

Summary of the bioinformatic pipeline

Process	Software	Options	Source
Quality assessment of reads and trimming	FastQC obicut	- -	FastQC OBITools
Paired-end merging	illuminapairedend	Discarded score<40.00	OBITools
Primer trimming	cutadapt	-	CutAdapt
Length selection	obigrep	-p 'seq_length>310' -p 'seq_length<330' (COI) / -p 'seq_length>290' -p 'seq_length<370' (16S) -s '^ [ACGT]+\$'	OBITools
Dereplication	obiuniq	-m sample	OBITools
Reads abundance counting	obitab	-	OBITools
Chimera removal	uchime_denovo	--sizeout --nonchimeras <i>nonchimeras.fasta</i>	VSEARCH
Clustering	swarm	-d 13 (COI) / -d 6 (16S) - z -o <i>outputfile</i> -w <i>seeds.fasta</i>	SWARM
MOTUs abundance counting	owi_reciunt_swarm	-	Metabarpark

Matchlist seed sequences	usearch_global	<pre>--db seeds.fasta --self --id .84 --userout matchlist.txt --userfields query+target+id --maxaccepts 0 --query_cov .9 --maxhits 10</pre>	VSEARCH
Co-occurrence curation	lulu	-	LULU
Taxonomic identification	ecotag owi_add_taxonomy	-	OBITools Metabarpark
MOTUs table	owi_combine	-	Metabarpark
Low absolute and relative (per sample) abundance correction	refine_MOTU_table	<pre>-t 0 -r 10 -a yes</pre>	Metagusano

Sources

- OBITools: <https://pythonhosted.org/OBITools/welcome.html>
- Cutadapt: <https://cutadapt.readthedocs.io/en/stable/>
- VSEARCH: <https://github.com/torognes/vsearch>
- SWARM: <https://github.com/torognes/swarm>
- Metabarpark: https://github.com/metabarpark/R_scripts_metabarpark
- LULU: <https://github.com/tobiasgf/lulu>
- Metagusano: https://github.com/metagusano/metabarcoding_scripts