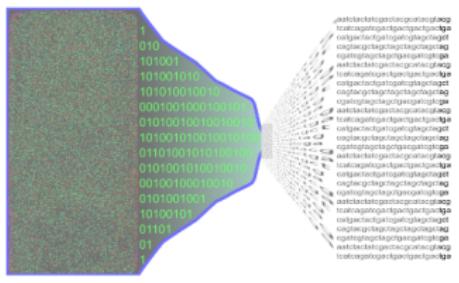


Pyrosequencing Shotgun Sequencing and Metatranscriptomics Analysis

Giuseppe D'Auria



FISABIO, Valencia

GENERALITAT
GIUSEPPE D'AURIA

Fundación para el Fomento de la

PYROSEQUENCING, SHOTGUN SEQUENCING, AND METATRANSCRIPTOMICS
ANALYSIS

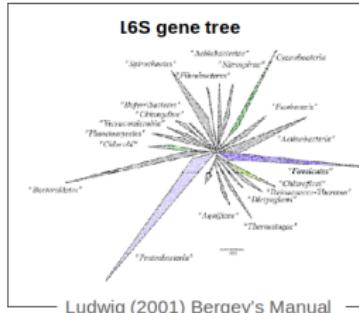
Norwich 12-16 October 2015

14/10/2015

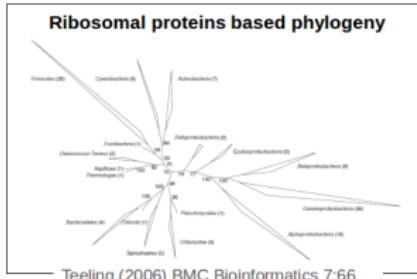
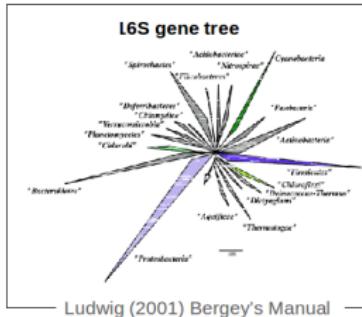
eu el Foment de la
Salut i Recerca
unitat Valenciana



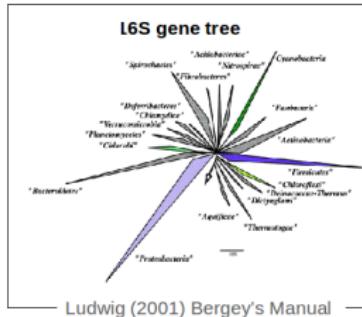
NAMING BACTERIA



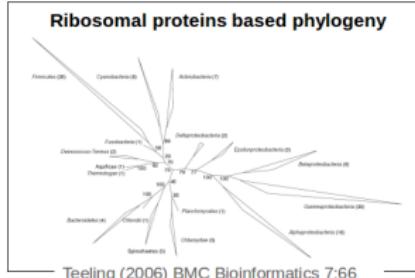
NAMING BACTERIA



NAMING BACTERIA



- Ludwig (2001) Bergey's Manual



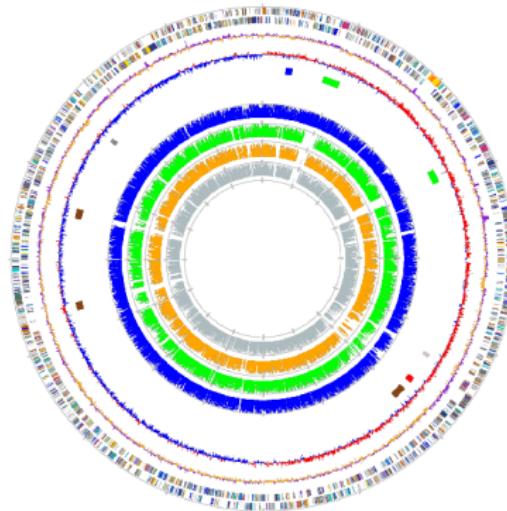
– Teeling (2006) BMC Bioinformatics 7:66



Biology

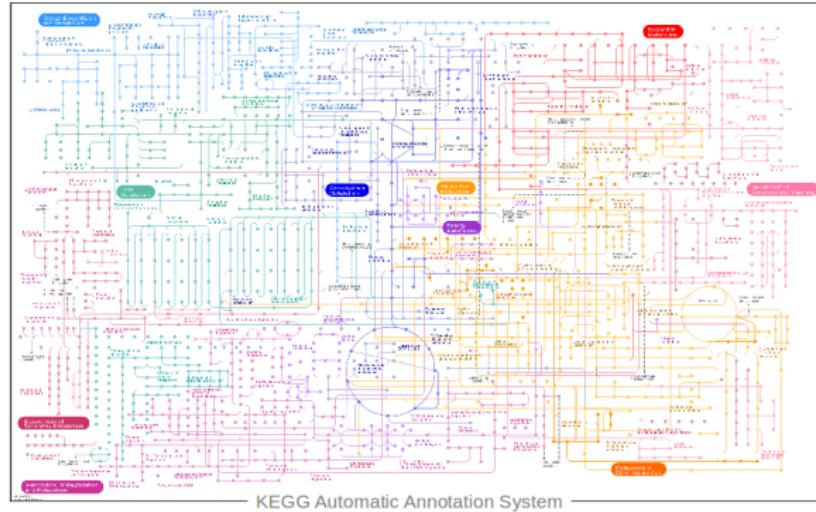
SOME MORE INFO - GENOMICS

GENOMICS: A WIDER SCREENSHOT

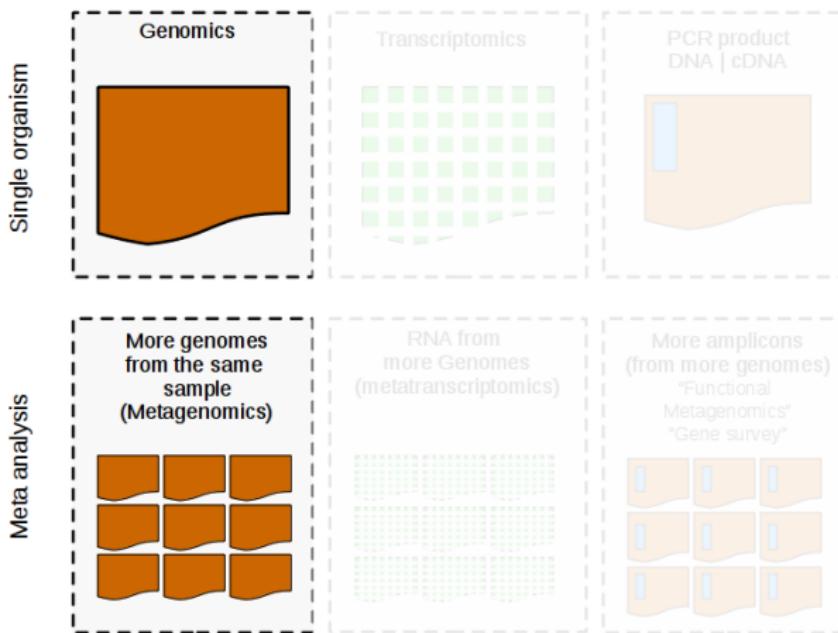


SOME MORE INFO - GENOMICS

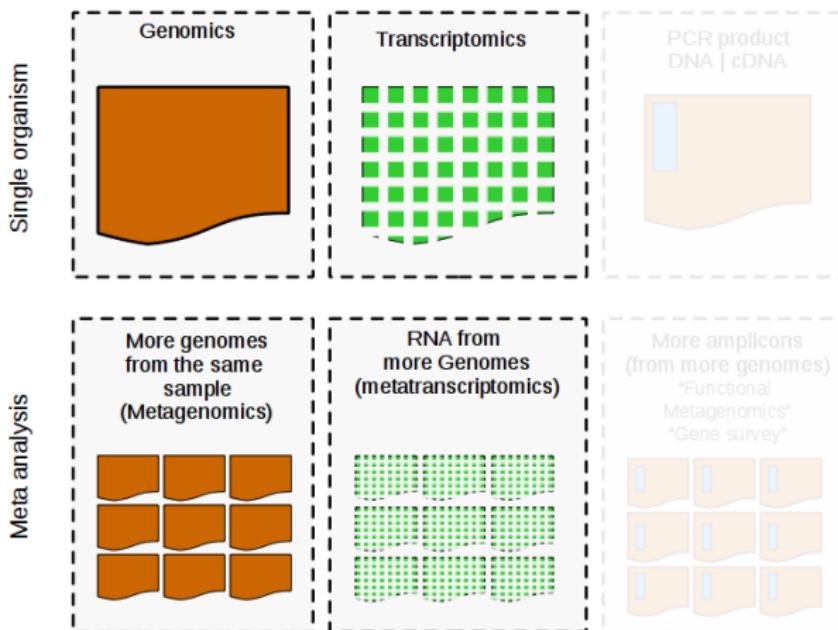
GENOMICS: A WIDER Screenshot



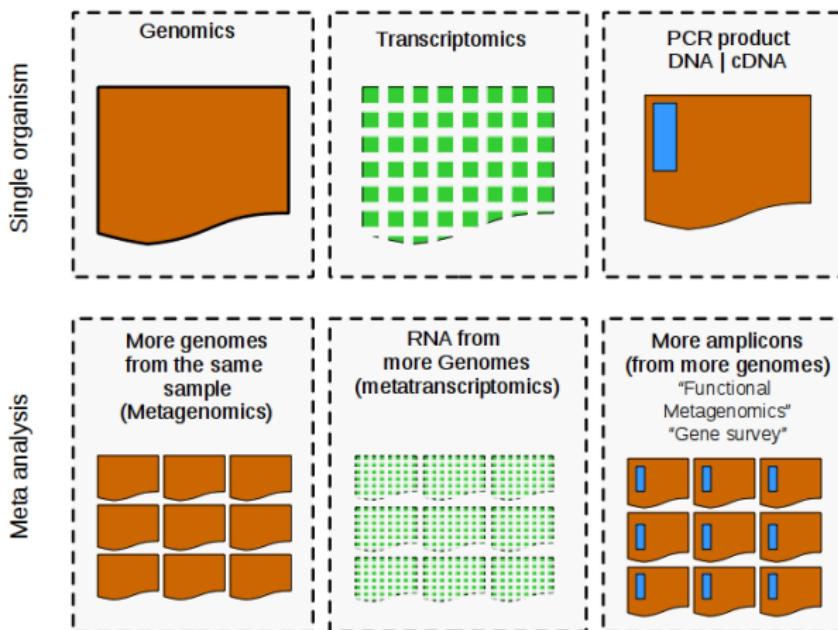
METEGENOMICS OVERVIEW



METEGENOMICS OVERVIEW



METEGENOMICS OVERVIEW



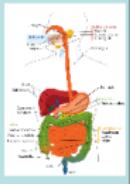
METEGENOMICS OVERVIEW

THE SAMPLING



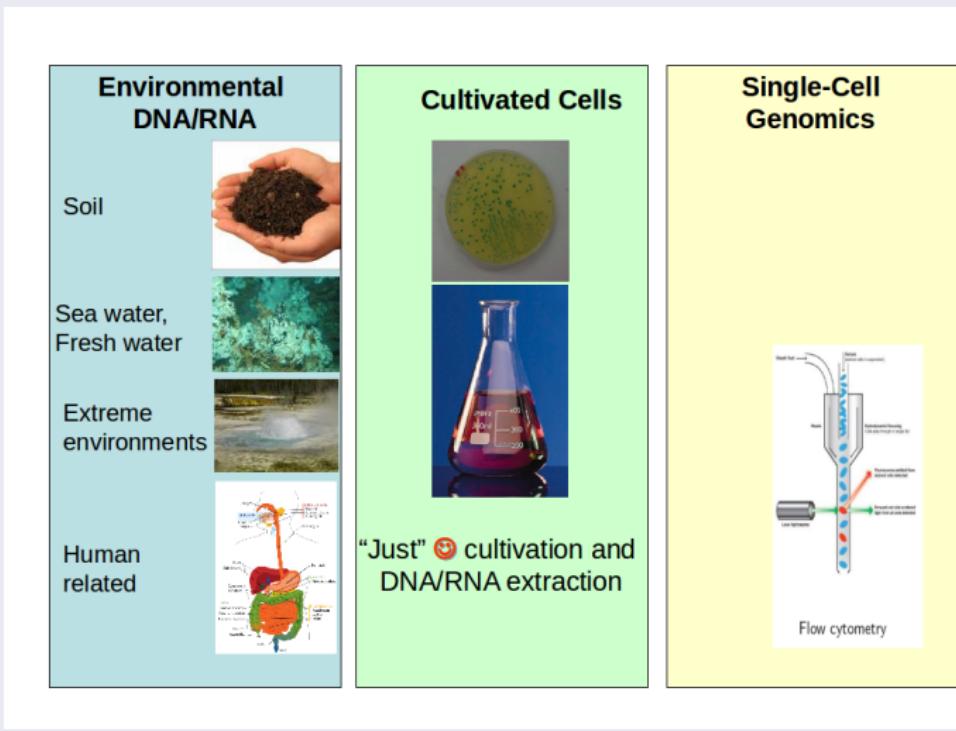
METEGENOMICS OVERVIEW

THE SAMPLING

Environmental DNA/RNA		Cultivated Cells
Soil		
Sea water, Fresh water		
Extreme environments		
Human related		"Just" ☺ cultivation and DNA/RNA extraction

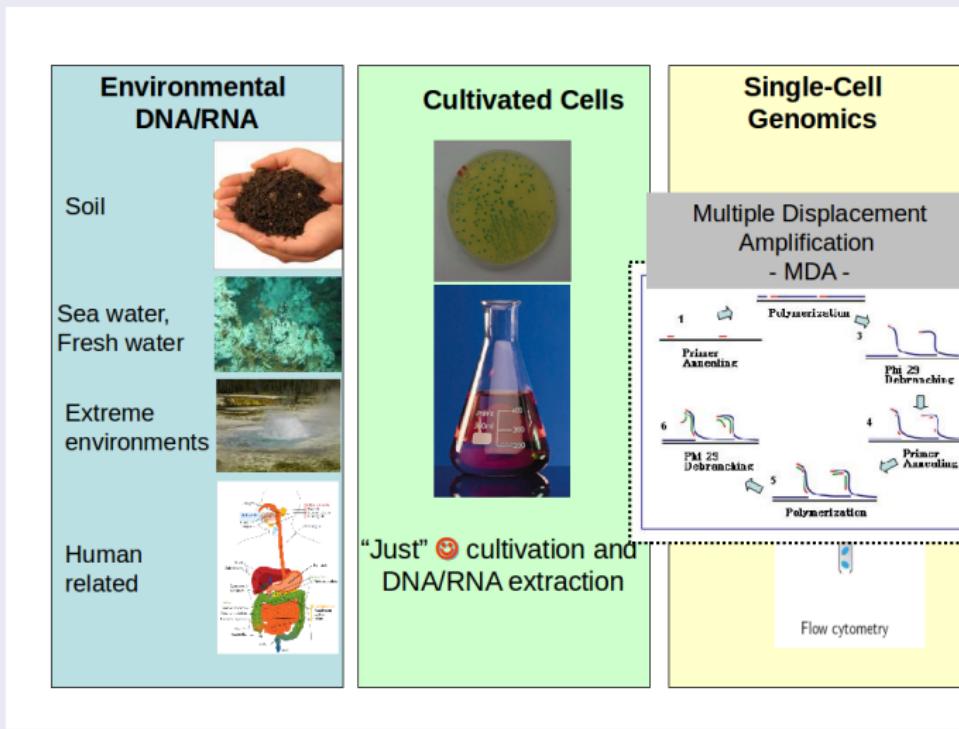
METEGENOMICS OVERVIEW

THE SAMPLING



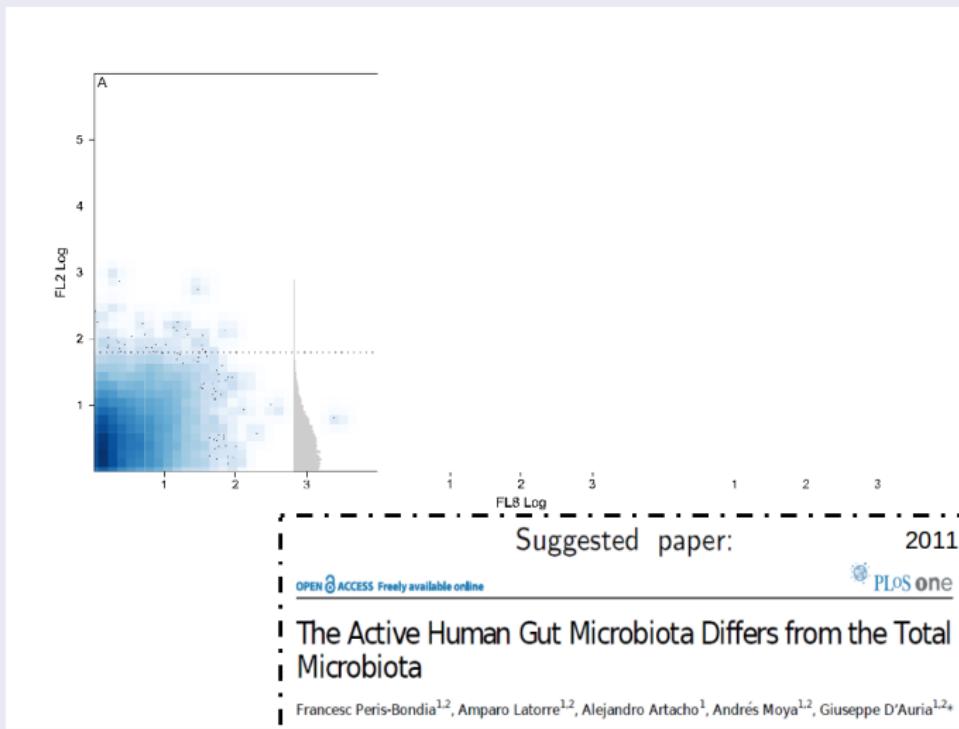
METEGENOMICS OVERVIEW

THE SAMPLING



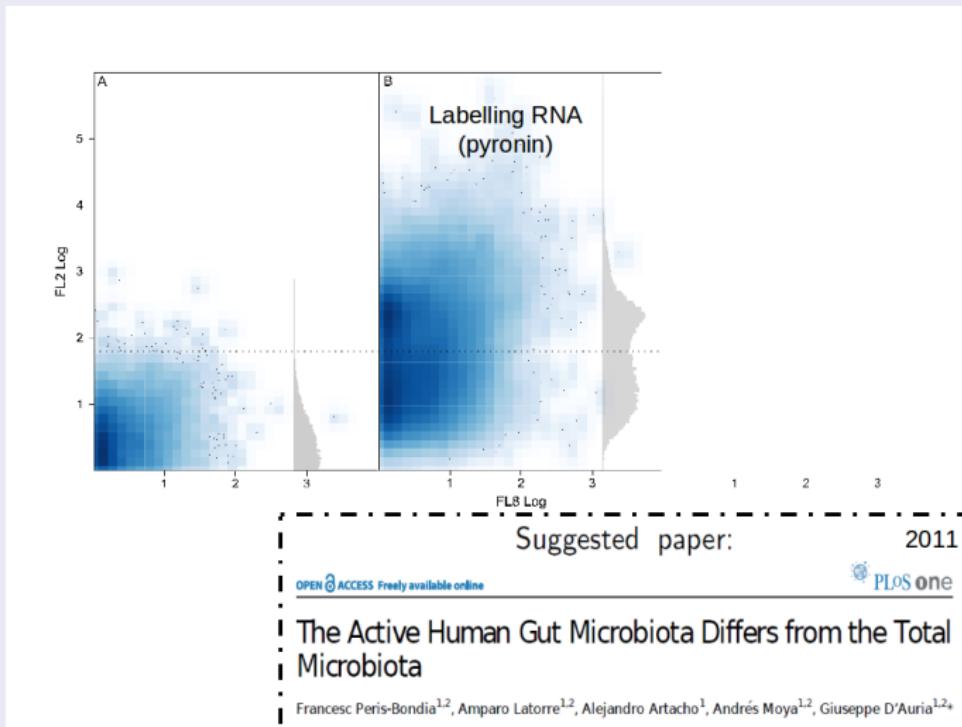
METEGENOMICS OVERVIEW

THE ACTIVE FRACTION



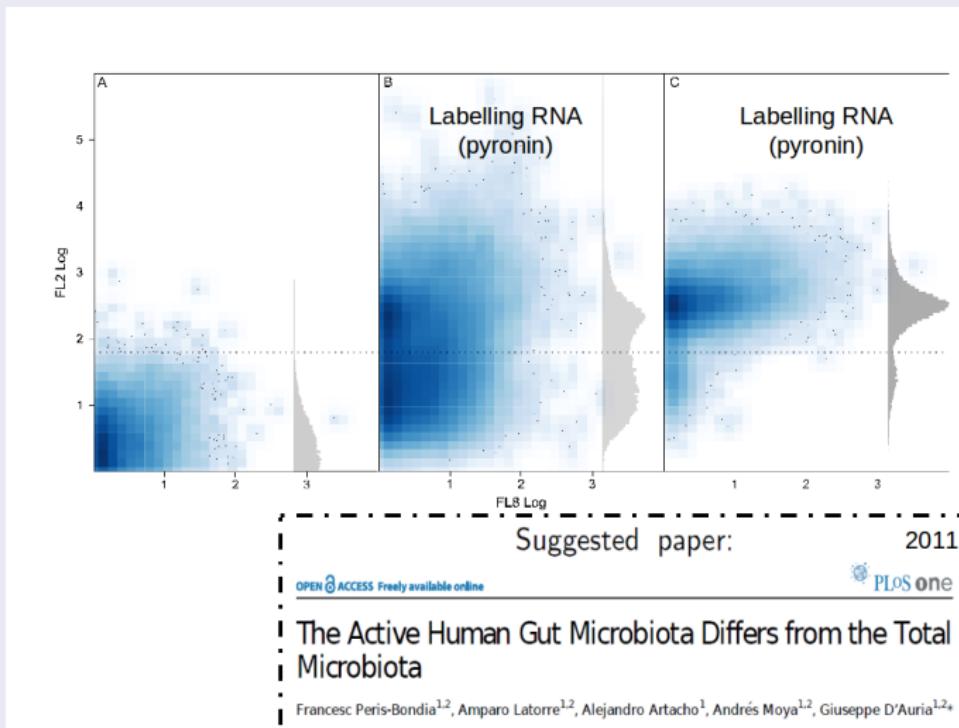
METEGENOMICS OVERVIEW

THE ACTIVE FRACTION



METEGENOMICS OVERVIEW

THE ACTIVE FRACTION



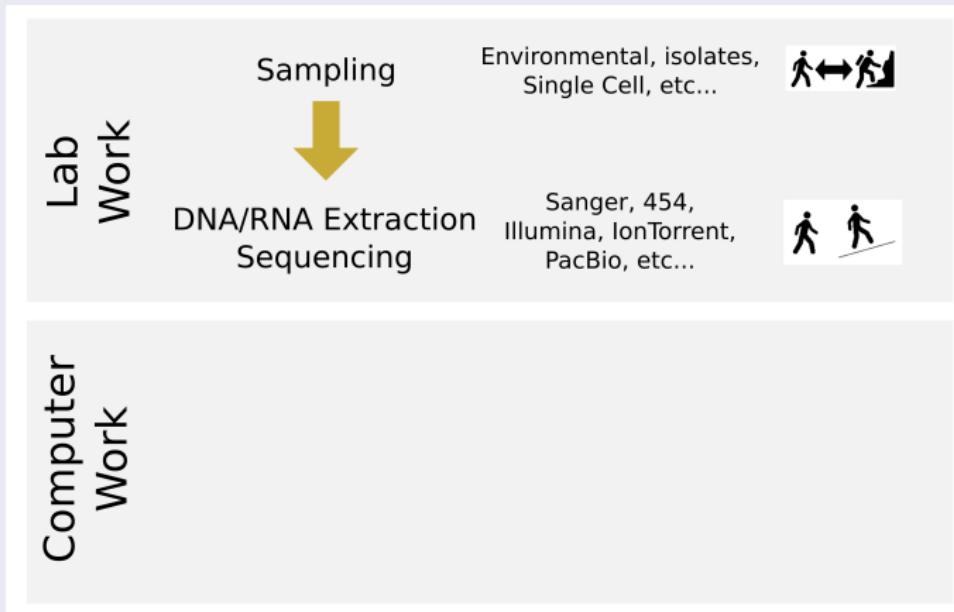
DIFFICULTY SCALE

FROM SAMPLING TO DATA ANALYSIS



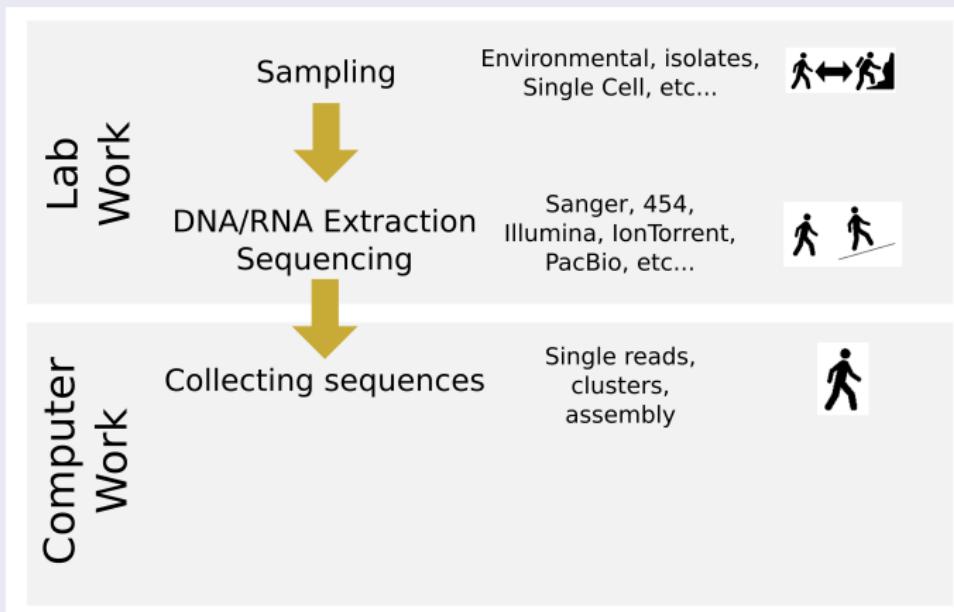
DIFFICULTY SCALE

FROM SAMPLING TO DATA ANALYSIS



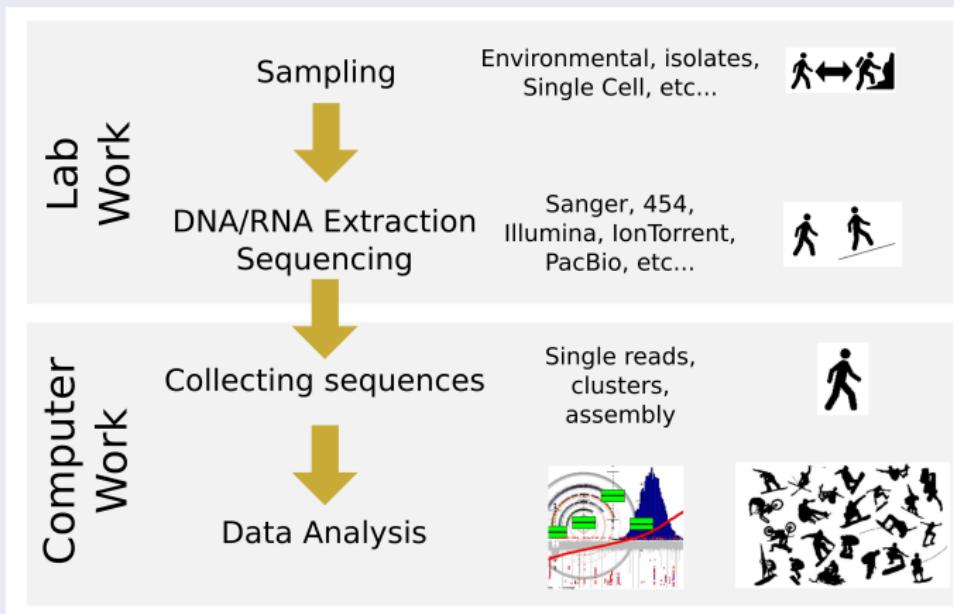
DIFFICULTY SCALE

FROM SAMPLING TO DATA ANALYSIS



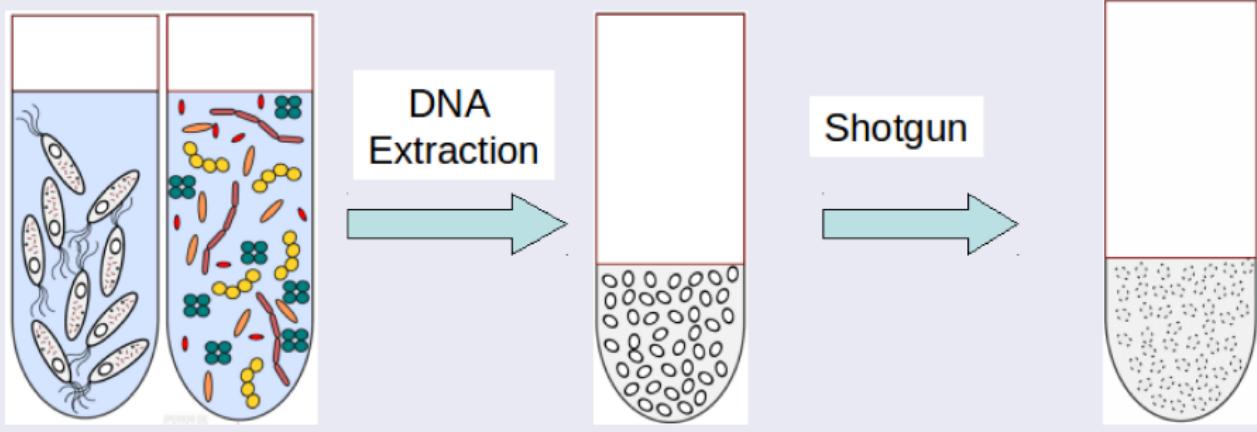
DIFFICULTY SCALE

FROM SAMPLING TO DATA ANALYSIS



WORK FLOW

GENOMICS AND METAGENOMICS

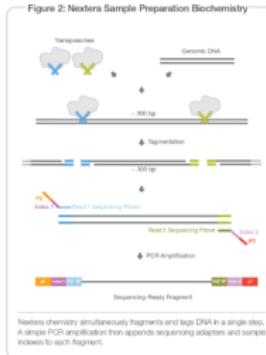


WORK FLOW

GENOMICS AND METAGENOMICS



Nebulization Hydroshearing



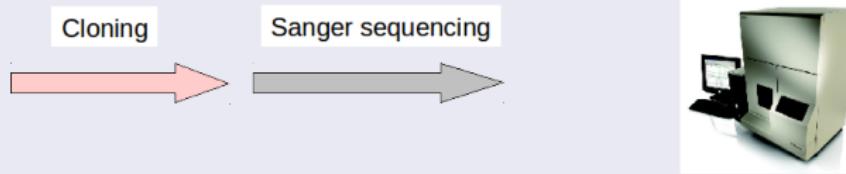
Tagmentation



Ultrasound

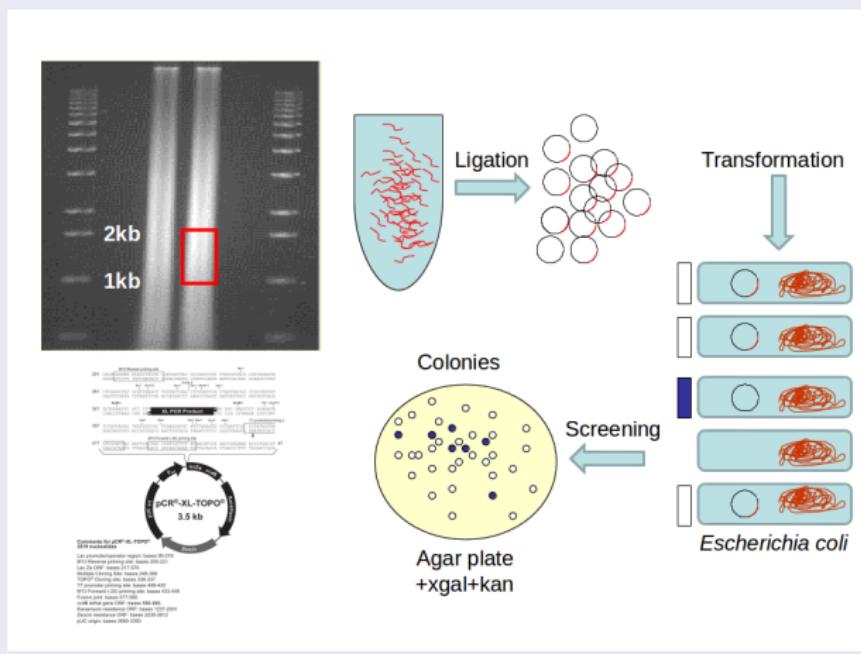
GENOMICS AND METAGENOMICS

CLONING PROTOCOL 1ST GEN



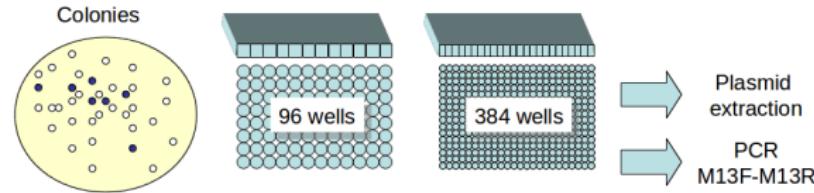
GENOMICS AND METAGENOMICS

CLONING PROTOCOL 1ST GEN



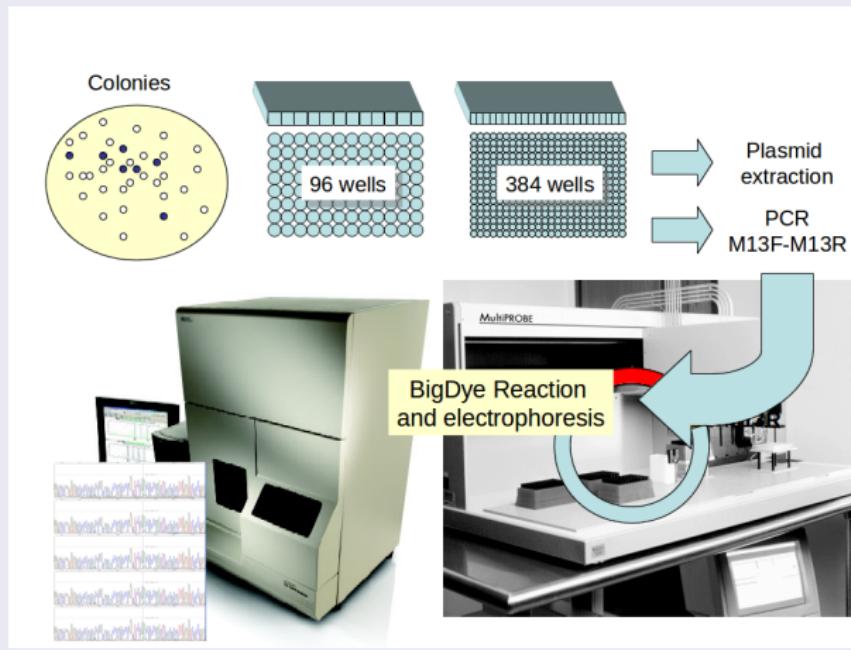
GENOMICS AND METAGENOMICS

CLONING PROTOCOL 1ST GEN



GENOMICS AND METAGENOMICS

CLONING PROTOCOL 1ST GEN

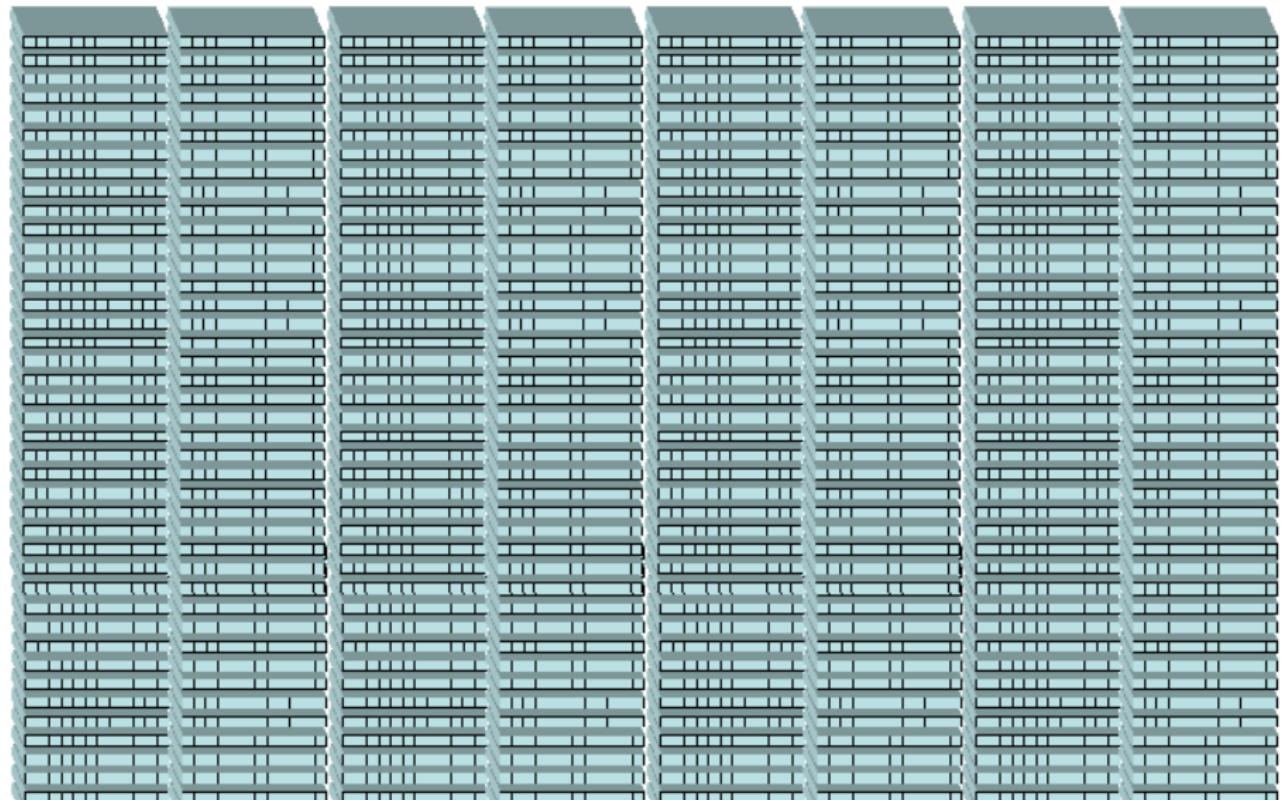


GENOMICS AND METAGENOMICS

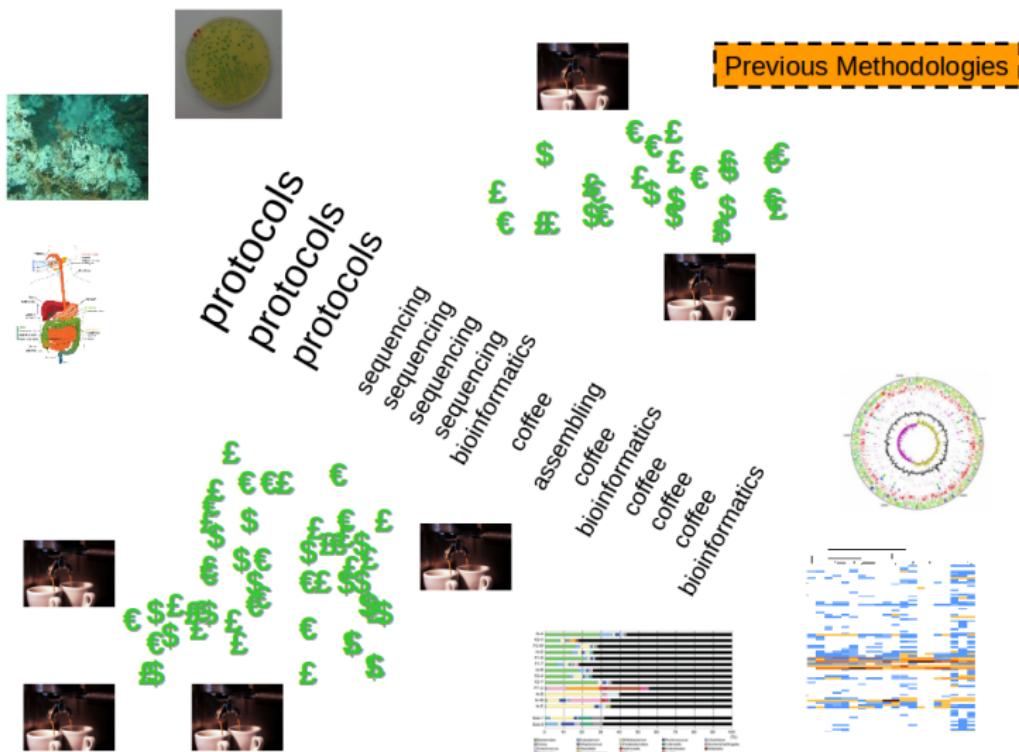
WHOLE DE NOVO GENOME PROJECT - 3.5 Mb

- 8x Sanger coverage
- $3.5 \text{ Mb} \times 8 = 28 \text{ Mb}$
- $28 \text{ Mb} / 800 \text{ bp} = 35.000 \text{ reads}$
- $35.000 \text{ reads} / 96 \text{ wells} = 364 \text{ plates}$
- $2 \text{ Euros} \times \text{read} = 192 \text{ Euros} / \text{Plate}$
- $364 \text{ Plates} = 70.000 \text{ Euros}$

GENOMICS AND METAGENOMICS

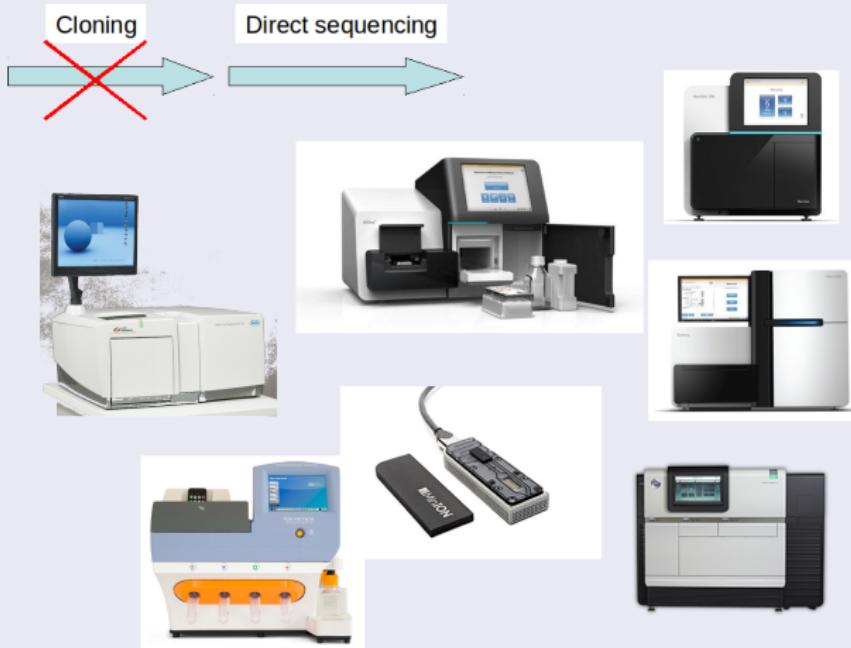


HISTORY



HISTORY

2ND - 3RD - X GENERATIONS

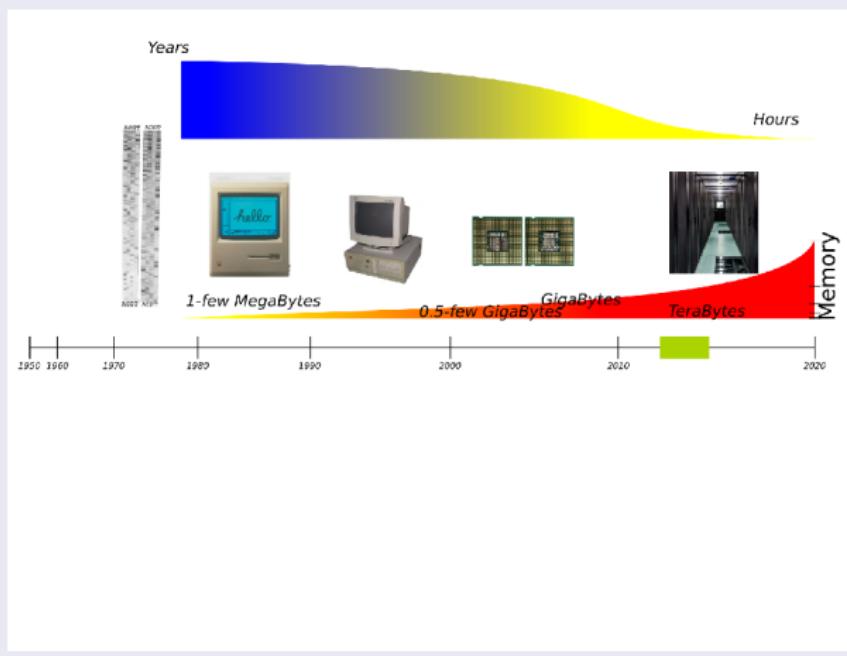


HISTORY



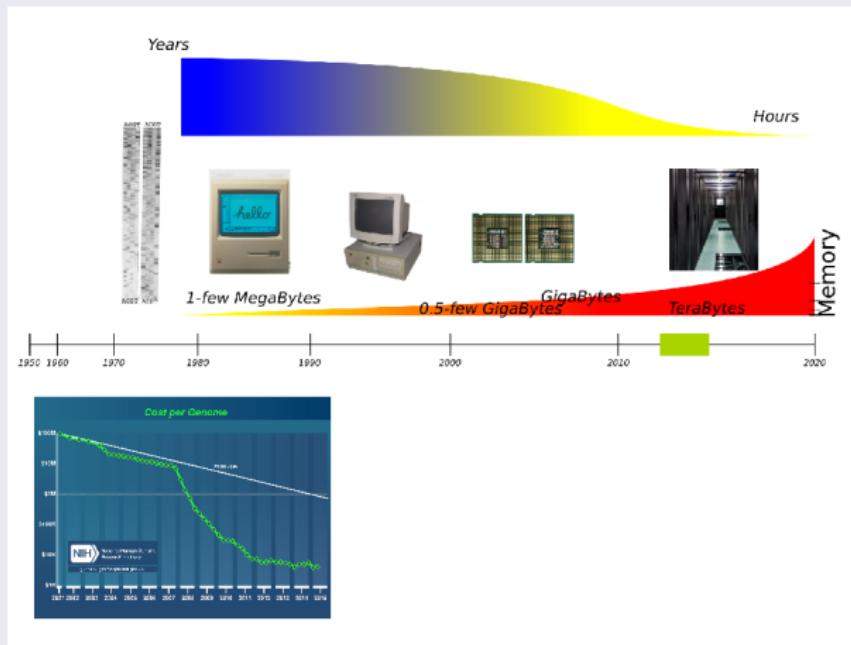
HISTORY

SEQUENCING EVOLUTION AND DEMOCRATIZATION



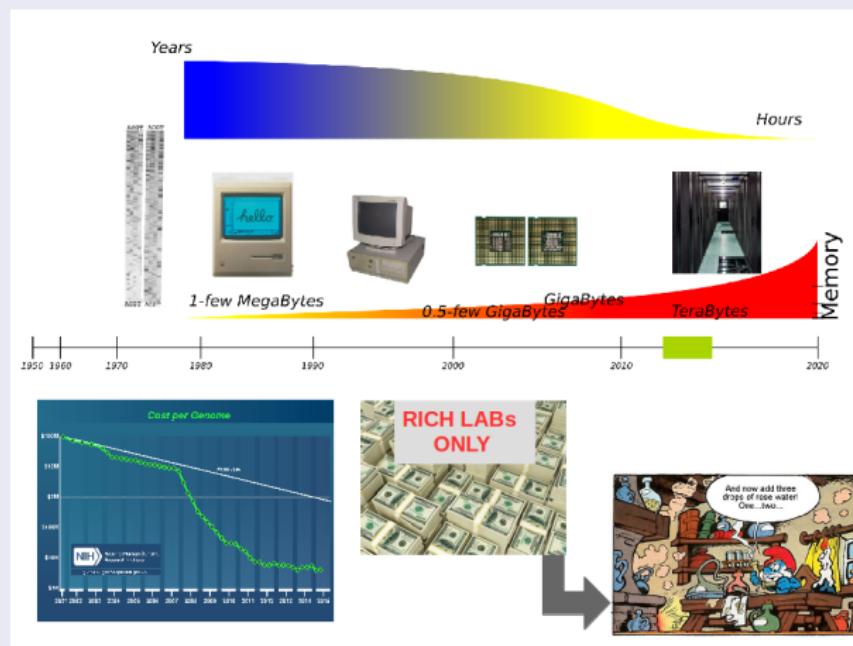
HISTORY

SEQUENCING EVOLUTION AND DEMOCRATIZATION



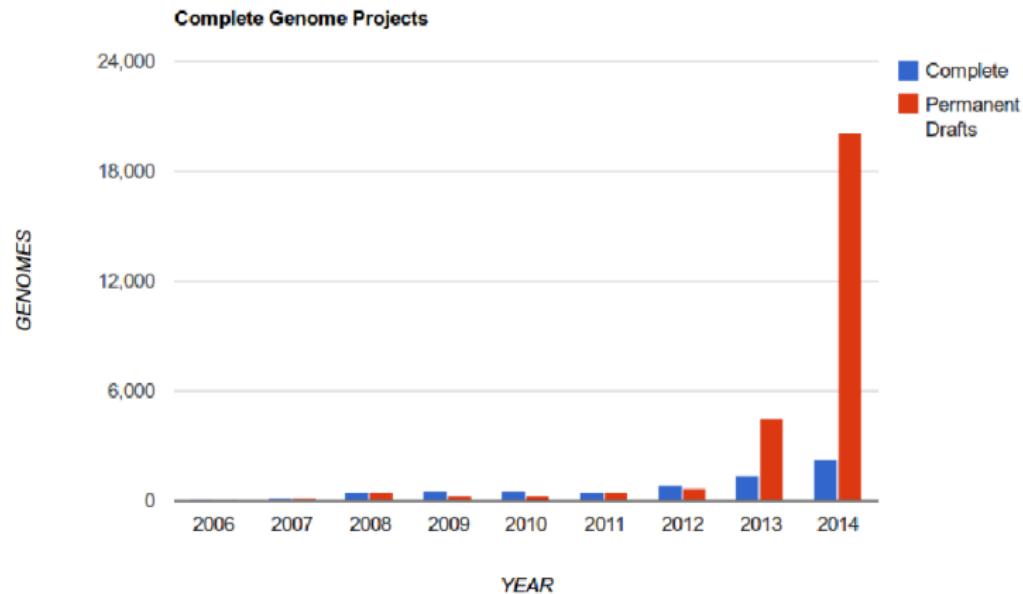
HISTORY

SEQUENCING EVOLUTION AND DEMOCRATIZATION

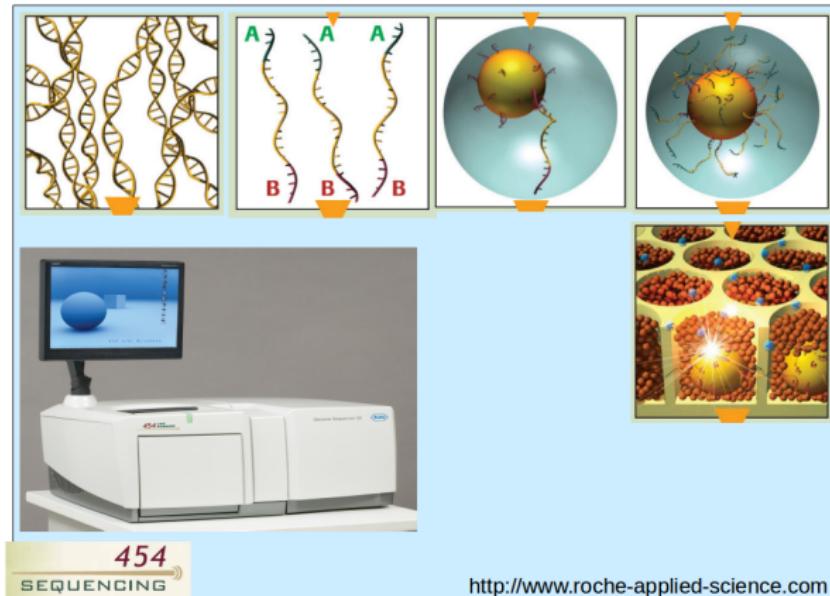


HISTORY

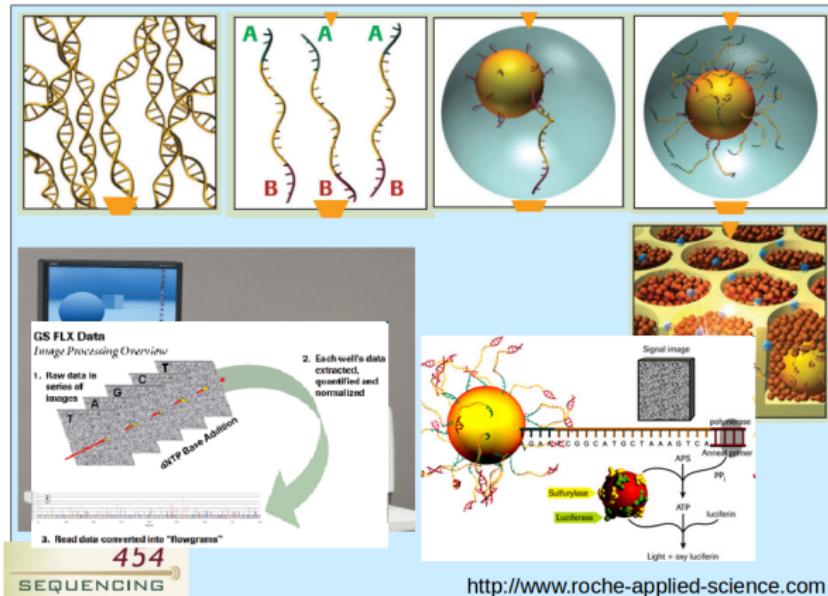
SEQUENCING EVOLUTION AND DEMOCRATIZATION



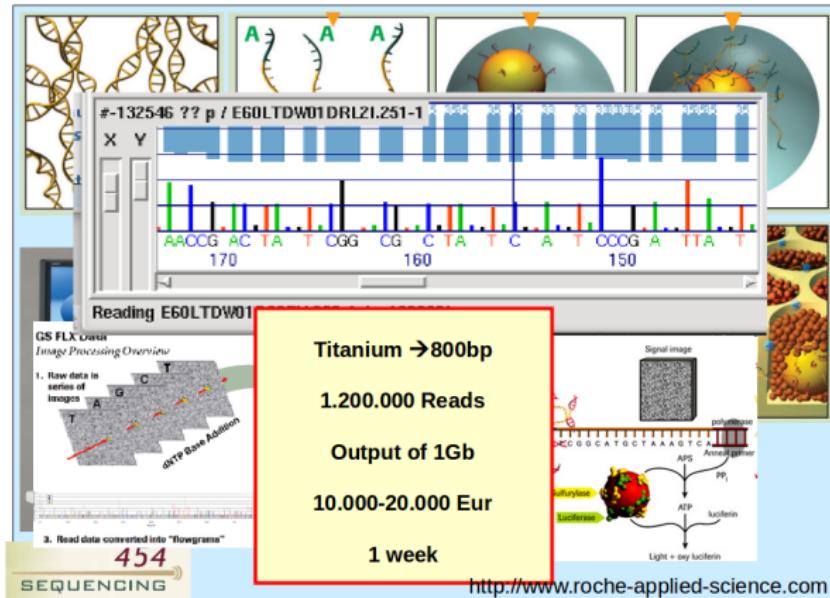
454 SEQUENCING



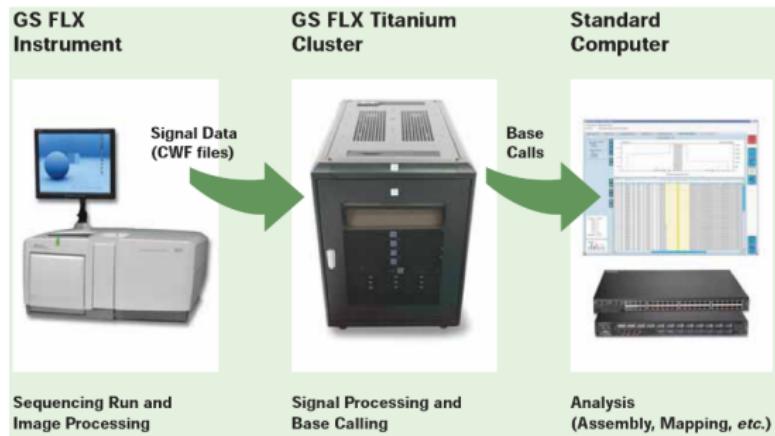
454 SEQUENCING



454 SEQUENCING



454 SEQUENCING



454 SEQUENCING PROBLEMS

SOMETIMES TWO OR MORE DIFFERENT FRAGMENTS BINDS THE BEAD

- **AmpliconNoise (open source)** is a collection of programs for the removal of noise from 454 sequenced PCR amplicons. It involves two steps the removal of noise from the sequencing itself and the removal of PCR point errors.

- ① Pyronoise (does flowgram-based clustering to spot misreads)
 - Calculate flowgram distances (PyroDist)
 - Complete linkage hierarchical clustering of flowgrams (FCluster)
 - Run EM algorithm (removes 454 error)
- ② SeqNoise (removes PCR point mutations)
- ③ Perseus (removes PCR chimeras without the need for a set of reference sequences)

454 SEQUENCING PROBLEMS

SOME TIMES A PCR PRODUCT IS GOING TO MOVE TO ANOTHER BEAD GENERATING ARTIFICIAL DUPLICATES. *Niu et al., BMC Bioinformatics 2010*

- Exact and nearly identical duplicates may make up 11~35% of the raw reads.
- In metagenomics, the amount of reads is used as an abundance measure, so artificial duplicates will introduce overestimation of abundance of taxon, gene, and function.
- **cdhit_454:** Identify artificial duplicates from metagenomic samples

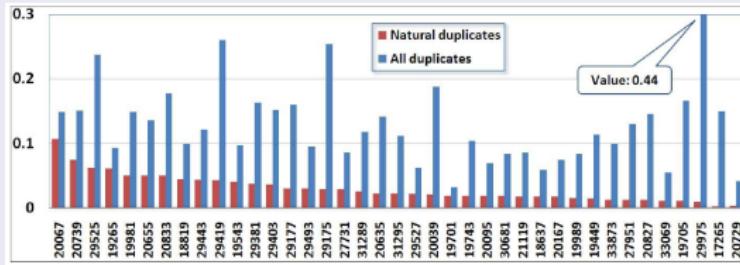
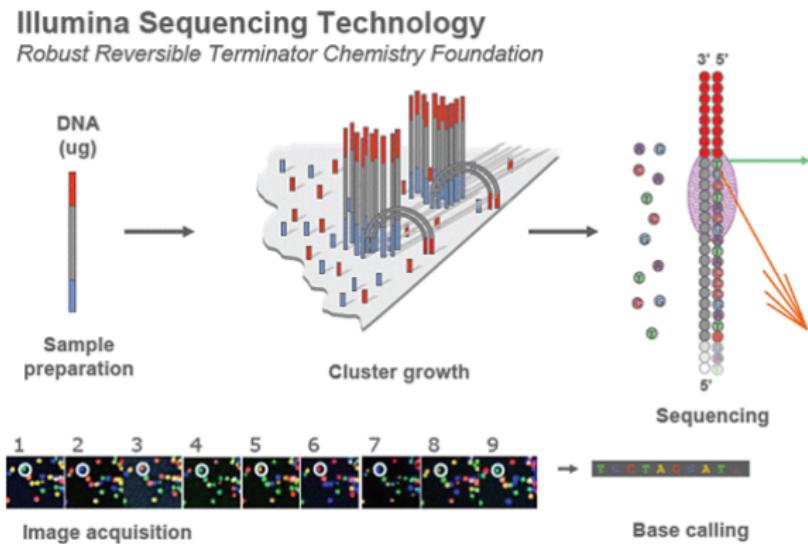


Figure 1 Ratio of all duplicates and average natural duplicates to all reads from genome projects. X-axis is project identifier of datasets, which are ordered by decreasing read density (number of reads divided by genome size). Y-axis is the ratio of duplicated reads to all reads.

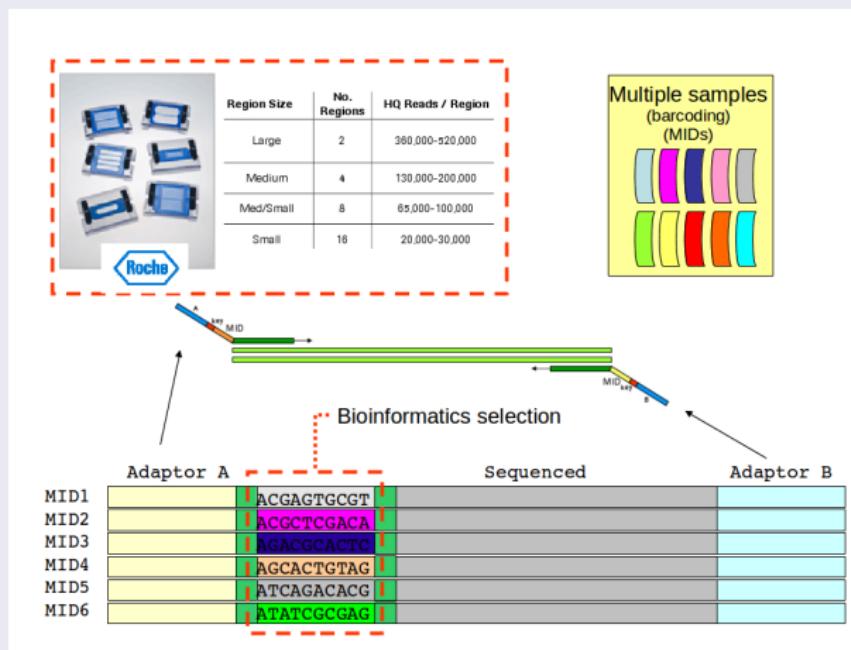
PYROSEQUENCING, SHOTGUN SEQUENCING, AND METATRANSCRIPTOMICS ANALYSIS

ILLUMINA SEQUENCING, JUST AN OVERVIEW



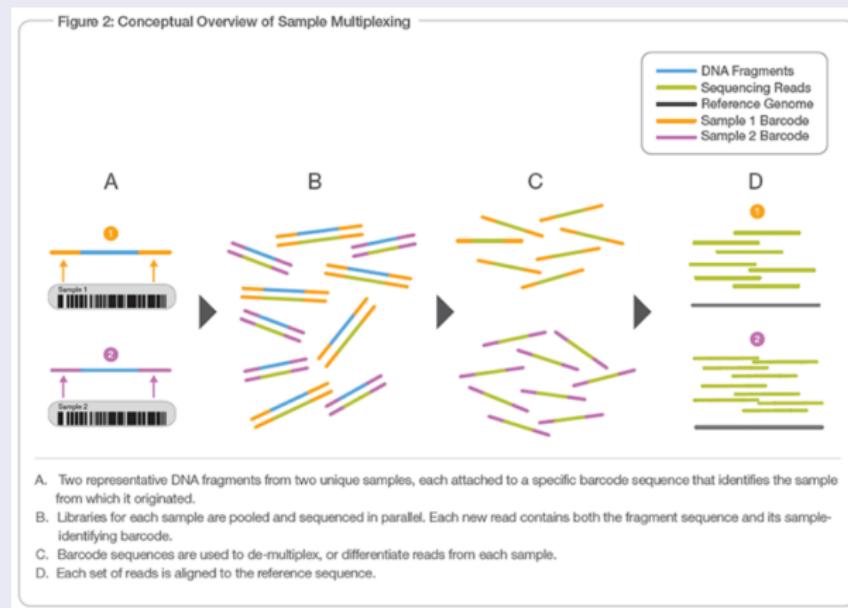
SEQUENCING - MULTIPLEXING

454 MULTIPLEXING



SEQUENCING - MULTIPLEXING

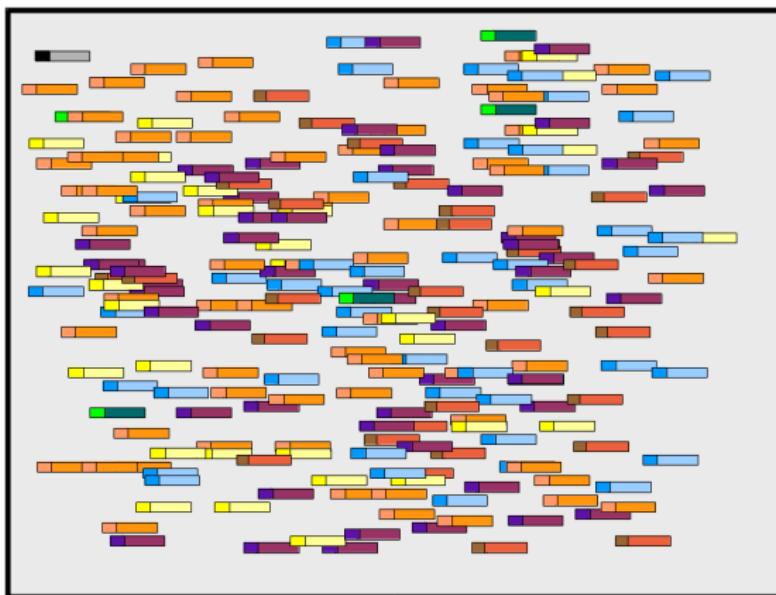
ILLUMINA MULTIPLEXING



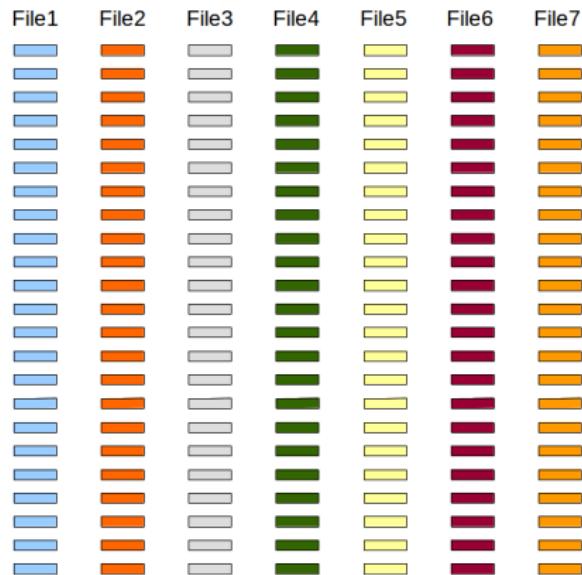
SEQUENCING - MULTIPLEXING

- Very important/common practice
- for 16S:
 - MiSeq: up to 384 samples pool
 - 454: less than 20 sample per 1/8 plate
- for metagenomics:
 - MiSeq: up to 40 (but it is already too much)
 - 454: one to ten metagenomes per run

MULTIPLEXING



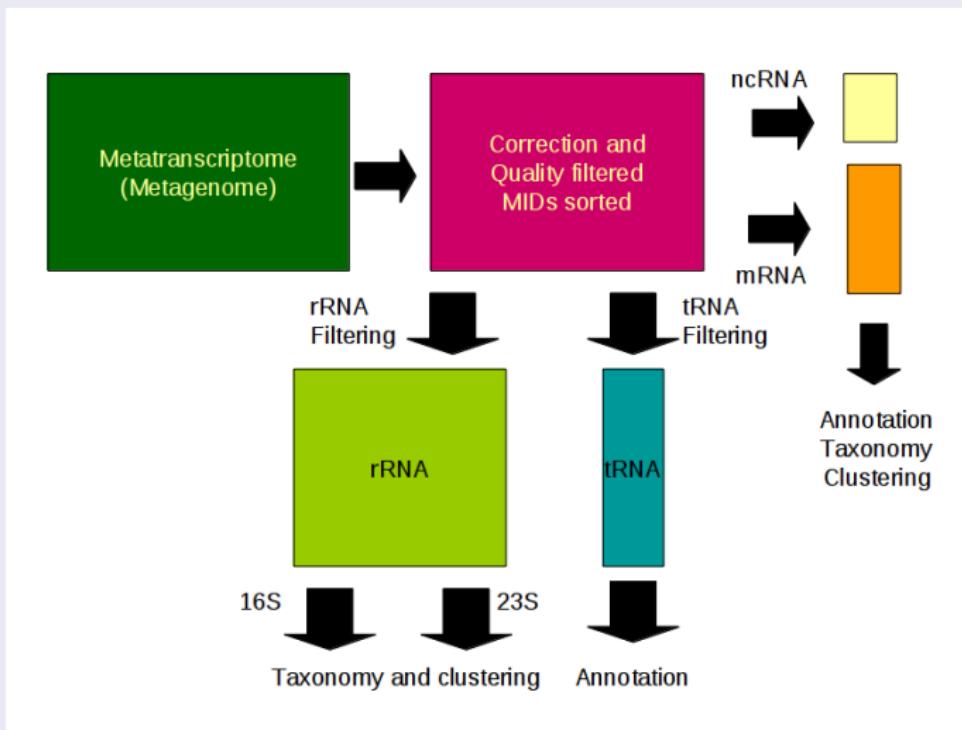
MULTIPLEXING



METAGENOMICS

METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES



METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES



METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES

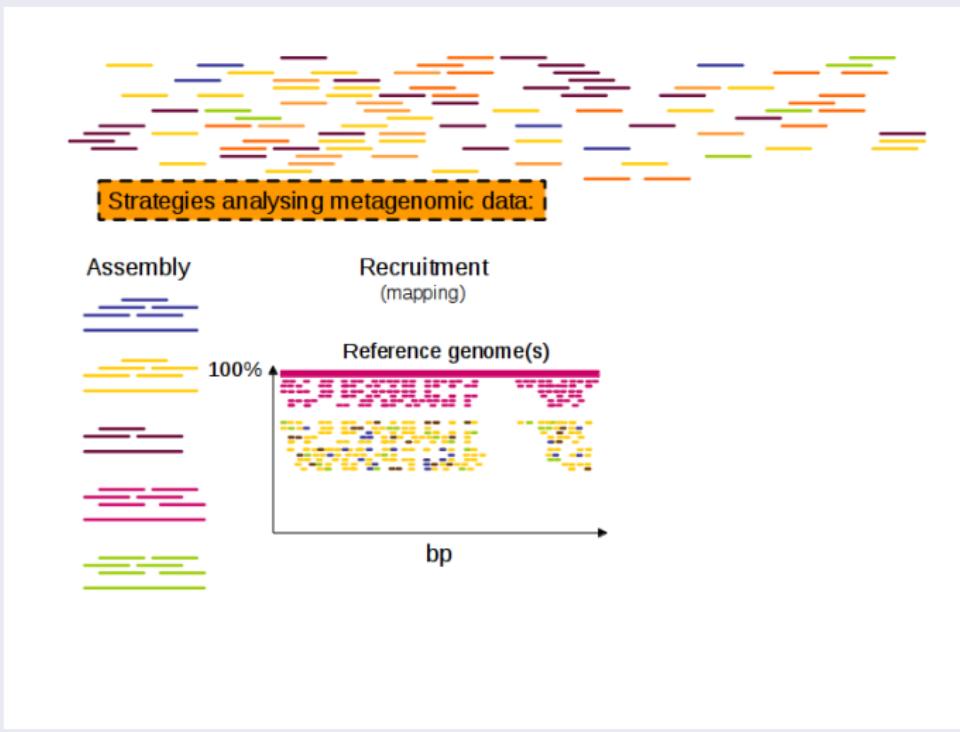


Assembly



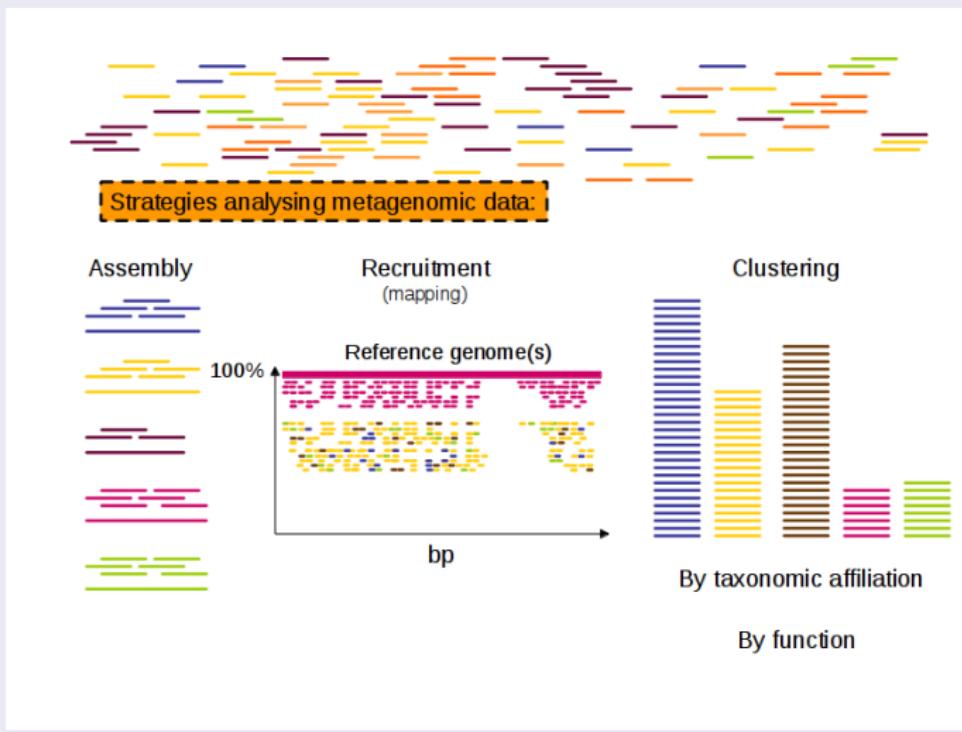
METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES

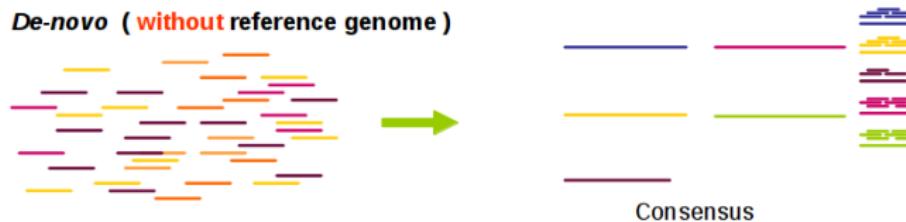


METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES



METEGENOMICS OVERVIEW: ASSEMBLY



METEGENOMICS OVERVIEW: ASSEMBLY

De-novo (without reference genome)

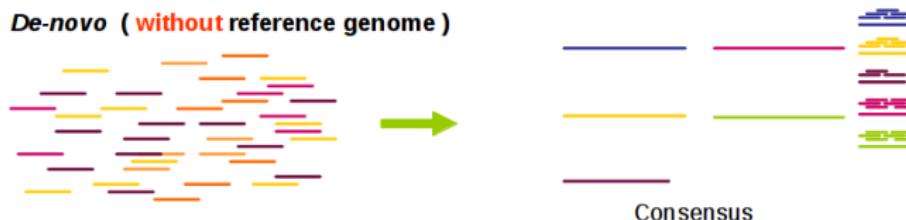


One of software outputs (Multi Fasta)

_____	>SEQ1 ATTACGATGGATCGATCGATCAGATAAATTGGATCAGTCAGCTAGTCGAT
_____	>SEQ2 CATGACTGCTAGCTAGCTAGTCGATCGTACGCTAGTAGCTAGCTAGC
_____	>SEQ3 ACTAGCTGATCGATGCCATGCGTAGCTACGTAGCTAGCTGTGTG
_____	>SEQ4 CATGATCGTACGTAGCTAGTCG
_____	>SEQ5 CATGACGTCAGTCATACGTCACTAGTAGACTGAC

METEGENOMICS OVERVIEW: ASSEMBLY

De-novo (without reference genome)



One of software outputs (Multi Fasta)

>SEQ1	ATTACGATGGATCGATCGATCAGATAAAATTGGATCAGTCAGCTAGTCGA
>SEQ2	CATGACATGCTAGCTAGTCGATCTACGCTAGTAGCTAGC
>SEQ3	ACTAGCTGATCGATGCCATGCGTAGCTACGTAGCTGTTG
>SEQ4	CATGATCGTACGTAGCTAGTCG
>SEQ5	CATGACGTCAGTCATACGTCAAGTAGACTGAC

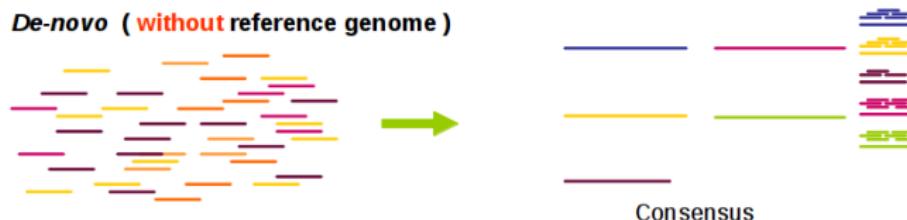
Prodigal
Glimmer
BLASTX

J. VALENCIANA

Fundación para el Fomento de la
Investigación Sanitaria y Biomédica
de la Comunitat Valenciana

METEGENOMICS OVERVIEW: ASSEMBLY

De-novo (without reference genome)



One of software outputs (Multi Fasta)

>SEQ1	ATTACGATGGATCGATCGATCAGATAAAATTGGATCAGTCAGCTAGTCGA
>SEQ2	CATGACATGCTAGCTAGTCGATCTACGCTAGTAGCTAGC
>SEQ3	ACTAGCTGATCGATGCCATGCGTAGCTACGTAGCTAGCTGTTG
>SEQ4	CATGATCGTACGTAGCTAGTCG

**Prodigal
Glimmer**

We almost lost coverage informations We have more info for annotation

CATGACGCTACGTACGTAGCTAGTCG

METEGENOMICS OVERVIEW: ASSEMBLY

```
# from a multifasta
>seq1
atgtcagtgcatacgtaacgtacgatcgag
>seq2
gcatacgatgcatacgtagctacgtacgtg
>seq3
cgatcgatcgatcgatgcatacgatcgatcg
>seqN
```



**A small demonstration in
the methodological section**

METEGENOMICS OVERVIEW: ASSEMBLY

STRAIN-RESOLVED COMMUNITY GENOMIC ANALYSIS OF GUT MICROBIAL COLONIZATION IN A PREMATURE INFANT MOROWITZ ET AL., 2011

Real Results.....

Real Results.....			
NCBI Whole Genome Shotgun sequencing project			
WGS record Assembly Show		List of all WGS projects	
What can be inferred in this field?			
AEQT000000000_1 Human gut metagenome			
Contig:	1,155	Show download	
Updated:	04-MAR-2011	HG-STR	
Length:	10	Assembly ID:	
Organism:	Human gut metagenome		
Biosource:	<i>host</i> _type = genetic DNA <i>isolation_source</i> = fecal sample from premature female infant <i>host</i> = Homo sapiens <i>isolate</i> = fecal sample <i>note</i> = metagenomic		
Keywords:	WGS		
WGS:	AEQT000000000_AEQT0001155		
<i>Statement of scientific validity: genomic analysis of gut microbial colonization in a premature infant.</i> Proc Natl Acad Sci U S A. 108(3), 1020-5. [PubMed: 21297941]			
Submitted:	04-MAR-2011	Earth and Planetary Sciences, University of California - Berkeley, 309 Hickey Hall, Berkeley, CA 94720, USA	
Note authors: Horwitz, J.A., Denner, C.J., Comstock, E., Thomas, C.K., Relman, D.A., and Bäckström, L.H.			
View Print Email Help Contigs 21-40 of 1,155		Search in project Show	
Accession	Length	Start	End
AEQT000000000_1_1	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_2	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_3	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_4	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_5	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_6	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_7	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_8	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_9	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_10	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_11	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_12	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_13	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_14	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_15	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_16	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_17	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_18	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_19	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_20	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_21	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_22	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_23	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_24	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_25	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_26	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_27	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_28	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_29	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_30	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_31	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_32	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_33	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_34	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_35	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_36	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_37	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_38	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_39	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_40	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_41	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_42	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_43	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_44	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_45	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_46	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_47	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_48	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_49	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_50	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_51	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_52	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_53	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_54	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_55	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_56	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_57	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_58	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_59	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_60	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_61	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_62	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_63	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_64	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_65	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_66	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_67	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_68	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_69	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_70	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_71	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_72	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_73	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_74	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_75	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_76	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_77	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_78	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_79	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_80	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_81	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_82	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_83	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_84	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_85	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_86	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_87	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_88	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_89	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_90	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_91	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_92	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_93	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_94	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_95	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_96	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_97	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_98	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_99	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_100	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_101	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_102	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_103	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_104	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_105	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_106	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_107	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_108	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_109	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_110	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_111	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_112	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_113	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5,167
AEQT000000000_1_114	5,167	1	5,167
Chrobovir sp. UCIC074	5,167	1	5

METEGENOMICS OVERVIEW: ASSEMBLY

STRAIN-RESOLVED COMMUNITY GENOMIC ANALYSIS OF GUT MICROBIAL COLONIZATION IN A PREMATURE INFANT MOROWITZ ET AL., 2011

Real Results....

Real Results.....

GenBank: AE0701000000
Title: Citrobacter sp. UC1C1Tii UC1C1Tii_01048, whole genome shotgun sequence

Project ID: AE0701000000
Title: Citrobacter sp. UC1C1Tii UC1C1Tii_01048, whole genome shotgun sequence

Organism: Citrobacter sp. UC1C1Tii (Unknown)

Accession: AE0701000000
Length: 1271 bp
Start: 1171
End: 2342
E-value: 0.00001
Bit score: 1000
Source: GenBank
Database: GenBank
Taxonomy: Bacteria; Gammaproteobacteria; Enterobacteroiaceae; Enterobacter; Citrobacter; Citrobacter

Authors: Farhadianpour, M.; Rasmussen, C.E.; Petroski, T.; Reichenbach, J.B., and Paulsen, I.K.

Journal: Proc Natl Acad Sci U S A. 2011 Dec 13;108(50):21139-21143. Epub 2011 Nov 15.
PMID: 22100000
DOI: 10.1073/pnas.1116601108

Abstract: We report the genome of *Citrobacter* sp. UC1C1Tii, isolated from a patient with nosocomial pneumonia. Analysis of gas-tight occlusion assays in a primary isolate revealed a minimum inhibitory concentration of 16 μg/ml amikacin. This is the first genome of a *Citrobacter* sp. strain isolated from a patient.

BLASTN search results:

1 hits found

1. Citrobacter sp. UC1C1Tii UC1C1Tii_01048, whole genome shotgun sequence

Search in project: []

GIUSEPPE D'AURIA

PYROSEQUENCING, SHOTGUN SEQUENCING, AND METATRANSCRIPTOMICS ANALYSIS Morowitz M J et al. PNAS 2011;

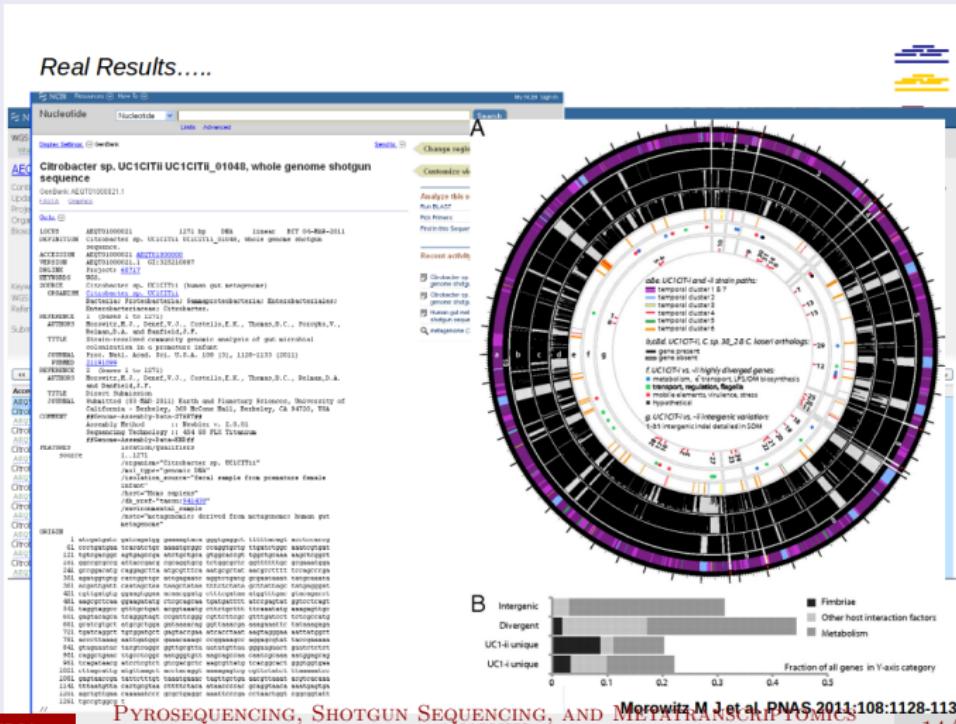
Morowitz M J et al. PNAS 2011;108:1128-1133

14/10/2015 28 / 39

METEGENOMICS OVERVIEW: ASSEMBLY

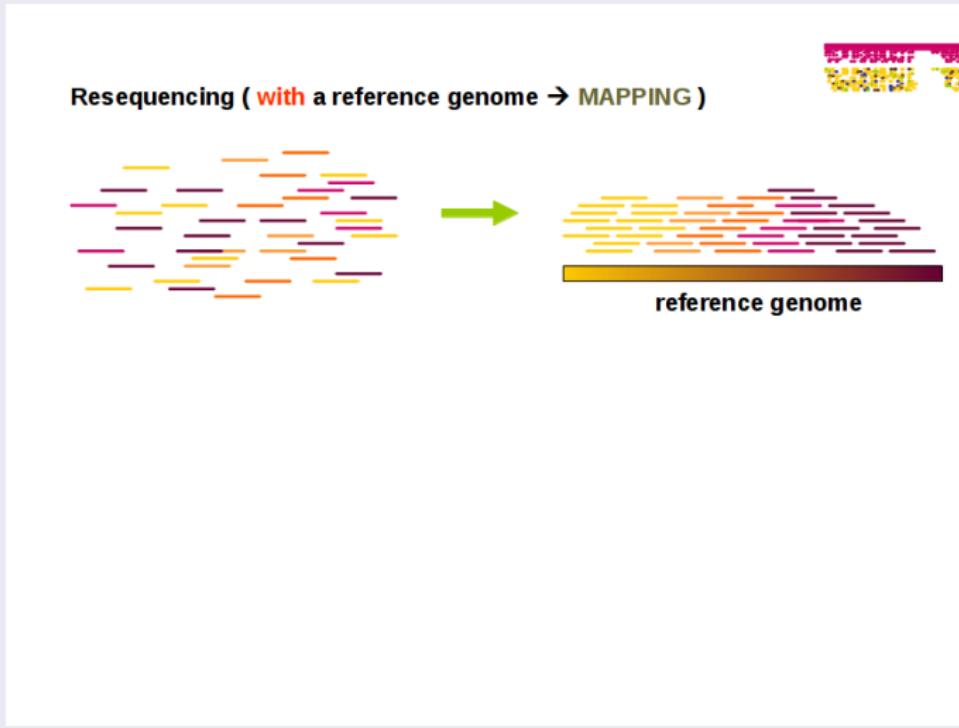
STRAIN-RESOLVED COMMUNITY GENOMIC ANALYSIS OF GUT MICROBIAL COLONIZATION IN A PREMATURE INFANT MOROWITZ ET AL., 2011

Real Results.....



METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES: MAPPING



METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES: MAPPING

Recruitment strategy

```
actagcttagctgatcgatgcitagc  
tgactgatcgtaaccatcgatgt  
agcttagtcgatcgtagcttagtgcg  
tcgatcgtagtcgatcgatgttag  
ctagctgactgatgttagcatctt  
tatcgccggccgcgcggcatagct  
agtcaatcgatcgtagcttagctag  
gcatcgatgcat.....
```



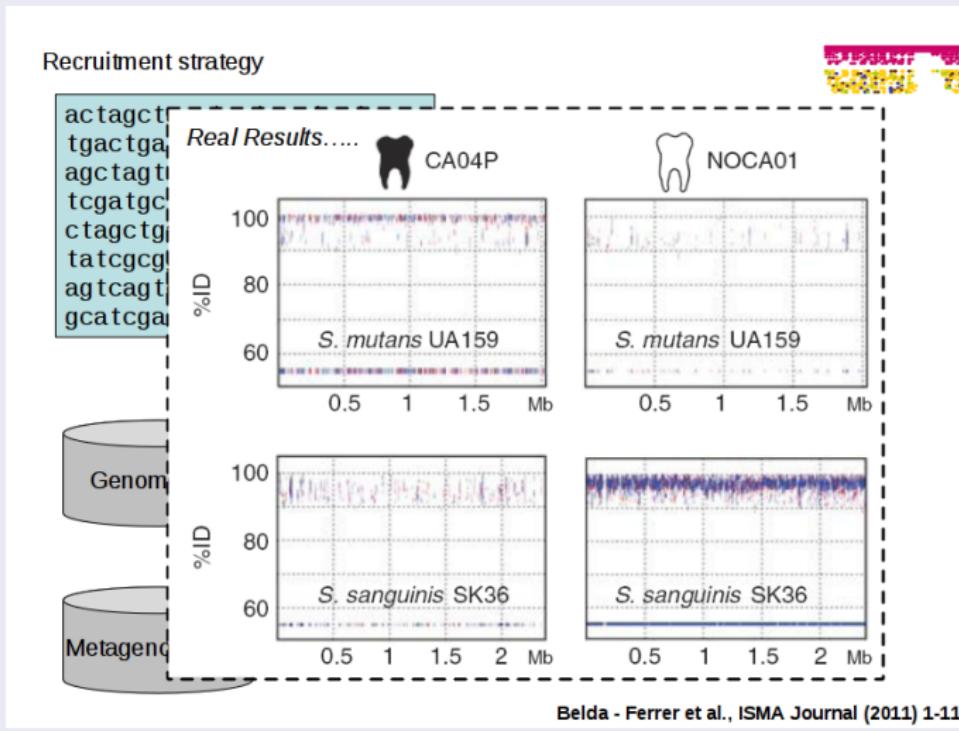
What's that?
Just a multifasta
collection of
complete genomes



Same as before but
metagenomes

METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES: MAPPING



METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES: CLUSTERING

Similarity based

CDHIT

BLASTCLUST

DOTUR

UCLUST

Etc

Database driven

BLAST

BLAST-LCA

RDP classifier

Etc...

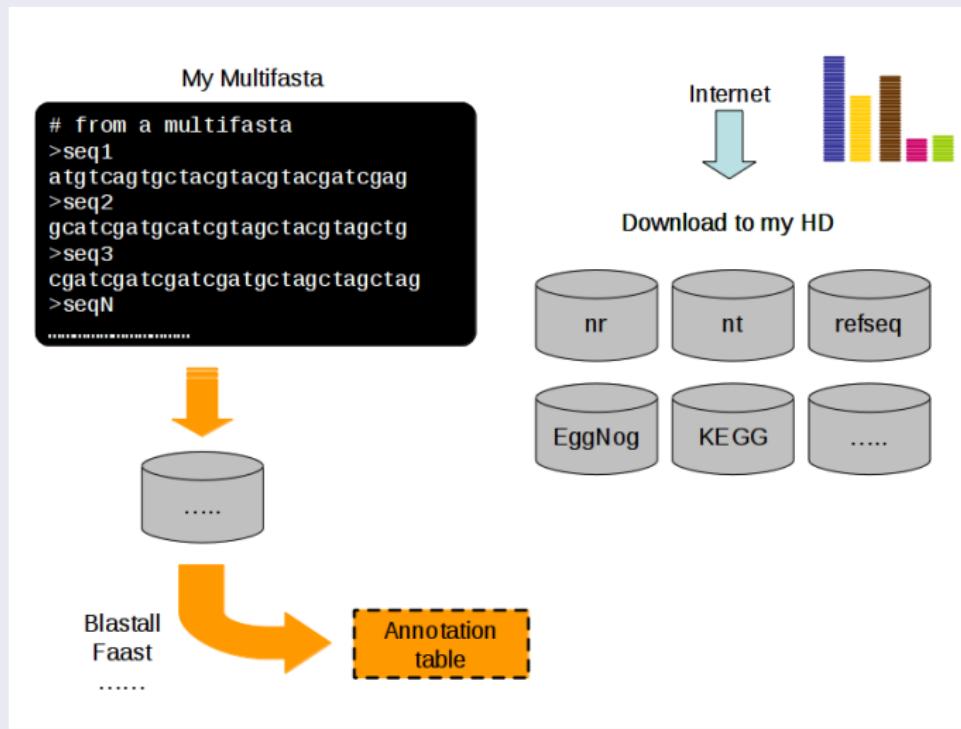


Clusters born from the proper dataset

Sequences are, in some way, labelled and labels will represents the basis for clustering

METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES: CLUSTERING



METEGENOMICS OVERVIEW

METEGENOMICS: STRATEGIES: CLUSTERING

16S based taxonomy

RDP classifier (<http://rdp.cme.msu.edu>)

```
java -jar /rdp_classifier-2.2.jar -q infile -o test
# from a multifasta
```

```
S002033089 Root norank 1.0 Bacteria domain 1.0 Proteobacteria phylum 1.0 Epsilonproteobacteria class 1.0 Campylobacterales order 1.0 Campylobacteraceae family 1.0 Arcobacter
S000379951 Root norank 1.0 Bacteria domain 1.0 Proteobacteria phylum 1.0 Epsilonproteobacteria class 1.0 Campylobacterales order 1.0 Campylobacteraceae family 1.0 Arcobacter
S002046413 Root norank 1.0 Bacteria domain 1.0 Proteobacteria phylum 1.0 Epsilonproteobacteria class 1.0 Campylobacterales order 1.0 Campylobacteraceae family 1.0 Arcobacter
S000900388 Root norank 1.0 Bacteria domain 1.0 Proteobacteria phylum 1.0 Epsilonproteobacteria class 1.0 Campylobacterales order 1.0 Campylobacteraceae family 1.0 Arcobacter
```



Functional clusters and annotation

There are several systems

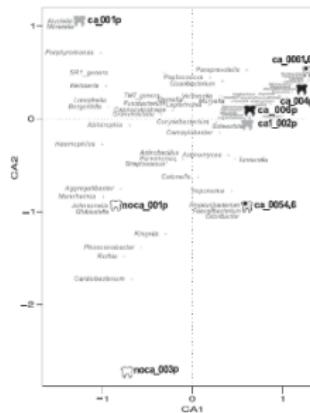
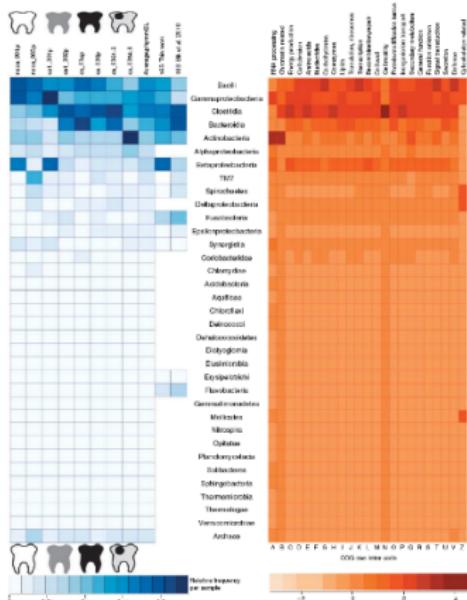
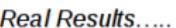
Most of them are based in homology searches vs reference databases

Most of them are freely accessible via web

KEGG, COG, RefSeq, Pfam, Tigrfam, CDD, EggNog

METEGENOMICS OVERVIEW

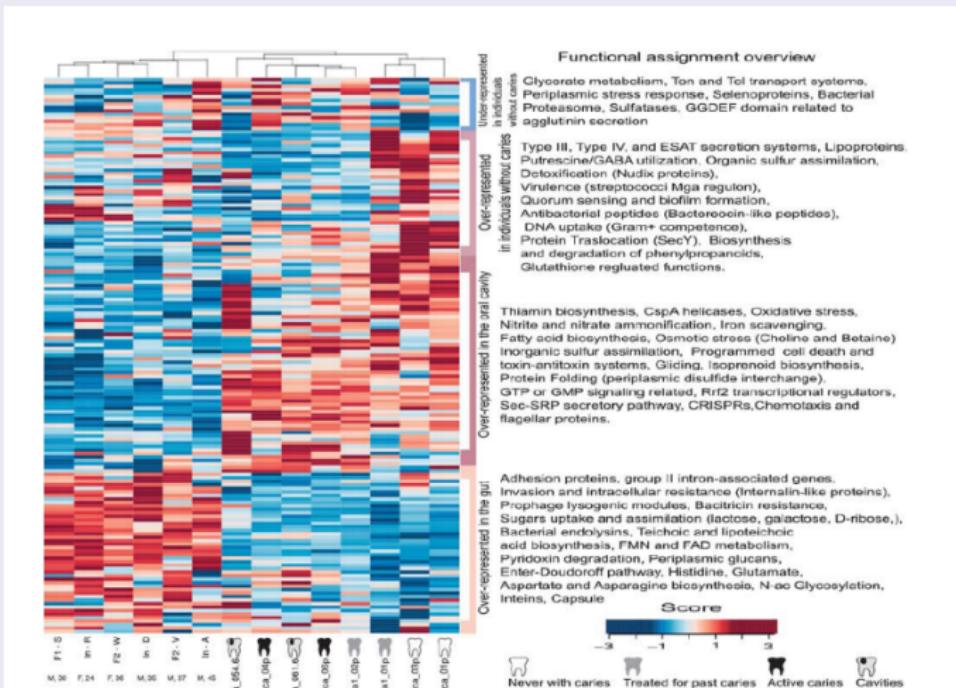
THE ORAL METAGENOME IN HEALTH AND DISEASE BELDA-FERRE ET AL., 2010



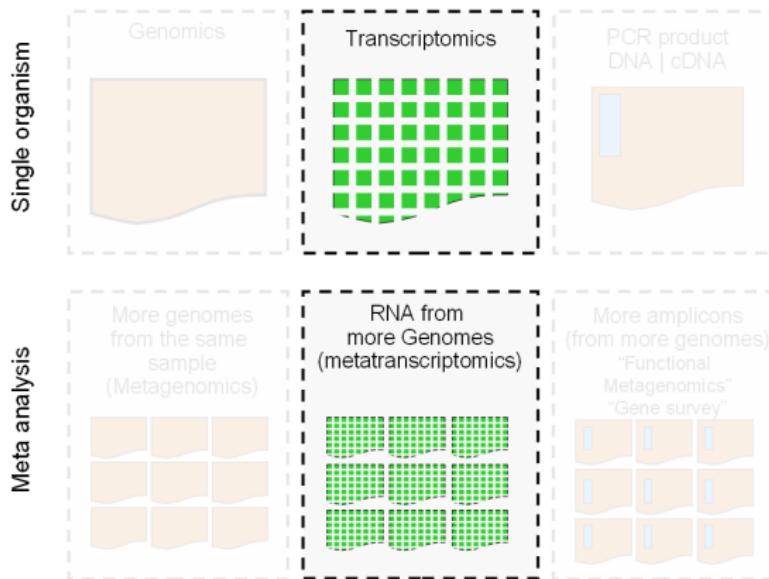
16S based Correspondance analysis

METEGENOMICS OVERVIEW

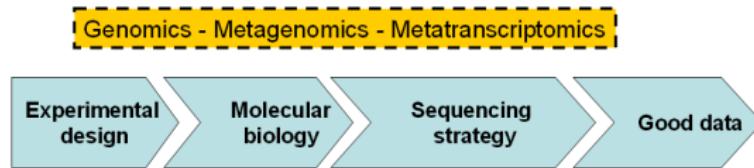
THE ORAL METAGENOME IN HEALTH AND DISEASE BELDA-FERRE ET AL., 2010



METETRANSCRIPTOMICS OVERVIEW



METETRANSCRIPTOMICS OVERVIEW

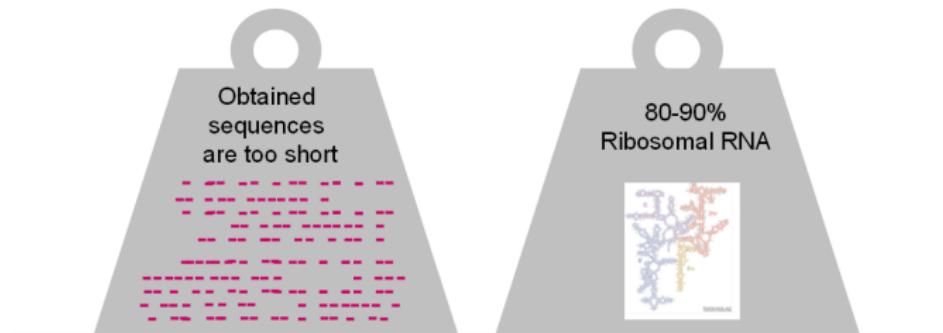


METETRANSCRIPTOMICS OVERVIEW

Genomics - Metagenomics - Metatranscriptomics



Metatranscriptomics: Two heavy problems

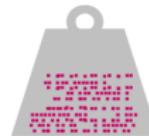


METATRANSCRIPTOMICS OVERVIEW

RIBOSOMAL RNA REMOVAL

Short sequences

This is mainly a methodological problem.
It relies on the protocol for retrotranscription.



Random priming



Retrotranscription by adaptors



METETRANSCRIPTOMICS OVERVIEW

RIBOSOMAL RNA REMOVAL

Ribosomal

Several kits exists trying to reduce rRNA amount

Mostly based on affinity (subtraction),
rRNA terminal digestion (restriction).

The MICROBExpress (Ambion)

mRNA-ONLY (Epicentre's)

METETRANSCRIPTOMICS OVERVIEW

RIBOSOMAL RNA REMOVAL

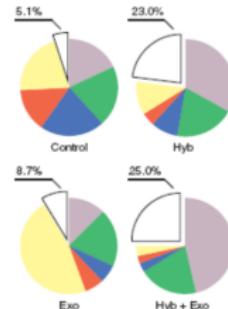
Validation of two ribosomal RNA removal methods for microbial metatranscriptomics

Shaomei He^{1,2,5}, Omri Wurtzel^{3,5}, Kanwar Singh¹, Jeff L Froula¹, Suzan Yilmaz¹, Susannah G Tringe¹, Zhong Wang¹, Feng Chen¹, Erika A Lindquist¹, Rotem Sorek³ & Philip Hugenholtz^{1,2,4}

Repeatability (by protocol and run)

a

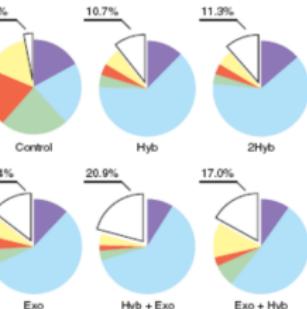
- Total non-rRNA
- Streptomyces rRNA
- Lactobacillus rRNA
- Desulfobacter rRNA
- Spirochaeta rRNA



rRNA removal

b

- Total non-rRNA
- *Candidatus rRNA*
- *Kangeliella rRNA*
- *Halorubidus rRNA*
- *Spirochaeta rRNA*



Methods, 2010

METETRANSCRIPTOMICS OVERVIEW

RIBOSOMAL RNA NORMALIZATION

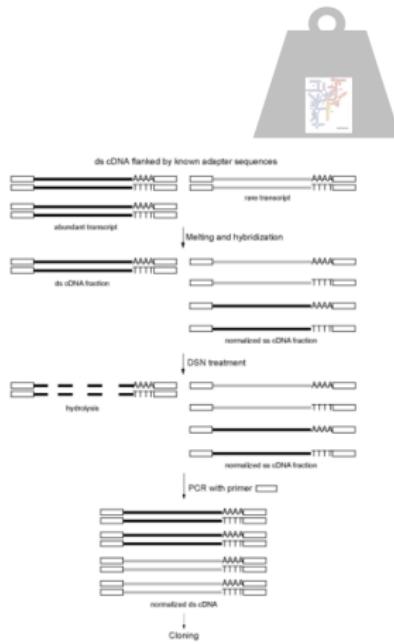
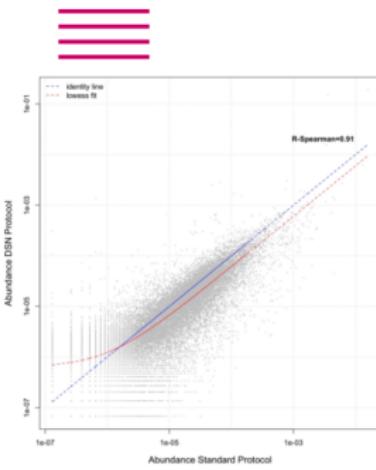
Other option is “normalization”



METETRANSCRIPTOMICS OVERVIEW

RIBOSOMAL RNA NORMALIZATION

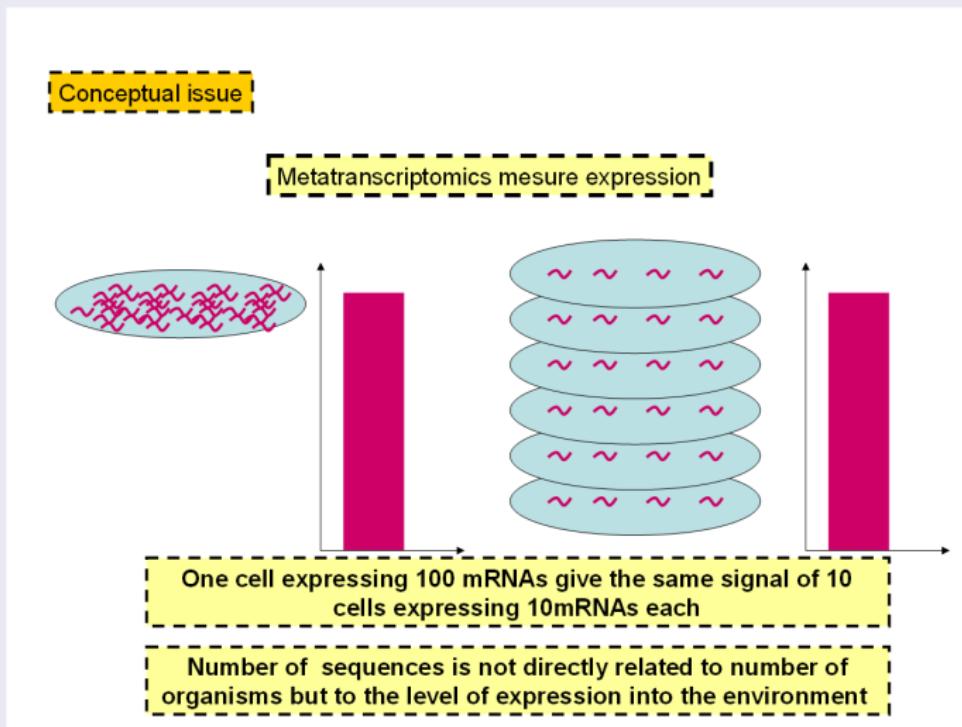
Other option is "normalization"



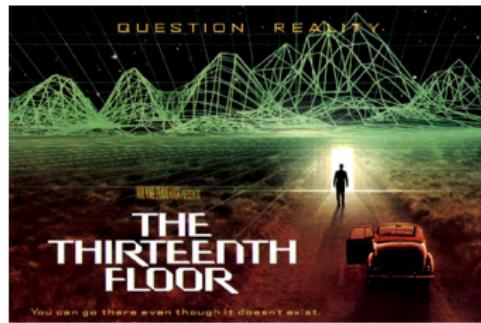
<http://www.prognosysbio.com/>

METETRANSCRIPTOMICS OVERVIEW

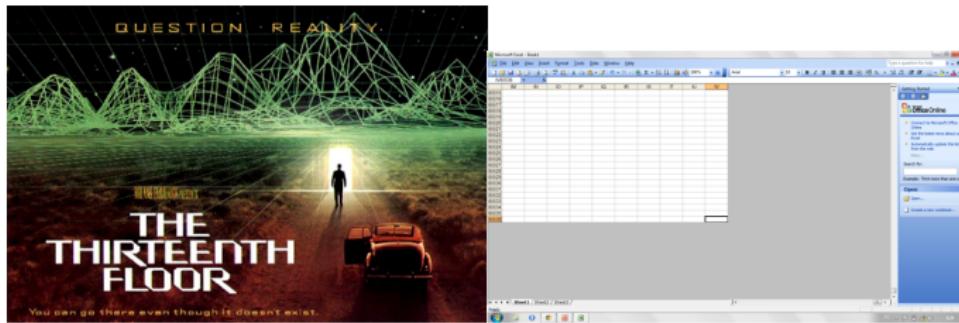
A CONCEPTUAL PROBLEM TO TAKE ALWAYS IN ACCOUNT



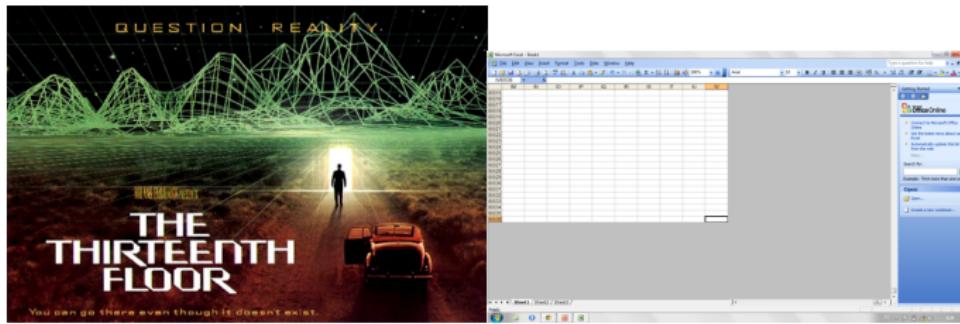
ADVICE: FOR INTREPID AND BRAVE PEOPLE



ADVICE: FOR INTREPID AND BRAVE PEOPLE



ADVICE: FOR INTREPID AND BRAVE PEOPLE



BIO-INFORMATICS VS INFO-BILOGISTS

- There is a continuous need of people with enough skills for managing big data projects...
- Well, when working with NGS we **ALREADY KNOW** that we are working with ***Giga-DATA***
- Forget about old offimatic world!! we are running out of memory/space/time.

ADVICE: FOR INTREPID AND BRAVE PEOPLE

GENOMICS/METAGENOMICS/WHATEVER RELATED WITH BIO-X



Perl is a scripting language widely used for system administration and programming on the World Wide Web.

It originated in the UNIX community and has a strong UNIX slant, but usage on Windows has grown rapidly.

ActivePerl is a quality-assured binary distribution of Perl for popular UNIX platforms and Windows.

perl (small 'p') is the program used to interpret the Perl language.

ADVICE: FOR INTREPID AND BRAVE PEOPLE

GENOMICS/METAGENOMICS/WHATEVER RELATED WITH BIO-X

<http://www.r-project.org/>

The R Project for Statistical Computing

PCA: 5 vars
princomp(x = data[, -c(1:2)])

Factor 1 [41%] Factor 3 [19%]

V. De Geerne

Clustering: 4 groups

Groups: 28 16 1 2

About R
What is R?
Contributors
Screenshots
What's new?

Download,
Packages
CRAN

R Project
Foundation
Members & Donors
Mailing Lists
Bug Tracking
Developer Page
Conferences
Search

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and Mac OS.

PEOPLE

**Foundation for the Promotion of
Health and Biomedical Research of
Valencian Region
(FISABIO)**



**Cavanilles Institute for Biodiversity
and Evolutionary Biology
(University of València)**



Maria Dzunkova



Thank You for your attention

```
ls -ltr
```