

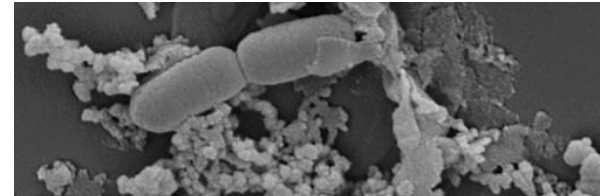


# Best practice microbiome data analysis at QBiC

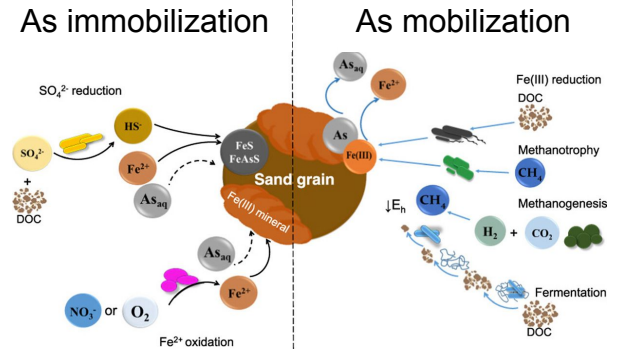


# Microbial communities are ubiquitous

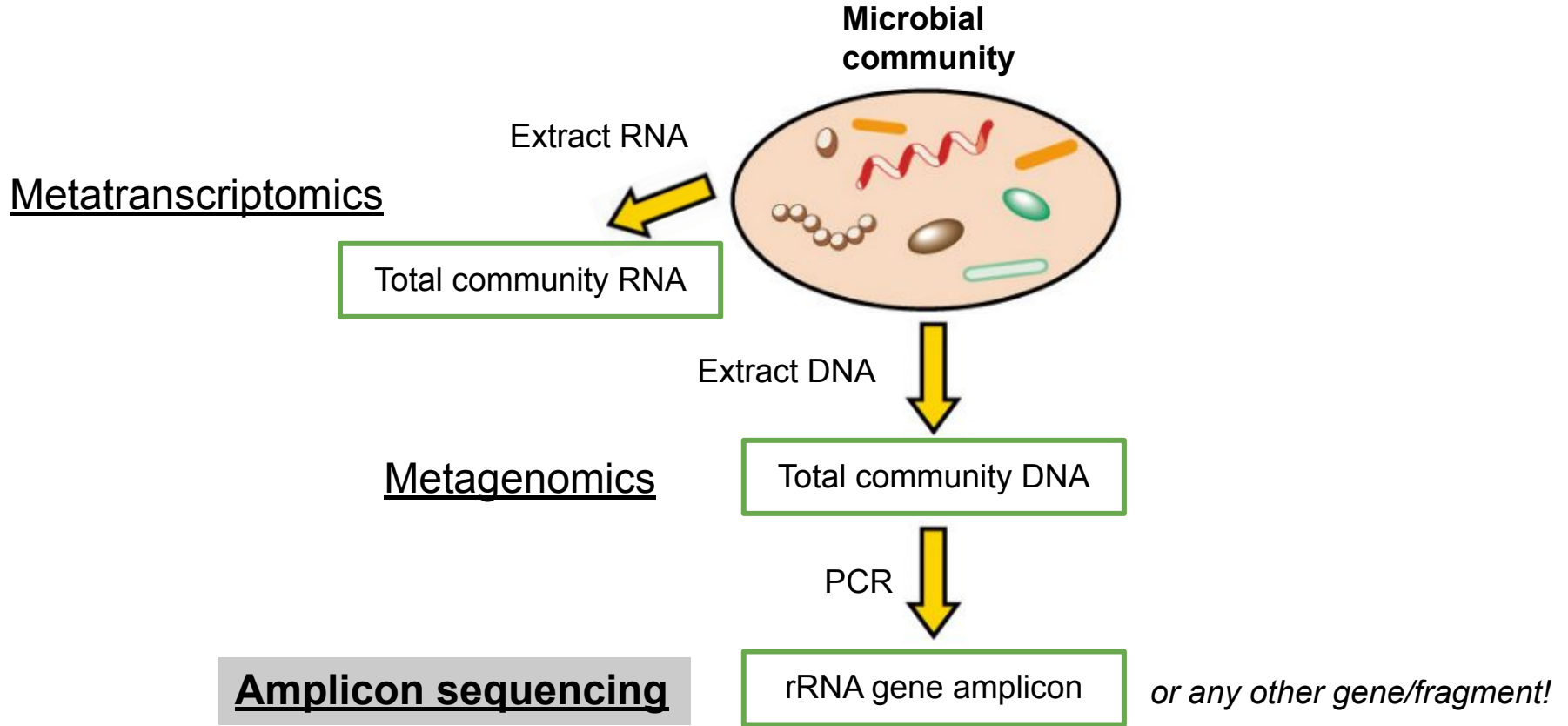
- Human: More bacterial cells than human cells
- Agriculture: Beneficial or pathogenic
- Environment: Global elemental cycles



Jakus et al. (2021) Environmental Microbiology

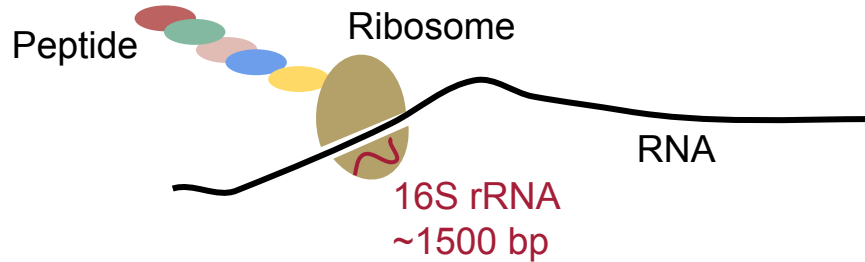


Glodowska et al. (2021) Journal of Hazardous Materials





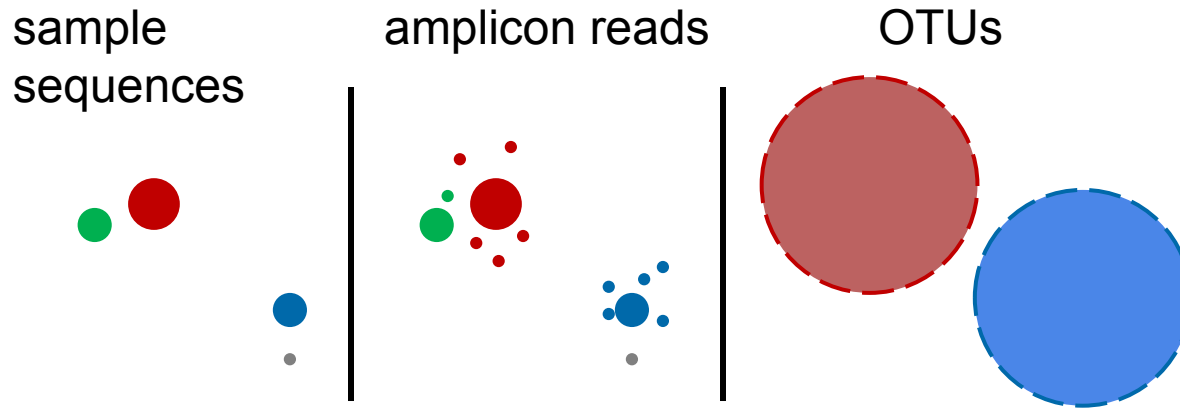
# The 16S rRNA gene is a good marker for bacteria



- Prokaryotes have 16S & 5S & 23S rRNA, eukaryotes 18S & 28S & 5.8S & 5S rRNA
- The 16S rRNA gene has conserved but also highly variable regions
- Variable regions of the 16S rRNA gene allow the discrimination of many bacterial taxa



# Error accumulation during library prep and sequencing



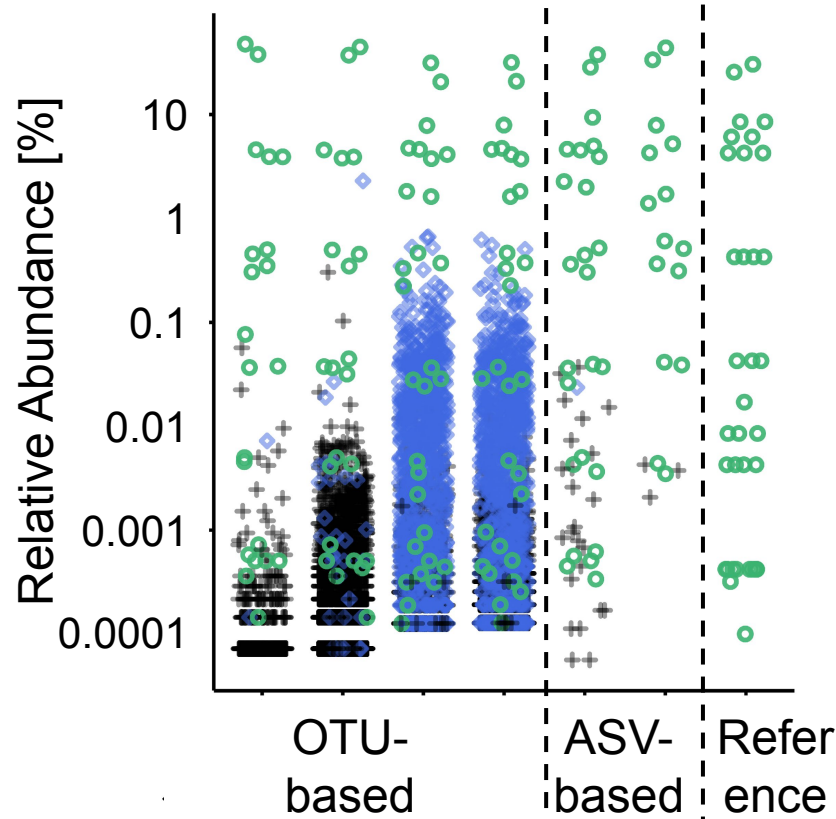
Errors  
 →  
 ← Make ASVs

Make OTUs  
 →

OTU: Operational taxonomic unit  
 ASV: amplicon sequencing variant



# Evaluation: sequences



Alignment  
to reference

- perfect
- ◇ one off
- ✦ other

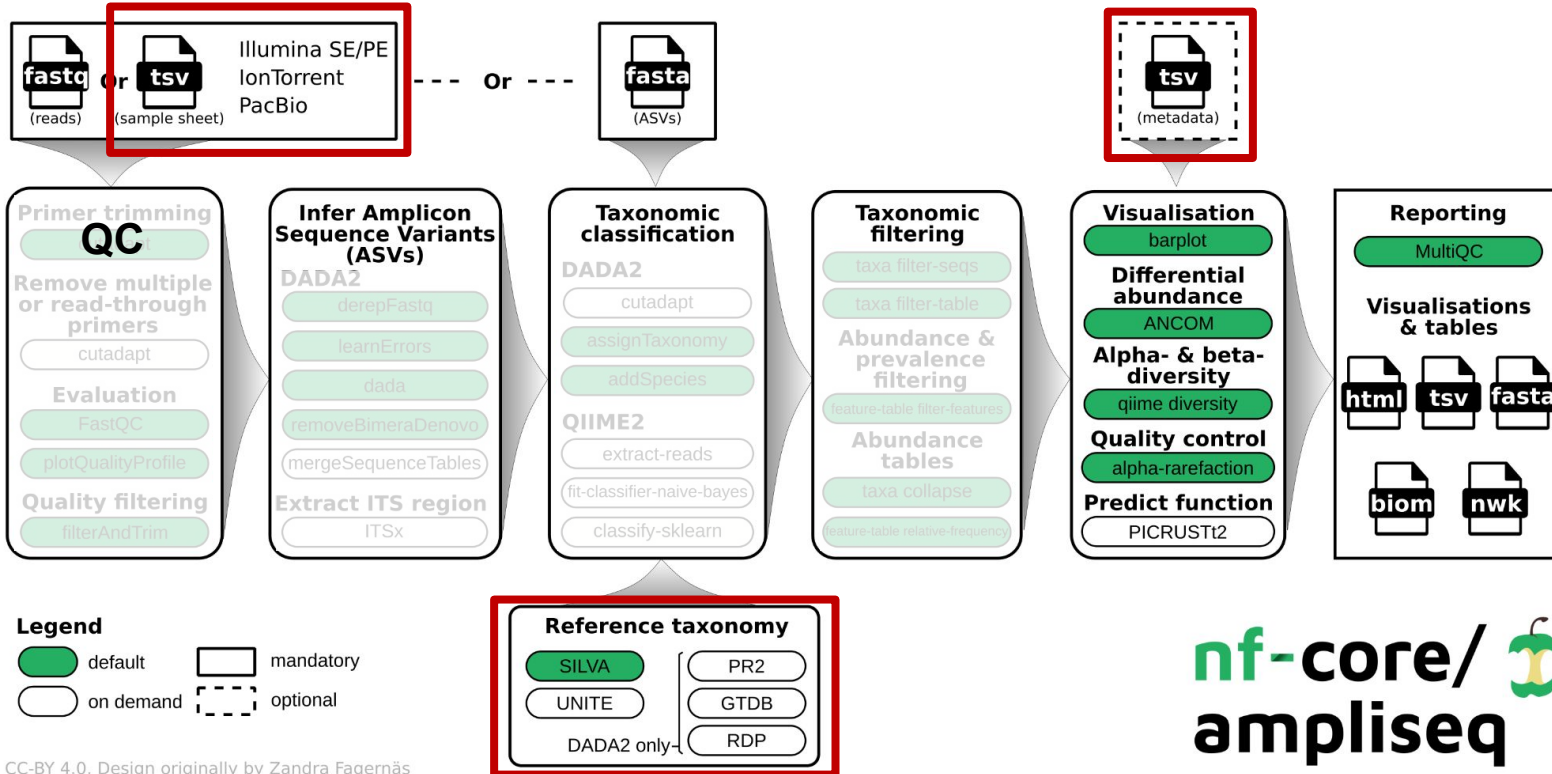
Reference:  
27 bacterial strains with  
35 expected amplicons  
(16S rRNA, V4 region)

# ASV-based DADA2 seems best compromise

	OTU-based				ASV-based	
	Mothur		QIIME1		QIIME2	
	99 <sup>a</sup>	97 <sup>a</sup>	99 <sup>a</sup>	97 <sup>a</sup>	DADA2 <sup>b</sup>	Deblur <sup>b</sup>
Precision (%)	1 ±1	6 ±3	2 ±2	3 ±3	49 ±18	72 ±13
Sensitivity (%)	69 ±7	72 ±11	87 ±6	87 ±6	85 ±12	71 ±25
Taxonomy <sup>c</sup>	49 ±23	49 ±22	63 ±18	63 ±19	77 ±8	80 ±9
Shannon index <sup>d</sup>	-11 ±12	-16 ±19	28 ±23	28 ±23	-6 ±5	-13 ±13



# Amplicon sequencing: nf-core/ampliseq v2.1.1

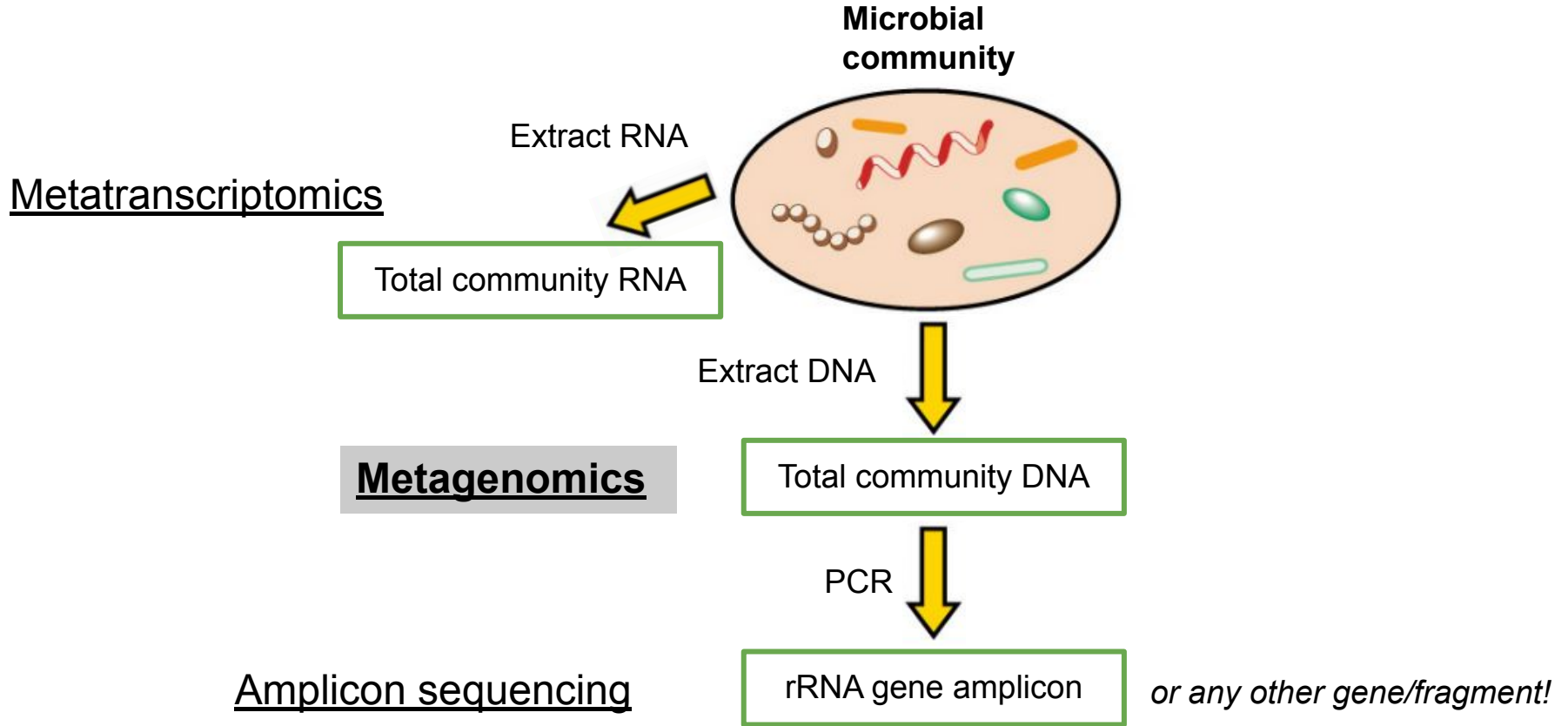


CC-BY 4.0. Design originally by Zandra Fagernäs

nf-core/  ampliseq

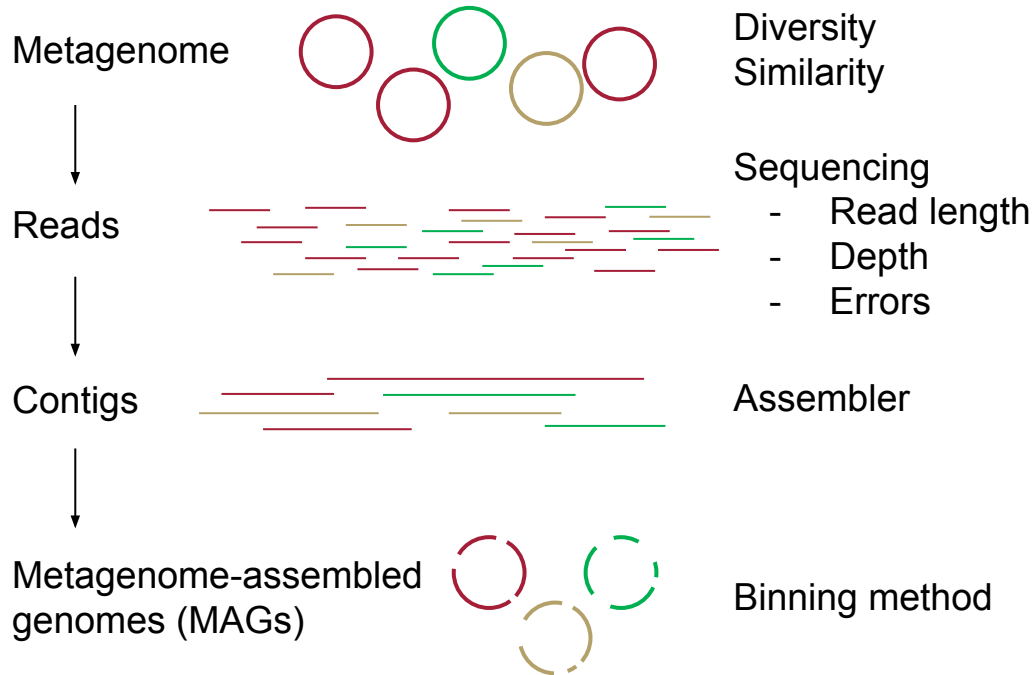








# Recovering genomes of microbial communities



## nature methods

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature methods](#) > [analyses](#) > article

[Open Access](#) | [Published: 02 October 2017](#)

### Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software

[Alexander Sczyrba](#) ✉, [Peter Hofmann](#), [...] [Alice C McHardy](#) ✉

*Nature Methods* **14**, 1063–1071 (2017) | [Cite this article](#)

32k Accesses | 276 Citations | 216 Altmetric | [Metrics](#)



Cold  
Spring  
Harbor  
Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

New Results

Posted July 12, 2021.

### Critical Assessment of Metagenome Interpretation - the second round of challenges

✉ F. Meyer, A. Fritz, Z.-L. Deng, ✉ D. Koslicki, A. Gurevich, G. Robertson, M. Alser, D. Antipov, ✉ F. Beghini, D. Bertrand, J. J. Brito, C.T. Brown, J. Buchmann, A. Buluç, B. Chen, R. Chikhi, P.T. Clausen, A. Cristian, P.W. Dabrowski, A.F. Darling, R. Fran, F. Fskan, F. Georgantas  
doi: <https://doi.org/10.1101/2021.07.12.451567>

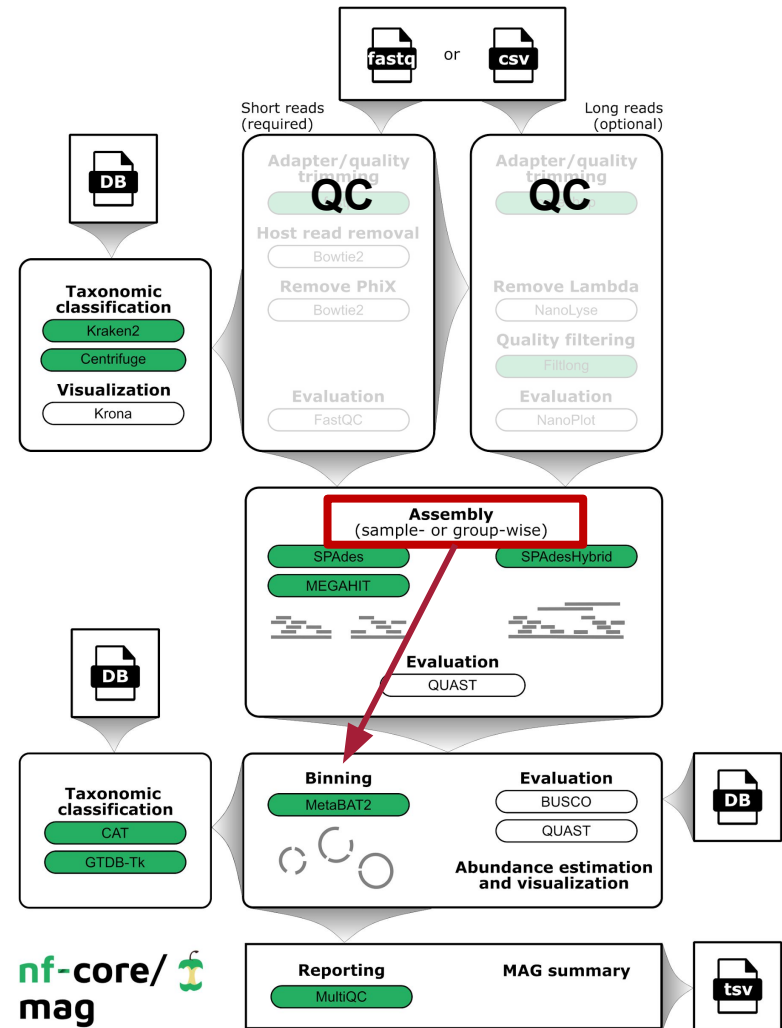


# Metagenome assembly: nf-core/mag v2.1.0

## Input data:

- Illumina (SE/PE)
- Nanopore

One to many samples



# Assembly statistics

## QUAST

QUAST is a quality assessment tool for genome assemblies, written by the Center for Algorithmic Biotechnology.

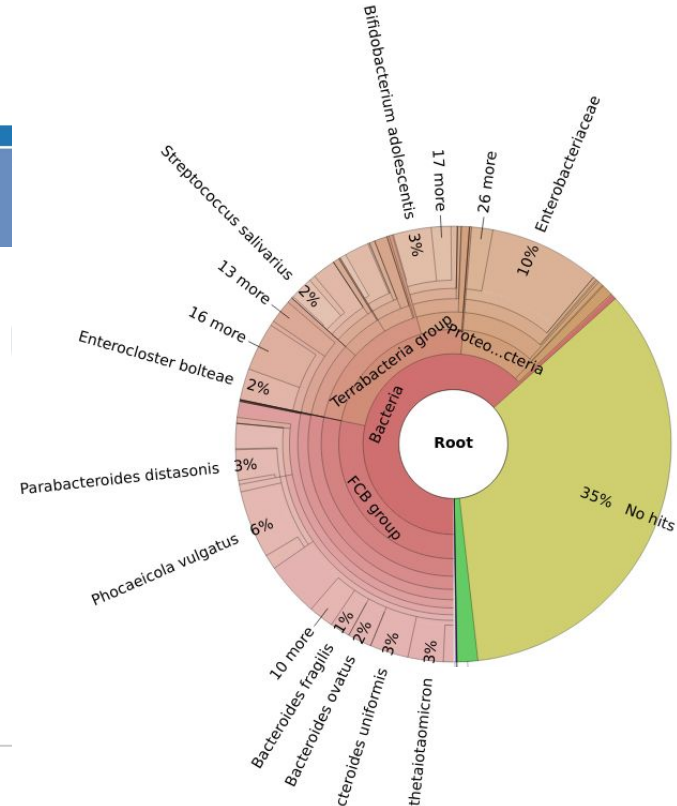
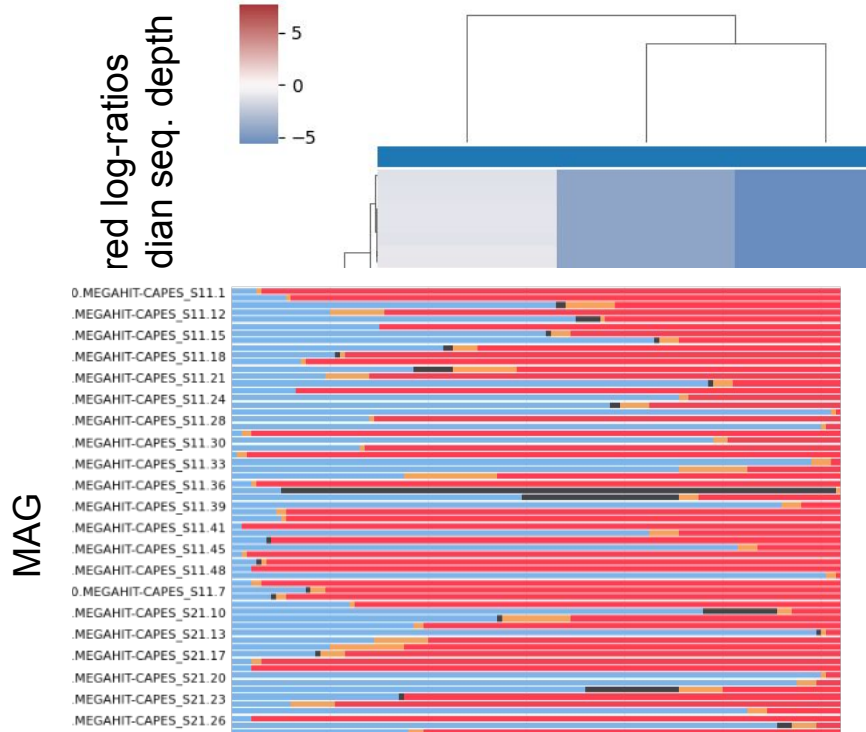
### Assembly Statistics

Showing  $\frac{9}{10}$  rows and  $\frac{6}{8}$  columns.

Sample Name	N50 (Kbp)	N75 (Kbp)	L50 (K)	L75 (K)	Largest contig (Kbp)	Length (Mbp)
MEGAHIT-CAPES_S11	6.0Kbp	1.3Kbp	3.0K	16 077.0K	211.6Kbp	122.3Mbp
MEGAHIT-CAPES_S21	10.3Kbp	1.7Kbp	1.1K	6 773.0K	278.8Kbp	77.9Mbp
MEGAHIT-CAPES_S7	12.9Kbp	3.1Kbp	0.6K	2 911.0K	416.3Kbp	52.8Mbp
SPAdes-CAPES_S11	4.5Kbp	1.2Kbp	3.7K	19 707.0K	183.6Kbp	125.5Mbp
SPAdes-CAPES_S21	9.3Kbp	1.5Kbp	1.2K	7 629.0K	396.3Kbp	79.4Mbp
SPAdes-CAPES_S7	16.0Kbp	3.5Kbp	0.5K	2 507.0K	478.3Kbp	52.9Mbp
SPAdesHybrid-CAPES_S11	7.5Kbp	1.5Kbp	2.7K	14 111.0K	332.9Kbp	135.5Mbp
SPAdesHybrid-CAPES_S21	10.0Kbp	2.1Kbp	1.1K	6 709.0K	486.7Kbp	87.5Mbp
SPAdesHybrid-CAPES_S7	47.4Kbp	11.1Kbp	0.2K	950.0K	3 915.6Kbp	64.4Mbp



# MAG summary includes depth, completeness and taxonomy





**Metatranscriptomics**

Extract RNA



Total community RNA

**Metagenomics**

Extract DNA



Total community DNA

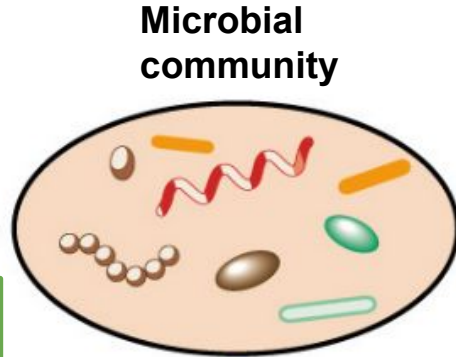
PCR



rRNA gene amplicon

*or any other gene/fragment!*

**Amplicon sequencing**



**Microbial  
community**

# Quantification of gene transcripts

What are active processes (under each condition)?

How are organisms adapting their gene expression?

Transcript measurements

RNA-Seq reads

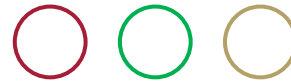


Quantification



Reference sequences

Genome(s)



OR

Contigs  
(Assembly)



& Gene  
coordinates

OR

Gene  
sequences



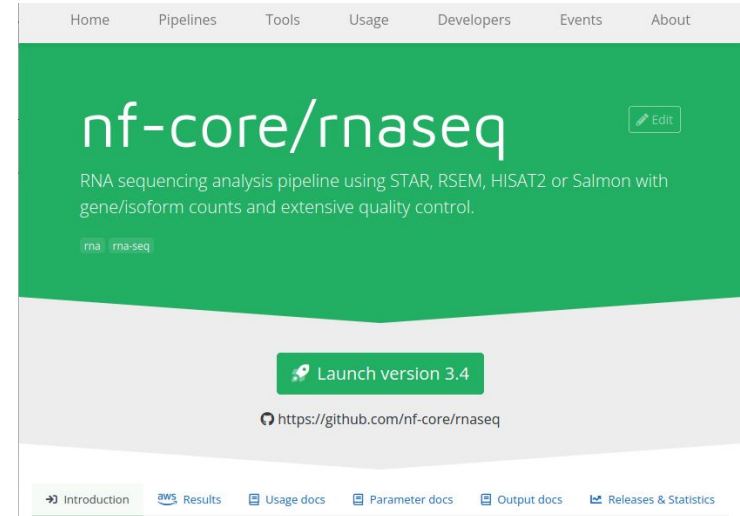




# Shotgun metatranscriptomics

## nf-core/rnaseq v3.4

- Read quality control
- Adapter and quality trimming
- Removal of ribosomal RNA
- Alignment and quantification
- Quality control



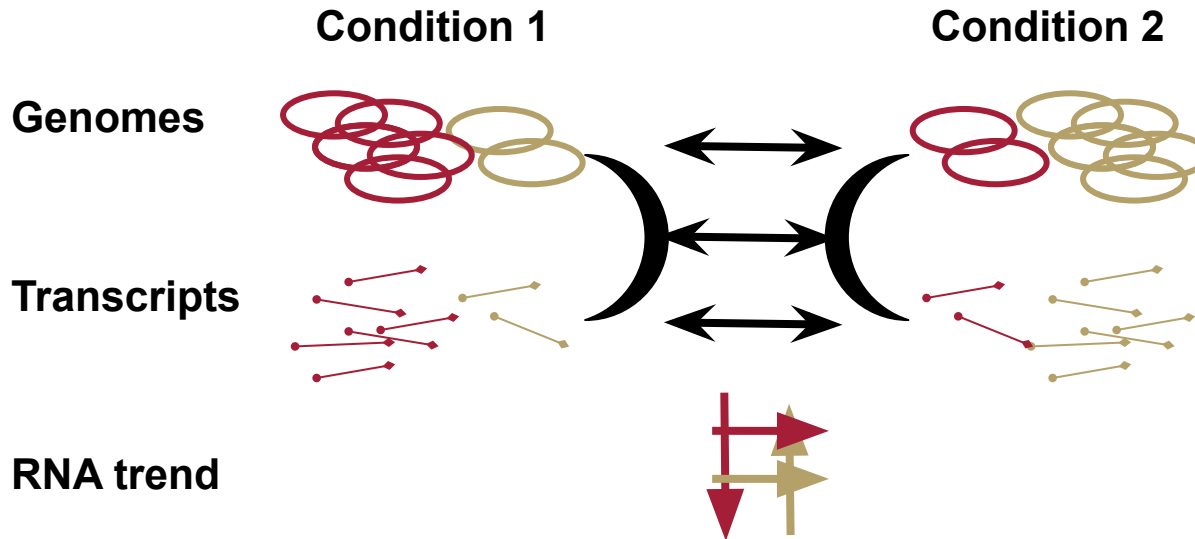
<https://nf-co.re/rnaseq>

Differential transcript abundances with DESeq2



# Differential transcript abundance / gene expression

Usual workflow → Ignores the underlying gene/genome abundances





# Metagenomics – metatranscriptomics pairs

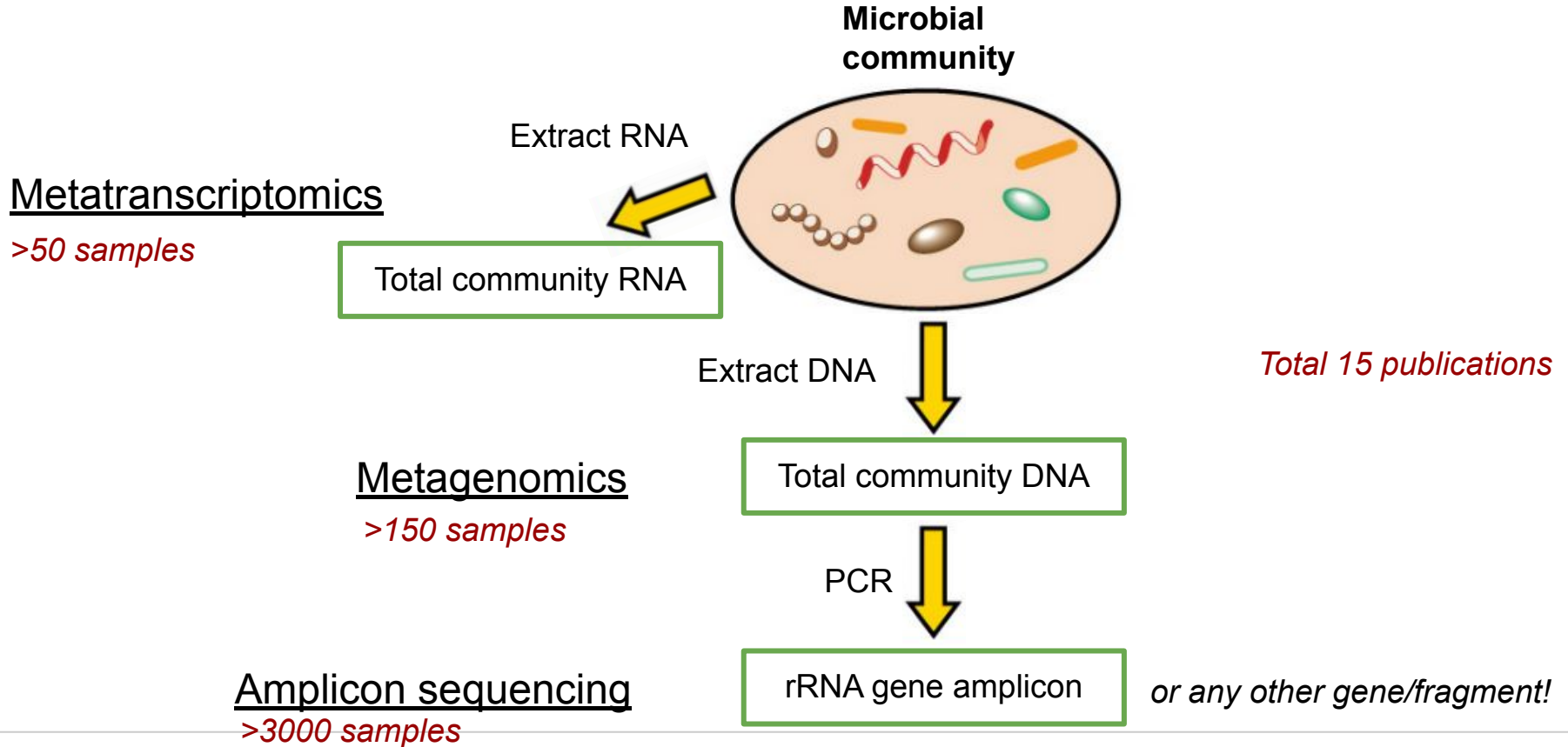
- What are active processes under each condition?
  - Transcript abundances: How many transcripts per condition
- What are organisms doing under each condition? How are organisms adapting their gene expression?
  - Gene expression: Transcript abundances per cell/genome per condition

Genomes



Transcripts







# Thanks for your attention!



Sabrina Krakau  
Alexander Peltzer  
Gisela Gabernet  
Sven Nahnsen

