

Microbiomes and metagenomics

Overview

- Microbiome sequencing – metabarcoding
 - Data cleaning, with or without UMIs
 - Software: UNOISE2, DADA2, deblur, QIIME2
- Metagenomics & -transcriptomics
 - Shotgun sequencing – short and long reads
 - Hi-C, binning, single cell approaches
- Mega-projects

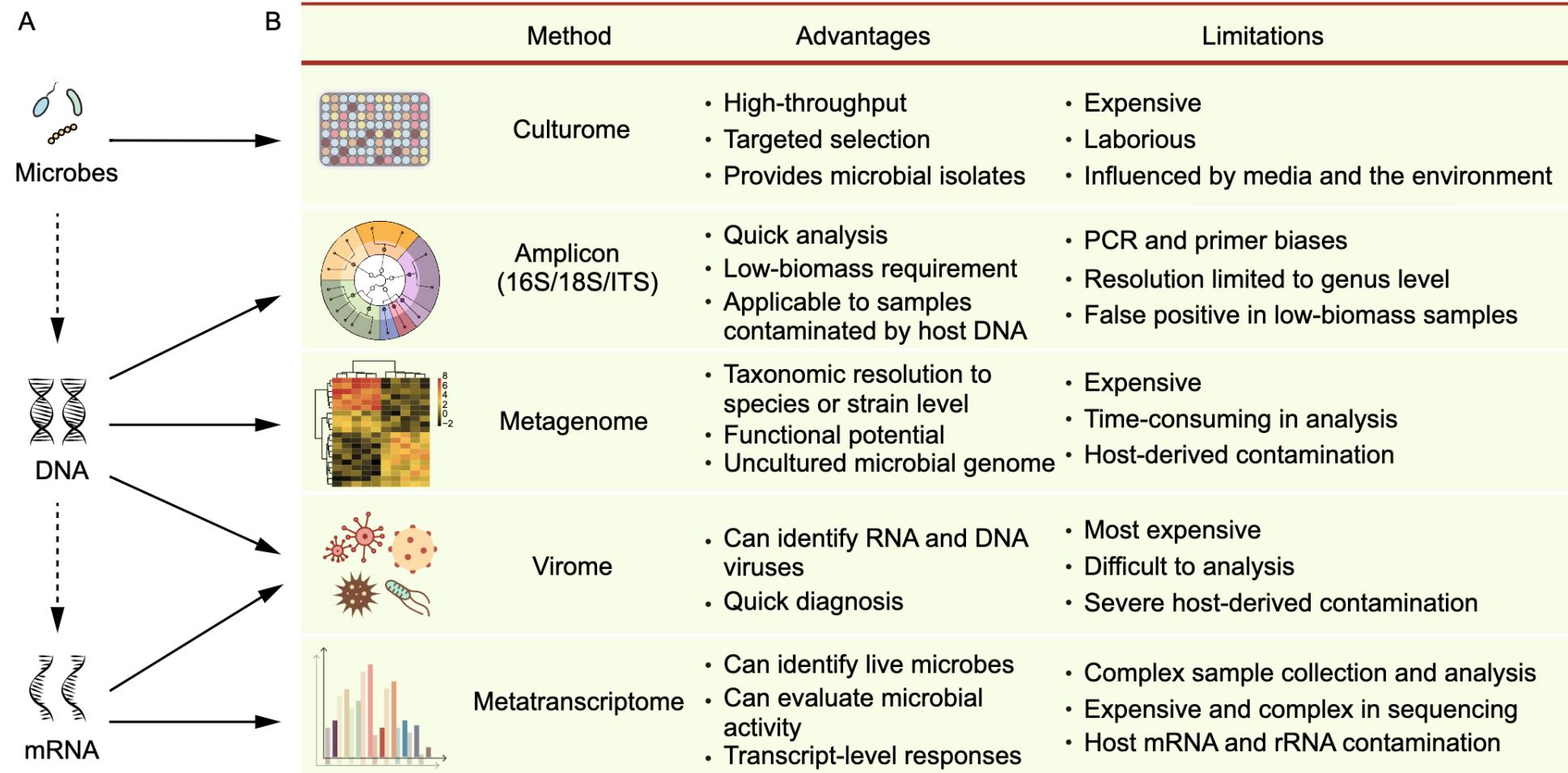
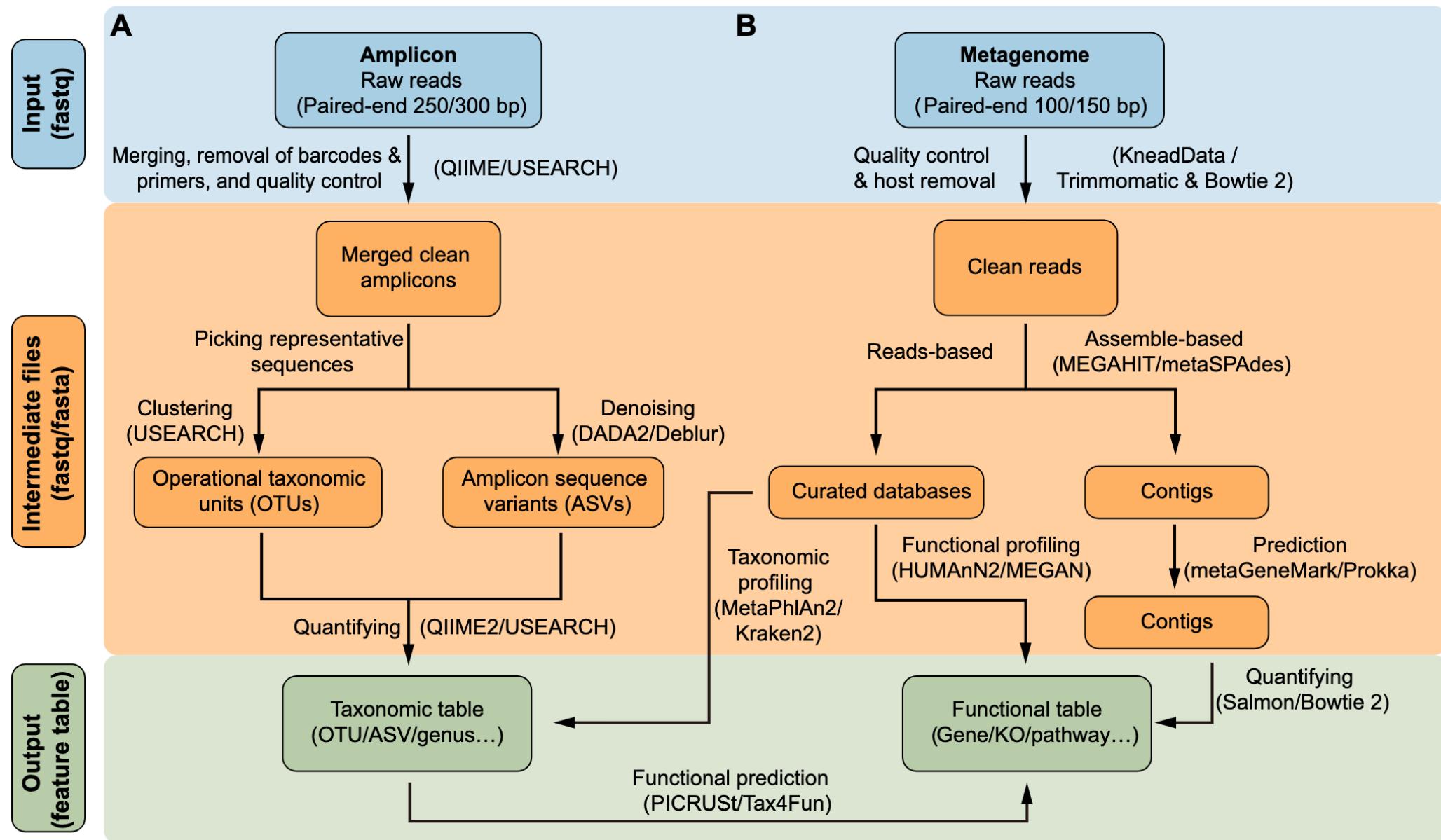


Figure 1. Advantages and limitations of HTS methods used in microbiome research. A Introduction to HTS methods for different levels of analysis. At the molecule-level, microbiome studies are divided into three types: microbe, DNA, and mRNA. The corresponding research techniques include culturome, amplicon, metagenome, metavirome, and metatranscriptome analyses. B The advantages and limitations of various HTS methods for microbiome analysis.

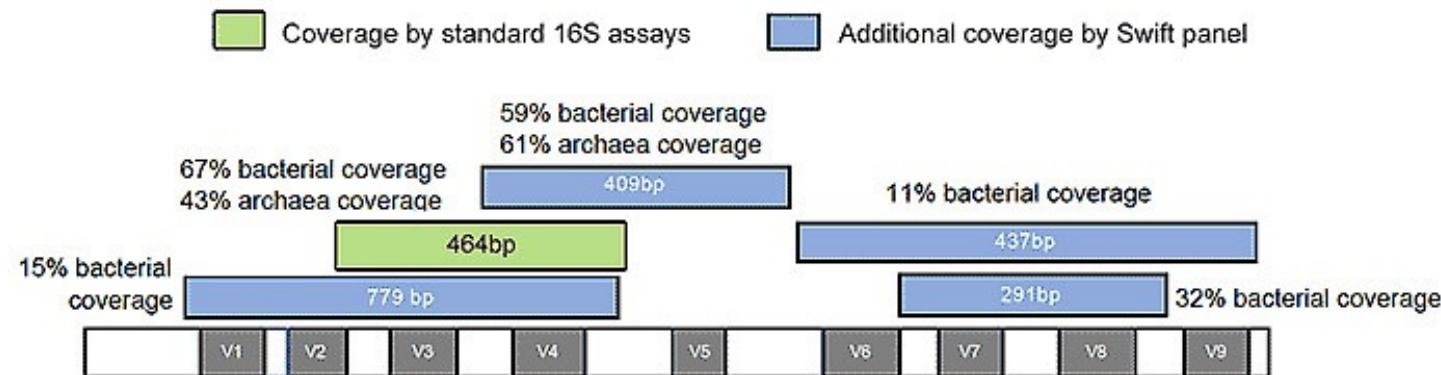
PMID: 32394199

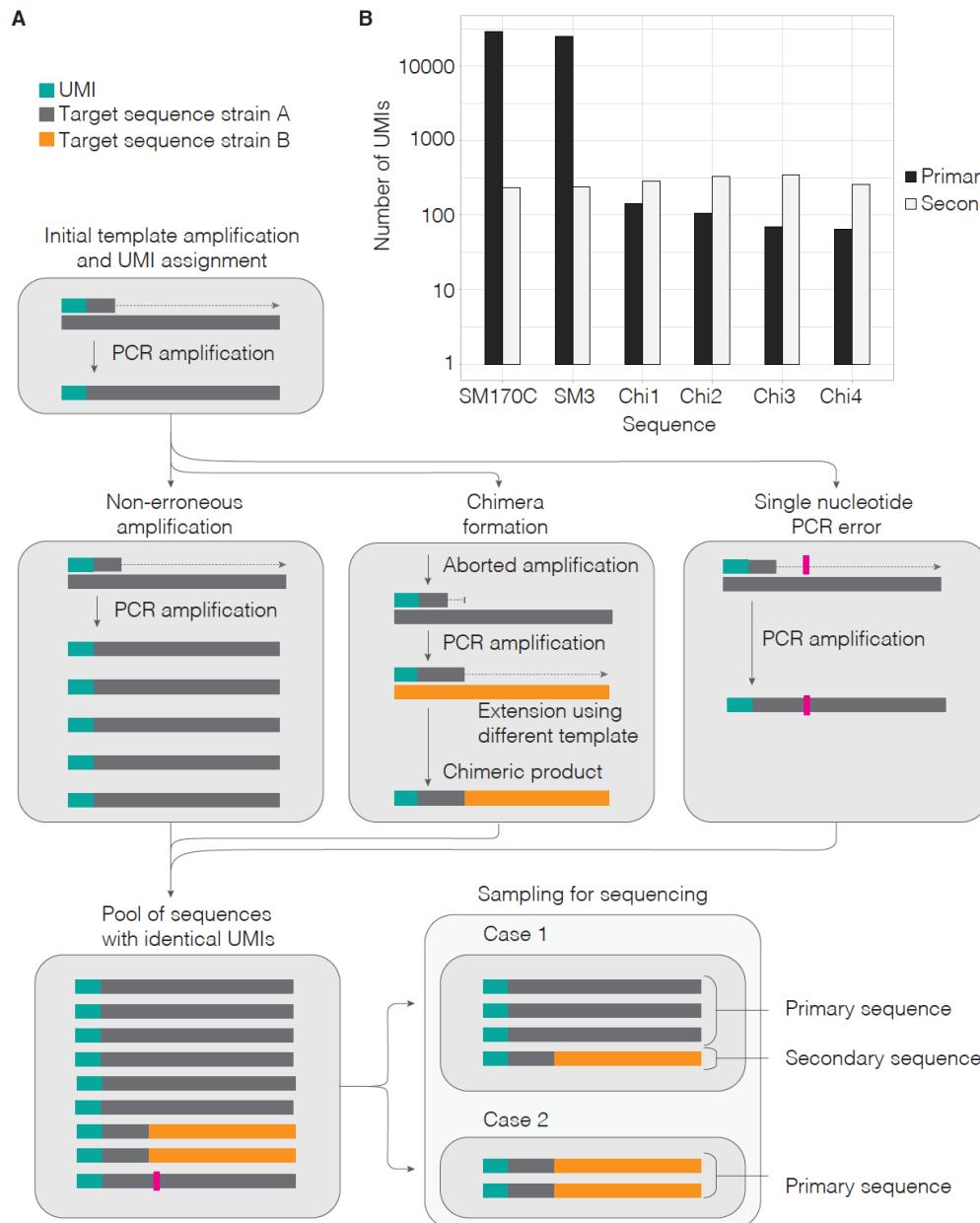


PMID: 32394199

16S ribosomal DNA profiling

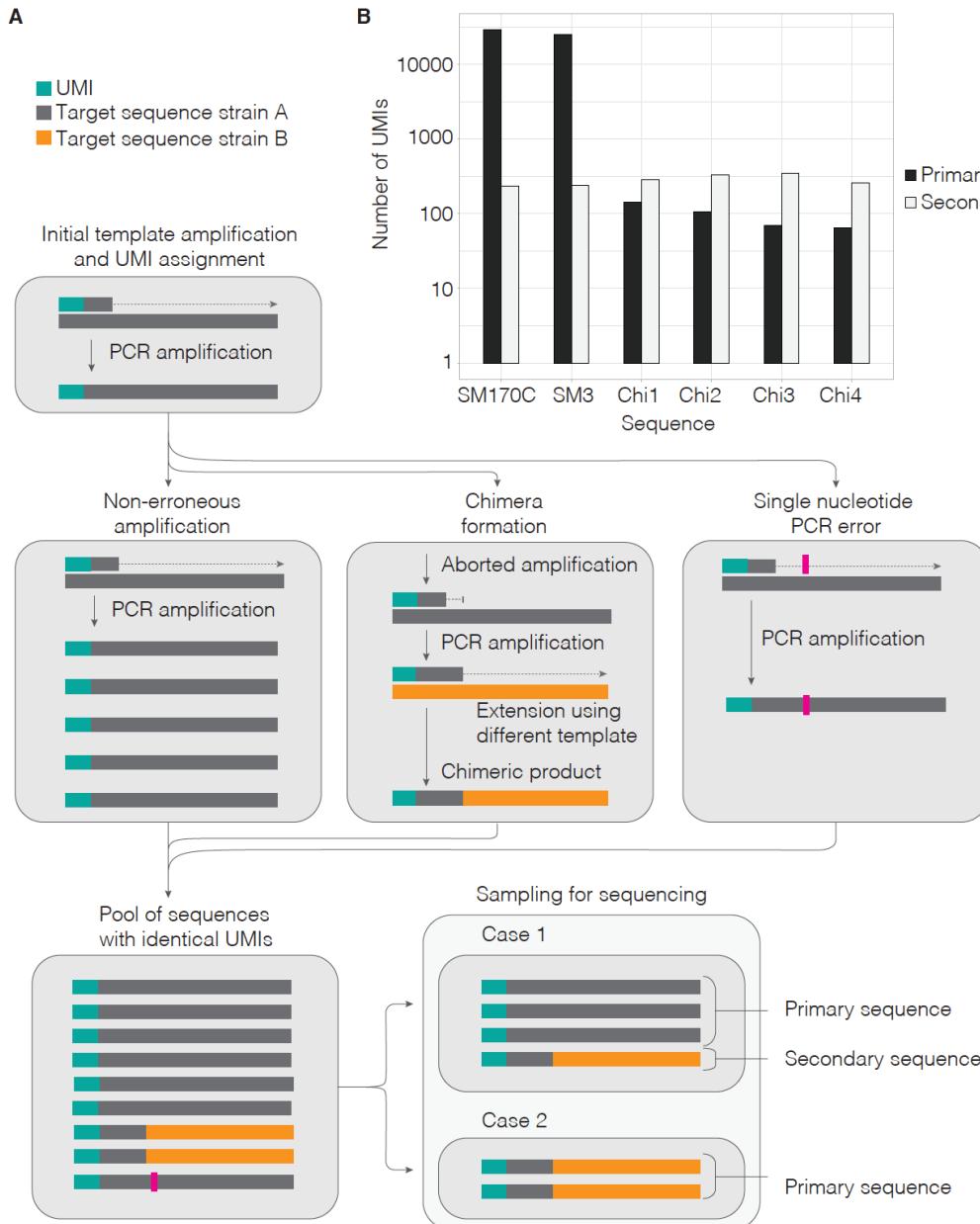
Amplicon coverage of 16S rRNA gene variable regions





PCR amplification introduces point mutations and formation of chimeric amplicons

Unique molecular identifiers (UMIs) can be used for error correction

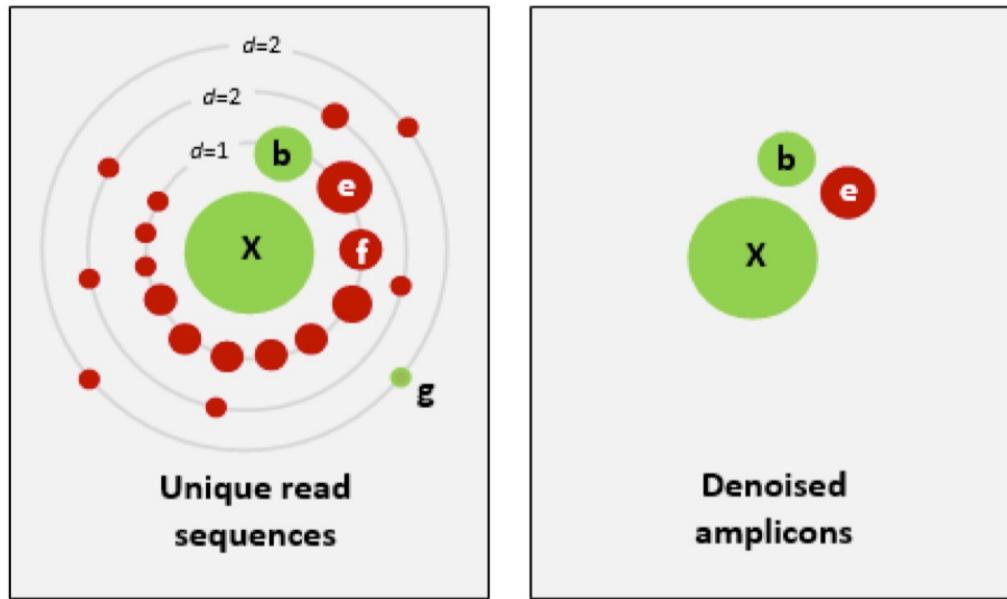


Software to clean up errors:
UNOISE2, Dada2, deblur
Guessing from read abundance distributions

Cost of including UMIs:
More coverage per sequence required.

Advantage:
Hard data on errors and error frequencies
Correction of amplification bias

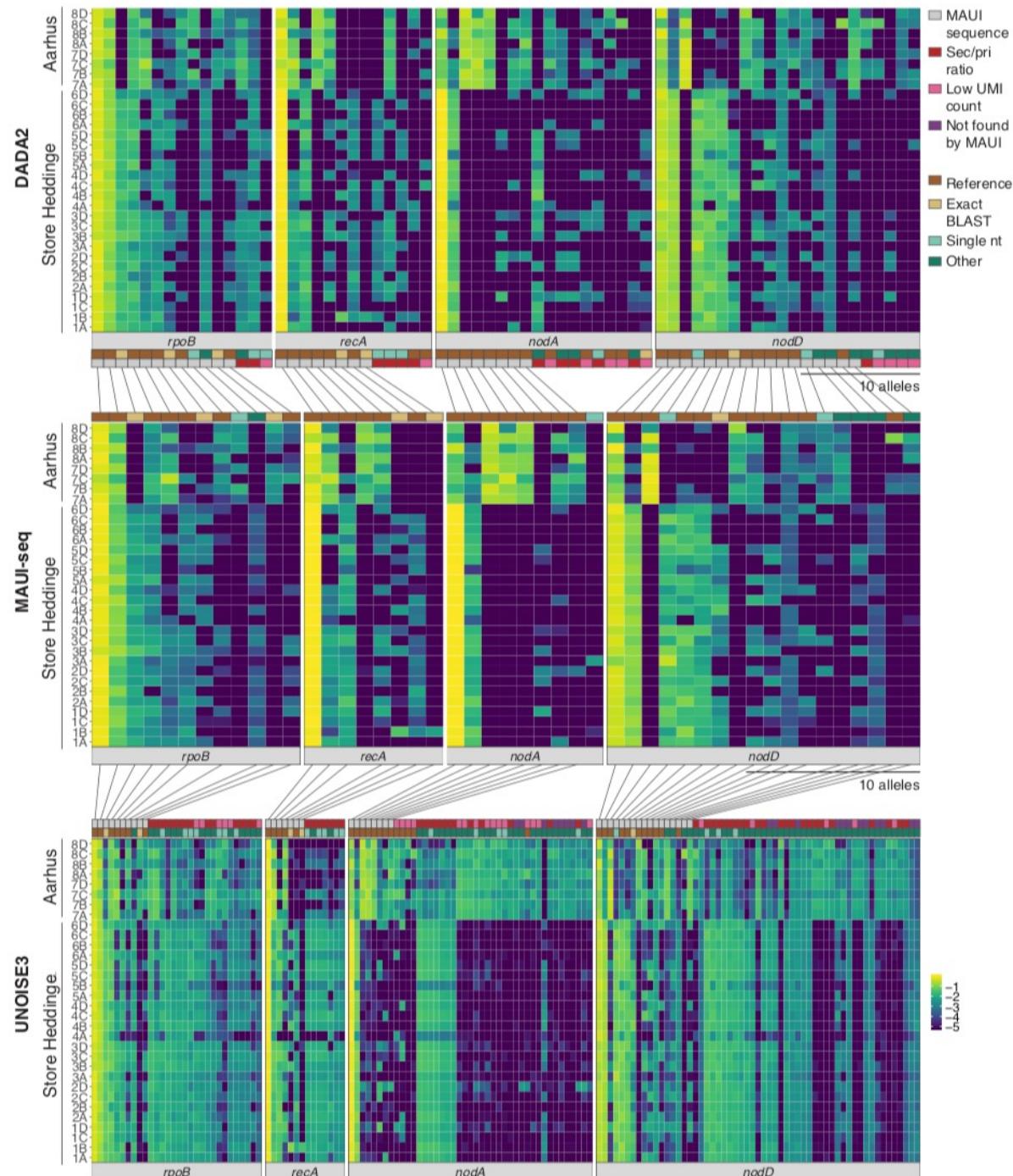
Usually, UMIs are not used, and a single 16S rDNA amplicon is profiled



Operational taxonomic units – OTUs:
Cluster sequences with 97% threshold

UNOISE2: zero-radius OTUs (ZOTUs)

Dada2: advanced error models for SNPs,
simple model for chimeras



Software to clean up errors:
UNOISE2, Dada2, deblur
Guessing from read abundance distributions

Cost of including UMIs:
More coverage per sequence required.

Advantage:
Hard data on errors and error frequencies
Correction of amplification bias

Usually, UMIs are not used, and a single 16S rDNA amplicon is profiled

Include mock community to get a better idea of error rates and biases introduced during sample prep and data analysis

Correspondence | Published: 24 July 2019

Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2

Evan Bolyen, Jai Ram Rideout, [...] J. Gregory Caporaso 

Nature Biotechnology 37, 852–857(2019) | [Cite this article](#)

25k Accesses | 328 Citations | 244 Altmetric | [Metrics](#)

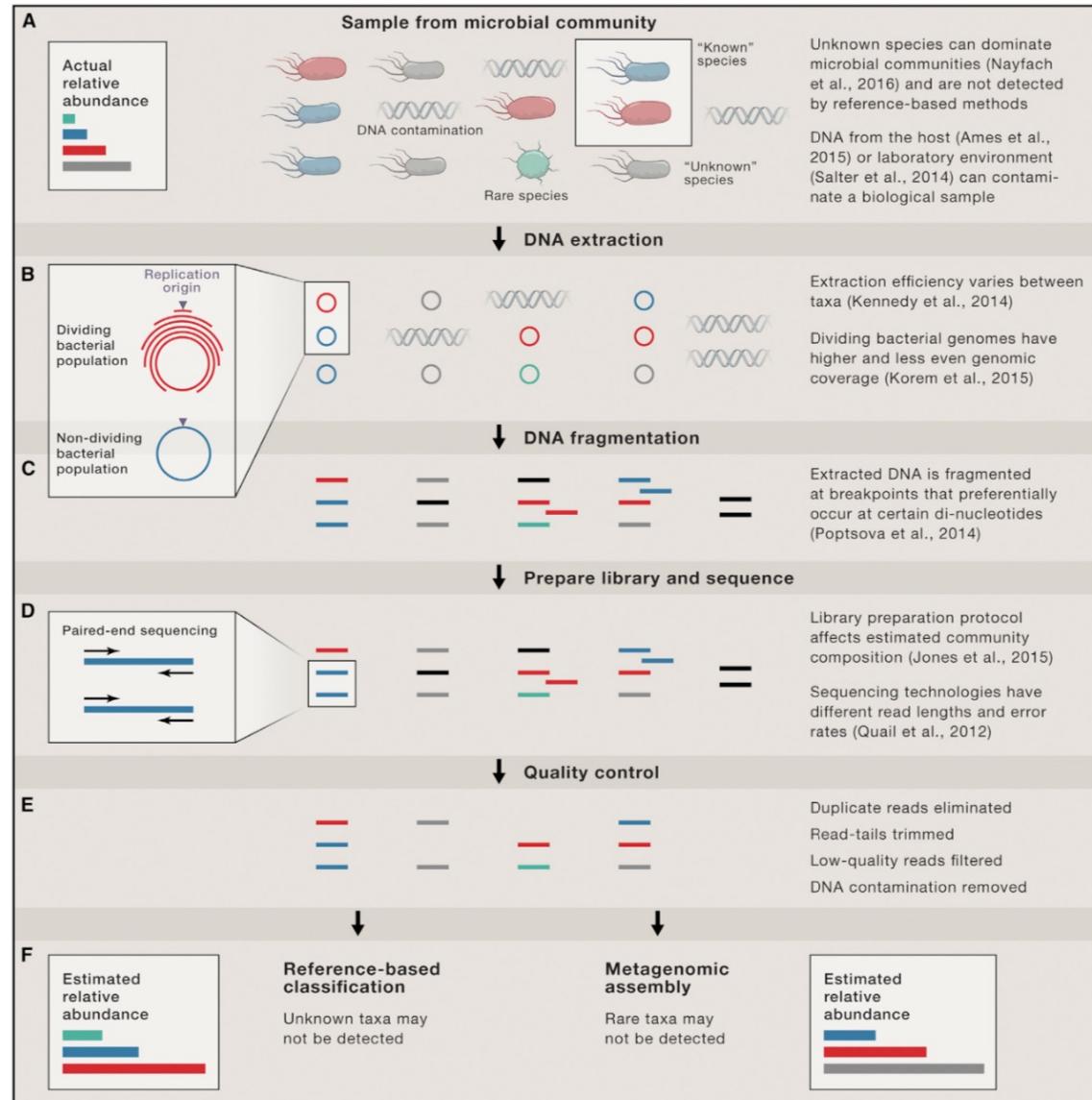
QIIME 2 uses plugins to combine different types of microbiome data analysis.

Is available through Galaxy.

Functionality is also being developed for metagenomics analysis

PMID: 31341288

Metagenomics



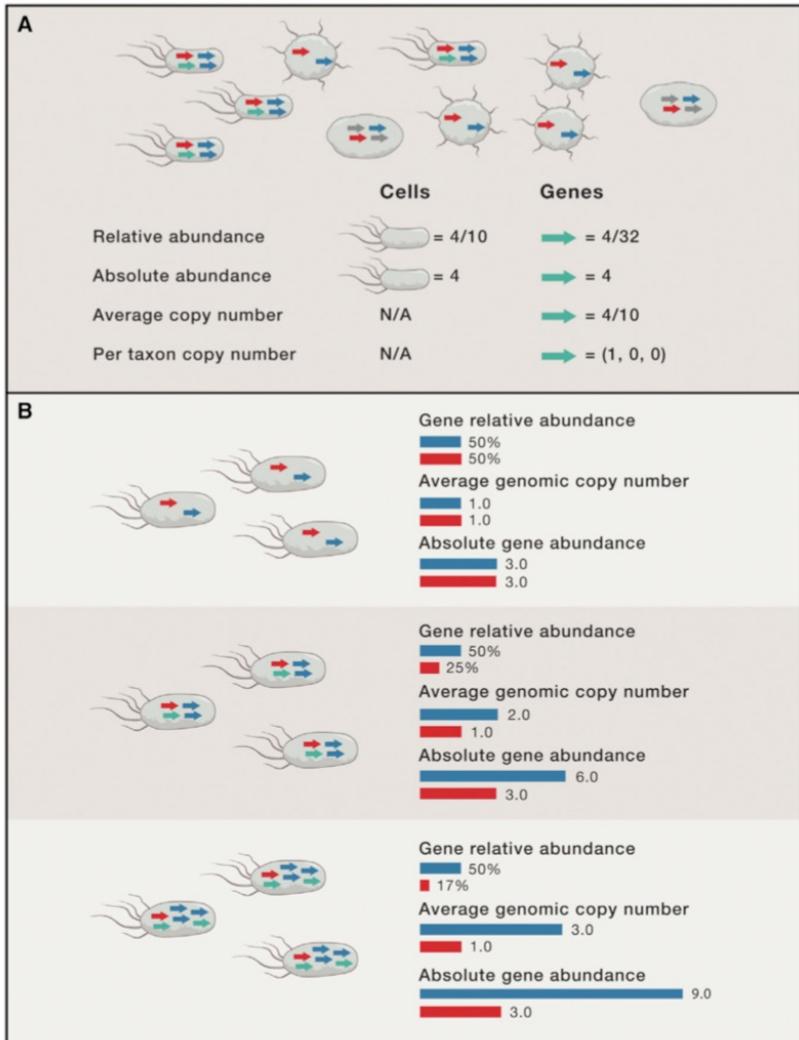
Sequencing full metagenomes

Each microbe is represented by its complete genome, not by a single gene

This means it requires many more reads to sample the same number of species as compared to 16S profiling

It can provide information on the gene repertoire of a microbial community

Metagenomics

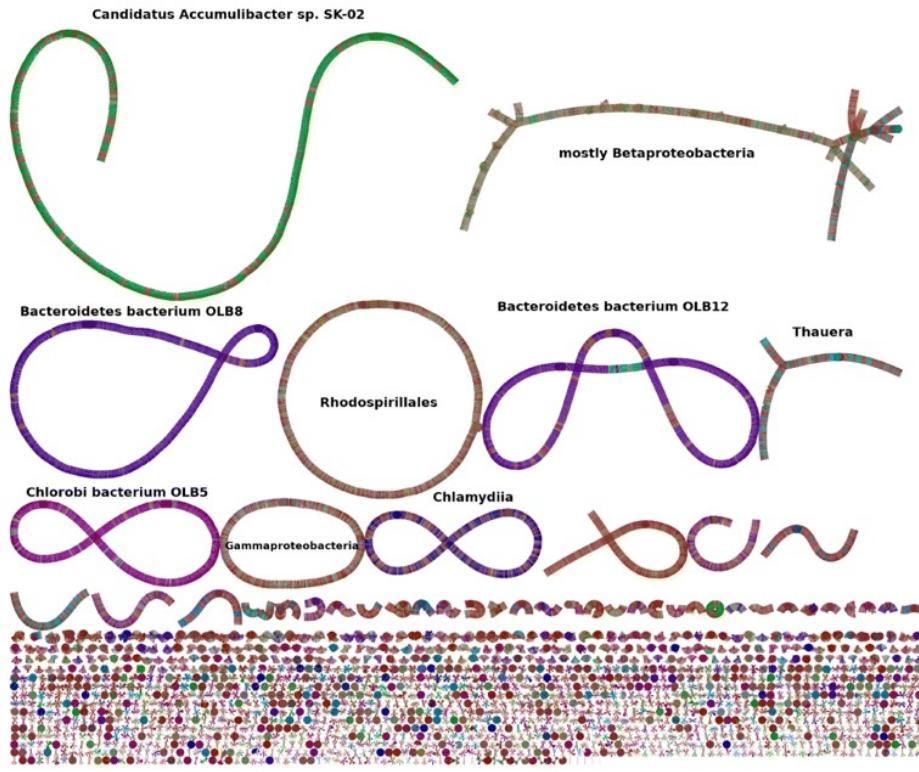


Abundance quantification can be challenging

What will the impact of PacBio Hifi reads be?

Metagenomics with long reads

Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data

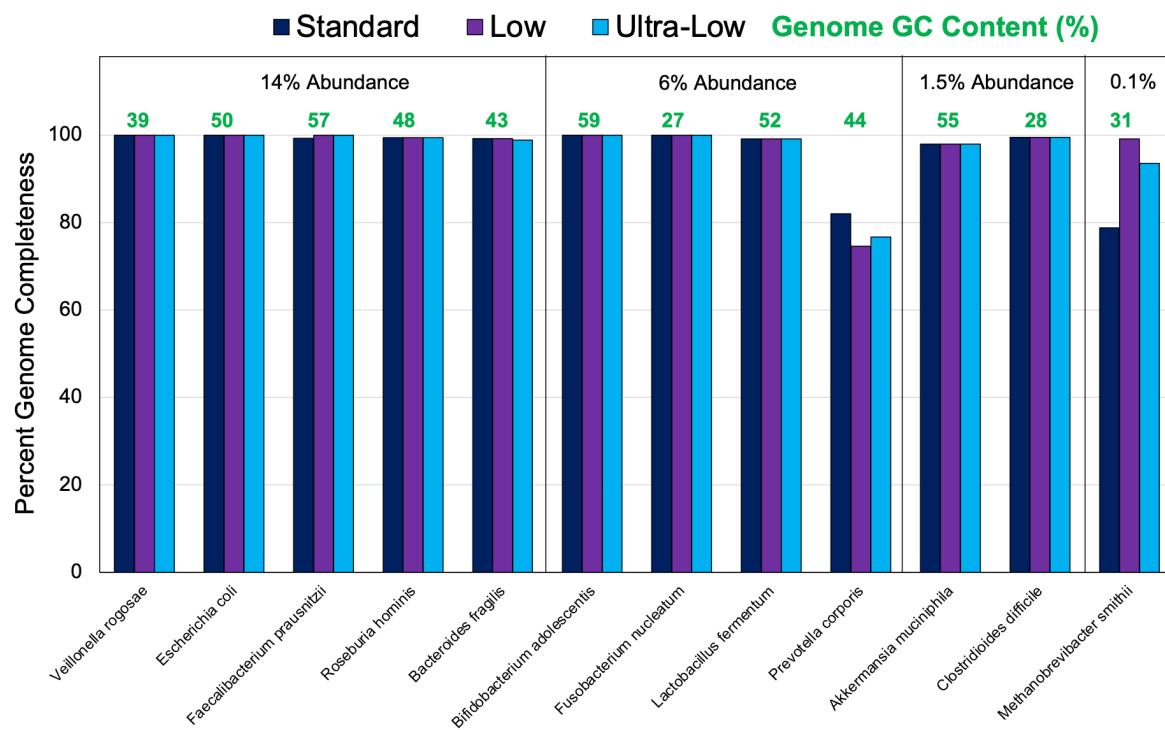


Nanopore applications: PMID 33815688

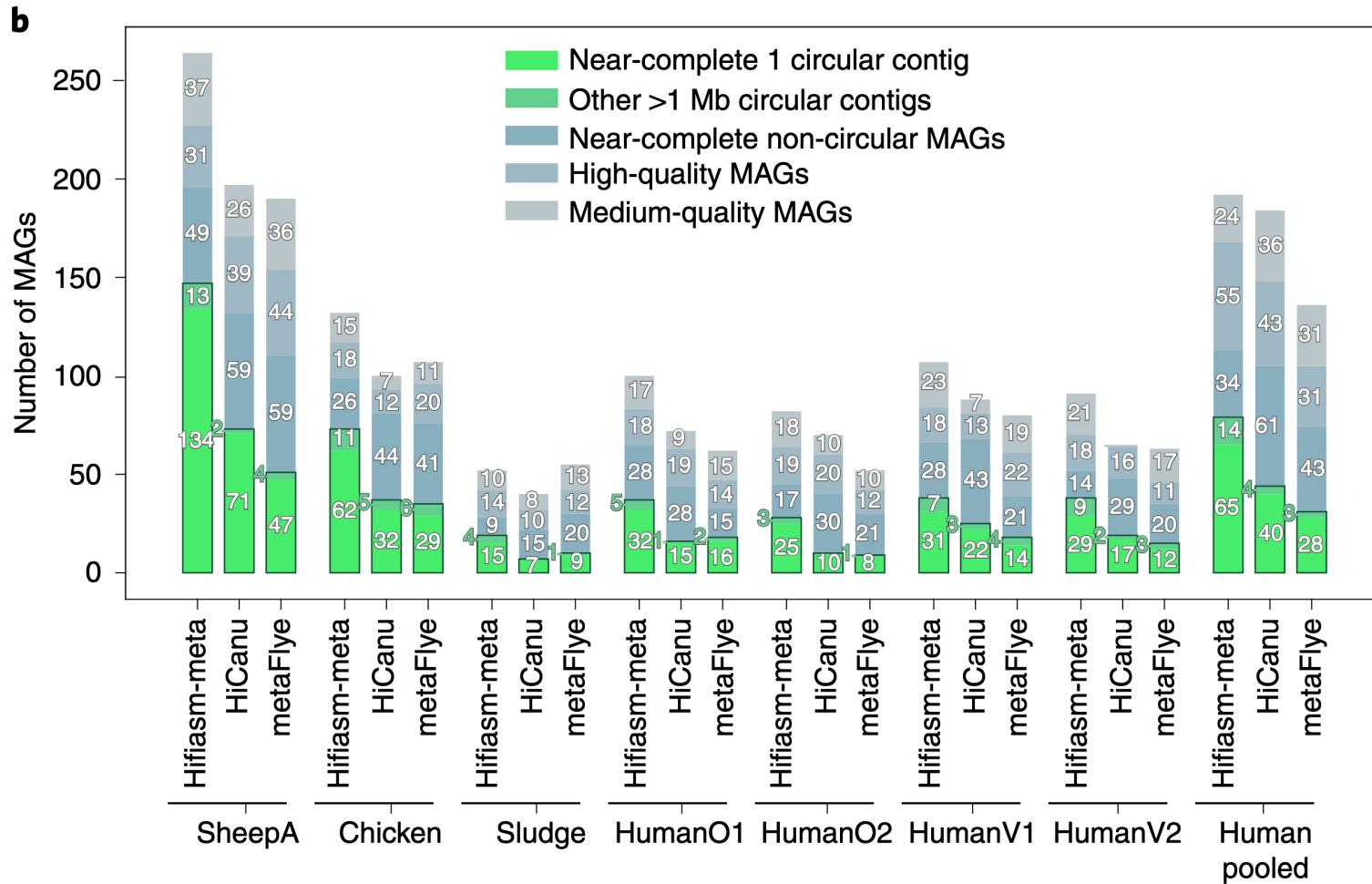
PMID: 30992083

Hifi reads

Figure 7. For all three preparation methods, the genomes of species with abundances as low as 1.5% were complete. As few as 2,500 reads or 10-fold coverage were sufficient to reconstruct a complete bacterial chromosome. Notably, a closed MAG was generated with 10-fold coverage for a 0.1% abundant species.



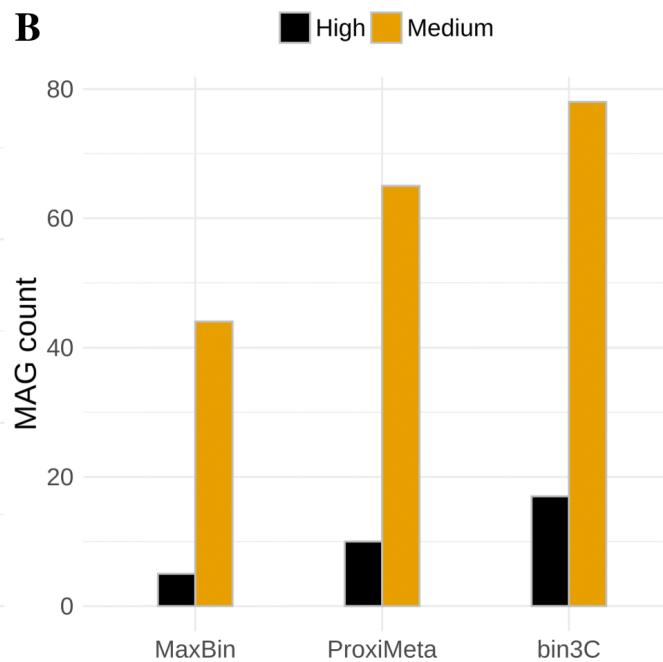
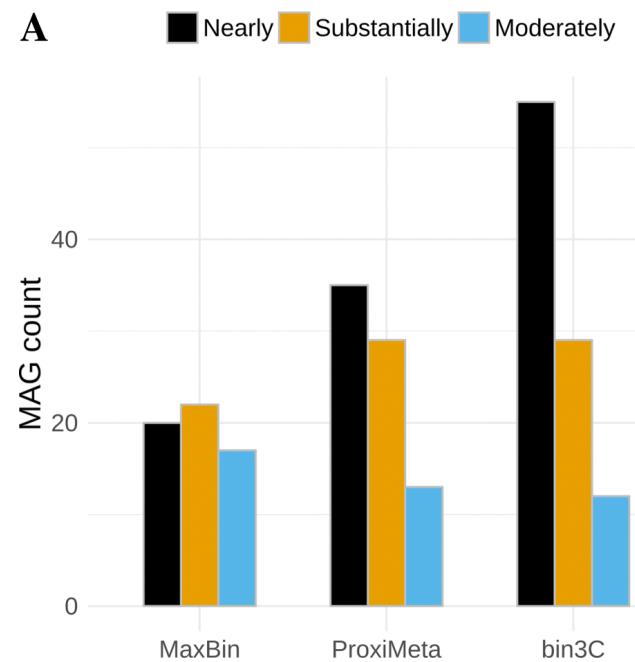
Hifi reads & Heng Li



MAG: metagenome-assembled genome

PMID: 35534630

Metagenomics with Hi-C data



MAG: metagenome-assembled genome

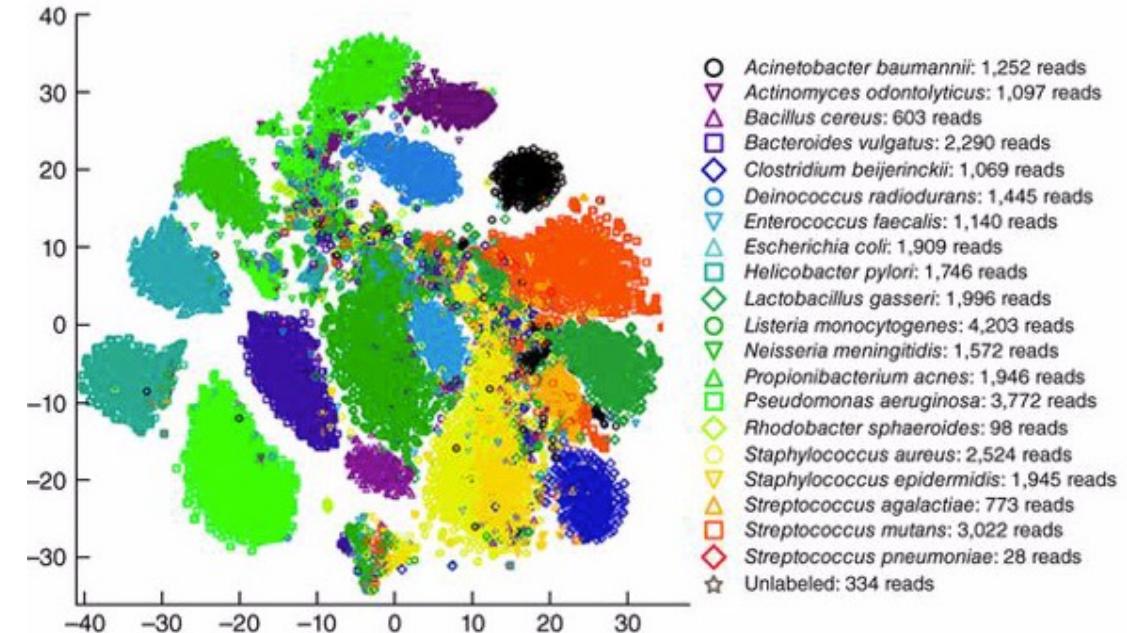
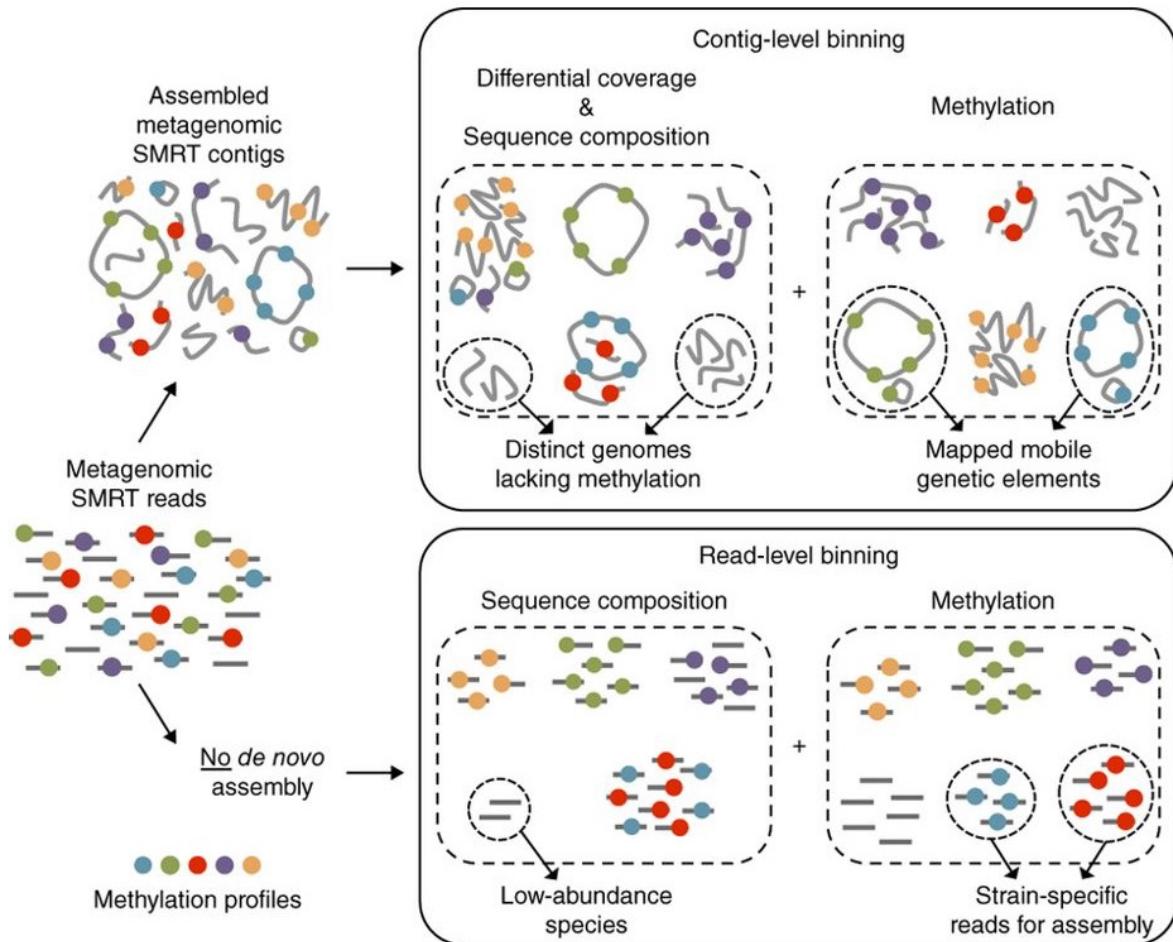
<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1643-1>

Hi-C data is more complicated to generate, but it can link bacterial plasmids and chromosomes

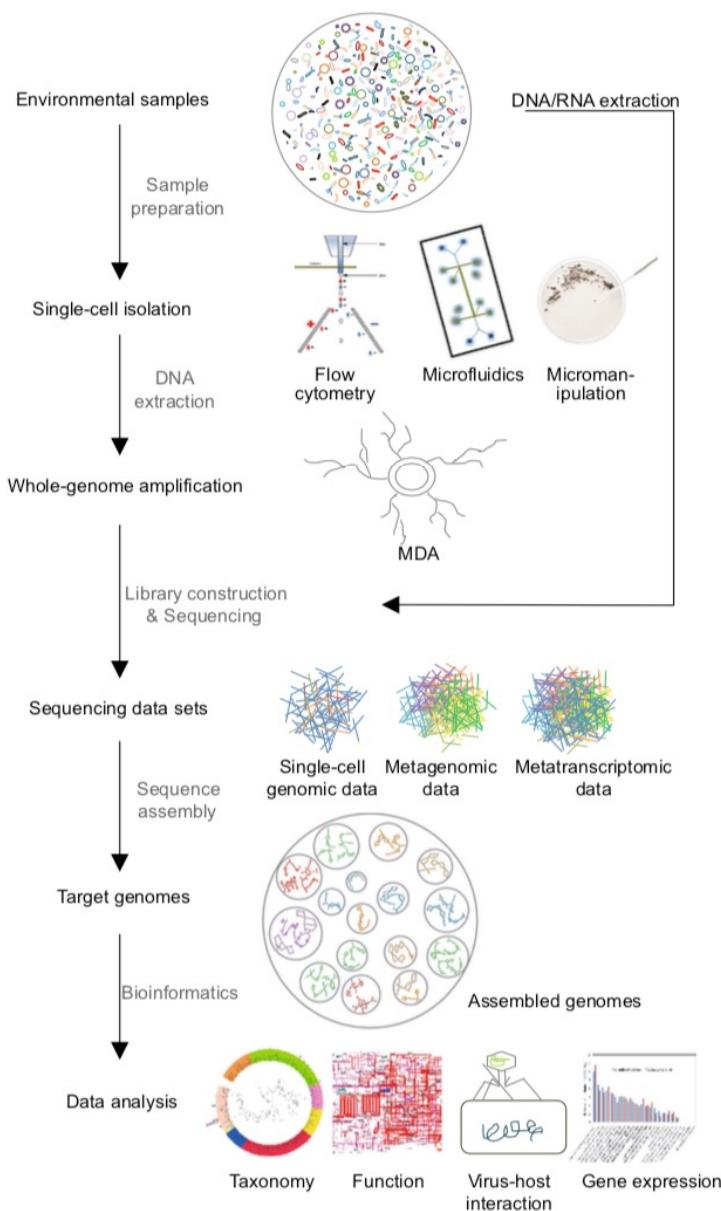
Genome scaffolding:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwjxrM6q64fjAhXJyaQKHQxjCQwQwqsBMAN6BAgKEAc&url=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DMxEw3IXUWU&usg=AOvVaw3We-pxsV-bUdpLij2_NXJd

Metagenomic deconvolution:
<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwjqmvt64fjAhUvMuwKHfhtCSAQwqsBMAB6BAgJEAQ&url=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DBa8O6rSoU8A&usg=AOvVaw34W3vB0BgPJwaxRscxCXGH>

Binning in metagenomics



Single cell metagenomics



Links metabolic functions to specific species

High quality genomes from lowly abundant species

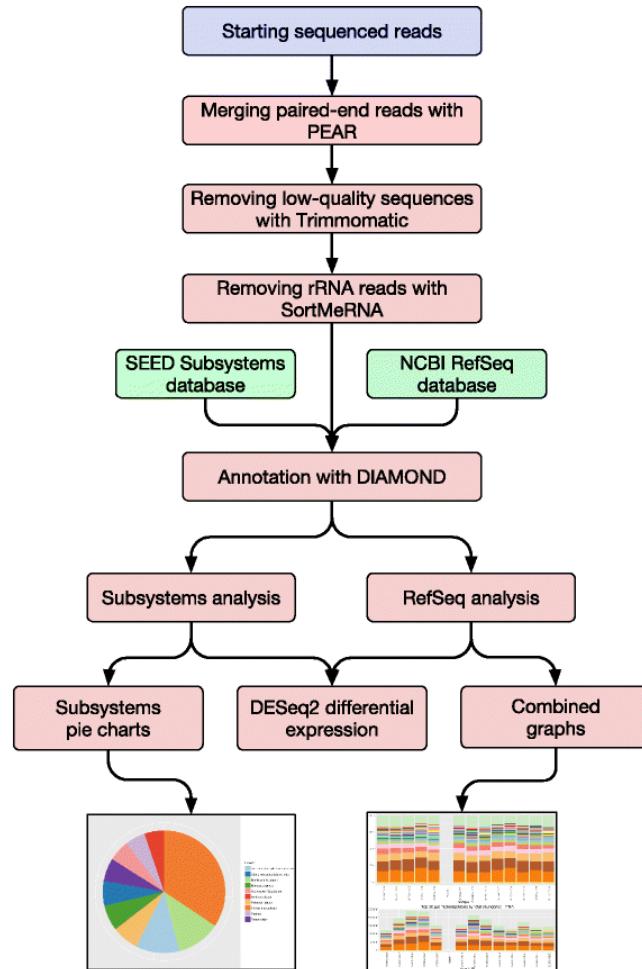
Complicated procedure for cell separation

Whole genome amplification required

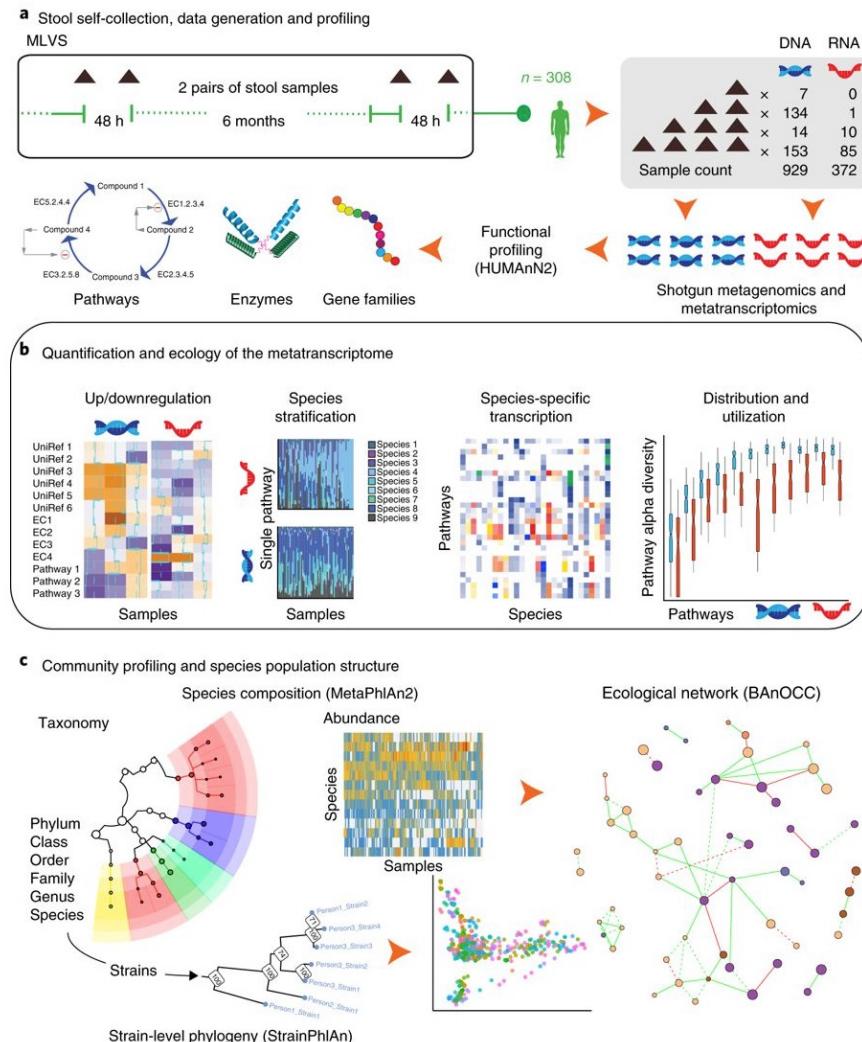
Data can help to guide metagenomic binning

<https://link.springer.com/article/10.1007/s13238-018-0544-5>

Metatranscriptomics



Deplete host RNA
and rRNA



<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2189-z>

<https://www.nature.com/articles/s41564-017-0084-4/figures/1>

Metagenomics mega-projects



Center*Healthy Cohort*

Demonstration Project "Disease Cohorts"

Reference microbial genomes

>2000 strains

[NCBI BioProject 28331](#)

Hundreds of strains

[NCBI BioProject 46305](#)

mWGS metagenomic sequence

Subset of the 300 subjects, multiple timepoints, 15+ bodysites

[NCBI BioProject 43017](#)

5 projects, each with unique, sampling sites, conditions, etc.

[NCBI BioProject 46305](#)

16S metagenomic sequence

300 subjects, multiple timepoints, 15+ bodysites

[NCBI BioProject 48489](#)

14 projects, each with unique, sampling sites, conditions, etc. 4 projects contain both 16S and mWGS components

[NCBI BioProject 46305](#)



iHMP
NIH Integrative Human Microbiome Project

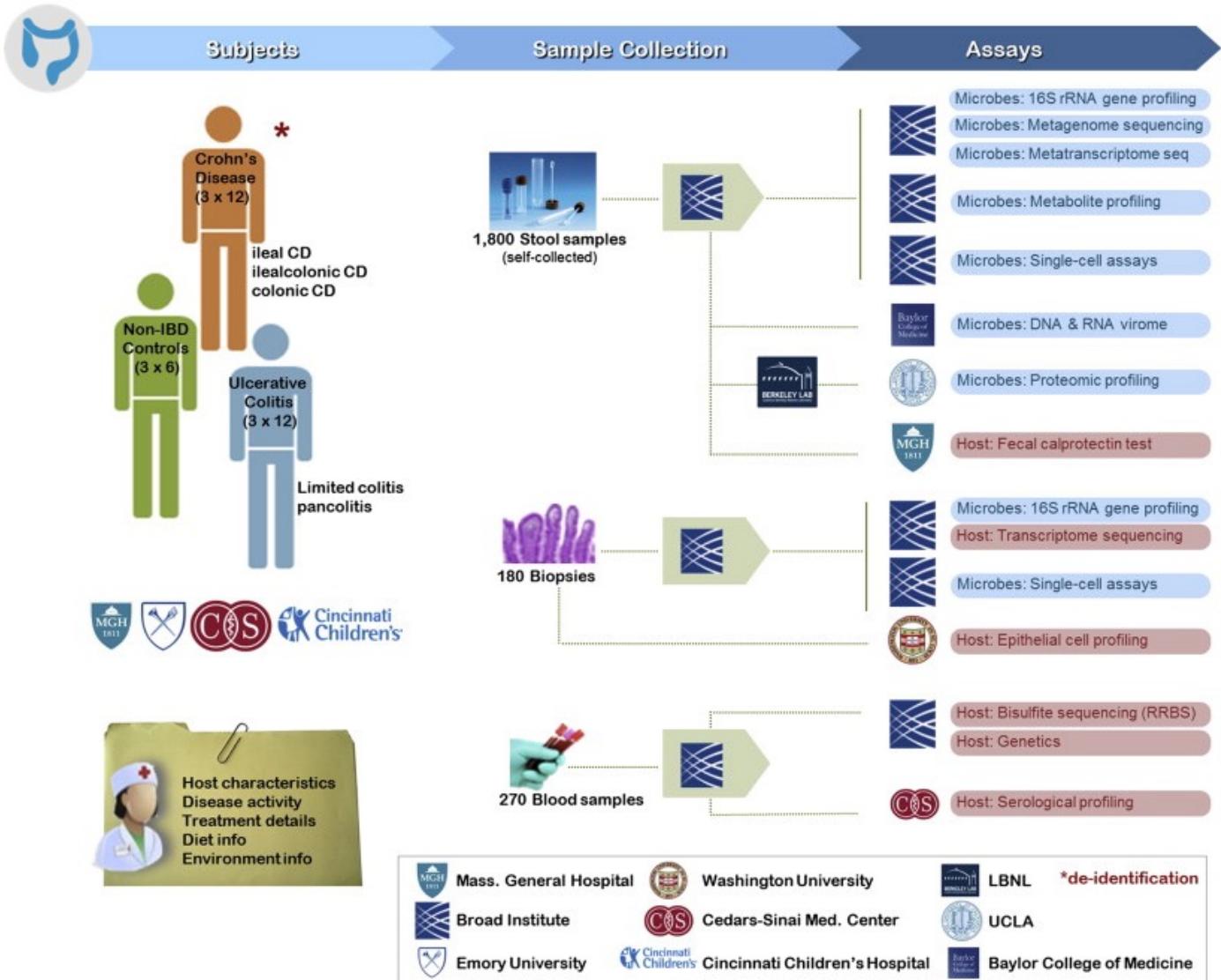


iHMP



SRA	dbGap	GEO	EBI PRIDE	ATCC/BEI	Study DB	EBI IntAct	Metabolomics Workbench
16S rRNA gene survey*	Subject genome sequences**	Whole transcriptome sequence	LC-MS/MS peptide profiles	Bacterial isolates	Cytokine profiles	Yeast two-hybrid binary protein complexes	Untargeted and targeted LC-MS metabolomic profiles
Whole metagenome shotgun sequences	Whole transcriptome sequences	Intestinal epithelial cell profiling response to bacterial metabolites	LC-MS/MS peptide profiles	Fecal calprotectin protein concentrations			Untargeted and targeted LC-MS lipidprofiles
Whole firome shotgun sequences	Human sequence from unfiltered total metagenomic sequences				Serology		
Bacterial single cell sequences	Subject phenotypes, clinical metadata, medical panels						
Single cell bacterial transcript sequences	Intestinal epithelial cell profiling response to bacterial metabolites***						
Whole metatranscriptome shotgun sequences							
Reduced representation bisulfite sequencing (RRBS) profiles							

Onset of Inflammatory Bowel Disease (IBD)



Novo Nordisk Foundation awards USD 30 million for a plant research project to kickstart the next green revolution



The world's population will increase by an additional 2 billion by 2050, and combined with rising incomes, the Food and Agriculture Organization of the United Nations expects that the demand for food will increase by about 50%. This poses new challenges for agricultural productivity and sustainability and is the basis for the Collaborative Crop Resilience Program (CCRP): a new 6-year research project to be carried out in Denmark and the United States for which the Novo Nordisk Foundation has awarded a grant of DKK 203 million (USD 30 million).

<https://novonordiskfonden.dk/en/news/novo-nordisk-foundation-awards-usd-30-million-for-a-plant-research-project-to-kickstart-the-next-green-revolution/>

A new international research project will create basic knowledge on the microorganisms that coexist with plants. The ambition is to increase crop production and help ensure sufficient food for a growing world population by creating smarter and more sustainable agriculture. This is the first major grant the Novo Nordisk Foundation has awarded within plant research.