Microbial Taxonomy and Phylogeny: Extending from rRNAs to Genomes

Dr. Kostas Konstantinidis

Department of Civil and Environmental Engineering &
Department of Biology (Adjunct),
Center for Bioinformatics and Computational Genomics

Georgia Institute of Technology

ICCC12 Conference

Florianópolis, Brazil 2010

Outline

- Translating old standards to sequence: The ANI approach
- New insights into the species issue from natural populations
- Assessing the whole prokaryotic diversity & taxonomy
- Conclusions & Perspectives

How are species/units defined?

The most popular species definition

"…a genomically coherent (discreet) group of strains based on the hybridization of their purified DNA molecules"

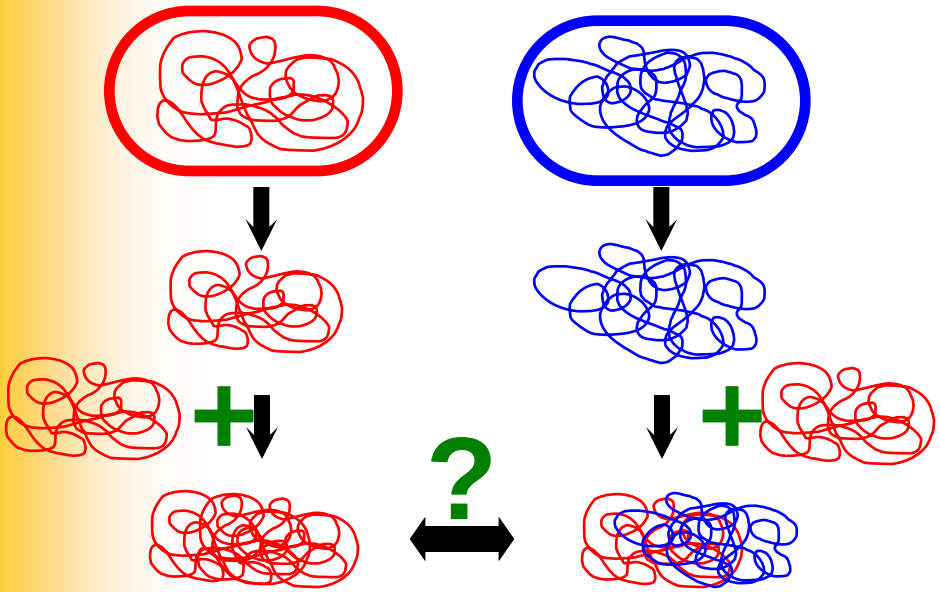
…Plus, a diagnostic phenotype

Wayne et al. IJSB, 1987

The DNA-DNA hybridization method

DDH general principle

- Isolate genomic DNA from strains A and B
- Random fragmentation
- Denature DNA
- Mix and let renature
- Quantify heteroduplex relative to homoduplex



- >70% => SAME species

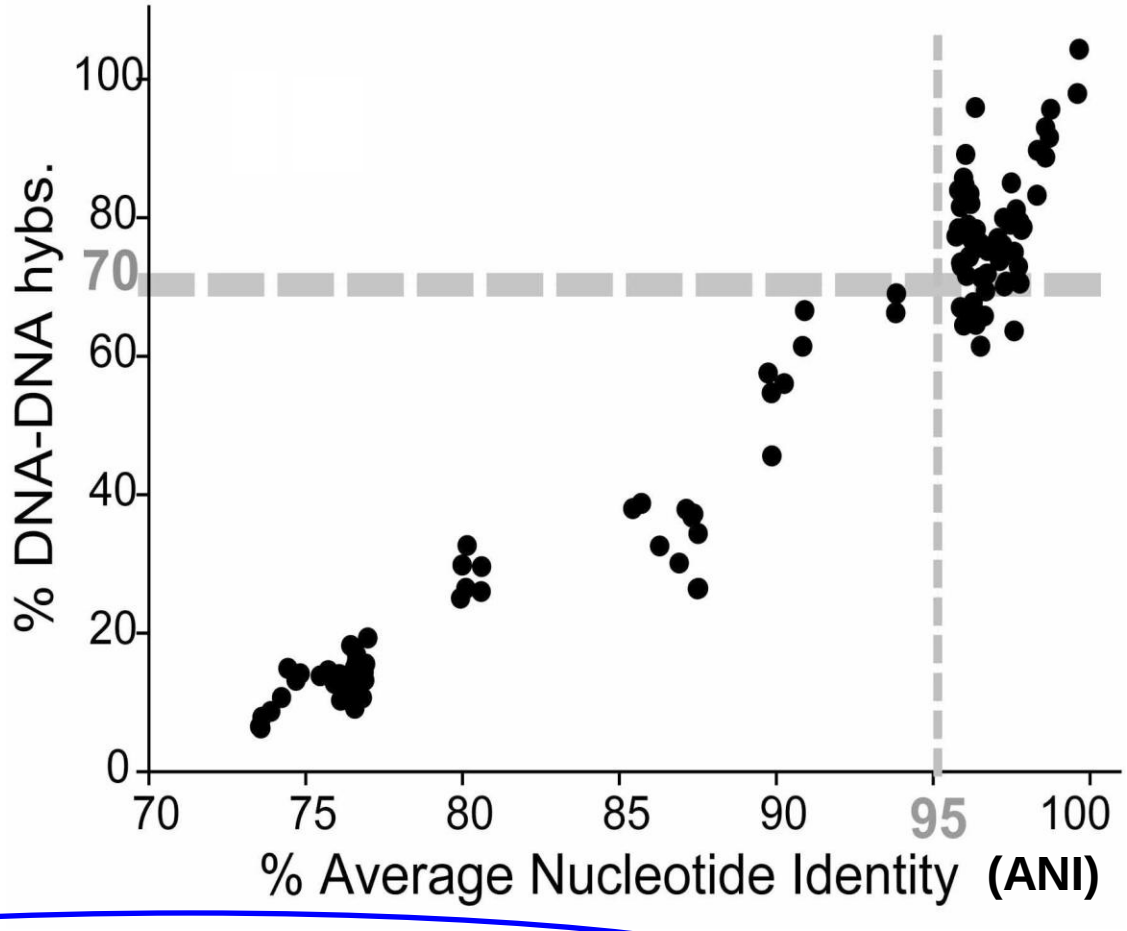
- Good correspondence with phenotypically coherent clusters of strains in *Enterobacteriaceae*

BUT

- Difficult to do!
- Unclear how it relates to whole-genome relatedness.
- Need to have isolates available…but only 1-2% of prokaryotic cells are cultivable! (the great plate count anomaly)

DDH vs. whole-genome sequence relatedness

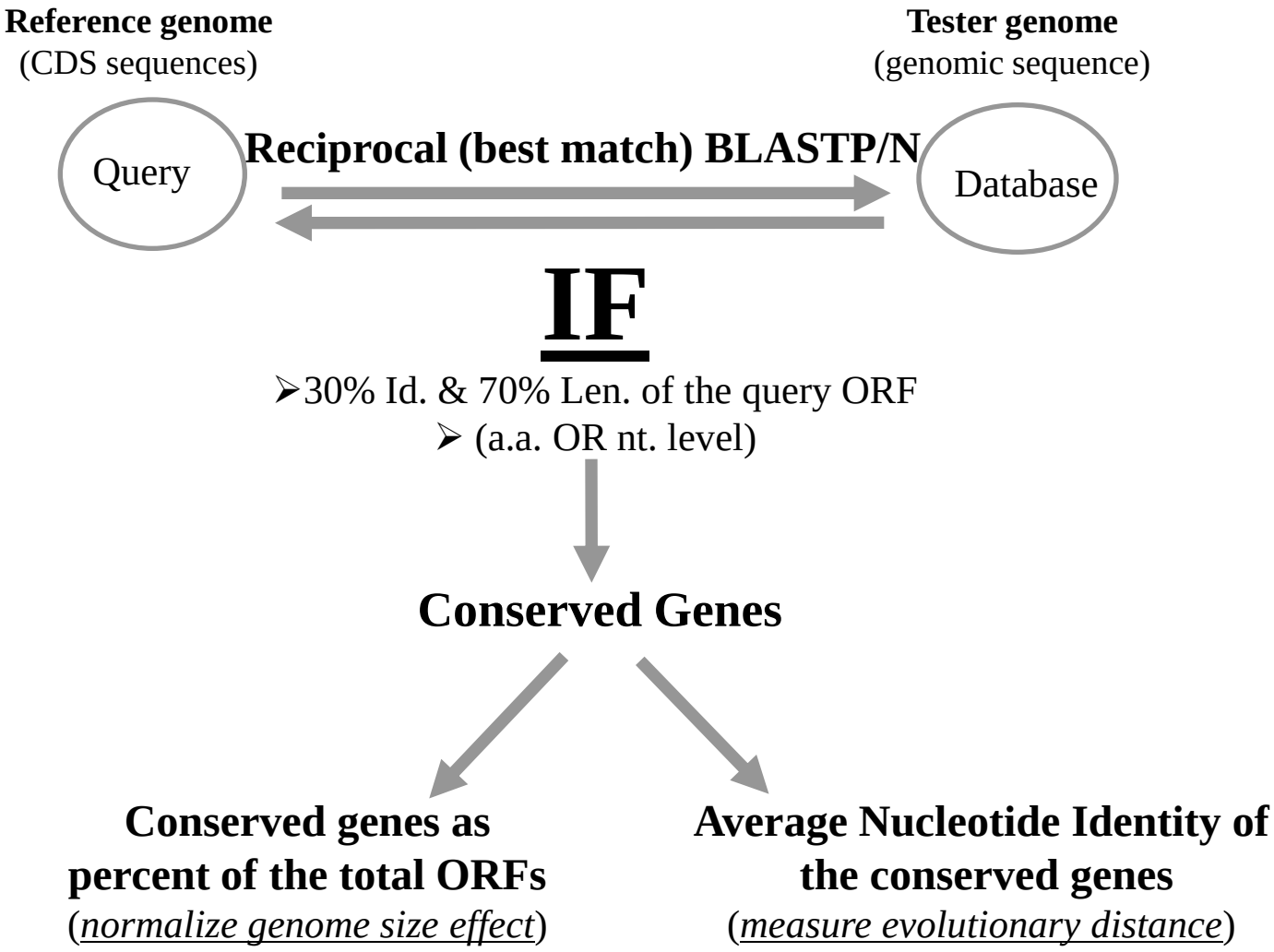
DDH vs. whole-genome Average Nucleotide Identity (ANI)



70% DDH <=> 95% ANI!

Goris, Konstantinidis, et al. IJSEM, 2007

ANI to measure relatedness



ANI vs. Maximum Likelihood (ML)

Within A Group

- Determine the conserved gene core based on RBM blast searches.
- Build a clustalw concatenated alignment of all core genes (>2,000).

↓

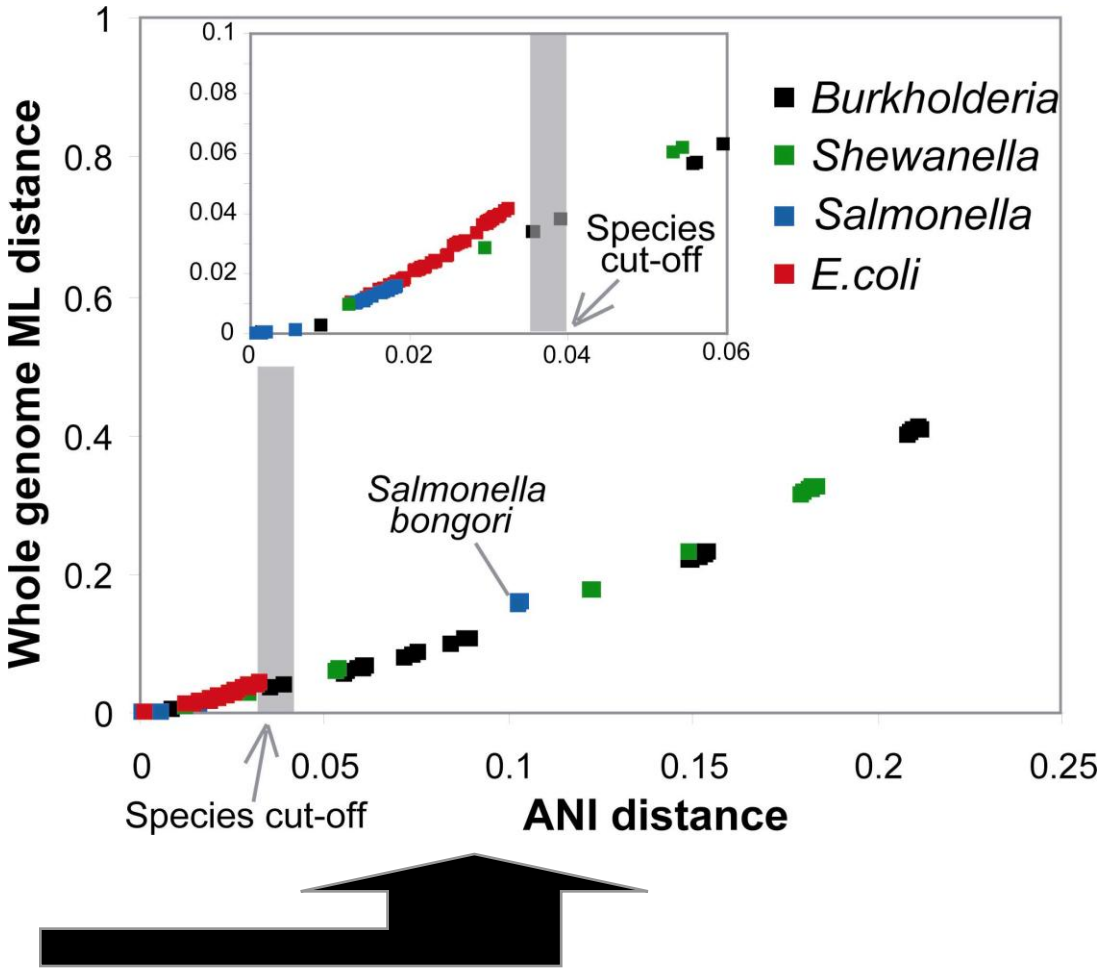
Determine the best model for sequence evolution by ModelTest.

↓

Build core-based ML phylogeny and calculate the distances between the genomes of the group.

↓

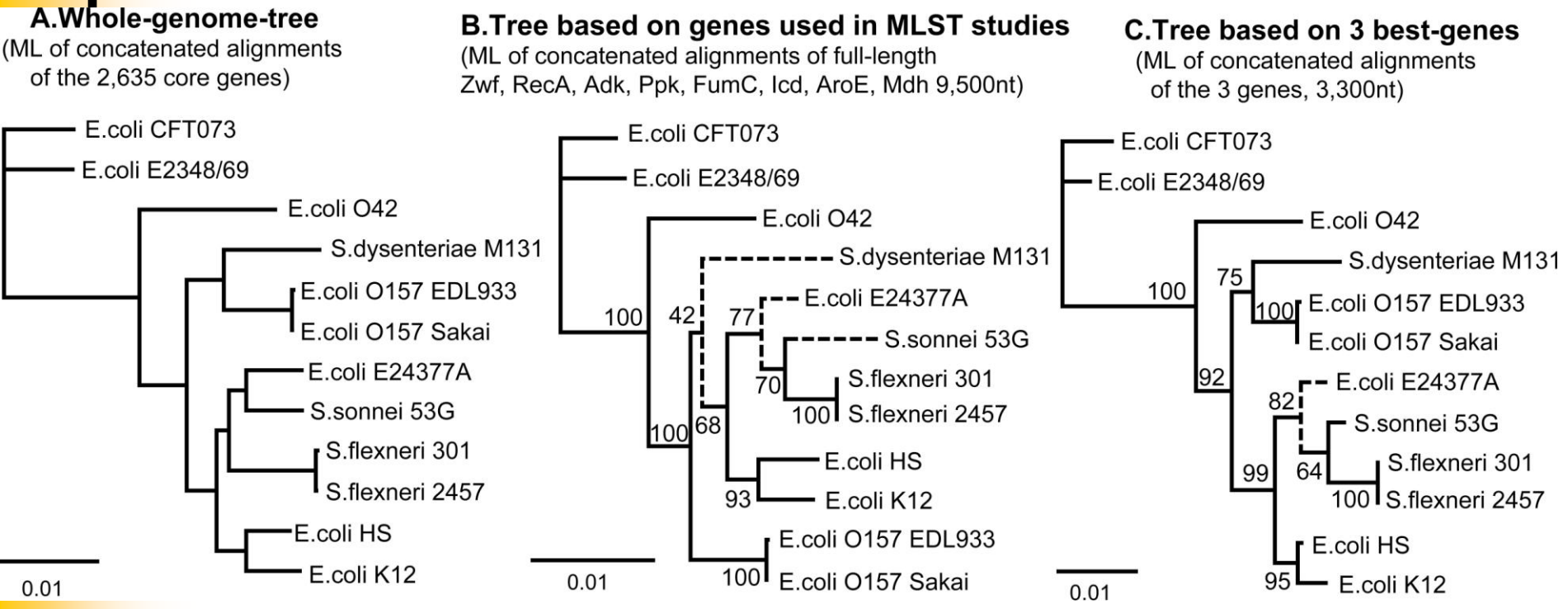
Compare ML to ANI (calculate on the same, core genes) distances.



- Highly robust,
- Simple and universal,
- Resolution within species

Konstantinidis, Ramette, &Tiedje, Applied & Environ. Micro. 2006.

MLSA vs. 3-best genesin the genome



- Classical MLST: Significantly different by KH test (p-value 0.02)
- 3 Best-genes: NOT significantly different by KH test (p-value 0.709)

- Resolution at the strain level with (just) three genes!

Konstantinidis, Ramette, &Tiedje, Applied & Environ. Micro. 2006.

How are species/units defined?

The most popular species definition

"…a genomically coherent (discreet) group of strains based on the hybridization of their purified DNA molecules"

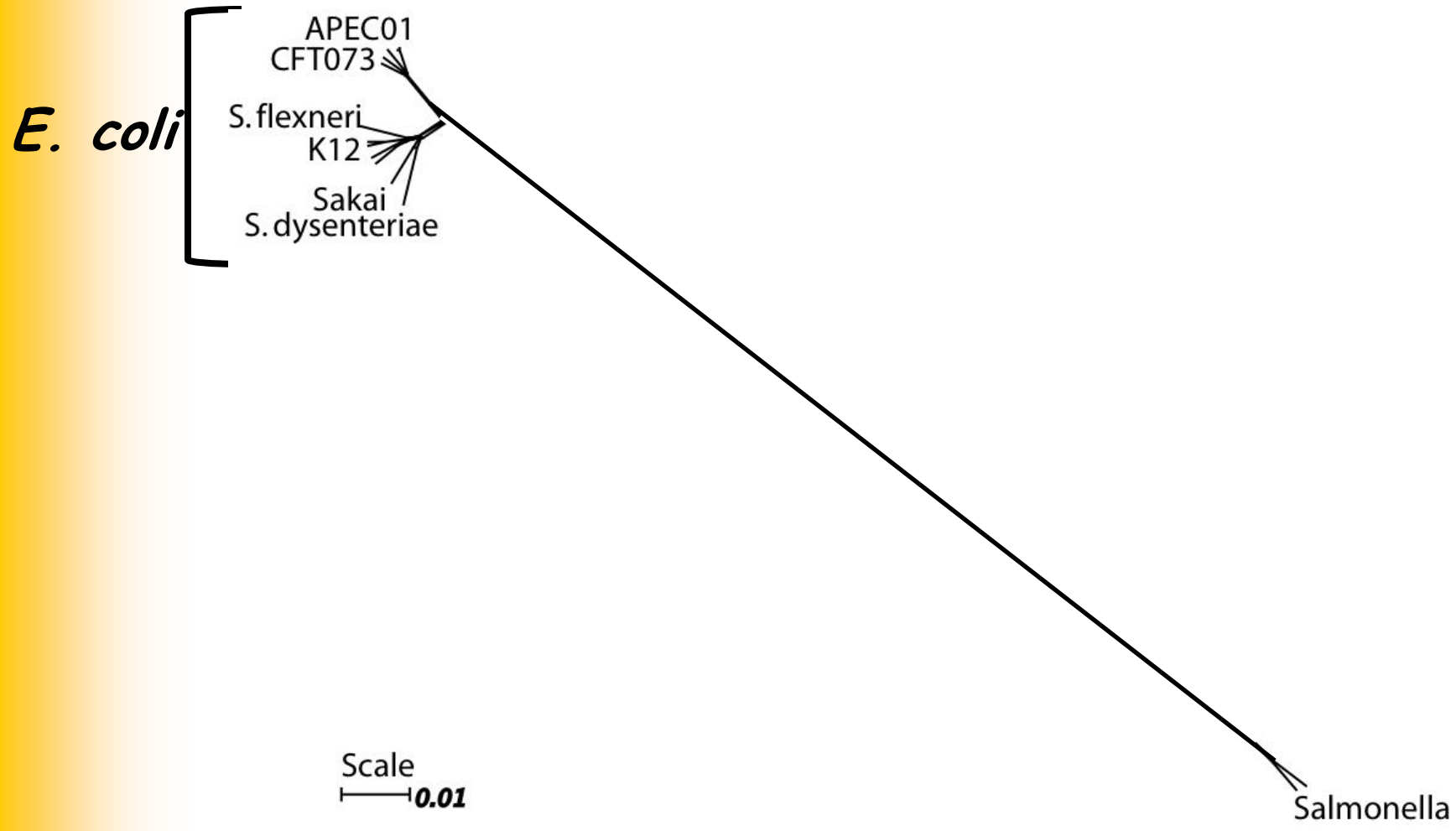
…Plus, a diagnostic phenotype

Wayne et al. IJSB, 1987

Genetic continuum OR discreet clusters?

Whole-genome Max. Lik. phylogeny

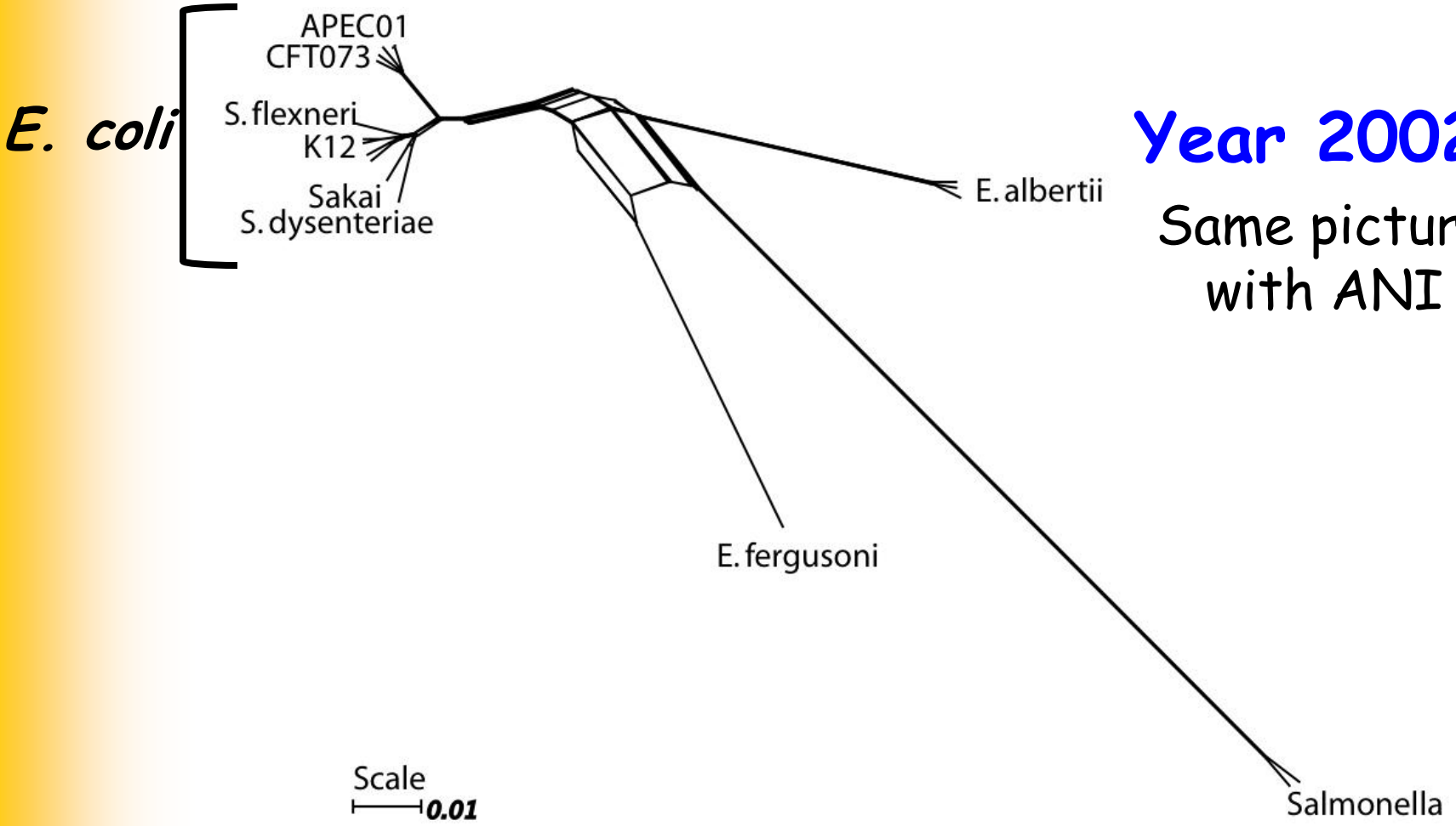
(based on 2,000 core genes)



Genetic continuum OR discreet clusters?

Whole-genome Max. Lik. phylogeny

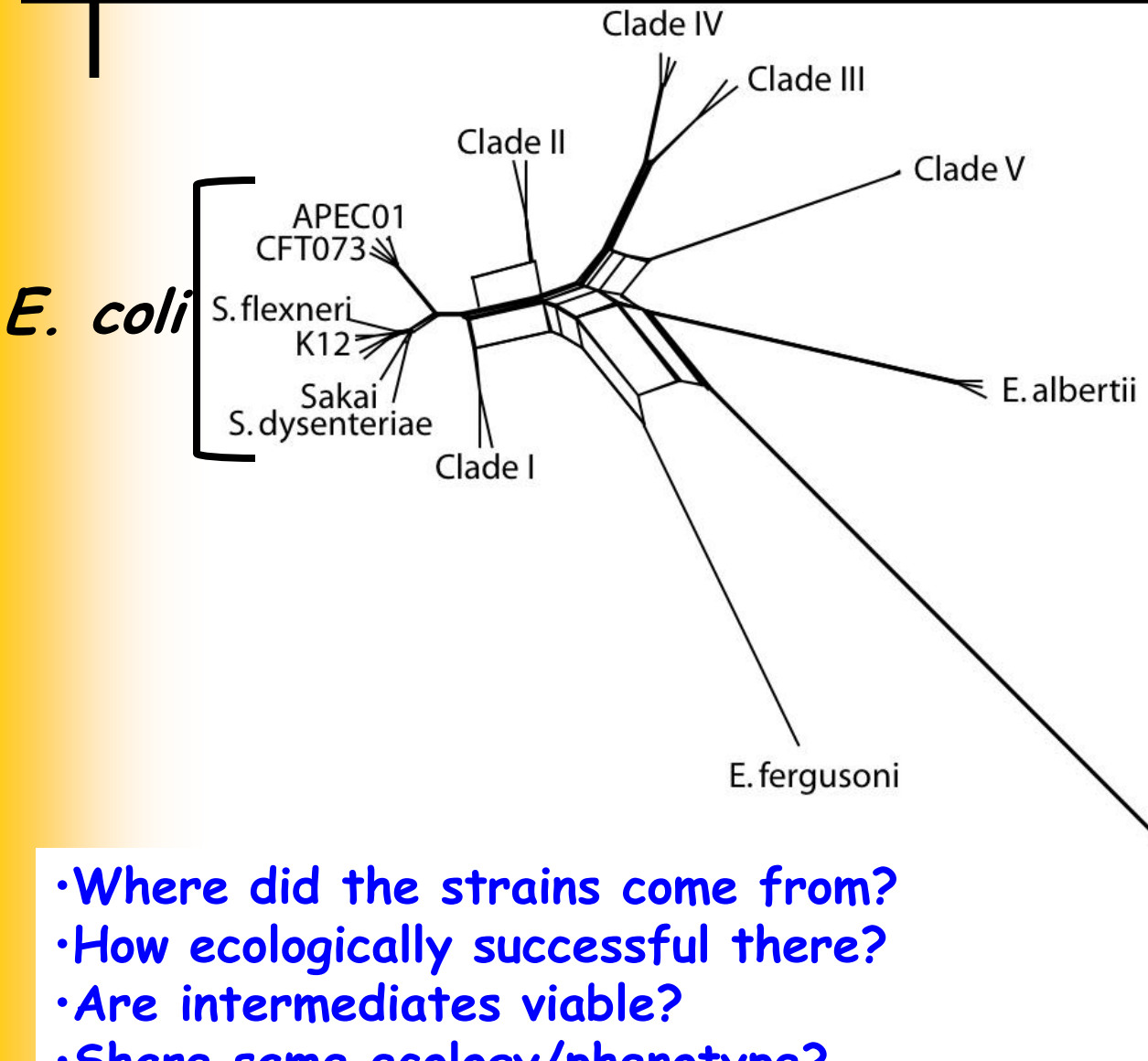
(based on 2,000 core genes)



Year 2002

Same picture with ANI

Genetic continuum OR discreet clusters?



Year 2010

Same picture with ANI

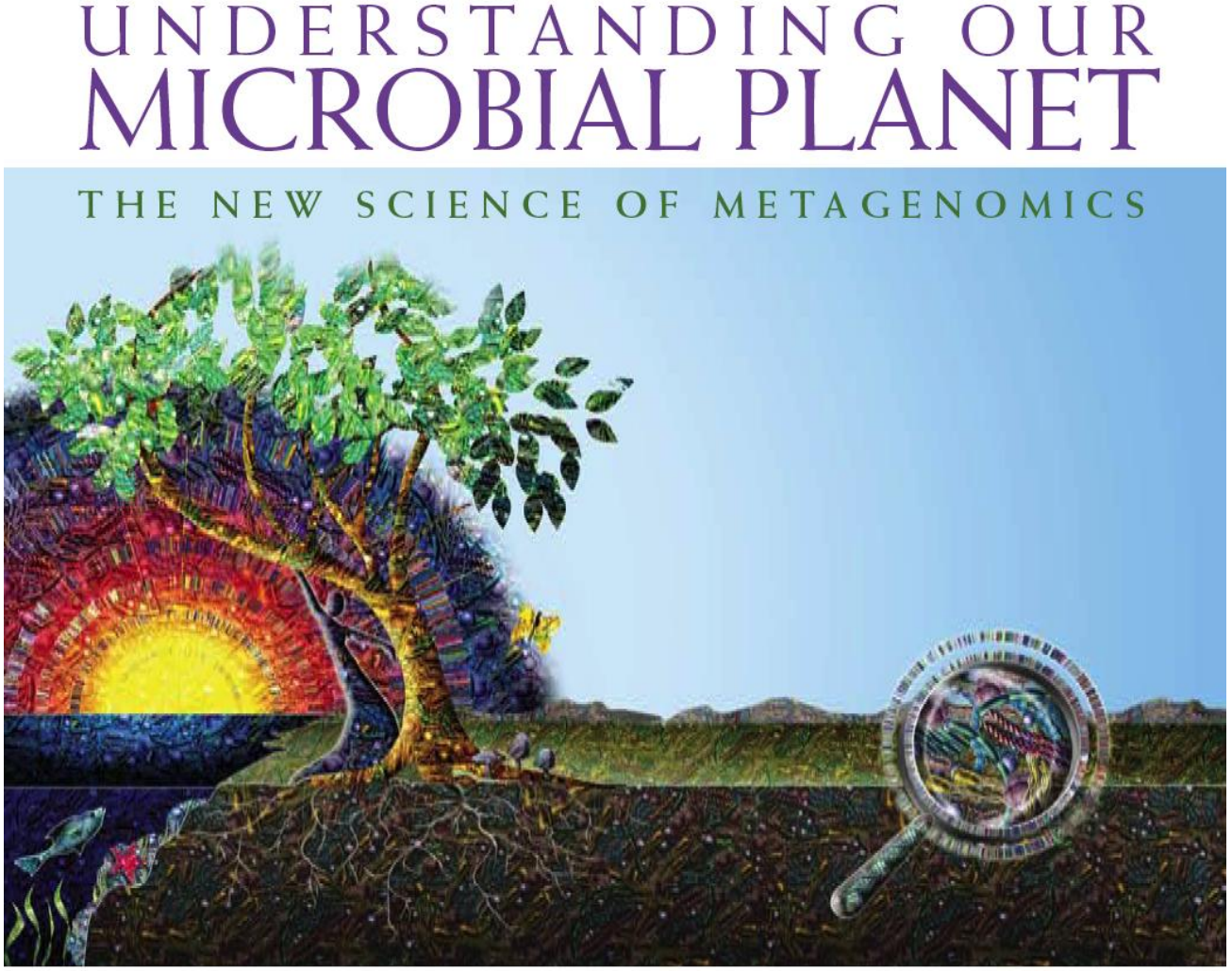
- Where did the strains come from?
- How ecologically successful there?
- Are intermediates viable?
- Share same ecology/phenotype?

Luo, Konstantinidis, et al., submitted

Outline

- Translating old standards to sequence: The ANI approach
- New insights into the species issue from natural populations
- Assessing the whole prokaryotic diversity & taxonomy
- Conclusions & Perspectives

How to study natural populations?



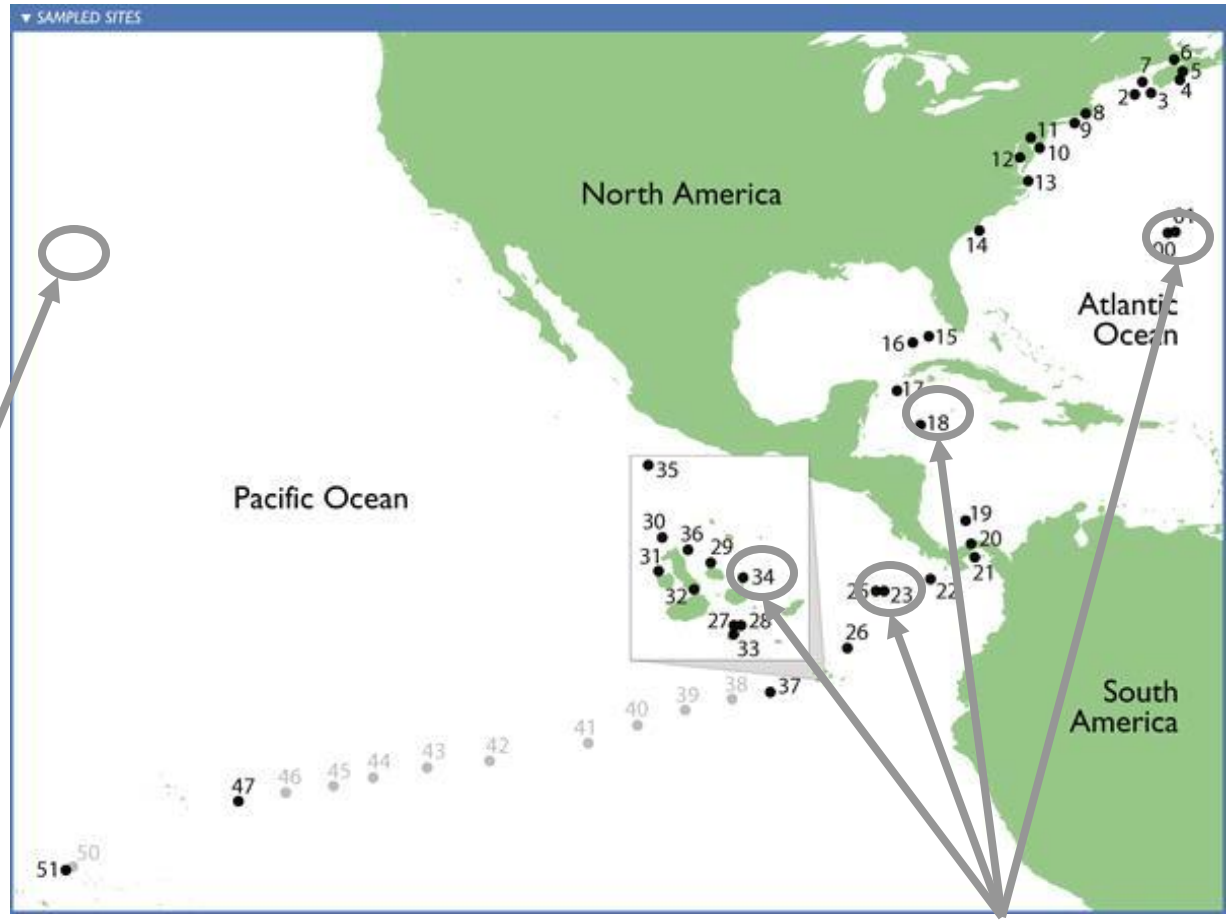
What is metagenomics?

"the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species"

Handelsman et al. Chemistry Biology, 1998

Metagenomic sampling of the Oceans

Global Ocean Survey (GOS) sampling sites

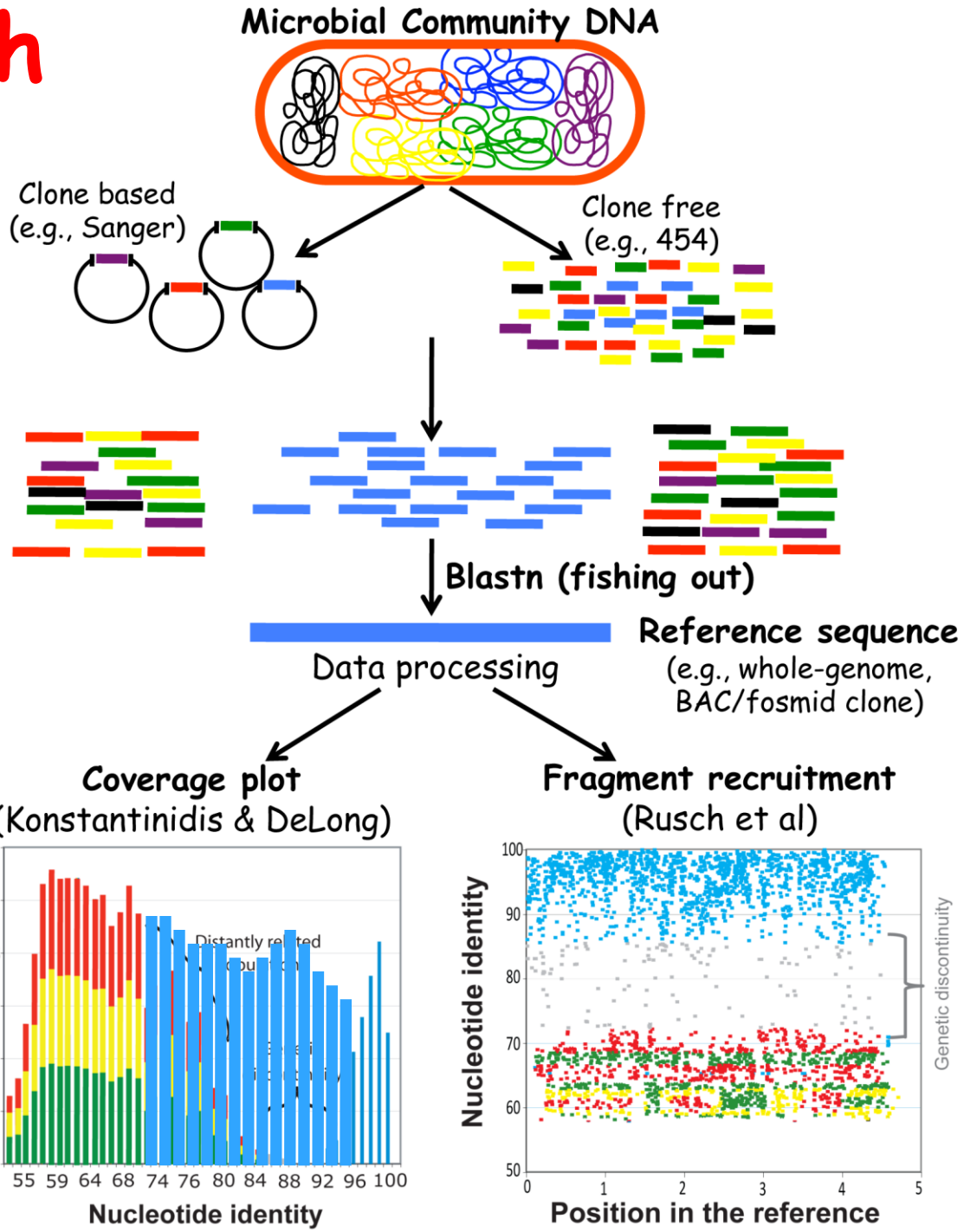


Hawaii Ocean Time Series
~200Mbp of shotgun library
Deep, 4000m depth
(Konstantinidis & DeLong, ISME 2008)

SAR3, SAR4, GOS18, GOS23, GOS34
~200-300Mbp of shotgun library
Surface, ~5m depth
(Rusch et al, PLoS Biology 2007)

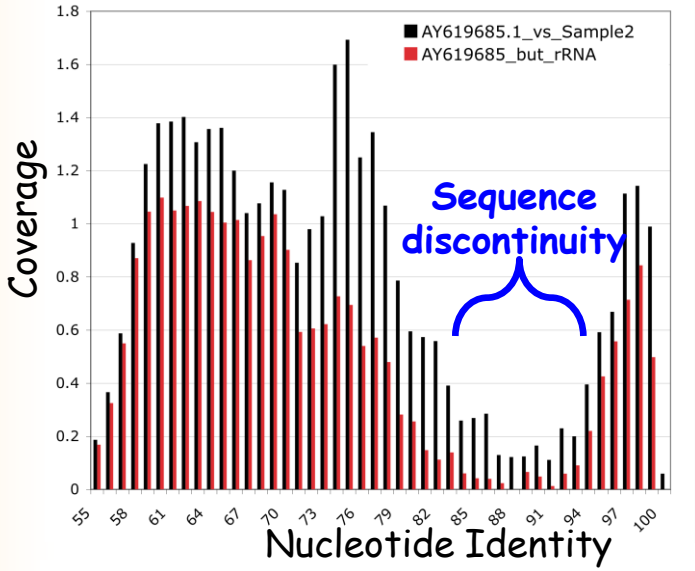
The approach

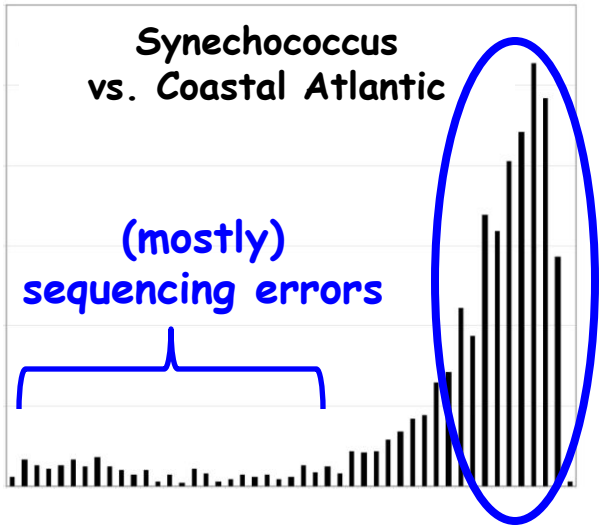
From Konstantinidis, 2010
In: Handbook of Molecular Microbial Ecology. Volume I Metagenomics and Complementary Approaches. Editor: Frans J. de Bruijn



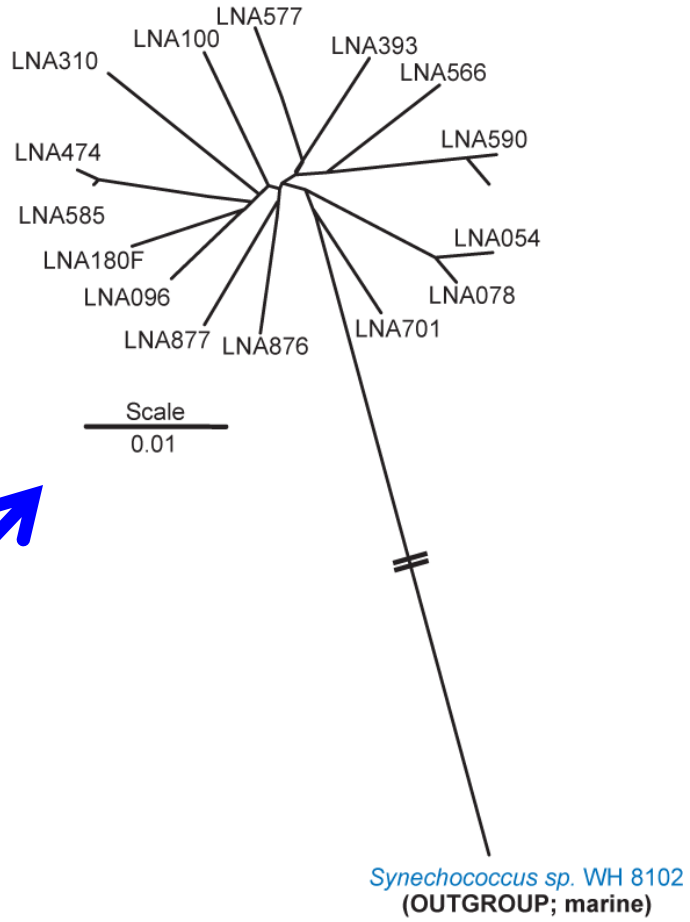
Clusters are ubiquitous in the Oceans!

BAC clone from uncultured γproteobacterium vs. Surface shotgun

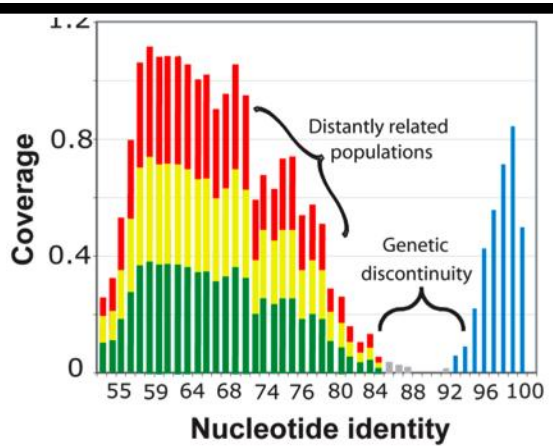




Based on a phylogenetic approach too!



Species-like populations!



?

- Sequence-discrete populations
- Smaller intra-population gene-content differences compared to several named species (<5% vs 20-30%)
- Detectable intra-genomic homologous recombination; albeit lower levels compared to biofilm communities (e.g., AMD)

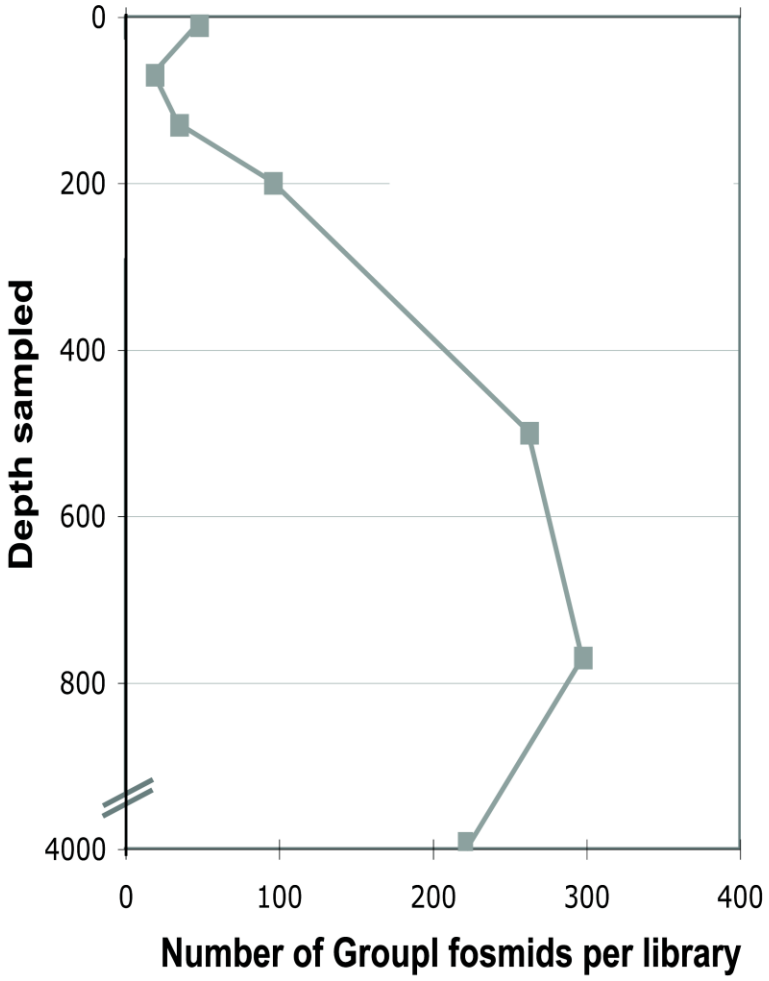
Genetic discontinuity frequently @ 95% ANI!

More details: Konstantinidis &DeLong, The ISME Journal, 2008

What about populations from different environments or sites?

GroupI *Crenarchaea* in the Oceans

Archaeal dominance in the mesopelagic zone of the Pacific Ocean



Constituting up to 20% of microbial cells in the sub-photic zone

Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*

Steven J. Hallam*†, Konstantinos T. Konstantinidis*, Nik Putnam‡, Christa Schleper§, Yoh-ichi Watanabe¶, Junichi Sugaharaǁ, Christina Preston**, José de la Torre††, Paul M. Richardson‡, and Edward F. DeLong*‡‡

Hallam Konstantinidis et al. PNAS 2006

Pathways of Carbon Assimilation and Ammonia Oxidation Suggested by Environmental Genomic Analyses of Marine *Crenarchaeota*

Steven J. Hallam[1], Tracy J. Mincer[1], Christa Schleper[2], Christina M. Preston[3], Katie Roberts[4], Paul M. Richardson[5], Edward F. DeLong[1]*

Hallam et al. Plos Biol. 2006

Isolation of an autotrophic ammonia-oxidizing marine archaeon

Martin Könneke[1]*†, Anne E. Bernhard[1]*†, José R. de la Torre[1]*, Christopher B. Walker[1], John B. Waterbury[2] & David A. Stahl[1]

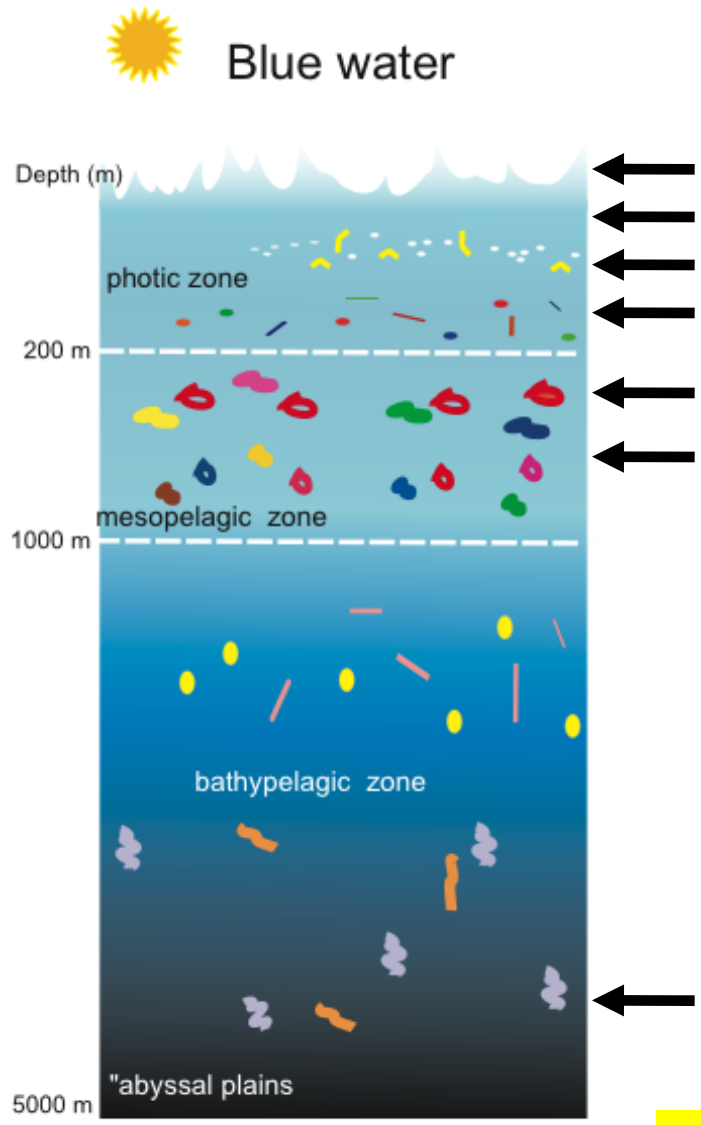
Wuchter et al. PNAS. 2006

Archaeal nitrification in the ocean

Cornelia Wuchter*, Ben Abbas*, Marco J. L. Coolen*†, Lydie Herfort*, Judith van Bleijswijk*, Peer Timmers*, Marc Strous*, Eva Telra*§, Gerhard J. Herndl*, Jack J. Middelburg¶, Stefan Schouten*, and Jaap S. Sinninghe Damsté*¶

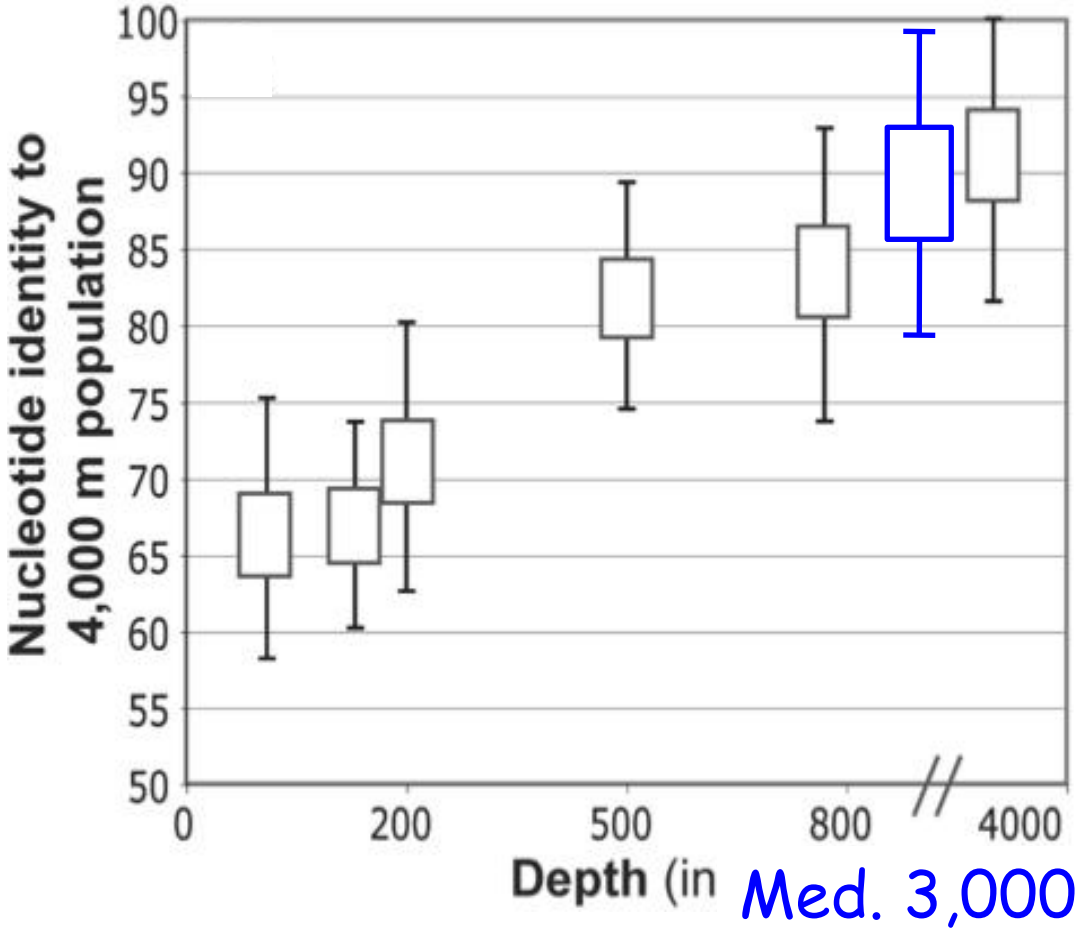
Konneke et al. Nature, 2005

How are they distributed with depth?



DeLong et. al. Science 2006

Nucleotide identity against 4,000m crenarchaeal population



Med. 3,000m

Konstantinidis et al.,
Applied & Environmental Microbiology 2009

Outline

- Translating old standards to sequence: The ANI approach
- New insights into the species issue from natural populations
- Assessing the whole prokaryotic diversity & taxonomy
- Conclusions & Perspectives

The current taxonomic system

- Prokaryotic Taxonomy: Classification, Identification, Nomenclature

- There is no official prokaryotic taxonomy. Bergey's is the closest approximation to this.

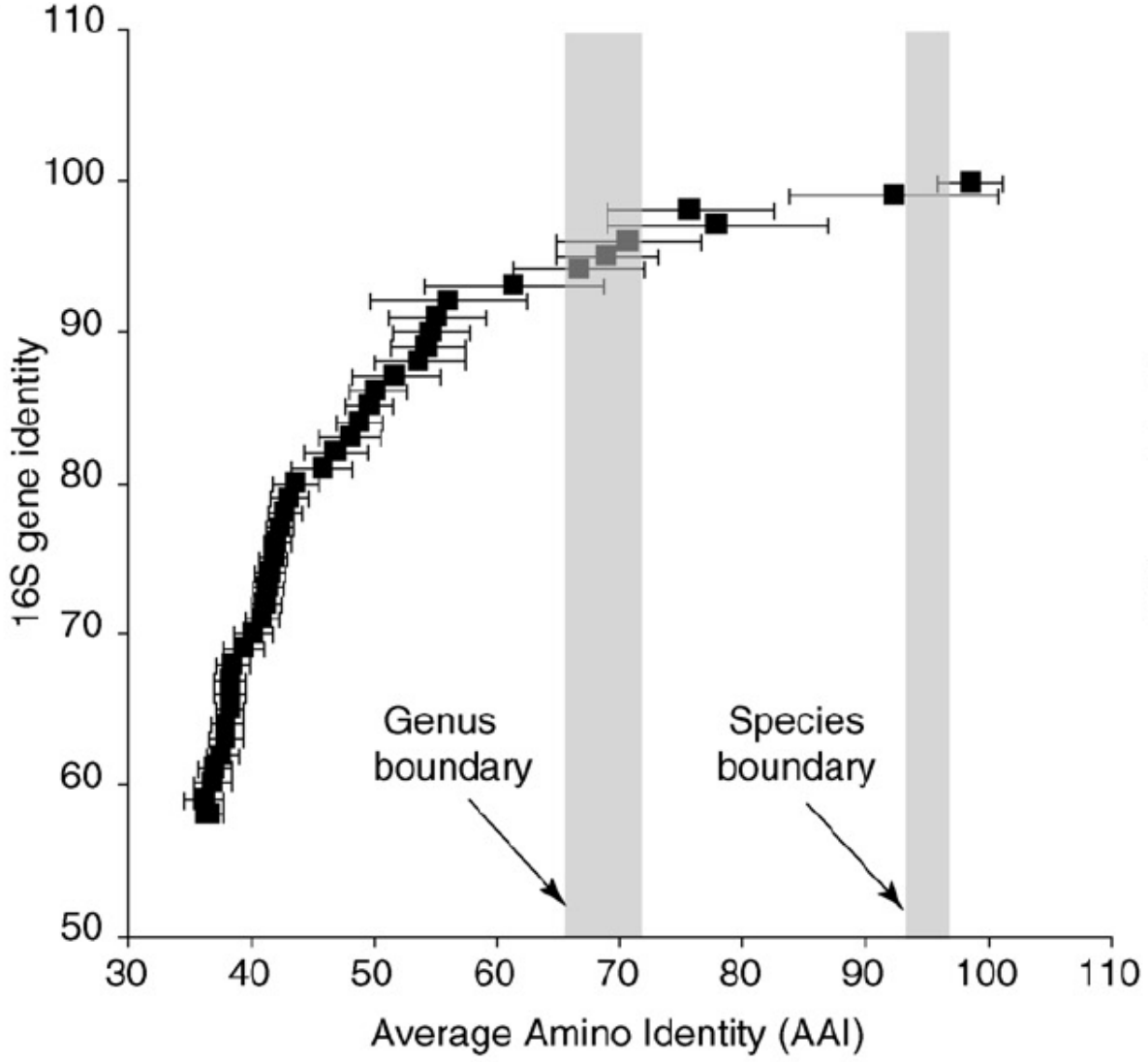
- 8 recognized taxonomic ranks.

Domain
Phylum
Class
Order
Family
Genus
Species
Subspecies

- Current classification is primarily based on 16S rRNA and secondarily on classical microscopic and biochemical observations.

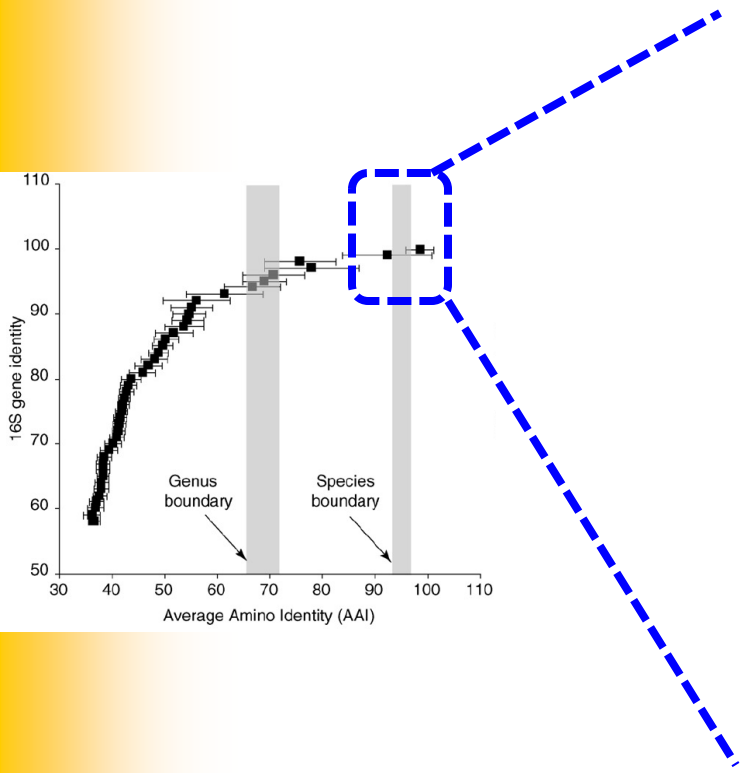
- How well does 16S rRNA represent the whole genome phylogeny?

16S rRNAvs whole-genome relatedness



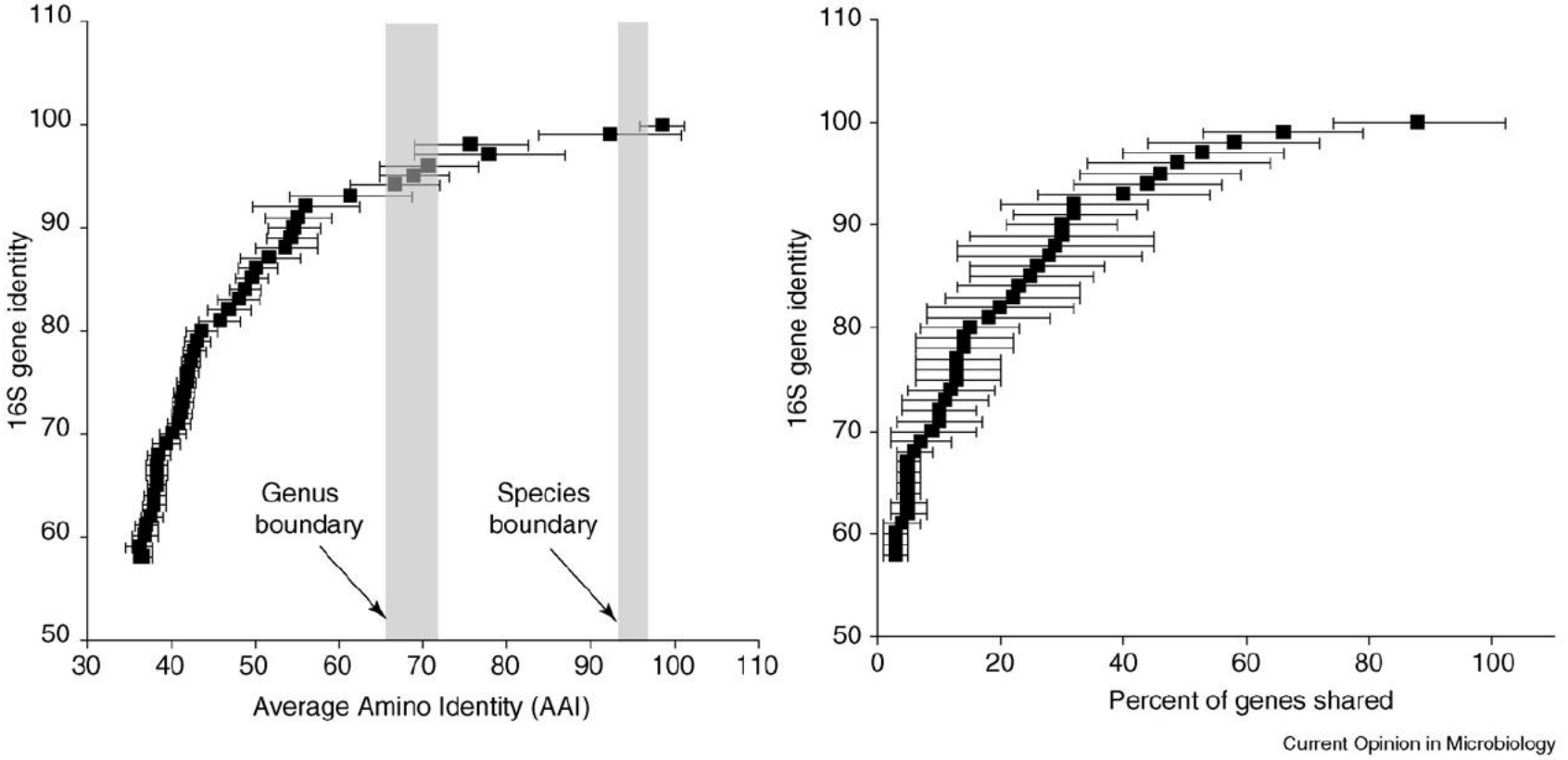
From Konstantinidis and Tiedje, Current Opinion in Microbiology, 2007

Zooming in at the species level



>70% DDH =>>95% ANI =>>98.5% 16S rRNA

16S rRNAvs gene-content relatedness



Genetic distances correspond to (even greater) gene content differences!!

From Konstantinidis and Tiedje, Current Opinion in Microbiology, 2007

The current taxonomic system

•Prokaryotic Taxonomy: Classification, Identification, Nomenclature

•There is no official prokaryotic taxonomy. Bergey's is the closest approximation to this.

•8 recognized taxonomic ranks.

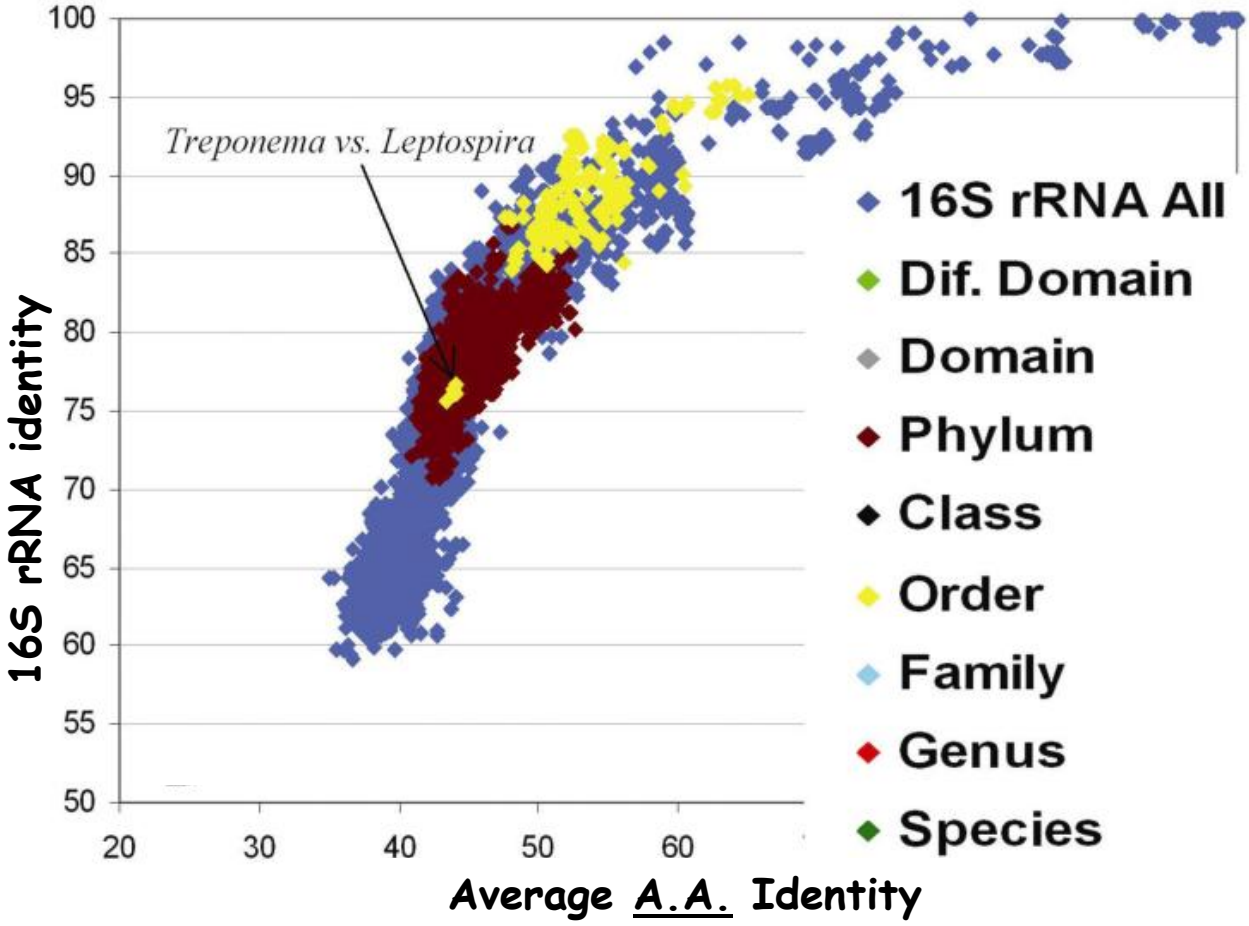
Domain
Phylum
Class
Order
Family
Genus
Species
Subspecies

•Current classification is primarily based on 16S rRNA and secondarily on classical microscopic and biochemical observations.

•Designation of higher than the species ranks is rather arbitrary (clustering by 16S rRNA but no standards on absolute differences).

Identifying outliers of taxonomy

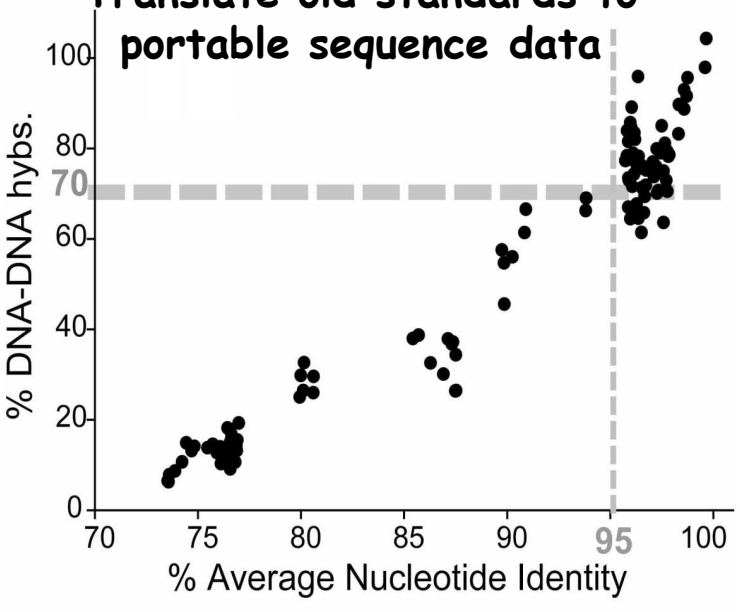
- 176X176 = 30,976 pairs
- Adjacent ranks show ~30% overlap
- Non-adjacent ranks overlap ~10 fold less frequently



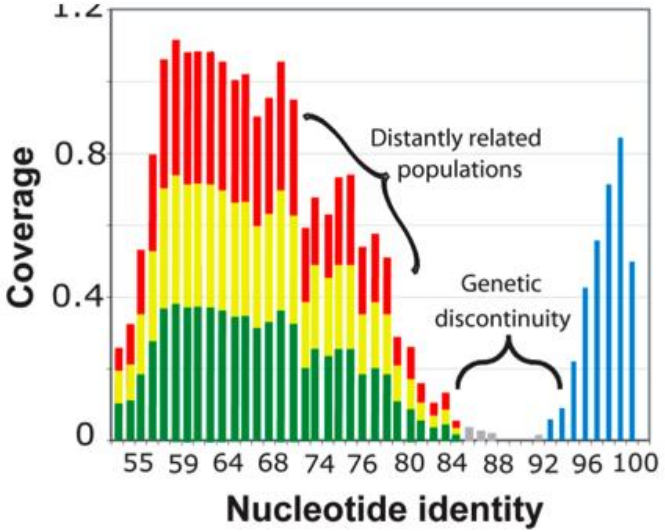
From Konstantinidis and Tiedje, Journal of Bacteriology, 2005

Summary-Conclusions

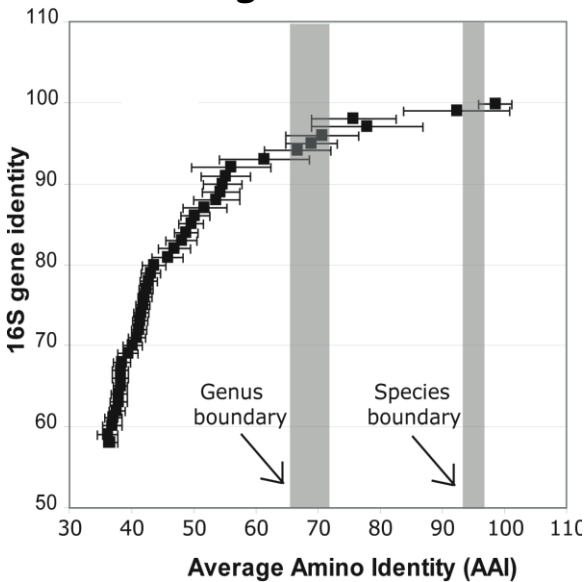
Translate old standards to portable sequence data



Sequence-distinct populations frequently @ 95% ANI level!



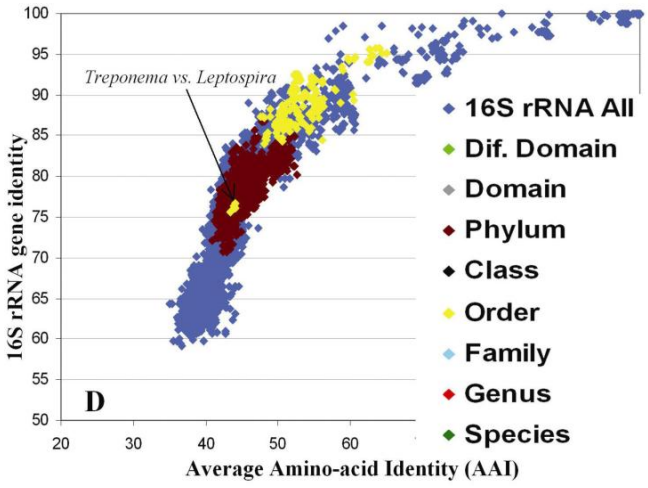
16S rRNA gene is reliable



95% ANI <=> 98.5% 16S

Curr. Opin. Microbiol. 2007

Genome-taxonomy for a more predictive classification system



Environmental Microbial Genomics Lab @ GaTech

kostas@ce.gatech.edu
www.enve-omics.gatech.edu

Natural & Engineered microbial systems
(Bioremediation)

Microbial & interdisciplinary research @ GaTech

Environmental Engineering Building Downtown Atlanta

Interested? Email
kostas@ce.gatech.edu

Georgia Tech

INTEGRATIVE BIOSYSTEMS INSTITUTE

Acknowledgments

People

Kostas' group @ Georgia Tech

- AlejanderLuo, Bioinformatics
- Alejandro Caro, Microbial Ecology/Bioinformatics
- Natasha DeLeong, Microbial Ecology
- Seung-Dae Oh, Env. Engineering
- DespoinaTsementzi, Env. Engineering
- Tate Nixon, Env. Engineering

DeLong's group @ M.I.T.

- Prof. Steven Hallam (now @ UBC)
- Dr. Virginia Rich (now @ U of Arizona)
- Dr. Gene Tyson (now @ U of Queensland)

Support

U.S. DEPARTMENT OF ENERGY

Genomes to Life Program

Award #DE-FG02-07ER64389

NSF

National Science Foundation

IOS-0919251 & CBET-0967130

Other collaborators

- Prof. **Frank Loeffler** (GaTech), on bioremediation communities
- Prof. **Spyros Pavlostathis** (Gatech), antibiotic resistance
- Prof **Hang Lu** (GaTech), on single-cell genomics
- Prof. **James Tiedje** (MSU), on the species concept
- Dr. **Alban Ramette** (Max Planck, Bremen), on biogeography

Phylogenetic relationships based on AAI

- **(Phylip) Neighbor joining tree of the full 176X176 matrix of AAI.**
  - Same results by weighted NG (Weighbor).

	Acinetob.	A.pernix	A.tumefaciens	…175 genomes
Acinetobacter sp.	1.00			
Aeropyrum pernix	0.95	1.00		
Agrobacterium tumefaciens	0.66	0.88	1.00	
…175 genomes total	…	…	…	…

- **All phyla and several classes are as deep branching as Archaea!**