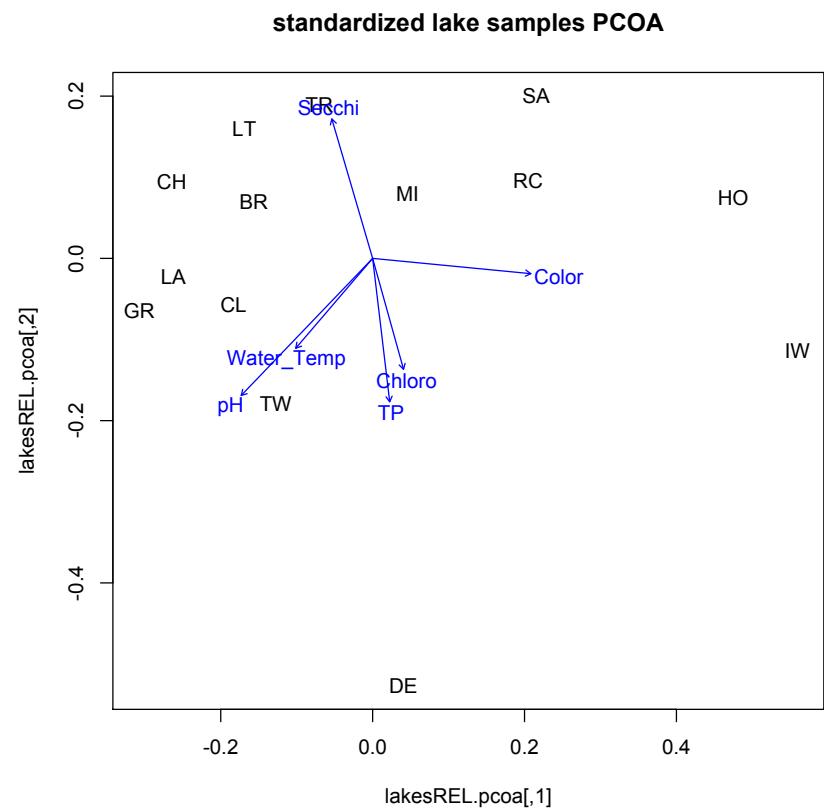
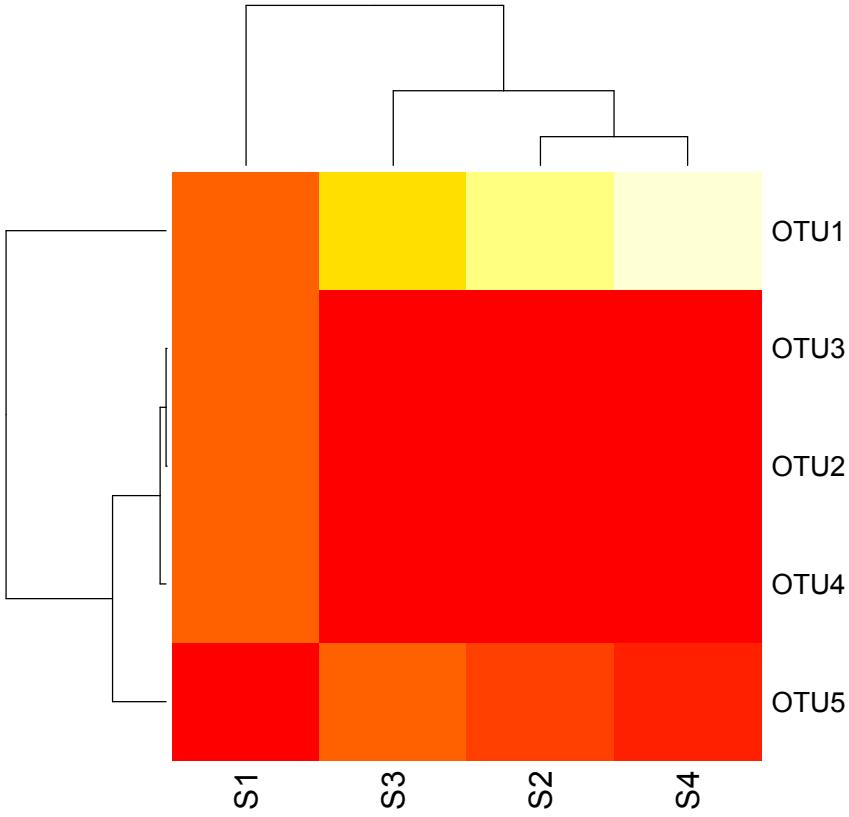
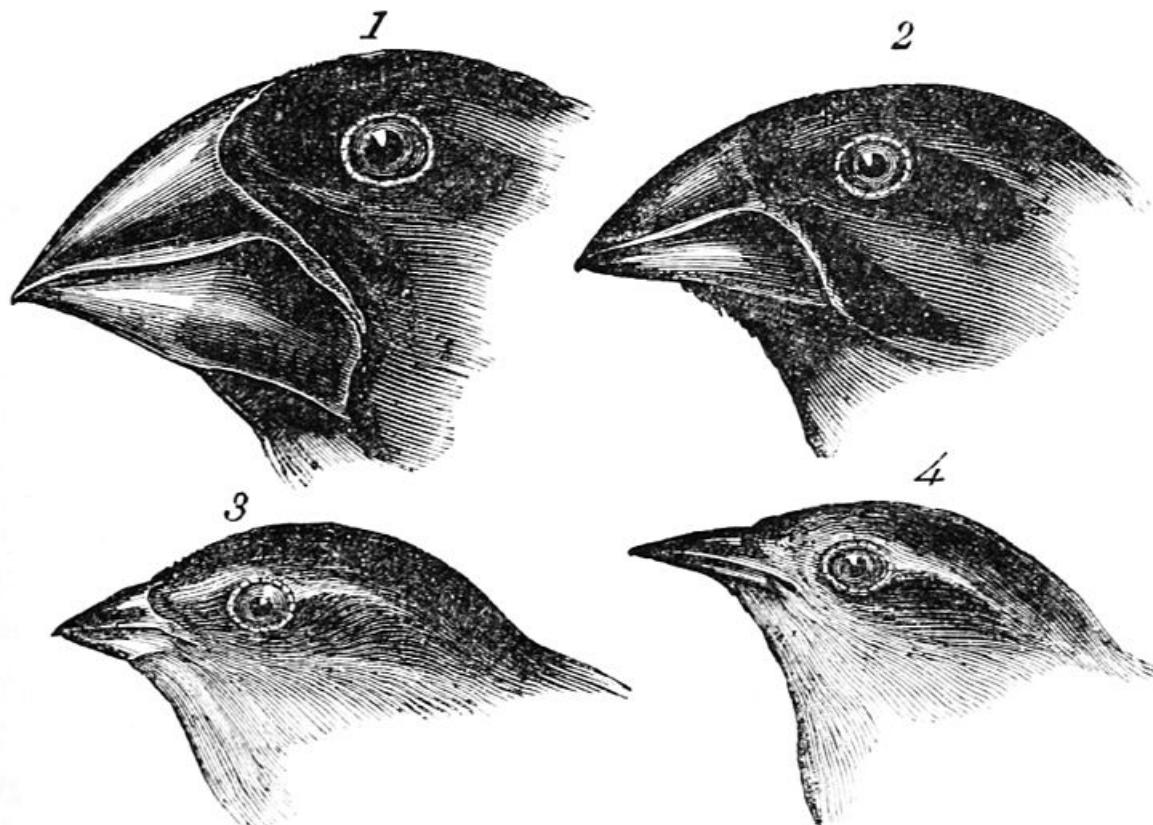


# Meta-genomic Signatures of Bacterial Life History Strategies

Stuart Jones  
University of Notre Dame



# Ecological traits



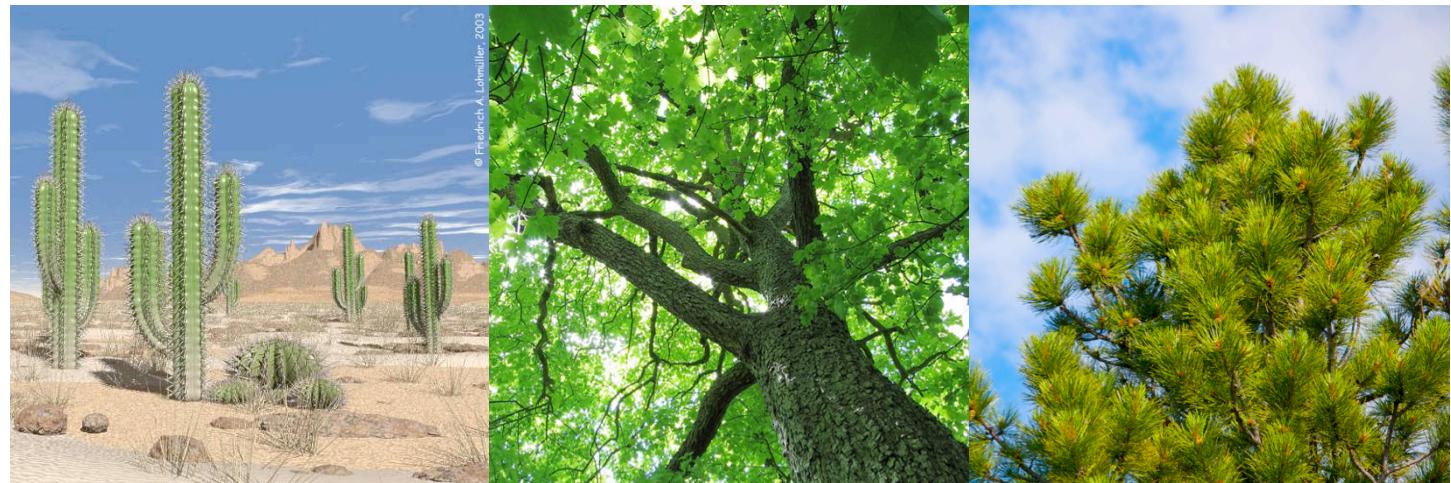
1. *Geospiza magnirostris*.  
3. *Geospiza parvula*.

2. *Geospiza fortis*.  
4. *Certhidea olivacea*.

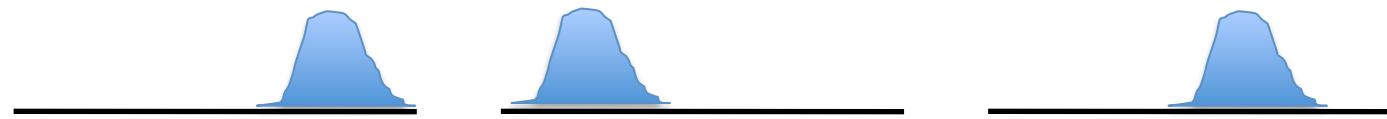
# **Traits, tradeoffs, and life history strategies**

- Related to fitness or at least growth rate
- A suite of traits = life history strategy, includes trade offs
- Enables understanding and prediction of distribution, species coexistence, and contributions to ecosystem services

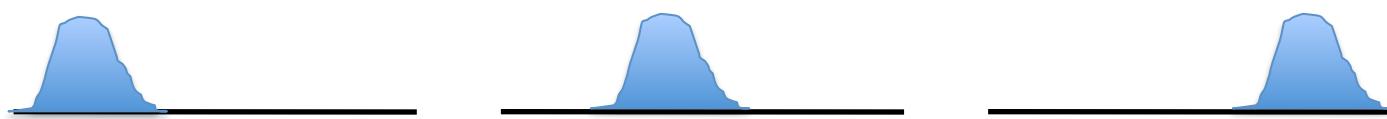
# A hopefully intuitive example



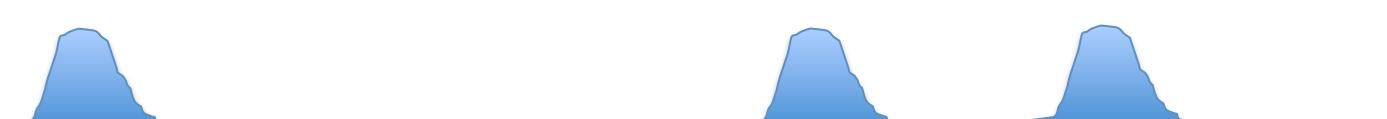
Water use  
efficiency



Freeze  
tolerance

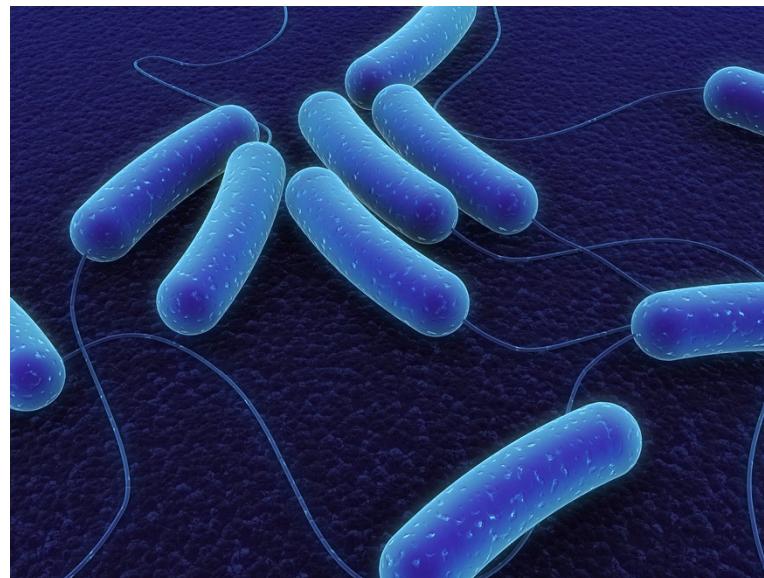


Specific  
Leaf area

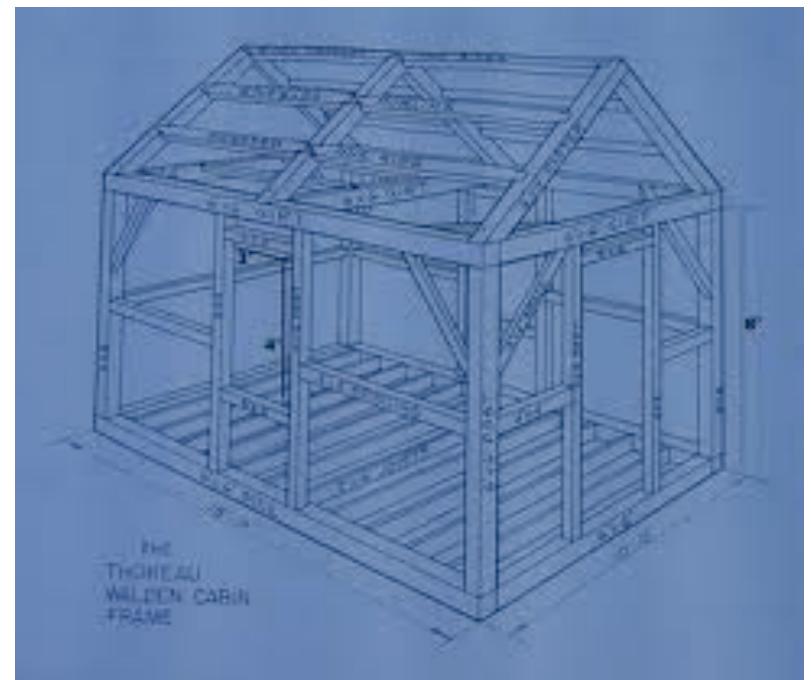
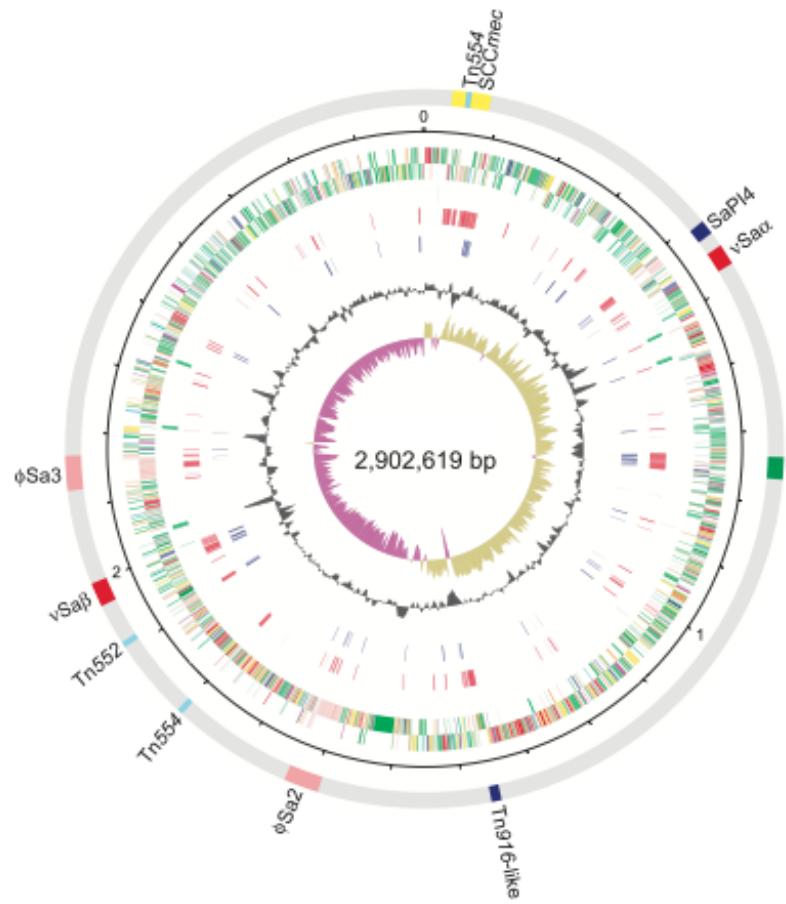


# Challenges presented by microbes

- Small
- Genetically and functionally diverse
- Morphologically homogeneous
- **Paucity of group-specific natural history**



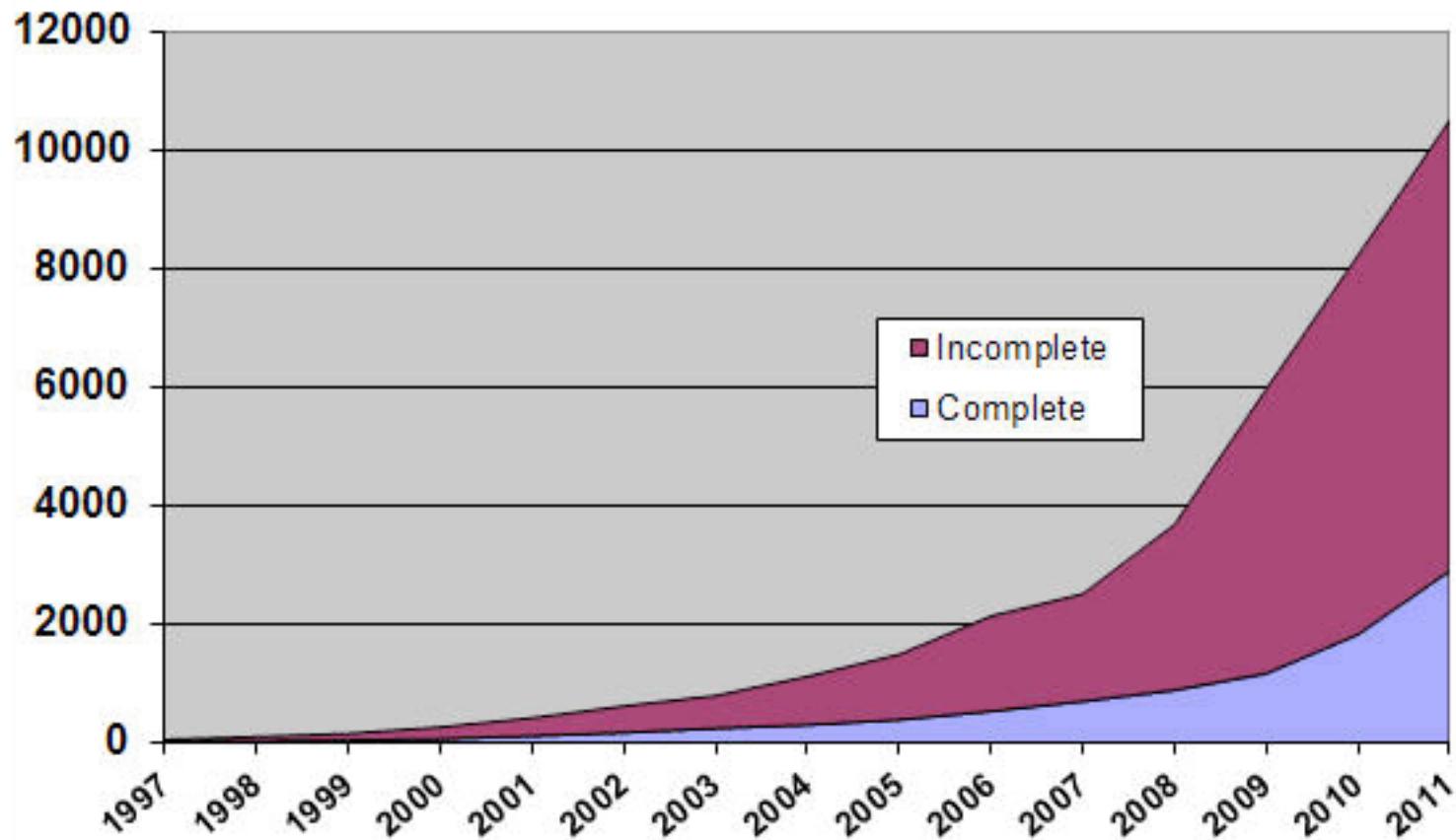
# Advances in genomics should help



# Data is becoming less limiting

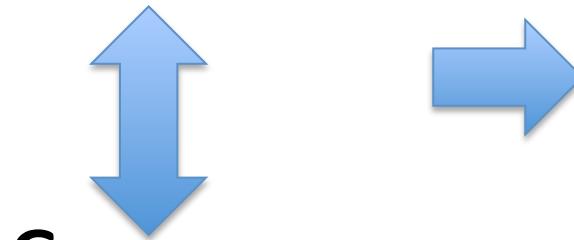
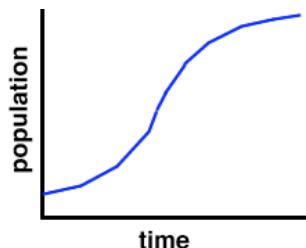


>24,000 as of  
yesterday

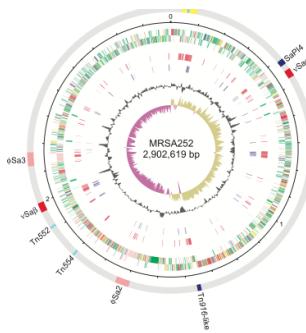


# Genomic features as markers of microbial traits and life histories

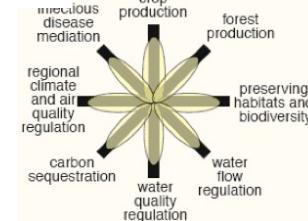
## Ecophysiological Traits



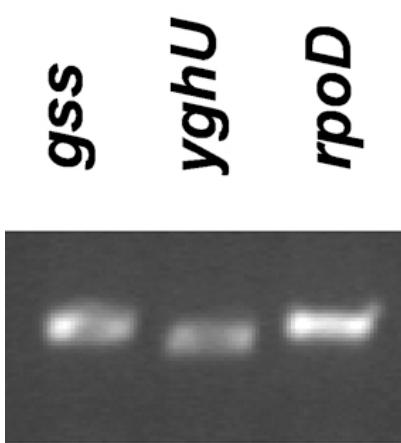
Genome



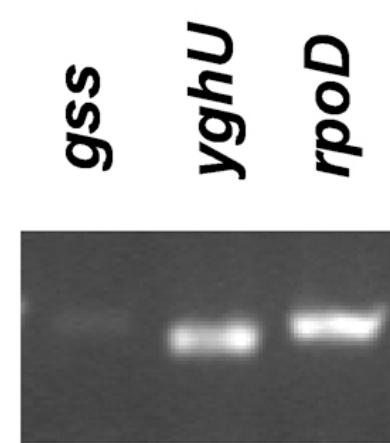
## Pattern & Process



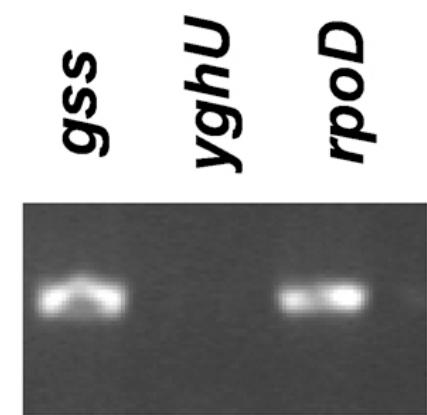
# Traditional one organism, one gene approach



Wild-type

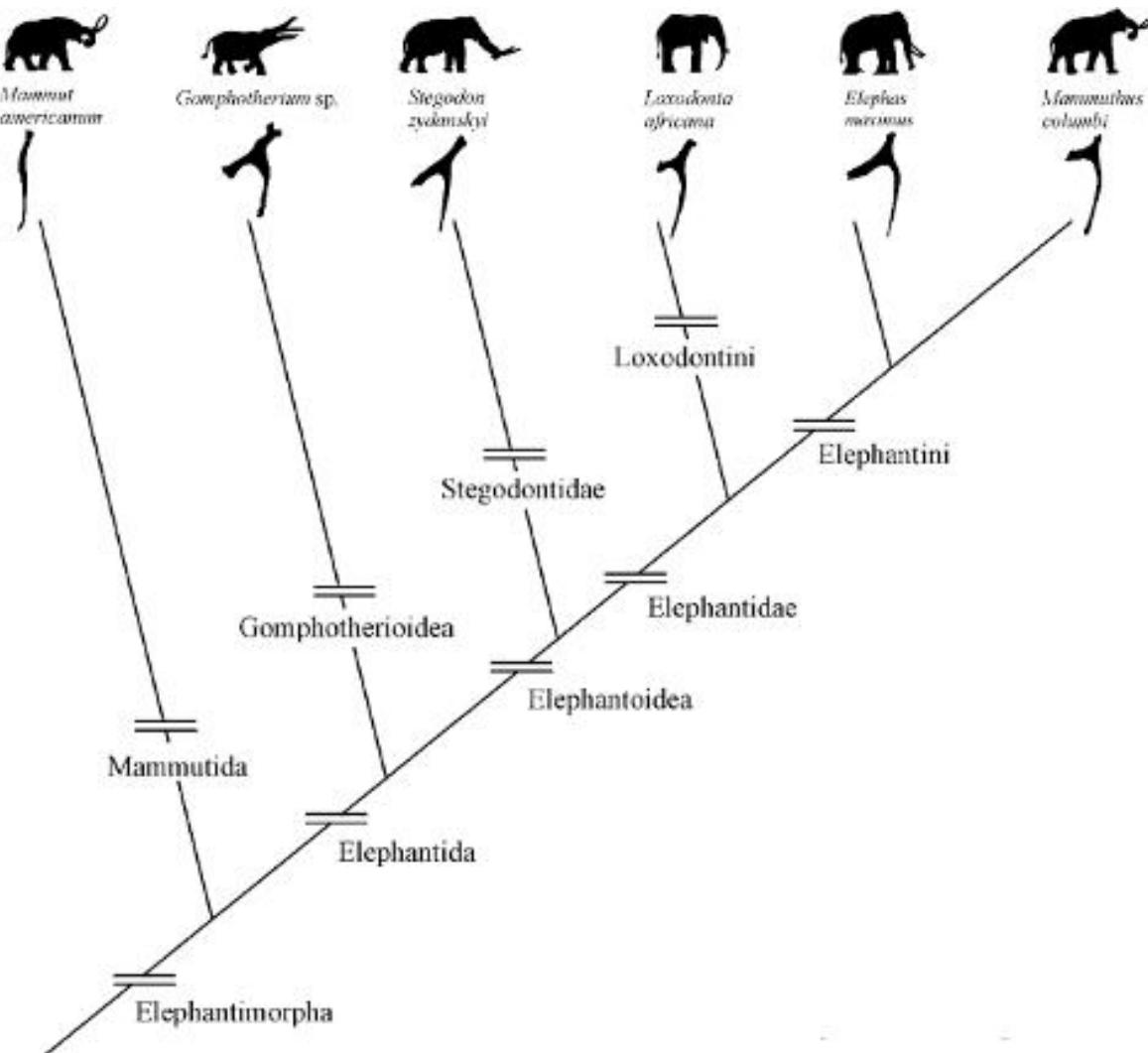


$\Delta gss$



$\Delta yghU$

# A comparative biology approach



# Three simple questions

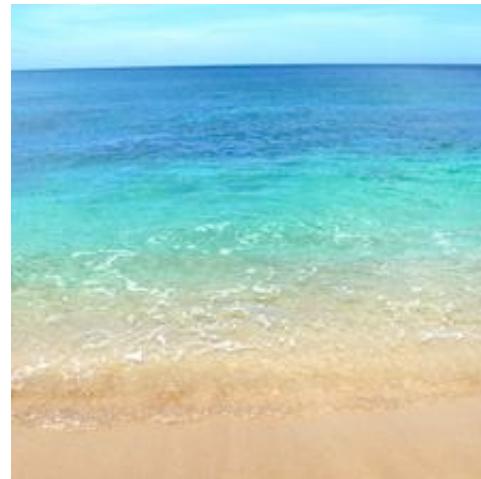
1. Can we find indications of habitat adaptation?
2. Can we develop general models for microbial life history strategies based on genomic markers?
3. What drives genome streamlining in free-living bacteria?

# Dataset

- 580 publicly available genomes
- Three environments



Human Gut

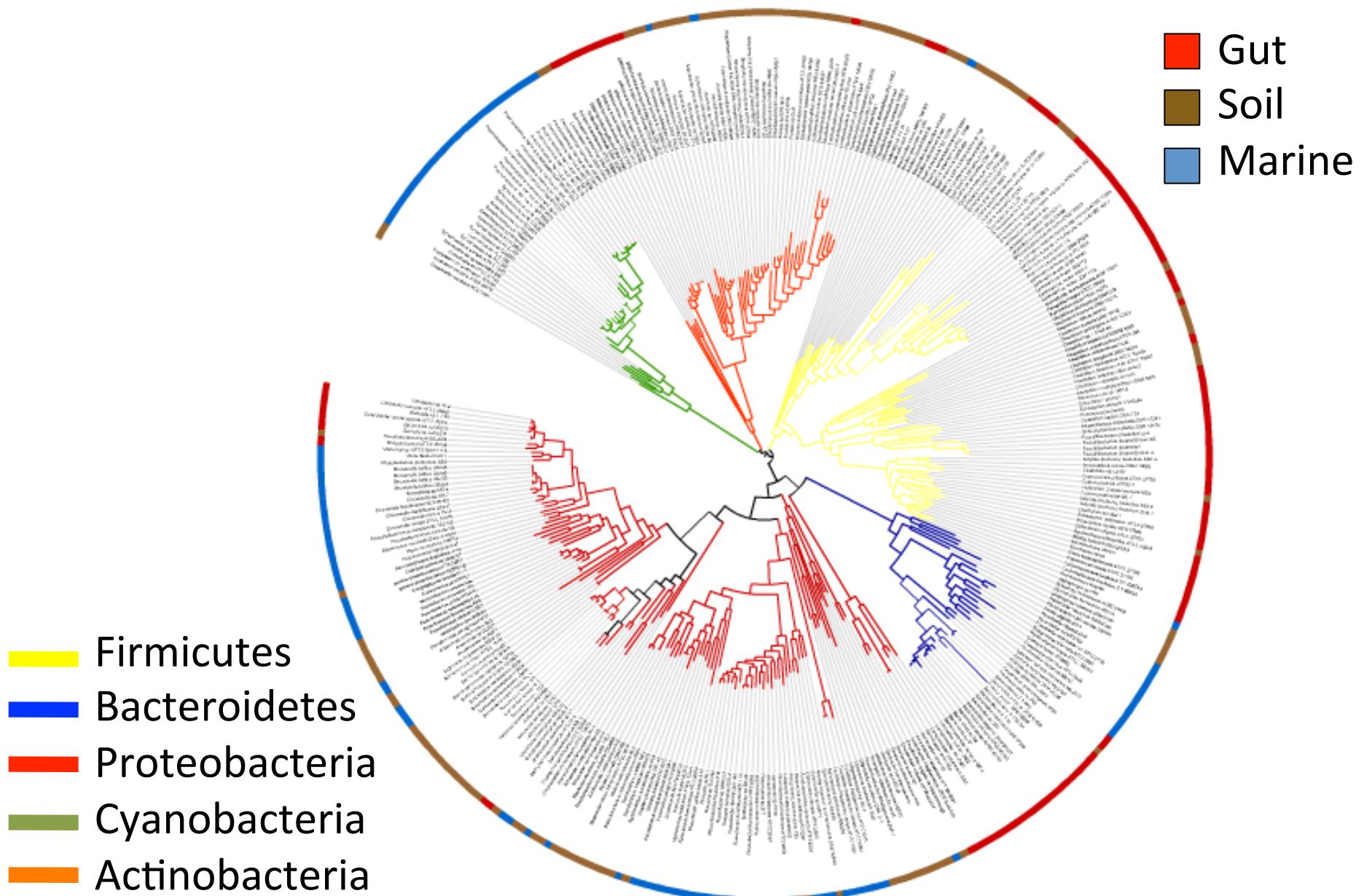


Marine

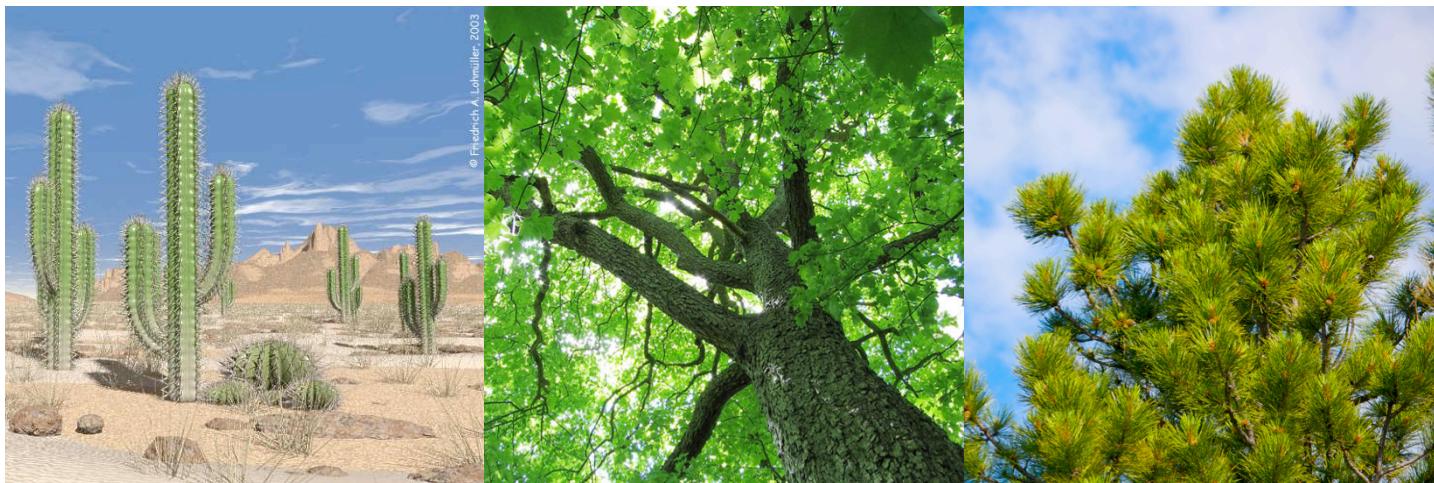


Soil

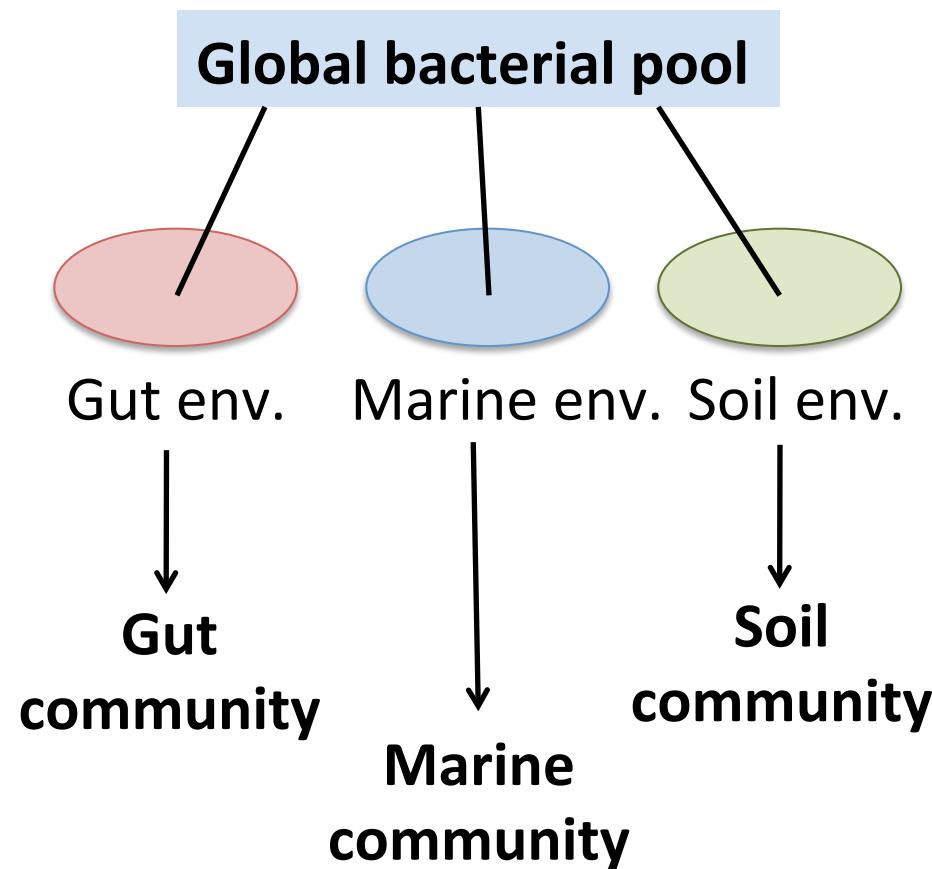
# Spans broad phylogenetic scales



# 1. Can we find indications of habitat adaptation?



# Environmental conditions across broad gradients drive adaptation



# General Approach

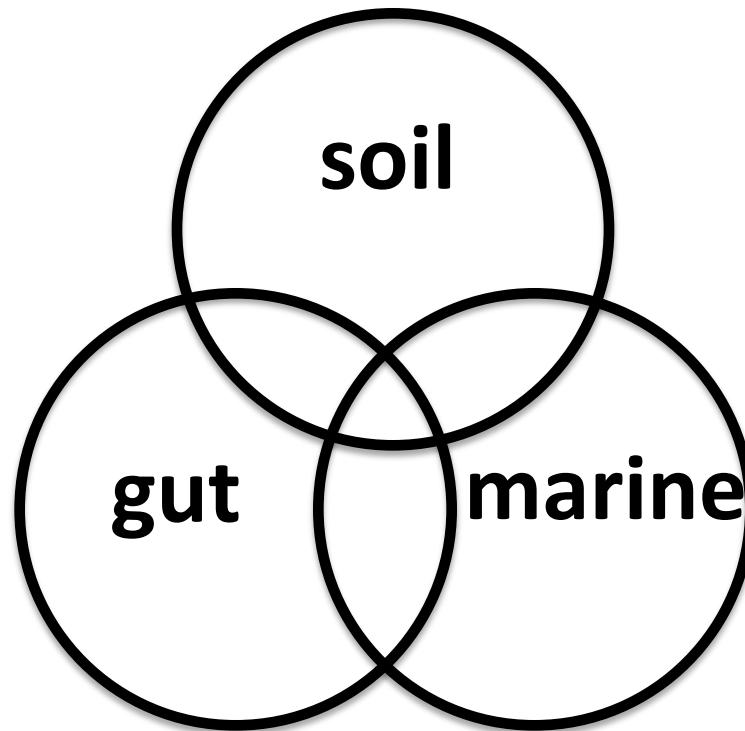


- 580 genomes
- 2 million open reading frames (ORFs)
- 135,000 *de novo* clusters of orthologous groups (COGs)

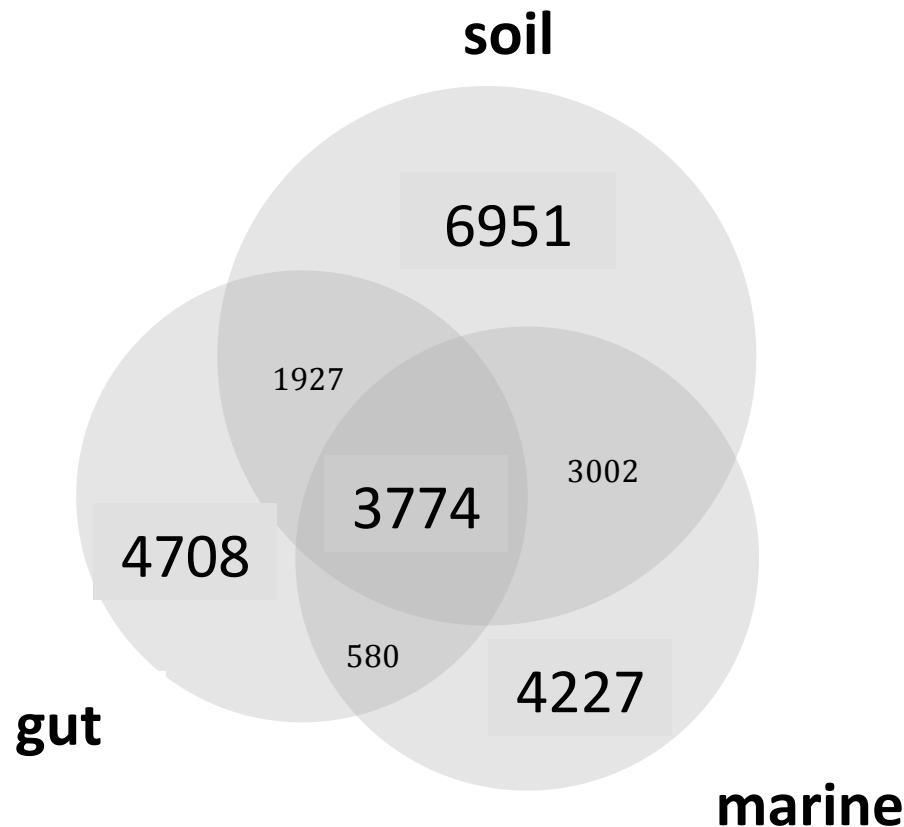
# Naïve Predictions for Selective Features

- Lack of oxygen in the gut
- Signal of symbiosis in gut
- Osmotic stress in soil and, perhaps, marine
- Temporal heterogeneity in soil and marine

# Environmental conditions across broad gradients drive genome content



# COGs distribution

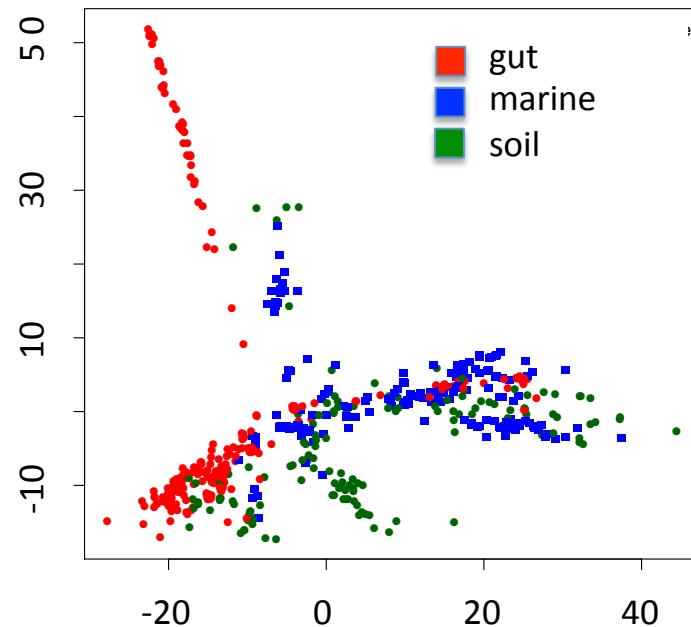


COGs present in more than 5% of genomes  
from a particular environment

# Feature selection

ANOSIM:

$p < 0.001$ ,  $R = 0.32$

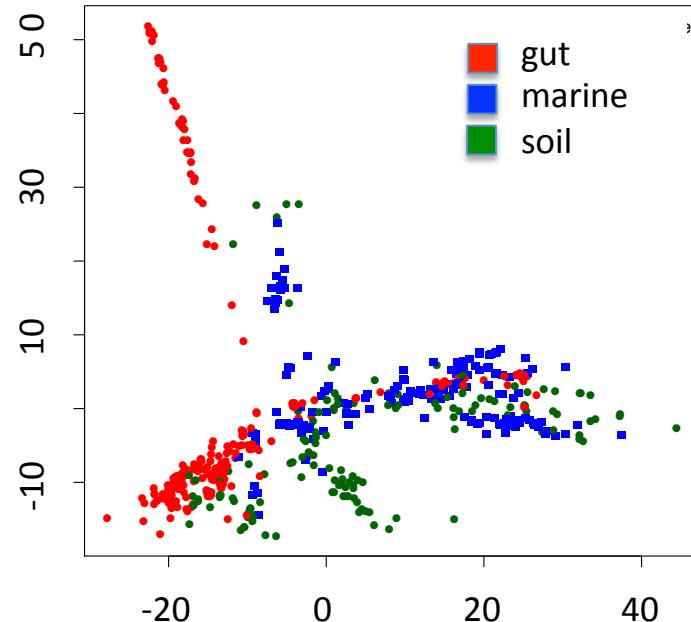


COGs present in more than 5% of genomes

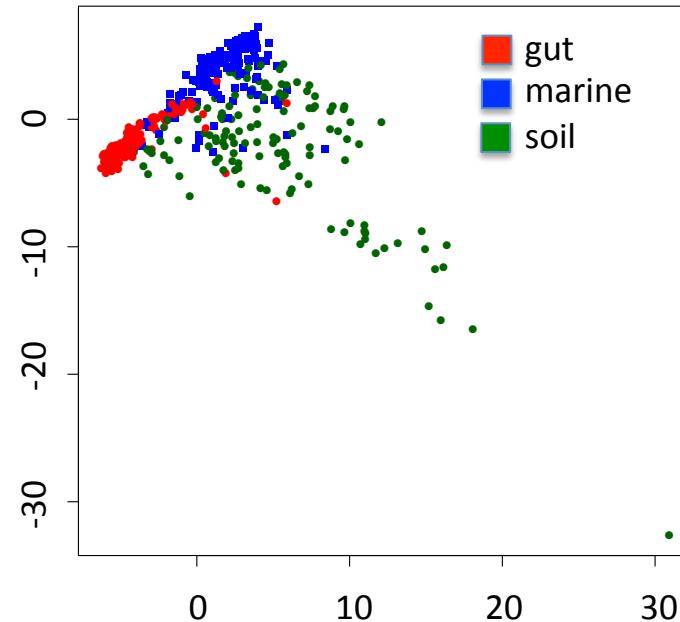
# Feature selection worked

ANOSIM:

$p < 0.001, R = 0.32$



$p < 0.001, R = 0.68$

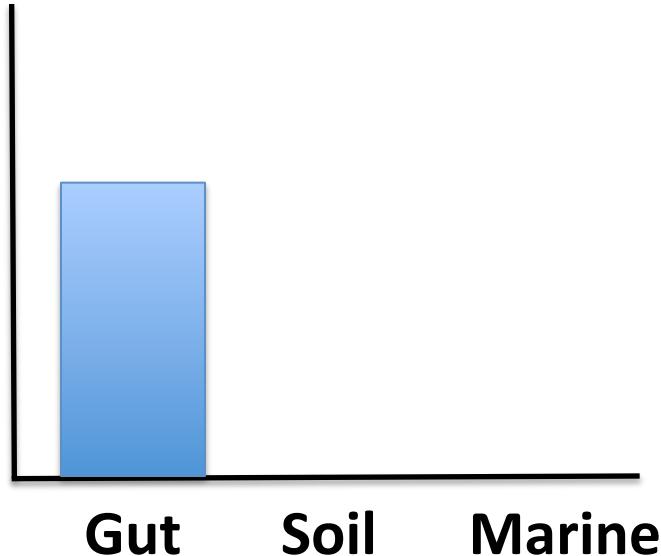


COGs present in more than  
5% of genomes

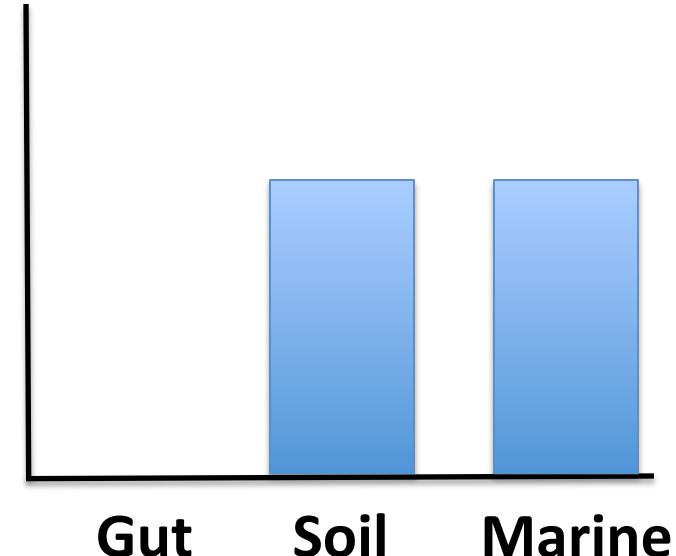
top 100 COGs

# COG presence or absence can be discriminating

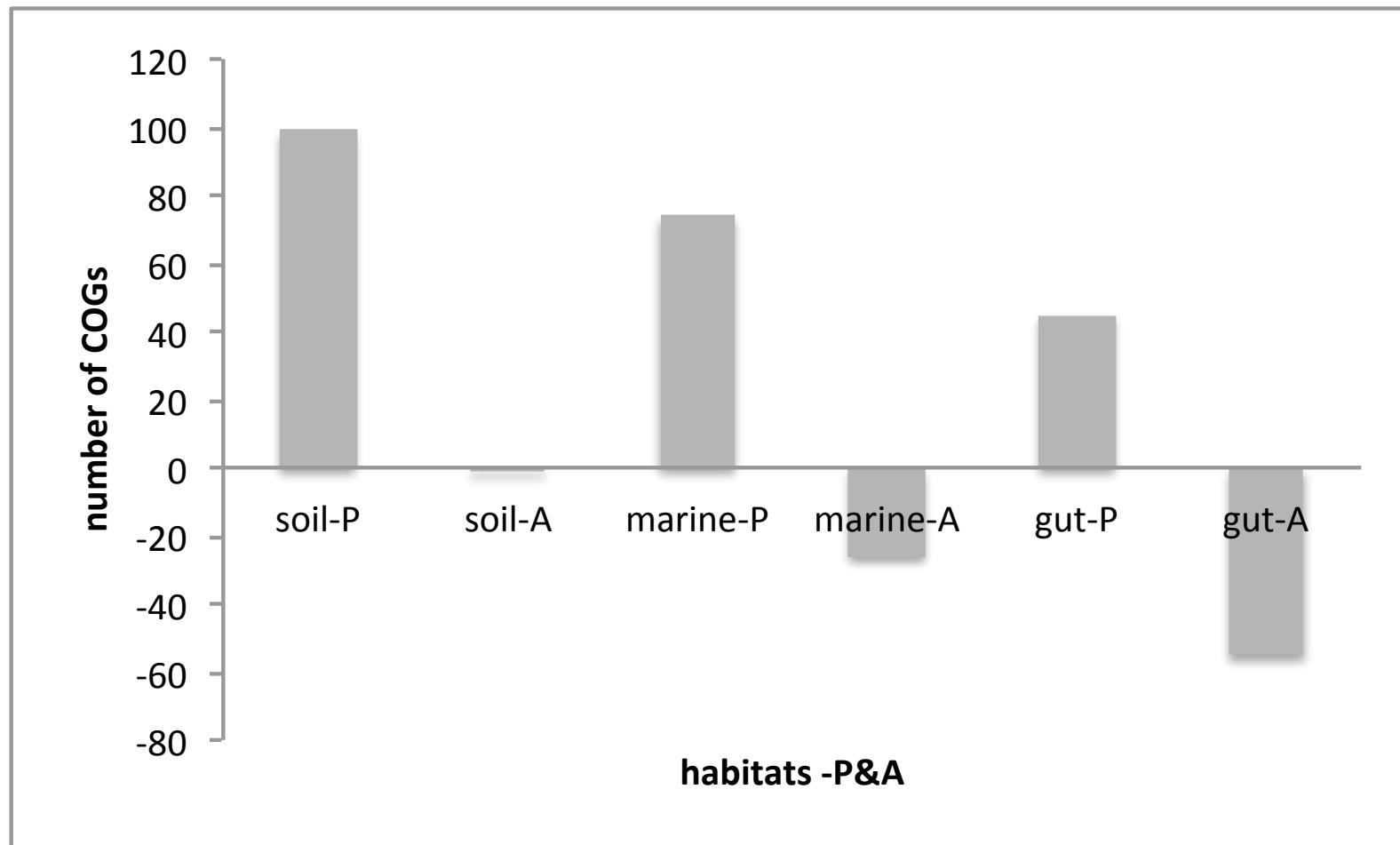
**PRESENT**



**ABSENT**

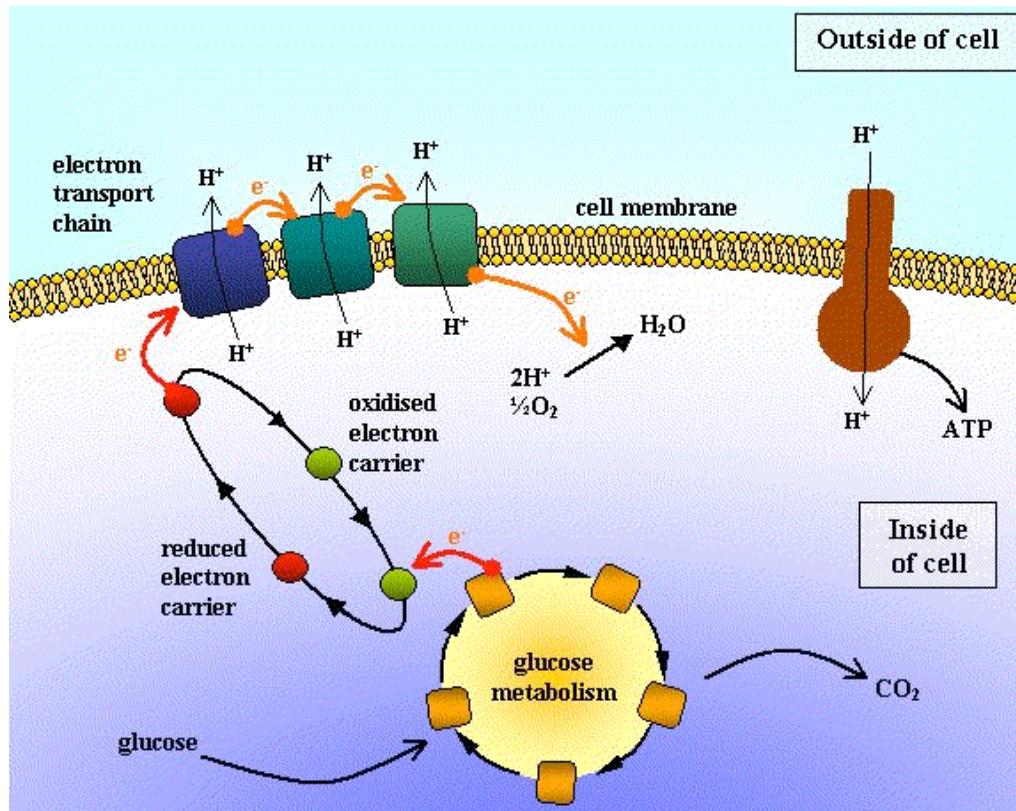


# Many discriminating COGs for gut were absent

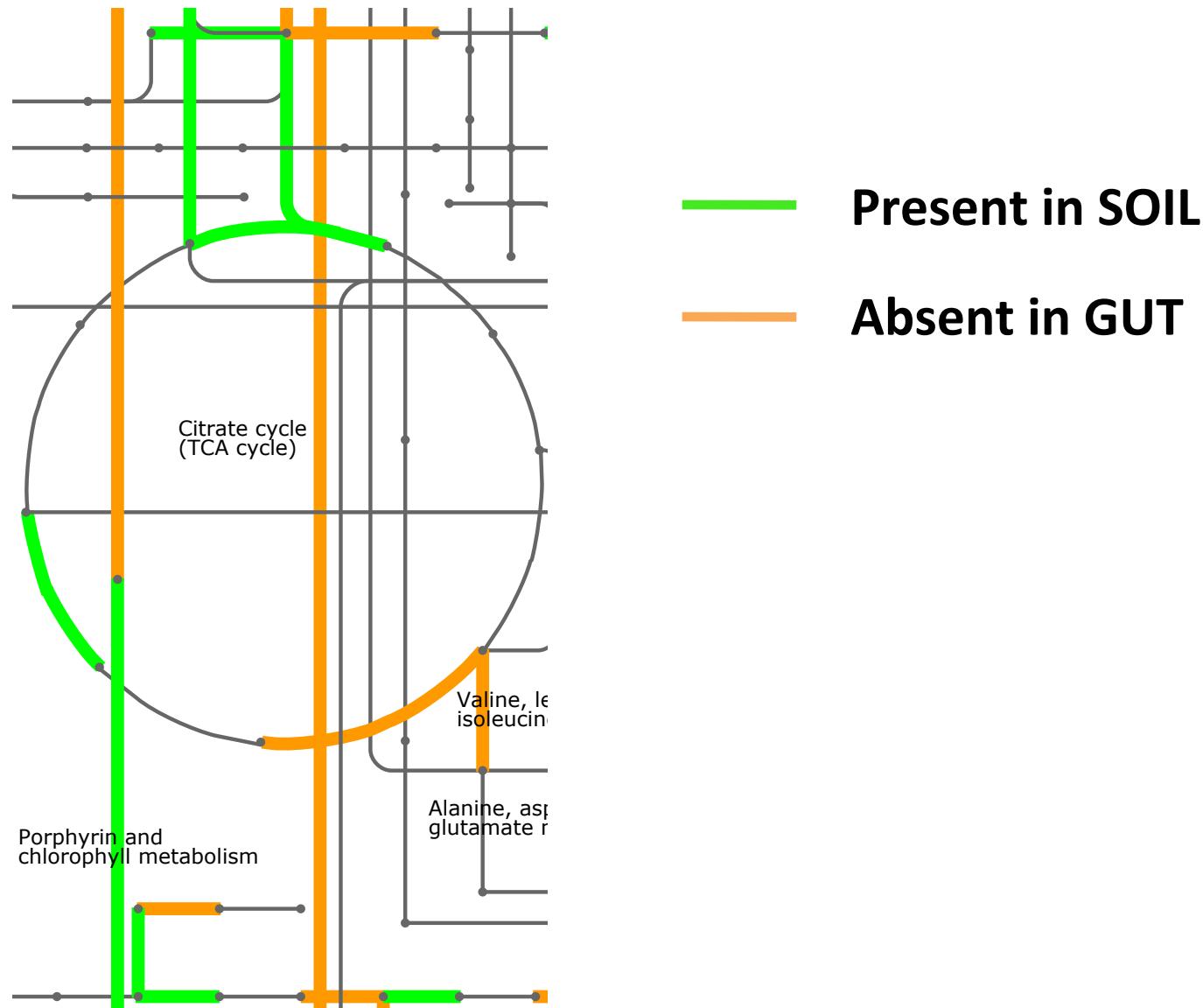


# Oxygen Availability

- TCA cycle, electron transport chain, glycolysis



# Discriminating COGs in TCA cycle



# Stress response

- DNA repair, ion transportation and oxidative stress response



oxidative stress

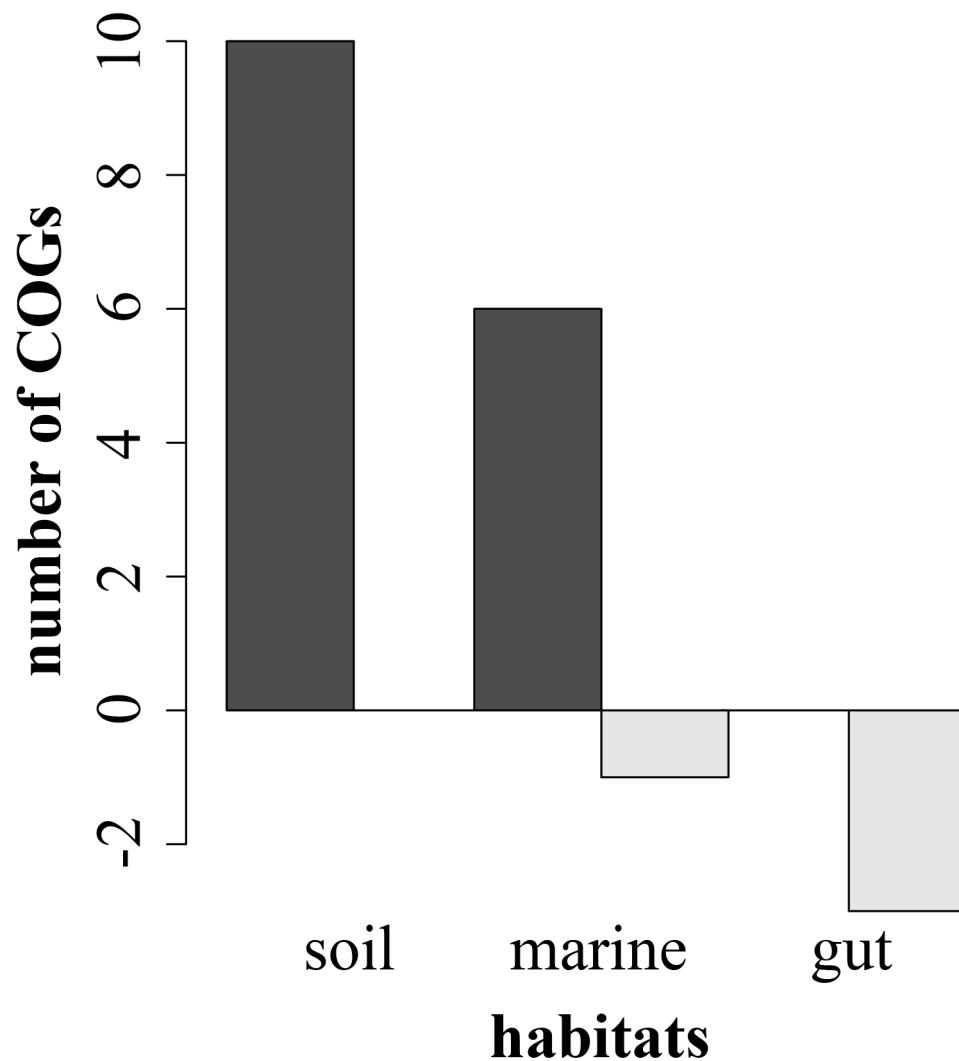


osmotic stress



UV light damage

# Greater stress response and DNA repair in marine and soil genomes

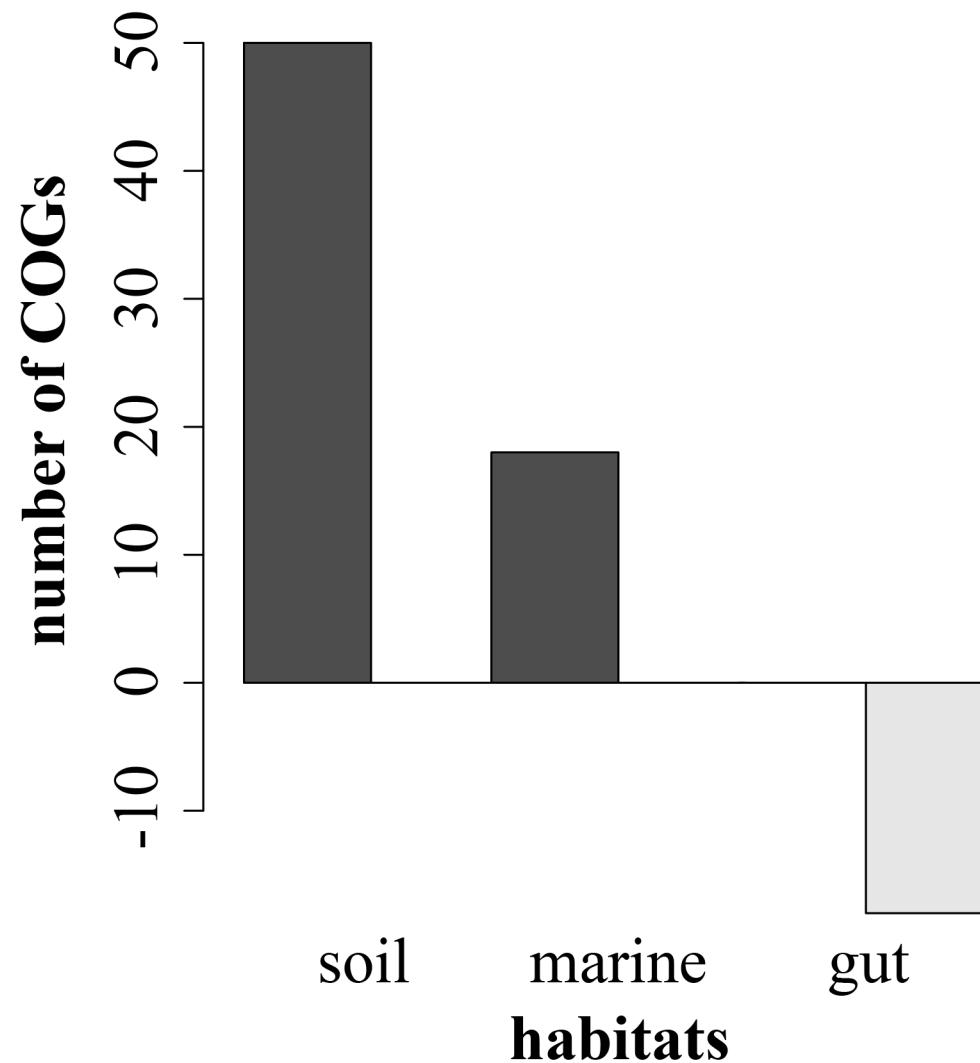


# Resource availability and energy storage

- lipid metabolism, amino acid biosynthesis



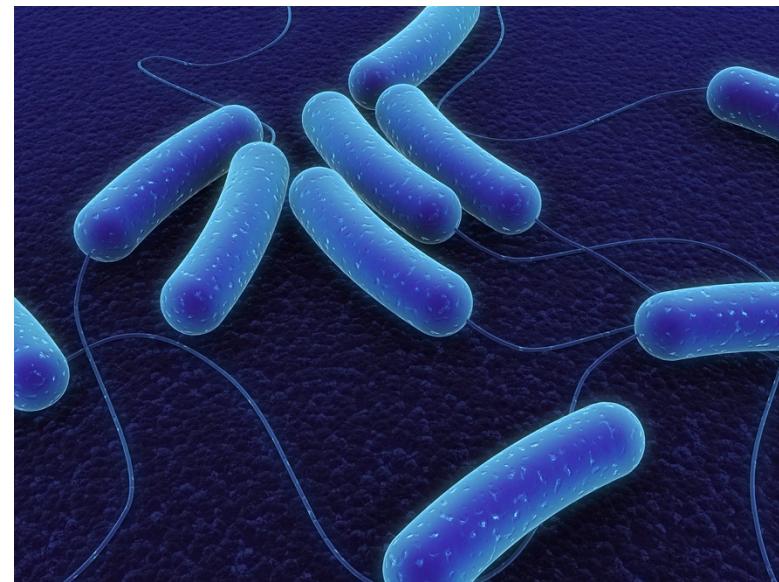
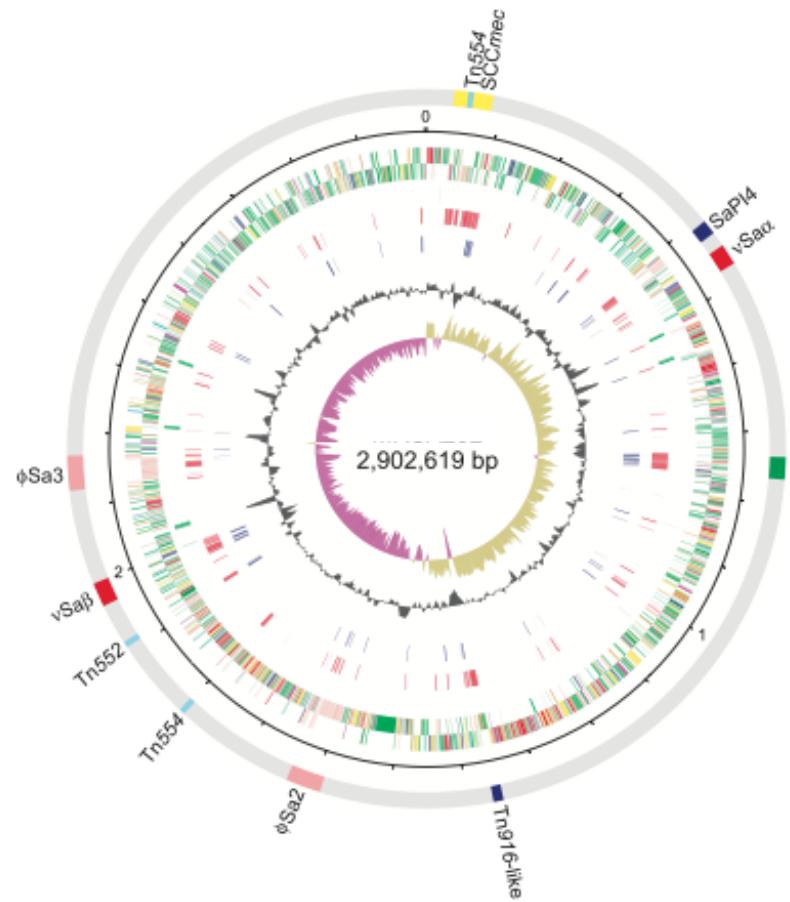
# **Soil and marine genomes have characteristic COGs associated with lipid metabolism**



# 1. Can we find indications of habitat adaptation?

- Observations are fairly consistent with naïve predictions
- Across broad gradients, we understand some of the important environmental features

## 2. Can we develop general models for microbial life history strategies based on genomic markers?



# Oligotrophs vs. Copiotrophs

Very old concept, but little quantitative support

## Oligotrophs

- Small
- Slow growing
- Gleaners
- Stress resistant
- Efficiency

## Copiotrophs

- Large
- High maximum growth rate
- Opportunists
- Require high resources
- Power

# Cross validated models

## The genomic basis of trophic strategy

### 1) in marine bacteria

Federico M. Lauro<sup>a,b</sup>, Diane McDougald<sup>c,1</sup>, Torsten Thomas<sup>c,1</sup>, Timothy J. Williams<sup>b</sup>, SuheLEN Egan<sup>b</sup>, Scott Rice<sup>c</sup>, Matthew Z. DeMaere<sup>b</sup>, Lily Ting<sup>b</sup>, Haluk Ertan<sup>b,d</sup>, Justin Johnson<sup>e</sup>, Steven Ferriera<sup>e</sup>, Alla Lapidus<sup>f</sup>, Iain Anderson<sup>f</sup>, Nikos Kyrpides<sup>f</sup>, A. Christine Munk<sup>f</sup>, Chris Detter<sup>g</sup>, Cliff S. Han<sup>g</sup>, Mark V. Brown<sup>a,b</sup>, Frank T. Robb<sup>h</sup>, Staffan Kjelleberg<sup>a,c</sup>, and Ricardo Cavicchioli<sup>a,b,2</sup>

OPEN  ACCESS Freely available online

PLOS GENETICS

### 2) The Systemic Imprint of Growth and Its Uses in Ecological (Meta)Genomics

Sara Vieira-Silva<sup>1,2\*</sup>, Eduardo P. C. Rocha<sup>1,2</sup>

# Cross validated models

## The genomic basis of trophic strategy in marine bacteria

1)

Federico M. Lauro<sup>a,b</sup>, Diane McDougald<sup>c,1</sup>, Torsten Thomas<sup>c,1</sup>, Timothy J. Williams<sup>b</sup>, Suhelen Egan<sup>b</sup>, Scott Rice<sup>c</sup>, Matthew Z. DeMaere<sup>b</sup>, Lily Ting<sup>b</sup>, Haluk Ertan<sup>b,d</sup>, Justin Johnson<sup>e</sup>, Steven Ferriera<sup>e</sup>, Alla Lapidus<sup>f</sup>, Iain Anderson<sup>f</sup>, Nikos Kyrpides<sup>f</sup>, A. Christine Munk<sup>f</sup>, Chris Detter<sup>g</sup>, Cliff S. Han<sup>g</sup>, Mark V. Brown<sup>a,b</sup>, Frank T. Robb<sup>h</sup>, Staffan Kjelleberg<sup>a,c</sup>, and Ricardo Cavicchioli<sup>a,b,2</sup>

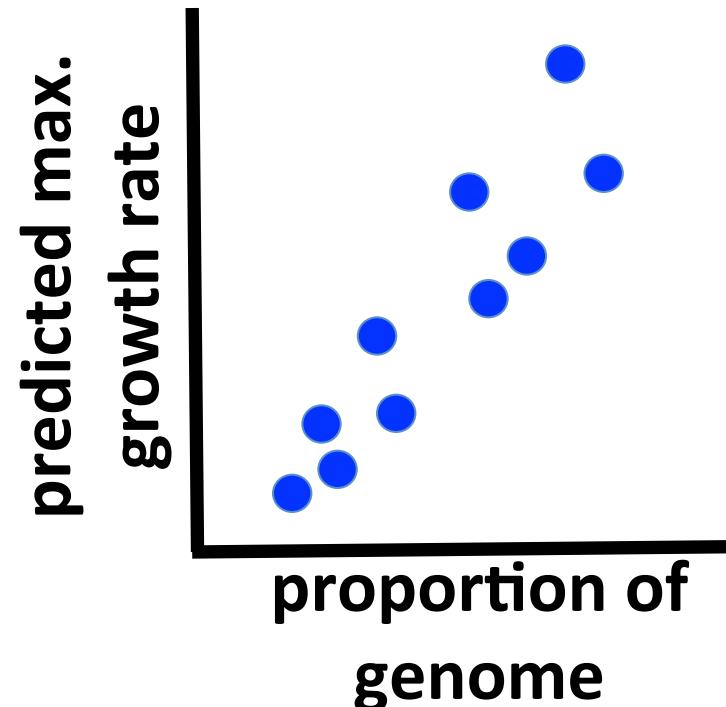
OPEN  ACCESS Freely available online

PLOS GENETICS

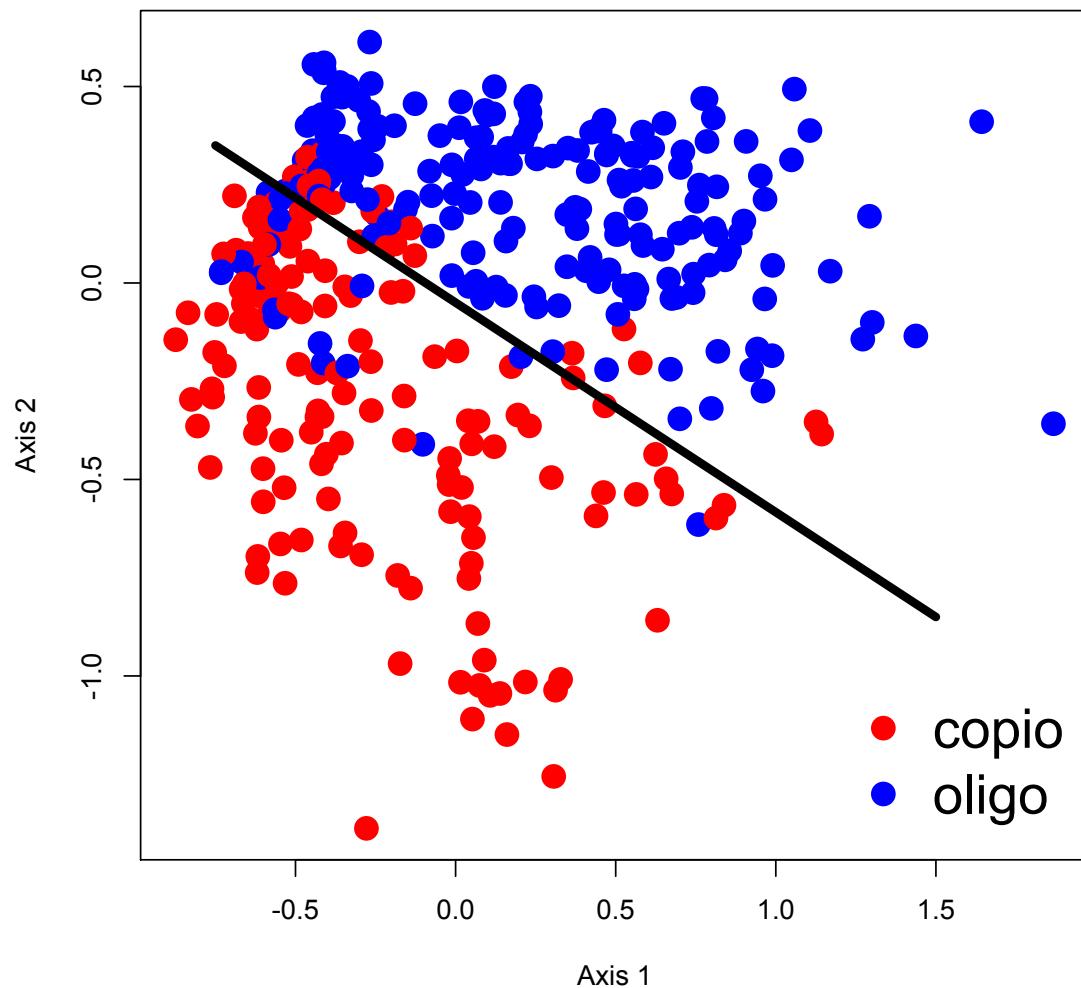
2)

## The Systemic Imprint of Growth and Its Uses in Ecological (Meta)Genomics

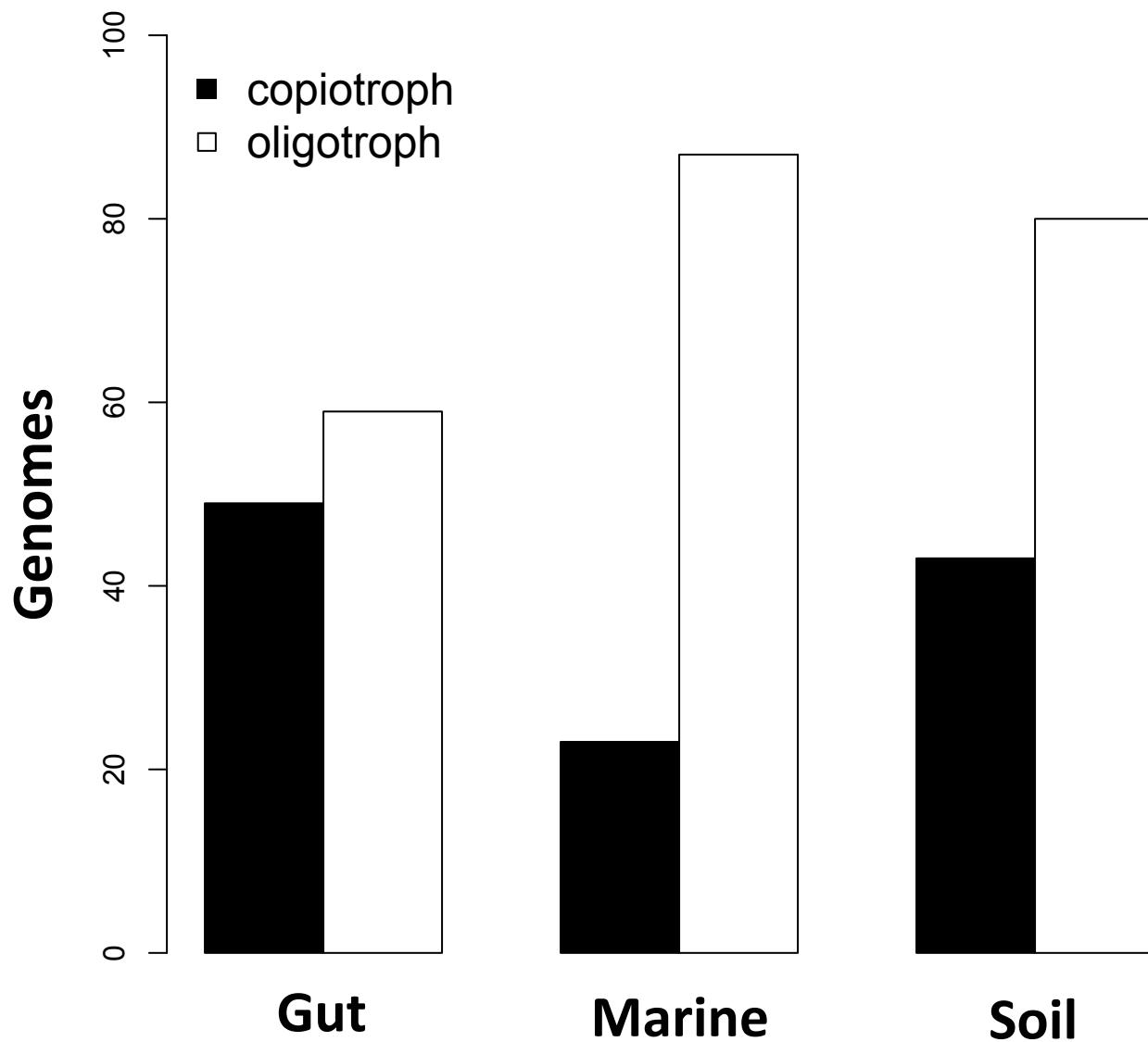
Sara Vieira-Silva<sup>1,2,\*</sup>, Eduardo P. C. Rocha<sup>1,2</sup>



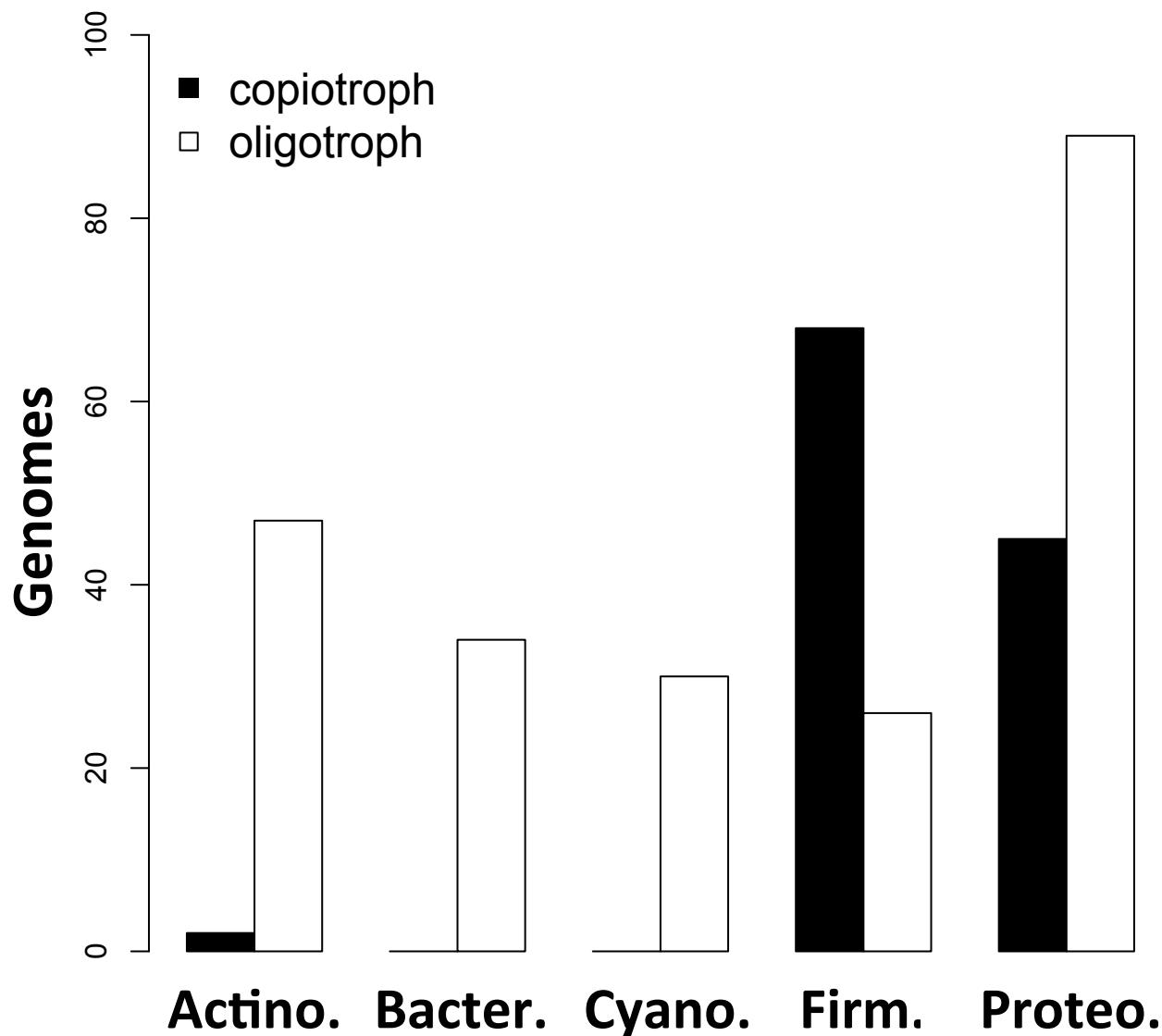
# A cross-validated model works well



# Environment or taxonomy as a driver?



# Environment or taxonomy as a driver?



# Other ecophysiological trade offs?

REVIEW

## Microbial Biogeography: From Taxonomy to Traits

Jessica L. Green,<sup>1\*</sup> Brendan J. M. Bohannan,<sup>1</sup> Rachel J. Whitaker<sup>2</sup>

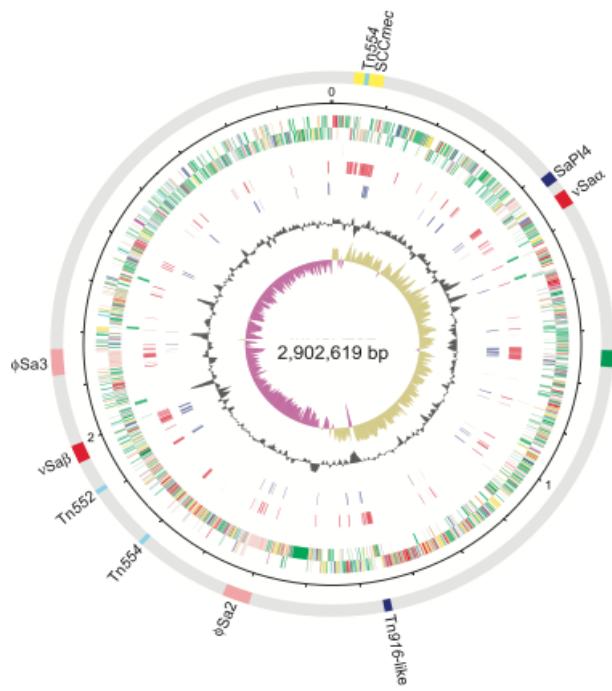
**Table 1.** Selected examples of microbial functional traits.

Trait	Potential consequence for fitness or performance	Measure	Example
Gene sequence identity	Protein function	PCR amplification	(19)
Ribosomal copy number	Growth rate, response time	Southern hybridization	(22)
Metabolite content	Survival	Mass spectrometry	(25)
Gene content	Metabolic potential, life history variation	Genomics	(26)
Genome size	Metabolic potential, life history variation	Genomics	(29)
Environmental resistance (antibiotic, viral, radiation, etc.)	Survival	PCR amplification, laboratory screening of isolates	(30)
Transcript abundance	Metabolic switching	Reverse transcription PCR, qPCR	(31)
Mutation rate	Adaptability	Laboratory assays of isolates	(32)
Cell size	Metabolic rate	Flow cytometry	(33)

# Other ecophysiological trade offs?

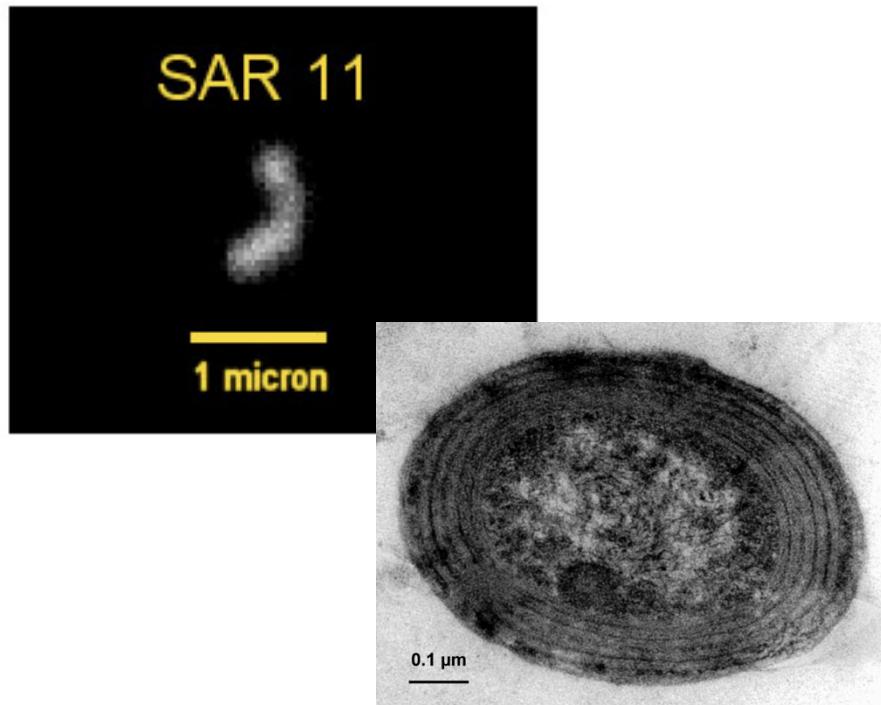
Trait	Genome feature	Sign (copio:oligo)	PGLS p
Ecophysiological flexibility	Gene number	+	0.0001
Growth capacity	16S rRNA gene number	-	0.0001
Growth regulation	Transcription factors	+	0.0001
Predation rate	Mean CRISPR repeats	+	0.0001
Environmental responses	Histidine kinases	+	0.0001
Resource acquisition	ABC superfamily	+	0.0001
Resource acquisition	Motility genes	+	0.0001
Growth efficiency	Amino acid molecular weight	+	0.0006

## 2. Can we develop general models for microbial life history strategies based on genomic markers?

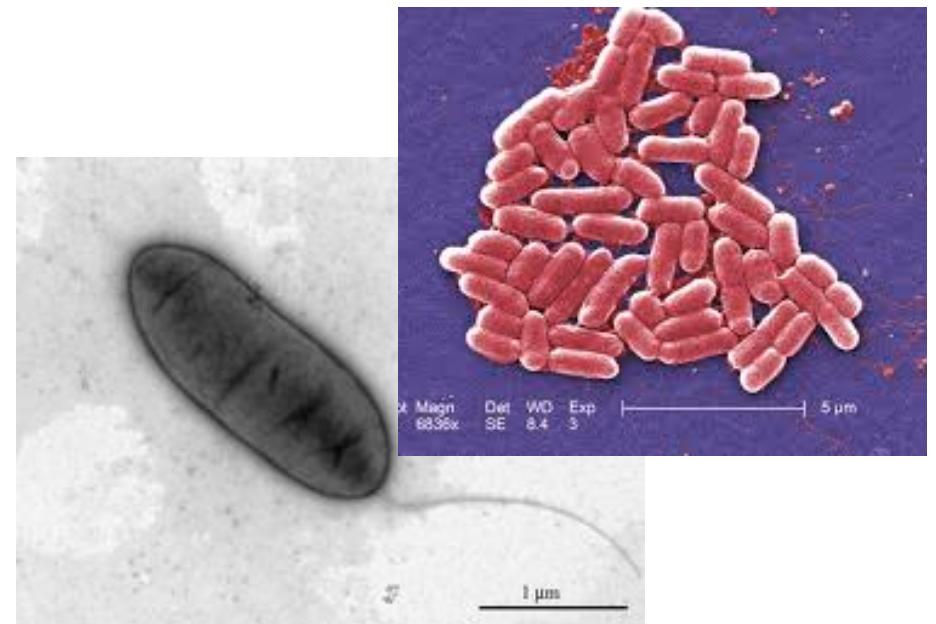


Yes, some support for long-held microbial lifestyles

### 3. What drives genome streamlining?

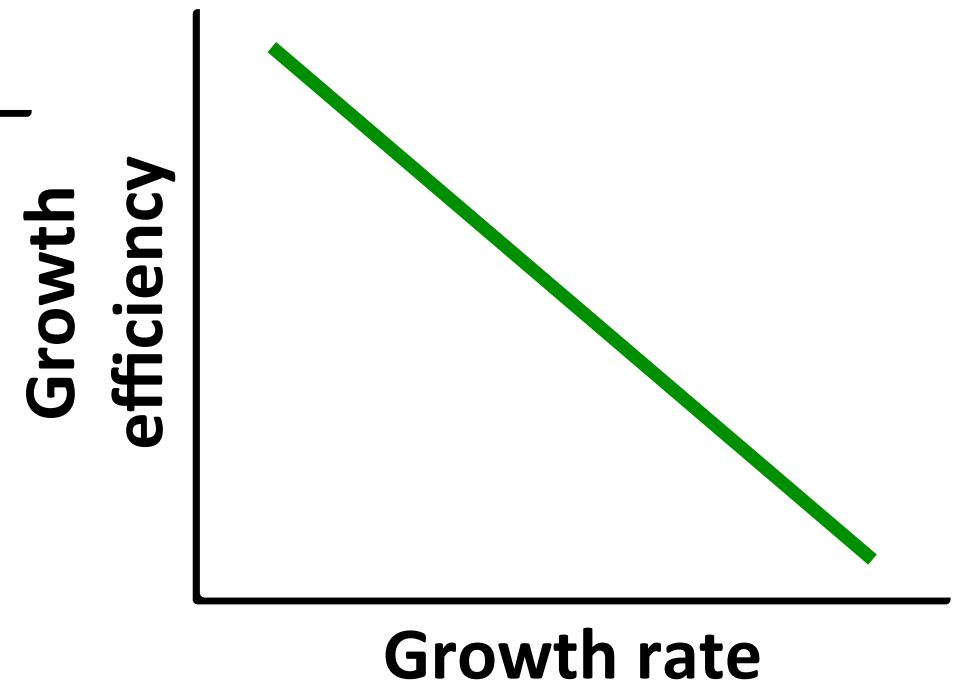
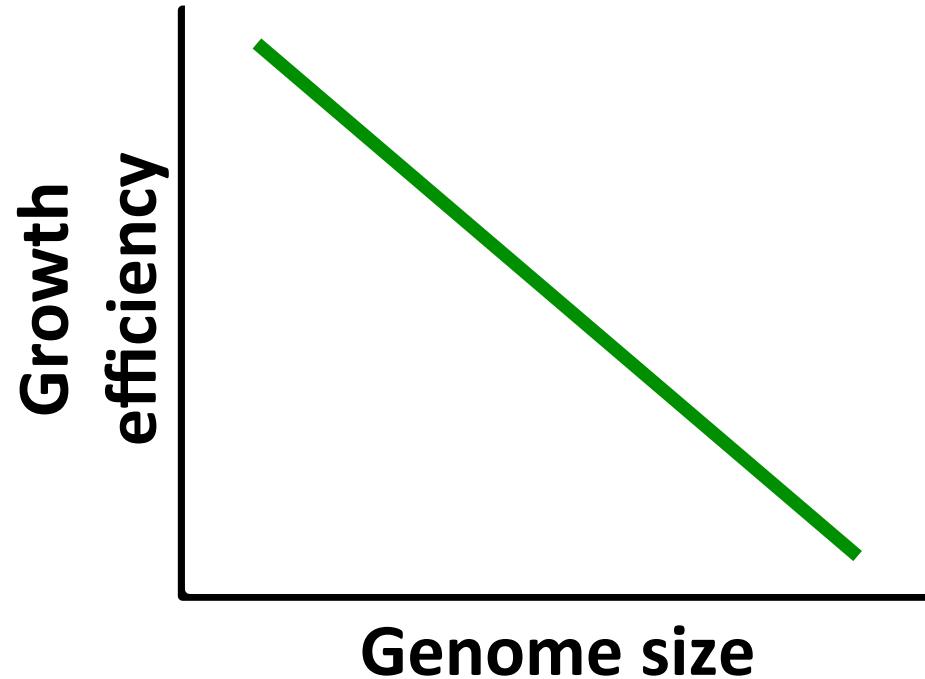


**SAR11 & *Prochlorococcus***  
cells << 1 micron  
genomes < 2.5 Mbp  
doubling time > 1 day



***E. Coli & Pseudomonas***  
cells >> 1 micron  
genomes > 4 Mbp  
doubling time < 1 hour

# Growth efficiency

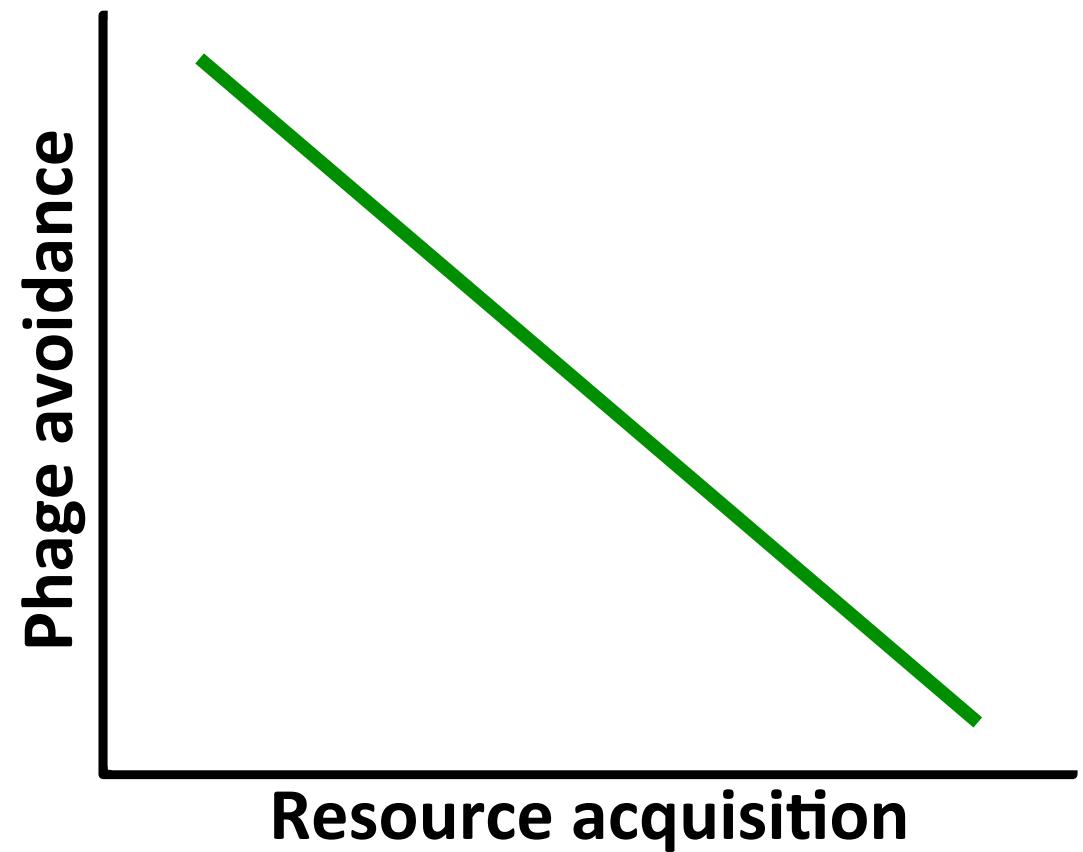
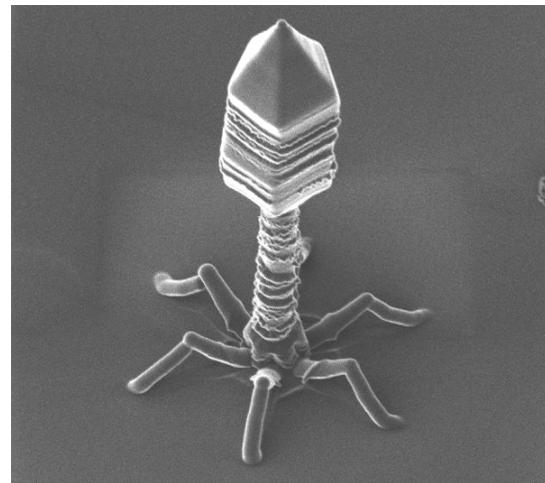


# Only half the story...

$$\frac{dN}{dt} = \text{births} - \text{deaths}$$



# Maybe it's predators, not efficiency



# Dataset

>1300 free-living bacterial genomes

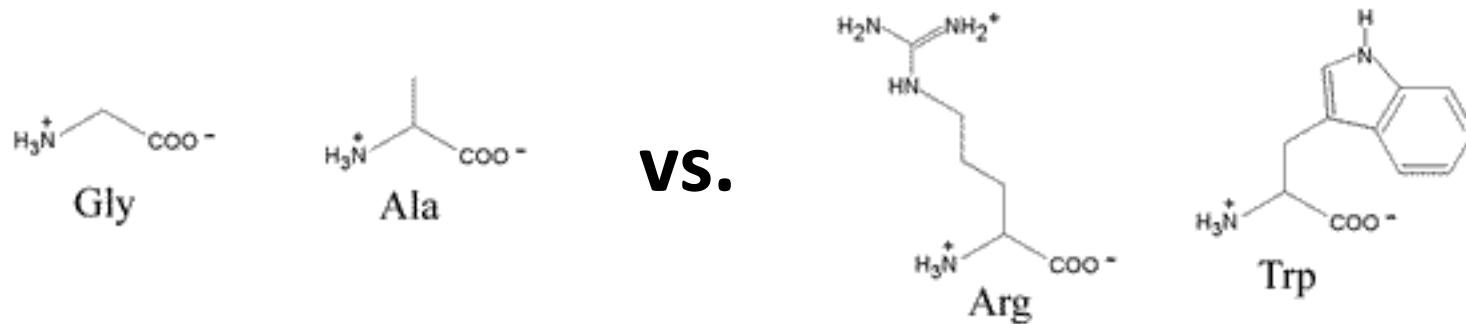
Diverse environments, including soil, marine,  
and freshwater



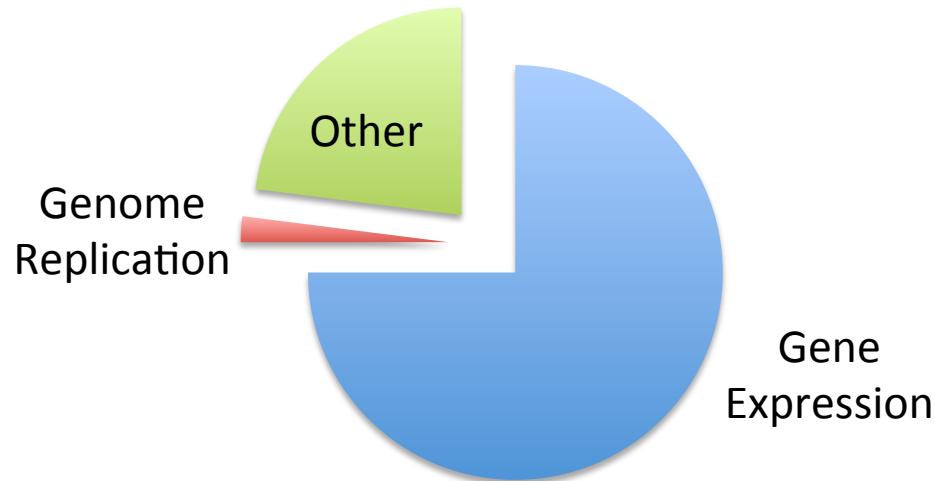
# If it is growth efficiency...

# Small genomes should have:

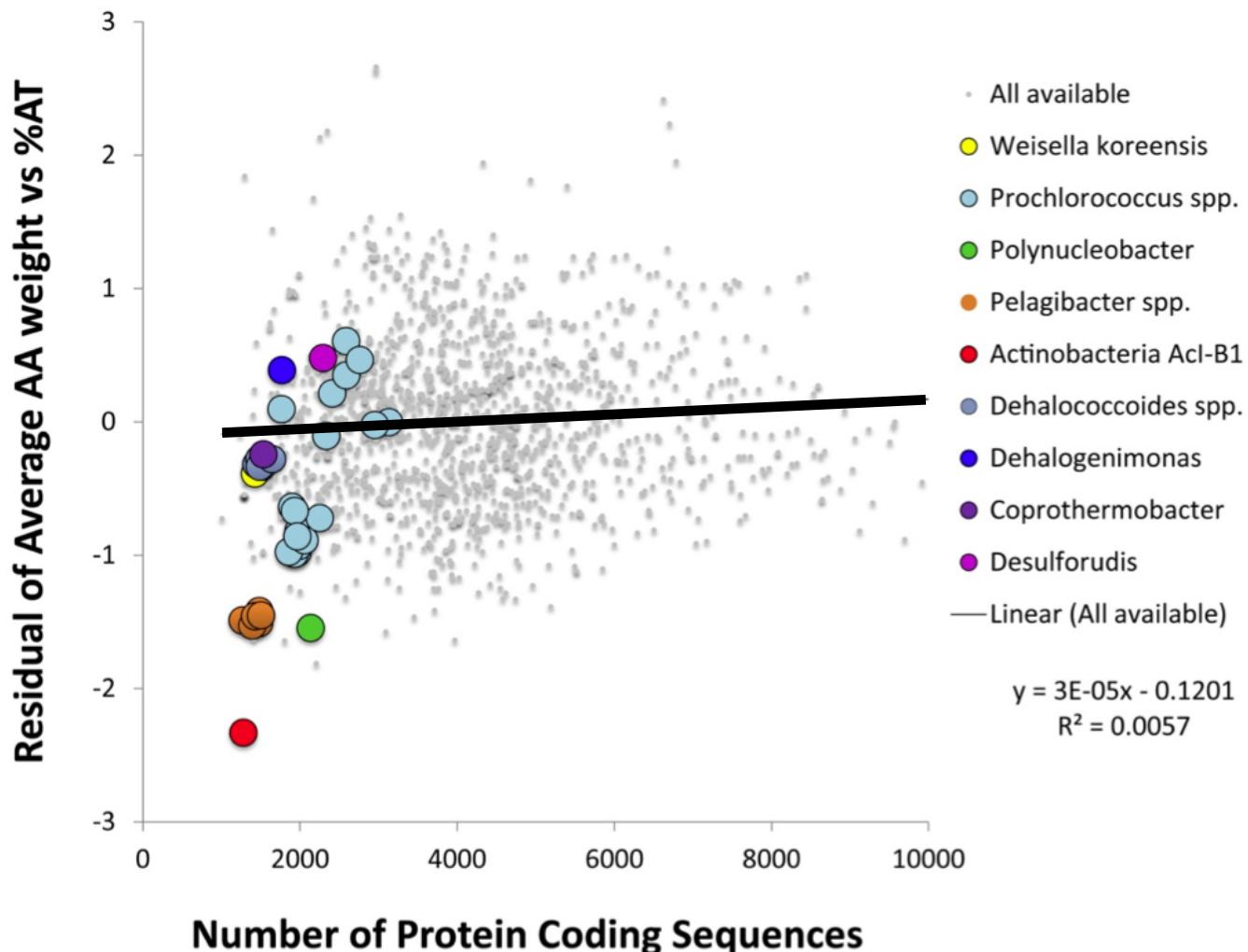
- ## 1) cheap amino acids



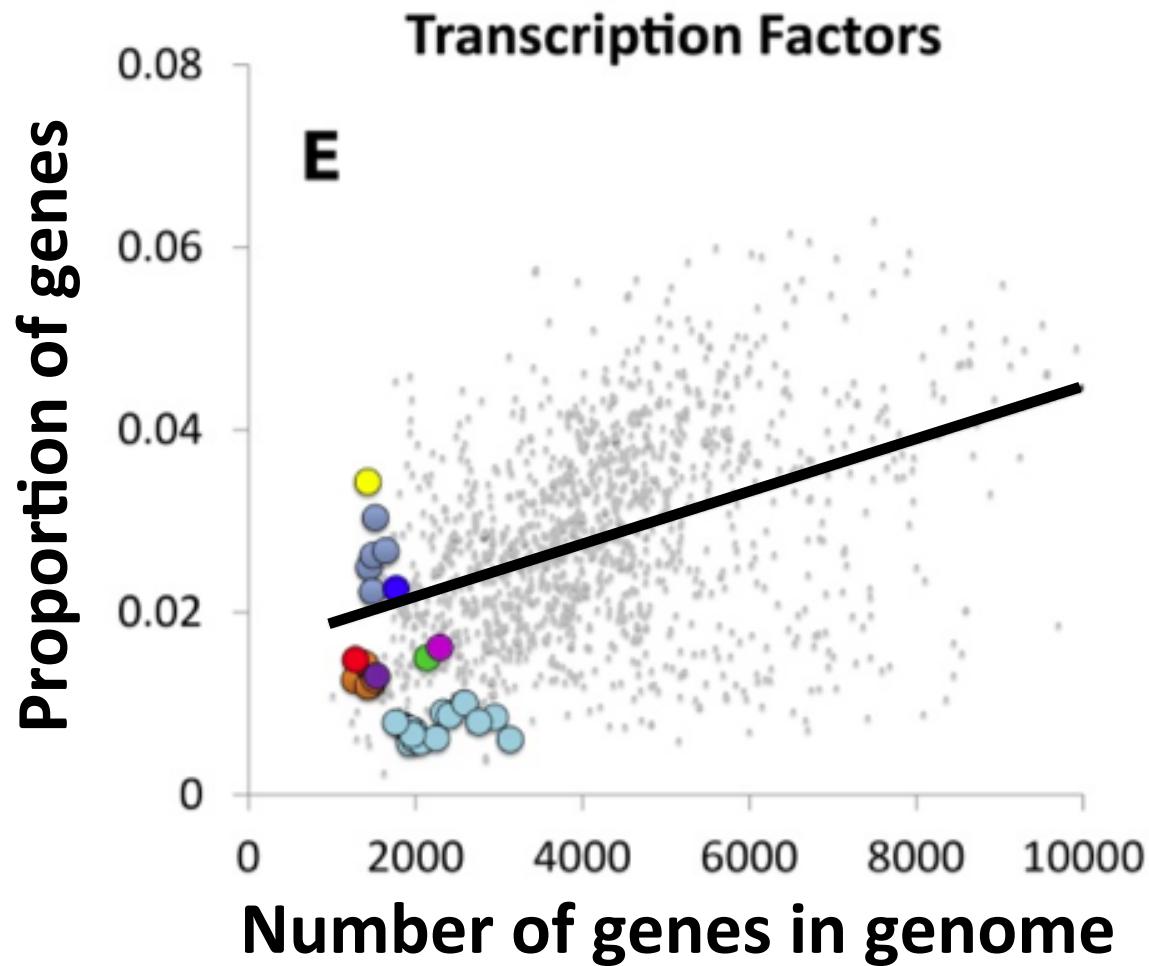
- ## 2) high gene transcription regulation



# 1) Weak, but positive relationship between genome size and AA weight



## 2) Reduced gene regulation in smaller genomes



# If it is predation...

Small genomes should have:

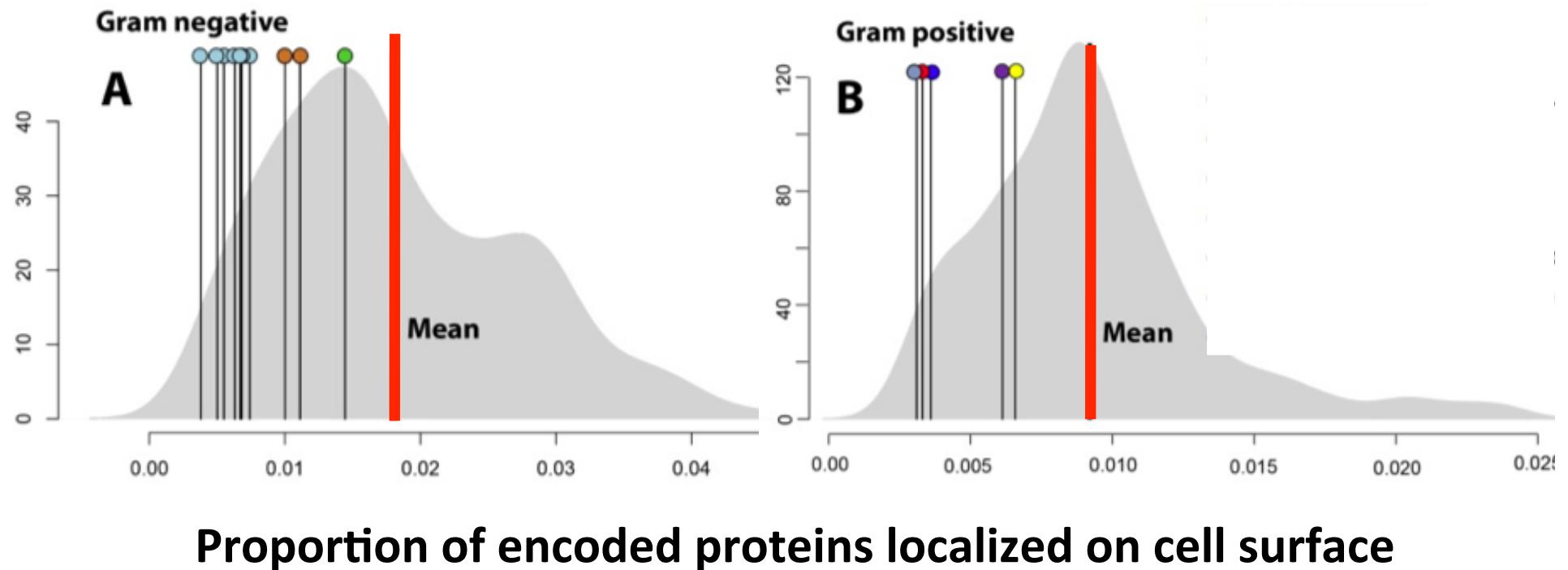
- 1) small number of surface proteins



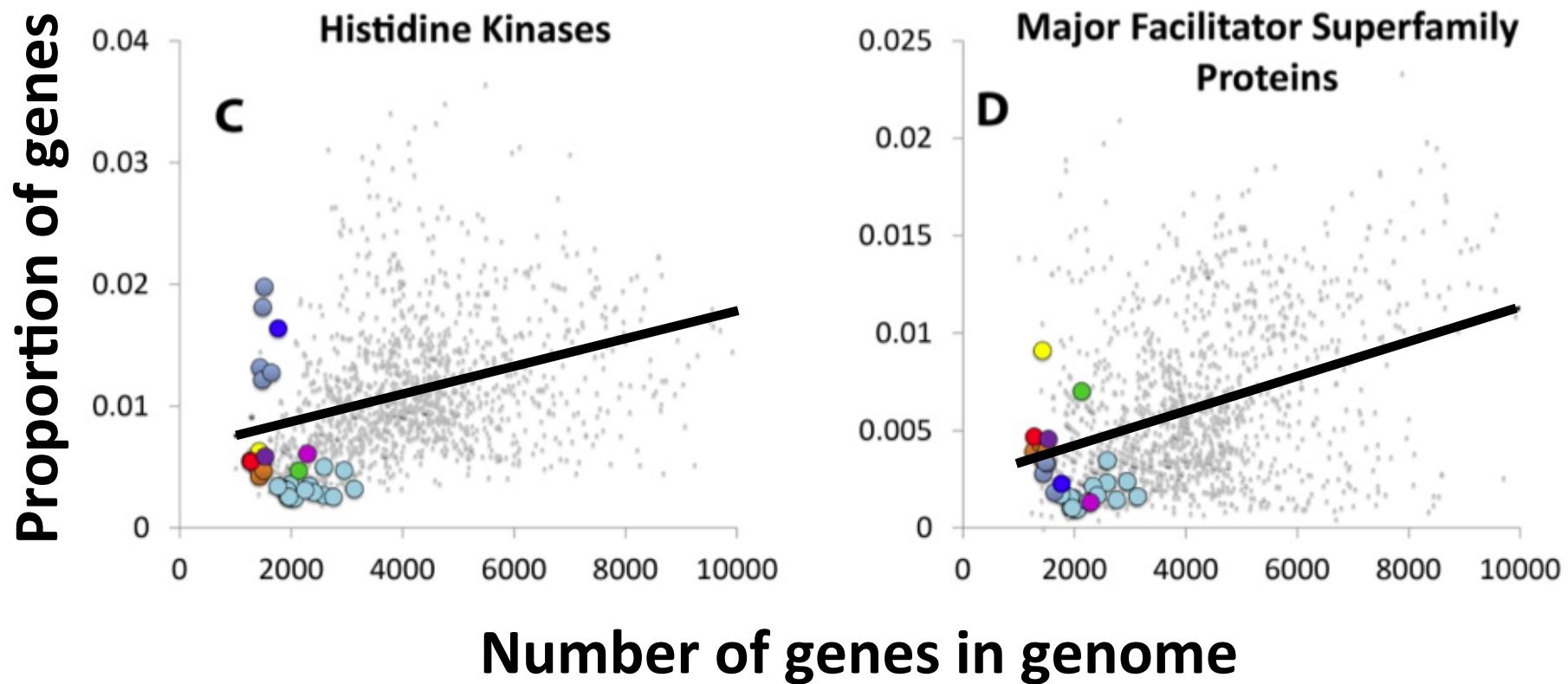
- 2) reduced genome allocation to transporter and environmental sensing machinery



# 1) Reduced surface proteins in “poster children”



## 2) Reduced sensory and transport machinery



# Our working model



**self-imposed  
shut-ins**

**vs.**

**live fast and die  
young**

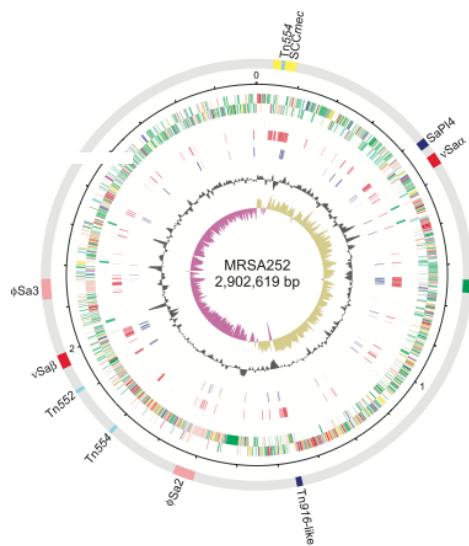


# Genomic markers of microbial life history strategies

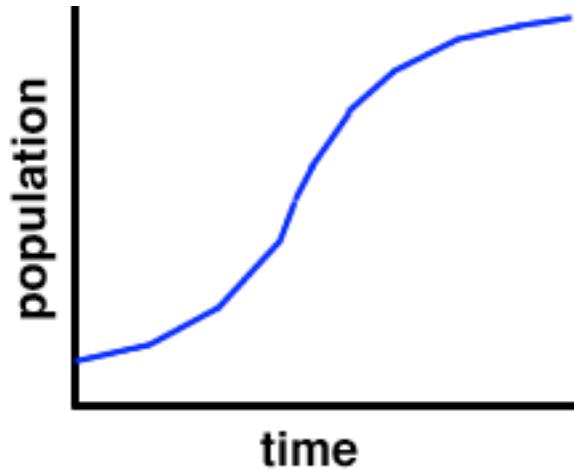
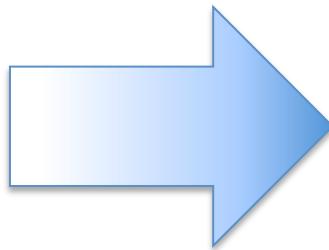
- Comparative genomics is a powerful tool
  - Provides taxonomy and functional capacity
  - Enables genotype-phenotype links
- Can identify putative tradeoffs and lifestyles
- We can't forget about physiology and natural history!

# How quantitative can we be?

Genome



Ecophysiological  
Traits





Dan Liu



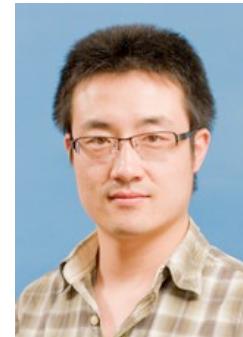
Joshua  
Livermore



Scott Emrich



Erliang Zeng



Wei Zhang



ENVIRONMENTAL  
CHANGE  
INITIATIVE

