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

adapted from:
METAGENOMICS: study microbial communities through direct DNA sequencing

Culture-Dependent Methods

Culture-Independent Methods

Susceptibility Testing

Genome Sequencing

PCR-based Screens

Direct DNA Sequencing
The Great Plate-Count Anomaly (Staley and Konopka, 1985)

1. Collect organism sample from water or soil.
2. View in microscope.
3. Culture sample on agar.
5. Count cells.
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 Pg = $10^{15}$ g)
  - 85-130 Pg of Nitrogen
  - 9-14 Pg of Phosphorous
- Diversity of Habitats: 340°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 5th, 2016 (GOLD, JGI):
  - ~56,000 prokaryotic genomes completed or in progress (~45,000 one year ago)

(File: genomesonline.org/cgi-bin/GOLD/index.cgi)
Two Means of Bacterial Evolution

Widespread HGT Can Obscure Phylogenetic Inference


[Diagram showing phylogenetic relationships and gene presence/absence for Bacillus and Clostridium species.]
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!
RECAP: why sequence microbial communities?

- Only a small proportion of organisms have been grown in culture
- Species do not live in isolation
- Clonal cultures fail to represent the natural environment of a given organism
- Many proteins and protein functions remain undiscovered
Q: Why sequence microbial communities?

A: They may distinguish/define habitats & hosts

Turnbaugh et al. 2006, An obesity associated gut microbiome with increased capacity for energy harvest. Nature 444 1027-1031
Results can translate to humans

10x more bacterial cells than human

100-fold more unique genes

Sequence-based Metagenomics
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/
Required:

MASSIVE COMPUTING RESOURCES
Sample

Extract DNA

16S rRNA gene-based surveys

PCR
Amplify
(sample specific barcodes)

Sequence

Shotgun sequencing of metagenome/transcriptome

Total DNA/cDNA

Sequence

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

V1/75-53 (22 mer)
V2/18S-176 (22 mer)
V3/18S-176 (22 mer)
V4/18S-1197 (17 mer)
V4/18S-1197 (17 mer)

Erlandsen SL et al. J Histochem Cytochem 2005;53:917-927

Circumvents the need to culture

Alan Walker, Sanger
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>

E.coli 16S SSU rRNA hyper-variable regions


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

Assign reads to communities

Visualize and compare community relationships
QIIME workflow schematic

$ beta_diversity_through_plots.py -i otu_table.txt -m Fasting_Map.txt -o betadivtree -t rep_set.tre
Measuring diversity of OTUs

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)

Evolution of Mammals and Their Gut Microbes

Ruth E. Ley, Micah Hamady, Catherine Lozupone, Peter J. Turnbaugh, Rob Roy Ramey, Stephen Birch, Michael L. Schlegel, Tammy A. Tucker, Mark D. Schrenzel, Rob Knight, Jeffrey I. Gordon

SCIENCE VOL 320 20 JUNE 2008
Mammalian Microbiomes Cluster by Diet

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

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
Shotgun whole metagenome sequencing to estimate community functions by database comparison.
Sequence-based Metagenomics: Shotgun environmental sequencing to characterize genetic diversity: 4+ Examples
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

- 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)

1. Primers bind to template
2. Polymerization begins
3. Polymerization continues
4. Strand displacement
5. New primers bind to newly formed DNA
6. Polymerization from new strands

GE Healthcare GenomiPhi Kit

Genome sequencing on single *Prochlorococcus* cells

<table>
<thead>
<tr>
<th></th>
<th>JGI</th>
<th>Agencourt</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># good seq reads</td>
<td>7,166</td>
<td>10,660</td>
<td></td>
</tr>
<tr>
<td>Average length (bp)</td>
<td>769.4</td>
<td>676.6</td>
<td></td>
</tr>
<tr>
<td>Total length (bp)</td>
<td>5,513,520</td>
<td>7,212,556</td>
<td></td>
</tr>
<tr>
<td># unknown seqs</td>
<td>12</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td># vectors</td>
<td>23</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td># other seqs</td>
<td>74 (MED4)</td>
<td>2 (MED4)</td>
<td></td>
</tr>
<tr>
<td>% genome sampled</td>
<td>62.47%</td>
<td>66.53%</td>
<td></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 Rinke\(^1\), Patrick Schwientek\(^1\), Alexander Sczyrba\(^1\)\(^2\), Natalia N. Ivanova\(^1\), Iain J. Anderson\(^1\)\(^3\), Jan-Fang Cheng\(^1\), Aaron Darling\(^3\)\(^4\), Stephanie Malfatti\(^1\), Brandon K. Swan\(^5\), Esther A. Gies\(^6\), Jeremy A. Dodsworth\(^7\), Brian P. Hedlund\(^2\), George Tsiamis\(^8\), Stefan M. Sievert\(^9\), Wen-Tso Liu\(^10\), Jonathan A. Eisen\(^3\), Steven J. Hallam\(^6\), Nikos C. Kyripides\(^1\), Ramunas Stepanauskas\(^8\), Edward M. Rubin\(^1\), Philip Hugenholtz\(^11\) & Tanja Woyke\(^1\)

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...
Example 4+: Cheaper sequencing – more info and more complete genomes from metagenomes

Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw

Rachel Mackelparg1,2, Mark P. Waldrop3, Kristen M. DeAngells4, Maude M. David4, Krystle L. Chavarria4, Steven J. B Edward M. Rubin5,6 & Janet K. Jansson7,8


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>
<thead>
<tr>
<th>Function-driven versus sequence-driven screening strategies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td><strong>Disadvantages</strong></td>
</tr>
<tr>
<td>Completely novel genes can be recovered</td>
<td>Dependent on expression of the cloned genes by the bacterial host</td>
</tr>
<tr>
<td>Selects for full-length genes</td>
<td>Requires production of a functional gene product by the bacterial host</td>
</tr>
<tr>
<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></td>
</tr>
<tr>
<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>Similar screening strategies can be used for different targets, for example, colony hybridization and PCR</td>
<td>Partial genes can be cloned</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

Diverse

More Stable

Higher Activity

<table>
<thead>
<tr>
<th>Enzyme Substrate</th>
<th>$K_{m}$ (μM)</th>
<th>Catechol</th>
<th>3-MC</th>
<th>4-MC</th>
<th>4-CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NahH</td>
<td>4.9</td>
<td>2.7</td>
<td>3.2</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>2.1</td>
<td>0.9</td>
<td>0.7</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td>2.0</td>
<td>0.5</td>
<td>0.6</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

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)
Antibiotic Resistant Infections Are A Leading Cause of Death

700,000 deaths in 2014

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

Cooper & Schlaes, 2011

A Perfect Storm
As bacterial infections grow more resistant to antibiotics, companies are pulling out of antibiotics research and fewer new antibiotics are being approved.

(UK Prime Minister's AMR Report, 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

<table>
<thead>
<tr>
<th></th>
<th>Previously discovered</th>
<th>Not previously discovered</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td><img src="Image1" alt="Pie chart for Human" /></td>
<td></td>
</tr>
<tr>
<td>Sommer, Dantas et al, Science 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Soil</strong></td>
<td><img src="Image2" alt="Pie chart for Soil" /></td>
<td></td>
</tr>
<tr>
<td>Forsberg...Dantas, Science 2012</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Uncultured resistance genes

<table>
<thead>
<tr>
<th></th>
<th>Previously discovered</th>
<th>Not previously discovered</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td><img src="Image3" alt="Pie chart for Human" /></td>
<td></td>
</tr>
<tr>
<td>Sommer, Dantas et al, Science 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Soil</strong></td>
<td><img src="Image4" alt="Pie chart for Soil" /></td>
<td></td>
</tr>
<tr>
<td>Forsberg...Dantas, Nature 2014</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Methods for studying antibiotic resistance in microbial communities

**KNOWN**

- CULTURE & PHENOTYPE
  - 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**

- Antibiotic Resistance Reservoir (RESISTOME)

- Known, Readily Cultured
- Known, Not Readily Cultured
- Unknown
Increasing functional metagenomic throughput via next-gen sequencing


Current methods are biased and slow at annotating antibiotic resistance functions in microbial communities.

1. Biased
2. Slow

ACGTCTCAATCGCTCATTTTCATTTCCTTTAGATTTCGGATTTTCATATTGGATCGACATATCGTCTCAATC

? | GCTTCGGT

Antibiotic Resistance Functions

BLAST
Resfams improves our ability to study antibiotic resistance in microbial communities.

**Resfams**: antibiotic resistance specific profile hidden Markov model (pHMM) database.

Antibiotic Resistance Functions
Resfams improves our ability to study antibiotic resistance in microbial communities

<table>
<thead>
<tr>
<th>Known Antibiotic Resistance Proteins Class</th>
<th>Novel Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAF</td>
<td>YAF</td>
</tr>
<tr>
<td>CGW</td>
<td>CDW</td>
</tr>
<tr>
<td>CDY</td>
<td></td>
</tr>
<tr>
<td>CVK</td>
<td></td>
</tr>
<tr>
<td>CKY</td>
<td></td>
</tr>
</tbody>
</table>

Gibson MK, Forsberg KJ, and Dantas G, 2015 (The ISME Journal)

www.dantaslab.org/resfams
Resfams predicted > 95% of resistance genes, while BLAST to resistance-specific databases predicted < 34% of resistance genes in hand curated gold standard set.
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

- USA1
- USA2
- Peru
- El Salvador
- Malawi
- Venezuela
A network of potential resistome exchange
GOAL
Investigate role of commensal microbiota in evolution of resistance in pathogens

HYPOTHESIS
• antibiotic exposure enriches commensal resistome
• commensals exchange resistomes with pathogens

EXPERIMENTS
Characterize human commensal resistome diversity, dynamics, and evolution with functional metagenomics
Analysis of 2 human gut resistomes by functional metagenomics

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

Resistance genes from CULTURED gut aerobes highly similar to pathogens

TEM-1 beta lactamase from *Neisseria meningitidis*

<table>
<thead>
<tr>
<th>PATHOGEN</th>
<th>COMMENSAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGAGTATCACCAATTTCGTGCGCCTTTATCCCTTTTGTGCCGATT</td>
<td>ATGAGTATCACCAATTTCGTGCGCCTTTATCCCTTTTGTGCCGATT</td>
</tr>
</tbody>
</table>
| TTGCTTCCTGGTTTTGCTGACCCAGAAGCTGTGAAGTGAAATGAGT | TTGCTTCCTGGTTTTGCTGACCCAGAAGCTGTGAAGTGAAATGAGT 

CTX-M-15 beta lactamase from *Klebsiella pneumoniae*

<table>
<thead>
<tr>
<th>PATHOGEN</th>
<th>COMMENSAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGGTTAAAAATATCAGGCGGCTCTTCAAGGCTGATGACGCGACGCGGAT</td>
<td>ATGGTTAAAAATATCAGGCGGCTCTTCAAGGCTGATGACGCGACGCGGAT</td>
</tr>
<tr>
<td>CAGCGTCTGTGTTAAAGTGATGCCGCTGATSCGAACGCGGGGACGTTAC</td>
<td>CAGCGTCTGTGTTAAAGTGATGCCGCTGATSCGAACGCGGGGACGTTAC</td>
</tr>
<tr>
<td>AGCAAAACCTGCGGATTAGACGCGCGCTCGAGGCAAGCGACTGCGGTGTTG</td>
<td>AGCAAAACCTGCGGATTAGACGCGCGCTCGAGGCAAGCGACTGCGGTGTTG</td>
</tr>
<tr>
<td>GCATTGATATTACAGACAGATTACGCGATTACGTCTGATGAGTA</td>
<td>GCATTGATATTACAGACAGATTACGCGATTACGTCTGATGAGTA</td>
</tr>
</tbody>
</table>

Resistance genes from UNCCULTURED majority are mostly novel 
BUT fully functional in Proteobacterial host

95 METAGENOMIC Resistance Genes

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

Microbiota is dynamic for first 3 years of life, followed by relative stability in adults.

High antibiotic use in US kids, often unwarranted.

Factors that may influence resistome establishment and dynamics:

- Antibiotic (ab)use
- Mode of delivery
- Breast milk vs formula
- Term vs pre-term birth
- Acute/chronic illness
- Environment (parents, soil, animals)
# Functional metagenomic screening of 22 kids: resistome is early and substantial

<table>
<thead>
<tr>
<th>SUBJECT (ASCENDING AGE)</th>
<th>0-3mo</th>
<th>3-6mo</th>
<th>6-12mo</th>
<th>1-3yo</th>
<th>4-6yo</th>
<th>6-9yo</th>
<th>10-19yo</th>
<th>Genes per GB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANTIBIOTIC CLASS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Median  Range</td>
</tr>
<tr>
<td>Aminoglycoside</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 0-206</td>
</tr>
<tr>
<td>β-lactam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40 7-310</td>
</tr>
<tr>
<td>+ β-lactamase inhibitor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3 0-54</td>
</tr>
<tr>
<td>2nd gen cephalosporin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 0-74</td>
</tr>
<tr>
<td>3rd gen cephalosporin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 0-25</td>
</tr>
<tr>
<td>4th gen cephalosporin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 0-116</td>
</tr>
<tr>
<td>Monobactam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 0-54</td>
</tr>
<tr>
<td>Carbapenem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 0-33</td>
</tr>
<tr>
<td>Tetracycline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>146 4-946</td>
</tr>
<tr>
<td>Quinolone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 0-232</td>
</tr>
<tr>
<td>Amino-acid derivative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>459 95-3000</td>
</tr>
<tr>
<td>Amphenicol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38 1-109</td>
</tr>
<tr>
<td>Polymyxin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 0-1</td>
</tr>
<tr>
<td>Pyrimidine derivative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>552 113-4000</td>
</tr>
<tr>
<td>+ sulfonamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>178 45-529</td>
</tr>
</tbody>
</table>

**SCALE:** Estimated Unique Resistance Genes per GB

0 1-10 10-50 50-150 150-400 >400

---

>2500 resistance genes from 22 pediatric microbiota:
Resistome trends in antibiotic type, spectrum, age

Ecology of the resistome WITHOUT anthropogenic antibiotics?

Woolhouse, Science (2013)
An uncontacted, antibiotic naïve, human community
(collaboration with MG Dominguez-Bello, R Knight, J Clemente)

Study population:
- Yanomami Amerindians
- Only prior contact with Amerindians
- No known exposure to anthropogenic antibiotics ("antibiotic-naïve")

Swabs (fecal, oral, skin) from 35 individuals prior to antibiotic and vaccine administration
Antibiotic resistance is present in antibiotic naïve human microbiota

- Oral and fecal samples from **4 subjects**
- Whole metagenome amplification (MDA)
- Functional selection against **15 antibiotics**
- **30 resistance genes** identified

<table>
<thead>
<tr>
<th></th>
<th>Library size (GB)</th>
<th>Penicillin</th>
<th>Piperacillin</th>
<th>Pip-Tazo</th>
<th>Cefotaxime</th>
<th>Ceftriaxone</th>
<th>Cefepime</th>
<th>Meropenem</th>
<th>Aztreonam</th>
<th>Chloramphenicol</th>
<th>Tetracycline</th>
<th>Tigecycline</th>
<th>Gentamicin</th>
<th>Ciprofloxacin</th>
<th>Colistin</th>
<th>Nitrofurantoin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject 3</strong> (Male, 11)</td>
<td>Fecal 3.97 - 5.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oral 3.21 - 4.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subject 5</strong> (Male, 20)</td>
<td>Fecal 0.44 - 0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oral 4.38 - 5.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subject 6</strong> (Male, 48)</td>
<td>Fecal 2.54 - 3.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oral 0.20 - 0.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subject 23</strong> (Female, 7)</td>
<td>Fecal 1.38 - 2.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oral 1.65 - 2.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### From fecal library:

Class A extended-spectrum β-lactamase (ESBL), subclass cblA

Top blastx hit is *Bacteroides oleciplenus* (90-91%)

- **Resistance to 5 late-generation antibiotics** including:
  - 4<sup>th</sup> Generation cephalosporin
  - Synthetic monobactam

<table>
<thead>
<tr>
<th>Antibiotic Class</th>
<th>Penicillins</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Gen. Cephalosporin</th>
<th>4&lt;sup&gt;th&lt;/sup&gt; Gen. Cephalosporin</th>
<th>Carbapenems</th>
<th>Aztreonam</th>
<th>Susceptible to β-lactamase inhibitors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ESBL</strong></td>
<td></td>
<td>Semi-synthetic</td>
<td>Semi-synthetic</td>
<td></td>
<td></td>
<td>YES</td>
</tr>
<tr>
<td><strong>Yanomami genes</strong></td>
<td></td>
<td>A</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>V</td>
</tr>
</tbody>
</table>

Adapted from table by Dr. Carey-Ann Burnham

**S** = Susceptible

**V** = Variable

**R** = Resistant

* Bootstrap value ≥ 90%

**Ref:** Smith 1994

**Ref:** Contreras 2010

**Ref:** Tapsall 2009

**Ref:** Zaura 2009

---

### From oral libraries:

Penicillin-binding proteins

- **Resistance to 3<sup>rd</sup>-generation cephalosporin ceftazidime**
- **High identity to oral commensals** from Western and Amerindians hosts

A network of potential resistome exchange

Dantas et al., American Scientist (2014)
The soil resistome is ancient, diverse, and growing:

Is it in RECENT exchange with pathogens?

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

Producer Hypothesis
*Benveniste, Davies, PNAS (1973)*

Increased Gene Abundance
*Knapp…Graham, PNAS (2010)*

**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**
GOAL
Investigate the soil microbiota for genetic evidence for recent resistome exchange with pathogens

HYPOTHESIS
- Soil resistome is ancient and diverse
- Resistance levels have been increasing in the soil
- Multidrug resistant soil bacteria may have increased potential for resistome exchange with pathogens

EXPERIMENTS
Characterize resistome of highly multidrug resistant soil bacteria using functional metagenomics
Media + 1000μg/mL of 18 antibiotics

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

Extract Metagenomic DNA

Pool cultures

95 cultures

Extract Total Metagenomics/Chips from Sample 1

Transform fragment library in E.coli to insert

Functional selections on plates with antibiotic concentration from 10μg/mL to 1000μg/mL

Reads ( myśli by Illumina)

PARF/MS: Parallel Annotation and Reassembly of Functional Metagenomic Selections


Kevin Forsberg

Alejandro Reyes

### 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, K. pneumoniae, P. aeruginosa, S. typhimurium</em></td>
</tr>
<tr>
<td>AB95_CH_13.1</td>
<td>1</td>
<td>Amphenicol</td>
<td>efflux</td>
<td><em>A. baumannii, 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, 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, E. coli, K. pneumoniae, 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, K. pneumoniae, P. aeruginosa, 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, E. cloacae, K. pneumoniae, P. aeruginosa, S. typhimurium, 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, K. pneumoniae, P. aeruginosa, 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

AB95_GE_3

AB95_CH_13

AB95_TE_2

AB95_PI_68

Gentamicin

Chloramphenicol

Tetracycline

Piperacillin

= RESISTANCE gene

= MOBILIZATION gene

= 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.

• **Two major mechanisms** of tetracycline resistance:
  – Active Efflux (1)
  – Ribosomal Protection (2)
  – Both prevalent in pathogens

• **3rd mechanism**: tetracycline inactivation
  – 3 genes from human commensals
  – Tet(X) only characterized enzyme
  – Not seen in pathogens until 2013
  – Oxidizes drug via FAD cofactor

• Drug inactivation is large clinical threat
  – e.g. β-lactamases, acetyltransferases
  – Allows survival of “cheaters”
  – Eliminates drug, energetically favorable

Adapted from Hillen (2002)
NINE new tetracycline resistance enzymes (*Tet-Destructases*) from SIX soils

**Predicted Function:** “FAD-Dependent Oxidoreductase”

**Sequence** unlike any tetracycline resistance gene

**Resistance up to 64-fold higher than vector-only control**

**Only homolog from pathogenic *Legionella***

**Resistance conferred by tetracycline inactivation**

Tetracycline destructases encode novel activity and mechanism

OUR MICROBIAL PLANET

Microbes—tiny forms too small to see—play a surprising 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 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 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.

Make air breathable. Without microbes, we wouldn’t have oxygen to breathe. This is because many microbes are photosynthetic—like plants, they harvest their 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.

Provide sources of new medicines. Hundreds of medicines available today were derived from chemicals that are 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 gut. Microbes also play a major role in creating many of the foods we love, such as cheese, yogurt, and bread.

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.

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.

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 President’s Circle Communications Initiative of the National Academies. Artwork by Nicole Siger (https://www.dav-dit.com/). Photos used are all by Ted Crawford, University of Hawaii at Manoa. Bottom photos by Jeff Miller (University of Michigan—Ann Arbor).

The National Academies: Advancing Science, Engineering, and Medicine for the国

Visit www.nationalacademies.org to learn more.

www.nationalacademies.org