



Metagenomics of bacterial communities

Or: Do microbial communities behave as meta-genomes?

Codon-usage mediated translational optimization in metagenomes

Outline and Concepts

- Microbial community genomics - metagenomics
 - High-throughput sequencing
 - Inventorying genomic content
- Translational optimization of prokaryote gene expression
 - Predicting expressivity
 - Linking function and expressivity
- Linking it all together
 - Do metagenomes behave as meta-genomes?

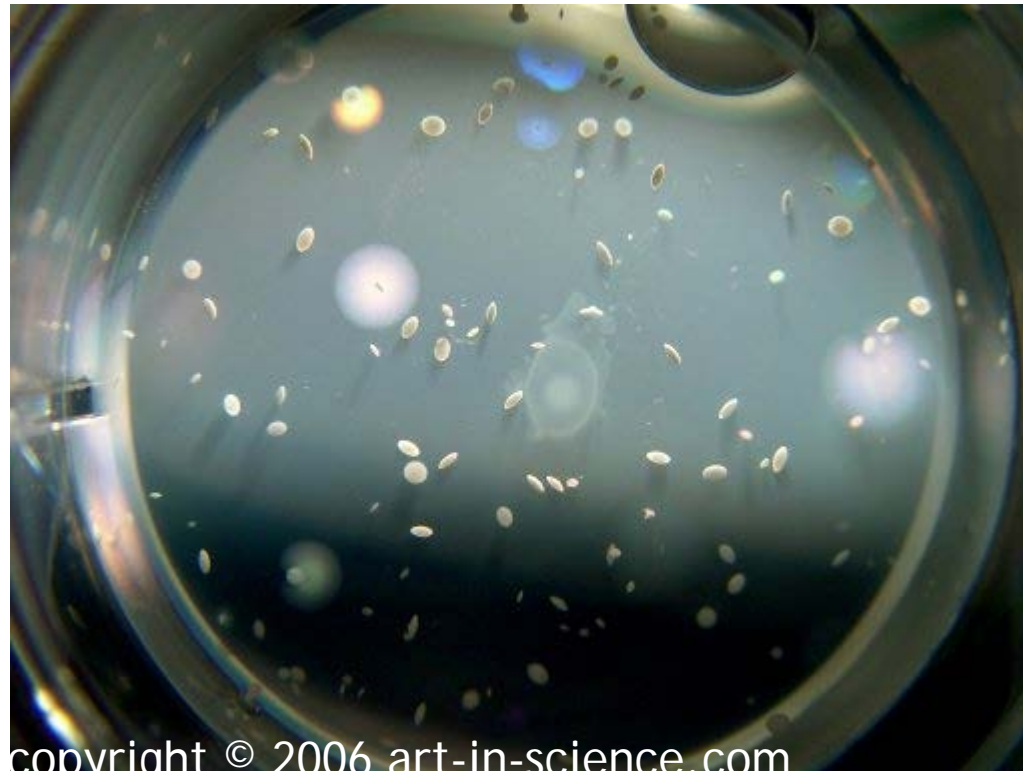


Chapter I

MICROBIAL METAGENOMICS

Bugs Are Hard to Culture...

- 1% of all microorganisms are lab cultivable



copyright © 2006 art-in-science.com

And, No *microbe* is an island...

John Donne (1572-1631), Meditation XVII.



EARLY LIFE FORMS

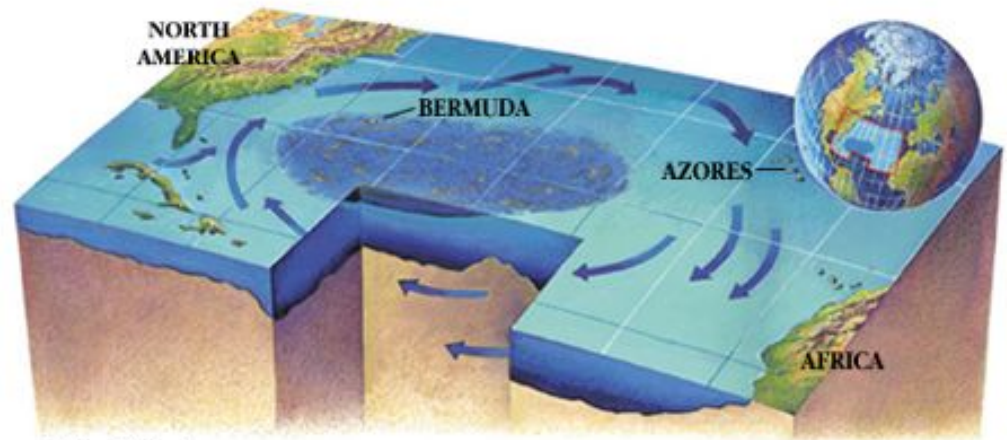
<http://www.nearingzero.net>

Enter Metagenomics!

- Capt'n Craig set sail for the Bermudas....



© Michael Scholz, California Wild



©Harold Smelcer

© Smithsonian mag, nov 98

Metagenomics

SMASH 'N' GRAB

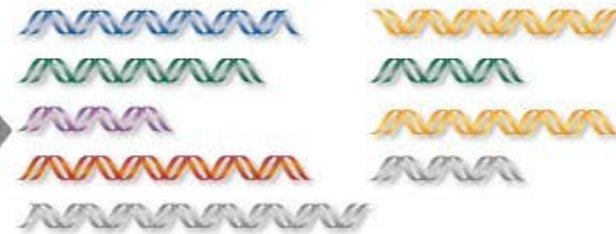
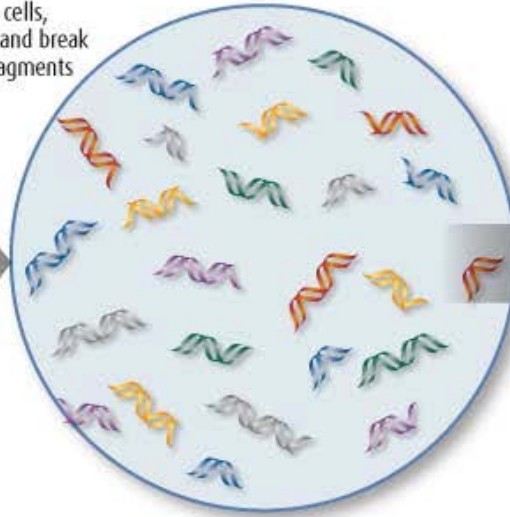
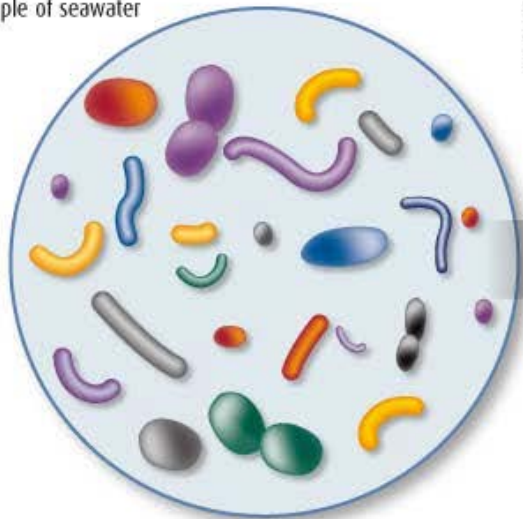
Metagenomics allows you to sequence new organisms that can't be cultivated in the lab

©New Scientist '07

Sample of seawater

Smash open cells,
extract DNA and break
into small fragments

Sequence fragments and
reassemble sequences



Take environmental
sample

Sample DNA

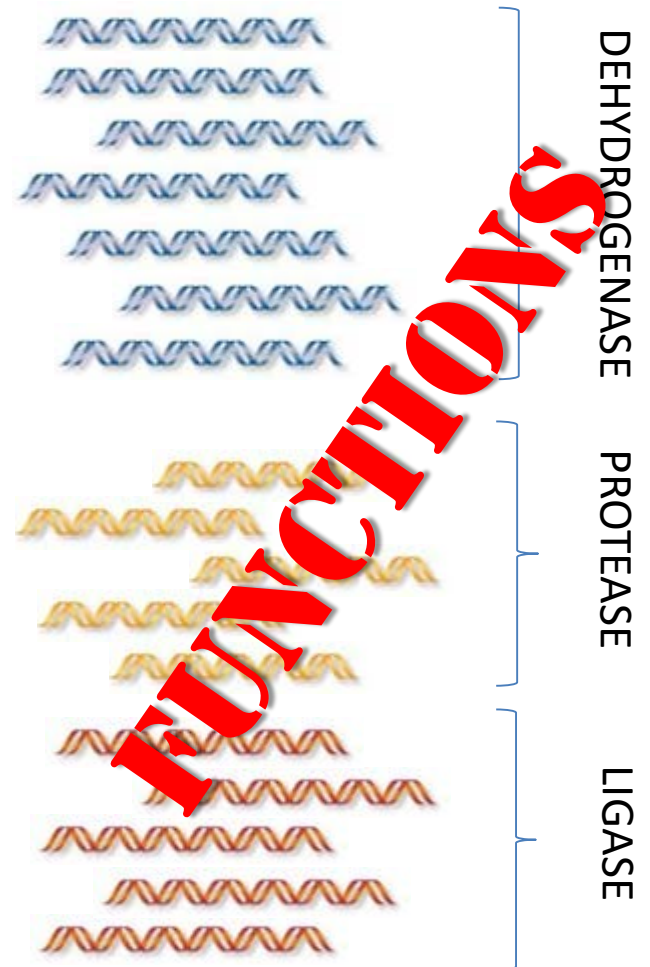
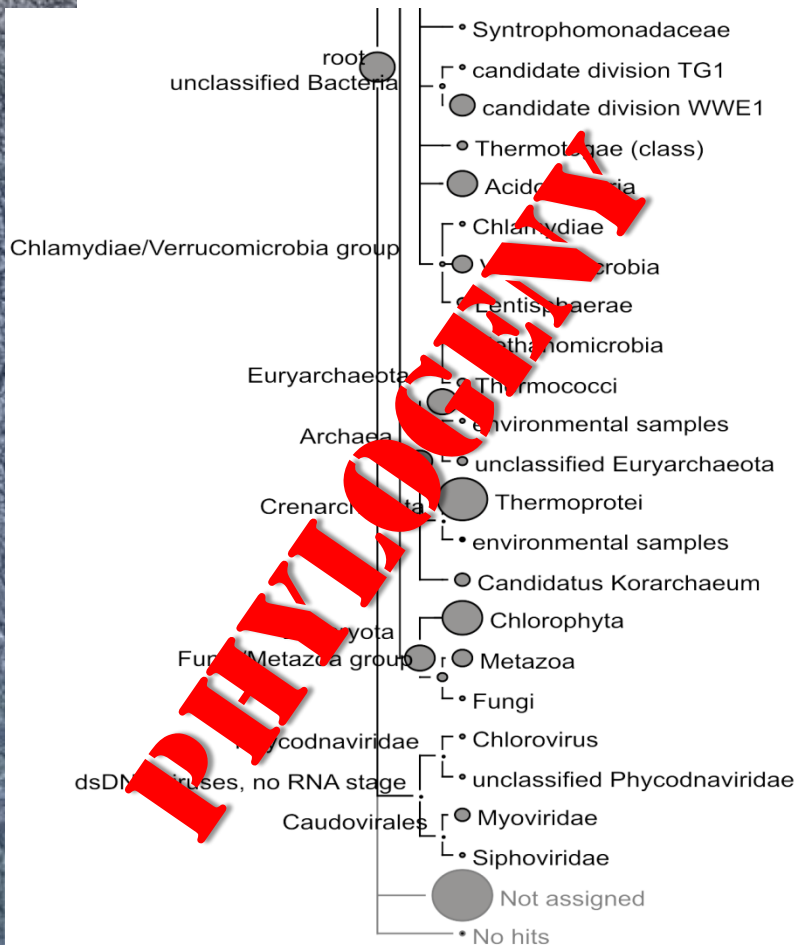
Sequence and assemble
fragments

Some assembly required...



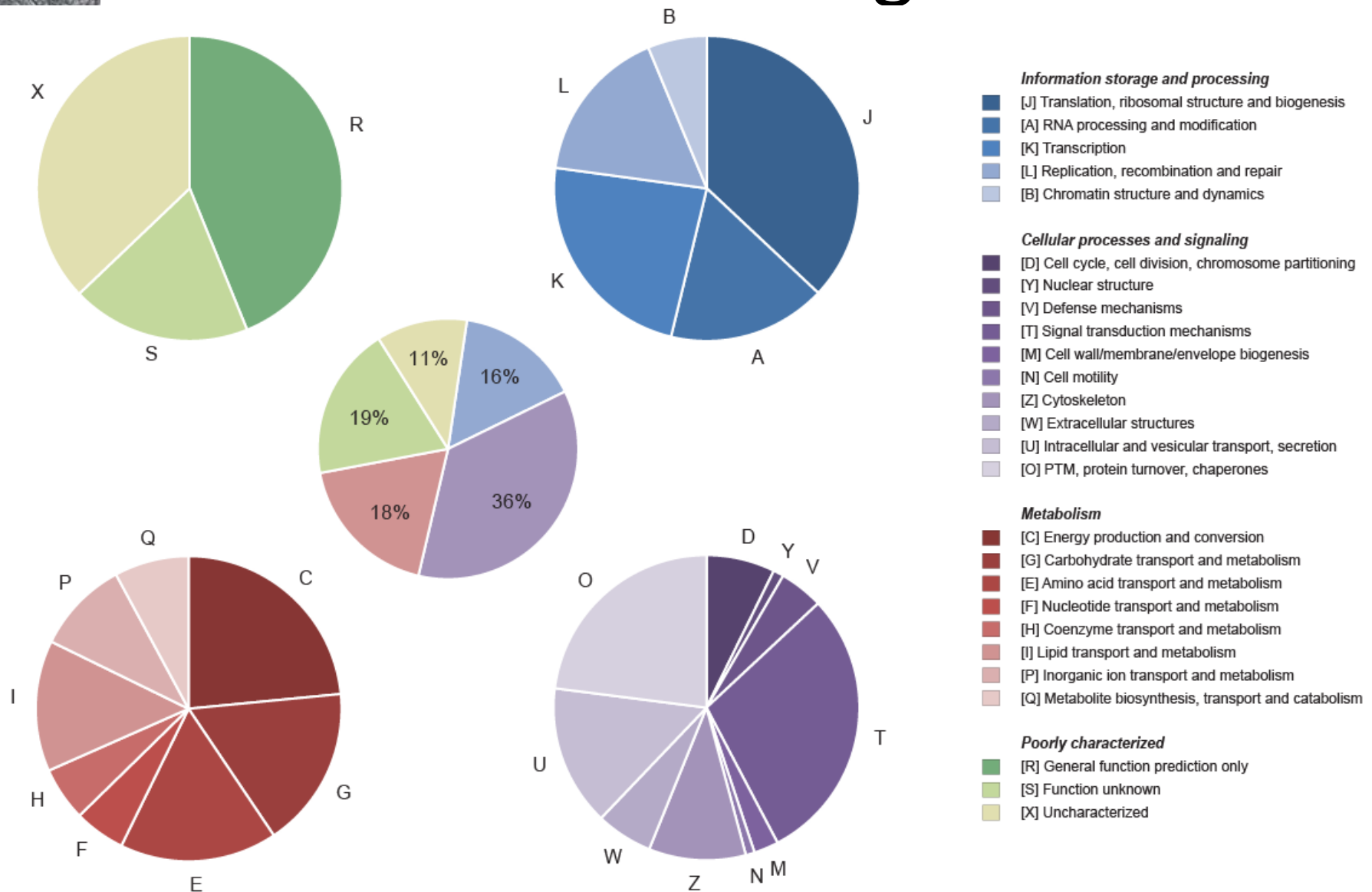
© Drew Sheneman

Making sense of it all



...

Clusters of Orthologous Genes





Chapter II

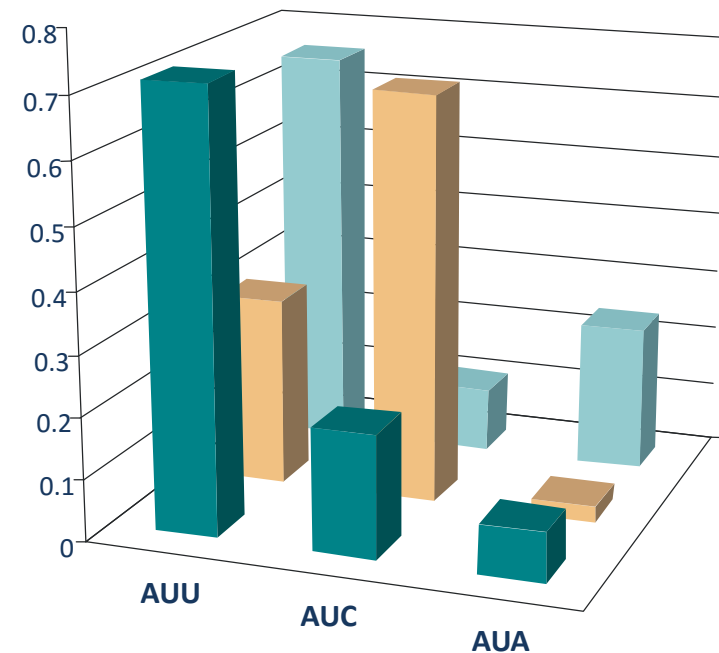
OPTIMIZATION OF TRANSLATION

Synonymous codon usage

1st position (5' end)	2nd position				3rd position (3' end)
↓	U	C	A	G	↓
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr Stop Stop	Cys Cys Stop Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Related by Transitions in the 3rd Position

ILE

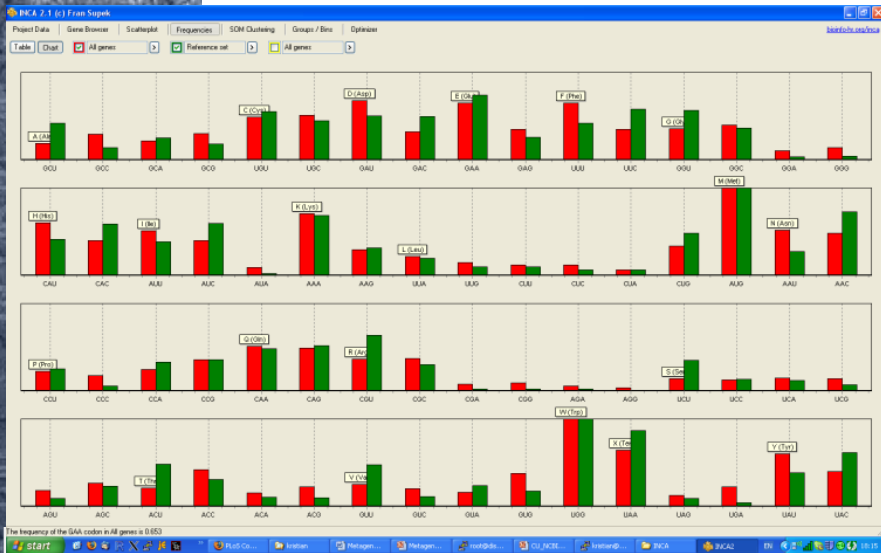


- *Mycoplasma pulmonis*
- *Deinococcus radiodurans*
- *Haemophilus influenzae*

CU bias in microbial genomes

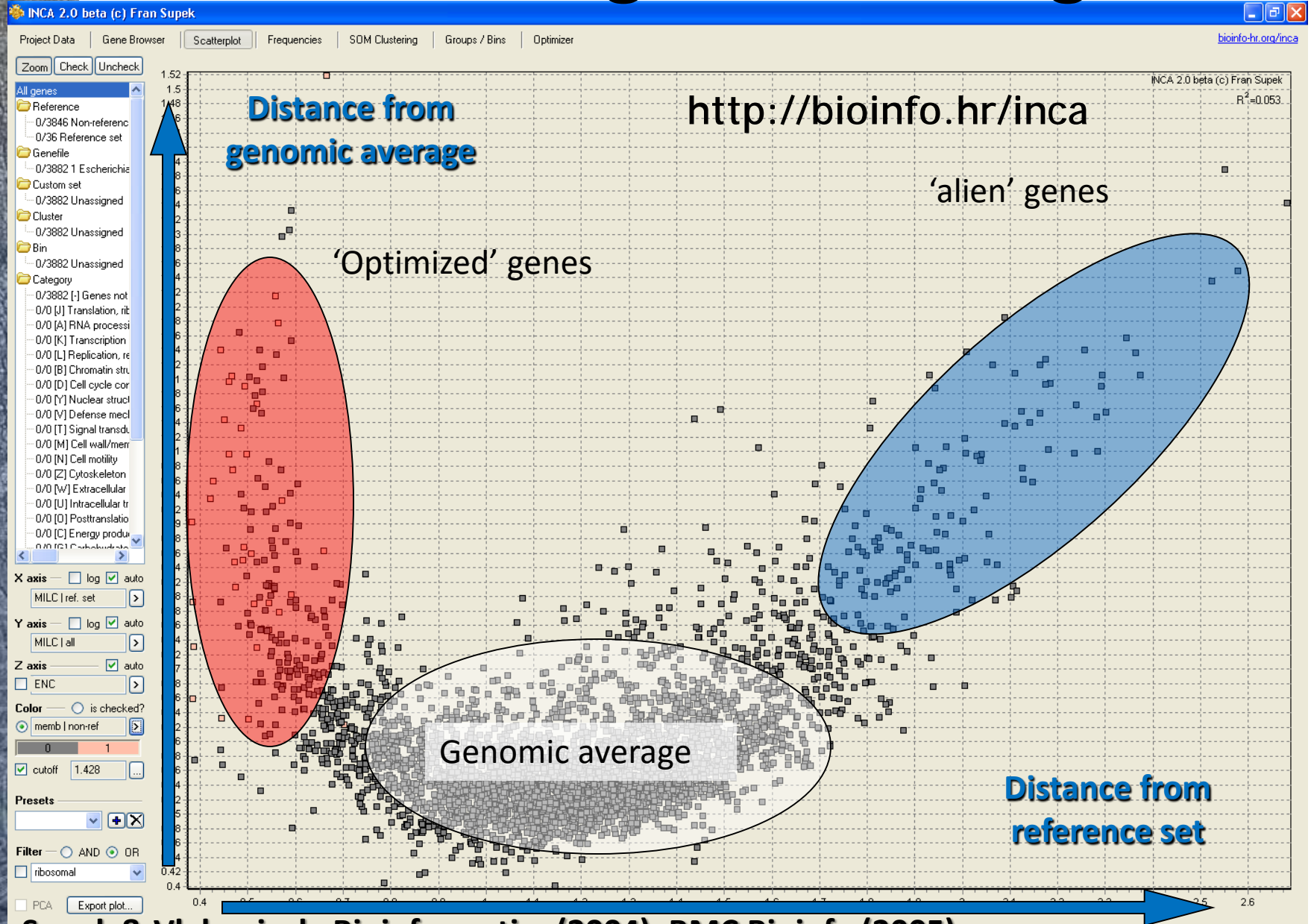
- Synonymous codons used differently
 - Between different species
 - GC content and AA composition
 - Within a single genome
 - “optimaly” encoded genes choose codons compatible to tRNA abundance and mRNA folding
 - Ribosomal proteins
 - Elongation factors
 - Chaperones
 - Background selection for “lifestyle specific” functions

Single sequence
Whole genome
Reference sequence set



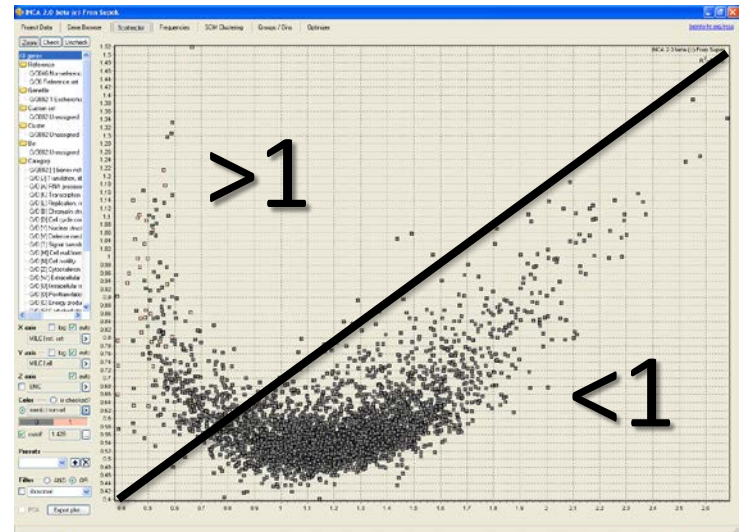
MILC – measure
independent of length and
composition

Measuring Codon Usage



Predicted expressivity

- Distance to genomic mean vs. distance to the reference set



- 'good' codons ensure optimal expression rates
 - Synechocystis sp.: photosynthesis genes
 - M. janaschii: methanogenesis genes
 - D. radiodurans: membrane and detox proteins
- MELP – MILC-based expression level predictor



Chapter III

TRANSLATION OPTIMIZATION AT THE METAGENOME LEVEL

Microbial communities

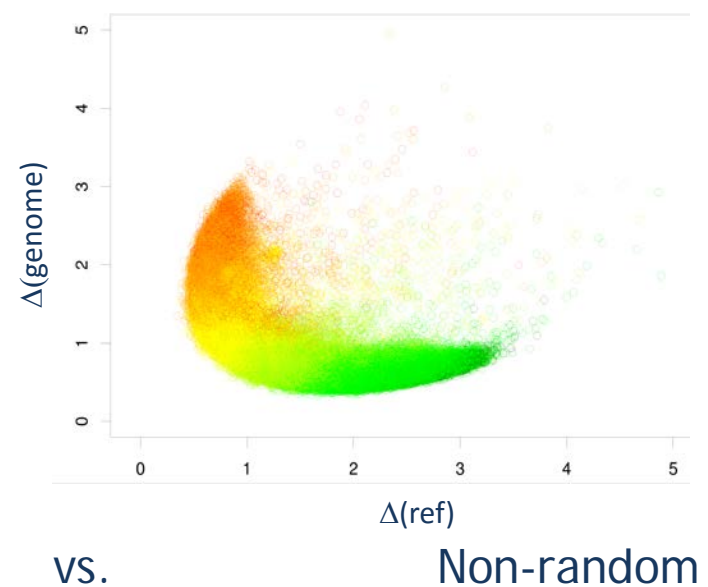
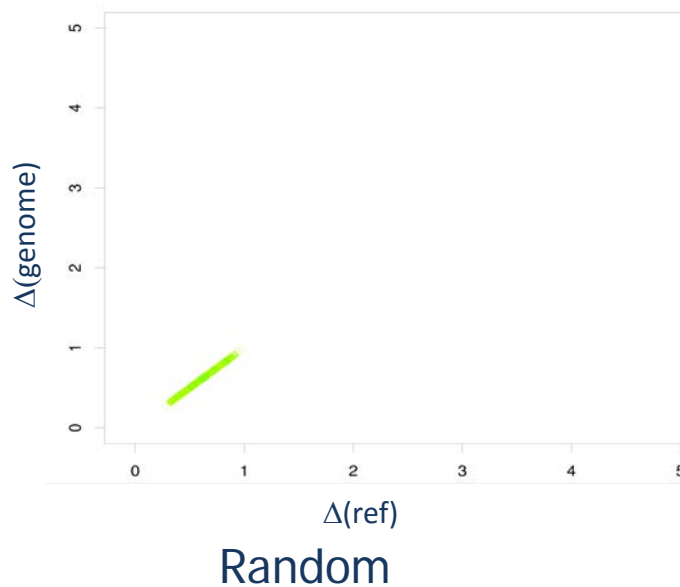
- Hypothesis: microbial communities behave as meta-genomes
 - Co-evolution introduces bias in CU
 - Habitat-wide optimization of function
 - Horizontal gene transfer easier
- Test on environmental shotgun sequence projects

Metagenomes analysed

- Sargasso sea (688539 sequences)
 - [Venter JC et al.](#), "Environmental genome shotgun sequencing of the Sargasso Sea.", *Science*, 2004 Apr 2;304(5667):66-74
- Minnesota soil sample (102879 sequences)
 - [Tringe SG et al.](#), "Comparative metagenomics of microbial communities.", *Science*, 2005 Apr 22;308(5721):554-7
- Whale carcass (66165 sequences)
 - [Tringe SG et al.](#), "Comparative metagenomics of microbial communities.", *Science*, 2005 Apr 22;308(5721):554-7
- Acid mine drainage (10214 sequences)
 - [Tyson GW et al.](#), "Community structure and metabolism through reconstruction of microbial genomes from the environment.", *Nature*, 2004 Mar 4;428(6978):37-43
- Human gut microbiome (79613 sequences)
 - [Gill SR et al.](#), "Metagenomic analysis of the human distal gut microbiome.", *Science*, 2006 Jun 2;312(5778):1355-9
- Lean and obese mice gut microbiomes (570k seq)
 - [Turnbaugh PJ et al.](#), "An obesity-associated gut microbiome with increased capacity for energy harvest.", *Nature*, 2006 Dec 21;444(7122):1027-31

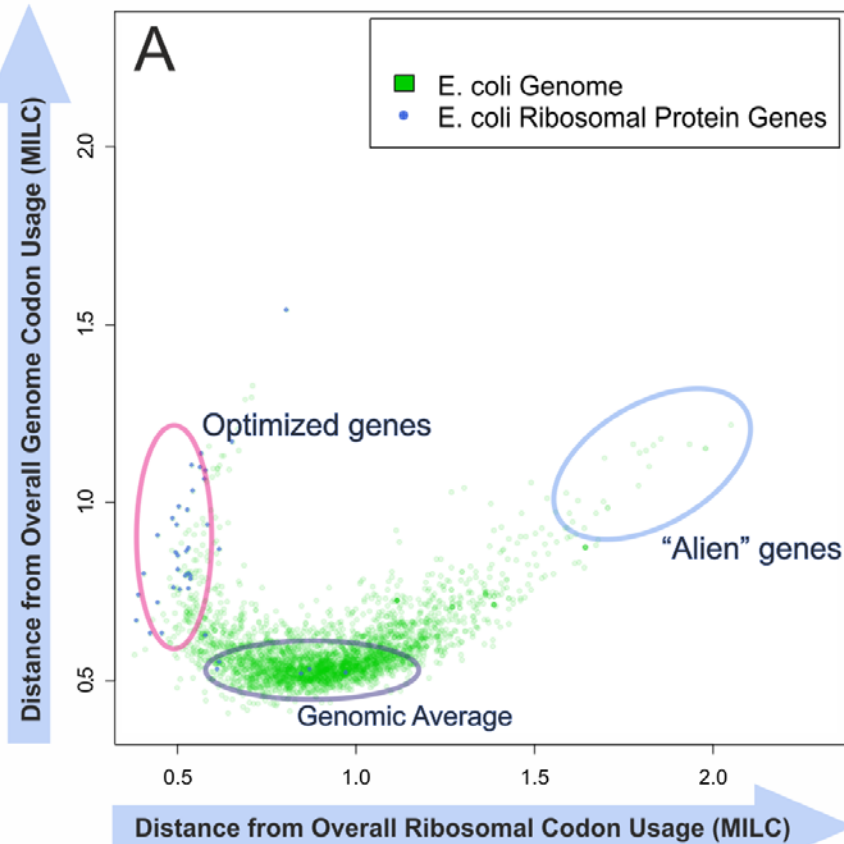
So, what about CU in metagenomes?

- Can we spot the bias in codon usage throughout the entire metagenome?

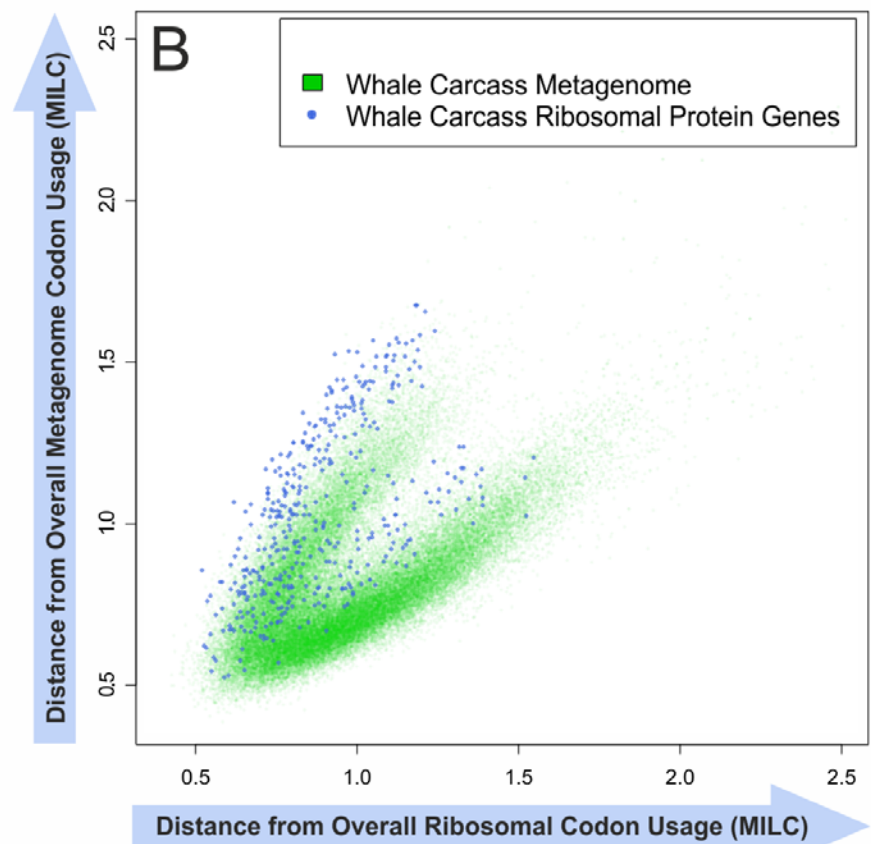


Yes, we can!

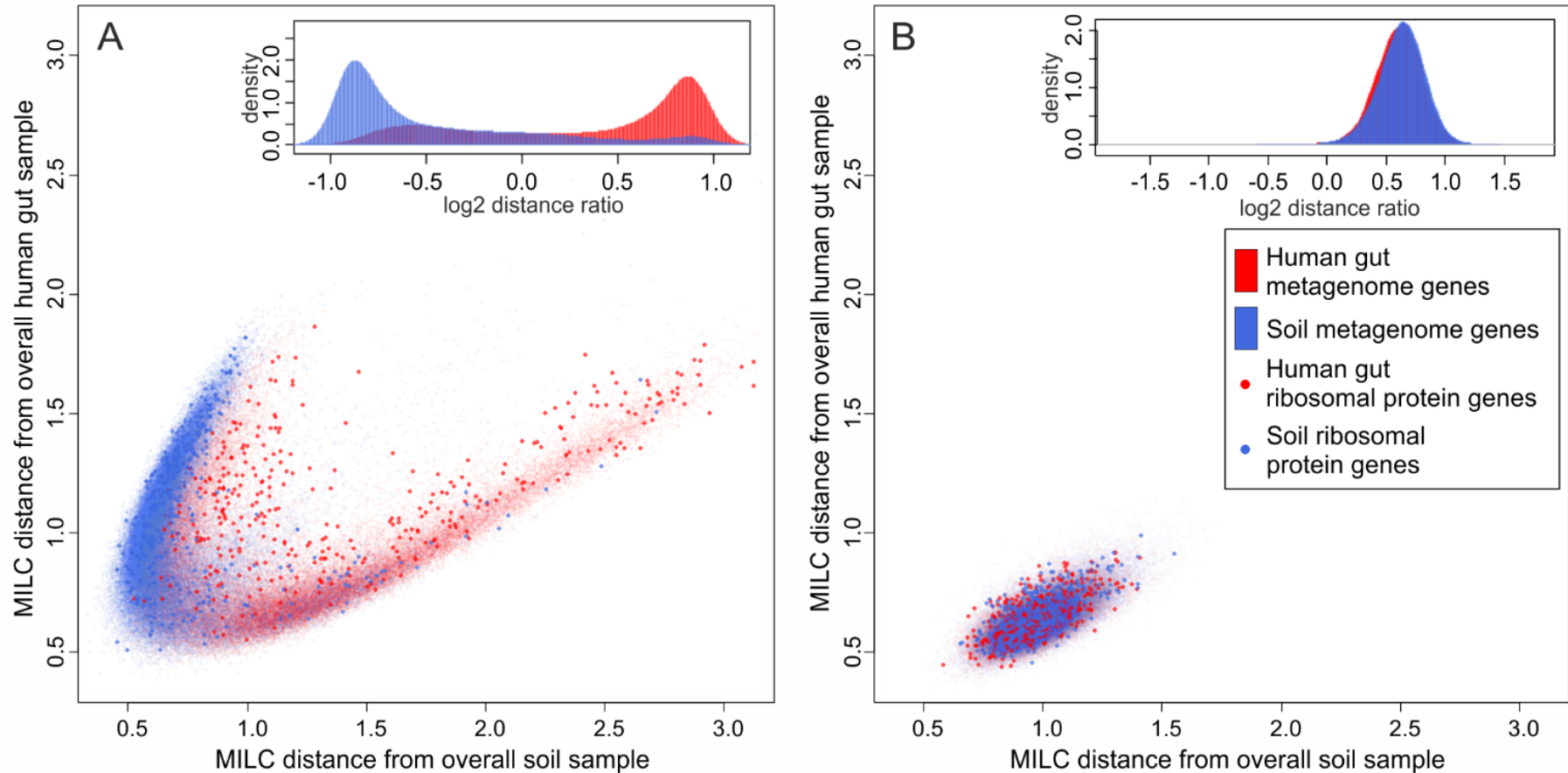
E. coli Genome



Whale Carcass Metagenome



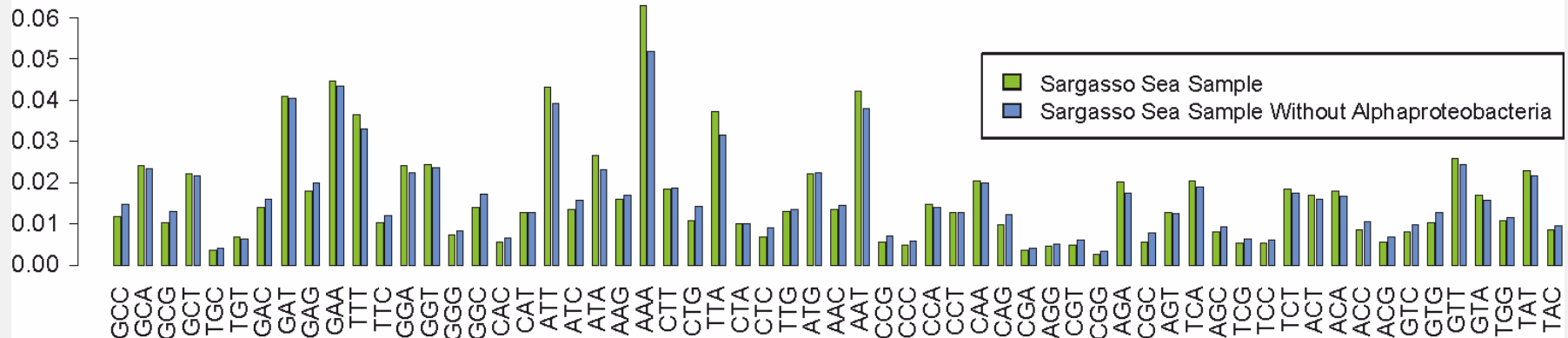
Environment-specific CU



Is this a composition effect?

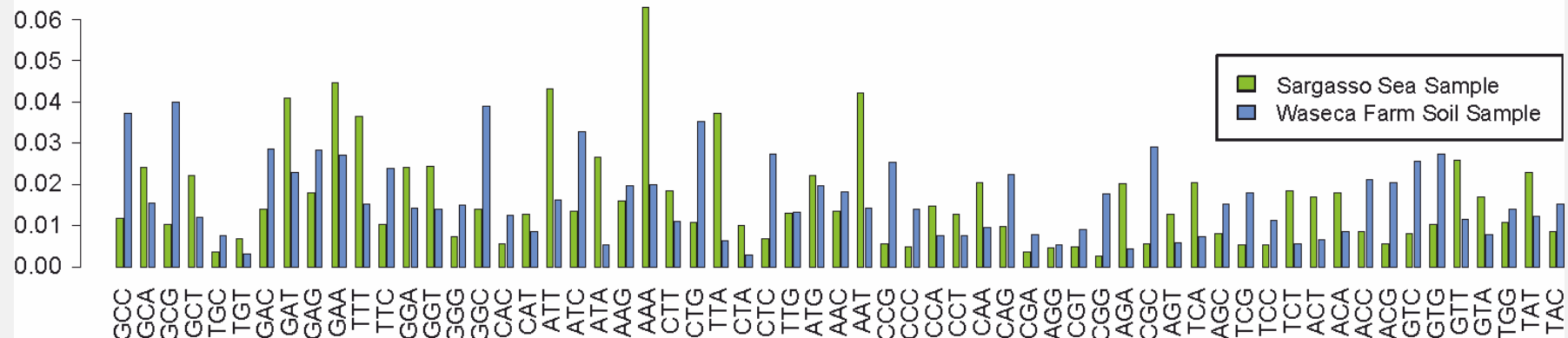
Sargasso Sea Sample vs. Sargasso Sea Sample Without Alphaproteobacteria Codon Frequencies

ICC= 0.978



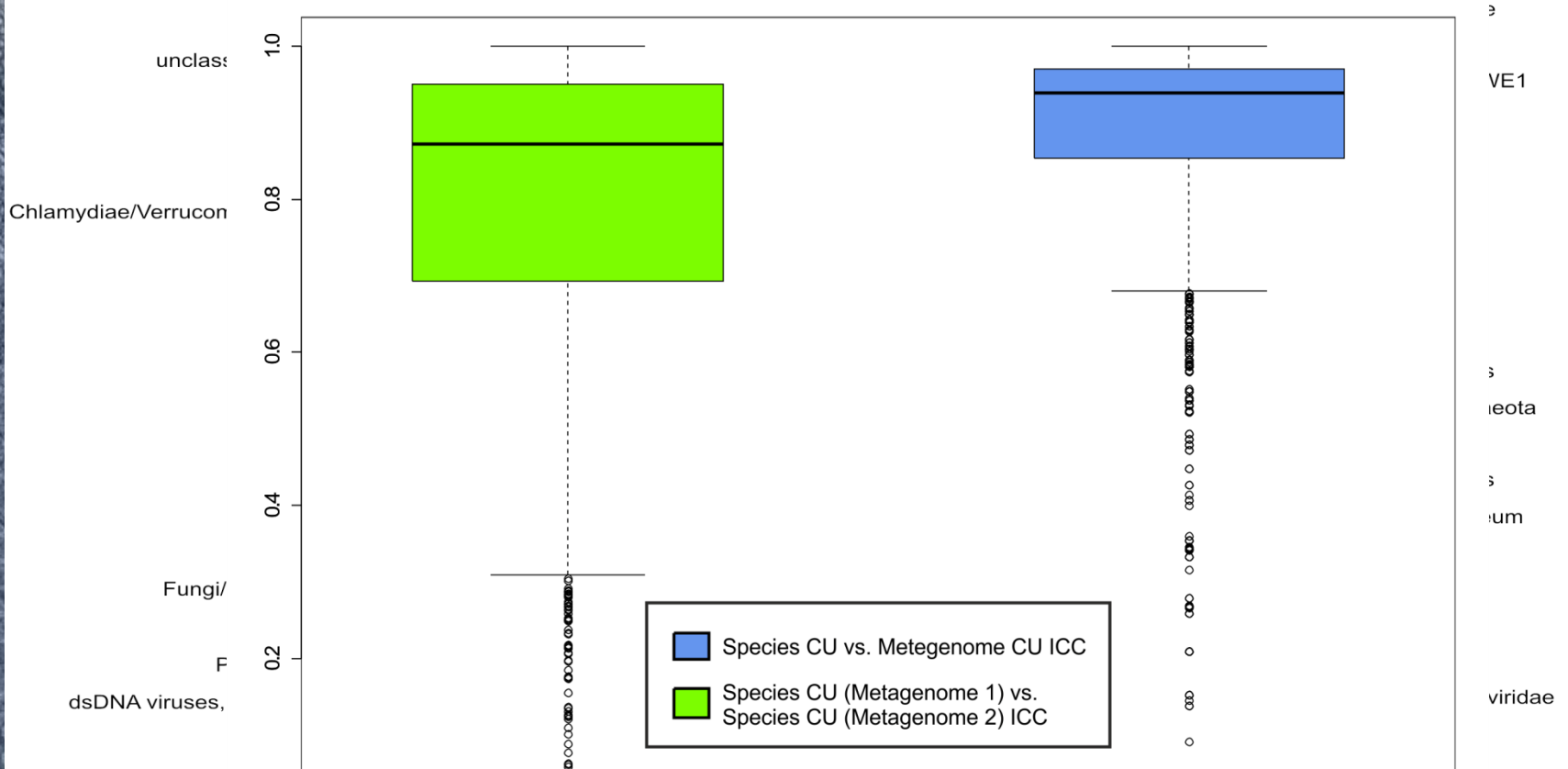
Sargasso Sea Sample vs. Waseca Farm Soil Sample Codon Frequencies

ICC= -0.035



Is this a composition effect?

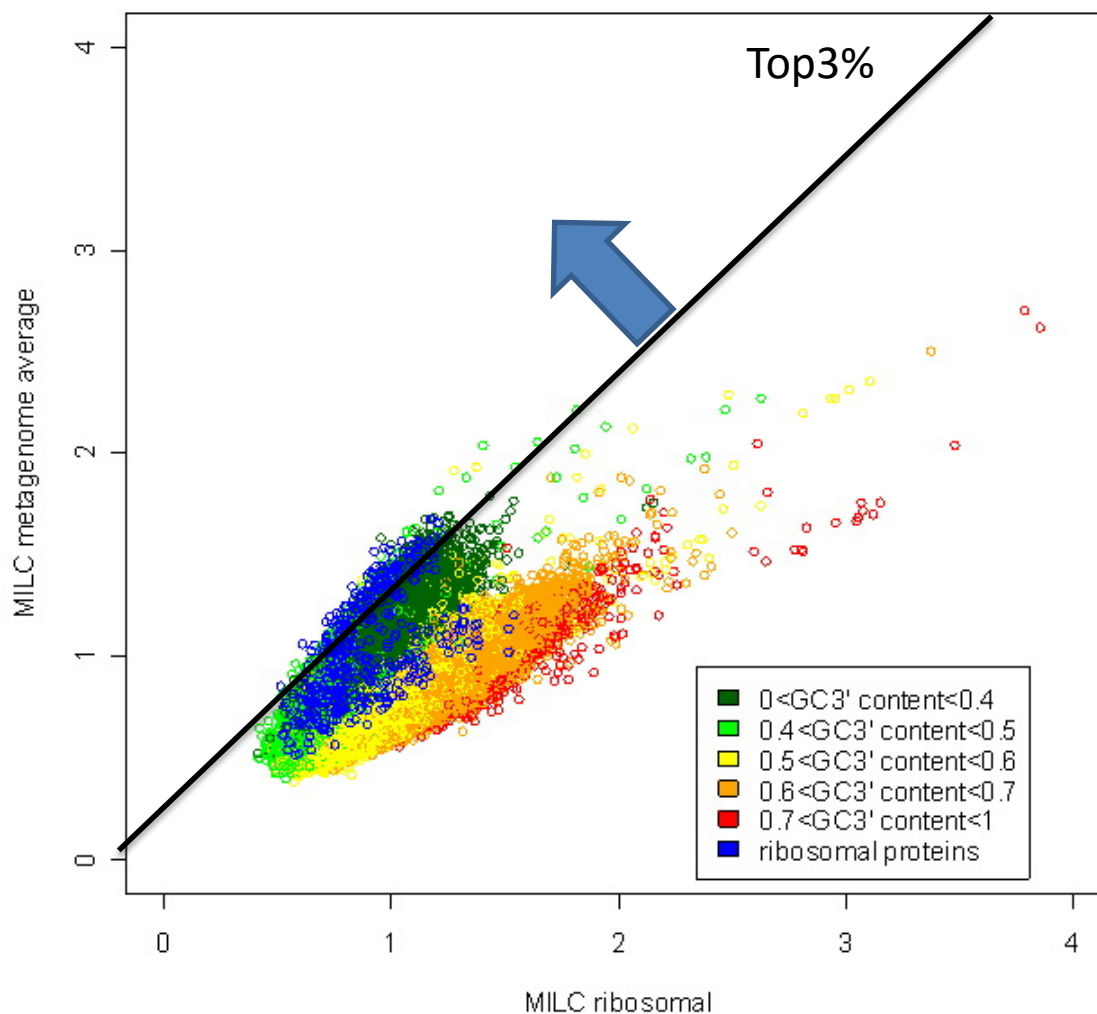
CU of phylogenies across metagenomes (ICC Comparison)



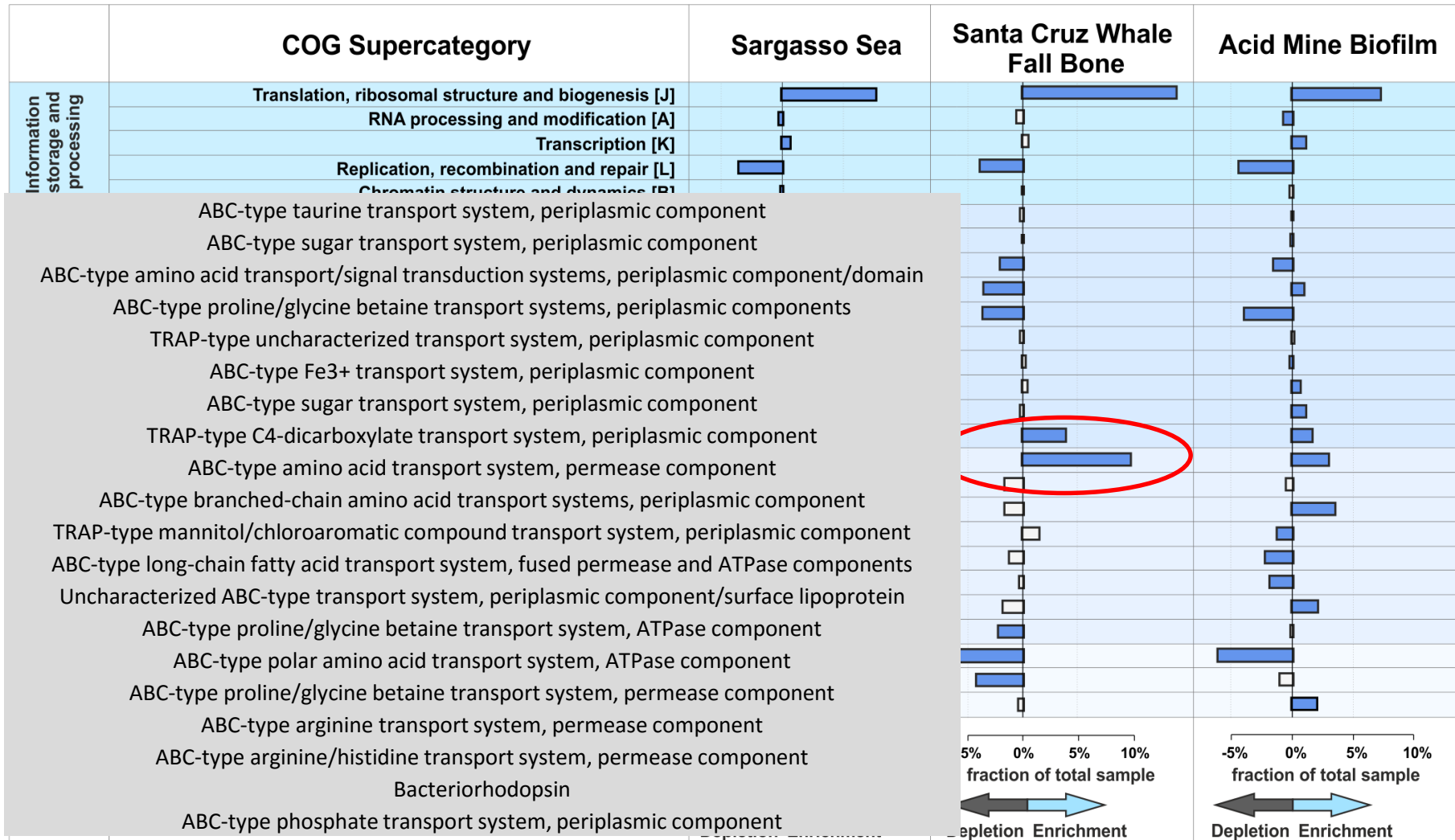
Kolmogorov-Smirnov Test
D: 0.2256 p-value < 2.2e-16
N: 1029 species

Enrichment of functions

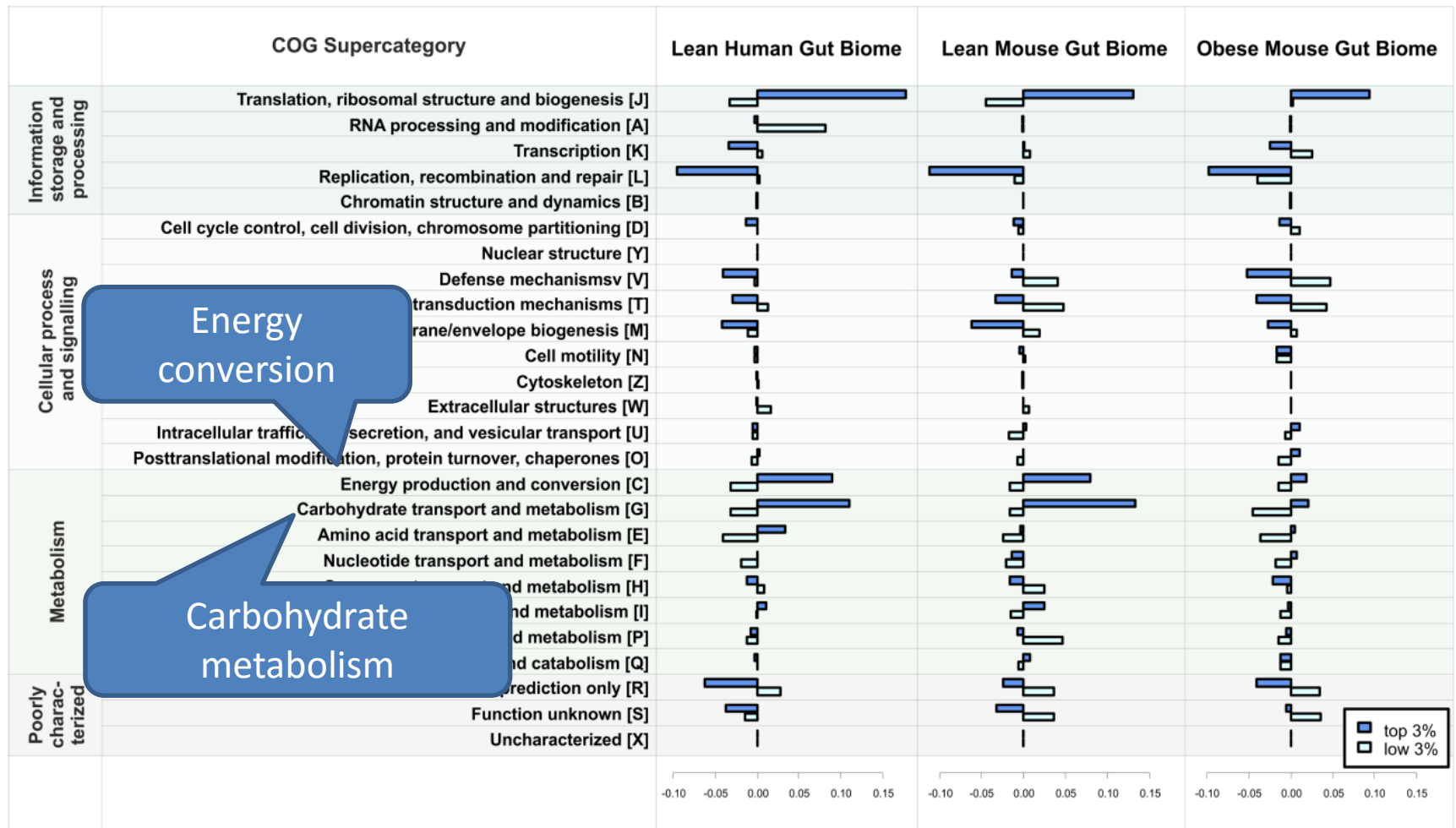
MILC metagenome average vs MILC ribosomal for whale carcass



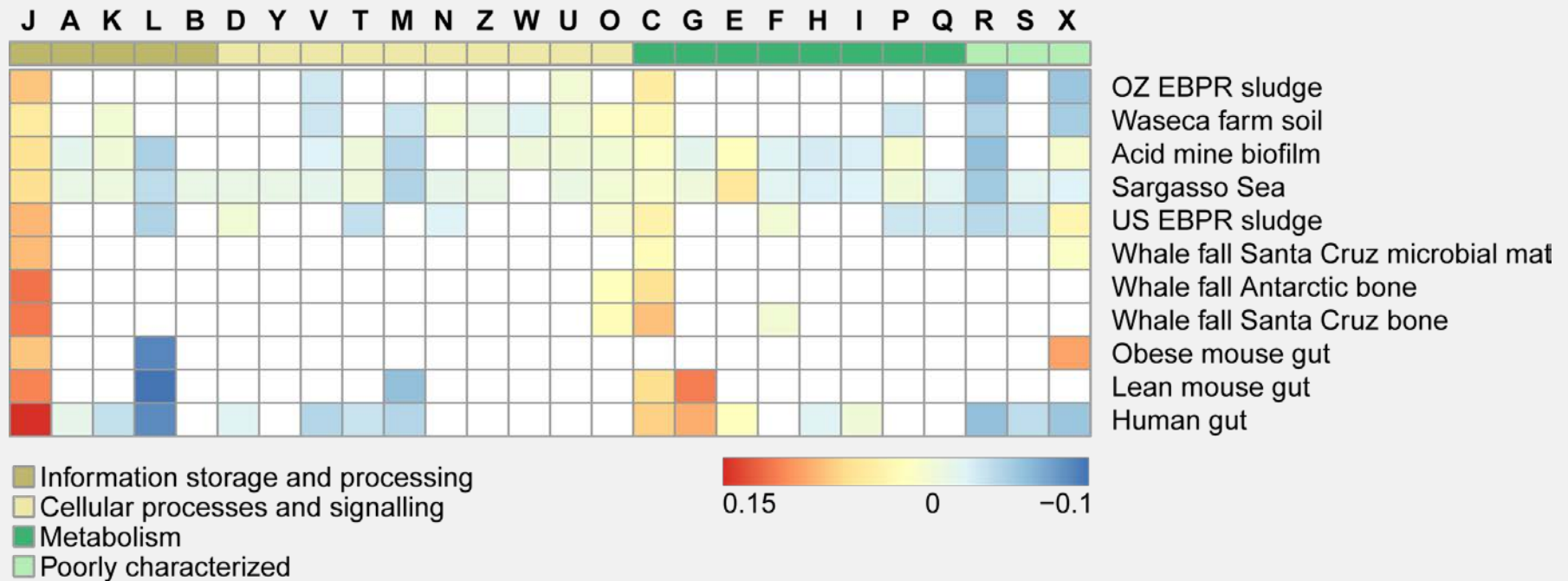
Environment-specific functions



Environment-specific functions



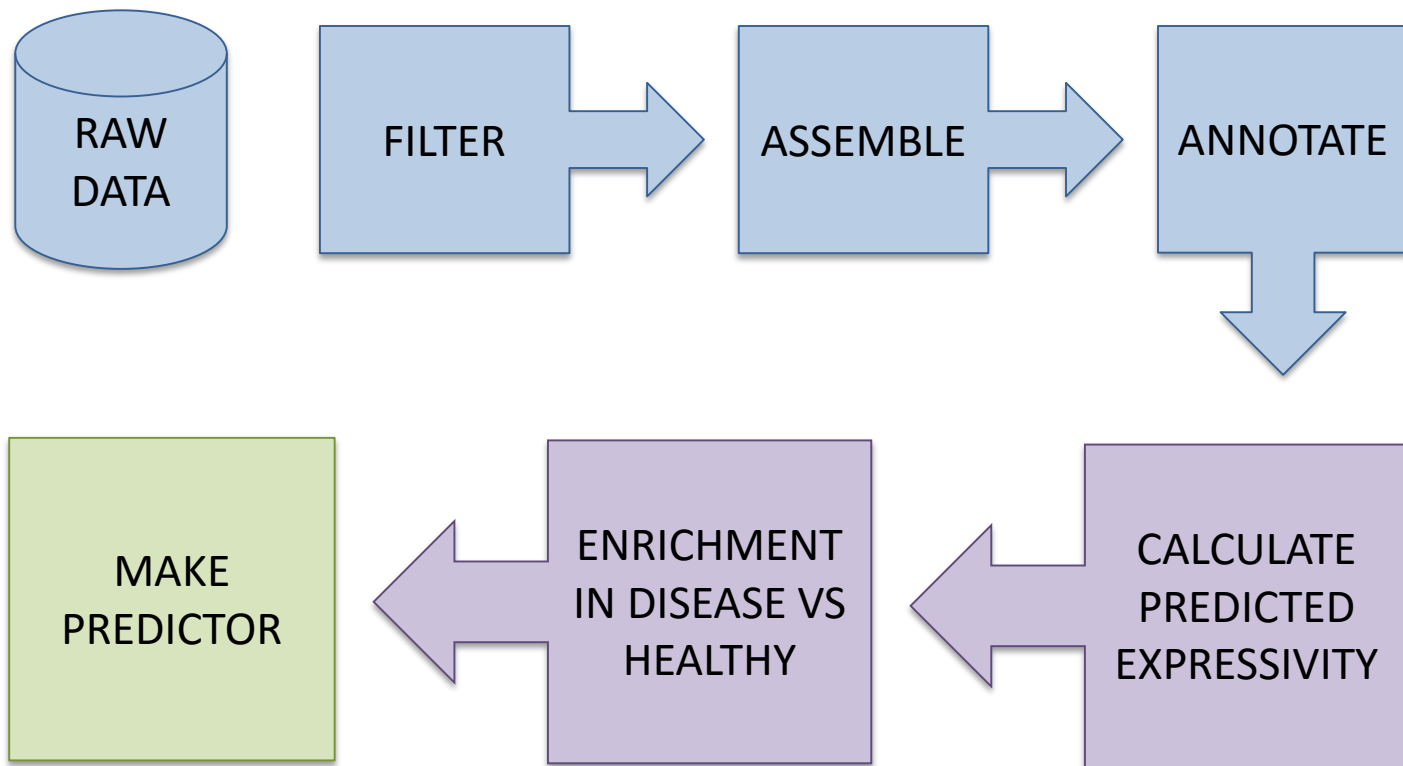
Functional profiles



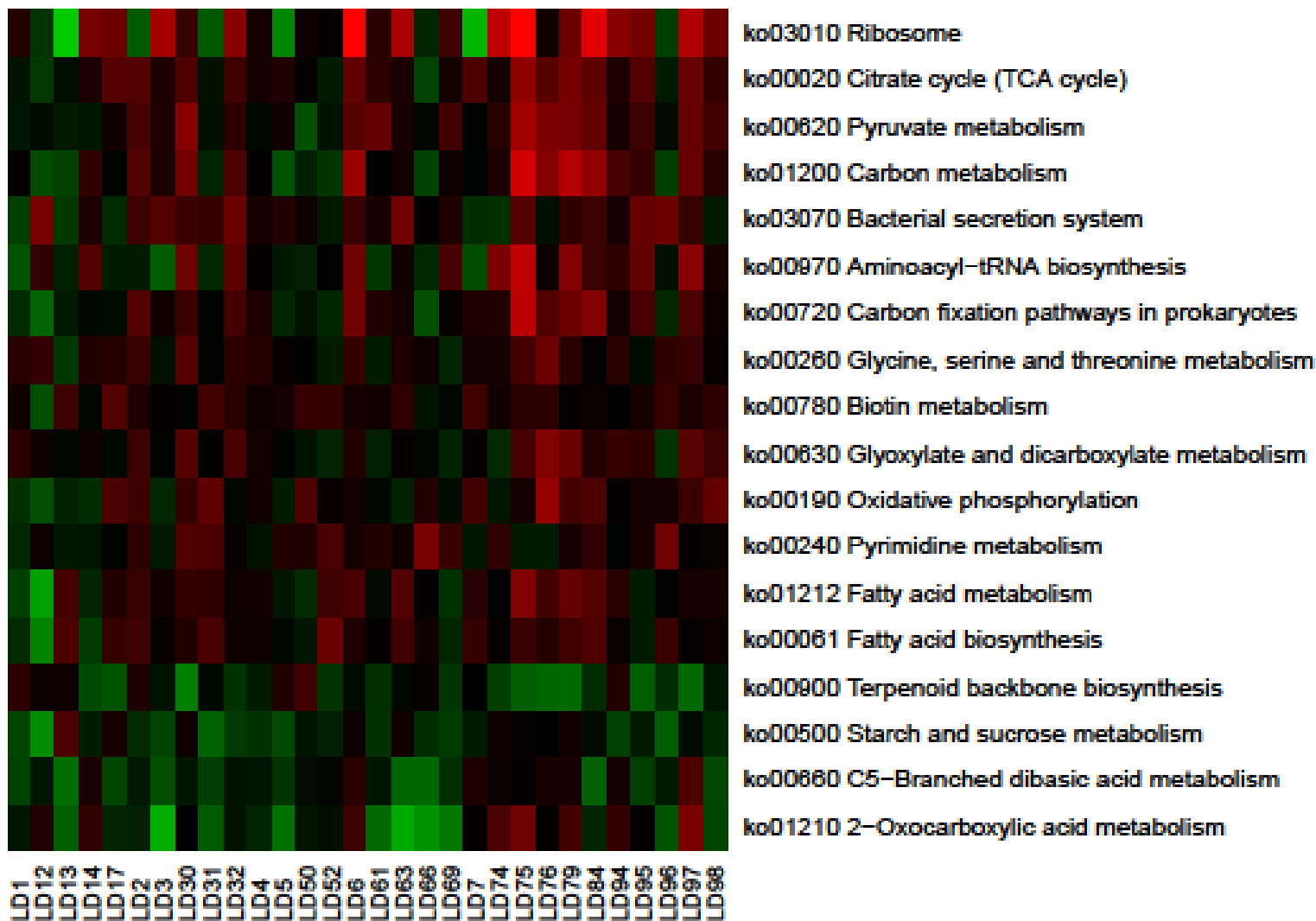
Liver Cirrhosis

- Liver damage – many and diverse causes
- Changes in gut microbiota
- Can we use translational optimization?
 - Identify important genes/functions for disease status
 - Predict and classify individuals based on microbial community sample

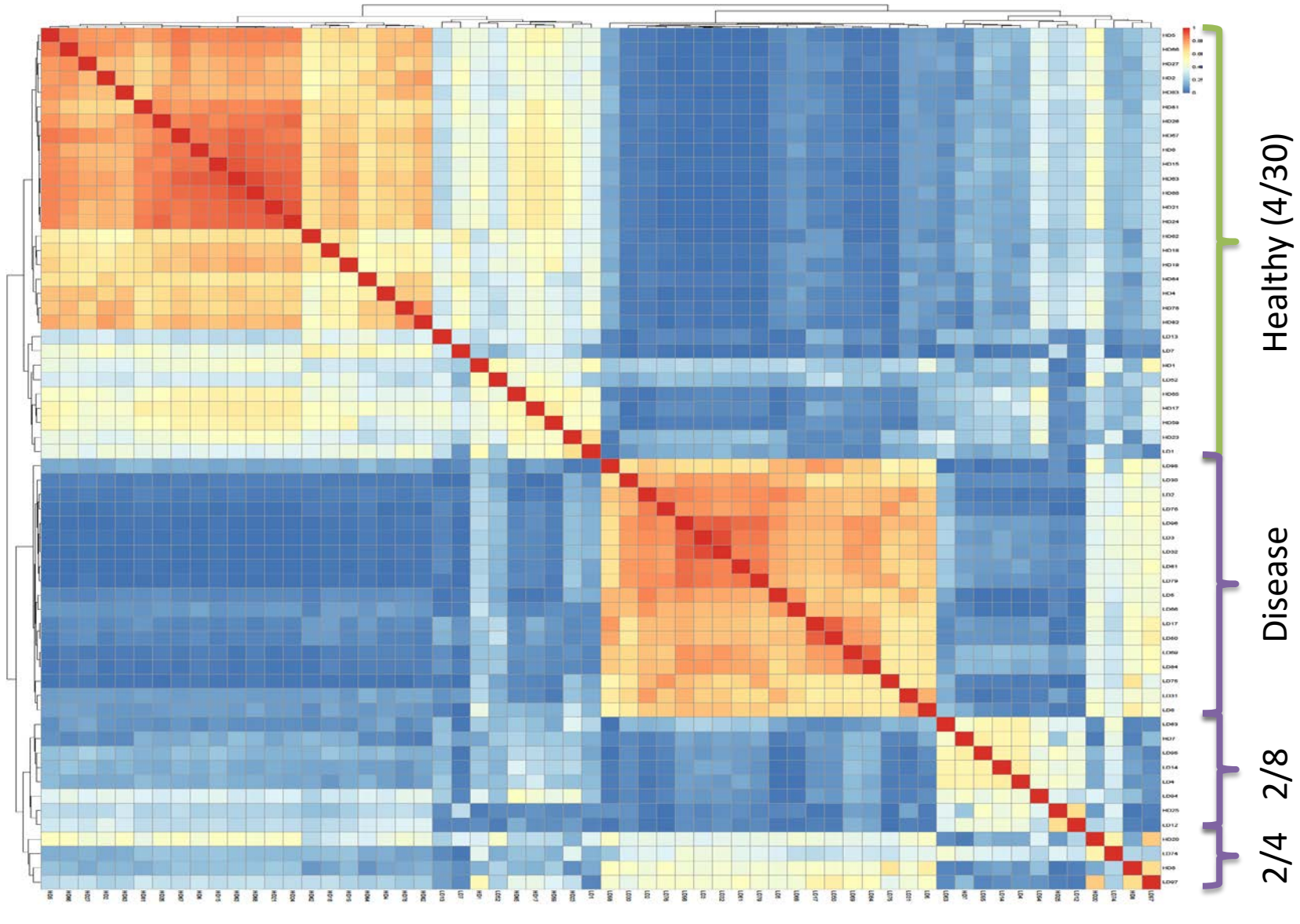
Analysis Workflow



Disease-relevant functions



Disease status classification



Healthy (4/30)

Disease

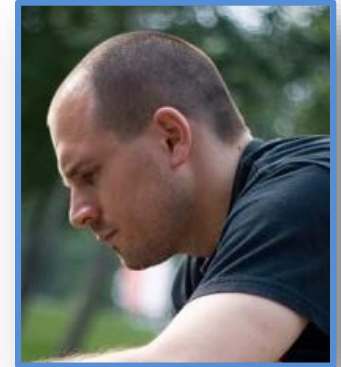
2/4 2/8

Take home messages

- **There are always new angles to look at existing data**
- Microbial communities behave as 'meta'-genomes
 - 'Background' translational optimization of expression
 - Community level optimization of function
 - Relevant clinical biomarker
 - Possible ways to assess horizontal gene transfer

Acknowledgements

- Maja Fabijanic
 - Vedran Lucic
 - Masha Roller
 - Tina Perica
 - Istvan Nagy (Szeged)
- Bioinfo group
- Rosa Karlic, Petar Jager, Ivan Franulovic



EMBO YIP/Cro NSF; ICGEB; Cro MSES; FP6 CompSoLS, FP7 IntegraLife

Our motivation



Thank You!

kristian@bioinfo.hr