



nf-core/ampliseq

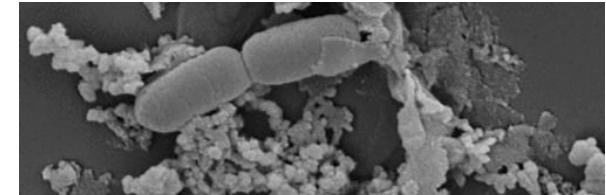
<https://nf-co.re/ampliseq>

by Daniel Straub

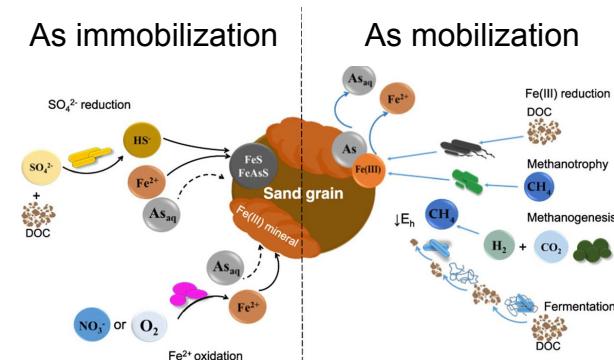


Microbial communities are ubiquitous

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



Jakus et al. (2021) Environmental Microbiology

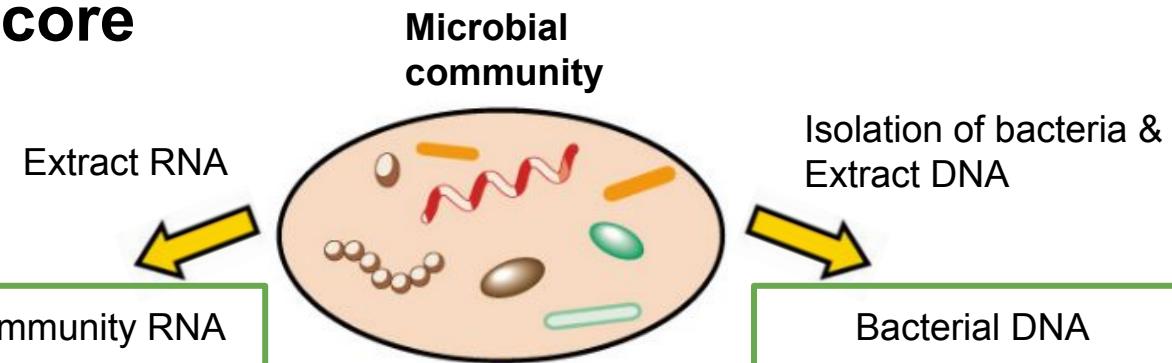


Glodowska et al. (2021) Journal of Hazardous Materials



Meta-omics at nf-core

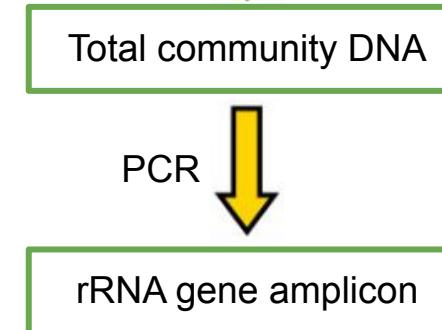
Metatranscriptome
nf-core/rnaseq



Metagenome assembly
nf-core/mag



Amplicon sequencing
nf-core/ampliseq



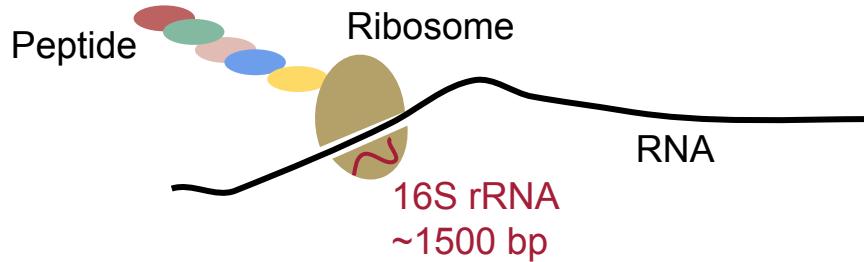
Isolation of bacteria &
Extract DNA

Bacterial DNA

Genome assembly
nf-core/bacass



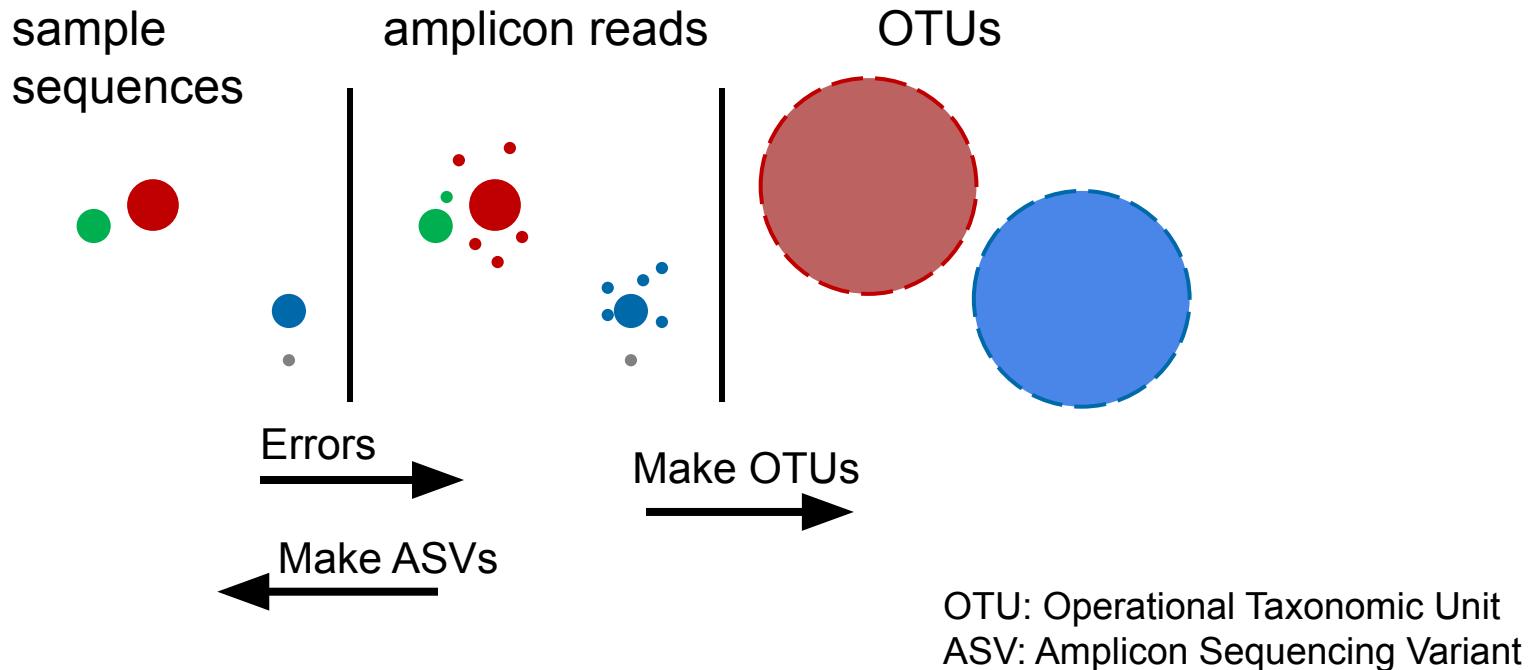
The 16S rRNA gene is a good marker for bacteria



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

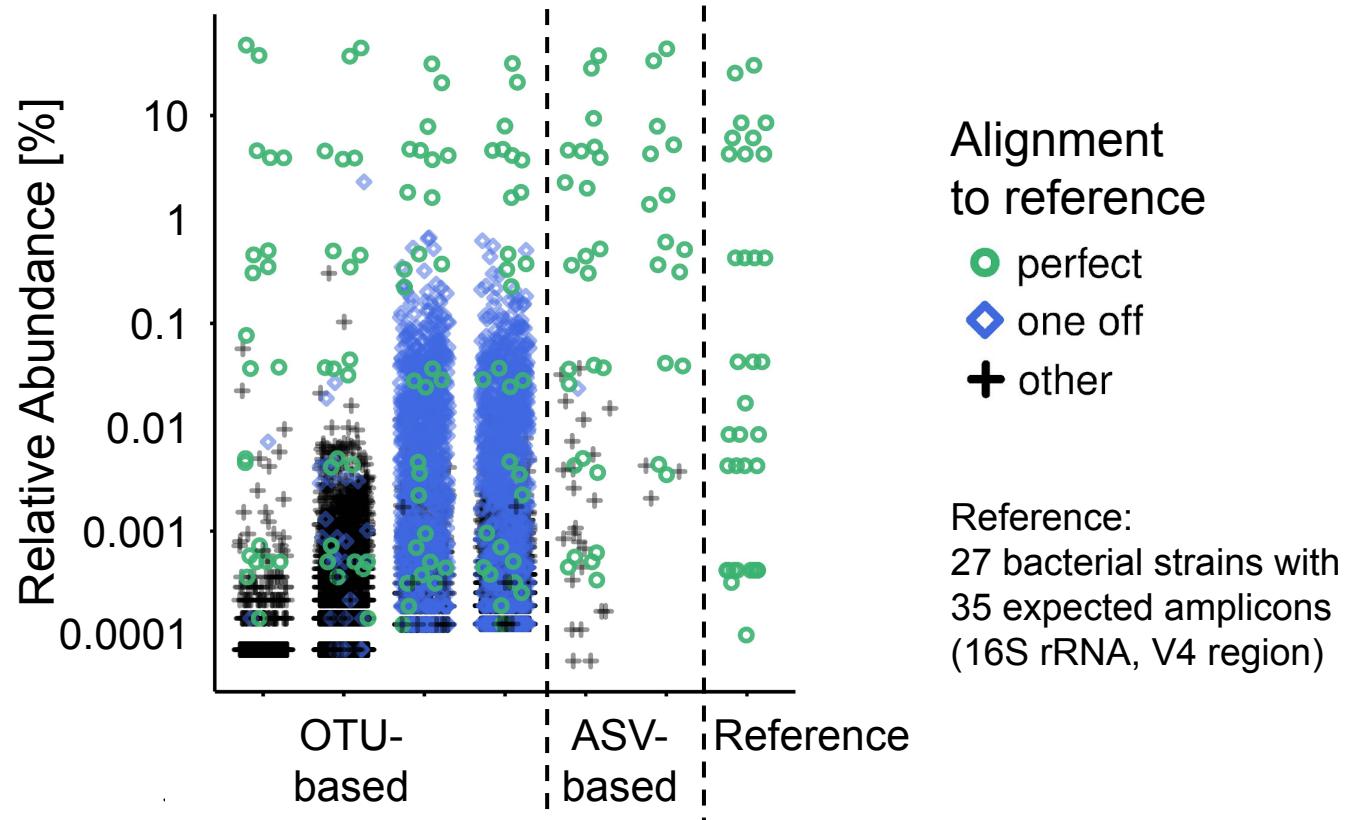


Error accumulation during library prep and sequencing





Evaluation: Sequences





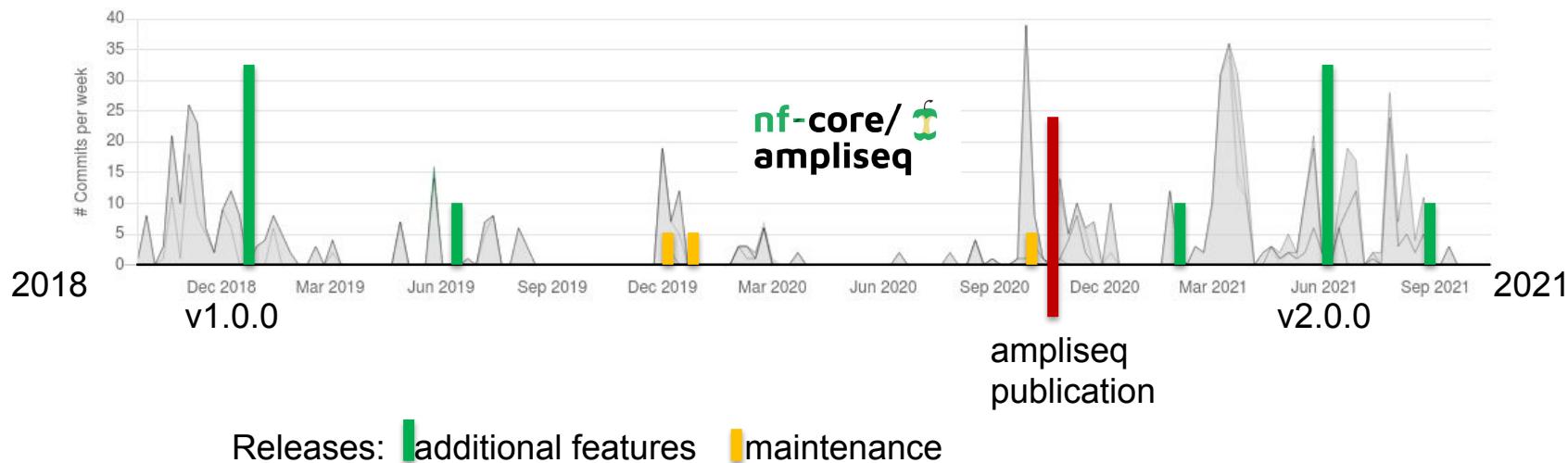
ASV-based DADA2 seems best compromise

	OTU-based		ASV-based		other	
	Mothur	QIIME1	QIIME2	DADA2 ^b	Deblur ^b	MEGAN
Precision (%)	99 ^a	97 ^a	99 ^a	97 ^a	49 ±18	72 ±13
Sensitivity (%)	1 ±1	6 ±3	2 ±2	3 ±3	49 ±18	n.d.
Taxonomy ^c	69 ±7	72 ±11	87 ±6	87 ±6	85 ±12	71 ±25
Shannon index ^d	49 ±23	49 ±22	63 ±18	63 ±19	77 ±8	80 ±9
	-11 ±12	-16 ±19	28 ±23	28 ±23	-6 ±5	-13 ±13
						-30 ±17

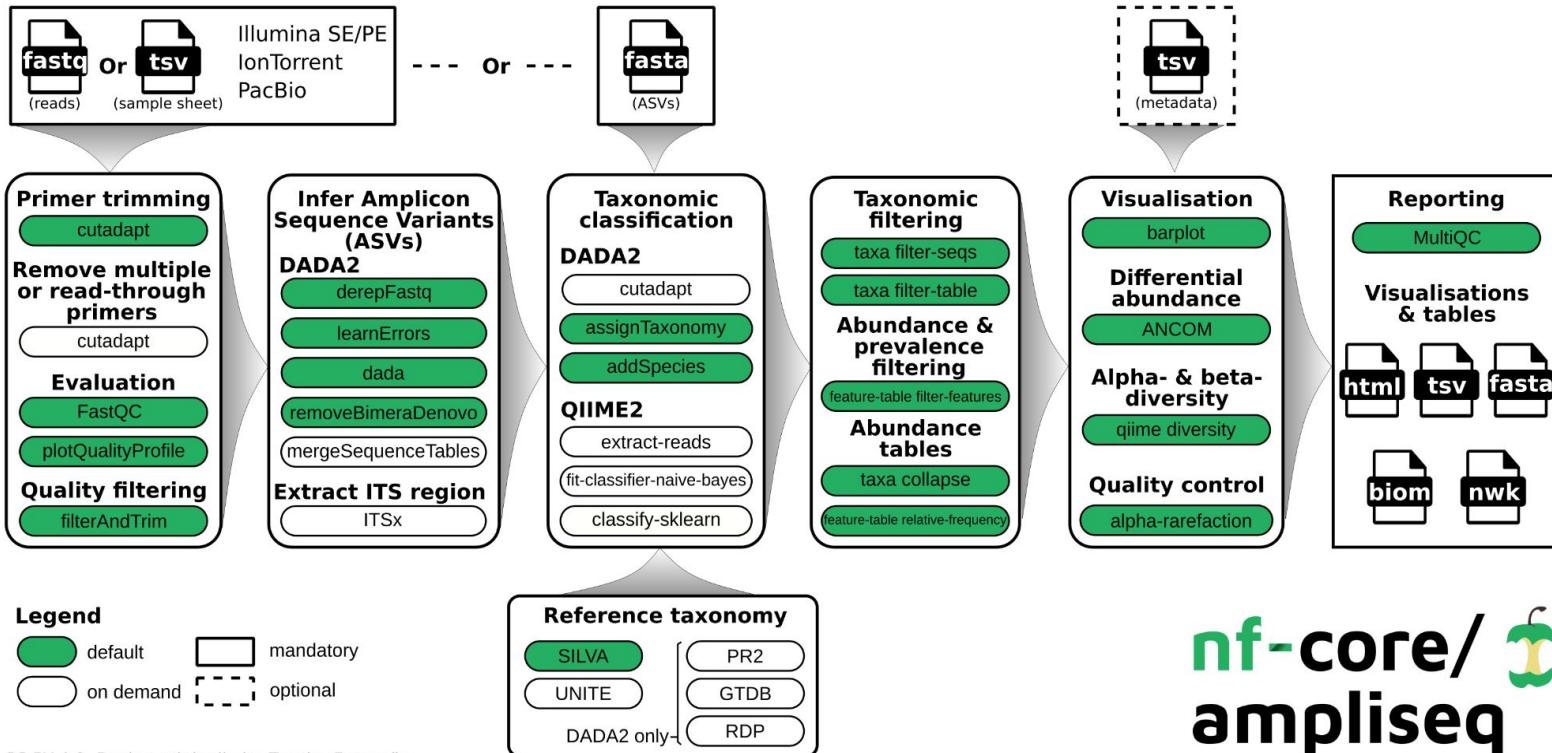


One pipeline to analyse all kinds of amplicon data

Wide range of input types, reference taxonomies, downstream analysis



Amplicon sequencing: nf-core/ampliseq v2.1.0

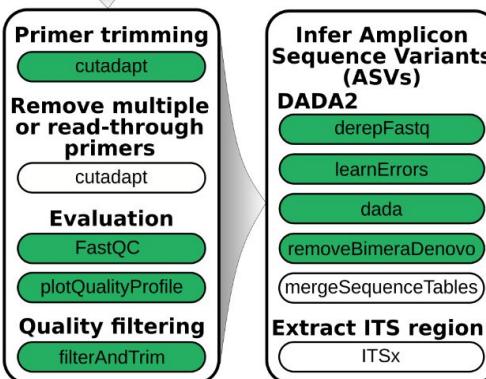
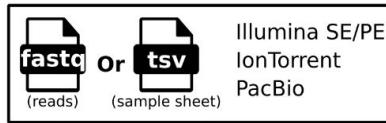


CC-BY 4.0. Design originally by Zandra Fagernäs

**nf-core/
ampliseq**



nf-core/ampliseq v2.1.0: Input & basic analysis



mandatory

sampleID	forwardReads	reverseReads	run
sample1	./data/S1_R1_001.fastq.gz	./data/S1_R2_001.fastq.gz	A
sample2	./data/S2_fw.fastq.gz	./data/S2_rv.fastq.gz	A
sample3	./S4x.fastq.gz	./S4y.fastq.gz	B
sample4	./a.fastq.gz	./b.fastq.gz	B

samplesheet.tsv

optional

Illumina PE: default
Illumina SE: --single_end
PacBio: --pacbio
IonTorrent: --iontorrent
Illumina PE ITS: --illumina_pe_its

```
nextflow run nf-core/ampliseq -r 2.1.0 -profile singularity --input 'samplesheet.tsv'  
--FW_primer 'GTGYCAGCMGCCGCGGTAA' --RV_primer 'GGACTACNVGGGTWTCTAAT'
```



nf-core/ampliseq v2.1.0: Basic output

results/dada2/DADA2_table.tsv

- results/dada2/ASV_seqs.fasta
- results/dada2/ASV_table.tsv

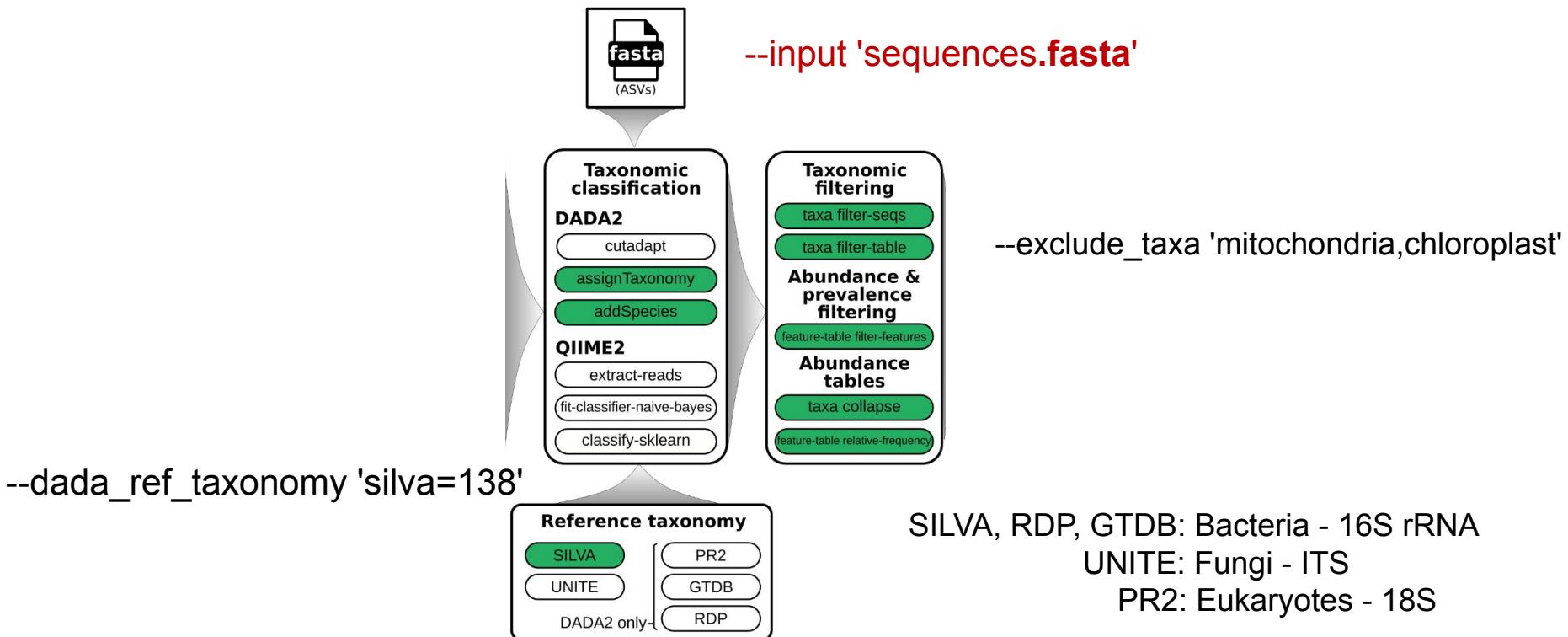
	A	B	C	D	E	
1	ASV_ID	sampleID_1	sampleID_1a	sampleID_2	sampleID_2a	sequence
2	b39be459cc96689db8cd6b00a0d86e8	153	0	317	0	TACGTAGGGTGCAGCGTTAA
3	5da32e051bdda695c26f4cc556c3aed	60	0	121	0	AAACCAGCACCTCAAGTGGCA
4	df915ccb4ee135b84e0b244e8af12bf92	0	97	0	80	CCCAGGAGCCACAAGTGGCA
5	3ebe761bf1238c87195d431f41bf976	86	16	57	0	TACAGAGGGTGCAGCGTTAA
6	38c27cead634984c1225a82648cf57	75	0	35	37	TACAGAGGGTGCAGCGTTAA
7	e8386d3a307c208c4b9f0a756259cd6b	81	0	44	18	TACAGAGGGTGCAGCGTTAA
8	b7d6cf3f7467ca45087ca083796a0af0	52	0	85	0	TACGTAGGGTGCAGCGTTAA
9	1f8747026500ce37ab66b935d9b1b53e	0	62	0	69	GACAGAGGTGGCAAGCGTTG
10	d88abff2ed5af388abe3bd71809f70e1	6	51	0	48	GACAGAGGTGGCAAGCGTTG
11	02b4cbe5c2955ab2f66dc82cdf6dad64	39	0	27	24	TACGTAGGGTGCAGCGTTAA
12	586ee8bd7ab4a9a8074517b760b32ac	0	36	0	42	GACAGAGGTGGCAAGCGTTG
13	3421464bea79a54e7ee62009457230b	0	38	0	36	GACGTAGGATGCAAGCGTTAC
14	637b9b3f4d1cbb1a10c07817619cdf69	21	12	21	17	TACAGAGGGTGCAGCGTTAA
15	7346ee25c4de4df48f80d736342c1c02	24	5	32	5	GACAGAGGTGGCAAGCGTTG
16	400b6000200310241400000000017d0600	0	0	0	24	CCCCACCCCCACACACTCCCCA

results/overall_summary.tsv

	A	B	D	I	J	K	L	M	N
1	sample	cutadapt_total_processed	cutadapt_passing_filters	DADA2_input	filtered	denoisedF	denoisedR	merged	nonchim
2	sampleID_1	2500	1812	1812	1446	1233	1171	1072	1072
3	sampleID_1a	2500	2221	2221	1986	1369	1328	1164	1164
4	sampleID_2	2500	2350	2347	1928	1579	1542	1409	1409
5	sampleID_2a	2500	2244	2244	2008	1479	1464	1253	1253
6									



nf-core/ampliseq v2.1.0: Taxonomic classification





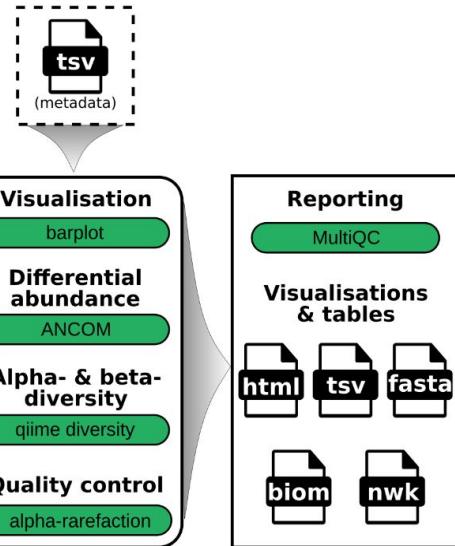
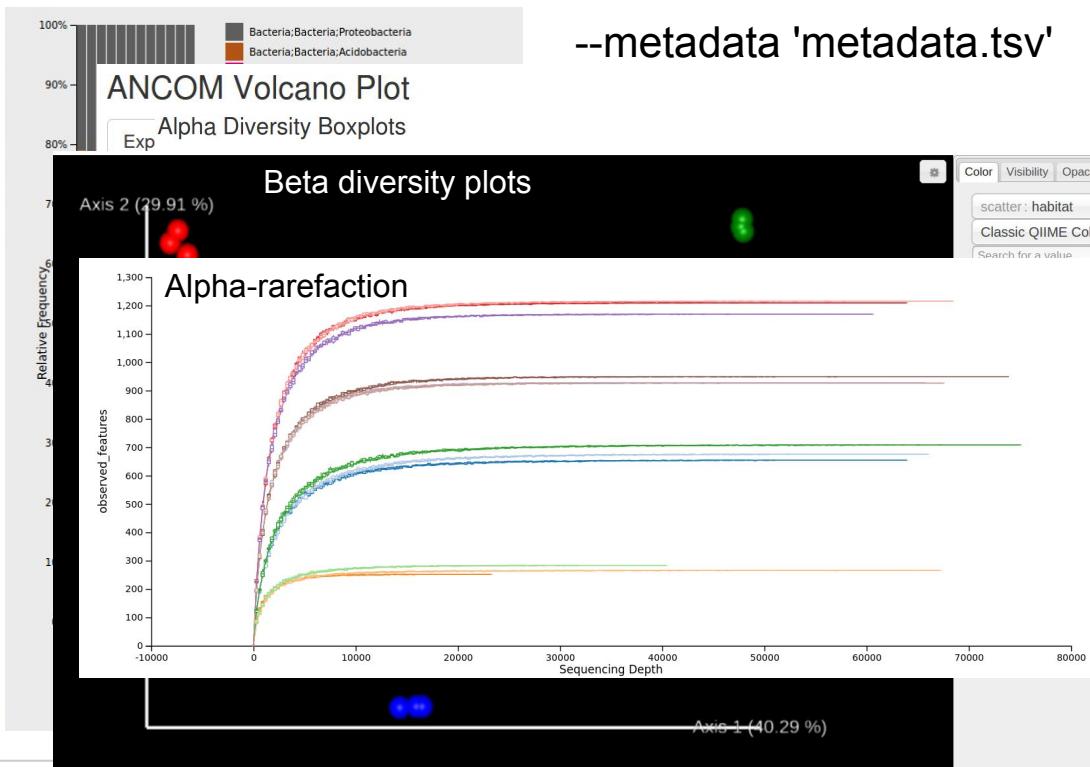
nf-core/ampliseq v2.1.0: Taxonomic classification output

	A	B	C	D	E	F	G	H	I	J	confidence	sequence
1	ASV_ID	Domain	Kingdom	Phylum	Class	Order	Family	Genus	Species			
2	99d7dd8ffd43cef331f256e86f097a82	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	NA	NA	NA		0.61	TACGTAGGGTGCAAG
3	b39be459cc96689db8cd600a0d86e0	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Gallionellaceae	Sideroxydans	NA		0.68	TACGTAGGGTGCGAG
4	7c342291a1c802d1945bf508a6ca867	Bacteria	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	NA		1	TACGGAGGATCCAAG
5	b7d6cf3f7467ca45087ca083796a0af0	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Polaromonas	vacuolata		0.57	TACGTAGGGTGCGAG
6	040bd74a5c9b1a37501cd9fc4c55432	Archaea	Archaea	Thaumarchaeota	Nitrosphaerales	Nitrosphaeraceae	Nitrosphaera	NA	NA		1	TACCATCAGCACCCGAG
7	cfd66c1310f85d586280b6dbd829635	Bacteria	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	NA		1	TACGGAGGATCCAAG
8	5da32e051bbda695c26f4cc556c3aed	Archaea	Archaea	Thaumarchaeota	Nitrosopumilales	Nitrosopumilaceae	Nitrosopumilus	NA	NA		1	AACCAGCACCTCAAG
9	3ebe761fb1238c87195d431f41bf976	Bacteria	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	NA		1	TACAGAGGGTGCAG
10	095dcca0alcacb655451100eb59c8ef	Bacteria	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Agitococcus	NA		0.56	TACAGAGGGTGCAG
11	6b235822a9e52c63e60c1bd0061cb11	Bacteria	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	NA		1	TACGGAGGATCCAAG
12	b0546a8c6a38415c52d2dc916319d3f	Bacteria	Bacteria	Verrucomicrobia	Spartobacteria	NA	NA	NA	NA		1	TACAGAGGGTCTCAAG
13	e8386d3a307c208c4b90fa756259cd6	Bacteria	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	NA		1	TACAGAGGGTGCAG
14	a9cdace376f468b22ea7ba7de7e029	Bacteria	Bacteria	Proteobacteria	Delta proteobacteria	NA	NA	NA	NA		0.75	TACGGAGGGTGCAG
15	dbe9c1a17d034dc956466d330b31880	Archaea	Archaea	Thaumarchaeota	Nitrosphaerales	Nitrosphaeraceae	Nitrosphaera	NA	NA		1	TACCATCAGCACCCGAG
16	f34cdcd73e535507d29c29b19d01f658	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Rhodoferax	NA		0.68	TACGTAGGGTGCGAG
17	27af01f1ebcc0867bb168193e6a9d27	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales_incep	Sphaerotilus	NA		0.89	TACGTAGGGTGCAAG
18	107273e73071282e4d0f4a0b46548da	Bacteria	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas	NA		1	TACGGAGGGTGCAG
19	38c27ceaed634984c1225a82648cf57	Bacteria	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	NA		0.99	TACAGAGGGTGCAG
20	868e0165fe32a8925b7f120af5e1dbfc	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Acidovorax	NA		0.96	TACGTAGGGTGCAAG
21	c58e5bf355be030d93e26cfa7e766dab	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Rhizobacter	NA		0.51	TACGTAGGGTGCAAG
22	ab32275a7fc06a23c531b394cf612f	Bacteria	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Malikia	NA		0.67	TACGTAGGGTGCAAG
23	25e119f38ff7d58bf3442bd40eaf00	Bacteria	Bacteria	NA	NA	NA	NA	NA	NA		1	TACGGAGGGTGCAG
24	2fac845da336657d551a9f350c6bc09	Bacteria	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Hyphomicrobiu	NA		0.84	TACGGAGGGACTAG
25	8562afa3955423d4eda6d71b8b277d8	Bacteria	Bacteria	Acidobacteria	Acidobacteria_Gp4	Gp4	NA	NA	NA		1	TACGGGGGGAGCAAG

results/dada2/ASV_tax_species.tsv



nf-core/ampliseq v2.1.0: Downstream analysis



<https://nf-co.re/ampliseq/results>



Thanks for your attention!



Alexander Peltzer
Sabrina Krakau



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

Daniel Lundin

Diego Brambilla
Emelie Nilsson
Jeanette Tångrot

