

# Preliminary 16S Analysis of samples from Montevideo, MetaSEW and MetaBEA

Gaston H. Gonnet  
Informatik, ETH Zurich  
MetaSUB Summit, July 1<sup>st</sup>, 2016



# The Data

# The data: collection

For MetaSEW the data was collected by the IMM (owns, controls and operates the Montevideo's sewage system)

For MetaBEA, the IMM collects samples regularly and we also collected our own samples.

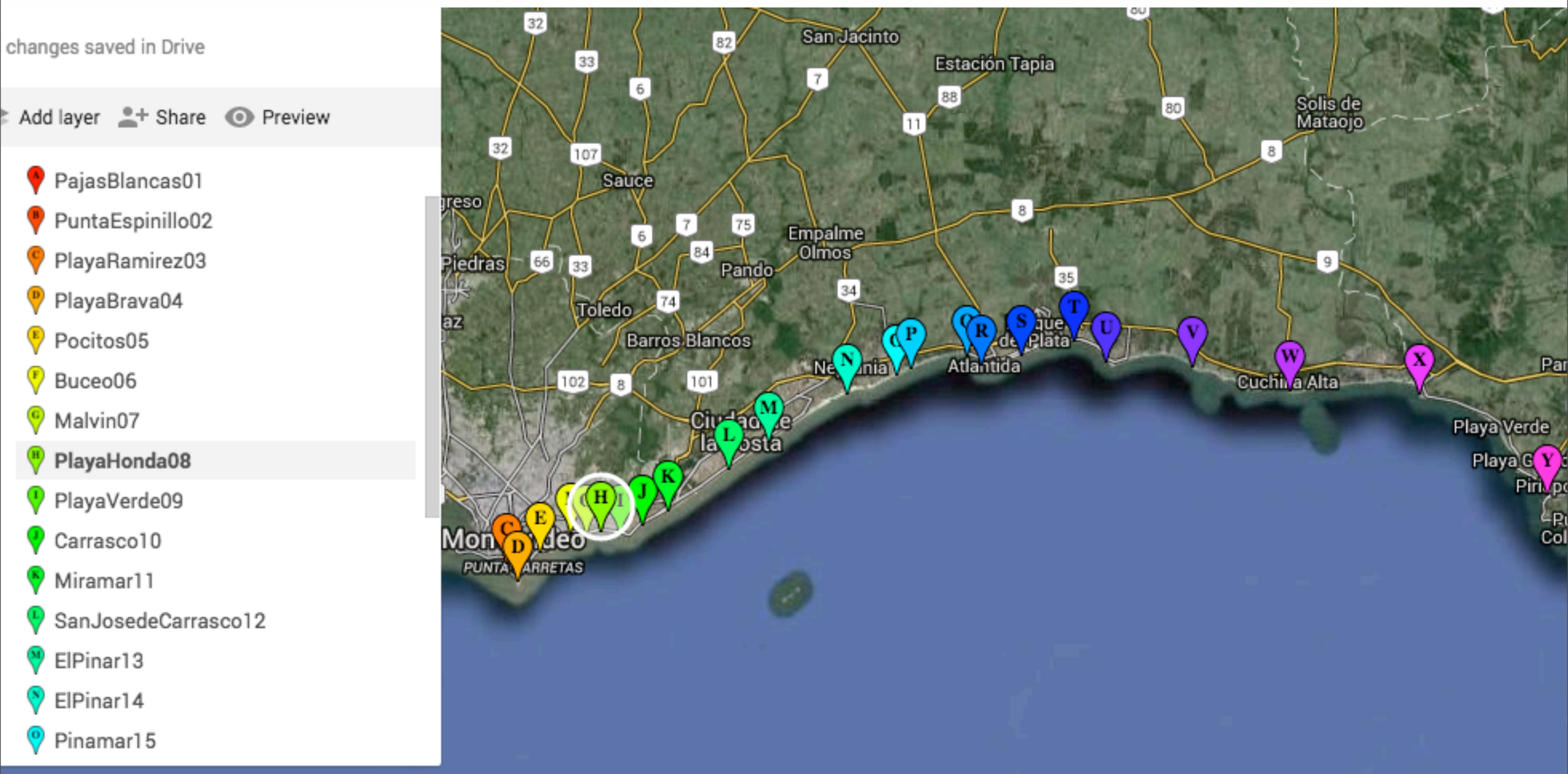


# The data: collection



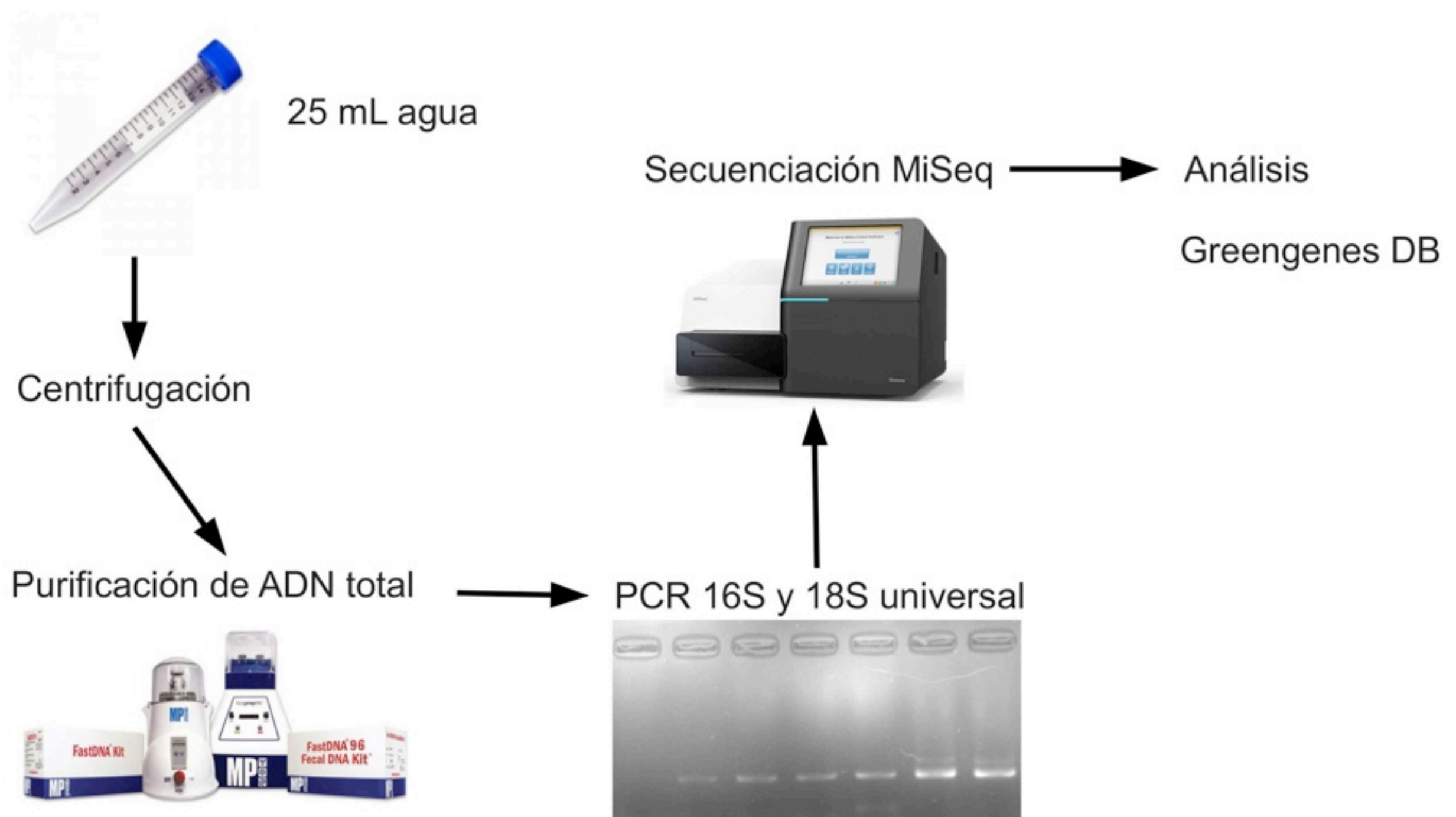


# The data: MetaBEA - sampling map



# The data: 16S and 18S analysis

## Procedimiento



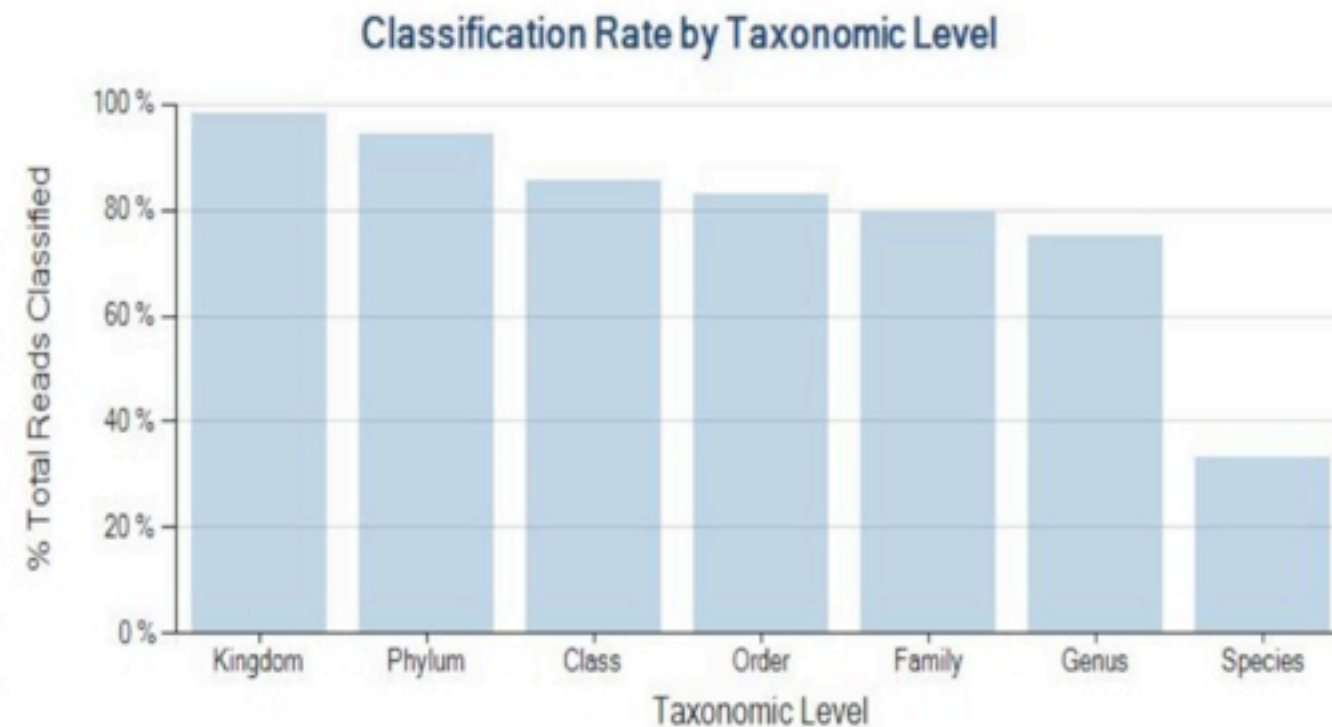
# Preliminary results



# Preliminary results

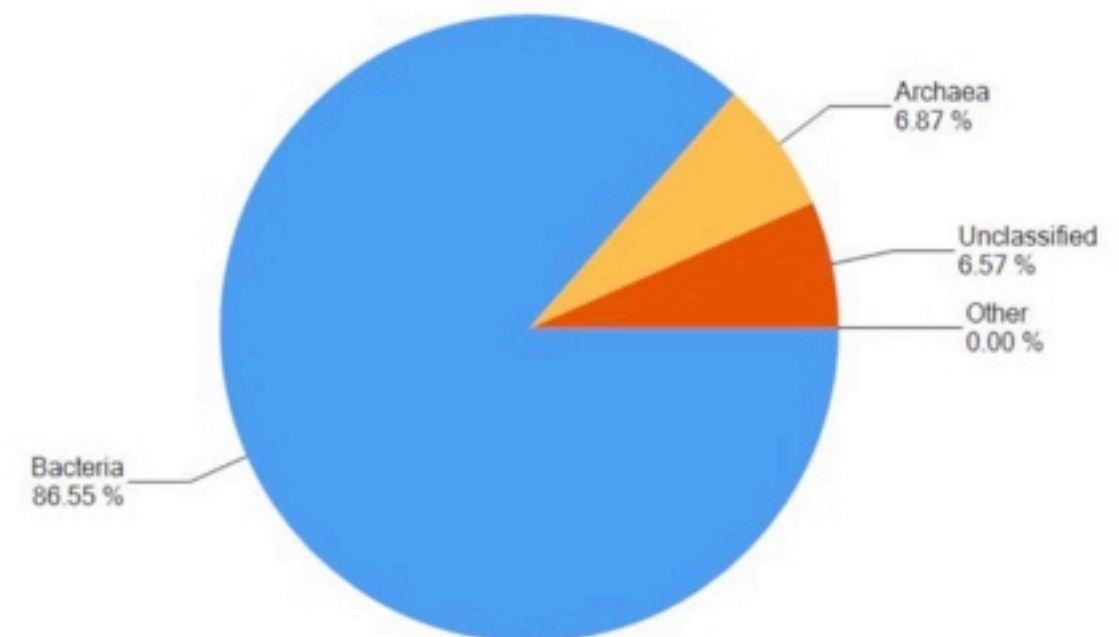
Total Reads	Reads Passing Quality Filtering	% Reads Passing Quality Filtering
562,015	535,192	95.2 %

Taxonomic Level	Reads Classified to Taxonomic Level	% Total Reads Classified to Taxonomic Level
Kingdom	526,088	98.30 %
Phylum	503,088	94.00 %
Class	458,546	85.68 %
Order	443,614	82.89 %
Family	425,424	79.49 %
Genus	402,176	75.15 %
Species	176,710	33.02 %



## Top Species Classification Results

Classification	Number of Reads	% Total Reads
Unclassified at Species level	358,482	66.98 %
<i>Nitrosopumilus maritimus</i>	35,816	6.69 %
<i>Nitrosococcus watsoni</i>	10,094	1.89 %
<i>Pseudoalteromonas issachenkonii</i>	7,968	1.49 %
<i>Methylostratum kenyense</i>	7,466	1.40 %
<i>Psychrobacter cibarius</i>	7,018	1.31 %
<i>Ectothiorhodospira haloalkaliphila</i>	5,807	1.09 %
<i>Aquimarina macrocephali</i>	3,596	0.67 %



Total Species-level Taxonomic Categories Identified: 987. This table shows the top 8 of 987 classifications.



# Preliminary results

## Resultados

Classification	Number of Reads	% Total Reads
Unclassified at Species level	165,678	36.94 %
<i>Pseudomonas lundensis</i>	135,371	30.18 %
<i>Akkermansia muciniphila</i>	14,860	3.31 %
<i>Thiohalorhabdus denitrificans</i>	6,183	1.38 %
<i>Lactobacillus mucosae</i>	5,904	1.32 %
<i>Alkaliphilus crotonatoxidans</i>	5,777	1.29 %
<i>Methanosaeta concilii</i>	4,674	1.04 %
<i>Chryseobacterium taichungense</i>	4,651	1.04 %

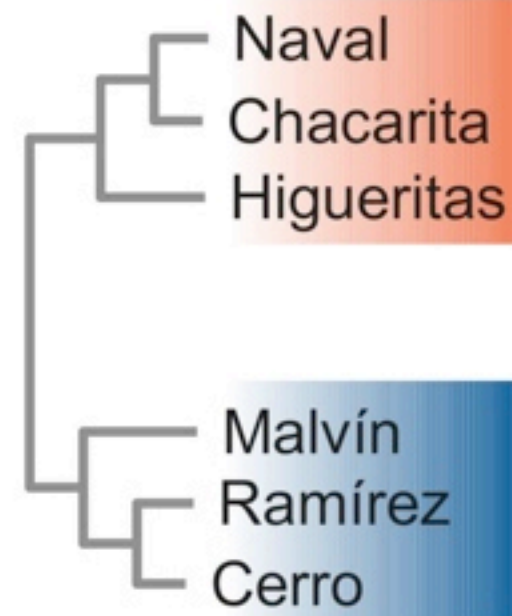
Classification	Number of Reads	% Total Reads
Unclassified at Species level	242,748	35.63 %
<i>Pseudomonas lundensis</i>	226,010	33.17 %
<i>Akkermansia muciniphila</i>	20,693	3.04 %
<i>Streptococcus fryi</i>	14,730	2.16 %
<i>Alkaliphilus crotonatoxidans</i>	5,702	0.84 %
<i>Thiohalorhabdus denitrificans</i>	5,692	0.84 %
<i>Bifidobacterium stercoris</i>	4,997	0.73 %
<i>Pseudomonas azotoformans</i>	4,907	0.72 %

Classification	Number of Reads	% Total Reads
Unclassified at Species level	358,482	66.98 %
<i>Nitrosopumilus maritimus</i>	35,816	6.69 %
<i>Nitrosococcus watsoni</i>	10,094	1.89 %
<i>Pseudoalteromonas issachenkonii</i>	7,968	1.49 %
<i>Methylobacterium kenyaense</i>	7,466	1.40 %
<i>Psychrobacter cibarius</i>	7,018	1.31 %
<i>Ectothiorhodospira haloalkaliphila</i>	5,807	1.09 %
<i>Aquimarina macrocephali</i>	3,596	0.67 %

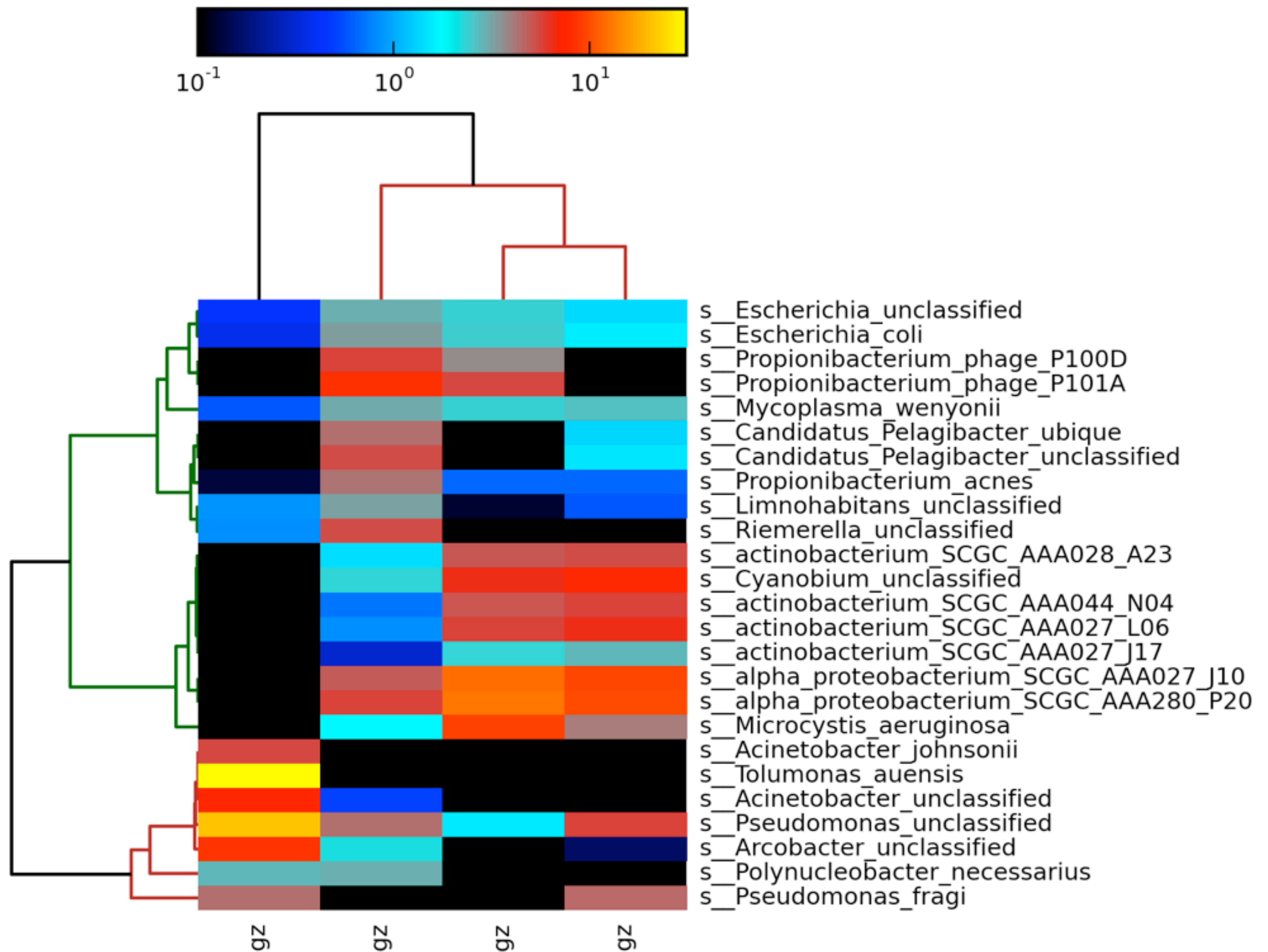
Classification	Number of Reads	% Total Reads
<i>Pseudomonas lundensis</i>	226,237	36.99 %
Unclassified at Species level	189,392	30.96 %
<i>Streptococcus fryi</i>	17,663	2.89 %
<i>Akkermansia muciniphila</i>	17,425	2.85 %
<i>Serratia entomophila</i>	7,492	1.22 %
<i>Megasphaera hominis</i>	6,858	1.12 %
<i>Acidovorax temperans</i>	6,682	1.09 %
<i>Chryseobacterium taichungense</i>	5,809	0.95 %

Classification	Number of Reads	% Total Reads
Unclassified at Species level	465,682	65.18 %
<i>Nitrosopumilus maritimus</i>	46,415	6.50 %
<i>Ruegeria lacuscaerulensis</i>	12,730	1.78 %
<i>Thiothrix fructosivorans</i>	10,314	1.44 %
<i>Aquimarina macrocephali</i>	10,259	1.44 %
<i>Nitrosococcus watsoni</i>	7,046	0.99 %
<i>Nisaea nitritireducens</i>	6,021	0.84 %
<i>Planomicrobium alkanoclasticum</i>	5,553	0.78 %

Classification	Number of Reads	% Total Reads
Unclassified at Species level	476,963	72.80 %
<i>Alteromonas litorea</i>	15,520	2.37 %
<i>Nitrosopumilus maritimus</i>	15,219	2.32 %
<i>Ruegeria lacuscaerulensis</i>	13,650	2.08 %
<i>Hymenobacter xinjiangensis</i>	9,203	1.40 %
<i>Aquimarina macrocephali</i>	7,476	1.14 %
<i>Arcobacter defluvi</i>	6,455	0.99 %
<i>Thiothrix fructosivorans</i>	5,168	0.79 %



# Preliminary results (kindly provided by Ebrahim)





# Plan



# Plan

How can we make this useful?

# Plan: depth

We need to estimate the depth of sequencing, that is the expected number of times that each DNA position will be sequenced.

This will depend on the relative representation precision that is desirable and on a model of the distribution of representation of species.

# Plan: rarity

We are interested in species that will appear with very low frequency, i.e. in the samples they are going to be extremely rare.



# Plan: baseline

We need to determine the baseline of each collection place. A “baseline” is a statistical quantification (averages and variances) of each species that appear with regularity.

Samples which deviate significantly from the baseline for problematic species should produce early warning for the health authorities

# Plan: modes

A mode is a baseline which happens with certain regularity given some external influences, e.g. summer/winter, dry/rainy, etc.

# Plan: calibration

We need to calibrate our results, in particular for MetaSEW. “Calibration” in this context means finding the relation between the amount of species at the source and what we measure from the collected samples.



16S vs  
shotgun

# I 6S vs shotgun

In the long run we want to do shotgun sequencing for several good reasons:

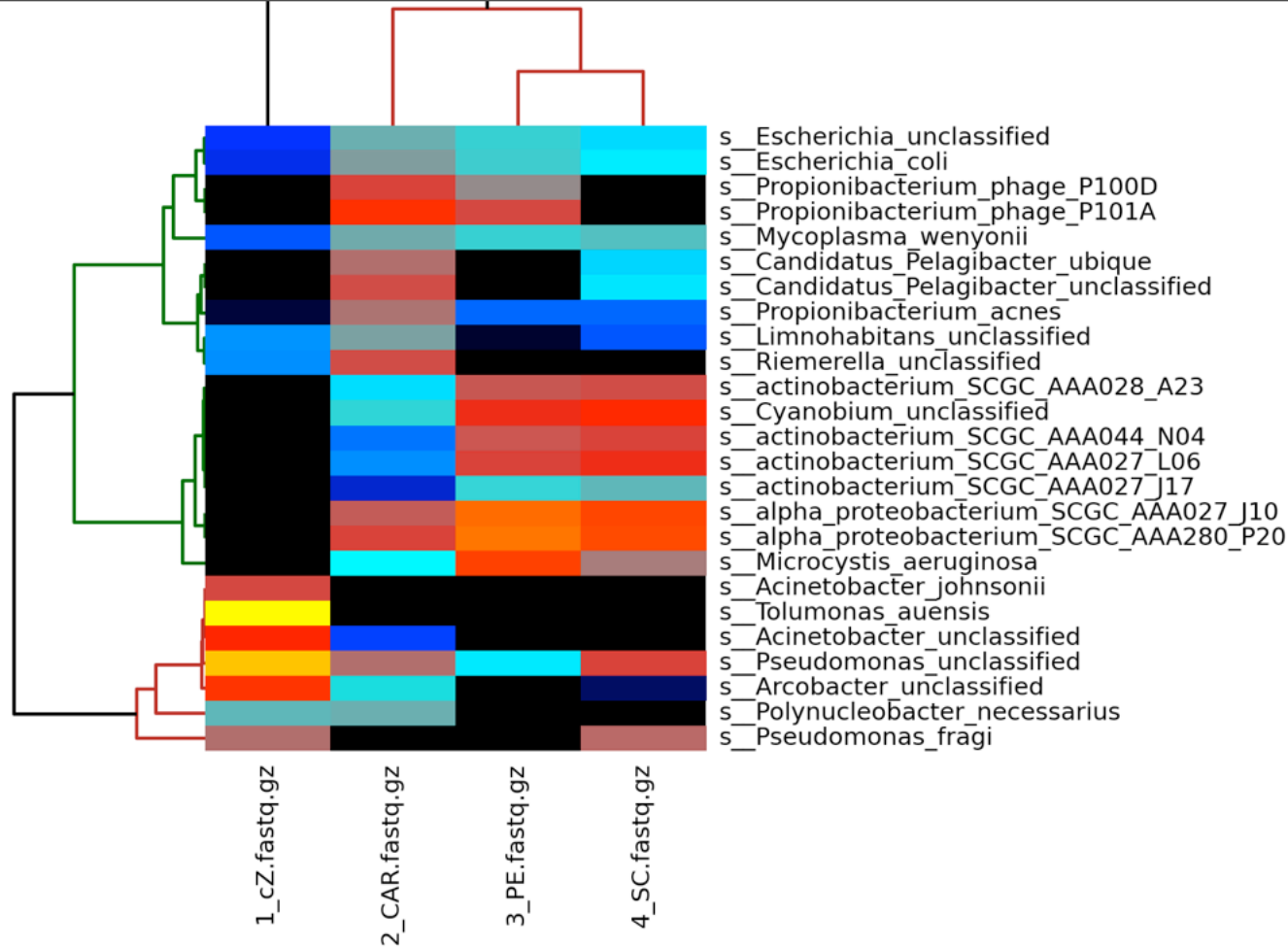
- to reconstruct unknown genomes,
- to determine strain variations,
- to determine antibiotic resistance,
- etc.

# 16S vs shotgun

16S or 18S analysis deserves a closer look, it requires some additional lab work, but delivers a lot of unbiased information at a very modest sequencing cost.



# Big thank you, organizers and audience



ETH Zürich



anges saved in Drive

