

From ecosystems to genes: metagenomics

Claudia Acquisti

Evolutionary Functional Genomics
Institute for Evolution and Biodiversity

Claudia.Acquisti@uni-muenster.de

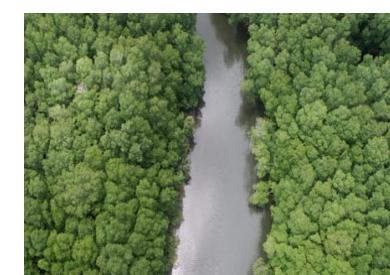
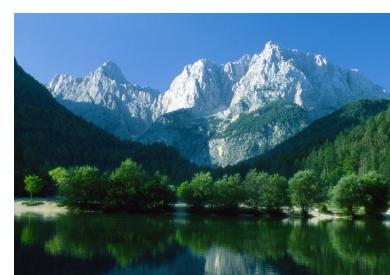
Datum	Thema	Referent
10.12.2015	Einführung & Enzyme der Gentechnik	Antje von Schaewen
17.12.2015	Methoden der Gentechnik	Maik Böhmer
07.01.2016	Frei	
14.01.2016	Vektoren & Klonierung	Antje von Schaewen
21.01.2016	Reportergene	Oliver Batistic
28.01.2016	Transgene Organismen	Stefan Weinl
04.02.2016	Genomik und moderne Methoden	Iris Finkemeier
11.02.2016	Proteine	Iris Finkemeier

Metagenomics

WHAT?
WHY?
HOW?



A door to the molecular understanding of biodiversity



Metagenomics = μετα + genomics

“A NEW TOOL, AND A WAY OF THINKING”

Jo Handelsman

1998 Handelsman, Clardy, Goodman, 1998, Chemistry Biology 5:245–249."

Friederich Miescher



1869
**Discovery
of DNA
(nuclein)**

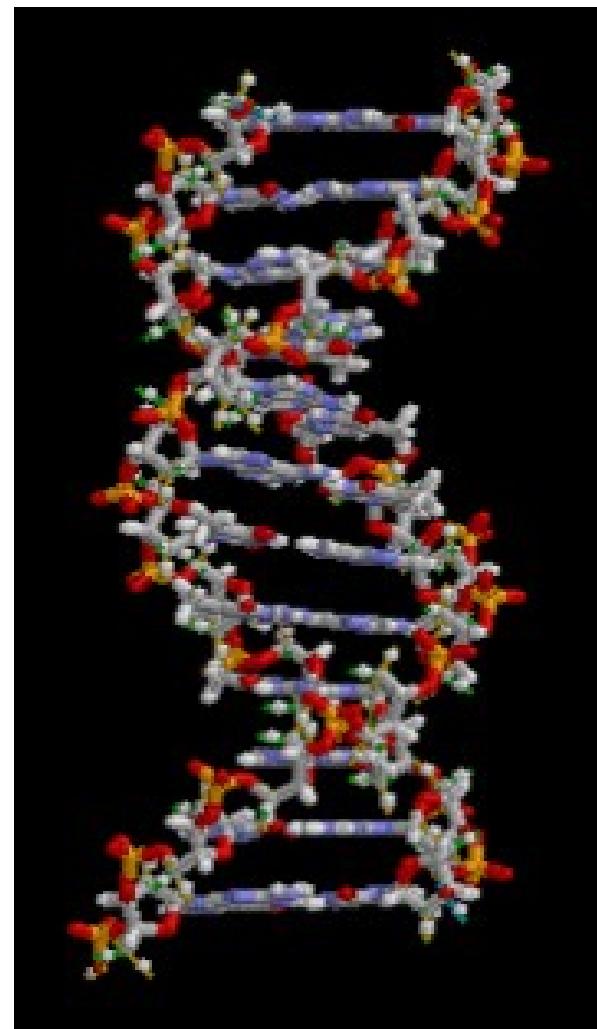


R. Dahm / Developmental Biology 278 (2005) 274–288

Watson and Crick



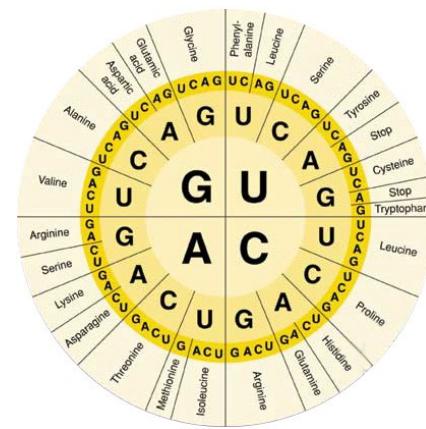
1953: DNA structure



Crick

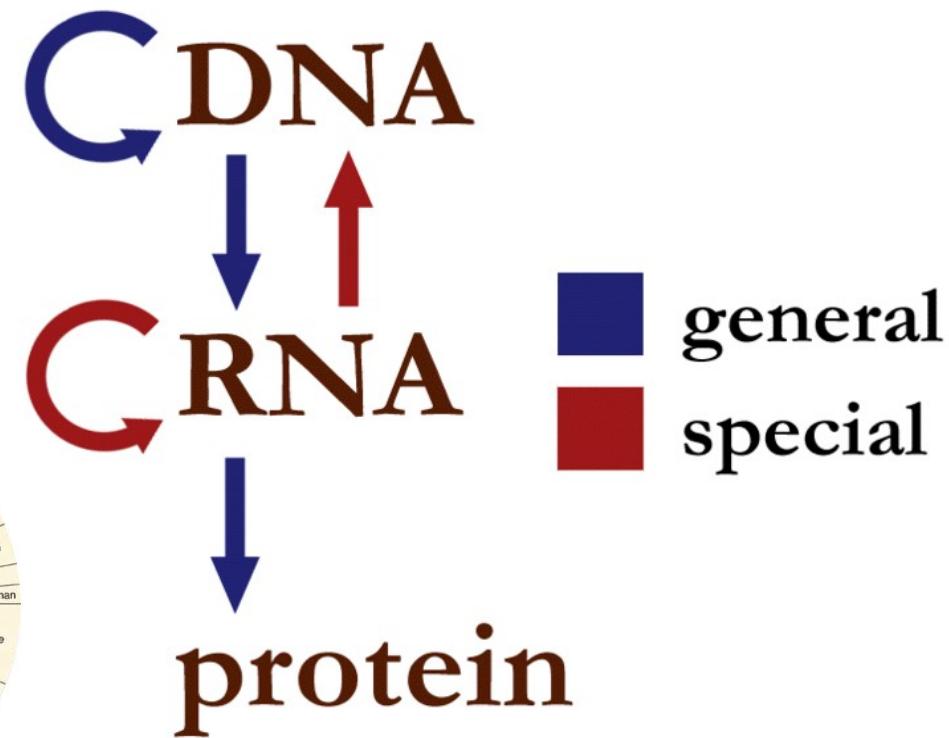


Crick Brenner Barnett



1958: DNA function

Central Dogma of Molecular Biology

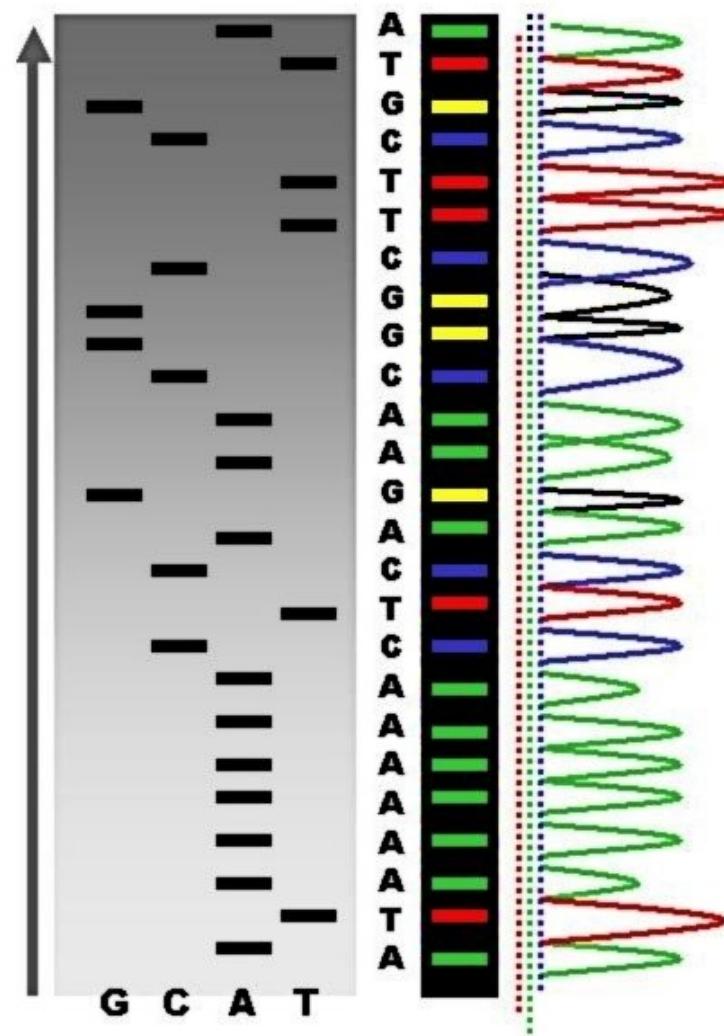


1977: DNA Sanger sequencing

Frederick Sanger



Chain termination method



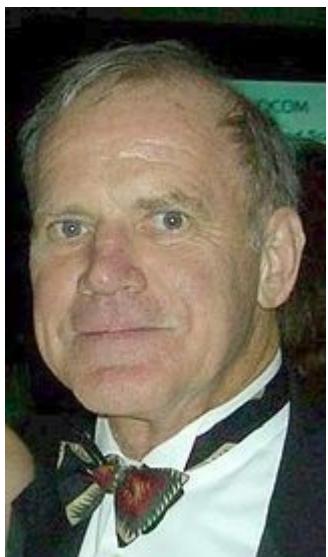
From model organisms to model populations

2008 The 1000 Genomes Project Consortium

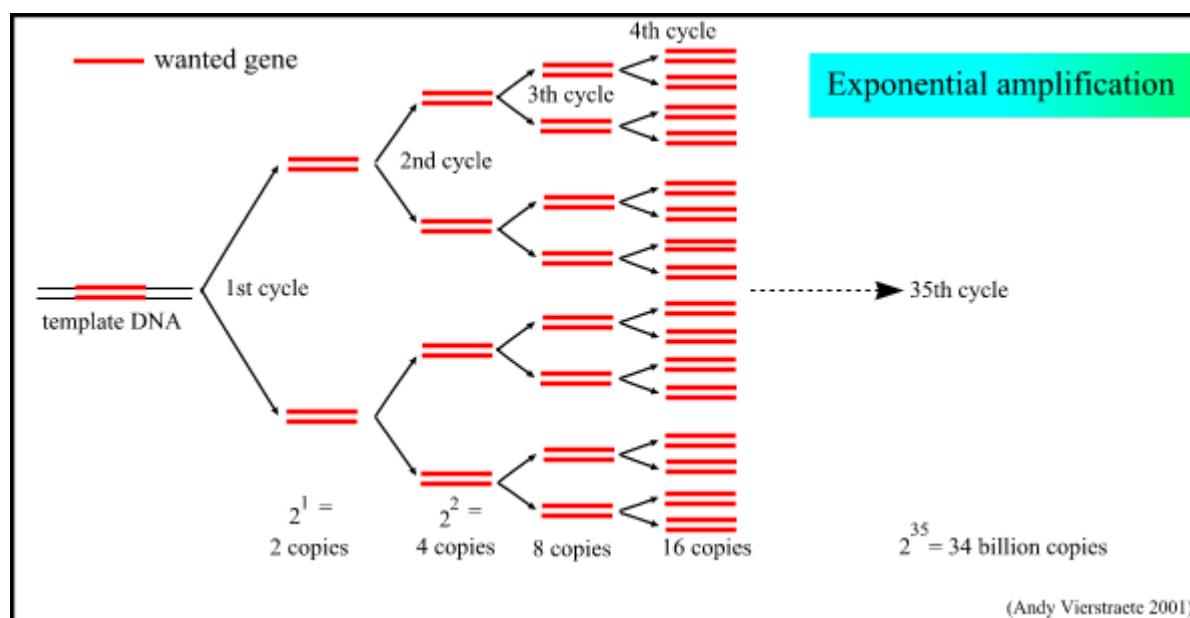
Summary of 1000 Genomes Phase 1 data

	Autosomes	Chromosome X	GENCODE regions ^a
Samples	1092	1092	1092
Total raw bases (Gb)	19,049	804	327
Mean mapped depth (x)	5.1	3.9	80.3
SNPs			
No. sites overall	36.7 M	1.3 M	498 K
Novelty rate ^b	58%	77%	50%
No. Syn / NonSyn / Nonsense	NA	4.7 / 6.5 / 0.097 K	199 / 293 / 6.3 K
Avg. no. SNPs per sample	3.60 M	105 K	24.0 K
Indels			
No. sites overall	1.38 M	59 K	1,867
Novelty rate ^b	62%	73%	54%
No. in-frame / frameshift	NA	19 / 14	719 / 1,066
Avg. no. indels per sample	344 K	13 K	440
Genotyped large deletions			
No. sites overall	13.8 K	432	847
Novelty rate ^b	54%	54%	50%
Avg. no. variants per sample	717	26	39

Kary Mullis

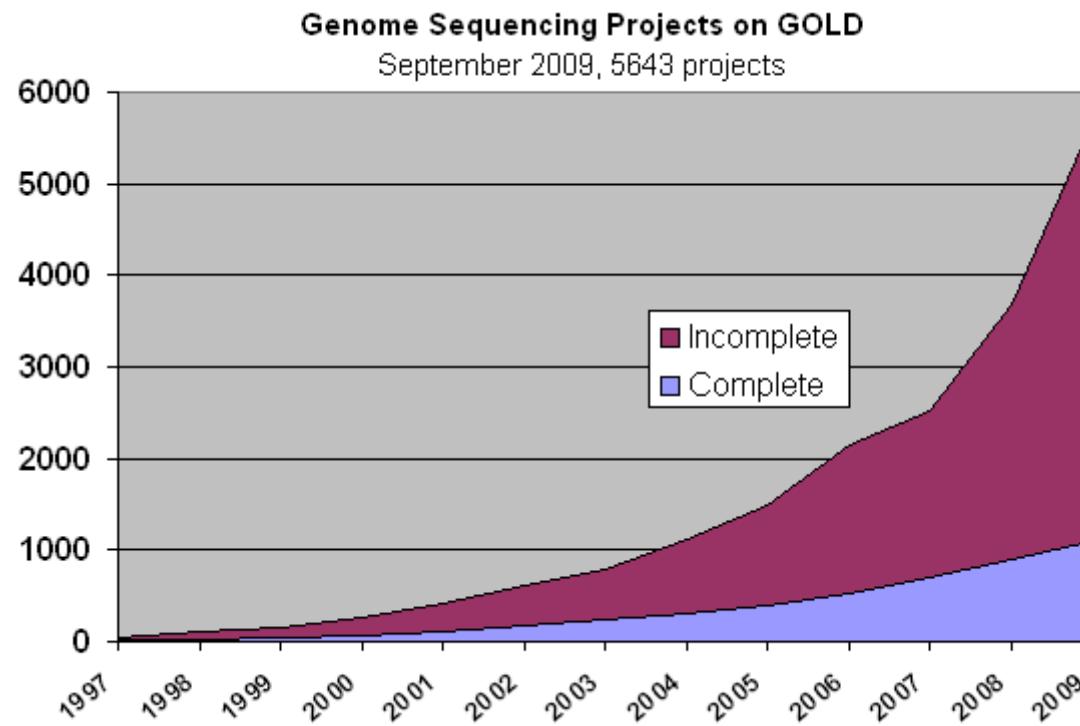
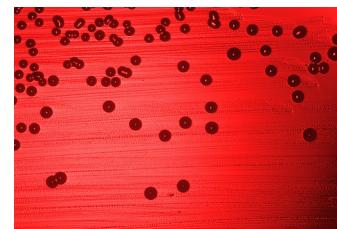


Polymerase Chain Reaction (PCR)



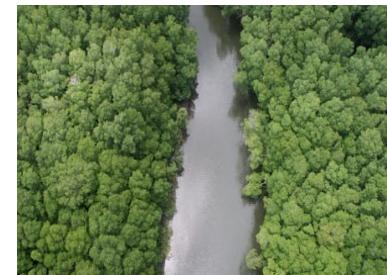
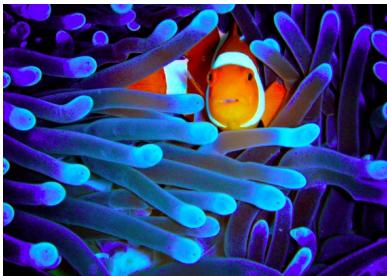
GENOMIC ERA

1995: The genome of *Haemophilus influenzae*

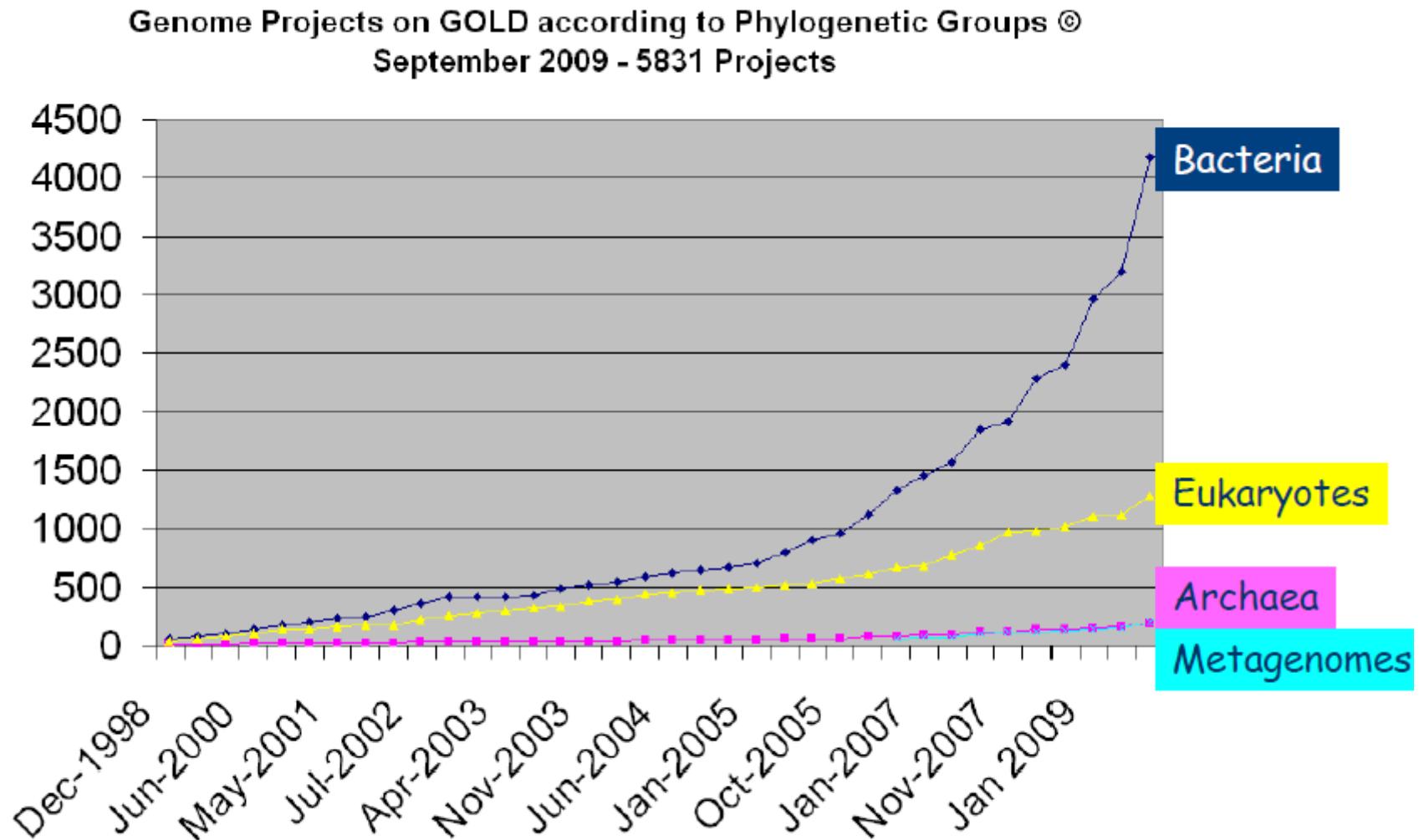


Genomes OnLine Database
<http://www.genomesonline.org>

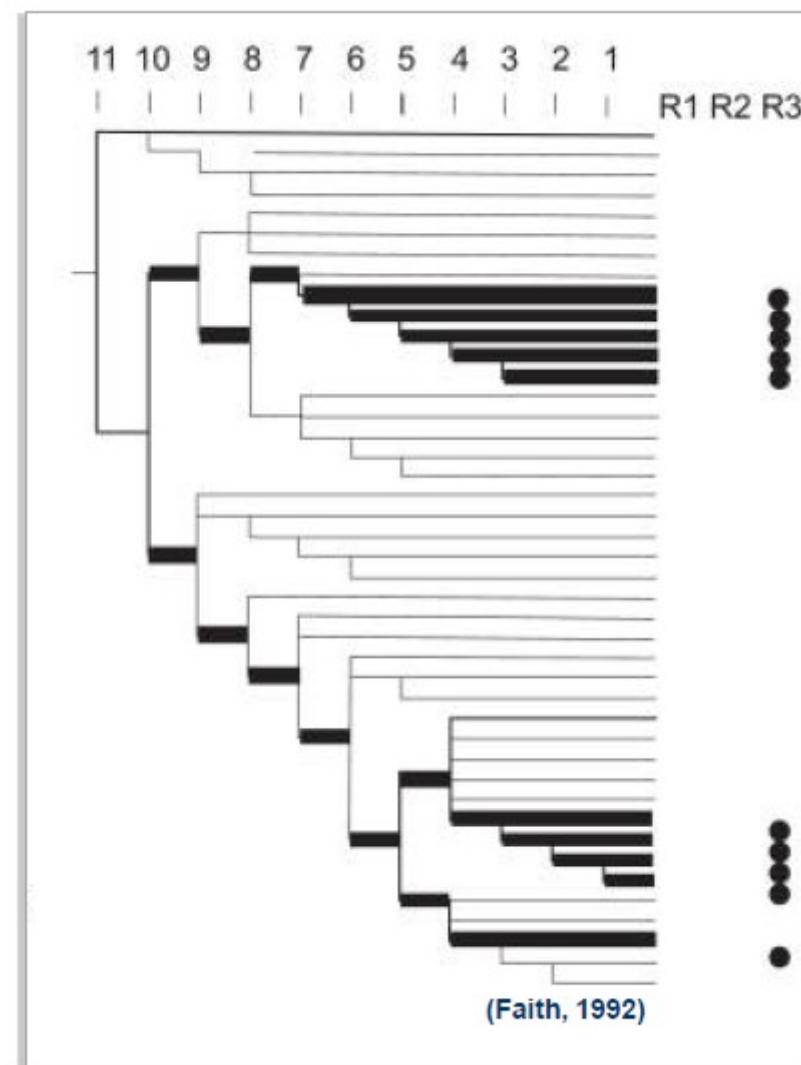
... but, how much diversity is there?



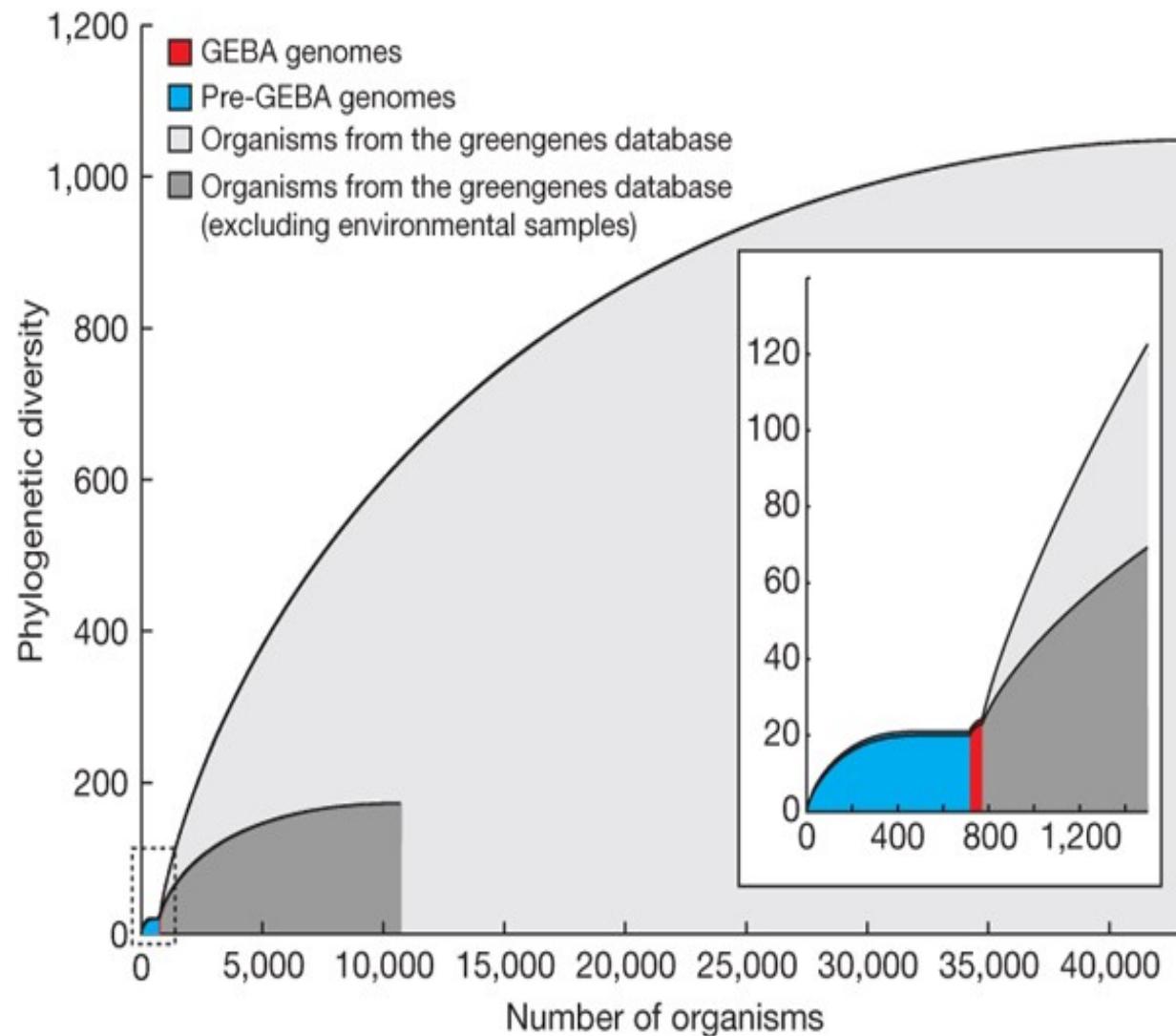
... but, how much diversity is there?



Phylogenetic diversity



A Genomic Encyclopedia of Bacteria and Archaea



Metagenomics
to understand
biodiversity

From model organisms to model populations

2008 The 1000 Genomes Project Consortium

Summary of 1000 Genomes Phase 1 data

	Autosomes	Chromosome X	GENCODE regions ^a
Samples	1092	1092	1092
Total raw bases (Gb)	19,049	804	327
Mean mapped depth (x)	5.1	3.9	80.3
SNPs			
No. sites overall	36.7 M	1.3 M	498 K
Novelty rate ^b	58%	77%	50%
No. Syn / NonSyn / Nonsense	NA	4.7 / 6.5 / 0.097 K	199 / 293 / 6.3 K
Avg. no. SNPs per sample	3.60 M	105 K	24.0 K
Indels			
No. sites overall	1.38 M	59 K	1,867
Novelty rate ^b	62%	73%	54%
No. in-frame / frameshift	NA	19 / 14	719 / 1,066
Avg. no. indels per sample	344 K	13 K	440
Genotyped large deletions			
No. sites overall	13.8 K	432	847
Novelty rate ^b	54%	54%	50%
Avg. no. variants per sample	717	26	39

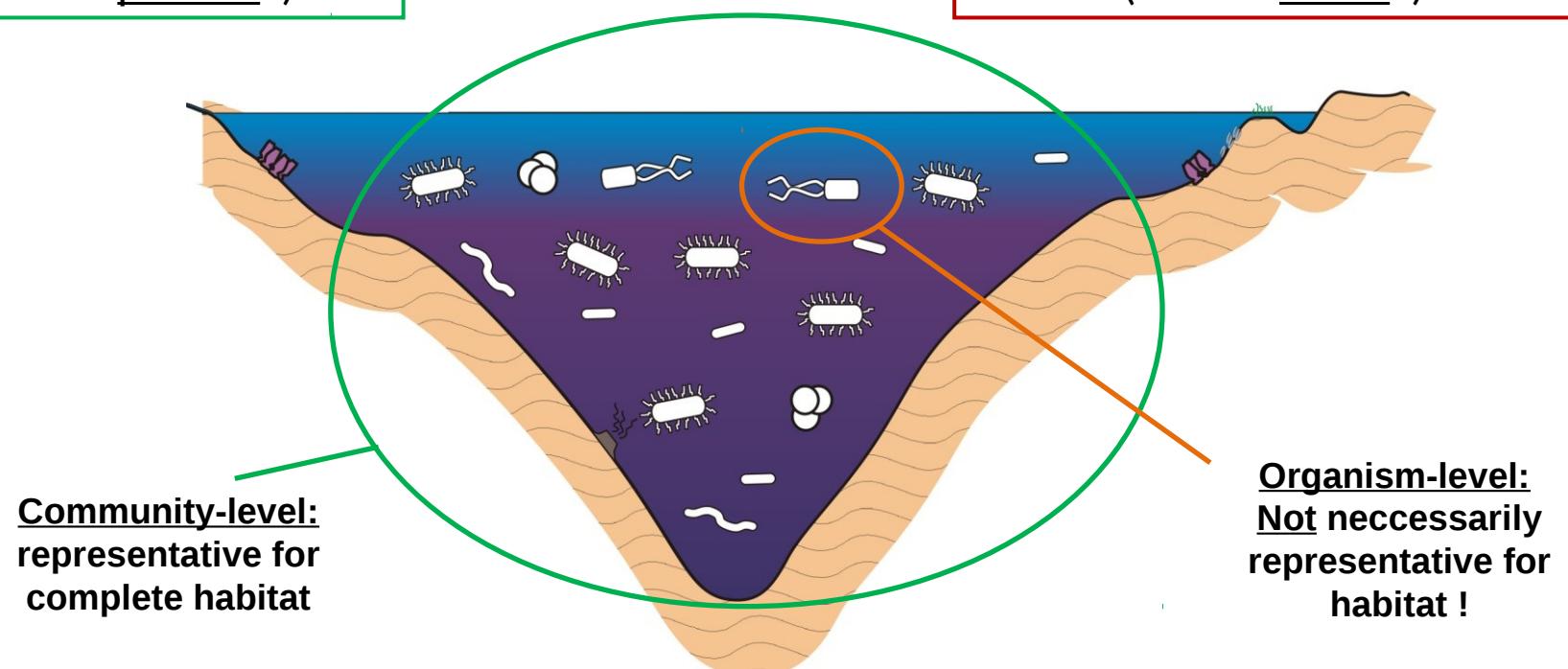
Meta-omics (genomics/ transcriptomics)

-ome = The *totality* of something (Genome = the totality of genes in an Organism)

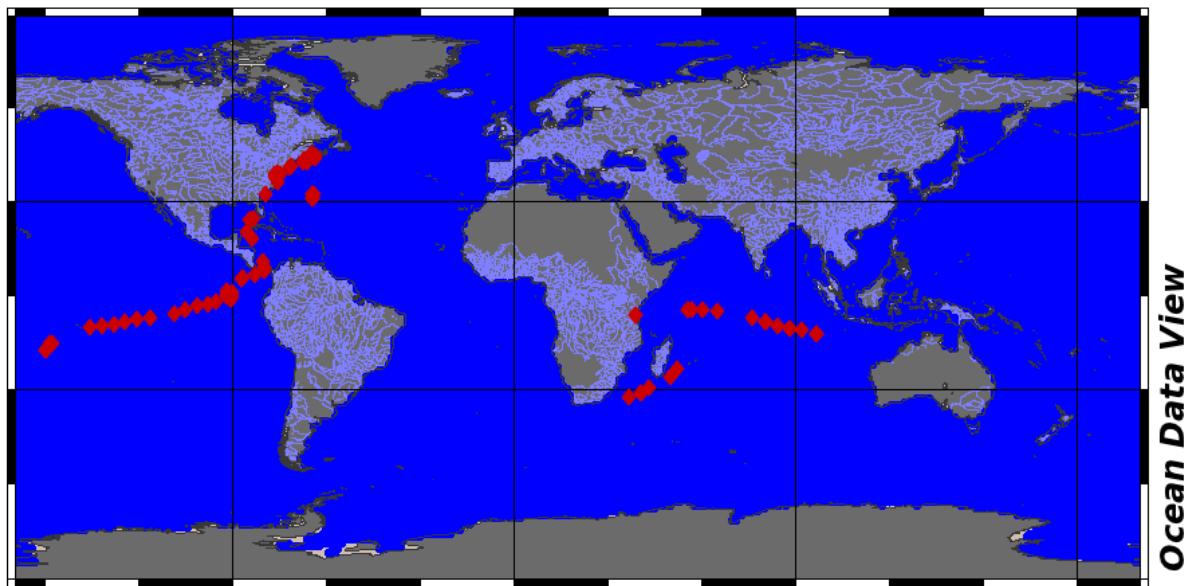
-omics = specialized fields in biological research (genome/genomics, proteome/proteomics, etc.)

Metagenomics : **DNA-level**
(what is present?)

Metatranscriptomics **RNA-level**
(what is active?)

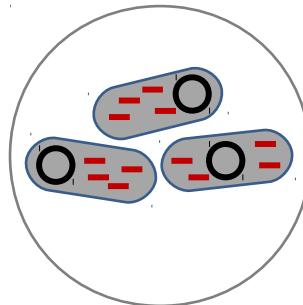


The Sorcerer II Global Ocean Sampling Expedition



Sample Preparation

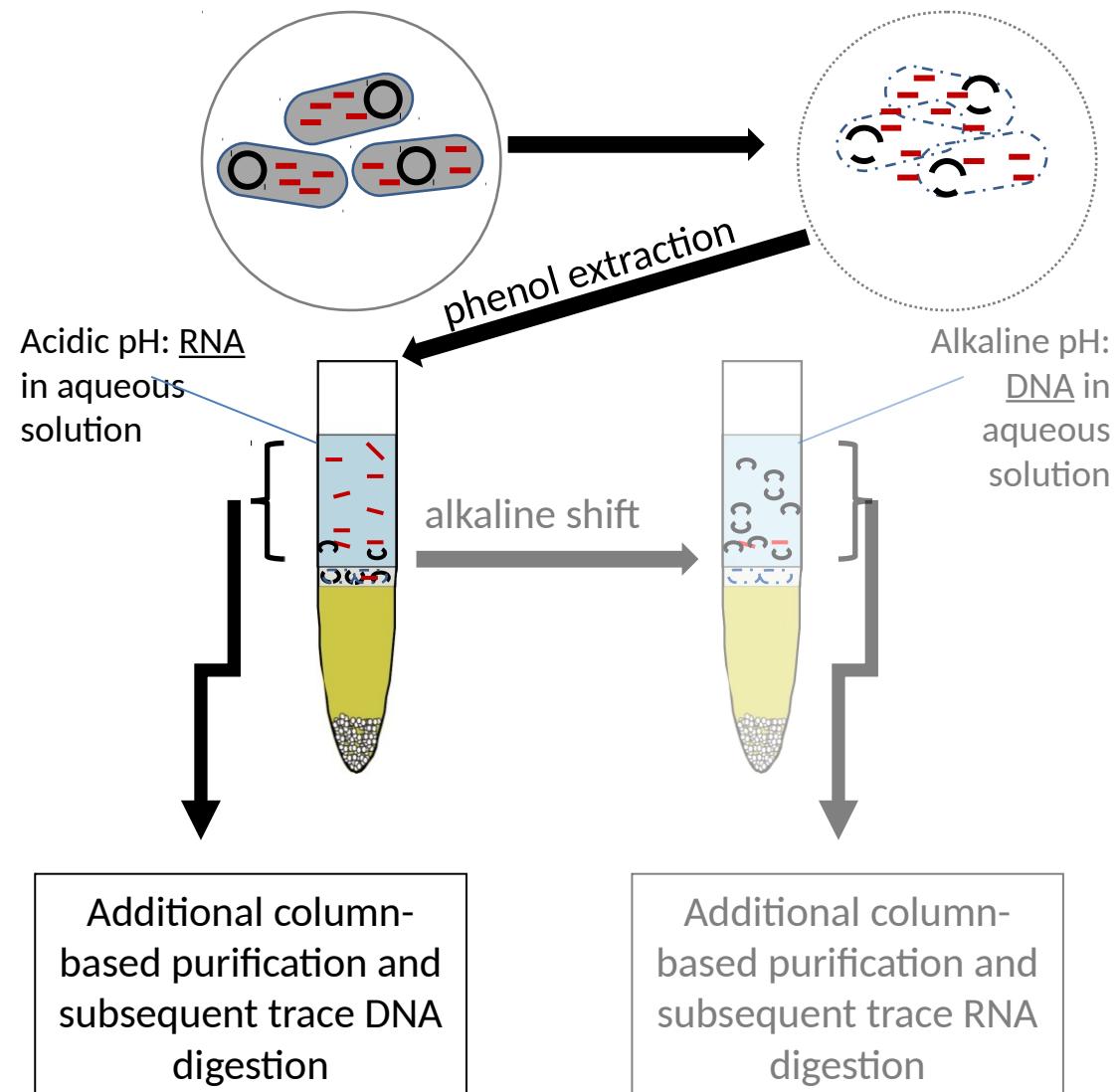
Co-extraction of RNA and DNA



Sample Preparation

Co-extraction of RNA and DNA

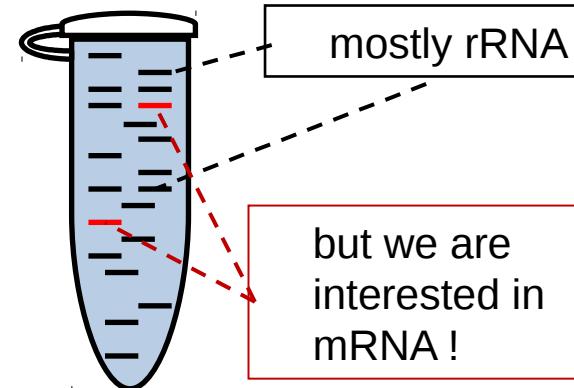
- Cell Disruption
- Phenol-based extraction of nucleic acids (DNA+RNA)
- Additional Column-based purification (this also removes 5S rRNA and tRNA)
- Enzymatical removal of trace DNA or RNA, respectively.



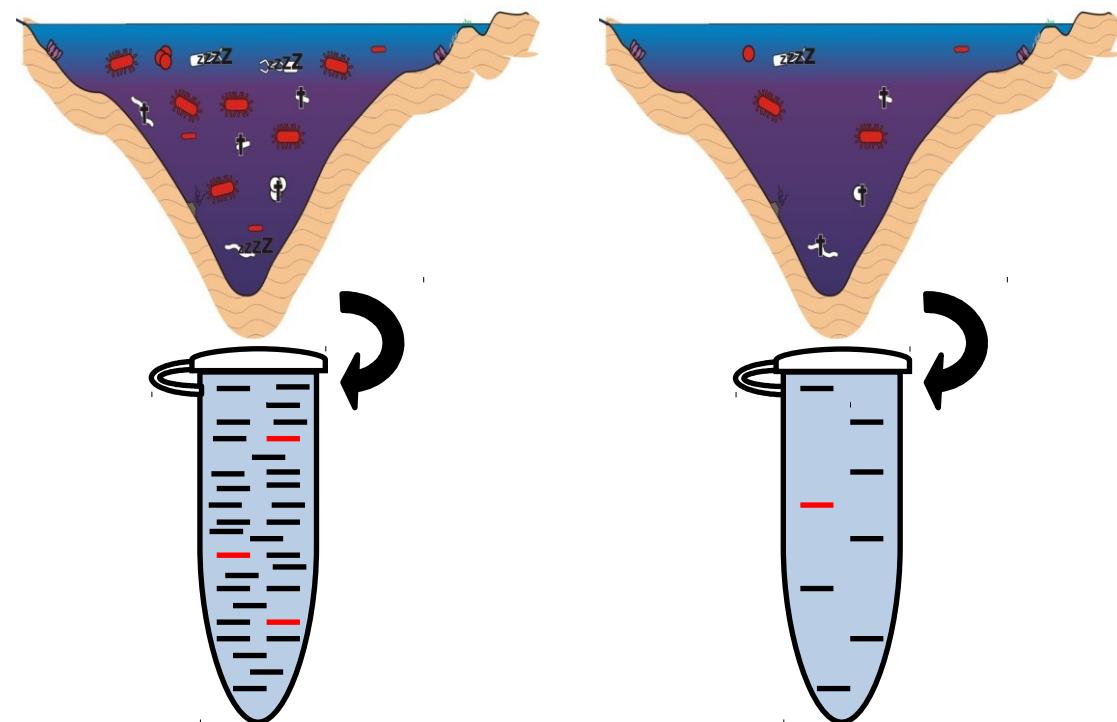
RNA Samples

What to expect

- No mRNA without ribosomes!
- Ribosomal RNA (rRNA) over-represented in samples (90-99%)
- RNA-yield dependent on activity and density of community
- Low RNA-yield at low active cell densities (e.g. Deep Sea)



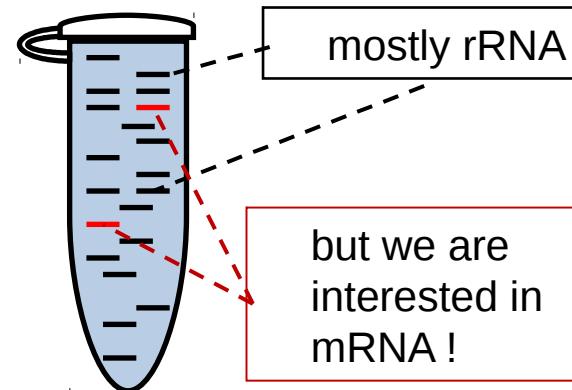
Solution: rRNA depletion
(if yield is high enough)



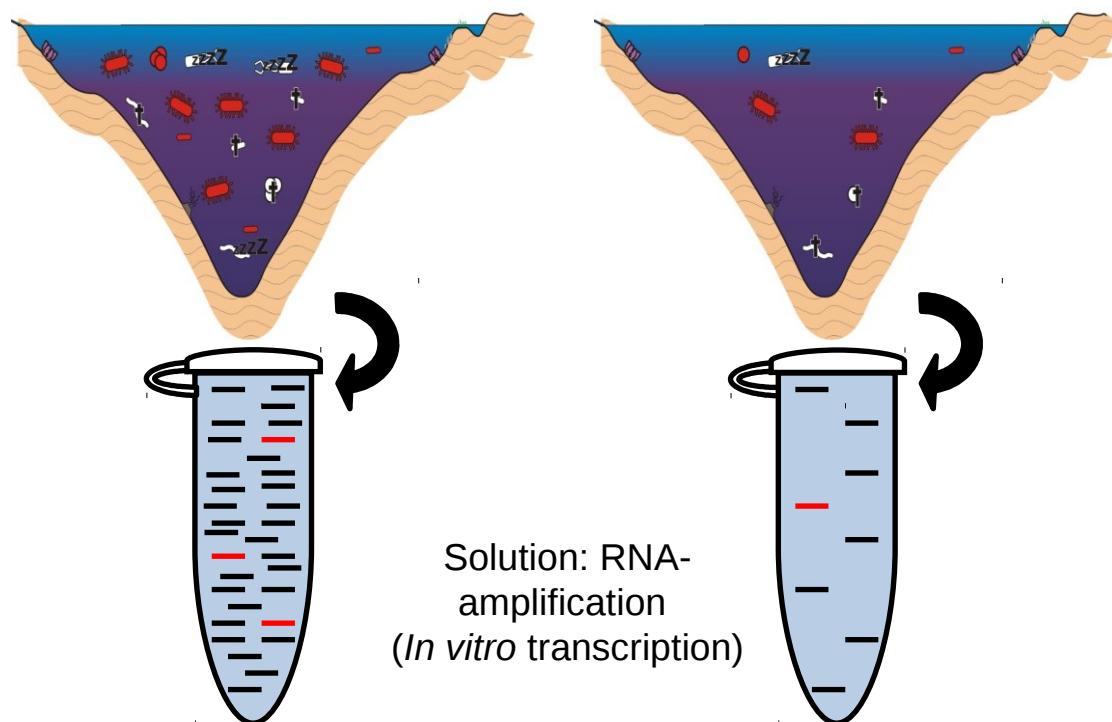
RNA Samples

What to expect

- No mRNA without ribosomes!
- Ribosomal RNA (rRNA) over-represented in samples (90-99%)
- RNA-yield dependent on activity and density of community
- Low RNA-yield at low active cell densities (e.g. Deep Sea)



Solution: rRNA depletion
(if yield is high enough)



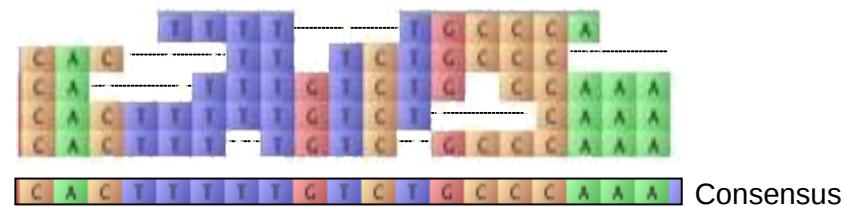
Sample processing

- No RNA amplification
- No rRNA-depletion
- Sequencing „as it is“
(Minimization of methodical bias)

- RNA amplification
- rRNA-depletion
- Maximization of mRNA output
and sequencing depth

estimate bias by RNA amplification methods

Assemble reads into (almost)
complete transcripts



Identify and annotate transcripts
of known protein-coding genes



Megan

Identify homologs between samples (Sequence-based and annotation-based)