OUR MICROBIAL PLANET

MICROBES—life forms too tiny to see—play a surprisingly large role in life on Earth. Microbes are everywhere, and they do a lot of good for human health and our planet. In fact, disease-causing microbes make up only a very tiny fraction of the millions of types of microbes. Microbes...

Keep us healthy. Amazingly, only about 1 out of 10 cells in the human body is actually a human cell; most of the cells in our bodies are microbes! Some of the microbes living in our bodies actually help us fight disease-causing bacteria by competing against them for space. This mutually beneficial relationship helps to protect us from getting diseases while giving the “good” microbes a place to live.

Keep our environment clean. Because of their special adaptations, some microbes can help clean up gasoline leaks, oil spills, sewage, nuclear waste, and many other types of pollution.

Make air breathable. Without microbes, we wouldn’t have oxygen to breathe. This is because many microbes are photosynthetic—like plants, they harvest energy from the sun, releasing oxygen into the air. Billions of years ago, photosynthetic microbes gradually added oxygen to Earth’s atmosphere, making it possible for larger forms of life—including humans—to live.

Support and protect crops. Microbes living in soil help protect plants from pests and diseases. They also are essential for converting nitrogen and other nutrients into forms that plants can use to grow.

Provide sources of new medicines. Hundreds of medicines available today were derived from chemicals first found in microbes. Microbes naturally produce an amazing variety of chemicals, which scientists can use to create new medicines.

Help us digest food. Many of the foods we eat would be indigestible without the 10-100 trillion microbes living within our guts. Microbes also play a major role in creating many of the foods we love, such as cheese, yogurt, and bread.

The science of metagenomics is shedding new light on the microbial world. Scientists estimate that less than 1% of Earth’s millions of microbial species can be grown in the laboratory. Using metagenomics, scientists can now study how whole communities of microbes function without having to grow each species separately—making more microbes accessible to science than ever before.

Visit www.nationalacademies.org/microbes to learn more!

Support for this publication generously provided by the Pew Charitable Trusts and the Gordon and Betty Moore Foundation.

THE NATIONAL ACADEMIES
Advisors to the Nation on Science, Engineering, and Medicine

The National Academies—National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council—are independent, nonprofit institutions that are committed to providing nonpartisan advice to the nation on issues that affect people’s lives across the entire spectrum of science and technology. Their work helps guide informed decisions by people who influence public policy and manage public resources. From the frontiers of space to the depths of the ocean, from the beginning of time to the end, the National Academies are engaged in providing the knowledge people need to ensure a healthy and sustainable future.
METAGENOMICS:
The Genomic Study of Uncultured Organisms

A research field &
A set of research techniques

MERRIAM-WEBSTER
Main Entry: meta-
Function: prefix
Meaning: more comprehensive : transcending
—usually used with the name of a discipline to designate a new but related discipline designed to deal critically with the original one
Metagenomics: A Set of Research Techniques

- Experimental and computational methods
- Overcome diversity and unculturability of most microbes
- Maximize understanding of genetic composition and activities of communities
- Representatively sample complexity that cannot be exhaustively characterized
Metagenomics: A Research Field

- Understand biology at the aggregate level
- Transcend the individual organism
- Describe community function
- Understand networks of genes working across organisms to serve collective functions
- *Systems biology for communities and the biosphere*
The Tyranny of Koch’s Postulates (1884): Proof of Microbial Causation

• The microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy animals.
• The microorganism must be isolated from a diseased organism and grown in pure culture.
• The cultured microorganism should cause disease when introduced into a healthy organism.
• The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.
The Pure-Culture Paradigm

• Axenic cultures are the Gold-Standard for good microbiological practice
• Multiple species in culture = contamination
• “Planktonic” growth in rich media = “natural”
• Slimy complex biofilm = “accident” and “unworthy of scientific study”
• Genomics: reinforced use/need for pure culture
Microbes are traditionally studied by **culture-dependent** methods.

99 - 99.9% of bacteria in most habitats are **not easily cultured**!

“...animalcules were in such enormous numbers, that all the water...seemed to be alive.” — van Leeuwenhoek (1683)
Microbial communities are highly diverse, abundant, and serve important functions in almost every environment.
Highly connected network of microbial communities across habitats

METAGENOMICS: study microbial communities through direct DNA sequencing

Culture-Dependent Methods

Susceptibility Testing

Genome Sequencing

PCR-based Screens

Direct DNA Sequencing

Culture-Independent Methods
The Great Plate-Count Anomaly (Staley and Konopka, 1985)
Molecular Phylogeny: 16s/18s SSU rRNA

Tree of Life

Most Life and Most Biological Evolution is Microbial

- Multicellular lineages (red) rare, not diverse as measured by SSU rRNA
- Most molecular diversity can be found in microbes
- Most (99%+) microbes can’t be cultured: known only from sequences

Figure adapted from Pace 1997 Science 276:734-740.
It’s a Prokaryotic World!!!

• Number of Species: Millions?
• Number of Individuals/Cells: $5 \times 10^{30}$
• Largest reservoirs of biomass:
  – 350-550 Petagrams of Carbon ($1 \text{ Pg} = 10^{15} \text{ g}$)
  – 85-130 Pg of Nitrogen
  – 9-14 Pg of Phosphorous
• Diversity of Habitats: $340^\circ \text{C, radioactive, 6km underground, soil, ocean, terrestrial subsurface}$
• Diversity of Cellular Chemistries
• Environmental Impact: remediation, disease, commensals, energy, food, photosynthesis
Prokaryotic Genomes Are Rapidly Being Sequenced!

- Mar 22\textsuperscript{nd}, 2019 (GOLD, JGI):
  - \(\sim 156,000\) prokaryotic genomes completed or in progress (\(\sim 100,000\) one year ago)

(https://gold.jgi.doe.gov/)
Two Means of Bacterial Evolution

Widespread HGT Can Obscure Phylogenetic Inference

It’s WAY more diverse than we thought

"What's in a name? That which we call a rose
By any other name would smell as sweet."

*Romeo and Juliet (II, ii, 1-2), W. Shakespeare*

- K12: Harmless lab strain
- CTF073: Uropathogenic strain, causes 70–90% of the 7 million cases of acute cystitis and 250,000 cases of pyelonephritis annually in US
- O157:H7: Enterohemorrhagic, 3-5% fatality
- Only 39.2% of their combined (nonredundant) set of proteins actually are common to all three strains!
- SAME SPECIES BY 16S rRNA!
Sequence-based Metagenomics
DNA Sequencing Keeps Getting Cheaper

DNA synthesis costs continue to drop much faster than Moore’s Law.

Note *useful* costs (e.g. finishing a genome) are higher than *raw* costs.

NHGRI; [http://www.genome.gov/sequencingcosts/](http://www.genome.gov/sequencingcosts/)
Required:

MASSIVE COMPUTING RESOURCES
Sample

Extract DNA

PCR Amplify (sample specific barcodes)

Sequence

Shotgun sequencing of metagenome/transcriptome

Total DNA/cDNA

Sequence

16S rRNA gene-based surveys

Who's there?

What are they doing?

Sequence-based Metagenomics
Sequence-based Metagenomics:
16S rRNA based
Characterization of
Microbial Communities
16S sequencing redefined the tree of life

16S rRNA / SSU sequencing

- 16S rRNA forms part of ALL bacterial ribosomes.

- Contains regions of highly conserved and highly variable sequence.

- Variable sequence can be thought of as a molecular “fingerprint” – can be used to identify bacterial genera and species.

- Large public databases available for comparison. – Ribosomal Database Project contains >1.5 million rRNA sequences.

- Conserved regions can be targeted to amplify broad range of bacteria from environmental samples.

- Not quantitative due to copy number variation

Circumvents the need to culture

Alan Walker, Sanger

Erlandsen S L et al. J Histochem Cytochem 2005;53:917-927
Specific hyper-variable regions (tags) are sequenced

<table>
<thead>
<tr>
<th>Region</th>
<th>Position</th>
<th># b.p.</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>69-99</td>
<td>30</td>
</tr>
<tr>
<td>V2</td>
<td>137-242</td>
<td>105</td>
</tr>
<tr>
<td>V3</td>
<td>338-533</td>
<td>195</td>
</tr>
<tr>
<td>V4</td>
<td>576-682</td>
<td>106</td>
</tr>
<tr>
<td>V5</td>
<td>822-879</td>
<td>57</td>
</tr>
<tr>
<td>V6</td>
<td>967-1046</td>
<td>79</td>
</tr>
<tr>
<td>V7</td>
<td>1117-1173</td>
<td>56</td>
</tr>
<tr>
<td>V8</td>
<td>1243-1294</td>
<td>51</td>
</tr>
<tr>
<td>V9</td>
<td>1435-1465</td>
<td>30</td>
</tr>
</tbody>
</table>


16S by NGS experimental workflow

Extract DNA and amplify marker gene tag with barcoded primers

Pool amplicons and sequence tags

Visualize and compare community relationships

Assign millions of tags to “SPECIES” (OTUs)
How do we define a species?

“No single definition has satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species”

Charles Darwin,
On the Origin of Species, 1859
How do we define a species for tag data?

Species concept works for sexually reproducing organisms

- Breaks down when applied to bacteria and fungi
  - Plasmids
  - Horizontal gene transfer
  - Transposons/Viruses

- **Operational Taxonomic Unit (OTU)**
  - An arbitrary definition of a taxonomic unit based on sequence divergence
  - OTU definitions matter
How do we define a species for tag data?

• Search for sequence similarity between 16S/18S variable regions (e.g. V3-V4) or particular genes (e.g rpoB)

• These genes are ‘house-keeping’ genes which are less likely to be involved in horizontal transfer

• However, note that 16S/18S sequences are known to have variable copy numbers which can bias results

www.mlst.net
Binning tags

Tags may be analysed in one of two ways:

• **Composition-based binning**
  • Relies on comparisons of gross-features to species/genus/families which share these features
    – GC content
    – Di/Tri/Tetra/... nucleotide composition (kmer-based frequency comparison)
    – Codon usage statistics

• **Similarity-based binning**
  • Requires that most sequences in a sample are present in a reference database
    – Direct comparison of OTU sequence to a reference database
    – Identity cut-off varies depending on resolution required (though inexact)
      – “Genus” - 90%
      – “Family” - 80%
      – “Species” - 97%
    – Multiple marker genes used for finer sub-strain identification (MLST)
    – Too stringent cut-off selection will lead to excessive diversity being reported
      – Sequencing errors
      – Sample prep issues
Extract DNA and amplify marker gene with barcoded primers

Pool amplicons and sequence

Assign millions of sequences from thousands of communities to OTUs

Visualize and compare community relationships

Assign reads to communities

>GCACCTGAGGACAGGCATGAGGAA...
>GCACCTGAGGACAGGGGAGGAGGA...
>TCACATGAACCTAGGCAGGACGAA...
>CTACCGGAGGACAGGCATGAGGAT...
>TCACATGAACCTAGGCAGGAGGAA...
>GCACCTGAGGACAGGCATGAGGAA...
>CTACCGGAGGACACACAGGAGGAA...
>GAACCTTCACATAGGCAGGAGGAT...
>TCACATGAACCTAGGGGCAAGGAA...
>GCACCTGAGGACAGGCAGGAGGAA...
QIIME2 workflow schematic

**Input**
- Sequencing output (454, Illumina, Sanger)
  - fasta, qual, sff/trace files

**Metadata**
- mapping file

**Pre-processing**
- e.g., remove primer(s), demultiplex, quality filter

**Denoise 454 Data**
- PyroNoise, Denoiser

**MG-RAST Submission**
- Submit sequences and metadata to MG-RAST

**Pick OTUs and representative sequences**
- Reference based:
  - BLAST, UCLUST

  - De novo:
  - e.g., UCLUST, CD-HIT, MOTHUR

**Assign taxonomy**
- BLAST, RDP Classifier

**Align sequences**
- e.g., PyNAST, INFERNAL, MUSCLE, MAFFT

**Build ‘OTU table’**
- i.e., per sample OTU counts

**Build phylogenetic tree**
- e.g., FastTree, RAxML, ClearCut

**OTU Table**
- (i.e., per sample OTU counts)

- α-diversity and rarefaction
  - e.g., Phylogenetic Diversity, Chao1, Observed Species

- β-diversity and rarefaction
  - e.g., Weighted and unweighted UniFrac, Bray-Curtis, Jaccard

**Phylogenetic Tree**
- Evolutionary relationship between OTUs

**Visualization**
- e.g., 2D and 3D PCoA plots, distance histograms, taxonomy pie charts, rarefaction plots, OTU network visualization, jackknifed hierarchical clustering.

**beta_diversity_through_plots.py**

```
$ beta_diversity_through_plots.py -i otu_table.txt -m Fasting_Map.txt -o betadivtree -t rep_set.tre
```

https://qiime2.org/
**dada2**: infers exact amplicon sequence variants (ASVs) instead of OTUs

**Advantages**

**Resolution**: ASVs resolve biological differences of even 1 or 2 nucleotides.

**Accuracy**: fewer false positive sequence variants than other methods report false OTUs.

**Comparability**: ASVs can be directly compared between studies, without reprocessing.

**Computational Scaling**: Compute time scales linearly with sample number

**dada2**: infers exact amplicon sequence variants (ASVs) instead of OTUs

---

**Accuracy: Simulated data**

3% OTUs (average linkage) vs DADA2

https://benjjneb.github.io/dada2/index.html
Measuring diversity of OTUs/ASVs

Two primary measures for sequence based studies:

• Alpha diversity
  – What is there? How much is there?
  – Diversity *within* a sample

• Beta diversity
  – How similar are two samples?
  – Diversity *between* samples
Measuring diversity: alpha

Alpha diversity

• Diversity within a sample
• Simpson’s diversity index (also Shannon, Chao indexes)
• Gives less weight to rarest species

\[ D = 1 - \frac{\sum_{i=1}^{S} n_i(n_i - 1)}{N(N - 1)} \]

- \( S \) is the number of species
- \( N \) is the total number of organisms
- \( n_i \) is the number of organisms of species \( i \)

Measuring diversity: beta

Beta diversity
• Diversity *between* samples
• Sorensen’s index

\[
\beta = \frac{2c}{S_1 + S_2}
\]

- \(S_1\) is the number of species in sample 1
- \(S_2\) is the number of species in sample 2
- \(c\) is the number of species present in both samples

Beta diversity: Unifrac

Beta diversity

- Diversity *between* samples
- Unifrac distance (*shared branch length on phylo tree*)
- Percentage observed branch length unique to either sample

Clustering by Unifrac distance
An early example of 16S community analysis (by 454)
Mammalian Microbiomes Cluster by Diet

Circles are different mammals, colors different categories.
Grey rounded squares are shared OTUs.
Thinner edge = fewer sequences.
Thicker edge = more sequences (shared OTUs only).
Larger nodes = more connections.
Each line indicates that an OTU was found in a given mammal.
Diamonds contain OTUs unique to individual samples (bigger = more OTUs).

Ley et al. 2008 Science 320:1647
...with the carnivores clustering together...
...two types of herbivores, foregut and hindgut fermenters...
...and one group of ominovores...
...that includes humans despite our diverse habitat and diet.
A human gut microbial gene catalogue established by metagenomic sequencing

Human gut microbiome viewed across age and geography

Figure 4 | Bacterial species abundance differentiates IBD patients and healthy individuals. Principal component analysis with health status as

Cohabiting family members share microbiota with one another and with their dogs

Figure 1 | Differences in the fecal microbial communities of Malawians, Amerindians and US children and adults. a, UniFrac distances between
Sequence-based Metagenomics: Shotgun environmental sequencing to characterize genetic diversity: 4+ Examples
Shotgun whole metagenome sequencing to estimate community functions by database comparison

<table>
<thead>
<tr>
<th>Sample</th>
<th>gene A</th>
<th>gene B</th>
<th>gene C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>546</td>
<td>205</td>
<td>0</td>
</tr>
<tr>
<td>Sample 2</td>
<td>0</td>
<td>48</td>
<td>988</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample n</td>
<td>0</td>
<td>700</td>
<td>0</td>
</tr>
</tbody>
</table>
Example 1: Massive Metagenomic Diversity

Environmental Genome Shotgun Sequencing of the Sargasso Sea


• 1.045 Gb sequenced
• 1.8 millions new genes found
• 148 new 16s rRNA sequences
• Sampled ~1800 different species

The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families


• 6.12 million predicted proteins
• **Doubled** total number of known proteins
• Covered ~all prokaryotic protein families
• 1,700 clusters with no homology to known families
• 6,000 previous ORFans have homologs in GOS data
Example 2: Reduced Metagenomic Diversity

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


- Acid-mine drainage biofilm
- 76.2 Mb shotgun sequence
- Bimodal GC distribution
- 16S rRNA: 3 bacterial and 3 archaeal lineages
- Assembled genomes for 5 lineages
- Some lineages are clonal
- Some lineages show extensive recombination
- Harsh environment (acid mine) reduces diversity
Example 3: Single cell sequencing of uncultured organisms

Multiple Displacement Amplification (MDA)

Genome sequencing on single *Prochlorococcus* cells

<table>
<thead>
<tr>
<th></th>
<th>JGI</th>
<th>Agencourt</th>
</tr>
</thead>
<tbody>
<tr>
<td># good seq reads</td>
<td>7,166</td>
<td>10,660</td>
</tr>
<tr>
<td>Average length (bp)</td>
<td>769.4</td>
<td>676.6</td>
</tr>
<tr>
<td>Total length (bp)</td>
<td>5,513,520</td>
<td>7,212,556</td>
</tr>
<tr>
<td># unkown seqs</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td># vectors</td>
<td>23</td>
<td>44</td>
</tr>
<tr>
<td># other seqs</td>
<td>74 (MED4)</td>
<td>2 (MED4)</td>
</tr>
<tr>
<td>% genome sampled</td>
<td>62.47%</td>
<td>66.53%</td>
</tr>
</tbody>
</table>

- “Unsampled” genomic regions were detected by PCR amplification from plones
- Primers can be designed from sequenced plones and metagenomic data

Insights into the phylogeny and coding potential of microbial dark matter

Christian Rinke1, Patrick Schwientek1, Alexander Sczyrba1,2, Natalia N. Ivanova1, Iain J. Anderson1,3, Jan-Fang Cheng1, Aaron Darling1,4, Stephanie Malfatti1, Brandon K. Swan5, Esther A. Gies6, Jeremy A. Dodsworth7, Brian P. Hedlund7, George Tsiamis8, Stefan M. Sievert9, Wen-Tso Liu10, Jonathan A. Eisen3, Steven J. Hallam9, Nikos C. Kyriides1, Ramunas Stepanauskas9, Edward M. Rubin1, Philip Hugenholtz11 & Tanja Woyke1

apply single-cell genomics to target and sequence 201 uncultivated archaeal and bacterial cells from nine diverse habitats belonging to 29 major mostly uncharted branches of the tree of life, so-called ‘microbial dark matter’. With this
Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw

Rachel Mackelprang1,2, Mark P. Waldrop3, Kristen M. DeAngells4, Maude M. David4, Krystle L. Chavarria4, Steven J. B Edward M. Rubin5,6 & Janet K. Jansson2,4


Permafrost contains an estimated 1672 Pg carbon (C), an amount roughly equivalent to the total currently contained within land plants and the atmosphere1–

Sequencing yielded a total of 176 million reads and 39.8 Gb of raw sequence.

transition from a frozen to a thawed state there are rapid shifts in many microbial, phylogenetic and functional gene abundances and

We also construct the first draft genome from a complex soil metagenome, which corresponds to a novel methanogen.

From these data, we identified 27,755 putative carbohydrate-active genes and expressed 90 candidate proteins, of which 57% were enzymatically active.

We also assembled 15 uncultured microbial genomes,
Function-based Metagenomics
The Promise of Metagenomics: Harness and Describe the Microbial “Bucket Brigade”
Functional Metagenomics: Strategy

Considerations:
- **Source:**
  - soil vs marine
  - pristine vs polluted
- **Library size:**
  - small insert plasmid (~5Kb)
  - large insert (fosmid, 40Kb), (BAC, 100Kb)
- **Host:**
  - “Model” organism
  - “Industrial” organism
- **Selection strategy:**
  - Clonal screening vs library selection
  - Liquid vs solid

## Screening (Selecting) Interesting Clones

### Table 4 | Function-driven versus sequence-driven screening strategies

<table>
<thead>
<tr>
<th></th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Function-driven screening method</strong></td>
<td>Completely novel genes can be recovered</td>
<td>Dependent on expression of the cloned genes by the bacterial host</td>
</tr>
<tr>
<td></td>
<td>Selects for full-length genes</td>
<td>Requires production of a functional gene product by the bacterial host</td>
</tr>
<tr>
<td></td>
<td>Selects for functional gene products</td>
<td>Dependent on the design of a simple activity-based screening strategy</td>
</tr>
<tr>
<td><strong>Sequence-driven screening method</strong></td>
<td>Independent of expression of the cloned genes by the bacterial host used</td>
<td>Recovered genes are related to known genes</td>
</tr>
<tr>
<td></td>
<td>Similar screening strategies can be used for different targets, for example,</td>
<td>Partial genes can be cloned</td>
</tr>
<tr>
<td></td>
<td>colony hybridization and PCR</td>
<td>Not selective for functional gene products</td>
</tr>
</tbody>
</table>

(Screen, OR SELECT, based on resistance, biochemical signal, etc.)

(PCR, Hybridization—must know target sequence/gene family of interest)

Sequence-driven Screening of Metagenomic Libraries: 2 Examples
Making Enzyme Chimeras by Degenerate PCR of Functionally Enriched Metagenomes

- **Enrich** microbial communities to degrade phenol or crude-oil
- Amplify central **fragment** of ring-cleavage enzyme (catechol 2,3-dioxygenase) from metagenome by **degenerate PCR**
- **Ligate to conserved** 5’ and 3’ flanking regions of known enzyme
- Screen and analyze hybrids

---

**Okuta et al., Gene (212) 1998**
Making Pathway Hybrids by Degenerate PCR of Metagenomes

- Polyketide synthases: natural assembly lines for antibiotic synthesis
- Pathways appear very modular
- Harness modularity by combinatorial assembly
- Same idea as before: flanking genes in pathway are “conserved” and internal gene(s) can be varied through PCR of metagenomic DNA
- Diversity increased, activity increased

Seow et al., J. Bacteriology (179) 1997
Function-driven Selection of Metagenomic Libraries: Examples from our work (www.dantaslab.org)
What are Antibiotics?

ANTIBIOTICS are chemicals that block the growth of bacteria and fungi.

WARNING: Antibiotics don’t work for viruses like colds and the flu. Using them for viruses will NOT make you feel better or get back to work faster.
ANTIBIOTICS are small molecules which kill (bacteriocidal) or inhibit the growth (bacteriostatic) of bacteria.
Antibiotic Resistant Infections Are A Leading Cause of Death

700,000 deaths in 2014

Deaths attributable to AMR every year by 2050

10 Million estimated deaths in 2050

(UK Prime Minister’s AMR Report, 2014)

Treatment of Antibiotic Resistant Infections Is Expensive

$55 Billion cost to the US economy in 2013

(US CDC, 2013)

$100 Trillion estimated cost to global economy by 2050

(UK Prime Minister’s AMR Report, 2014)

Resistant Infections Are Increasing BUT New Antibiotic Discovery Is Decreasing

(Tschäberle & Hack, 2014)
Horizontal gene transfer enables rapid transfer of resistance.


Enriched for antibiotic resistance genes

Traditional **culture-based methods** underestimate the resistance problem

### 1) Culture-bias vastly underestimates **bacteria**

- **1x vs. 100-1000x**
- **99.0-99.9% of bacteria are not cultured!**

### 2) Culture-bias vastly underestimates **antibiotic resistance**

**Cultured resistance genes**

- **Previously discovered**
- **Not previously discovered**

**Uncultured resistance genes**

- **Human**
  - Sommer, Dantas et al, *Science* 2009
- **Soil**
  - Forsberg...Dantas, *Science* 2012
- **Human**
  - Sommer, Dantas et al, *Science* 2009
- **Soil**
  - Forsberg...Dantas, *Nature* 2014
Methods for studying antibiotic resistance in microbial communities

**KNOWN**

- Clinical resistance levels
- Direct clone to (multidrug) resistance connection
- Culture bias

**SHOTGUN METAGENOME SEQUENCING**

- No culture bias
- Large sampling depth
- Only previously identified genes
- Relative abundance

**FUNCTIONAL METAGENOMIC SELECTIONS**

- No culture bias
- Large sampling depth
- Function confirmed
- Can identify novel genes

**UNKNOWN**

- Known, Readily Cultured
- Known, Not Readily Cultured
- Unknown

**Antibiotic Resistance Reservoir (RESISTOME)**
Functional metagenomic selections identify novel antibiotic resistance genes in microbial communities

Increasing functional metagenomic throughput via next-gen sequencing

**Functional Metagenomics**

1. Extract Total Metagenomic DNA From Sample 1
2. Shear and Select for 2-5 kb Fragments
3. Transform Fragment Library Into Indicator Strain
4. Functional Selection on plates with antibiotic concentration toxic to wild-type host
5. Barcode Selected Fragments

Sample 2

Sample n


Increasing functional metagenomic throughput via next-gen sequencing

Functional Metagenomics

1. Extract Total Metagenomic DNA From Sample 1
2. Shear and Select for 2-5 kb Fragments
3. Transform Fragment Library Into Indicator Strain
4. Functional Selection on plates with antibiotic concentration toxic to wild-type host
5. Barcode Selected Fragments
6. Pool Selected Fragments

PARFuMS: Parallel Annotation and Reassembly of Functional Metagenomic Selections

1. Annotate Functions
2. Long Read Assembly
   - Phrap
3. Short Read Assembly
   - Velvet
   - Reads Binned by Barcode

Next-Generation Sequencing

Illumina


ShortBRED: improved annotation accuracy from short reads

1. Create a short protein marker database for functionally selected antibiotic resistance genes (ARGs)

   ![Diagram of ARGs, Reference database of bacterial proteins, Cluster proteins into families, Identify unique protein sequences (markers) for each family of interest, True Markers: no overlap with any reference sequence or consensus, Junction Markers: minor overlap with any reference sequence or consensus, Quasi Markers: least overlap possible; overlapping families are merged.]

2. Quantify relative abundance of antibiotic resistance genes (ARGs) in whole metagenome shotgun data

   ![Diagram of mapping short reads to unique markers using USEARCH, Normalize results to produce a relative abundance profile for ARGs across samples.]

Culture-independent methods for measuring microbial community response to antibiotics

1. Measure all resistance genes (FUNCTIONAL METAGENOMICS)

2. Identify all bacteria (16S rRNA SEQUENCING)

3. Link resistance to bacteria, estimate abundance (SHOTGUN SEQUENCING)

Prediction of genes and species most likely to exchange between the environment and humans

Prediction of community response to antibiotic perturbation.

Geospatial data
Transmission networks of microbiomes and resistomes across habitats

adapted from:
Dantas and Sommer, American Scientist (2014)
30,000 year-old preserved DNA
Resistance to 3 antibiotic classes:
• β-lactams  
  e.g. Penicillin, Amoxicillin
• Tetracyclines  
  e.g. Tetracycline, Minocycline
• Glycopeptides  
  e.g. Vancomycin

Most clinical antibiotics are produced by soil bacteria

Producer resistome billions of years old

Antibiotic resistance in soil enriched over last 70+ years

Antibiotic resistance is ancient  
D’Costa…Wright, Nature (2011)

Producer Hypothesis  
Benveniste, Davies, PNAS (1973)

Increased Gene Abundance  
Knapp…Graham, PNAS (2010)

The soil resistome is ancient, diverse, and growing:

Is it in RECENT exchange with pathogens?
Resistome of highly multi-drug resistant soil cultures

11 U.S. soils
(urban, farm, pristine)

Media + 1000\(\mu\)g/mL of 18 antibiotics

Pool cultures

Extract Metagenomic DNA

95 cultures

Media + 1000\(\mu\)g/mL of 18 antibiotics

Abx Passage 7 days

7 days

Abx Passage 7 days


Soil resistance genes with 100% ID to globally-distributed pathogenic isolates

<table>
<thead>
<tr>
<th>Gene Name</th>
<th># of Selections</th>
<th>Antibiotic Class</th>
<th>Resistance Mechanism</th>
<th>Identical Pathogenic Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB95_PI_68.1</td>
<td>4</td>
<td>β-lactam</td>
<td>enzymatic cleavage</td>
<td><em>A. baumannii,</em>  <em>K. pneumoniae,</em>  <em>P. aeruginosa,</em>  <em>S. typhimurium,</em>  <em>P. mirabilis</em></td>
</tr>
<tr>
<td>AB95_CH_13.1</td>
<td>1</td>
<td>Amphenicol</td>
<td>efflux</td>
<td><em>A. baumannii,</em>  <em>P. aeruginosa</em></td>
</tr>
<tr>
<td>AB95_TE_2.2</td>
<td>3</td>
<td>Tetracycline</td>
<td>efflux</td>
<td><em>A. baumannii,</em>  <em>S. typhimurium</em></td>
</tr>
<tr>
<td>AB95_TE_1.1</td>
<td>3</td>
<td>Tetracycline</td>
<td>efflux</td>
<td><em>A. baumannii,</em>  <em>E. coli,</em>  <em>K. pneumoniae,</em>  <em>S. typhimurium</em></td>
</tr>
<tr>
<td>AB95_GE_3.3</td>
<td>2</td>
<td>Aminoglycoside</td>
<td>covalent modification</td>
<td><em>E. cloacae,</em>  <em>K. pneumoniae,</em>  <em>P. aeruginosa,</em>  <em>S. typhimurium</em></td>
</tr>
<tr>
<td>AB95_GE_3.1</td>
<td>2</td>
<td>Sulfonamide</td>
<td>target modification</td>
<td><em>C. diptheriae,</em>  <em>E. cloacae,</em>  <em>K. pneumoniae,</em>  <em>P. aeruginosa,</em>  <em>S. typhimurium,</em>  <em>Yersinia pestis</em></td>
</tr>
<tr>
<td>AB95_CH_21.1</td>
<td>1</td>
<td>Aminoglycoside</td>
<td>covalent modification</td>
<td><em>A. baumannii,</em>  <em>K. pneumoniae,</em>  <em>P. aeruginosa,</em>  <em>S. typhi</em></td>
</tr>
</tbody>
</table>

- 5 antibiotic classes
- **ALL major resistance mechanisms**
- Pathogens are *Gram +/-* clinical isolates from **across the world**
- Pathogen genes on plasmids and chromosomes
Soil bacteria and human pathogens share MULTIDRUG resistance clusters

**PATHOGENS**

- *Pseudomonas aeruginosa*
- *Acinetobacter baumannii*
- *Salmonella enterica*
- *Salmonella typhimurium*
- *Salmonella typhimurium*

**SOIL GENES**

- *E. coli Int1*
- *S. aureus Int1*
- *Salmonella int1 partial*
- *C. jejuni Cfla PP FLO*

**RESISTANCE genes**

- Gentamicin
- Chloramphenicol
- Tetracycline
- Piperacillin

**MOBILIZATION genes**

- 100% ID

**OBSERVATION:**
Multidrug-Resistant Proteobacteria from the Soil are Exchanging Antibiotic Resistance Genes with Pathogens in the Clinic

**QUESTION:**
Is this the **RULE** or the **EXCEPTION**?

Kevin Forsberg  
Sanket Patel
Soil resistomes are structured by bacterial phylogeny and have lower HGT potential than human pathogens.

18 Soils From 2 Long-Term Experimental Plots

- **Successional Grassland (CC) (9 soils)**
  - 8 Soils with 16S phylogenetic data
  - 9 Soils with resistome data

- **Agricultural Field (KBS) (9 soils)**
  - 9 Soils with 16S phylogenetic data
  - 7 Soils with resistome data

219 of 324 selections yielded >3000 antibiotic resistance genes

Soil resistance genes are co-localized with fewer mobilization genes than pathogens

Transmission networks of microbiomes and resistomes across habitats

adapted from:
Dantas and Sommer, American Scientist (2014)
Resistance spreads across habitats
Antibiotic perturbation of the human microbiome can be dysbiotic

Analysis of 2 human gut resistomes by functional metagenomics

Fecal samples from 2 unrelated healthy adults
• no antibiotic therapy >1 year

Resistomes of 2 healthy adults: diverse and undersampled

95 METAGENOMIC Resistance Genes

- Resistance genes from UNCULTURED bacteria are mostly novel, but fully functional in cultured host
- Previous UNDERSAMPLING

105 AEROBIC CULTURED Resistance Genes

- Resistance genes from aerobic CULTURED Proteobacteria are highly similar to pathogenic resistance genes

Antibiotics are the most prescribed medication for preterm infants. Preterm birth is the leading cause of infant death. Preterm infants are highly susceptible to infections.

99% of VLBW infants receive antibiotics in the 1st two days of life.

Gut microbiomes of preterm infants are dominated by MDROs. We can predict microbiome and resistome responses to antibiotics.

Preterm birth is the leading cause of infant death.

85% prediction accuracy based on 4 variables.

Ecology of the resistome WITHOUT anthropogenic antibiotics?

Functional resistance genes to modern antibiotics are present in antibiotic-naïve microbiomes of remote communities.
Transmission networks of microbiomes and resistomes across habitats

adapted from:
Dantas and Sommer, American Scientist (2014)
Cross-Habitat Resistome Dynamics in Resource-Poor Human Environments in Central and South America

Erica Pehrsson

Pablo Tsukayama
Dissemination of bacteria and antibiotic resistance genes across interconnected habitats in low-income settings in Latin America

Village in Rural El Salvador (RES)  Peri-urban Shanty-Town (PST) in Peru

• Most microbiome studies from extremes of global population: industrialized nations or remote hunter-gatherers

• Low-income, resource-poor settings represent $2/3$rd of global population

• Models for global population growth (pop. density, living conditions)

• Lower hygiene standards = higher rates of infectious disease and bacterial exchange

• Frequent misuse of antibiotics

SAMPLES ANALYZED FROM RES and PST:

• 263 fecal samples from 115 individuals from 27 houses

• 209 environmental samples from animal feces, soils, water, sewage

RES and PST cluster with other human gut microbiota by lifestyle across industrializing gradient

Adonis $R^2 = 37.6\%, P < 0.001$

RES and PST microbiomes and resistomes cluster by habitat

Adonis $R^2 = 22.4\%, \ P < 0.001$

Adonis $R^2 = 41.9\%, \ P < 0.001$
Frequency of AR gene detection per library significantly correlated to (1) mobilization, (2) multidrug resistance, (3) multiple genetic contexts.

1100 UNIQUE ANTIBIOTIC RESISTANCE PROTEINS FROM ALL FUNCTIONAL SELECTIONS

Spearman rho = 0.59, p < 0.05

TEM β-lactamase found in 25 different contexts co-localized with mobilization elements and other AR genes
Identification of resistome dissemination hotspots may help with surveillance

Chicken coops (El Salvador) and Sewage treatment plant (Peru) were hotspots for resistome exchange between humans and the environment.

OUR MICROBIAL PLANET

Think microbes are bad guys? Think again.

Keep us healthy. Amazingly, only about 1 out of 10 cells in the human body is actually a human cell; most of the cells in our bodies are microbes. Some of the microbes living in our bodies actually help us fight disease-causing microbes by competing against them for space. This mutually beneficial relationship helps to protect us from getting diseases while giving the "good" microbes a place to live.

Keep our environment clean. Because of their special adaptations, some microbes can help clean up gasoline leaks, oil spills, sewage, nuclear waste, and many other types of pollution.

Make air breathable. Without microbes, we wouldn't have oxygen to breathe. This is because many microbes are photosynthetic—like plants, they harvest energy from the sun, releasing oxygen into the air. Billions of years ago, photosynthetic microbes gradually added oxygen to Earth's atmosphere, making it possible for large forms of life—including humans—to live.

Provide sources of new medicines. Hundreds of medicines available today were derived from chemicals not found in microbes. Microbes naturally produce an amazing variety of chemicals, which scientists can use to create new medicines.

Support and protect crops. Microbes living in soil help protect plants from pests and diseases. They also are essential for converting nitrogen and other nutrients into forms that plants can use to grow.

Help us digest food. Many of the foods we eat would be indigestible without the 10-100 trillion microbes living within our guts. Microbes also play a major role in creating many of the foods we love, such as cheese, yogurt, and bread.

The science of metagenomics is shedding new light on the microbial world. Scientists estimate that less than 1% of Earth's millions of microbial species can be grown in the laboratory. Using metagenomics, scientists can now study how whole communities of microbes function without having to grow each species separately—making more microbes accessible to science than ever before.

Visit www.nationalacademies.org/microbes to learn more!