TerraGenome: a consortium for the sequencing of the soil metagenome

Timothy M. Vogel

www.GenomEnviron.org

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Scientific Committee on Problems of the Environment

SCOPE

http://www.icsu-scope.org/

↓

SCOPE program on Microbial Environmental Genomics

MicroEnGen III

↓

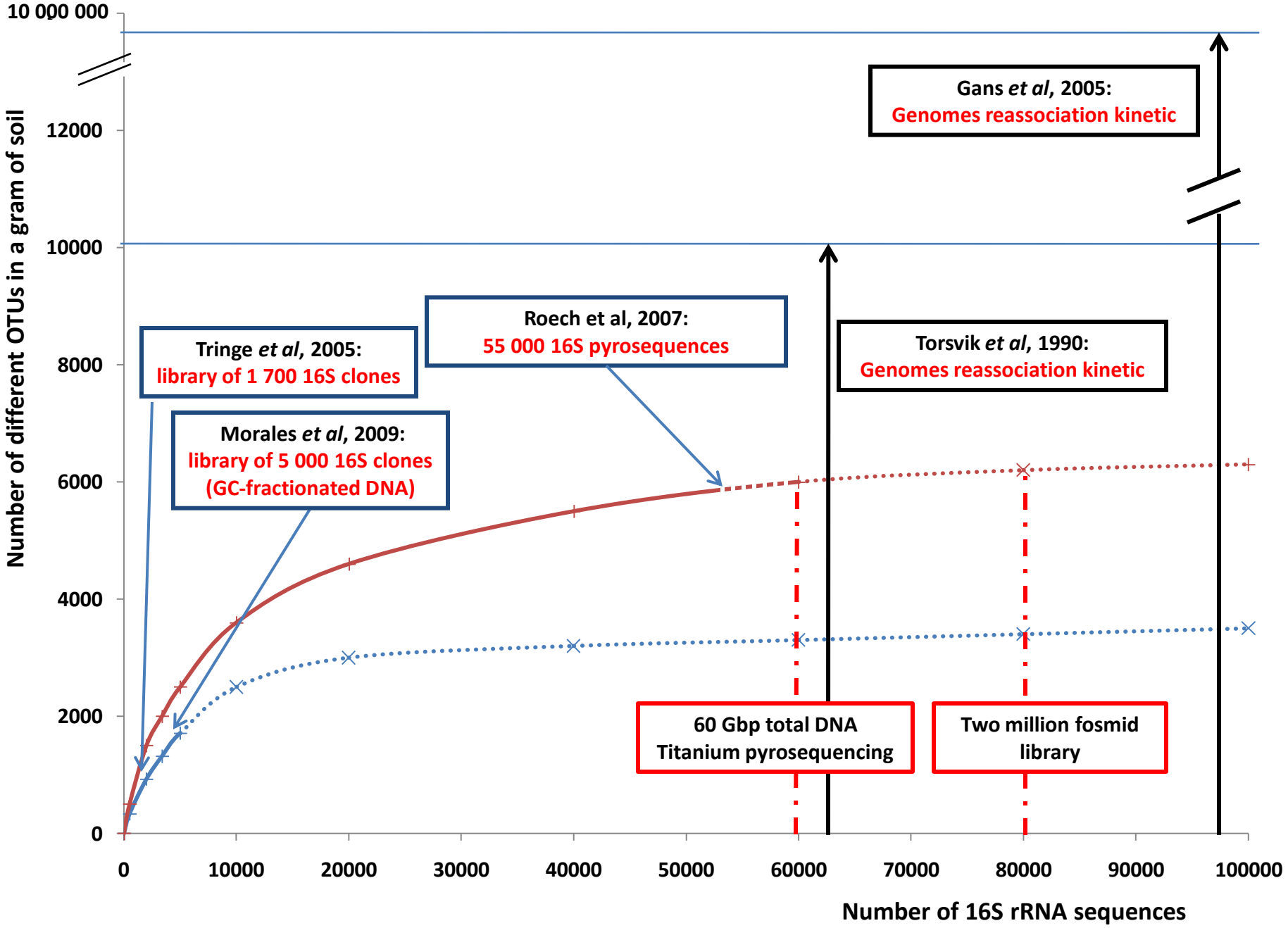
Soil Metagenome International Consortium

METASTED

↓

TERRAGENOME

http://www.terragenome.org/

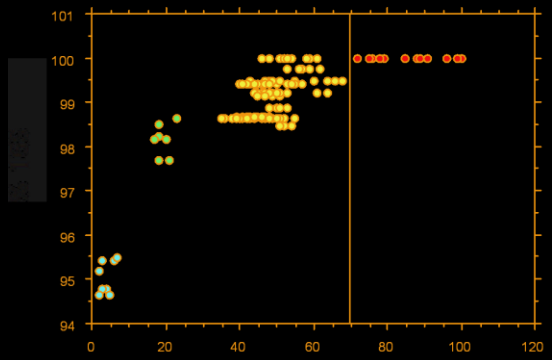

10 000 000
12000
10000
8000
6000
4000
2000
0
Number of different OTUs in a gram of soil
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Number of 16S rRNA sequences
Gans et al, 2005:
Genomes reassociation kinetic
Tringe et al, 2005:
library of 1 700 16S clones
Roech et al, 2007:
55 000 16S pyrosequences
Torsvik et al, 1990:
Genomes reassociation kinetic
Morales et al, 2009:
library of 5 000 16S clones
(GC-fractionated DNA)
60 Gbp total DNA
Titanium pyrosequencing
Two million fosmid
library

TAXONOMIC RESOLUTION

+

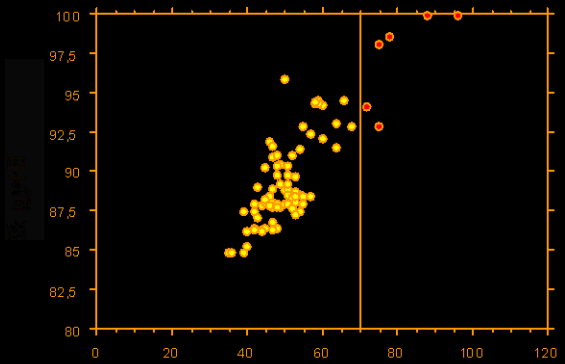
Macro-phylogeny

rrs



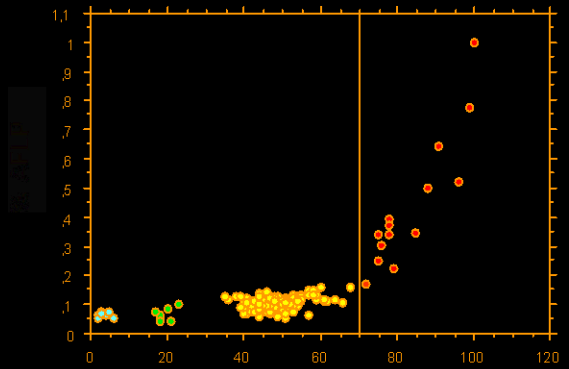
micro-phylogeny

gyrB



Population genetics

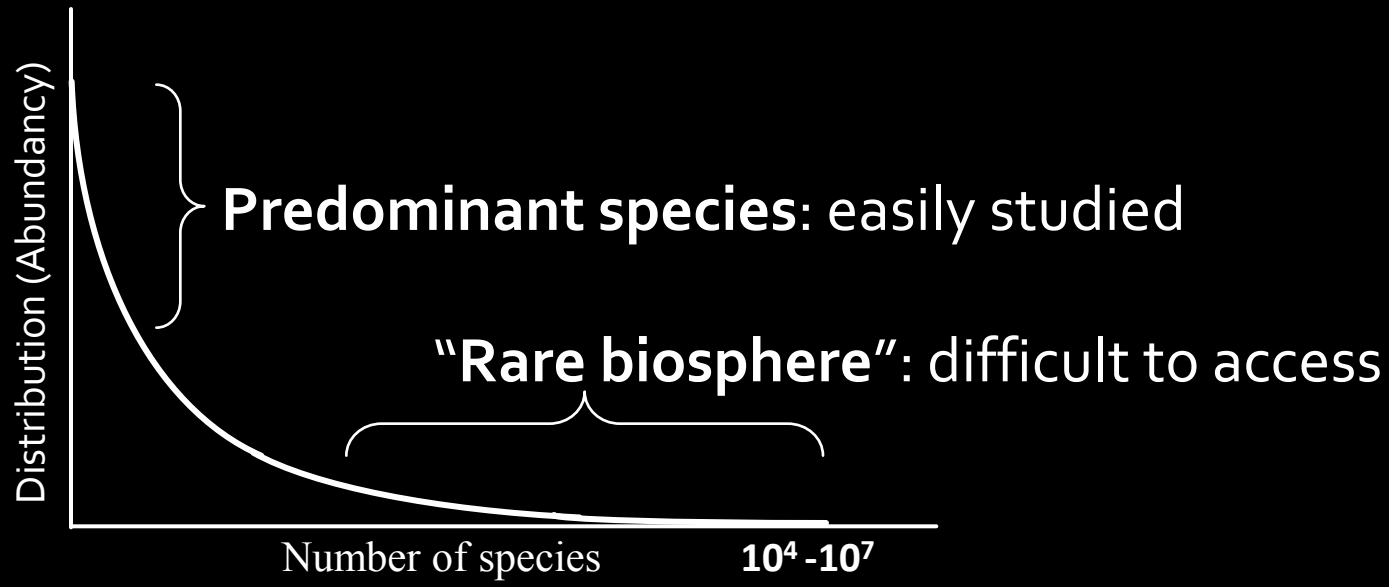
AFLP



- infra-specific
- between closely related species
- infra-generic
- between genera

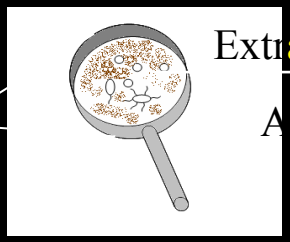
Nesme et al 2003

- **Soil: 10^4 - 10^7 species / gram**
- **Uneven distributed**



- **Terragenome: To sequence and assemble an entire metagenome of soil (Rothamsted, UK).**

Metagenomic approach:

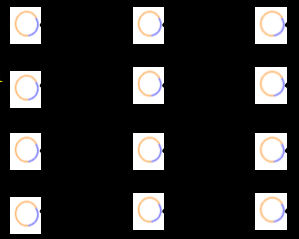


Extraction

ADN

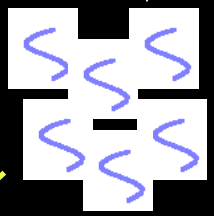
Cloning

Transformation

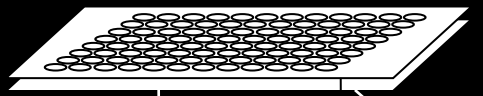


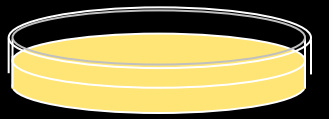
vector

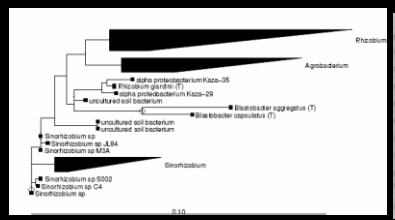
PCR

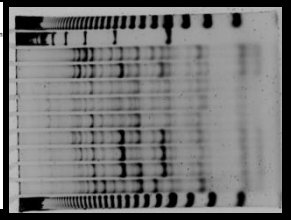


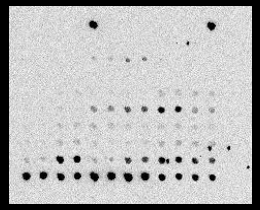
Clone Library

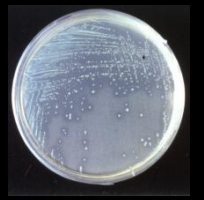












Culture *in vitro*

Cloning and/or sequencing

RISA, T-RFLP DGGE, Phylochip

Molecular screening

Chemical screening

Biological activity

Lombard *et al.*, 2006

The long-term experimental site in the UK: Rothamsted http://www.rothamsted.ac.uk/

- Extensive metadata
- From 50 to 140 years of controlled experiments

A few problems

"Species"/g soil		Number of bp		Number of pyroseq Runs (fosmid clones)
10^4	→	4×10^{10}	→	10^2 (10^6)
10^7	→	4×10^{13}	→	10^5 (10^9)

Total b&a/g soil		Number of bp		Number of runs
10^9	→	4×10^{15}	→	10^7 (10^{11})

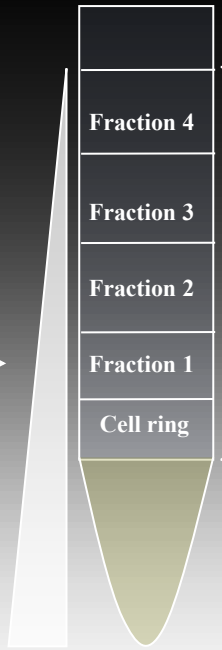
-Need to create strategies to improve soil nucleic diversity accessibility

-Need to find variables that fraction DNA to access parts of the rare biosphere to limit sequencing redondancy

Searching for genomic diversity

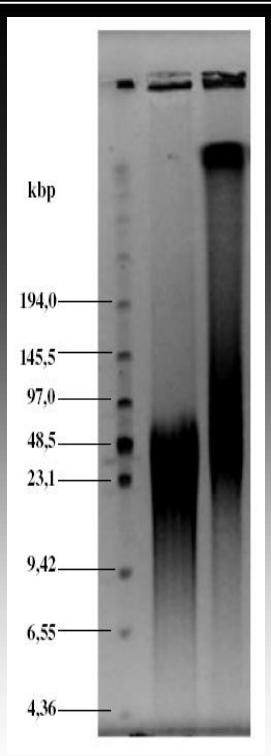
Vertical soil fractionation

Cellular fractionation in a density gradient



Cell lysis stringency

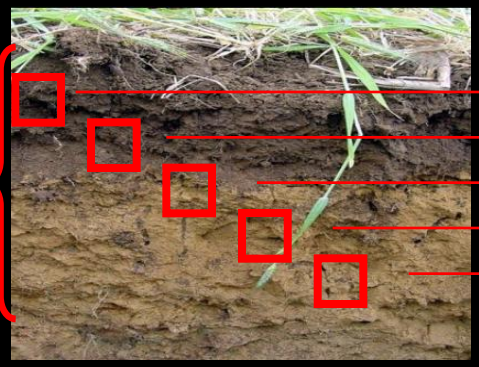
Molecular DNA weight fractionation



Some people think one needs to do composite sampling

Some people sieve soil and technique is everything

Of course this could vary from person to person



MIX

Plot 1

Plot 2

Plot 3

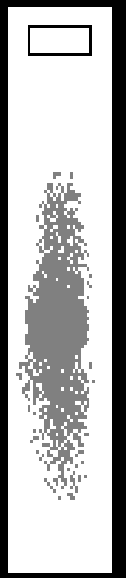
Plot 4

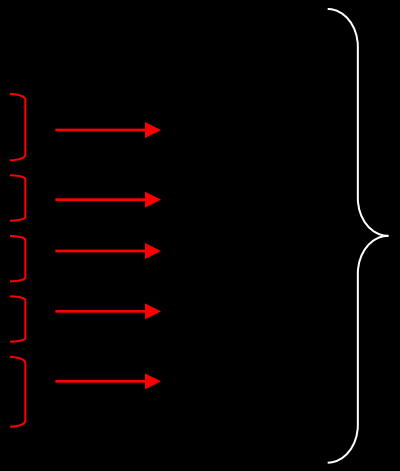
Plot 5

1 EXTRACTION

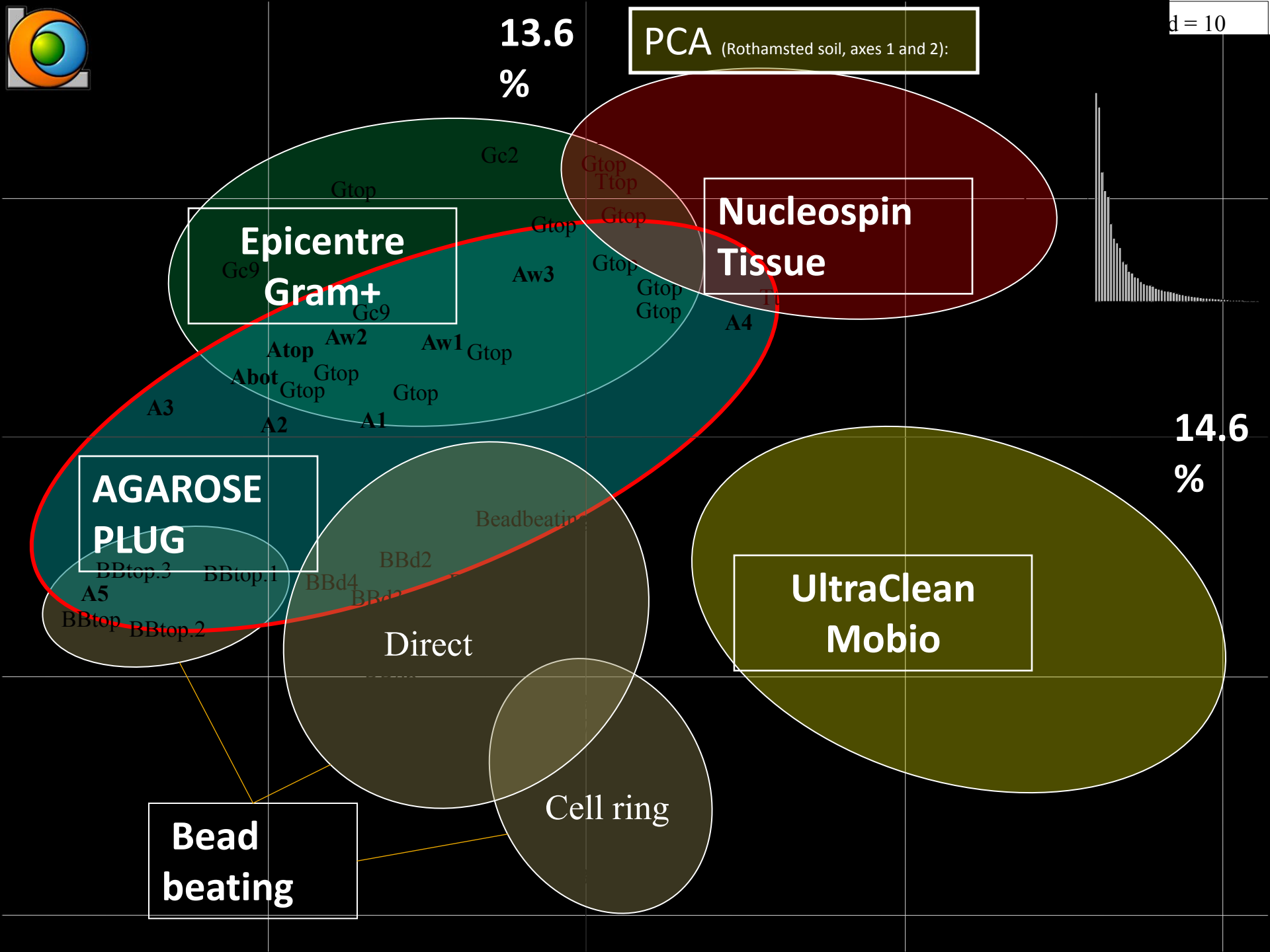
5 EXTRACTIONS

PFGE:

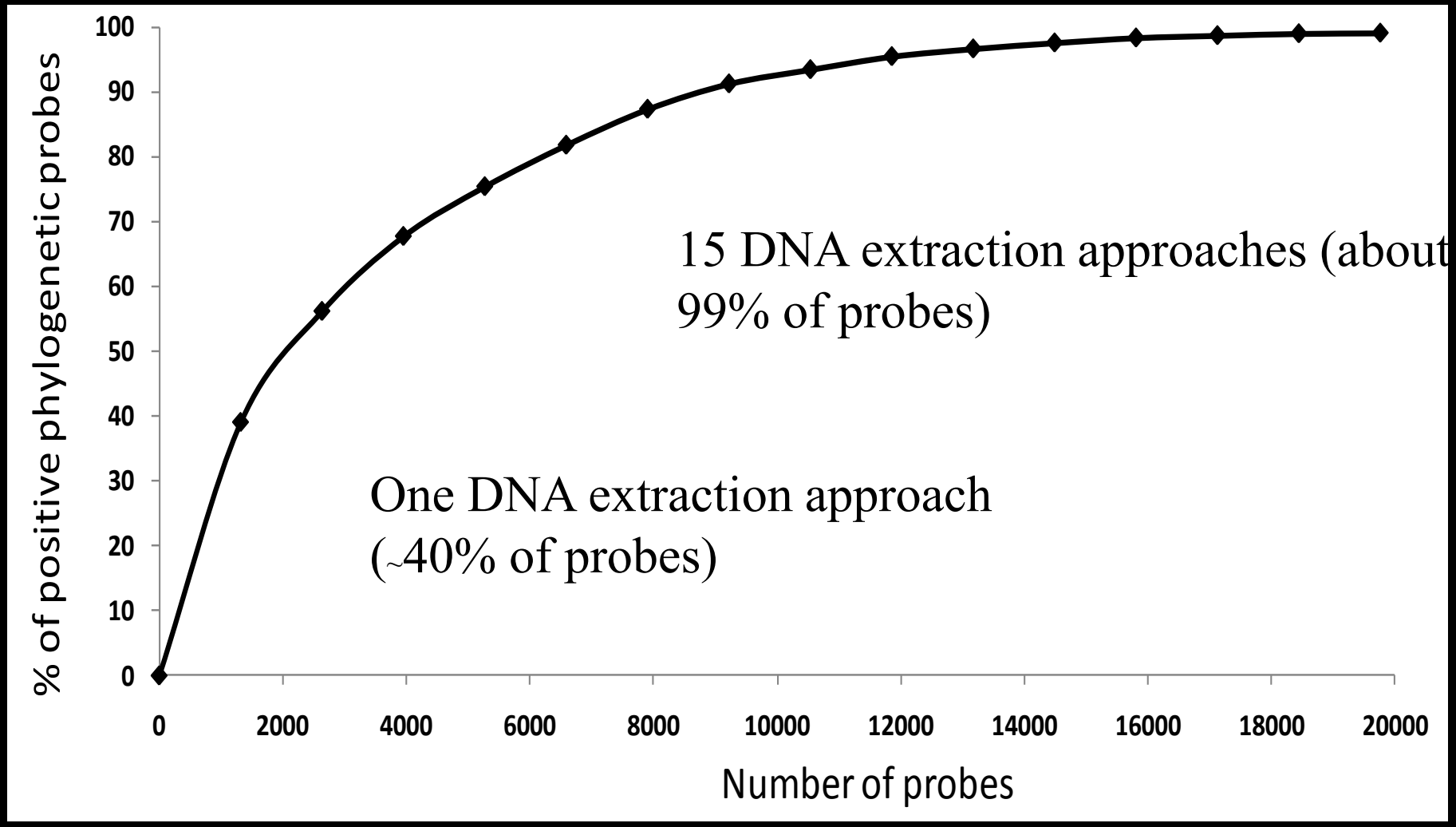




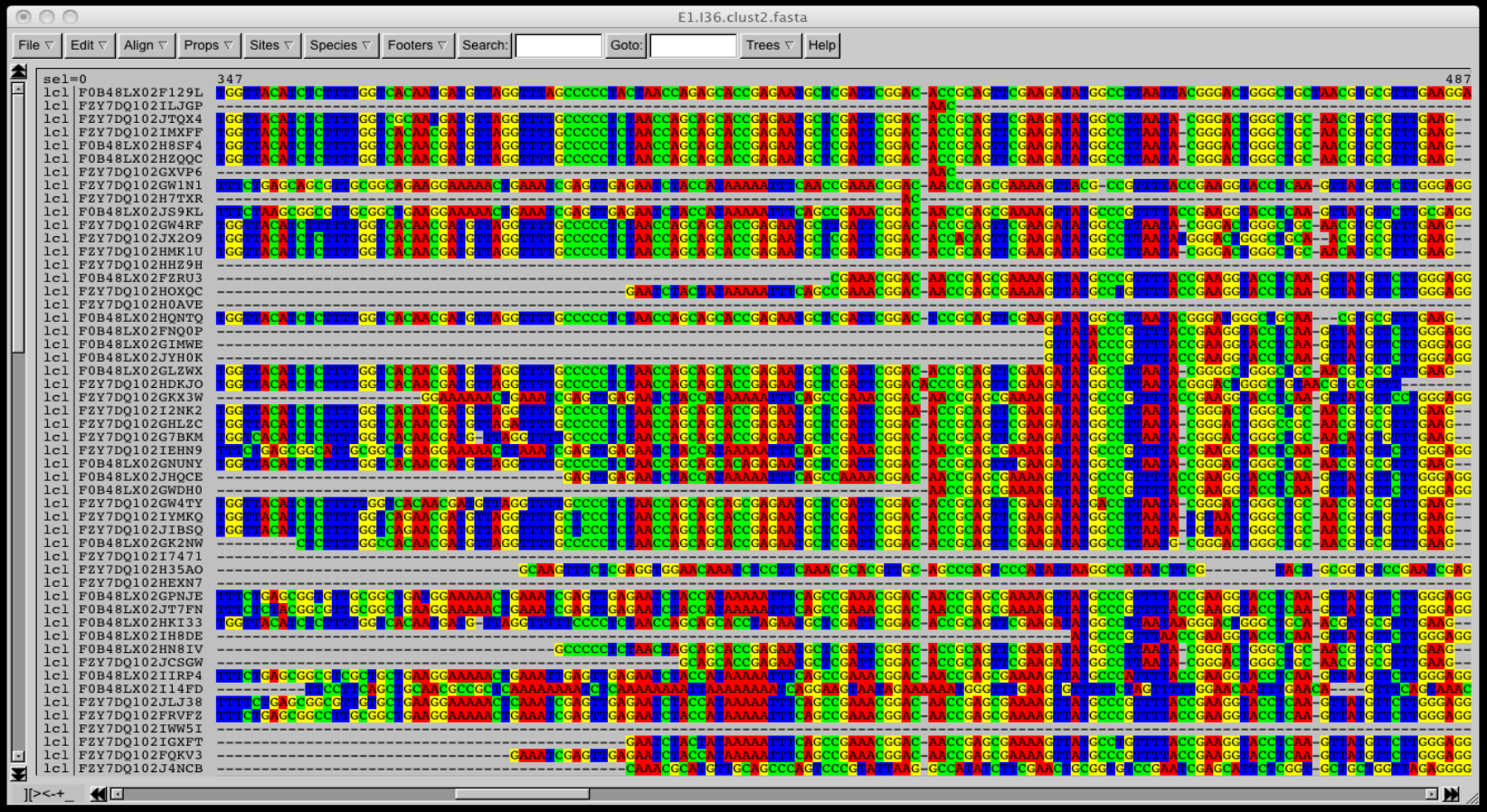
Diversity differences and distribution

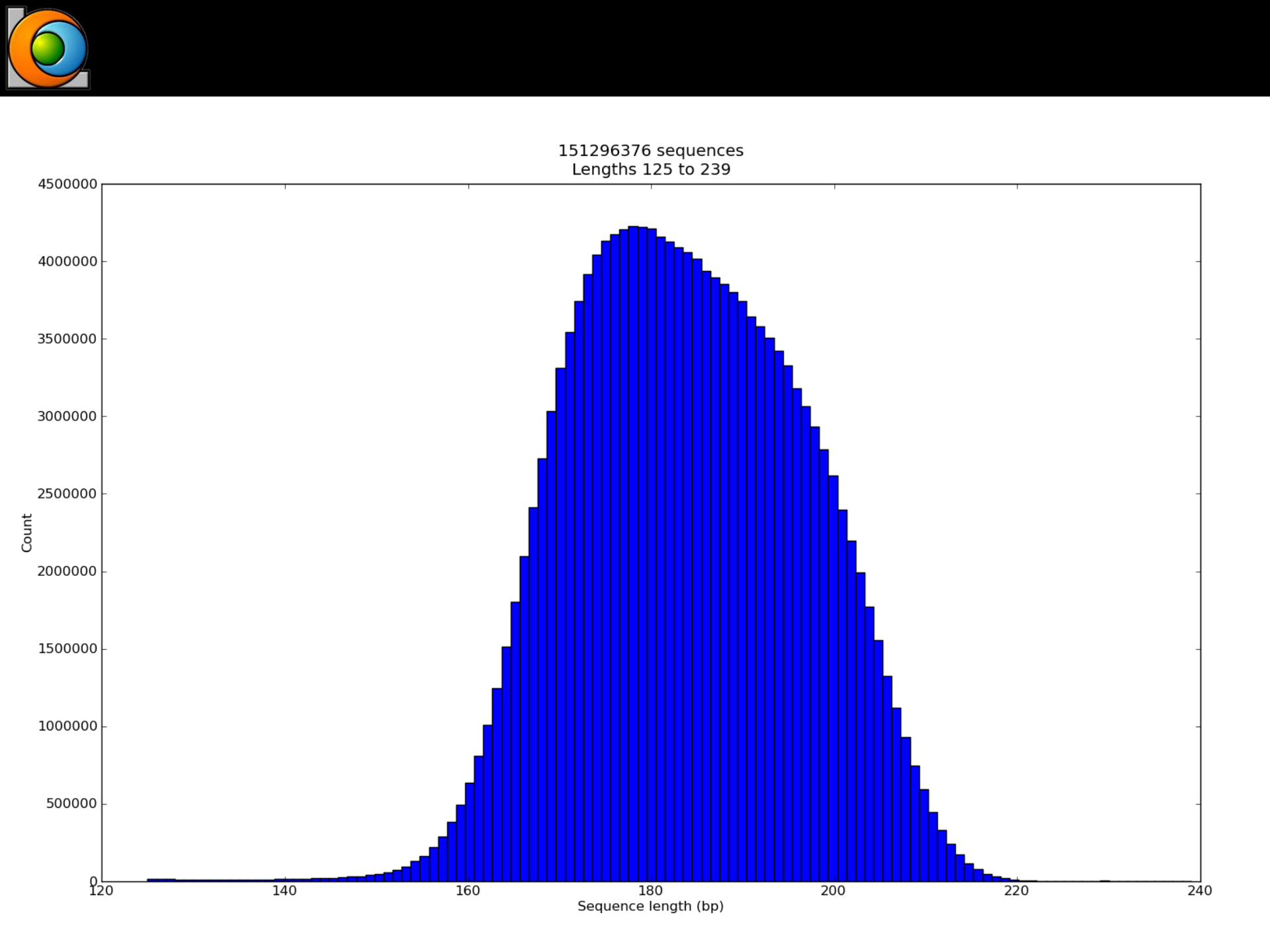

13.6 %
PCA (Rothamsted soil, axes 1 and 2):
d = 10
Gc2
Gtop
Ttop
Gtop
Gtop
Gt

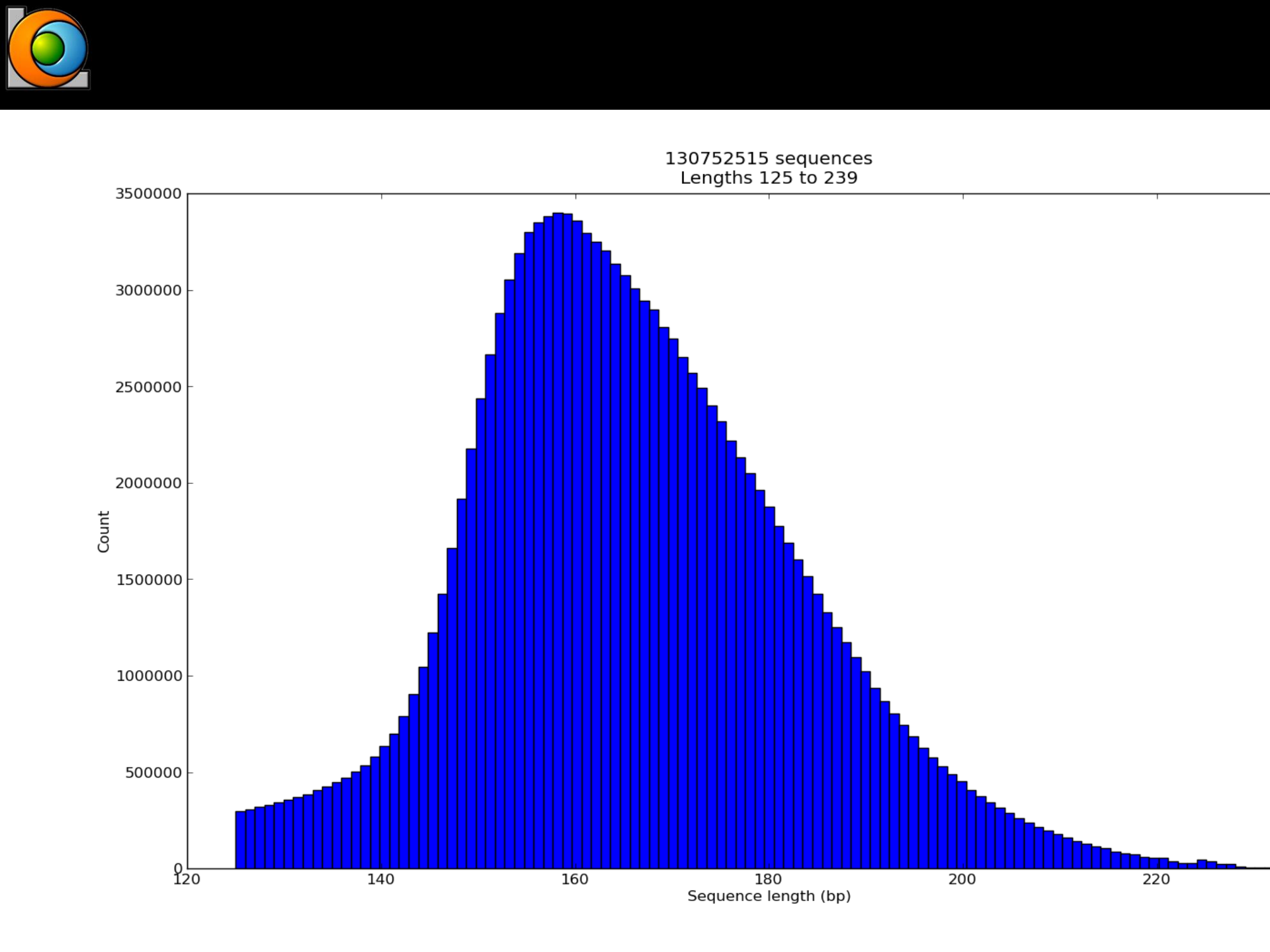
Rothamsted soil phylochip saturation curve



Pyrosequencing (10^6 sequences 400pb)




151296376 sequences
Lengths 125 to 239
Count
Sequence length (bp)
0
500000
1000000
1500000
2000000
2500000
3000000
3500000
4000000
4500000
120
140
160
180
200
220
240


130752515 sequences
Lengths 125 to 239
Count
Sequence length (bp)
0
500000
1000000
1500000
2000000
2500000
3000000
3500000
120
140
160
180
200
220

Jobs Overview

The overview below list all genomes currently processed and the progress on the annotation. To get a more detailed report on an annotation job, plea on the progress bar graphic in the overview.

In case of questions or problems using this service, please contact: mg-rast@mcs.anl.gov.

Progress bar color key:

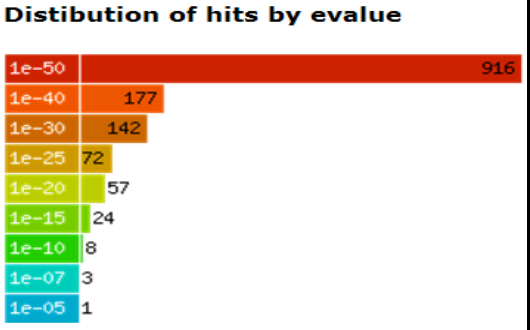
not started
queued for computation
in progress
load in progress
requires user input
failed with an error
successfully completed

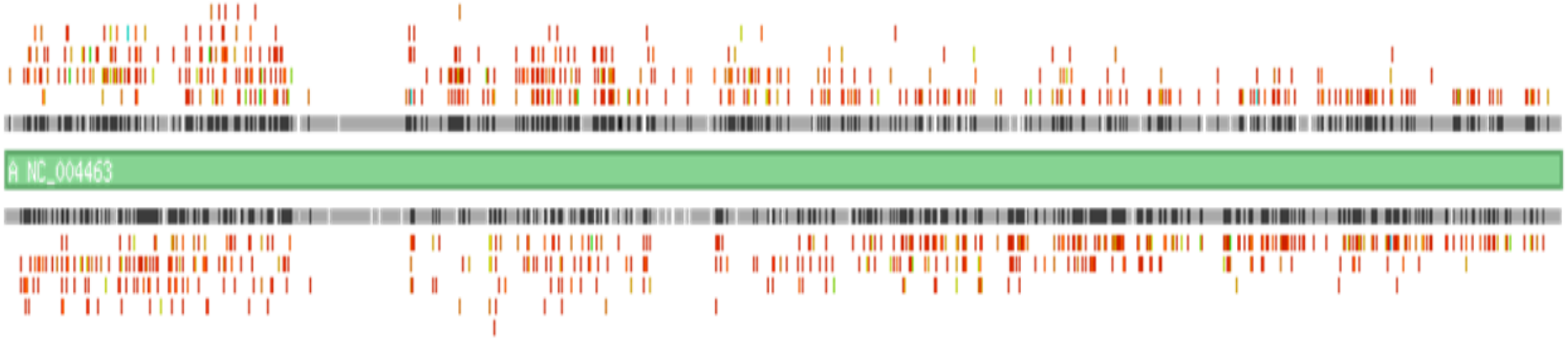
Jobs you have access to :

You currently have access to 568 public jobs.

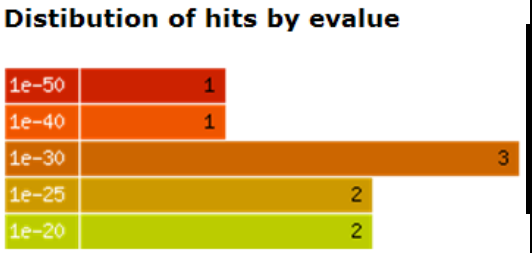
Job	Owner	ID	Name	Size (bp)	Creation Date	Annotation Progress
12857	delmont, tom		E4B-454	0	2010-12-06	view details
12846	delmont, tom	4453247.3	E4A	470523038	2010-12-03	view details
12845	delmont, tom	4453246.3	E1-454	358393415	2010-12-03	view details
12844	delmont, tom	4453245.3	Integrons 78 met	2185452	2010-12-03	view details
11850	delmont, tom	4452163.3	F64	408531491	2010-10-27	view details
11849	delmont, tom	4452162.3	N57	478979896	2010-10-27	view details
11848	delmont, tom	4452161.3	N56	422554969	2010-10-27	

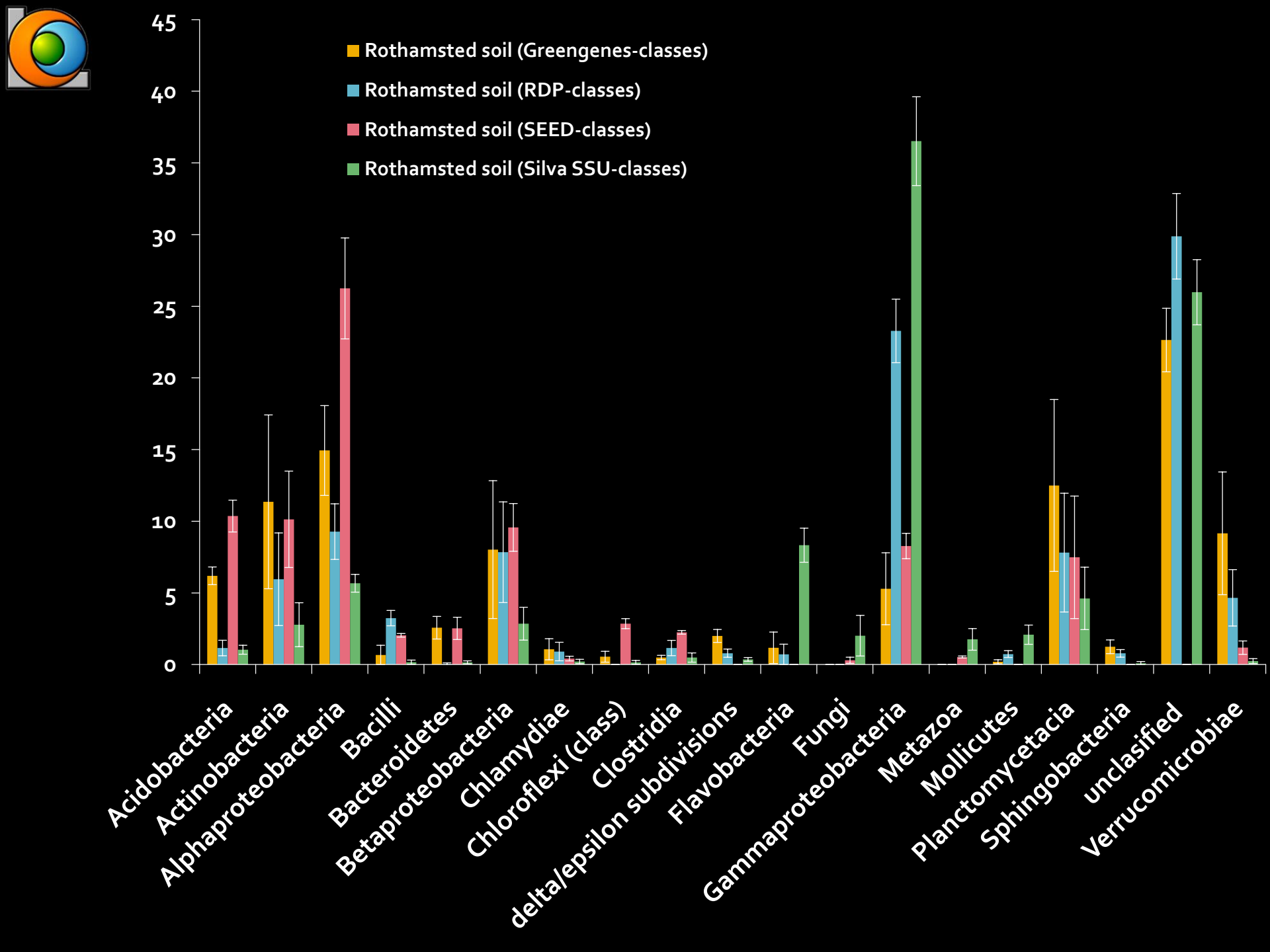
Bradyrhizobium japonicum USDA 110



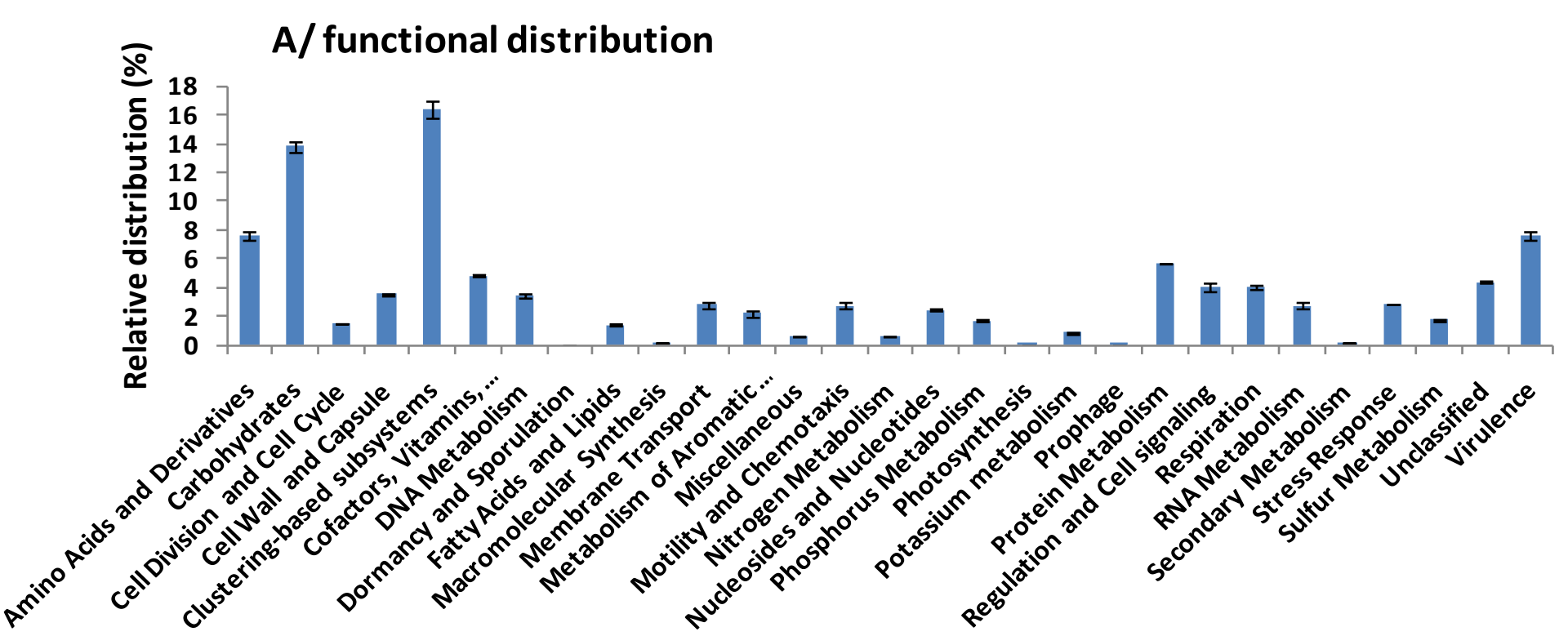


Blastopirellula marina DSM 3645




Rothamsted soil (Greengenes-classes)
Rothamsted soil (RDP-classes)
Rothamsted soil (SEED-classes)
Rothamsted soil (Silva SSU-classes)
45
40
35
30
25
20
15
10
5
0
Acidobacteria
Actinobacteria
Alphaproteobacteria
Bacilli
Bacteroidetes
Betaproteobacteria
Chlamydiae
Chloroflexi (class)
Clostridia
delta/epsilon subdivisions
Flavobacteria
Fungi
Gammaproteobacteria
Metazoa
Mollicutes
Planctomycetacia
Sphingobacteria
unclasified
Verrucomicrobiae

Functional Subsystem Distribution



Is Everything Everywhere?

{What is everything? Species, genes, function, sequences}
{ Where is everywhere? 1 g; 1kg; 1 ton of soil }

Is Everything Everywhere? Yes:

Soil community "adaptation" is growth or gene rearrangement and/or exchange

Is Everything Everywhere? No:

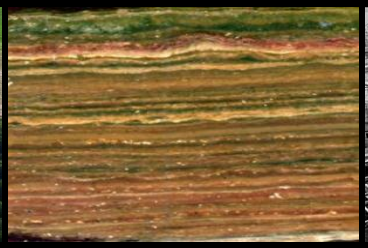
"Adaptation is *de novo* gene synthesis w/ or w/o gene exchange (Or microbial movement)

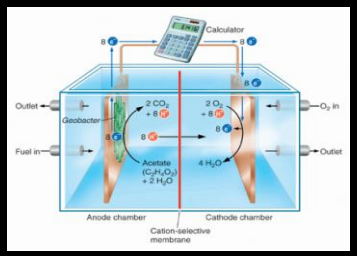
Soil compared to other « environments »

Oceans
Coral atolls
Deep oceans
Antarctic lakes
Arctic snows

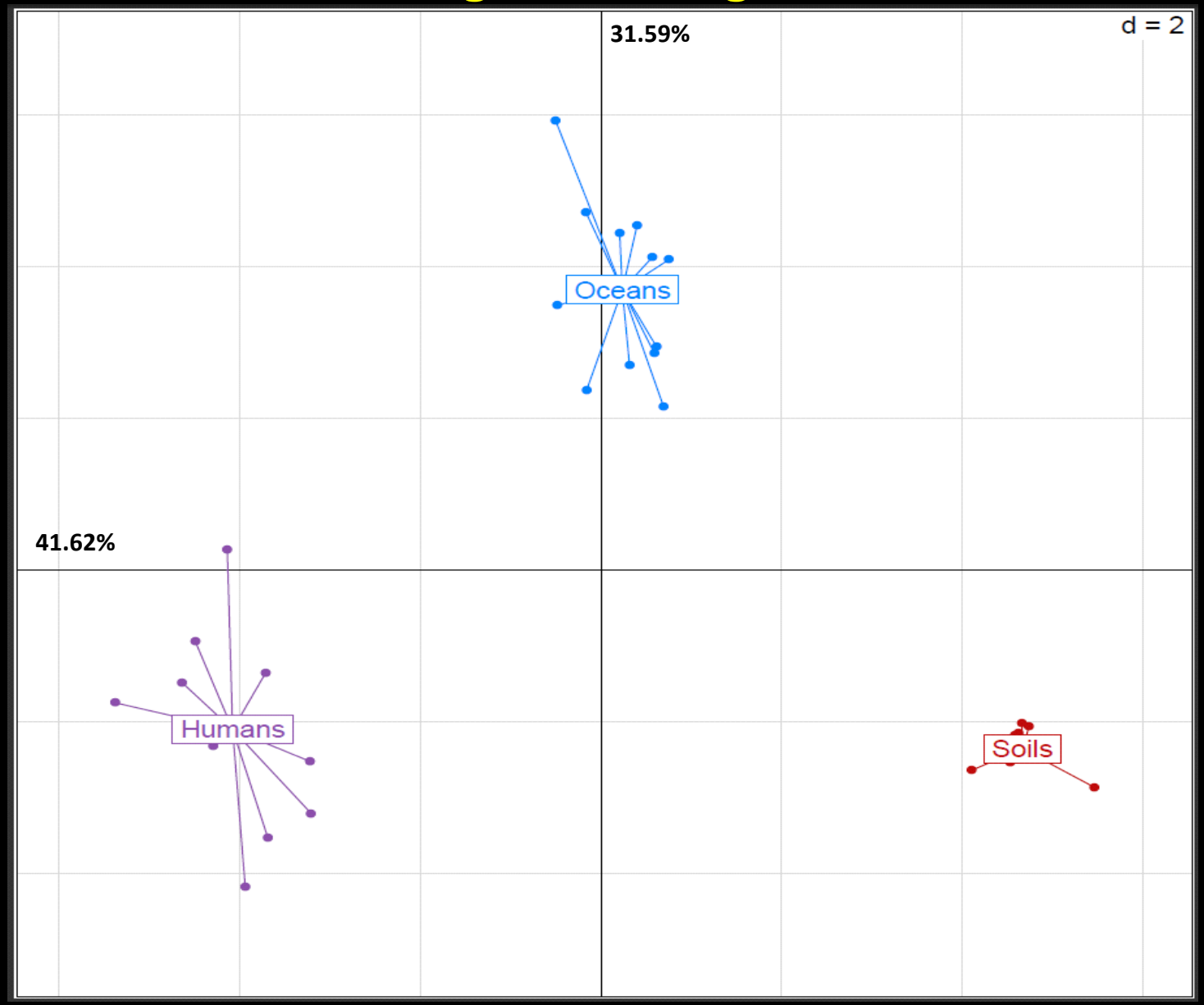
Soils
Hypersaline sediments
Sludges
Microbial fuel cells
Acid mine drainage biofilms

Polluted air
Human feces
Chicken ceacum
Mouse ceacum
Cow rumen

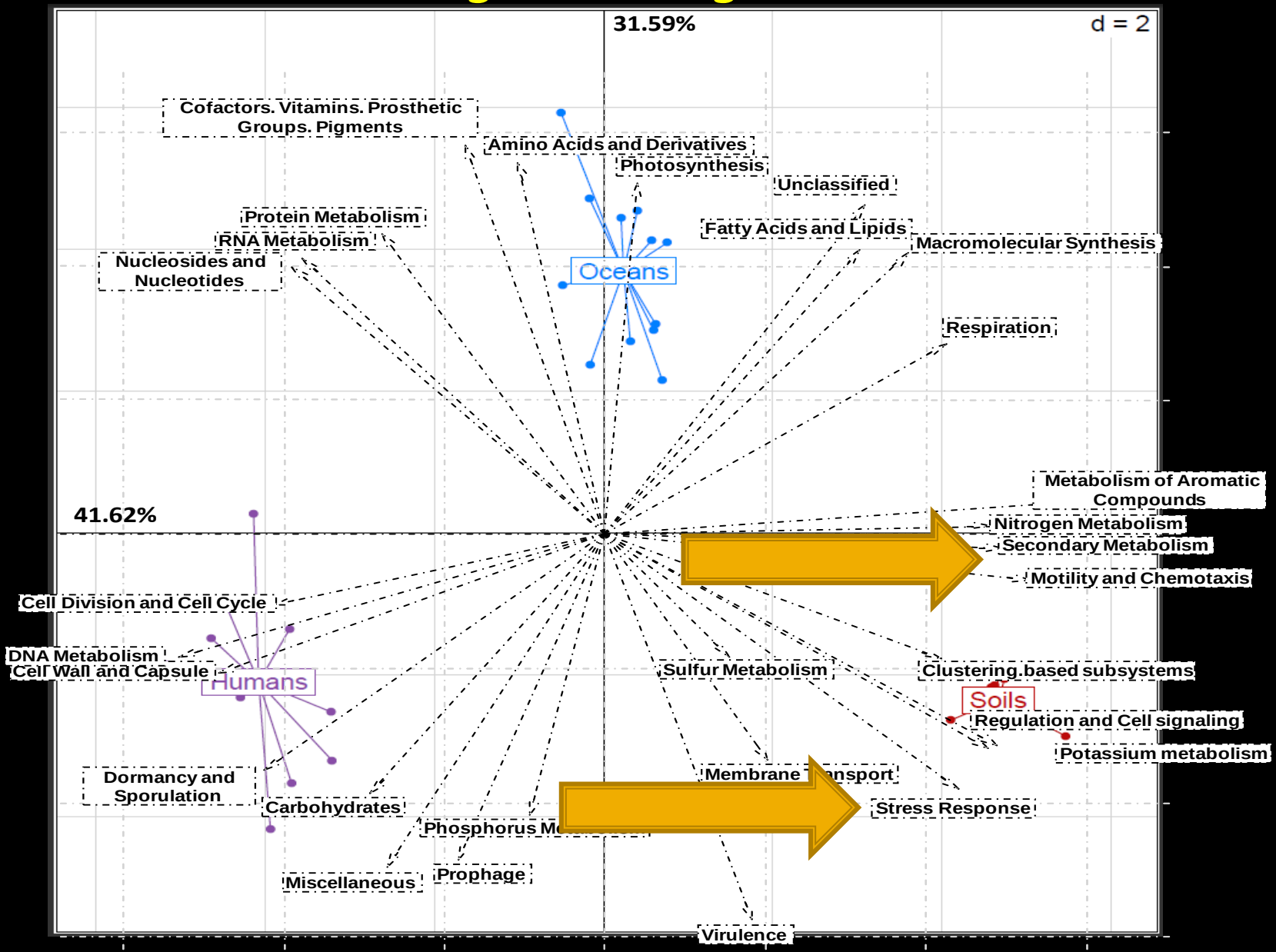




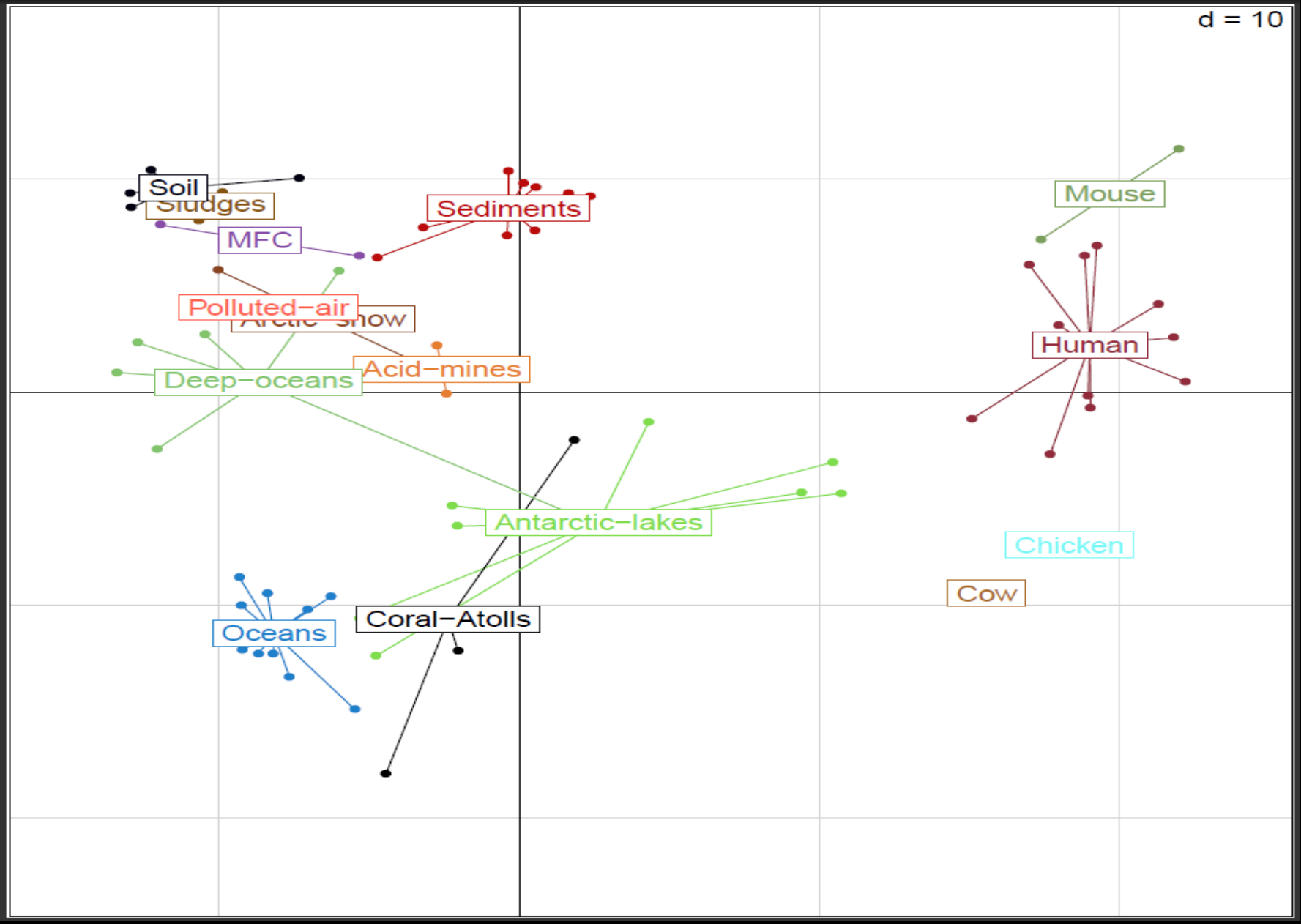
PCA comparing general functional subsystems distributions among 32 metagenomes



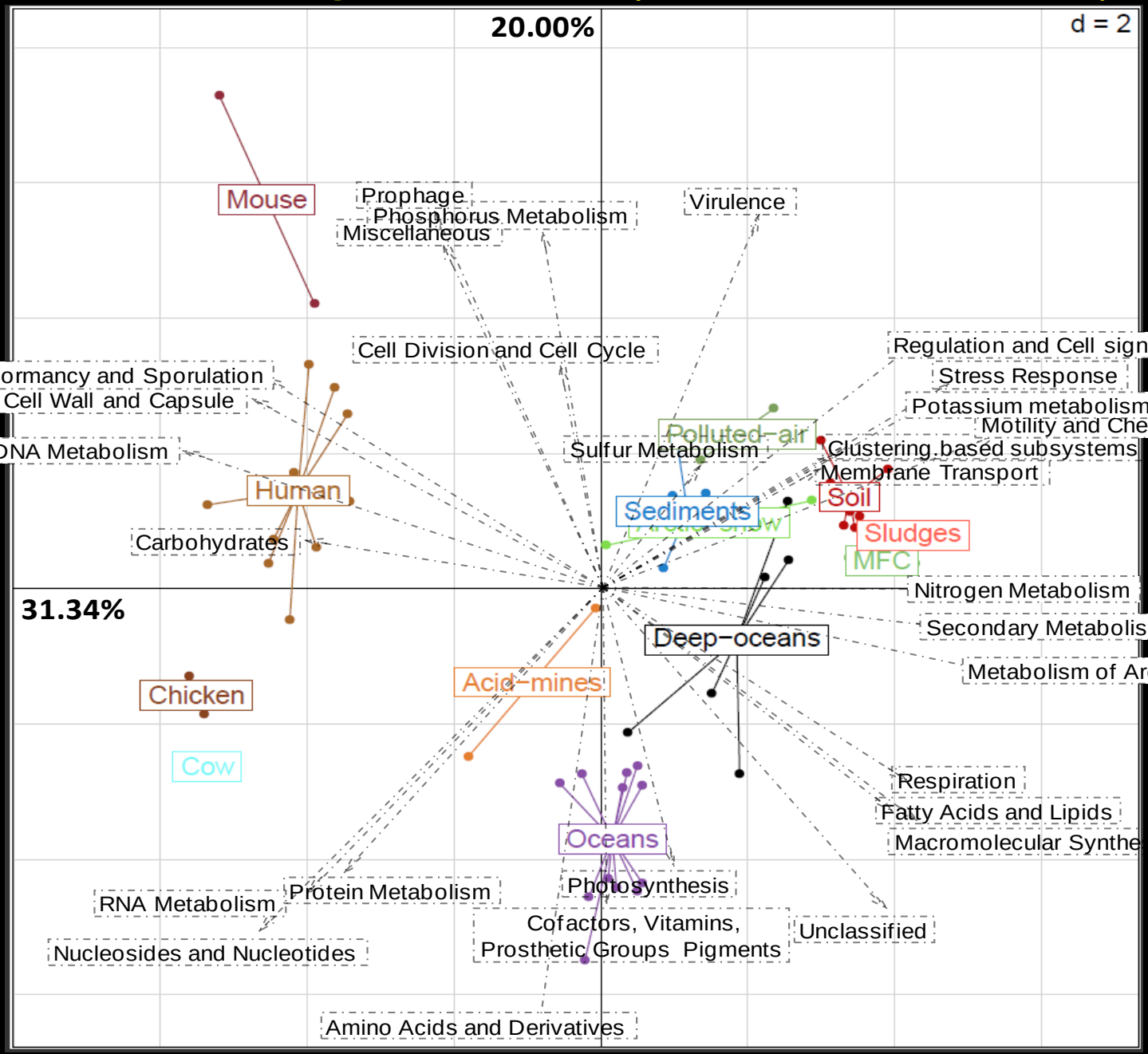
PCA comparing general functional subsystems distributions among 32 metagenomes



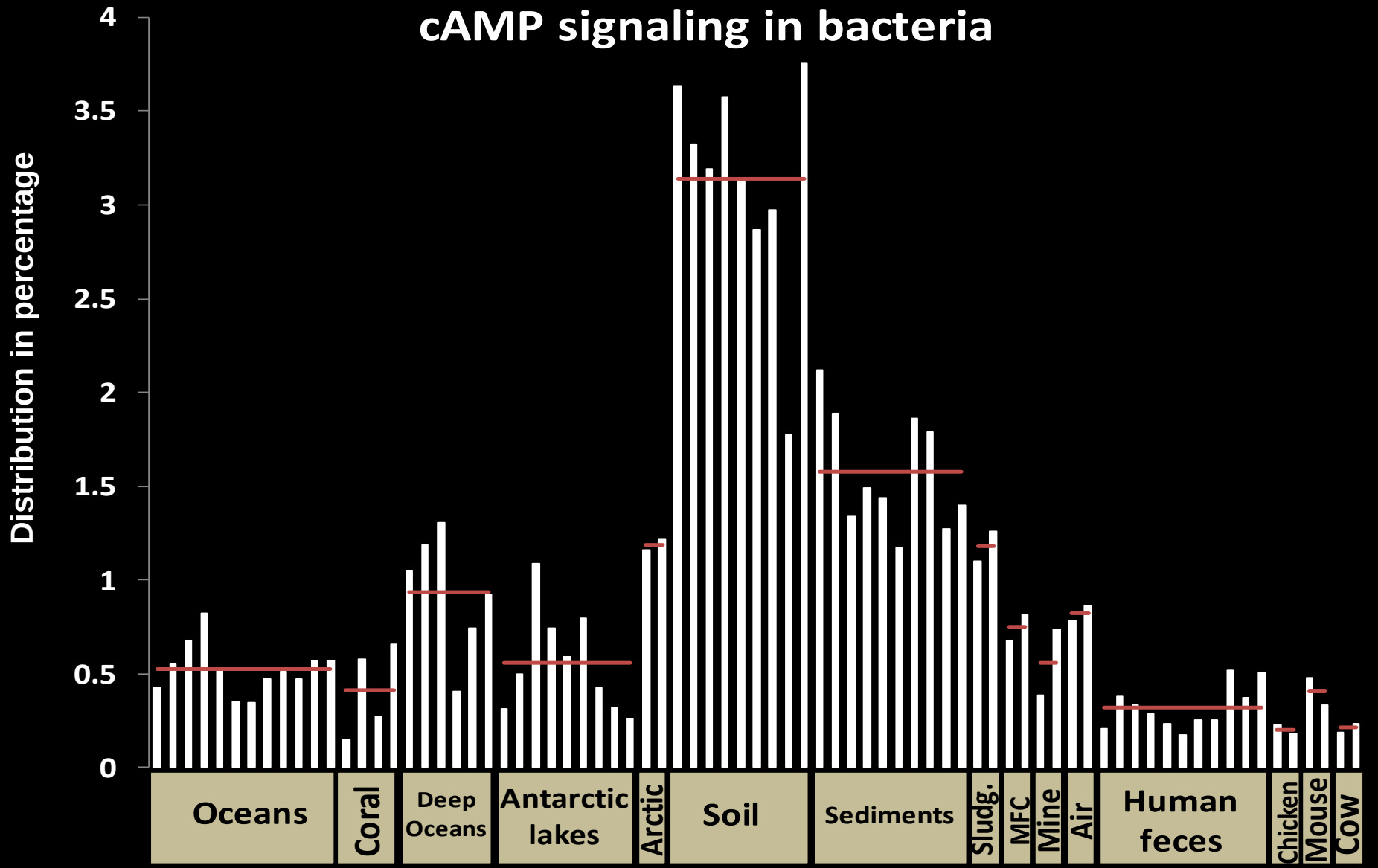
Global metagenomic comparison of 15 ecosystems



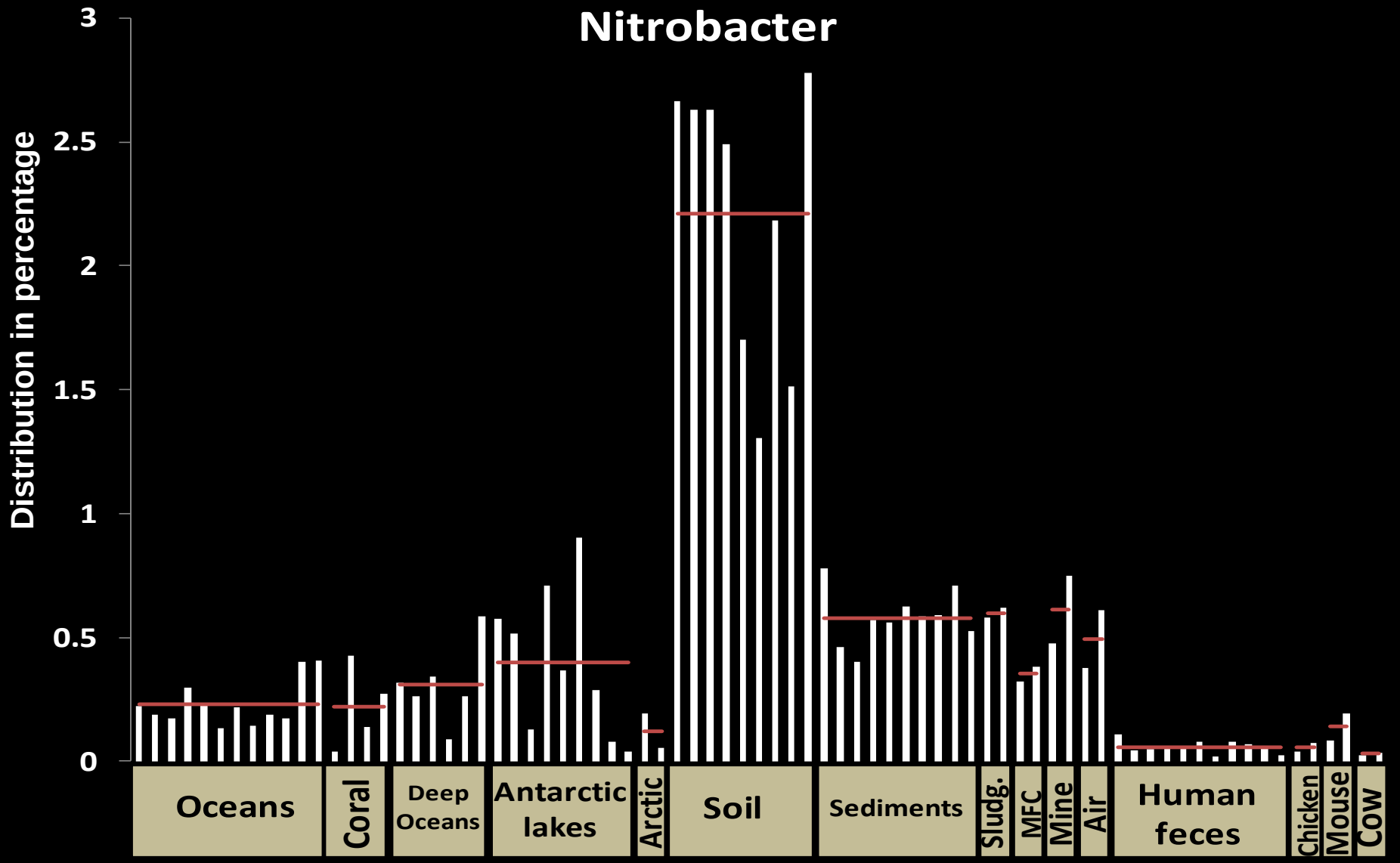
Global metagenomic comparison of 15 ecosystems

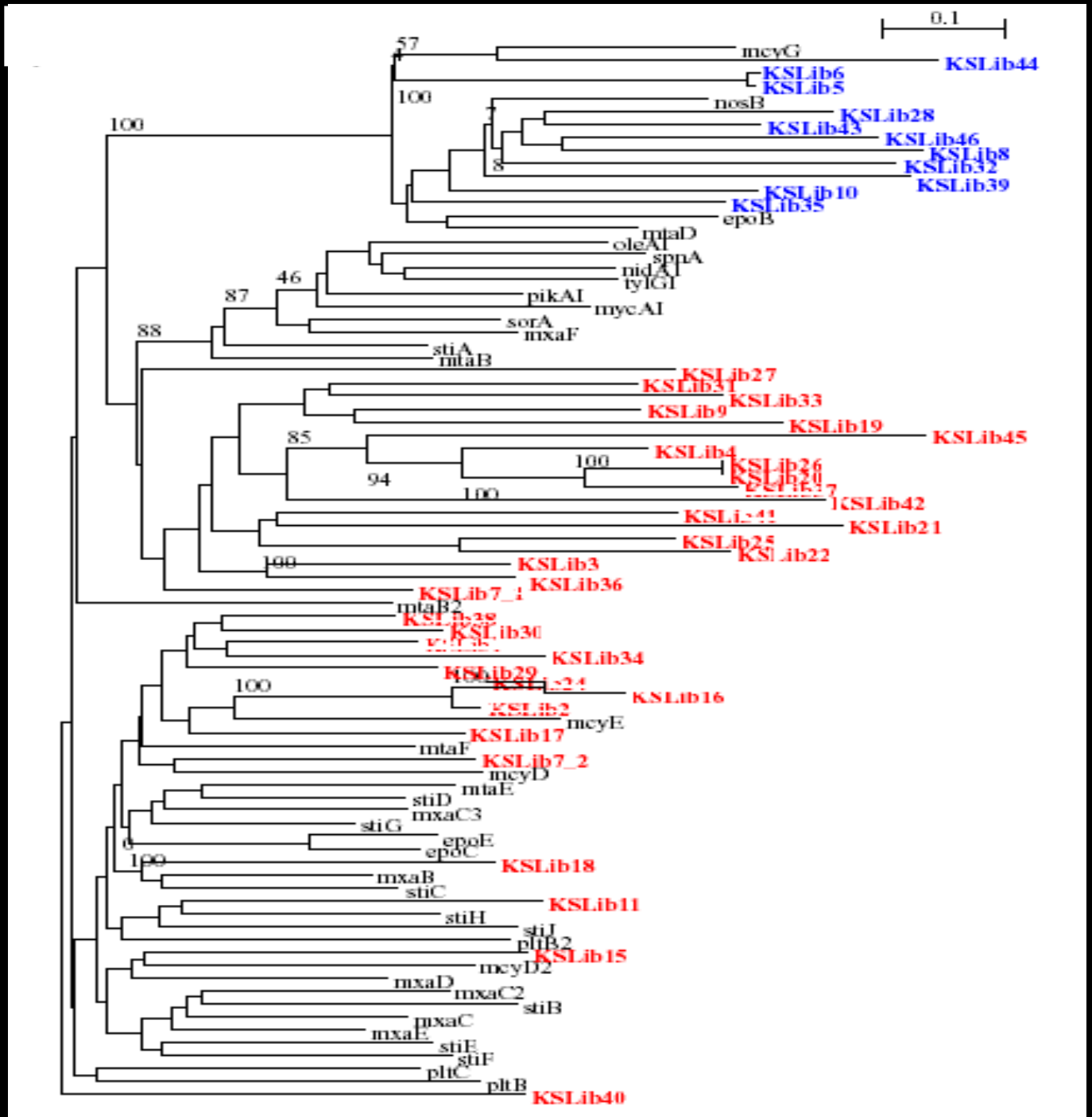


Functional distribution dominance in one ecosystem



Genus distribution dominance in one ecosystem

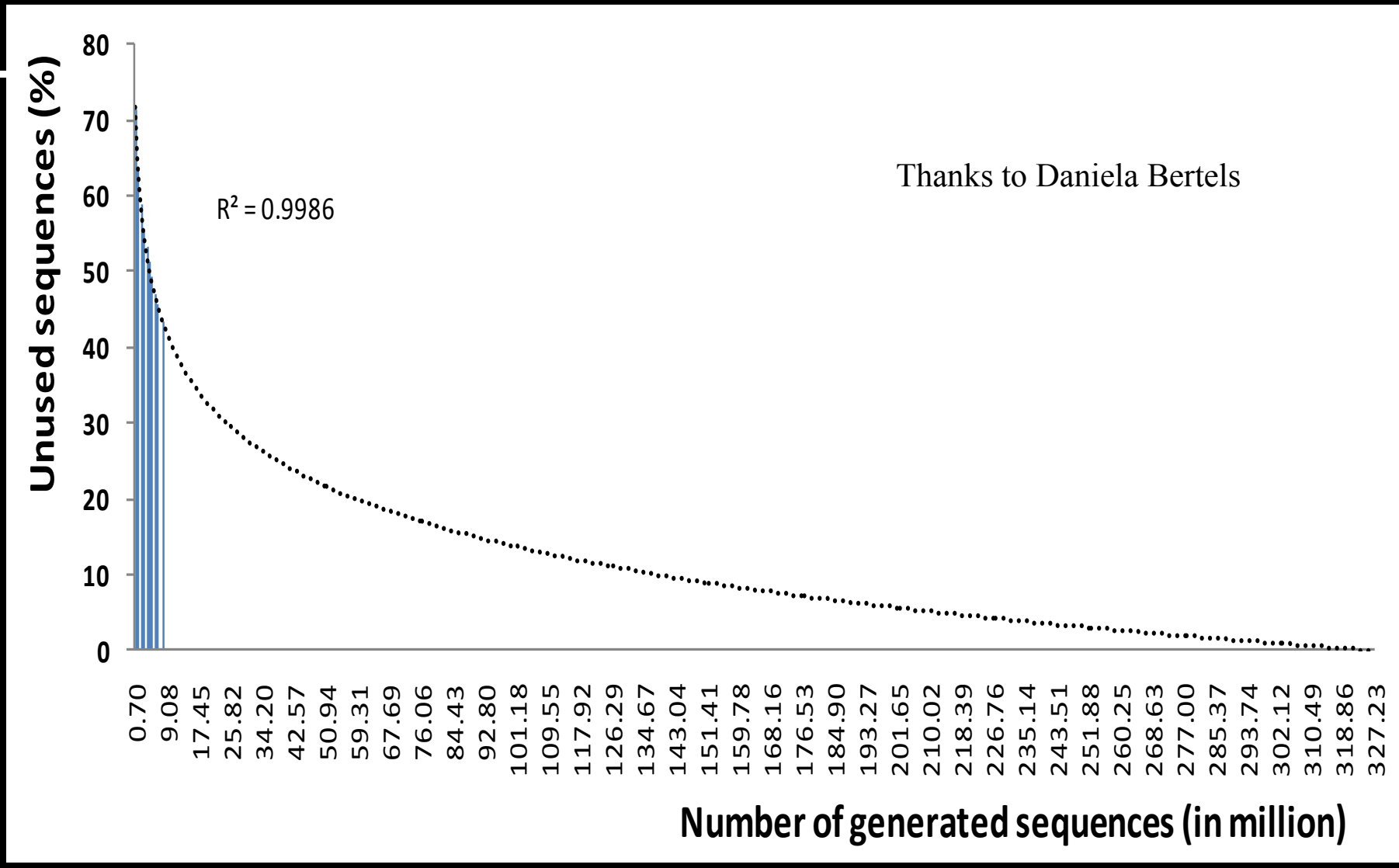




TREE vs Biodiversity

Ginolhac et al

Assembly vs Diversity



CONCLUSIONS

Soil Metagenomics provides access to microbial biodiversity

1. Sequencing helps uncover hidden diversity
    - Sequence interpretation is heavily dependent on genome sequences in database – need more genomes sequenced.

2. Clone librairies provide contiguous sequences
    - Long term possibility of genomic assembly.

3. Definition of sequence and genomic variance affects biodiversity estimates.

Environmental Microbial Genomics Group

www.genomenviron.org