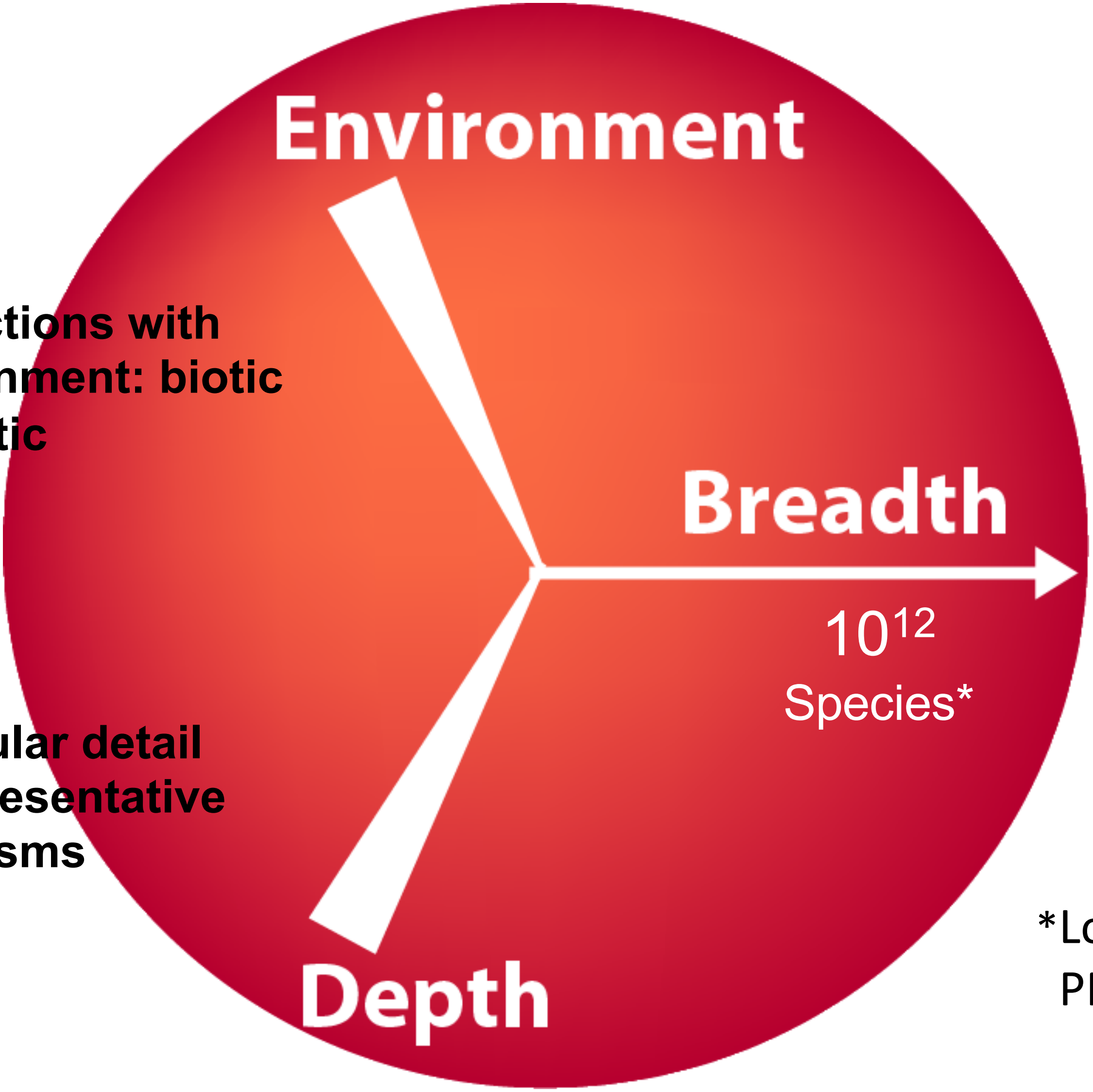
Past and Future of Microbiome Science

James M. Tiedje
Departments of Microbiology and Molecular Genetics
and of Plant, Soil & Microbial Sciences
Michigan State University

Bioremediation - Expanding the Toolbox:
Session I - The Microbiome
Webinar
September 30, 2019

MICHIGAN STATE
UNIVERSITY

To Understand the Microbial World

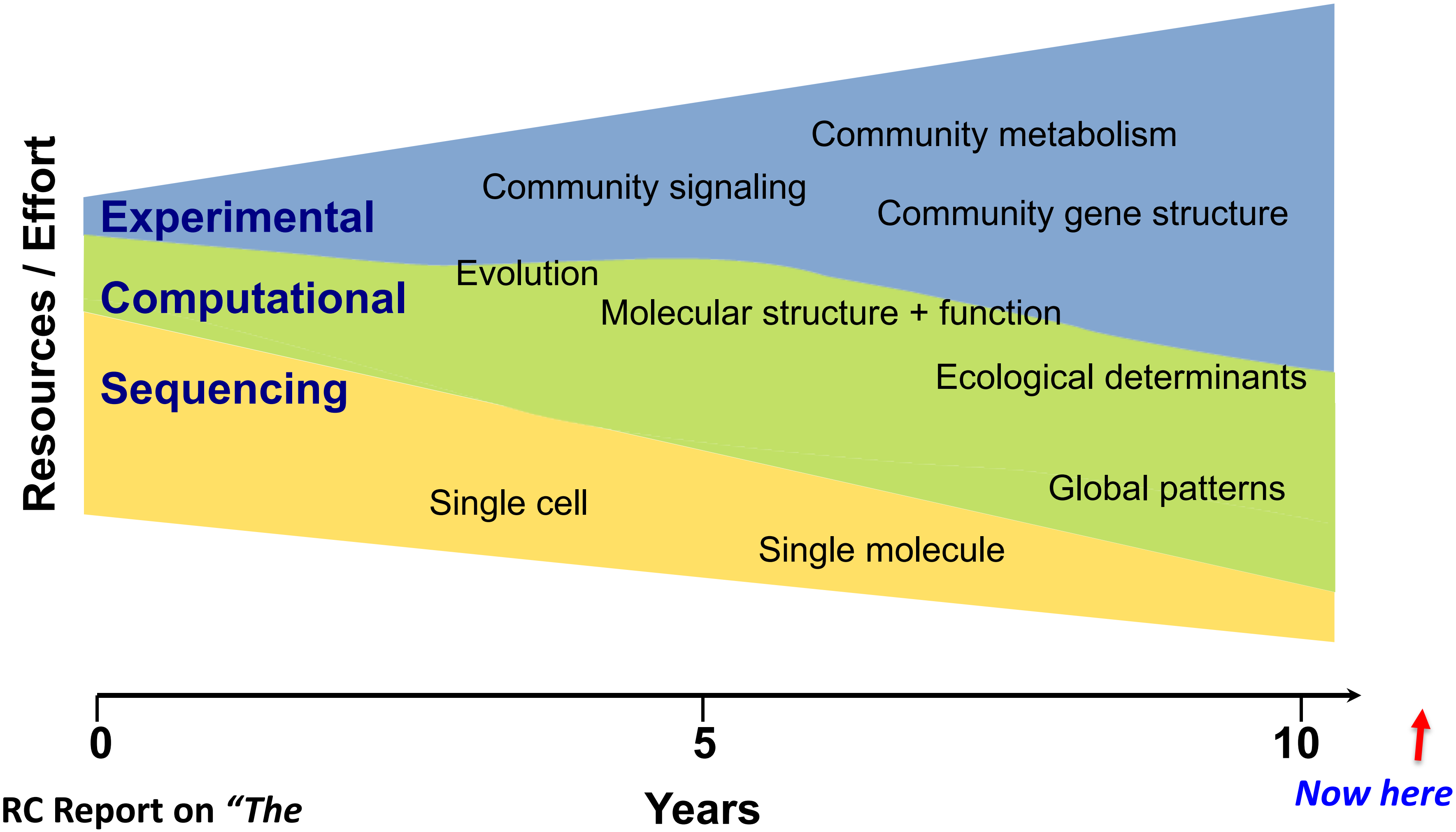


Interactions with environment: biotic & abiotic

Molecular detail of representative organisms

*Locey and Lennon PNAS 2016

Possible Shifts in Emphasis as the Field of Metagenomics Develops



Now here

NRC Report on *"The New Science of Metagenomics"* *2007*

Shift in branding to "Microbiome"

Era's in Microbial Ecology: Technology Enabled

1960-70's: Functions/processes; incubations*, isolates;
isotopes, gas/liq chromatography, auto analyzers
1980's: Math models, kinetics*; PCs
1990's: Genes: DNA extraction, hybridization, 16S,
machine sequencers, internet
2000's: Genome sequences**: HTP sequencing, assembly
2010's: MetaG, T, P, metabolomics: Low cost HTP sequence, HPCs
2020's: ???

*Suflita, Horowitz, Shelton, Tiedje, 1982, Dehalogenation: Novel pathway for anaerobic biodegradation. *Science*, 218;1115

**Chain…Tiedje, et.al., 2006, *Burkholderia xenovorans* LB400 (PCB degrader) harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. PNAS, 103:15280

My Themes

1. Untapped organismal and genetic diversity
2. Technology enabled, what is next?
3. Microbiomes; community-based activities

Genetic Traits Important to Biodegradation

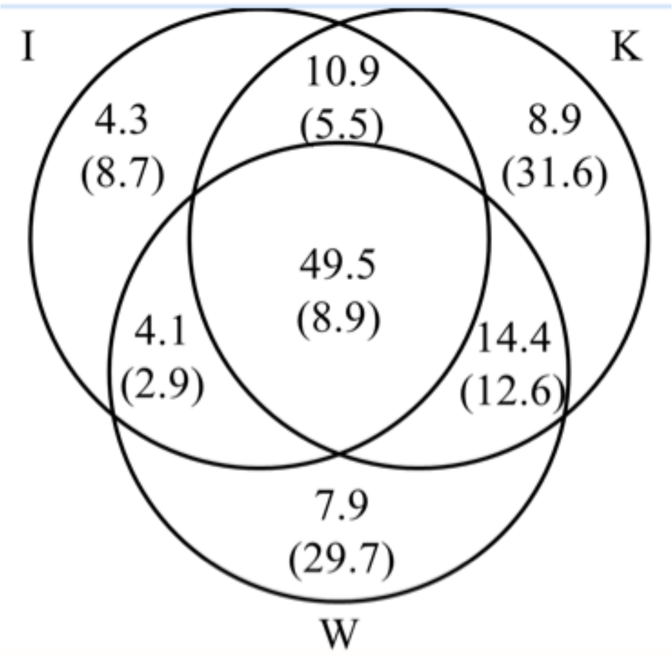
1. Enzyme (the catalyst for bioconversion)
   - Its kinetics
2. Regulation (of gene expression)
3. Permease (transport into the cell)
4. Detergent production
5. Chemotaxis (movement to the pollutant)
6. Resistances
   - heavy metals
   - solvents
   - acidity, salt
   - heat
7. Attach or non-attaching microbes

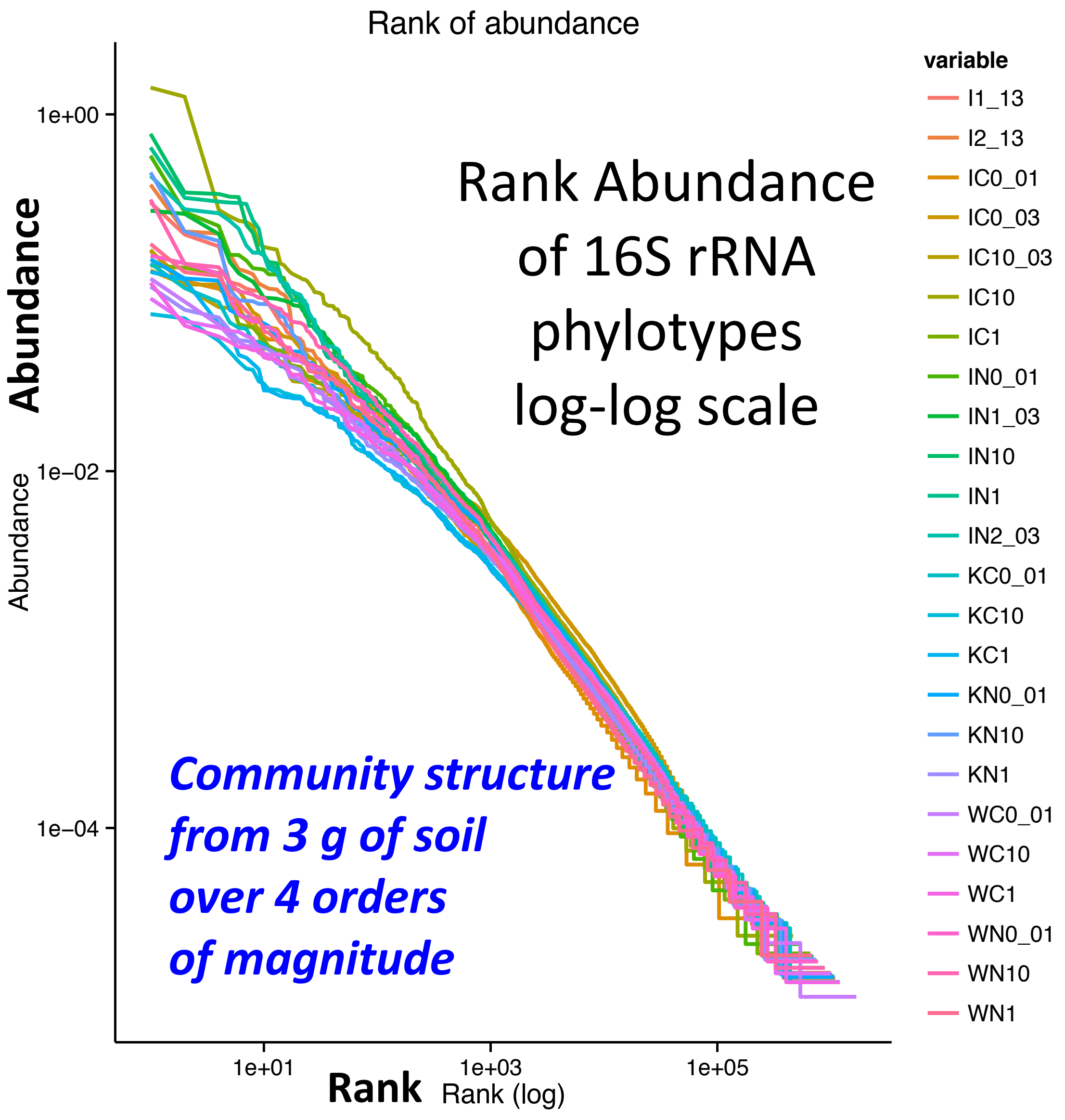
Commonality of Midwest Soil Bacterial Populations

Samples from:

- Kansas, Iowa, Wisconsin
- Native prairie
- >100 year of cultivation

Overlap by state in number of sequences in common phylotypes





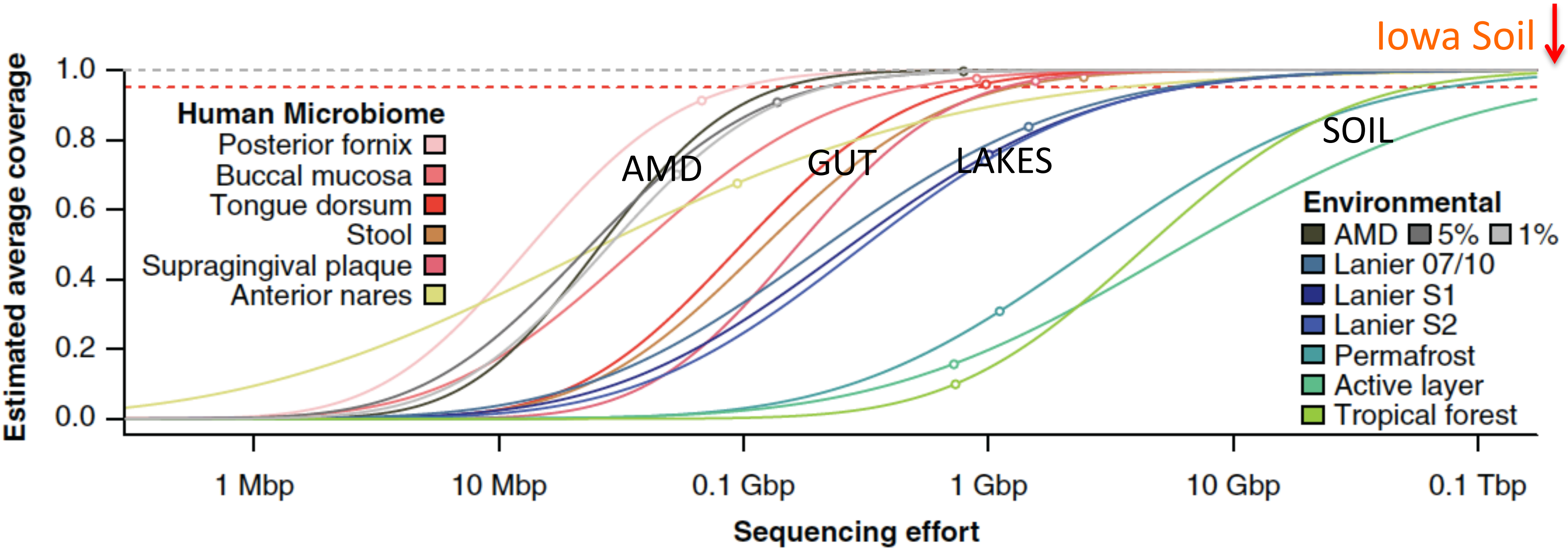
Abundance

Rank

Rank Abundance of 16S rRNA phylotypes log-log scale

Community structure from 3 g of soil over 4 orders of magnitude

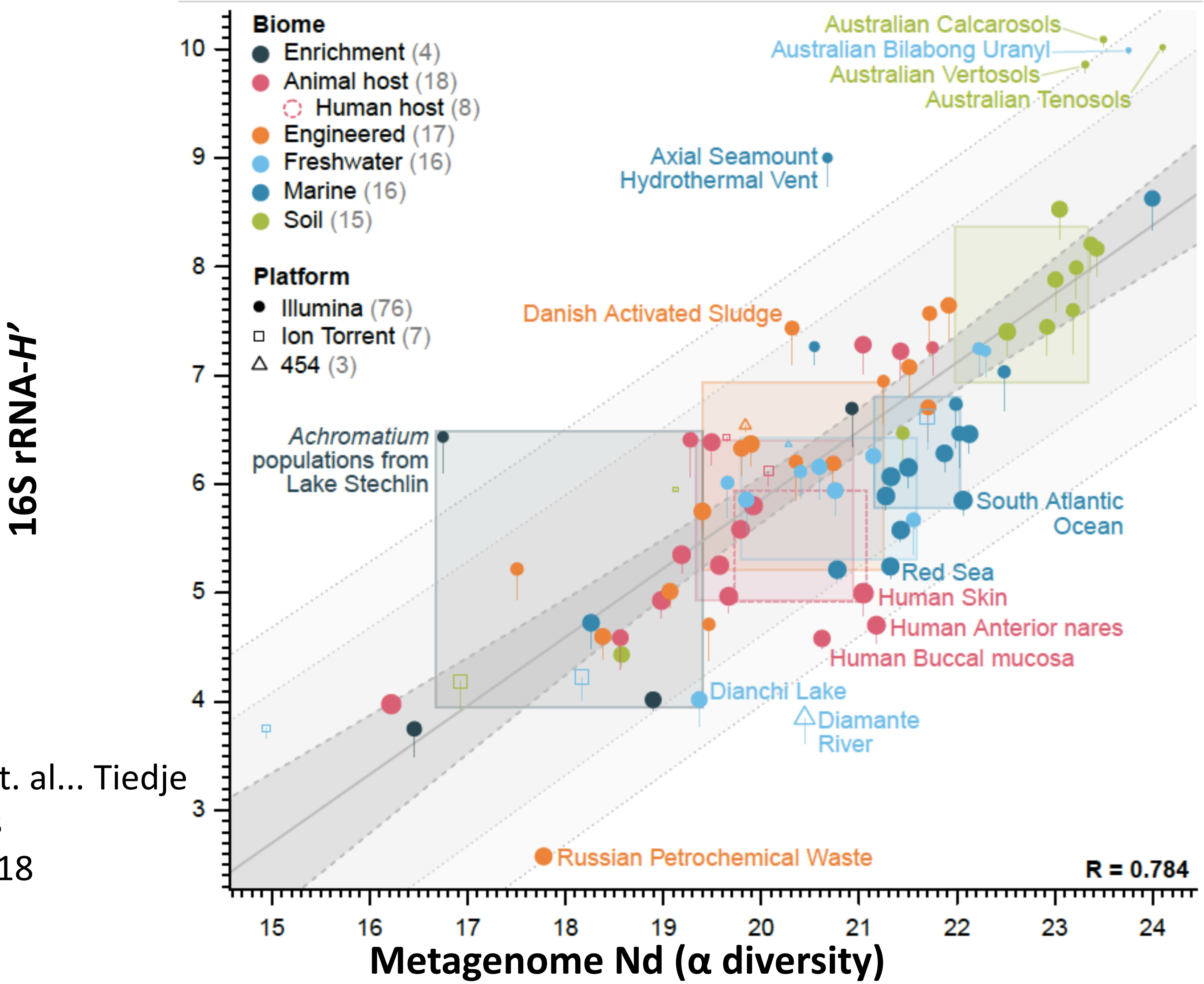
How Diverse are Microbiomes at the Gene Level ?



- ***Practical outcomes are related to microbiome complexity***
- ***0.6 – 1 Tbp for U.S Midwest prairie soil****

Fig: Rodriquez-R and Konstantinidis, *Bioinformatics*, 2013
Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets
*Rodriguez_R, et.al, *mSystems*, 2018; an advanced algorithm

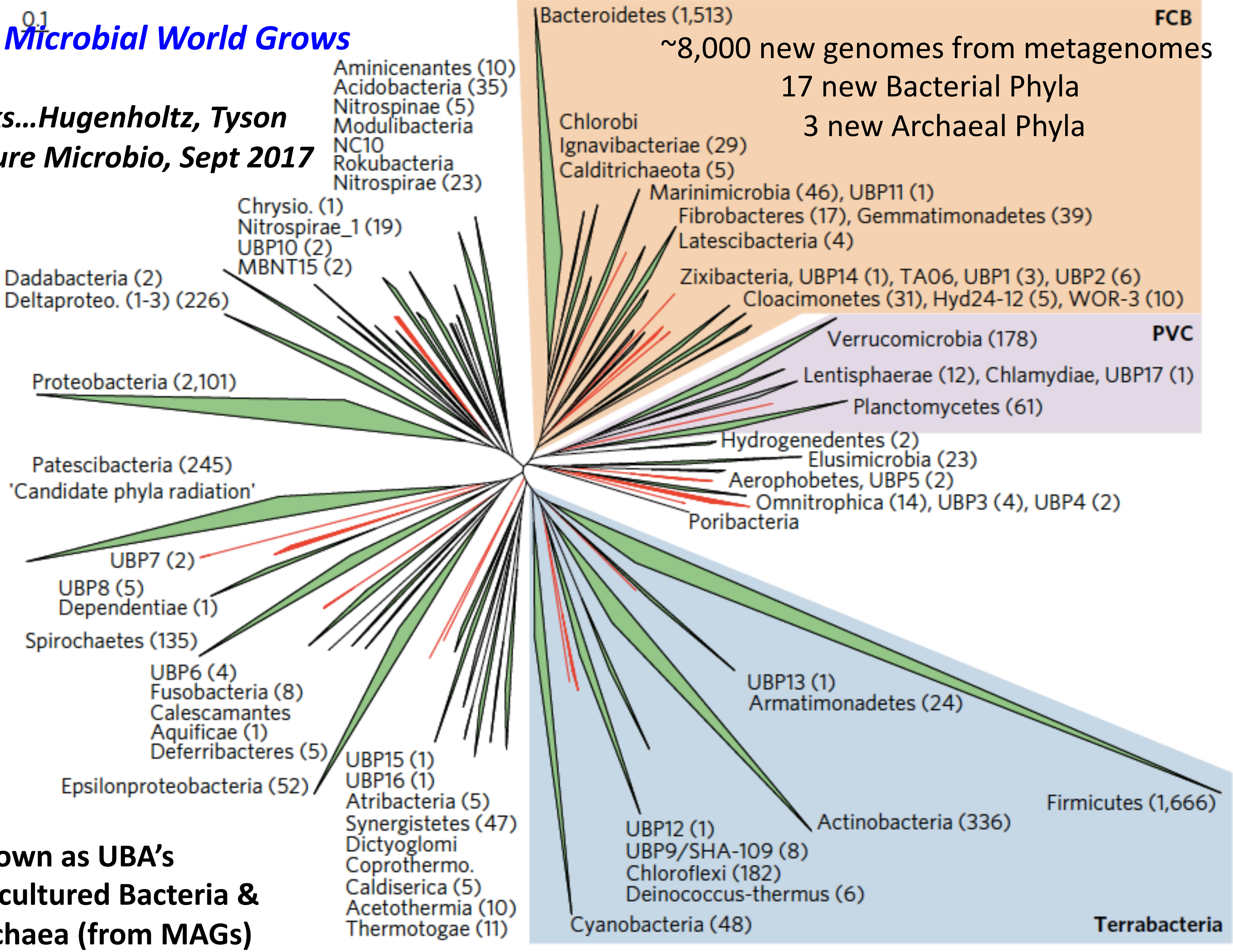
Comparison of Nonpareil sequence diversity (Nd) and 16S rRNA Shannon H’ taxonomic diversity indices on 86 metagenomes



Rodriguez-R et. al... Tiedje Konstantinidis
mSystems, 2018

The Microbial World Grows

Parks…Hugenholtz, Tyson
Nature Microbio, Sept 2017



~8,000 new genomes from metagenomes
17 new Bacterial Phyla
3 new Archaeal Phyla

Known as UBA's
Uncultured Bacteria &
Archaea (from MAGs)

MiGA

Microbial Genomes Atlas

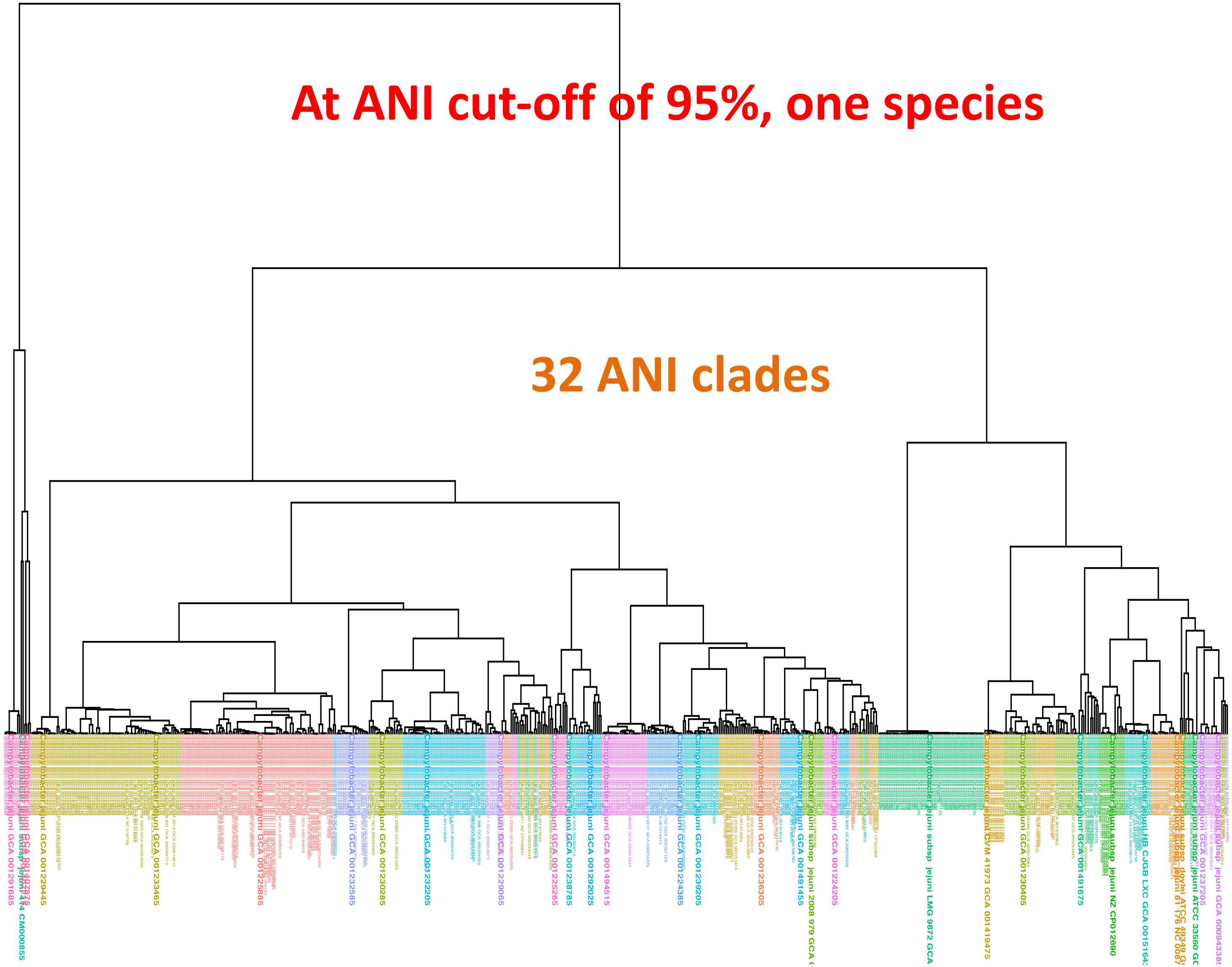
LM **Rodriguez-R [@_lrr_]** GT, S **Gunturu** MSU, C **Luo** GT,
JM **Tiedje** MSU, J **Cole** MSU, KT **Konstantinidis** GT

Nucl Acids Res, 2018

Georgia Tech

MICHIGAN STATE UNIVERSITY

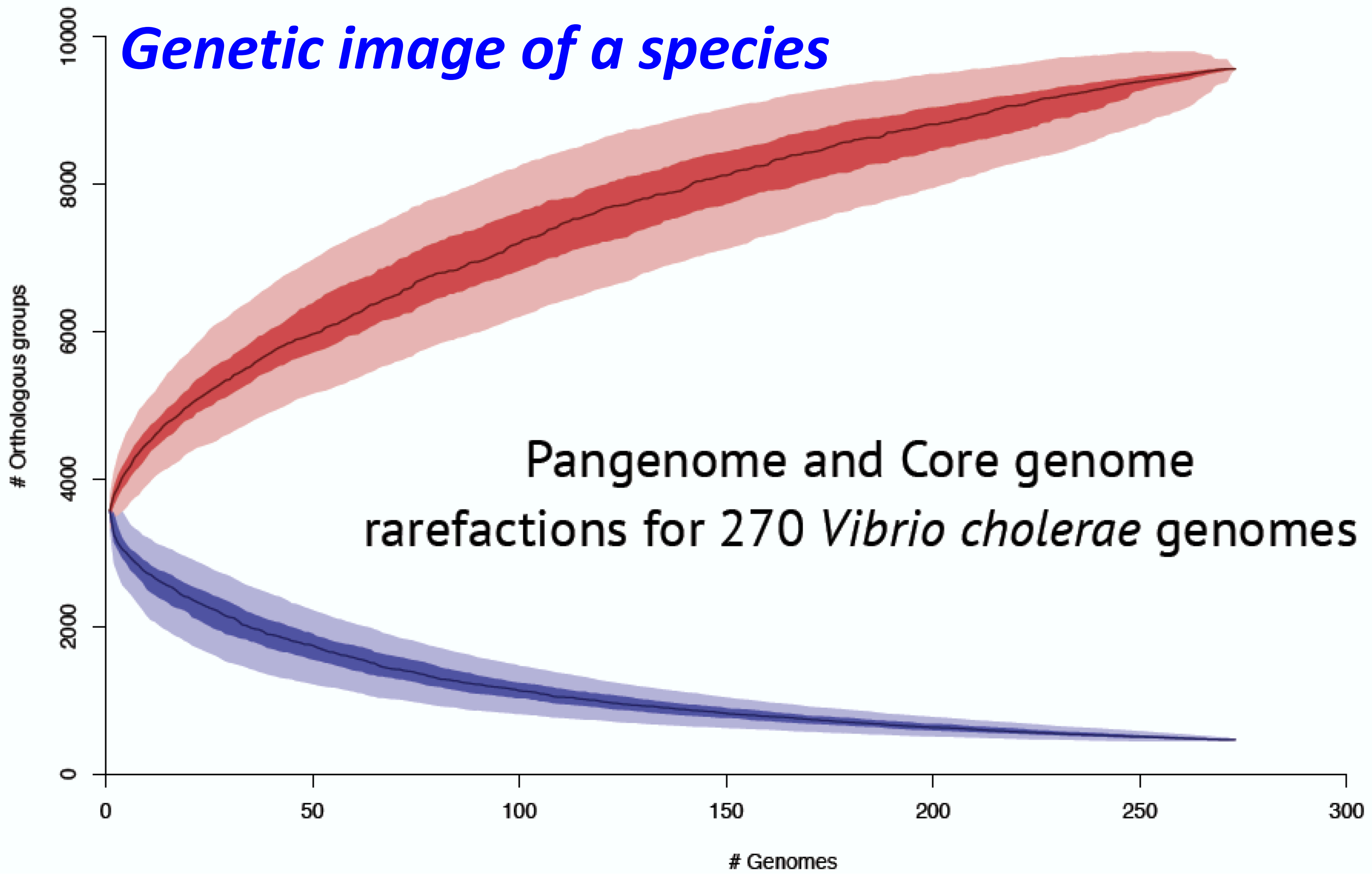
Whole genome ANI relationships of 682 Campylobacter jejuni strains



At ANI cut-off of 95%, one species

32 ANI clades

Genetic image of a species



Pangenome and Core genome rarefactions for 270 *Vibrio cholerae* genomes

All MiGA tools in the Enveomics Collection at:

www.enve-omics.gatech.edu

Ecological Framework for Community Outcome

What determines microbial community structure?

Deterministic ↔ Stochastic

Major drivers - Subtle - Physiological - Pure chance*

Substrate use & its kinetics, adaptions to stressful conditions

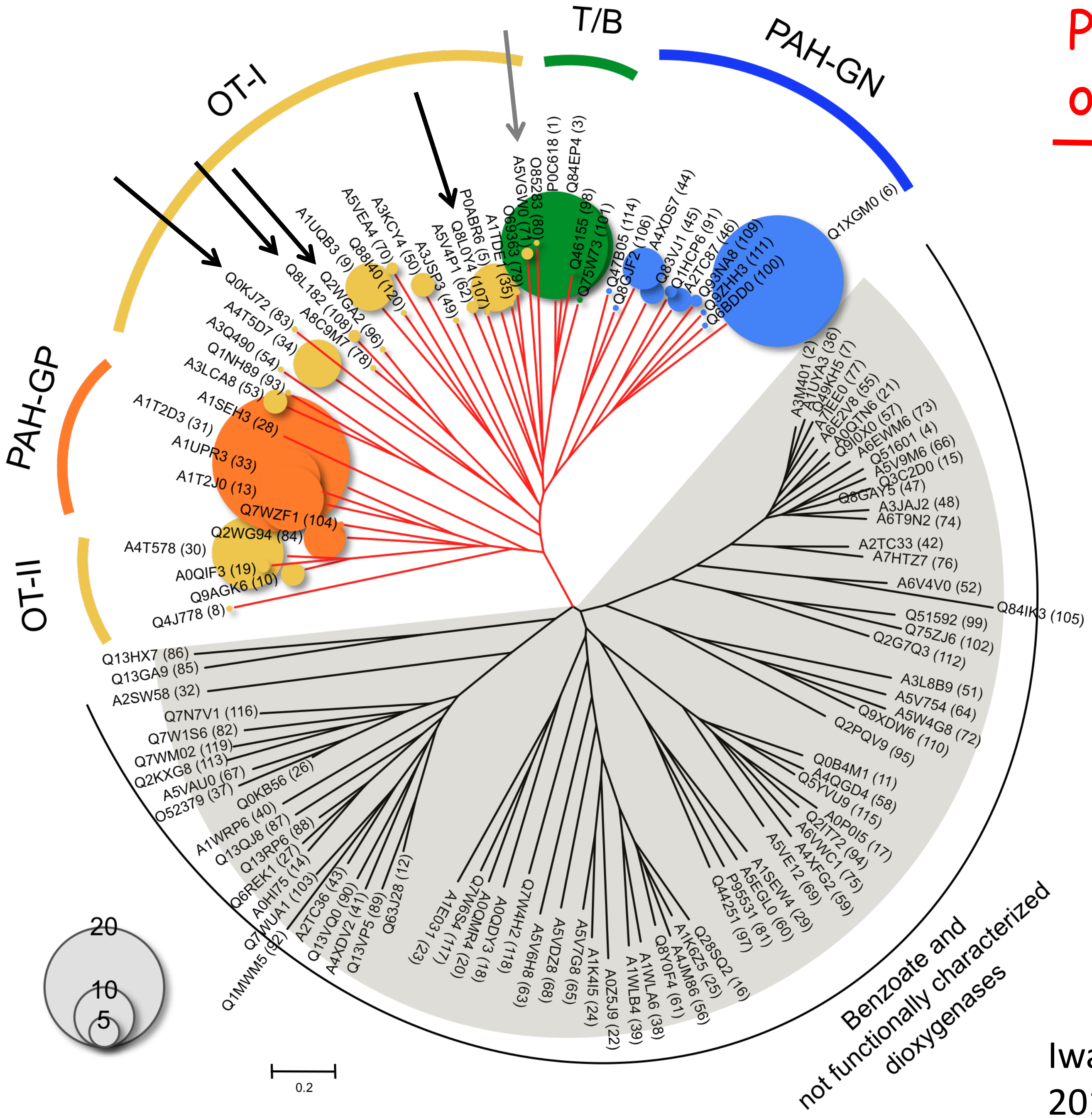
spores/dormancy, attachment/biofilms, phage attack, interactions

1. What features are deterministic?
2. How much is deterministic?
3. How much of the unexplained is the unmeasured?
4. How much is due to non-uniform inoculation (dispersal)?
5. Is system dynamics a primary determinant of this distribution?

* Laurent Philippot “Time for noise to become a signal”

How to ‘herd’ the community structure for desired outcomes

Protein tree of dioxygenases



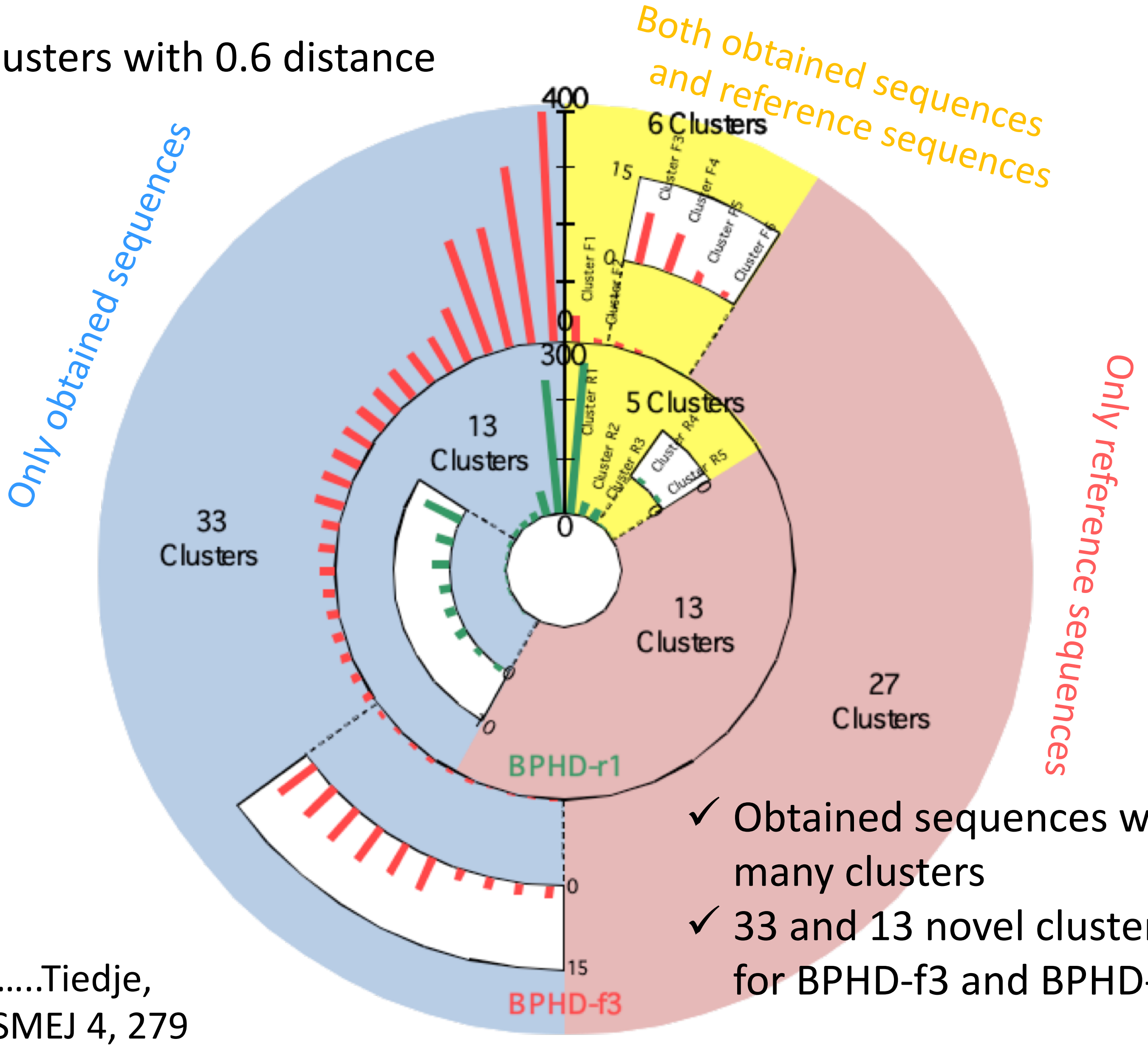
OT-I contains small clusters that attack dioxin (gray) and dibenzofuran (black) compounds (clusters 71, 107, 96, 108, 83).

OT-II is mainly phthalate dioxygenases (30)

Iwai, Johnson…Tiedje 2011, AEM, Appl. 77: 3551

Known and novel sequence clusters

Clusters with 0.6 distance

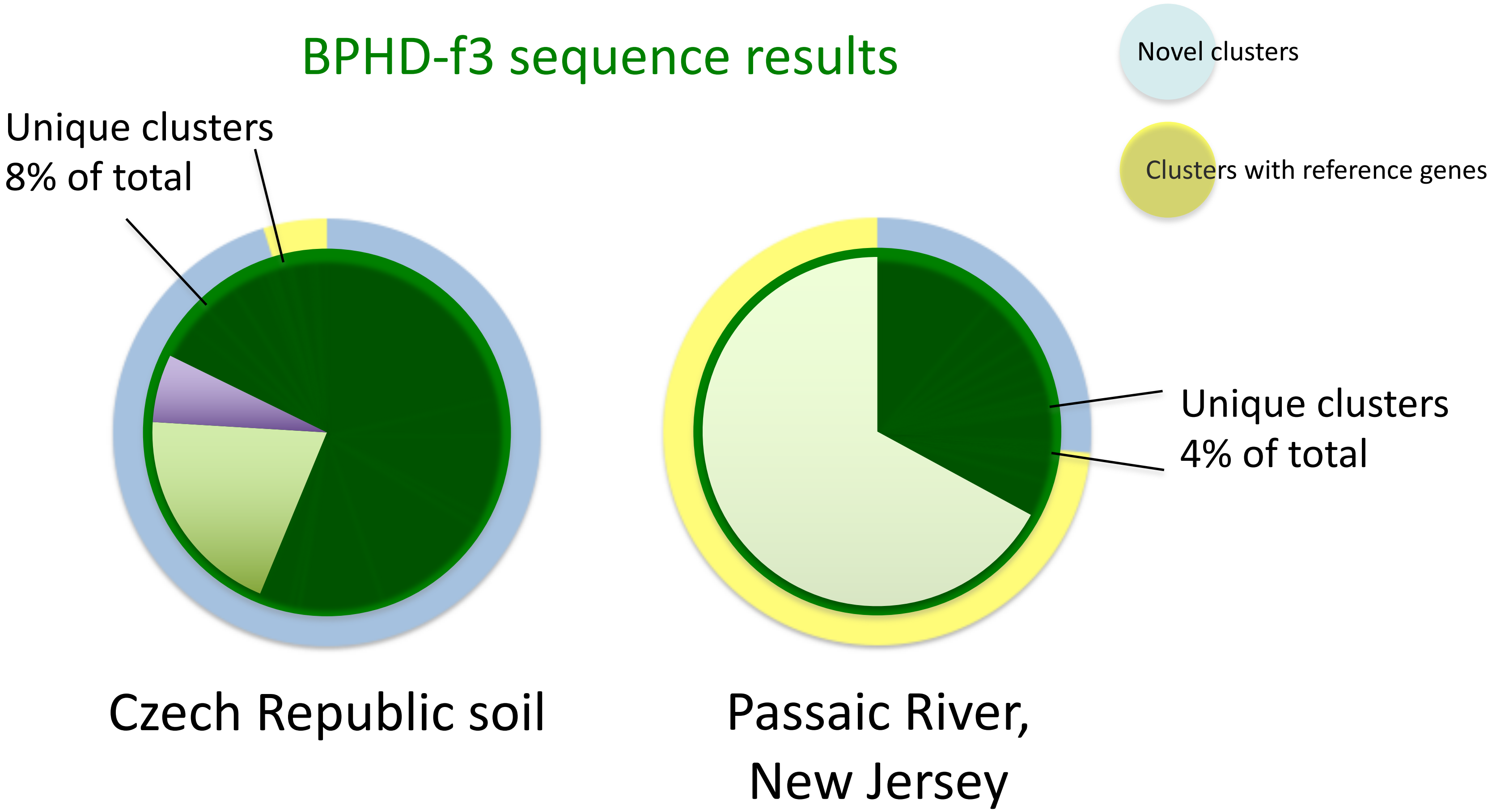


- ✓ Obtained sequences were distributed in many clusters
- ✓ 33 and 13 novel clusters were obtained for BPHD-f3 and BPHD-r1, respectively

Iwai, S.....Tiedje,
200, ISMEJ 4, 279

Dioxygenase cluster comparison of two sites

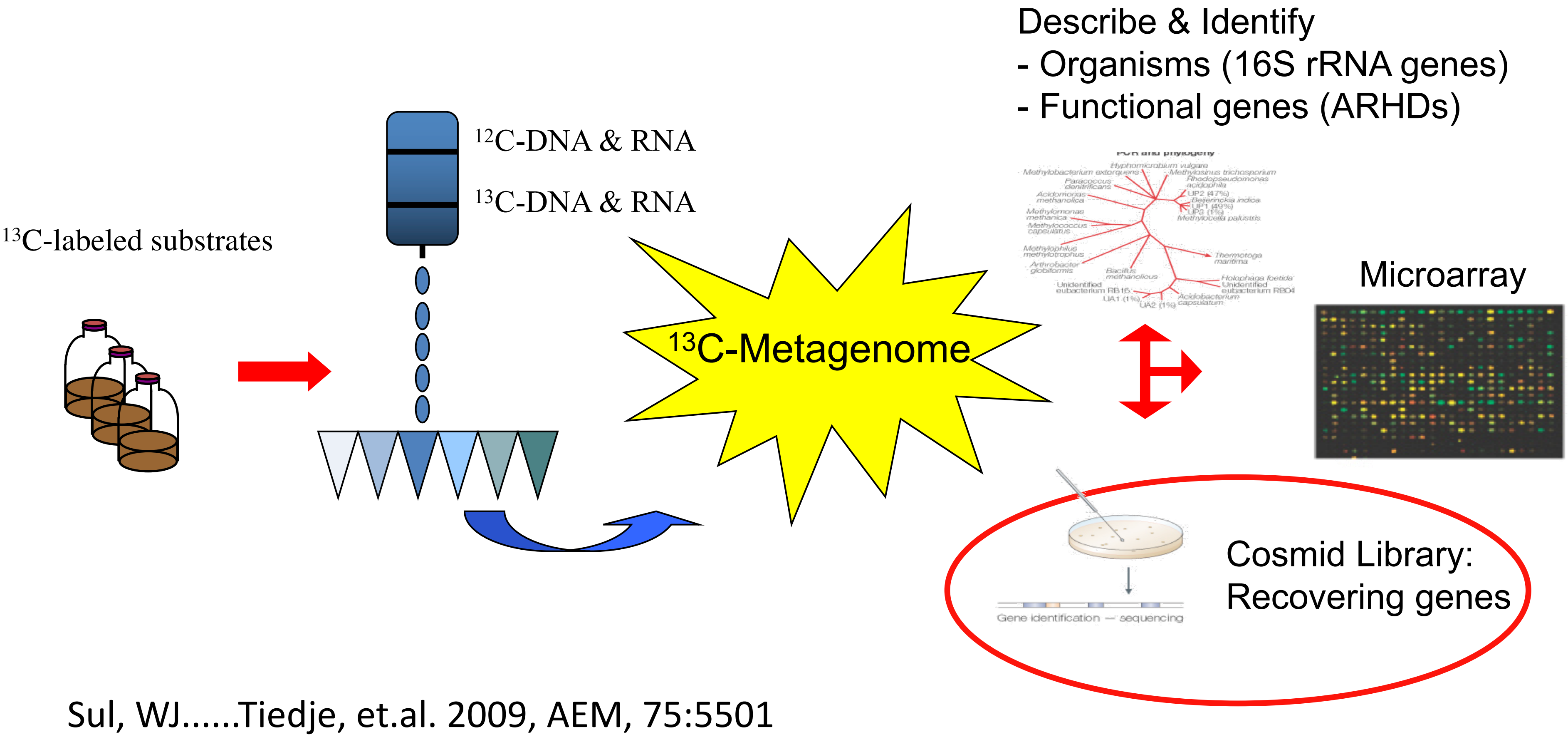
BPHD-f3 sequence results



- More than 90% of the total sequences were in common clusters
- Novel clusters found in Czech sample were also found in Passaic River
- Passaic River sequences were clustered more with reference genes

Stable Isotope Probing : Selecting Sub-Populations

- Separates and concentrates nucleic acids from populations that actively assimilate the labeled substrate
- Use ^{13}C-biphenyl, an analogue of PCB, as substrate.



^{13}C-labeled substrates

^{12}C-DNA & RNA

^{13}C-DNA & RNA

^{13}C-Metagenome

Describe & Identify
- Organisms (16S rRNA genes)
- Functional genes (ARHDs)

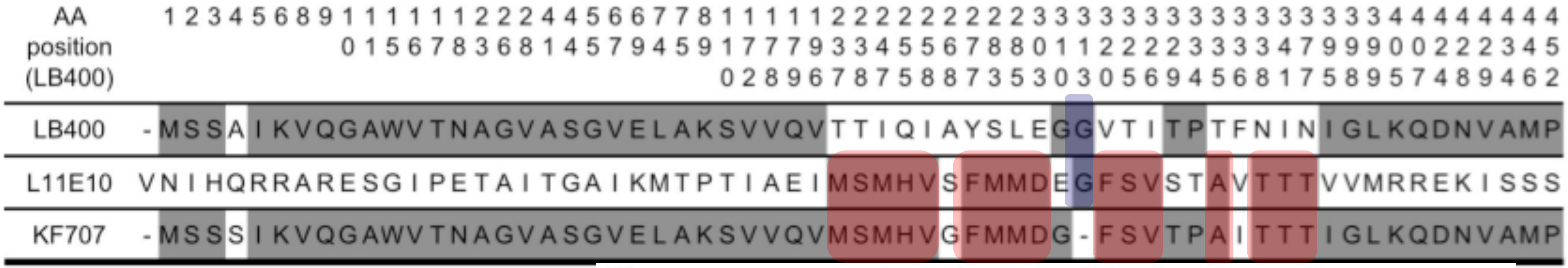
Microarray

Cosmid Library: Recovering genes

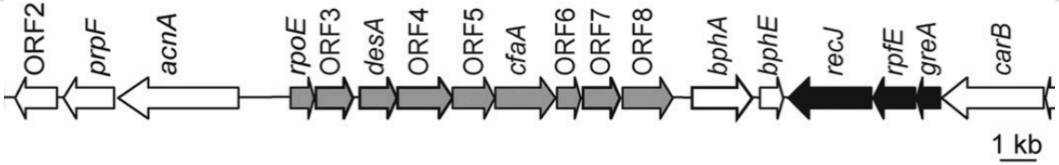
Sul, WJ......Tiedje, et.al. 2009, AEM, 75:5501

Structure and PCB Dioxyenase Activity of the Metagenomic Clone

Regions responsible for substrate specificity → I II III IV of BphA

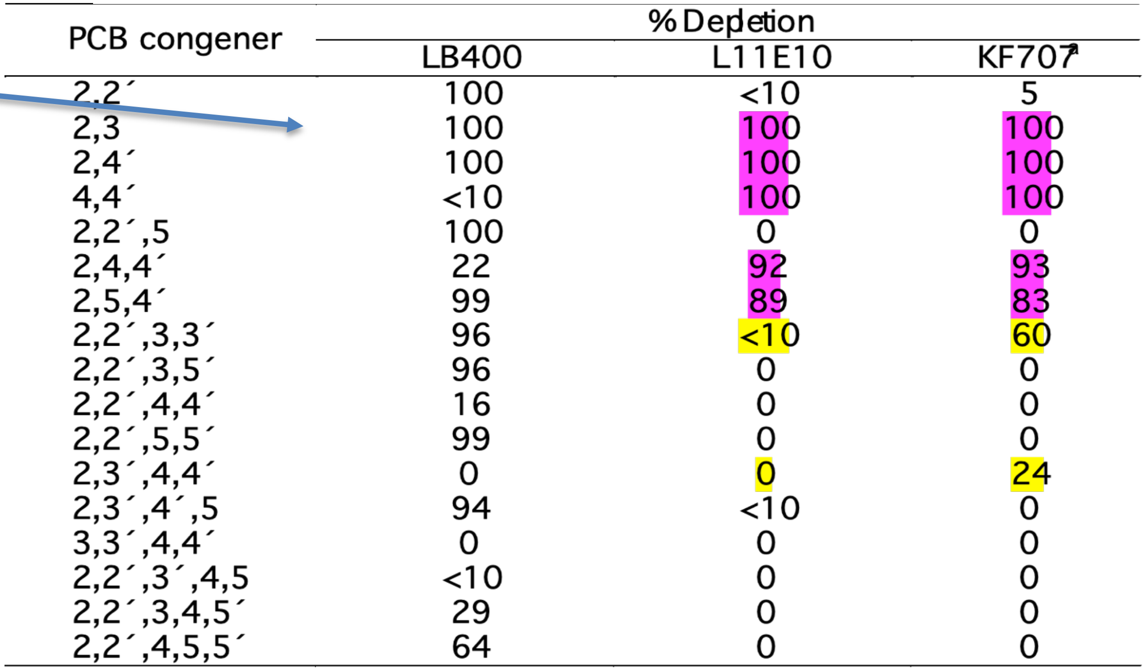


^{13}C enriched clone L11E10 →



- Dioxygenase activity on PCB congeners →

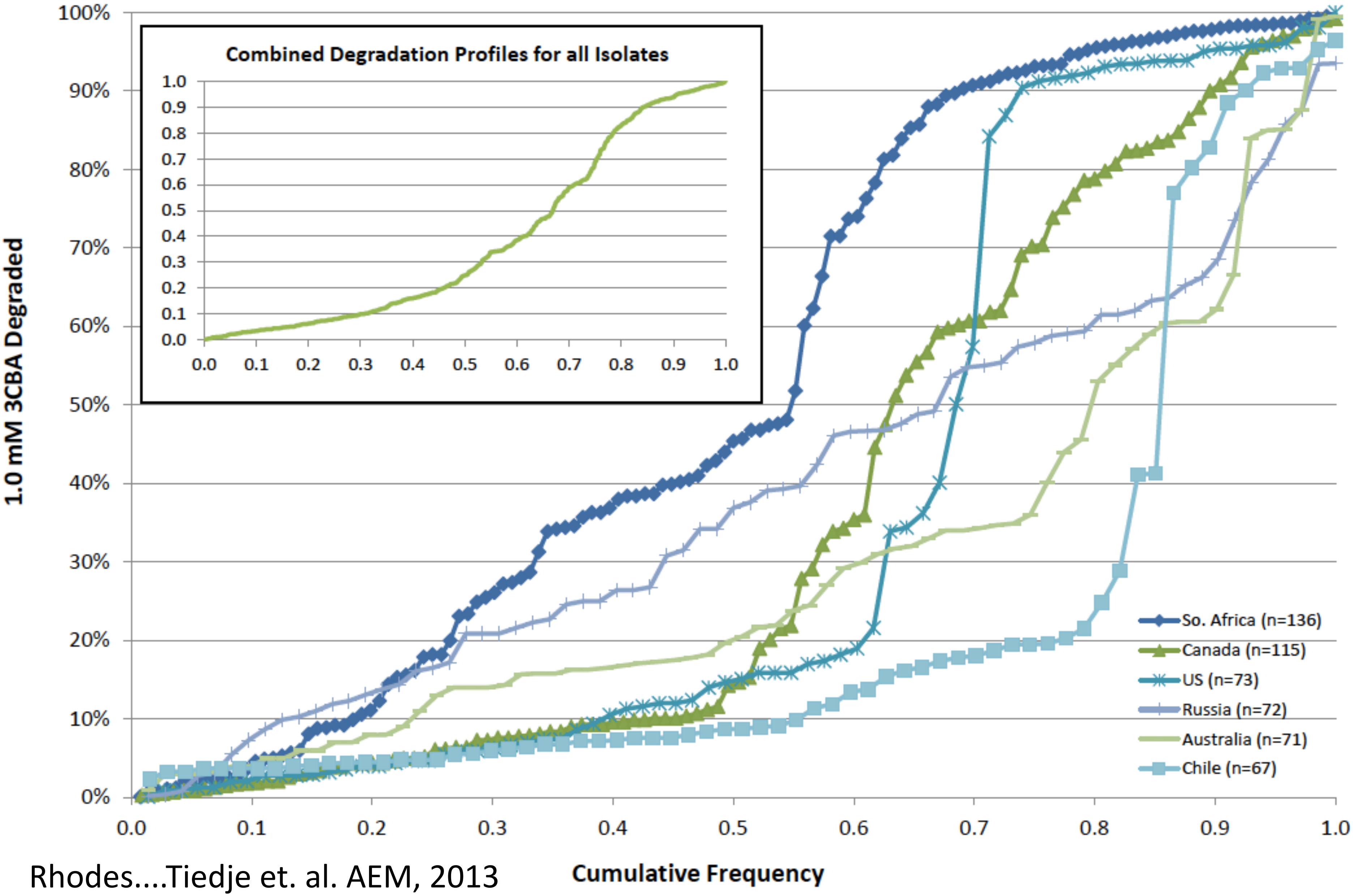
- L11E10-*bphAE* attack PCB rings that are **not substituted on one ring** or substituted at the **para position** only.



PCB congener	% Depletion		
	LB400	L11E10	KF707[a]
2,2´	100	<10	5
2,3	100	100	100
2,4´	100	100	100
4,4´	<10	100	100
2,2´,5	100	0	0
2,4,4´	22	92	93
2,5,4´	99	89	83
2,2´,3,3´	96	<10	60
2,2´,3,5´	96	0	0
2,2´,4,4´	16	0	0
2,2´,5,5´	99	0	0
2,3´,4,4´	0	0	24
2,3´,4´,5	94	<10	0
3,3´,4,4´	0	0	0
2,2´,3´,4,5	<10	0	0
2,2´,3,4,5´	29	0	0
2,2´,4,5,5´	64	0	0

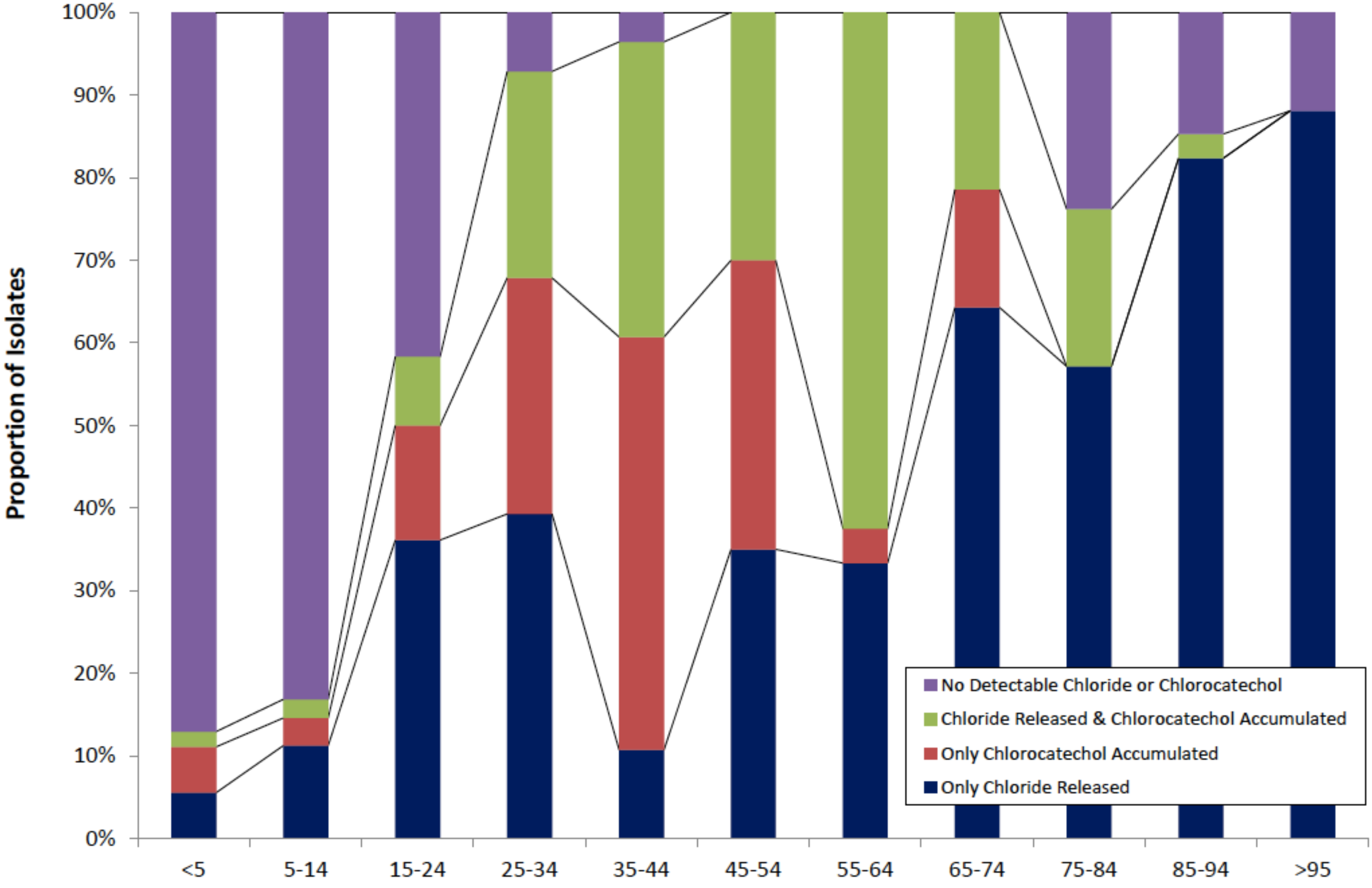
Sul,...Tiedje, et.al.
2009, AEM, 75:5501

Rank Order of Degradation Kinetics of 3-Chlorobenzoate Degraders



Rhodes....Tiedje et. al. AEM, 2013

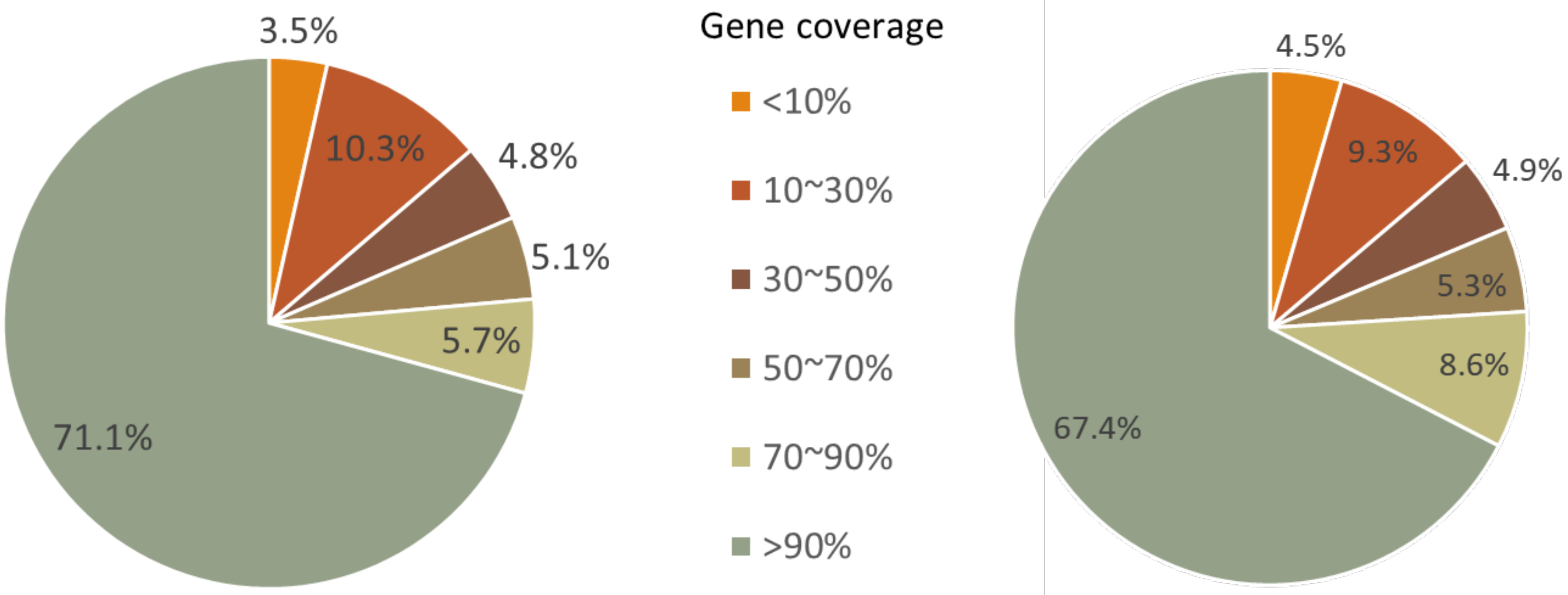
Many Strains Produce Intermediate Products (red and green colors)



Rhodes…Tiedje et. al. AEM, 2013

Advantages of long read sequencing: gene context

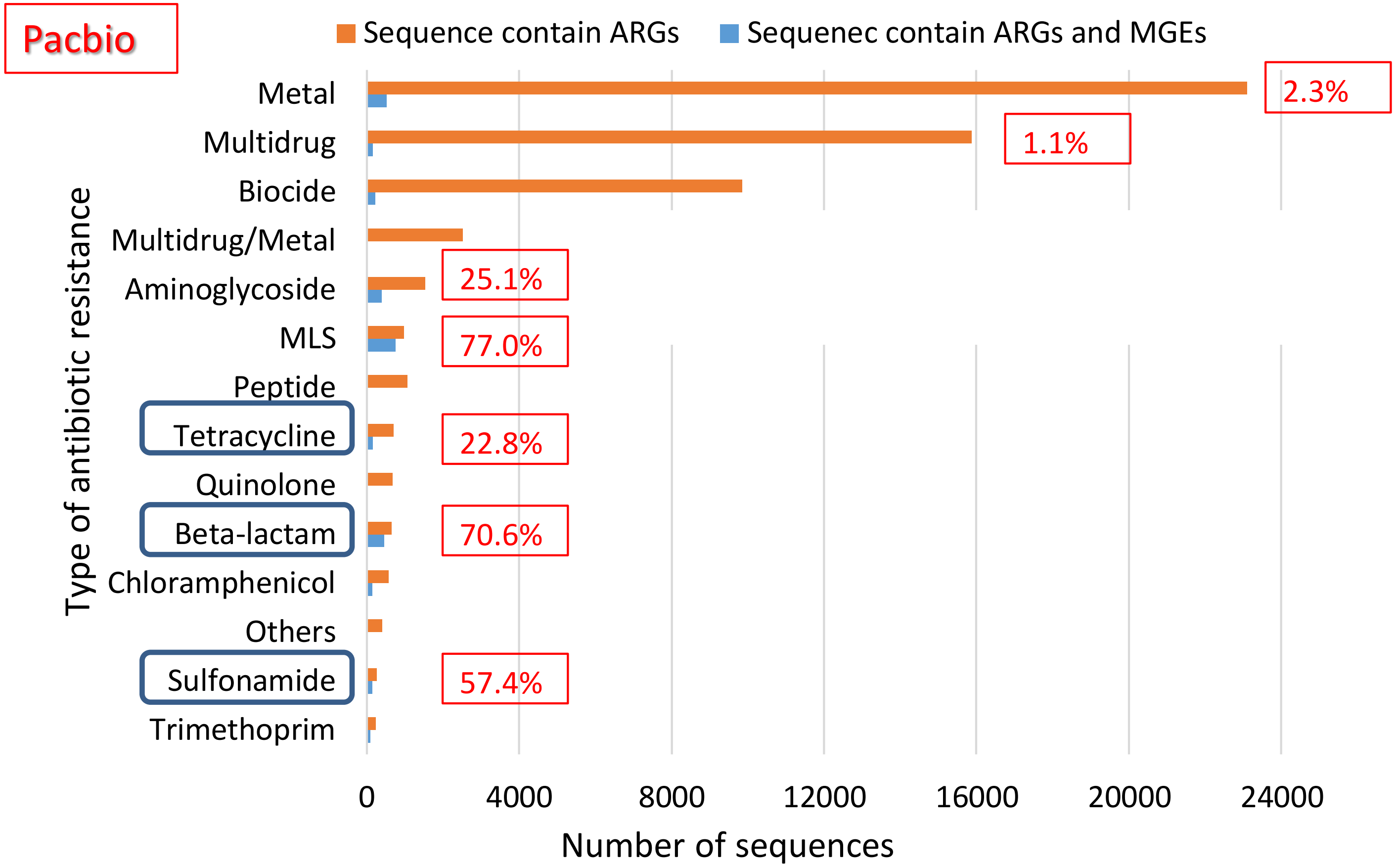
ARG gene coverage: distribution of Blastn hits



PacBio **Nanopore**

Reads:	232,649	1,920,385
# Sequences >10 ARGs & MGEs	212	49

Percentage of ARGs associated with MGEs



A small proportion of multidrug and metal resistance genes appear transferable

What About the Future?

What is Next for Our Science: 5 Years

- Return to the questions of the 1970s, 80s
  - Process measures, modeling, culture-based resolution
  - But with more resolving methods
- Hypothesis-driven microbiome science:
  - Ecology, Evolution, Function, Mechanism
- Long-read sequencing technology becomes standard
- Use of advanced microscopies, at the microbes' scale
    - image analysis
- More use of isotope methods, link to function
- Better habitat characterization & its reporting
- SAGs and MAGs become part of organized taxonomy
- Use of whole genome information, not only 16S rRNA
- Meta analyses of the decade of 'diversity' data

Longer–term Future Prospects

- In field, rapid, e-connected diagnostics, AI guidance
  - Image analysis
  - enhanced metabolite/chemical analyses
  - multiple omics markers
- AI, Machine Learning, Deep learning neural networks
- HTP culturing: ecol, physio & biochem characterization
- DNA markers for kinetic properties; competition
- HTP direct protein sequencing
- Real-time, e-linkage of omics with function
- CRISPR-*cas9* directed microbe design
- Particular microbiome & management ownership
- Synthetic ecobiology; designer communities
- Quantum computing: complex system-complex solutions

Predicting the Future is Hard

but

It is better to try

It helps you be prepared