in the pH 1.5 heavy metal fluid?
- How do they change over time?
- How do the organisms interact?

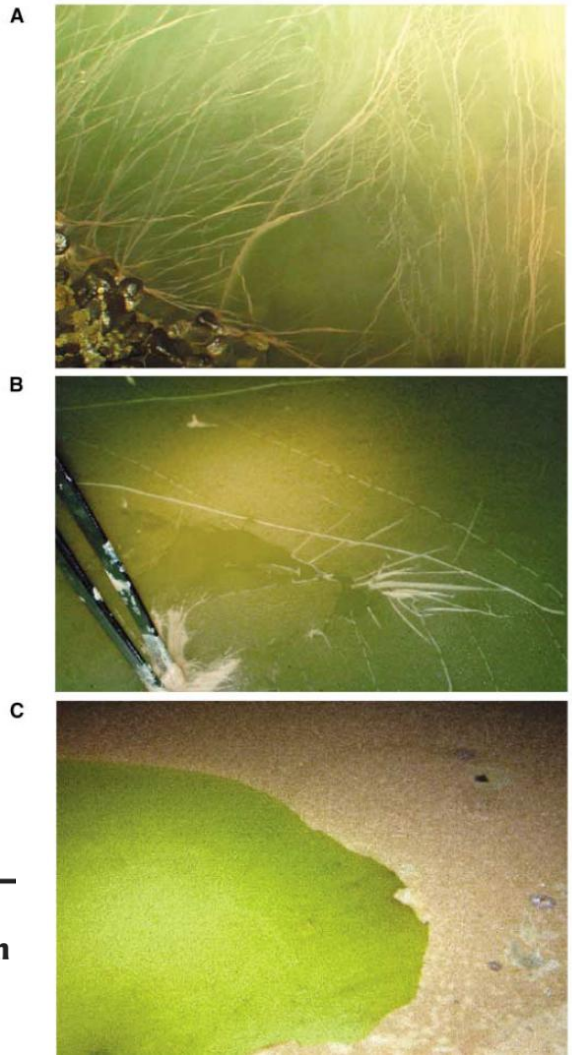


Fig. 1. (A) Photograph of the biofilm during collection in January 2004. The biofilm occurs as a continuous sheet over the surface of the AMD pool; wrinkles form because of movement of the solution. [Photograph taken from the AB end location (fig. S1).] (B) Close-up photograph during sample collection showing that the biofilm is thin and apparently homogeneous. (C) A thicker biofilm in the same location 5 months later, which suggests that the initial biofilm was actively growing when sampled. [Photographs by T. Johnson]

FEB 2004

nature

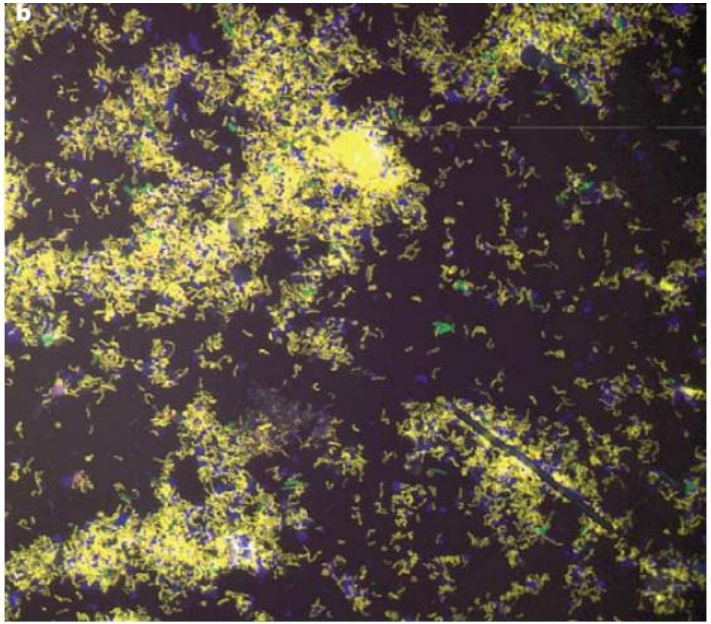
Article | Published: 01 February 2004

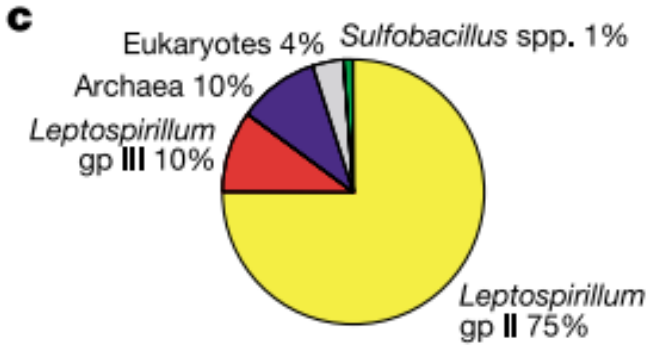
Community structure and metabolism through reconstruction of microbial genomes from the environment

Multiple 16S rRNA approaches:

amplicon barcoding (clone libraries are PCR based) and FISH

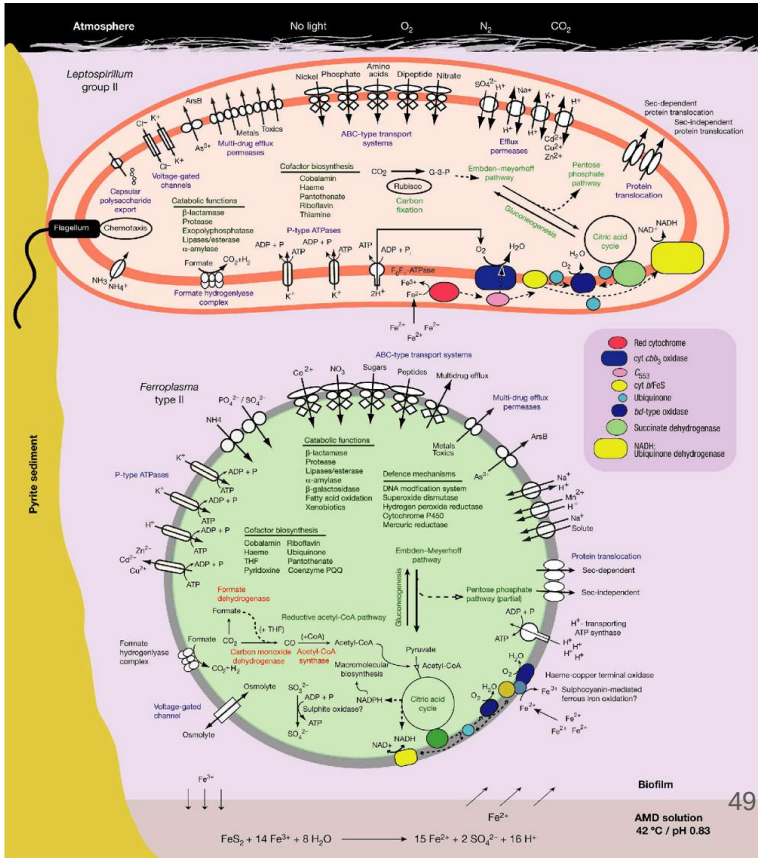
FISH: to quantify organisms in a community





Binned community genome: Metabolic profiling at organismal (genome) level

- Near complete genomes for the five dominant members of the biofilm community
    - Dominant Lepto group II
    - How do we assess genome quality? How complete the MAGS are?



MAG quality standards – Standardizing genome completion and quality

- High, medium, low
  - Estimated Completion and contamination
- Inventory Single copy genes- CheckM or other method
- Pull rRNA from MAG (5, 16, 30S)
- tRNA profile
- Contamination less than 10%

Binned community genome: Metabolic profiling at organismal (genome) level

- Showed only one organism had capacity for nitrogen fixation, thus supplied nitrogen to rest of community

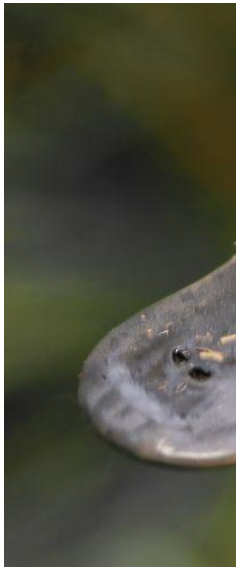
Genome-Directed Isolation of the Key Nitrogen Fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an Acidophilic Microbial Community

Analysis of assembled random shotgun sequence data from a low-diversity, subsurface acid mine drainage (AMD) biofilm revealed a single *nif* operon. This was found on a genome fragment belonging to a member of *Leptospirillum* group III, a lineage in the *Nitrospirae* phylum with no cultivated representatives. Based on the prediction that this organism is solely responsible for nitrogen fixation in the community, we pursued a selective isolation strategy to obtain the organism in pure culture. An AMD biofilm sample naturally abundant in *Leptospirillum* group III cells was homogenized, filtered, and serially diluted into a nitrogen-free liquid medium. The resulting culture in the terminal dilution grew autotrophically to a maximum cell density of $\sim 10^6$ cells/ml, oxidizing ferrous iron as the sole energy source. 16S rRNA-internal transcribed spacer region clone library analysis confirmed that the isolate is a member of *Leptospirillum* group III and that the culture is axenic. We propose the name *Leptospirillum ferrodiazotrophum* sp. nov. for this iron-oxidizing, free-living diazotroph. This study highlights how environmental sequence data can provide insights for culturing previously uncultured microorganisms.

Challenges of binned approach: then and now

- Computationally intensive- most assemblers require high memory
- AMD is a 5 member community, much more tractable for binning, could it scale to more complex communities?
- Workflows today are fairly well established, but at the time were using or modifying single genome tools
- How much sequencing is enough? How much of your reads go into your assembly? How much of your assembly goes into your bins? How well do your bins recruit your meta-T

Partially complete genomes are challenging



1 complete duck contains:
2 legs
2 wings
tail
beak

bin contains:
4 legs
0 wings
tail
beak

contaminated
Incomplete
But likely a duck

Partially complete genomes are challenging

1 complete duck contains:
2 legs
2 wings
tail
beak

bin contains:
4 legs
0 wings
tail
beak

contaminated
Incomplete
But likely a duck

Metagenome assembled genomes (MAGs) standards

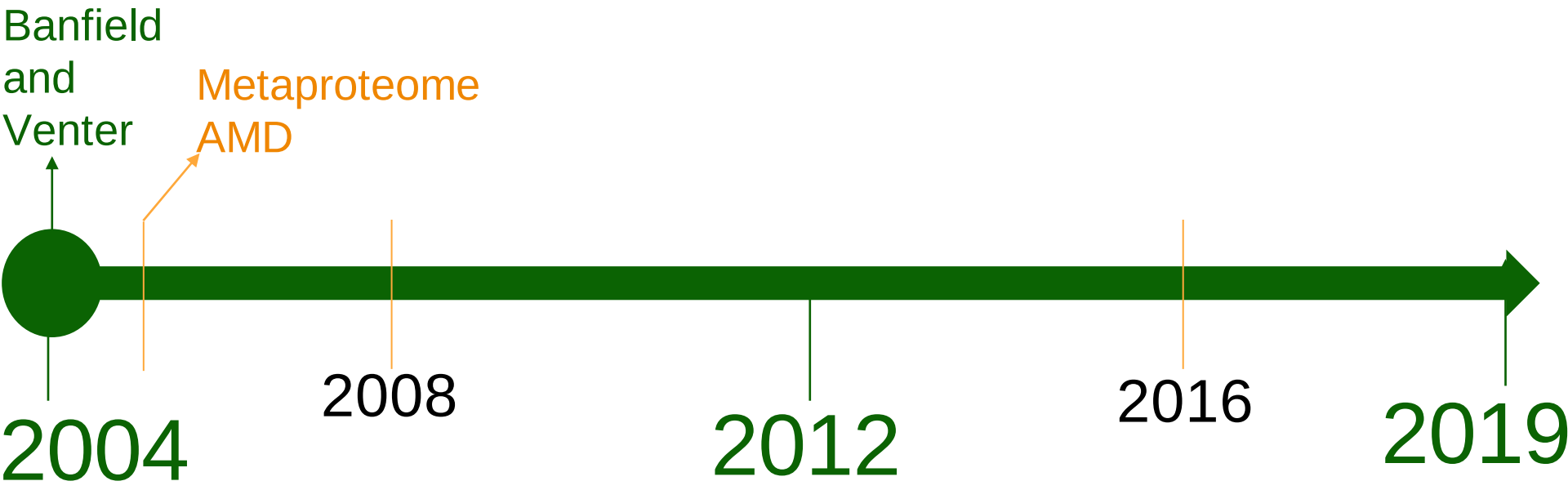
nature biotechnology

Perspective | Open Access | Published: 01 August 2017

Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea

We present two standards developed by the Genomic Standards Consortium (GSC) for reporting bacterial and archaeal genome sequences. Both are extensions of the Minimum Information about Any (x) Sequence (MIxS). The standards are the Minimum Information about a Single Amplified Genome (MISAG) and the Minimum Information about a Metagenome-Assembled Genome (MIMAG), including, but not limited to, assembly quality, and estimates of genome completeness and contamination. These standards can be used in combination with other GSC checklists, including the Minimum Information about a Genome Sequence (MIGS), Minimum Information about a Metagenomic Sequence (MIMS), and Minimum Information about a Marker Gene Sequence (MIMARKS). Community-wide adoption of MISAG and MIMAG will facilitate more robust comparative genomic analyses of bacterial and archaeal diversity.

Greatest hit Microbial Community Genomics



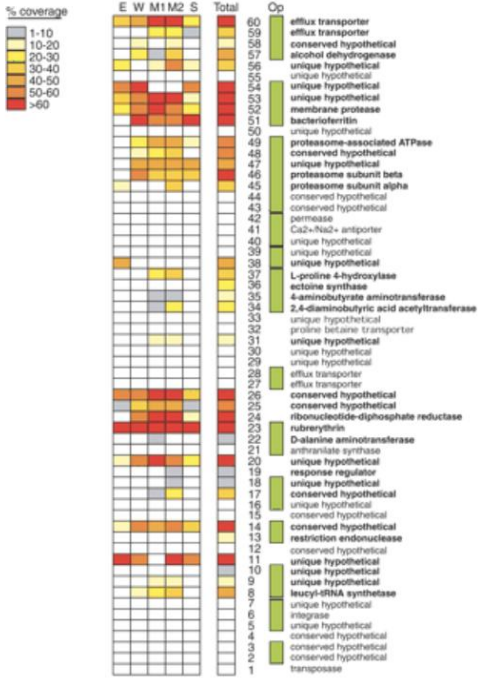


Fig. 5.

Characterization of a genome fragment using the proteome dataset. The diagram shows the annotation, putative operon (Op) structure, and gene number on *Leptospirillum* group II scaffold 21. If the protein encoded by a gene was confidently detected (i.e., matching of two or more peptides), its annotation is in bold type. Colored boxes convey the percentage of each protein detected via MS in extracellular (E), whole-cell (W), membrane (M1 and M2), and cytoplasmic (S) fractions, as well as in the combined biofilm fractions (T). Membrane fractions were prepared by using two different protocols (***8***).

REPORT

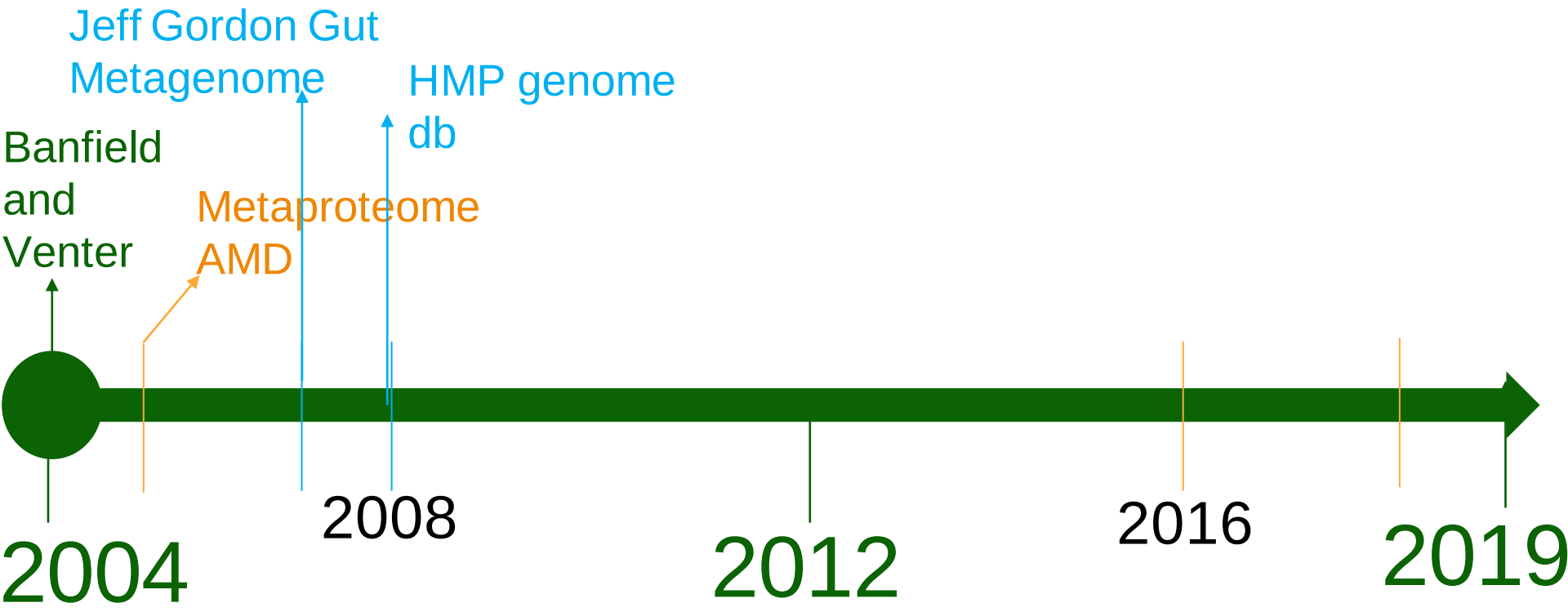
Community Proteomics of a Natural Microbial Biofilm

Rachna J. Ram[1], Nathan C. VerBerkmoes[3,4], Michael P. Thelen[1,6], Gene W. Tyson[1], Brett J. Baker[2], Robert C. Blake II[7], Manes…

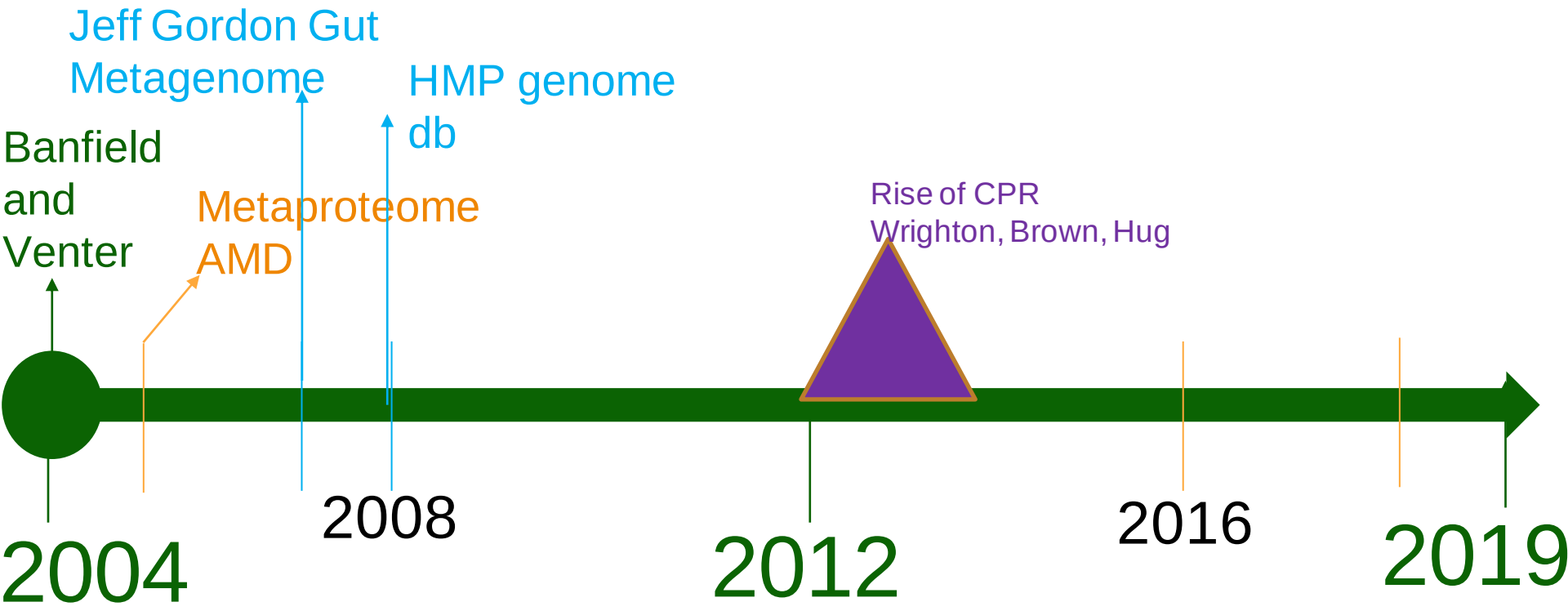
+ See all authors and affiliations

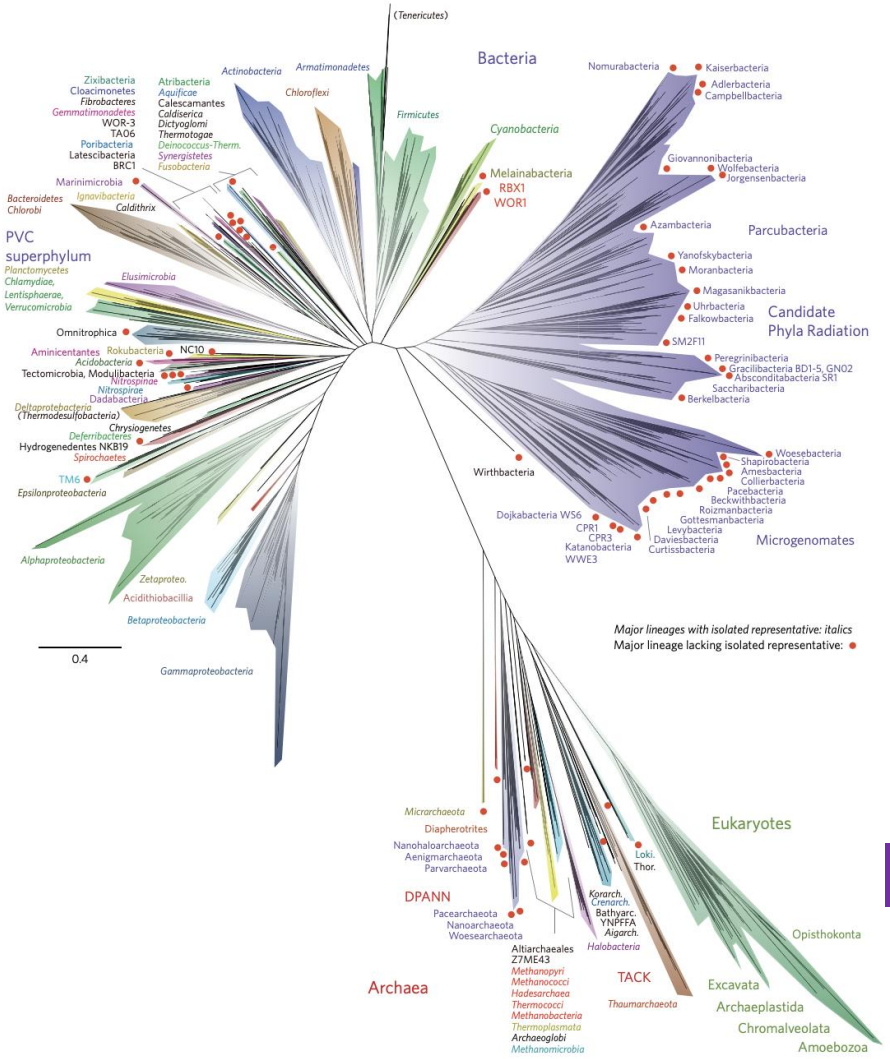
Science 24 Jun 2005:
Vol. 308, Issue 5730, pp. 1915-1920
DOI: 10.1126/science. 1109070

Greatest hit Microbial Community Genomics



Greatest hit Microbial Community Genomics





2017

nature microbiology

LETTERS

PUBLISHED: 11 APRIL 2016 | ARTICLE NUMBER: 16048 | DOI: 10.1038/NMICROBIOL.2016.48

OPEN

A new view of the tree of life

Laura A. Hug[1†], Brett J. Baker[2], Karthik Anantharaman[1], Christopher T. Brown[3], Alexander J. Probst[1], Cindy J. Castelle[1], Cristina N. Butterfield[1], Alex W. Hernsdorf[3], Yuki Amano[4], Kotaro Ise[4], Yohey Suzuki[5], Natasha Dudek[6], David A. Relman[7,8], Kari M. Finstad[9], Ronald Amundson[9], Brian C. Thomas[1] and Jillian F. Banfield[1,9*]

2015

nature

Letter | Published: 15 June 2015

Unusual biology across a group comprising more than 15% of domain Bacteria

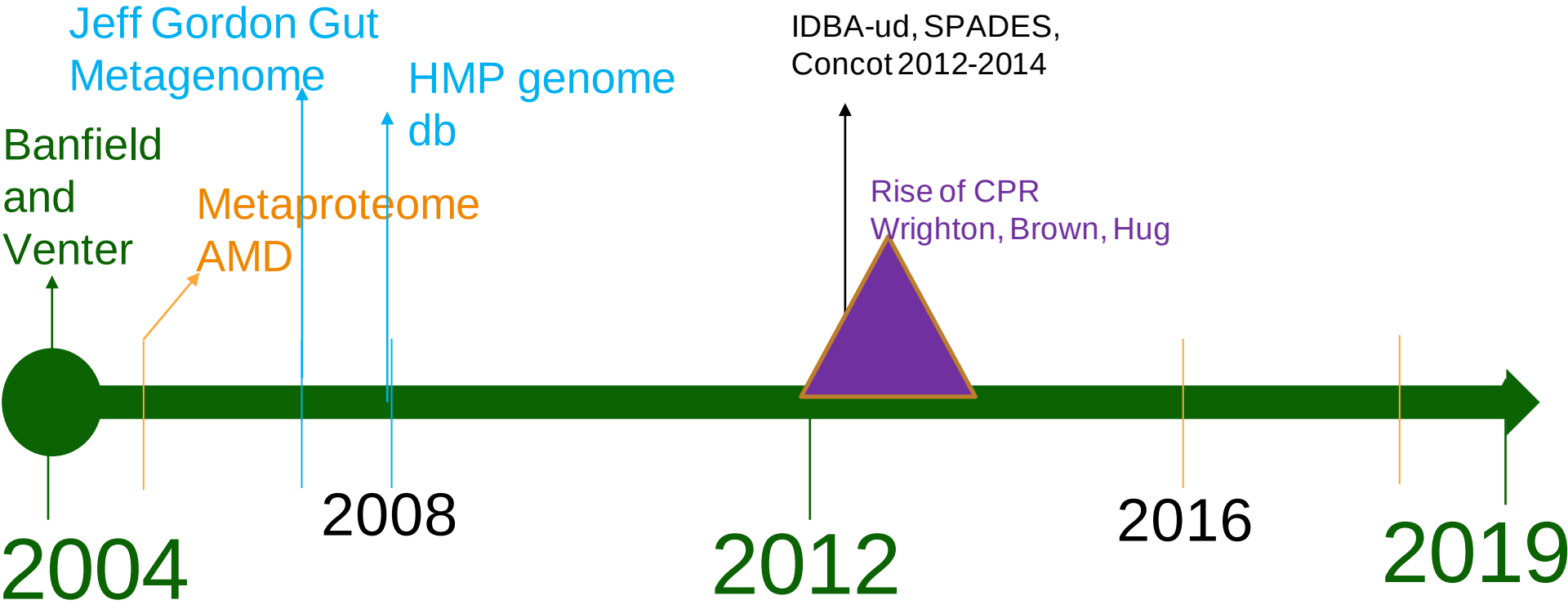
2012

REPORT

Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla

Kelly C. Wrighton[1], Brian C. Thomas[1], Itai Sharon[1], Christopher S. Miller[1], Cindy J. Castelle[2], Nathan C. VerBerkmoes[3],

Greatest hit Microbial Community Genomics



Bioinformatics. 2012 Jun 1;28(11):1420-8. doi: 10.1093/bioinformatics/bts174. Epub 2012 Apr 11. Sign in

IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth.

Peng Y[1], Leung HC, Yiu SM, Chin FY.

2012

2012

SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing

2015

Anton Bankevich, Sergey Nurk, [...], and Pavel A. Pevzner

nature methods

2014

Brief Communication | Published: 14 September 2014

Binning metagenomic contigs by coverage and composition

Johannes Alneberg, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z Ijaz, Leo Lahti, Nicholas J Loman, Anders F Andersson & Christopher Quince

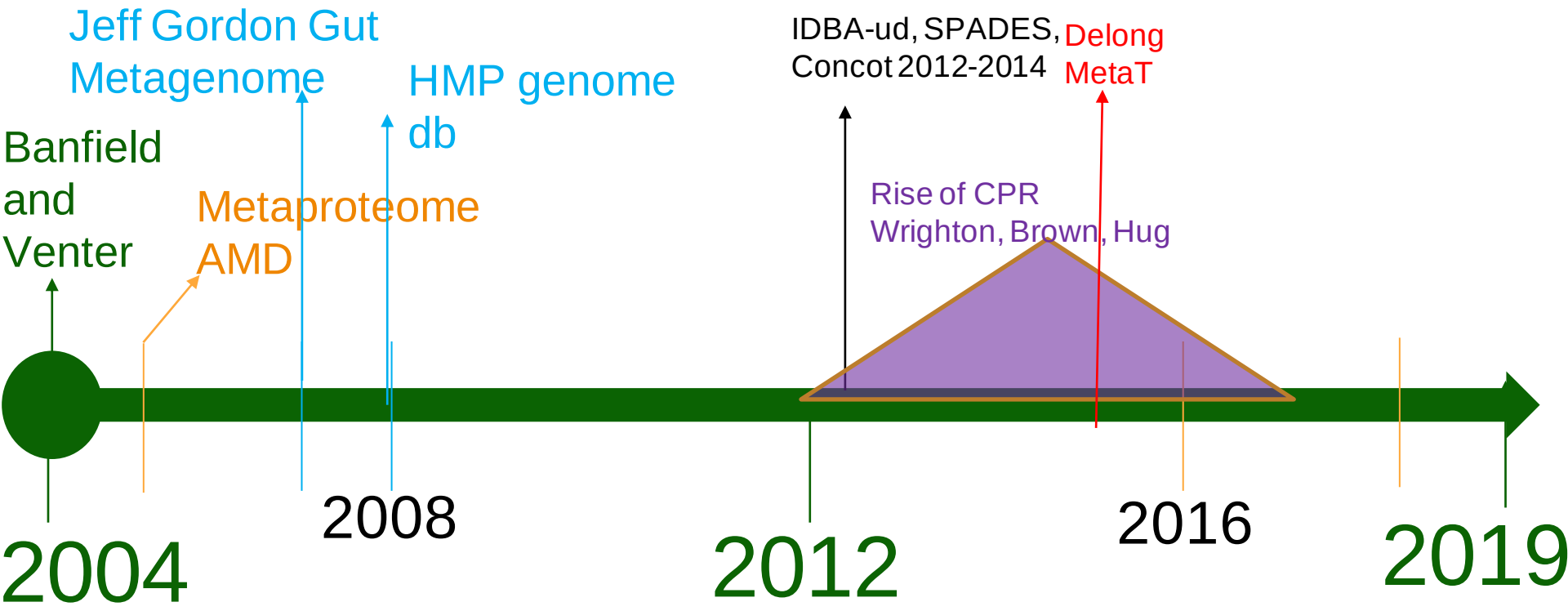
Bioinformatics. 2015 May 15;31(10):1674-6. doi: 10.1093/bioinformatics/btv033. Epub 2015 Jan 20. Sign in

MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.

2015

Li D[1], Liu CM[1], Luo R[1], Sadakane K[1], Lam TW[2].

Greatest hit Microbial Community Genomics



PNAS

Microbial community transcriptional networks are conserved in three domains at ocean basin scales

Frank O. Aylward[a], John M. Eppley[a], Jason M. Smith[b], Francisco P. Chavez[b], Christopher A. Scholin[b], and Edward F. DeLong[a,c,d,1]

[a]Daniel K. Inouye Center for Microbial Oceanography Research and Education (C-MORE), University of Hawaii at Manoa, Honolulu, HI 96822; [b]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039; [c]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; and [d]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Edward F. DeLong, February 13, 2015 (sent for review December 17, 2014; reviewed by Jeffrey I. Gordon, Curtis Huttenhower, and Thomas Schmidt)

nature microbiology

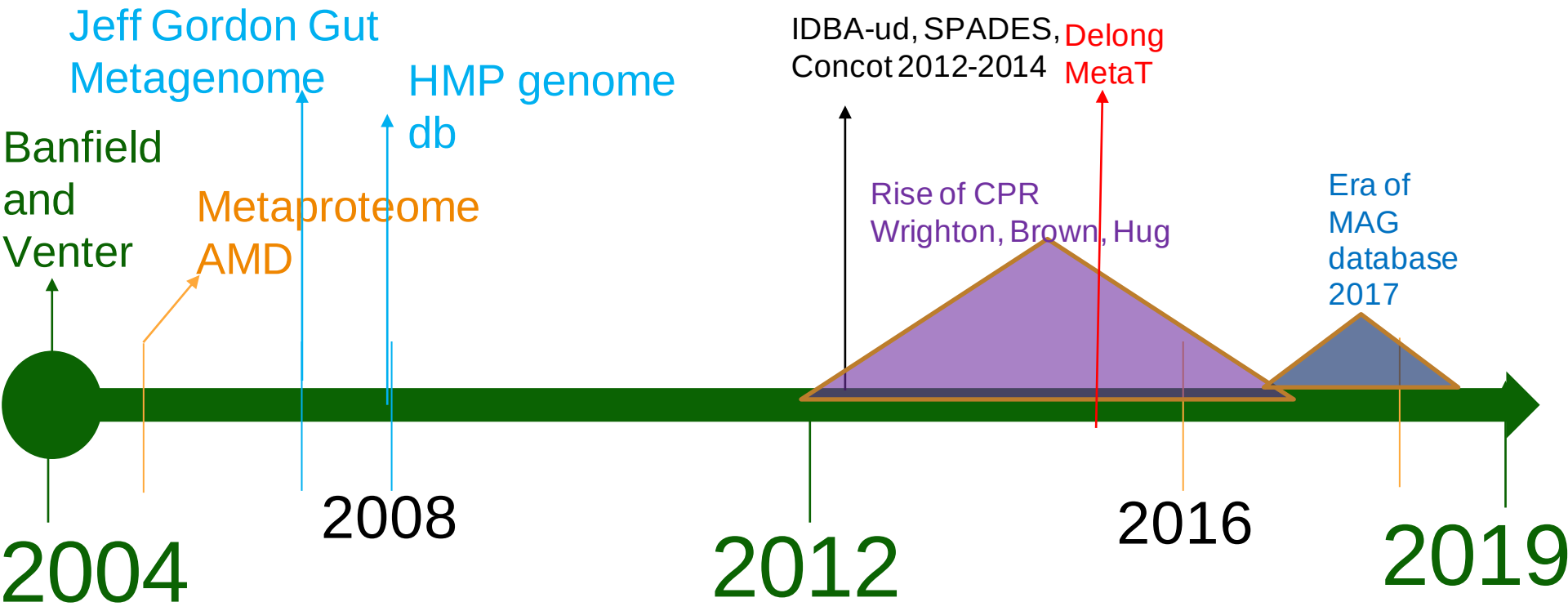
Article | Published: 08 January 2018

Dynamics of metatranscription in the inflammatory bowel disease gut microbiome

Melanie Schirmer, Eric A. Franzosa, Jason Lloyd-Price, Lauren J. McIver, Randall Schwager, Tiffany W. Poon, Ashwin N. Ananthakrishnan, Elizabeth Andrews, Gildardo Barron, Kathleen Lake, Mahadev Prasad, Jenny Sauk, Betsy Stevens, Robin G. Wilson, Jonathan Braun, Lee A. Denson, Subra Kugathasan, Dermot P. B. McGovern, Hera Vlamakis, Ramnik J. Xavier ✉ & Curtis Huttenhower ✉

Nature Microbiology **3**, 337–346(2018) | Cite this article

Greatest hit Microbial Community Genomics



nature biotechnology

Resource | Open Access | Published: 02 August 2019

Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery

Robert D. Stewart, Marc D. Auffret, Amanda Warr, Alan W. Walker, Rainer Roehe & Mick Watson ✉

Nature Biotechnology **37**, 953–961(2019) | Cite this article

nature microbiology

Article | Open Access | Published: 11 September 2017

Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life

Donovan H. Parks, Christian Rinke, Maria Chuvochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz ✉ & Gene W. Tyson ✉

nature

Article | Open Access | Published: 11 February 2019

A new genomic blueprint of the human gut microbiota

Alexandre Almeida ✉, Alex L. Mitchell, Miguel Boland, Samuel C. Forster, Gregory B. Gloor, Aleksandra Tarkowska, Trevor D. Lawley & Robert D. Finn ✉

Nature **568**, 499–504(2019) | Cite this article

CORRECTED PROOF

GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database

Pierre-Alain Chaumeil ✉, Aaron J Mussig, Philip Hugenholtz, Donovan H Parks ✉

Bioinformatics, btz848, https://doi.org/10.1093/bioinformatics/btz848

Published: 15 November 2019 **Article history** ▾

Some questions- choose a method- many work 16S rRNA, unbinned, binned

- If you wanted to address the who is there and relative changes in diversity across thousands of samples?

- If you wanted to rapidly profile the nitrogen fixing capacity in a community, and account for the recovery of novel genes (hint may not be picked up by primers)

- If you wanted to profile metabolic interdependencies between different organisms

- If you wanted to profile the archaeal and bacterial dsDNA viruses in your community

Terminology review: Word choice matters

- **Microbiota** [NOT MICROFLORA]: assemblage of organisms present in a defined environment
- **16S rRNA gene analyses** [NOT METAGENOMICS]: Survey of 16S rRNA genes present in an environment
- **Metagenome**: collection of genomes and genes from the microbiota
- **Microbiome**: entire habitat, including microbes (bacteria, archaea, euks, viruses), their genomes, and surrounding environment can include metabolome or chemistry data too.

From Microbiome Journal

Session 1 Learning objectives Let's Review

- What is microbial dark matter, why does it exist?
- What is metagenomics and how has it changed our understanding of microbial diversity (to uncover dark matter_
- What is the difference between 16S and metaG when to use each method
- What is the value of the unbinned metagenome approach- know example
- What is the value of the binned metagenome approach- know example
- Since these original papers, where is the field today
- [Fun discussion- where are we going? Areas for improvement? Challenges?]