## **Trimming with Trimgalore**

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_1\_12S\_1.fastq.gz ../raw\_data/1\_1\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_1\_COI\_1.fastq.gz ../raw\_data/1\_1\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_2\_12S\_1.fastq.gz ../raw\_data/1\_2\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_2\_COI\_1.fastq.gz ../raw\_data/1\_2\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_3\_12S\_1.fastq.gz ../raw\_data/1\_3\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_3\_COI\_1.fastq.gz ../raw\_data/1\_3\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_4\_12S\_1.fastq.gz ../raw\_data/1\_4\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_4\_COI\_1.fastq.gz ../raw\_data/1\_4\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_5\_12S\_1.fastq.gz ../raw\_data/1\_5\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_5\_COI\_1.fastq.gz ../raw\_data/1\_5\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_6\_12S\_1.fastq.gz ../raw\_data/1\_6\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/1\_6\_COI\_1.fastq.gz ../raw\_data/1\_6\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_1a\_12S\_1.fastq.gz ../raw\_data/2\_1a\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_1a\_COI\_1.fastq.gz ../raw\_data/2\_1a\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_1b\_12S\_1.fastq.gz ../raw\_data/2\_1b\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_1b\_COI\_1.fastq.gz ../raw\_data/2\_1b\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_2\_12S\_1.fastq.gz ../raw\_data/2\_2\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_2\_COI\_1.fastq.gz ../raw\_data/2\_2\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_3\_12S\_1.fastq.gz ../raw\_data/2\_3\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/2\_3\_COI\_1.fastq.gz ../raw\_data/2\_3\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_1\_12S\_1.fastq.gz ../raw\_data/3\_1\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_1\_COI\_1.fastq.gz ../raw\_data/3\_1\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_2a\_12S\_1.fastq.gz ../raw\_data/3\_2a\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_2a\_COI\_1.fastq.gz ../raw\_data/3\_2a\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_2b\_12S\_1.fastq.gz ../raw\_data/3\_2b\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_2b\_COI\_1.fastq.gz ../raw\_data/3\_2b\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_3\_12S\_1.fastq.gz ../raw\_data/3\_3\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/3\_3\_COI\_1.fastq.gz ../raw\_data/3\_3\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_1\_12S\_1.fastq.gz ../raw\_data/4\_1\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_1\_COI\_1.fastq.gz ../raw\_data/4\_1\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_2\_12S\_1.fastq.gz ../raw\_data/4\_2\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_2\_COI\_1.fastq.gz ../raw\_data/4\_2\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_3\_12S\_1.fastq.gz ../raw\_data/4\_3\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_3\_COI\_1.fastq.gz ../raw\_data/4\_3\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_4\_12S\_1.fastq.gz ../raw\_data/4\_4\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_4\_COI\_1.fastq.gz ../raw\_data/4\_4\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_5\_12S\_1.fastq.gz ../raw\_data/4\_5\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_5\_COI\_1.fastq.gz ../raw\_data/4\_5\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_6\_12S\_1.fastq.gz ../raw\_data/4\_6\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/4\_6\_COI\_1.fastq.gz ../raw\_data/4\_6\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/5\_1\_12S\_1.fastq.gz ../raw\_data/5\_1\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/5\_1\_COI\_1.fastq.gz ../raw\_data/5\_1\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/5\_2\_12S\_1.fastq.gz ../raw\_data/5\_2\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/5\_2\_COI\_1.fastq.gz ../raw\_data/5\_2\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/5\_3\_12S\_1.fastq.gz ../raw\_data/5\_3\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/5\_3\_COI\_1.fastq.gz ../raw\_data/5\_3\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_1a\_12S\_1.fastq.gz ../raw\_data/6\_1a\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_1a\_COI\_1.fastq.gz ../raw\_data/6\_1a\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_1b\_12S\_1.fastq.gz ../raw\_data/6\_1b\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_1b\_COI\_extracted\_1.fastq.gz ../raw\_data/6\_1b\_COI\_extracted\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_2\_12S\_1.fastq.gz ../raw\_data/6\_2\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_2\_COI\_extracted\_1.fastq.gz ../raw\_data/6\_2\_COI\_extracted\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_3\_12S\_1.fastq.gz ../raw\_data/6\_3\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/6\_3\_COI\_1.fastq.gz ../raw\_data/6\_3\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/H2O-89A\_\_12S\_extracted\_1.fastq.gz ../raw\_data/H2O-89A\_\_12S\_extracted\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/H2O-89B\_12S\_extracted\_1.fastq.gz ../raw\_data/H2O-89B\_12S\_extracted\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/H2O-90A\_\_COI\_1.fastq.gz ../raw\_data/H2O-90A\_\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/H2O90-B\_\_COI\_1.fastq.gz ../raw\_data/H2O90-B\_\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/nega\_ctrl\_1\_COI\_1.fastq.gz ../raw\_data/nega\_ctrl\_1\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/neg\_ctrl\_1\_12S\_1.fastq.gz ../raw\_data/neg\_ctrl\_1\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/neg\_ctrl\_2\_12S\_1.fastq.gz ../raw\_data/neg\_ctrl\_2\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/neg\_ctrl\_3\_12S\_1.fastq.gz ../raw\_data/neg\_ctrl\_3\_12S\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/neg\_ctrl\_3\_COI\_1.fastq.gz ../raw\_data/neg\_ctrl\_3\_COI\_2.fastq.gz

trim\_galore --paired --quality 30 --length 140 ../raw\_data/neg\_ctrll\_2\_COI\_1.fastq.gz ../raw\_data/neg\_ctrll\_2\_COI\_2.fastq.gz

## **remove PCR duplicates using Stacks**

also used to remove Ns and primer (COI: 6 Ns plus 20 bp primer; 12 S forward: 6 Ns plus 22 bp Primer; 12S reverse 6Ns plus 28 bp primer)

clone\_filter -1 ./1\_1\_COI\_1\_val\_1.fq.gz -2 ./1\_1\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./1\_2\_COI\_1\_val\_1.fq.gz -2 ./1\_2\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./1\_3\_COI\_1\_val\_1.fq.gz -2 ./1\_3\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./1\_4\_COI\_1\_val\_1.fq.gz -2 ./1\_4\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./1\_5\_COI\_1\_val\_1.fq.gz -2 ./1\_5\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./1\_6\_COI\_1\_val\_1.fq.gz -2 ./1\_6\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./2\_1a\_COI\_1\_val\_1.fq.gz -2 ./2\_1a\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./2\_1b\_COI\_1\_val\_1.fq.gz -2 ./2\_1b\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./2\_2\_COI\_1\_val\_1.fq.gz -2 ./2\_2\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./2\_3\_COI\_1\_val\_1.fq.gz -2 ./2\_3\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./3\_1\_COI\_1\_val\_1.fq.gz -2 ./3\_1\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./3\_2a\_COI\_1\_val\_1.fq.gz -2 ./3\_2a\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./3\_2b\_COI\_1\_val\_1.fq.gz -2 ./3\_2b\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./3\_3\_COI\_1\_val\_1.fq.gz -2 ./3\_3\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./4\_1\_COI\_1\_val\_1.fq.gz -2 ./4\_1\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./4\_2\_COI\_1\_val\_1.fq.gz -2 ./4\_2\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./4\_3\_COI\_1\_val\_1.fq.gz -2 ./4\_3\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./4\_4\_COI\_1\_val\_1.fq.gz -2 ./4\_4\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./4\_5\_COI\_1\_val\_1.fq.gz -2 ./4\_5\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./4\_6\_COI\_1\_val\_1.fq.gz -2 ./4\_6\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./5\_1\_COI\_1\_val\_1.fq.gz -2 ./5\_1\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./5\_2\_COI\_1\_val\_1.fq.gz -2 ./5\_2\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./5\_3\_COI\_1\_val\_1.fq.gz -2 ./5\_3\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./6\_1a\_COI\_1\_val\_1.fq.gz -2 ./6\_1a\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./6\_1b\_COI\_extracted\_1\_val\_1.fq.gz -2 ./6\_1b\_COI\_extracted\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./6\_2\_COI\_extracted\_1\_val\_1.fq.gz -2 ./6\_2\_COI\_extracted\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./6\_3\_COI\_1\_val\_1.fq.gz -2 ./6\_3\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./H2O-90A\_\_COI\_1\_val\_1.fq.gz -2 ./H2O-90A\_\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./H2O90-B\_\_COI\_1\_val\_1.fq.gz -2 ./H2O90-B\_\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./nega\_ctrl\_1\_COI\_1\_val\_1.fq.gz -2 ./nega\_ctrl\_1\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./neg\_ctrl\_3\_COI\_1\_val\_1.fq.gz -2 ./neg\_ctrl\_3\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./neg\_ctrll\_2\_COI\_1\_val\_1.fq.gz -2 ./neg\_ctrll\_2\_COI\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 26 --oligo\_len\_2 26

clone\_filter -1 ./1\_1\_12S\_1\_val\_1.fq.gz -2 ./1\_1\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./1\_2\_12S\_1\_val\_1.fq.gz -2 ./1\_2\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./1\_3\_12S\_1\_val\_1.fq.gz -2 ./1\_3\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./1\_4\_12S\_1\_val\_1.fq.gz -2 ./1\_4\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./1\_5\_12S\_1\_val\_1.fq.gz -2 ./1\_5\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./1\_6\_12S\_1\_val\_1.fq.gz -2 ./1\_6\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./2\_1a\_12S\_1\_val\_1.fq.gz -2 ./2\_1a\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./2\_1b\_12S\_1\_val\_1.fq.gz -2 ./2\_1b\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./2\_2\_12S\_1\_val\_1.fq.gz -2 ./2\_2\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./2\_3\_12S\_1\_val\_1.fq.gz -2 ./2\_3\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./3\_1\_12S\_1\_val\_1.fq.gz -2 ./3\_1\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./3\_2a\_12S\_1\_val\_1.fq.gz -2 ./3\_2a\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./3\_2b\_12S\_1\_val\_1.fq.gz -2 ./3\_2b\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./3\_3\_12S\_1\_val\_1.fq.gz -2 ./3\_3\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./4\_1\_12S\_1\_val\_1.fq.gz -2 ./4\_1\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./4\_2\_12S\_1\_val\_1.fq.gz -2 ./4\_2\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./4\_3\_12S\_1\_val\_1.fq.gz -2 ./4\_3\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./4\_4\_12S\_1\_val\_1.fq.gz -2 ./4\_4\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./4\_5\_12S\_1\_val\_1.fq.gz -2 ./4\_5\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./4\_6\_12S\_1\_val\_1.fq.gz -2 ./4\_6\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./5\_1\_12S\_1\_val\_1.fq.gz -2 ./5\_1\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./5\_2\_12S\_1\_val\_1.fq.gz -2 ./5\_2\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./5\_3\_12S\_1\_val\_1.fq.gz -2 ./5\_3\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./6\_1a\_12S\_1\_val\_1.fq.gz -2 ./6\_1a\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./6\_1b\_12S\_1\_val\_1.fq.gz -2 ./6\_1b\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./6\_2\_12S\_1\_val\_1.fq.gz -2 ./6\_2\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./6\_3\_12S\_1\_val\_1.fq.gz -2 ./6\_3\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./H2O-89A\_\_12S\_extracted\_1\_val\_1.fq.gz -2 ./H2O-89A\_\_12S\_extracted\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./H2O-89B\_12S\_extracted\_1\_val\_1.fq.gz -2 ./H2O-89B\_12S\_extracted\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./neg\_ctrl\_1\_12S\_1\_val\_1.fq.gz -2 ./neg\_ctrl\_1\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./neg\_ctrl\_2\_12S\_1\_val\_1.fq.gz -2 ./neg\_ctrl\_2\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

clone\_filter -1 ./neg\_ctrl\_3\_12S\_1\_val\_1.fq.gz -2 ./neg\_ctrl\_3\_12S\_2\_val\_2.fq.gz -i gzfastq -o ./filtered/ --inline\_inline --oligo\_len\_1 28 --oligo\_len\_2 34

## **merge read pairs with usearch**

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_1\_12S\_1\_val\_1.1.fq -reverse 1\_1\_12S\_2\_val\_2.2.fq -fastaout 1\_1\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_1\_COI\_1\_val\_1.1.fq -reverse 1\_1\_COI\_2\_val\_2.2.fq -fastaout 1\_1\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_2\_12S\_1\_val\_1.1.fq -reverse 1\_2\_12S\_2\_val\_2.2.fq -fastaout 1\_2\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_2\_COI\_1\_val\_1.1.fq -reverse 1\_2\_COI\_2\_val\_2.2.fq -fastaout 1\_2\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_3\_12S\_1\_val\_1.1.fq -reverse 1\_3\_12S\_2\_val\_2.2.fq -fastaout 1\_3\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_3\_COI\_1\_val\_1.1.fq -reverse 1\_3\_COI\_2\_val\_2.2.fq -fastaout 1\_3\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_4\_12S\_1\_val\_1.1.fq -reverse 1\_4\_12S\_2\_val\_2.2.fq -fastaout 1\_4\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_4\_COI\_1\_val\_1.1.fq -reverse 1\_4\_COI\_2\_val\_2.2.fq -fastaout 1\_4\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_5\_12S\_1\_val\_1.1.fq -reverse 1\_5\_12S\_2\_val\_2.2.fq -fastaout 1\_5\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_5\_COI\_1\_val\_1.1.fq -reverse 1\_5\_COI\_2\_val\_2.2.fq -fastaout 1\_5\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_6\_12S\_1\_val\_1.1.fq -reverse 1\_6\_12S\_2\_val\_2.2.fq -fastaout 1\_6\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 1\_6\_COI\_1\_val\_1.1.fq -reverse 1\_6\_COI\_2\_val\_2.2.fq -fastaout 1\_6\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_1a\_12S\_1\_val\_1.1.fq -reverse 2\_1a\_12S\_2\_val\_2.2.fq -fastaout 2\_1a\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_1a\_COI\_1\_val\_1.1.fq -reverse 2\_1a\_COI\_2\_val\_2.2.fq -fastaout 2\_1a\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_1b\_12S\_1\_val\_1.1.fq -reverse 2\_1b\_12S\_2\_val\_2.2.fq -fastaout 2\_1b\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_1b\_COI\_1\_val\_1.1.fq -reverse 2\_1b\_COI\_2\_val\_2.2.fq -fastaout 2\_1b\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_2\_12S\_1\_val\_1.1.fq -reverse 2\_2\_12S\_2\_val\_2.2.fq -fastaout 2\_2\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_2\_COI\_1\_val\_1.1.fq -reverse 2\_2\_COI\_2\_val\_2.2.fq -fastaout 2\_2\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_3\_12S\_1\_val\_1.1.fq -reverse 2\_3\_12S\_2\_val\_2.2.fq -fastaout 2\_3\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 2\_3\_COI\_1\_val\_1.1.fq -reverse 2\_3\_COI\_2\_val\_2.2.fq -fastaout 2\_3\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_1\_12S\_1\_val\_1.1.fq -reverse 3\_1\_12S\_2\_val\_2.2.fq -fastaout 3\_1\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_1\_COI\_1\_val\_1.1.fq -reverse 3\_1\_COI\_2\_val\_2.2.fq -fastaout 3\_1\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_2a\_12S\_1\_val\_1.1.fq -reverse 3\_2a\_12S\_2\_val\_2.2.fq -fastaout 3\_2a\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_2a\_COI\_1\_val\_1.1.fq -reverse 3\_2a\_COI\_2\_val\_2.2.fq -fastaout 3\_2a\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_2b\_12S\_1\_val\_1.1.fq -reverse 3\_2b\_12S\_2\_val\_2.2.fq -fastaout 3\_2b\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_2b\_COI\_1\_val\_1.1.fq -reverse 3\_2b\_COI\_2\_val\_2.2.fq -fastaout 3\_2b\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_3\_12S\_1\_val\_1.1.fq -reverse 3\_3\_12S\_2\_val\_2.2.fq -fastaout 3\_3\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 3\_3\_COI\_1\_val\_1.1.fq -reverse 3\_3\_COI\_2\_val\_2.2.fq -fastaout 3\_3\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_1\_12S\_1\_val\_1.1.fq -reverse 4\_1\_12S\_2\_val\_2.2.fq -fastaout 4\_1\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_1\_COI\_1\_val\_1.1.fq -reverse 4\_1\_COI\_2\_val\_2.2.fq -fastaout 4\_1\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_2\_12S\_1\_val\_1.1.fq -reverse 4\_2\_12S\_2\_val\_2.2.fq -fastaout 4\_2\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_2\_COI\_1\_val\_1.1.fq -reverse 4\_2\_COI\_2\_val\_2.2.fq -fastaout 4\_2\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_3\_12S\_1\_val\_1.1.fq -reverse 4\_3\_12S\_2\_val\_2.2.fq -fastaout 4\_3\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_3\_COI\_1\_val\_1.1.fq -reverse 4\_3\_COI\_2\_val\_2.2.fq -fastaout 4\_3\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_4\_12S\_1\_val\_1.1.fq -reverse 4\_4\_12S\_2\_val\_2.2.fq -fastaout 4\_4\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_4\_COI\_1\_val\_1.1.fq -reverse 4\_4\_COI\_2\_val\_2.2.fq -fastaout 4\_4\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_5\_12S\_1\_val\_1.1.fq -reverse 4\_5\_12S\_2\_val\_2.2.fq -fastaout 4\_5\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_5\_COI\_1\_val\_1.1.fq -reverse 4\_5\_COI\_2\_val\_2.2.fq -fastaout 4\_5\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_6\_12S\_1\_val\_1.1.fq -reverse 4\_6\_12S\_2\_val\_2.2.fq -fastaout 4\_6\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 4\_6\_COI\_1\_val\_1.1.fq -reverse 4\_6\_COI\_2\_val\_2.2.fq -fastaout 4\_6\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 5\_1\_12S\_1\_val\_1.1.fq -reverse 5\_1\_12S\_2\_val\_2.2.fq -fastaout 5\_1\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 5\_1\_COI\_1\_val\_1.1.fq -reverse 5\_1\_COI\_2\_val\_2.2.fq -fastaout 5\_1\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 5\_2\_12S\_1\_val\_1.1.fq -reverse 5\_2\_12S\_2\_val\_2.2.fq -fastaout 5\_2\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 5\_2\_COI\_1\_val\_1.1.fq -reverse 5\_2\_COI\_2\_val\_2.2.fq -fastaout 5\_2\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 5\_3\_12S\_1\_val\_1.1.fq -reverse 5\_3\_12S\_2\_val\_2.2.fq -fastaout 5\_3\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 5\_3\_COI\_1\_val\_1.1.fq -reverse 5\_3\_COI\_2\_val\_2.2.fq -fastaout 5\_3\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_1a\_12S\_1\_val\_1.1.fq -reverse 6\_1a\_12S\_2\_val\_2.2.fq -fastaout 6\_1a\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_1a\_COI\_1\_val\_1.1.fq -reverse 6\_1a\_COI\_2\_val\_2.2.fq -fastaout 6\_1a\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_1b\_12S\_1\_val\_1.1.fq -reverse 6\_1b\_12S\_2\_val\_2.2.fq -fastaout 6\_1b\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_1b\_COI\_extracted\_1\_val\_1.1.fq -reverse 6\_1b\_COI\_extracted\_2\_val\_2.2.fq -fastaout 6\_1b\_COI\_extracted\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_2\_12S\_1\_val\_1.1.fq -reverse 6\_2\_12S\_2\_val\_2.2.fq -fastaout 6\_2\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_2\_COI\_extracted\_1\_val\_1.1.fq -reverse 6\_2\_COI\_extracted\_2\_val\_2.2.fq -fastaout 6\_2\_COI\_extracted\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_3\_12S\_1\_val\_1.1.fq -reverse 6\_3\_12S\_2\_val\_2.2.fq -fastaout 6\_3\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs 6\_3\_COI\_1\_val\_1.1.fq -reverse 6\_3\_COI\_2\_val\_2.2.fq -fastaout 6\_3\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs H2O-89A\_\_12S\_extracted\_1\_val\_1.1.fq -reverse H2O-89A\_\_12S\_extracted\_2\_val\_2.2.fq -fastaout H2O-89A\_\_12S\_extracted\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs H2O-89B\_12S\_extracted\_1\_val\_1.1.fq -reverse H2O-89B\_12S\_extracted\_2\_val\_2.2.fq -fastaout H2O-89B\_12S\_extracted\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs H2O-90A\_\_COI\_1\_val\_1.1.fq -reverse H2O-90A\_\_COI\_2\_val\_2.2.fq -fastaout H2O-90A\_\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs H2O90-B\_\_COI\_1\_val\_1.1.fq -reverse H2O90-B\_\_COI\_2\_val\_2.2.fq -fastaout H2O90-B\_\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs nega\_ctrl\_1\_COI\_1\_val\_1.1.fq -reverse nega\_ctrl\_1\_COI\_2\_val\_2.2.fq -fastaout nega\_ctrl\_1\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs neg\_ctrl\_1\_12S\_1\_val\_1.1.fq -reverse neg\_ctrl\_1\_12S\_2\_val\_2.2.fq -fastaout neg\_ctrl\_1\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs neg\_ctrl\_2\_12S\_1\_val\_1.1.fq -reverse neg\_ctrl\_2\_12S\_2\_val\_2.2.fq -fastaout neg\_ctrl\_2\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs neg\_ctrl\_3\_12S\_1\_val\_1.1.fq -reverse neg\_ctrl\_3\_12S\_2\_val\_2.2.fq -fastaout neg\_ctrl\_3\_12S\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs neg\_ctrl\_3\_COI\_1\_val\_1.1.fq -reverse neg\_ctrl\_3\_COI\_2\_val\_2.2.fq -fastaout neg\_ctrl\_3\_COI\_assembled.fq

/home/f6nv175/usearch11.0.667 -fastq\_mergepairs neg\_ctrll\_2\_COI\_1\_val\_1.1.fq -reverse neg\_ctrll\_2\_COI\_2\_val\_2.2.fq -fastaout neg\_ctrll\_2\_COI\_assembled.fq

## **4. dereplicate with vsearch (keeping only one copy of each sequence, and keeping the count!)**

vsearch -derep\_fulllength 1\_1\_12S\_assembled.fq -output 1\_1\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_1\_COI\_assembled.fq -output 1\_1\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_2\_12S\_assembled.fq -output 1\_2\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_2\_COI\_assembled.fq -output 1\_2\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_3\_12S\_assembled.fq -output 1\_3\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_3\_COI\_assembled.fq -output 1\_3\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_4\_12S\_assembled.fq -output 1\_4\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_4\_COI\_assembled.fq -output 1\_4\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_5\_12S\_assembled.fq -output 1\_5\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_5\_COI\_assembled.fq -output 1\_5\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_6\_12S\_assembled.fq -output 1\_6\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 1\_6\_COI\_assembled.fq -output 1\_6\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_1a\_12S\_assembled.fq -output 2\_1a\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_1a\_COI\_assembled.fq -output 2\_1a\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_1b\_12S\_assembled.fq -output 2\_1b\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_1b\_COI\_assembled.fq -output 2\_1b\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_2\_12S\_assembled.fq -output 2\_2\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_2\_COI\_assembled.fq -output 2\_2\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_3\_12S\_assembled.fq -output 2\_3\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 2\_3\_COI\_assembled.fq -output 2\_3\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_1\_12S\_assembled.fq -output 3\_1\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_1\_COI\_assembled.fq -output 3\_1\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_2a\_12S\_assembled.fq -output 3\_2a\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_2a\_COI\_assembled.fq -output 3\_2a\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_2b\_12S\_assembled.fq -output 3\_2b\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_2b\_COI\_assembled.fq -output 3\_2b\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_3\_12S\_assembled.fq -output 3\_3\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 3\_3\_COI\_assembled.fq -output 3\_3\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_1\_12S\_assembled.fq -output 4\_1\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_1\_COI\_assembled.fq -output 4\_1\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_2\_12S\_assembled.fq -output 4\_2\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_2\_COI\_assembled.fq -output 4\_2\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_3\_12S\_assembled.fq -output 4\_3\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_3\_COI\_assembled.fq -output 4\_3\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_4\_12S\_assembled.fq -output 4\_4\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_4\_COI\_assembled.fq -output 4\_4\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_5\_12S\_assembled.fq -output 4\_5\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_5\_COI\_assembled.fq -output 4\_5\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_6\_12S\_assembled.fq -output 4\_6\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 4\_6\_COI\_assembled.fq -output 4\_6\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 5\_1\_12S\_assembled.fq -output 5\_1\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 5\_1\_COI\_assembled.fq -output 5\_1\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 5\_2\_12S\_assembled.fq -output 5\_2\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 5\_2\_COI\_assembled.fq -output 5\_2\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 5\_3\_12S\_assembled.fq -output 5\_3\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 5\_3\_COI\_assembled.fq -output 5\_3\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_1a\_12S\_assembled.fq -output 6\_1a\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_1a\_COI\_assembled.fq -output 6\_1a\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_1b\_12S\_assembled.fq -output 6\_1b\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_1b\_COI\_extracted\_assembled.fq -output 6\_1b\_COI\_extracted\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_2\_12S\_assembled.fq -output 6\_2\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_2\_COI\_extracted\_assembled.fq -output 6\_2\_COI\_extracted\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_3\_12S\_assembled.fq -output 6\_3\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength 6\_3\_COI\_assembled.fq -output 6\_3\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength H2O-89A\_\_12S\_extracted\_assembled.fq -output H2O-89A\_\_12S\_extracted\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength H2O-89B\_12S\_extracted\_assembled.fq -output H2O-89B\_12S\_extracted\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength H2O-90A\_\_COI\_assembled.fq -output H2O-90A\_\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength H2O90-B\_\_COI\_assembled.fq -output H2O90-B\_\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength nega\_ctrl\_1\_COI\_assembled.fq -output nega\_ctrl\_1\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength neg\_ctrl\_1\_12S\_assembled.fq -output neg\_ctrl\_1\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength neg\_ctrl\_2\_12S\_assembled.fq -output neg\_ctrl\_2\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength neg\_ctrl\_3\_12S\_assembled.fq -output neg\_ctrl\_3\_12S\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength neg\_ctrl\_3\_COI\_assembled.fq -output neg\_ctrl\_3\_COI\_assembled\_count.fq -sizeout

vsearch -derep\_fulllength neg\_ctrll\_2\_COI\_assembled.fq -output neg\_ctrll\_2\_COI\_assembled\_count.fq -sizeout

## **5. create “master” for OTU creation (adding to pre-existing count file)**

concatenate all derepliacated/counted files of each gene

* concatenating the two newly assembled files with all previously assembled files

cat \*COI\_\*.fq > COI\_all\_OTUs\_2020-06-09.txt

dereplicate the pooled files with sizeout + sizein to keep read abundances! Use minuniquesize 5 to remove singletons and potential contaminations!

vsearch -derep\_fulllength COI\_all\_sequences\_2020-06-09.txt -output COI\_all\_sequences\_count\_derep\_min5\_2020-06-09.fq -sizein -sizeout -minuniquesize 5

## **6) cluster OTUs mit usearch**

Cluster with 3% similarity on both strands!

/home/f6nv175/usearch11.0.667 -cluster\_otus COI\_all\_sequences\_count\_derep\_min5\_2020-06-09.fq -otus COI\_all\_OTUs\_2020-06-09.txt -uparseout COI\_all\_OTU\_out\_table\_2020-06-09.txt -relabel OTU\_ -strand both

compare dereplicated reads of each replica against the OTUs without chimeras (repeated for 12S with 99% clustering threshold

/home/f6nv175/usearch11.0.667 -usearch\_global 1\_1\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 1\_1\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 1\_2\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 1\_2\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 1\_3\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 1\_3\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 1\_4\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 1\_4\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 1\_5\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 1\_5\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 1\_6\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 1\_6\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 2\_1a\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 2\_1a\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 2\_1b\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 2\_1b\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 2\_2\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 2\_2\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 2\_3\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 2\_3\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 3\_1\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 3\_1\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 3\_2a\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 3\_2a\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 3\_2b\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 3\_2b\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 3\_3\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 3\_3\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 4\_1\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 4\_1\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 4\_2\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 4\_2\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 4\_3\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 4\_3\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 4\_4\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 4\_4\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 4\_5\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 4\_5\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 4\_6\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 4\_6\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 5\_1\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 5\_1\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 5\_2\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 5\_2\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 5\_3\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 5\_3\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 6\_1a\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 6\_1a\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 6\_1b\_2020-06-08\_COI\_extracted\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 6\_1b\_2020-06-09\_COI\_extracted\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 6\_2\_2020-06-08\_COI\_extracted\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 6\_2\_2020-06-09\_COI\_extracted\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global 6\_3\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out 6\_3\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global H2O-90A\_\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out H2O-90A\_\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global H2O90-B\_\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out H2O90-B\_\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global nega\_ctrl\_1\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out nega\_ctrl\_1\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global neg\_ctrl\_3\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out neg\_ctrl\_3\_COI\_OTUs.txt

/home/f6nv175/usearch11.0.667 -usearch\_global neg\_ctrll\_2\_COI\_assembled\_count.fq -db COI\_all\_OTUs\_2020-06-09.txt -strand both -id 0.97 -maxhits 1 -blast6out neg\_ctrll\_2\_COI\_OTUs.txt

## **7) BLAST**

blastn -db /n/holyscratch01/external\_repos/INFORMATICS/ref/ncbi/nt/20200203/nt -num\_threads 16 -max\_target\_seqs 5 -outfmt '6 qseqid sseqid evalue pident sacc stitle' -query COI\_all\_OTUs\_2020-06-09.txt -out COI\_all\_OTUs\_BLAST\_nt\_2020-06-10

## **8) Add taxonomy to Blast hits**

#remove BLAST hists with non standard NCBI numbers

awk 'length($5)<10' COI\_all\_OTUs\_BLAST\_nt\_2020-06-10 > COI\_all\_OTUs\_BLAST\_nt\_out\_no\_NCBI\_2020-06-15

#COI BLAST Table includes 5 hits per OTU, to remove duplicates and keep only first hit per OTU:

awk '!seen[$1]++' COI\_all\_OTUs\_BLAST\_nt\_out\_no\_NCBI\_2020-06-15 > COI\_all\_OTUs\_BLAST\_nt\_out\_noDup\_2020-06-15

# **reduce nucl\_gb.accession2taxid to include only matching entries**

awk 'NR==FNR{inFileA[$5]; next} ($1 in inFileA)' COI\_all\_OTUs\_BLAST\_nt\_out\_noDup\_2020-06-15 ../nucl\_gb.accession2taxid\_2020-06-25 > COI\_nt\_nucl\_gb.accession2taxid\_2020-07-06

#to add TaxID of NCBI to OTU table:

awk -v OFS='\t' 'NR==FNR{a[$1]=$3; next} ($5 in a) {print $0, a[$5]}' COI\_nt\_nucl\_gb.accession2taxid\_2020-07-06 COI\_all\_OTUs\_BLAST\_nt\_out\_noDup\_2020-06-15 > COI\_all\_OTUs\_BLAST\_nt\_out\_TaxID\_2020-07-06

## **9) Retrieve taxonomy and species information**

curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=taxonomy&amp;id=127563&amp;retmode=xml&amp;rettype=full" | xmllint --xpath "/TaxaSet/Taxon/Lineage/text()" - >> OTU\_1\_taxonomy\_2020-06-16

* the number highlighted in red is taxid; this gives the full taxonomy for this tax id! (to the next HIGHEST RANK; e.g., if species taxid is given it proves all to genus level

curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=taxonomy&amp;id=127563&amp;retmode=xml&amp;rettype=full" | xmllint --xpath "/TaxaSet/Taxon/ScientificName/text()" - >> OTU\_1\_species\_name\_2020-06-16

* the number highlighted in red is taxid; this gives the species name
* this creates two separate file for each OTU, one containing the taxonomy and the species name
* these need to concatenated into single files:

sed -e '$s/$//' -s OTU\*name\_2020-\* > COI\_nt\_all\_OTUs\_species\_names\_2020-07-06

sed -e '$s/$//' -s OTU\*taxonomy\_2020-\* > COI\_nt\_all\_OTUs\_full\_taxonomy\_2020-07-06

sort COI\_all\_OTUs\_BLAST\_nt\_out\_TaxID\_2020-07-06 -k1 > COI\_all\_OTUs\_BLAST\_nt\_out\_TaxID\_sorted\_2020-07-06

* COI\_all\_OTUs\_BLAST\_nt\_out\_TaxID\_sorted\_2020-07-06 COI\_nt\_all\_OTUs\_full\_taxonomy\_2020-07-06 and COI\_nt\_all\_OTUs\_species\_names\_2020-07-06 were merged in Excel
* final file with all information: COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06

## **10) sum OTU frequency in each sample**

To add taxon information to the sample specific OTU files, unique OTUs need to be collapsed and their frequency summed up; done in two steps, first using “cut” to remove remnants of Illumina headers; secondly remove duplicates and add the sum using awk; in the final file there will be two columns 1. OTU name, 2. its frequency in the respective sample

cut -sd'=' -f2 1\_1\_COI\_OTUs.txt > 1\_1\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 1\_2\_COI\_OTUs.txt > 1\_2\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 1\_3\_COI\_OTUs.txt > 1\_3\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 1\_4\_COI\_OTUs.txt > 1\_4\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 1\_5\_COI\_OTUs.txt > 1\_5\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 1\_6\_COI\_OTUs.txt > 1\_6\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 2\_1a\_COI\_OTUs.txt > 2\_1a\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 2\_1b\_COI\_OTUs.txt > 2\_1b\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 2\_2\_COI\_OTUs.txt > 2\_2\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 2\_3\_COI\_OTUs.txt > 2\_3\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 3\_1\_COI\_OTUs.txt > 3\_1\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 3\_2a\_COI\_OTUs.txt > 3\_2a\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 3\_2b\_COI\_OTUs.txt > 3\_2b\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 3\_3\_COI\_OTUs.txt > 3\_3\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 4\_1\_COI\_OTUs.txt > 4\_1\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 4\_2\_COI\_OTUs.txt > 4\_2\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 4\_3\_COI\_OTUs.txt > 4\_3\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 4\_4\_COI\_OTUs.txt > 4\_4\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 4\_5\_COI\_OTUs.txt > 4\_5\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 4\_6\_COI\_OTUs.txt > 4\_6\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 5\_1\_COI\_OTUs.txt > 5\_1\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 5\_2\_COI\_OTUs.txt > 5\_2\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 5\_3\_COI\_OTUs.txt > 5\_3\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 6\_1a\_COI\_OTUs.txt > 6\_1a\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 6\_3\_COI\_OTUs.txt > 6\_3\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 H2O-90A\_\_COI\_OTUs.txt > H2O-90A\_\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 H2O90-B\_\_COI\_OTUs.txt > H2O90-B\_\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 nega\_ctrl\_1\_COI\_OTUs.txt > nega\_ctrl\_1\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 neg\_ctrl\_3\_COI\_OTUs.txt > neg\_ctrl\_3\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 neg\_ctrll\_2\_COI\_OTUs.txt > neg\_ctrll\_2\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 6\_1b\_2020-06-09\_COI\_extracted\_OTUs.txt > 6\_1b\_COI\_OTUs\_shortened\_2020-06-16.txt

cut -sd'=' -f2 6\_2\_2020-06-09\_COI\_extracted\_OTUs.txt > 6\_2\_COI\_OTUs\_shortened\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 1\_1\_COI\_OTUs\_shortened\_2020-06-16.txt > 1\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 1\_2\_COI\_OTUs\_shortened\_2020-06-16.txt > 1\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 1\_3\_COI\_OTUs\_shortened\_2020-06-16.txt > 1\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 1\_4\_COI\_OTUs\_shortened\_2020-06-16.txt > 1\_4\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 1\_5\_COI\_OTUs\_shortened\_2020-06-16.txt > 1\_5\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 1\_6\_COI\_OTUs\_shortened\_2020-06-16.txt > 1\_6\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 2\_1a\_COI\_OTUs\_shortened\_2020-06-16.txt > 2\_1a\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 2\_1b\_COI\_OTUs\_shortened\_2020-06-16.txt > 2\_1b\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 2\_2\_COI\_OTUs\_shortened\_2020-06-16.txt > 2\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 2\_3\_COI\_OTUs\_shortened\_2020-06-16.txt > 2\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 3\_1\_COI\_OTUs\_shortened\_2020-06-16.txt > 3\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 3\_2a\_COI\_OTUs\_shortened\_2020-06-16.txt > 3\_2a\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 3\_2b\_COI\_OTUs\_shortened\_2020-06-16.txt > 3\_2b\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 3\_3\_COI\_OTUs\_shortened\_2020-06-16.txt > 3\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 4\_2\_COI\_OTUs\_shortened\_2020-06-16.txt > 4\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 4\_3\_COI\_OTUs\_shortened\_2020-06-16.txt > 4\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 4\_4\_COI\_OTUs\_shortened\_2020-06-16.txt > 4\_4\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 4\_5\_COI\_OTUs\_shortened\_2020-06-16.txt > 4\_5\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 4\_6\_COI\_OTUs\_shortened\_2020-06-16.txt > 4\_6\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 5\_1\_COI\_OTUs\_shortened\_2020-06-16.txt > 5\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 5\_2\_COI\_OTUs\_shortened\_2020-06-16.txt > 5\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 5\_3\_COI\_OTUs\_shortened\_2020-06-16.txt > 5\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 6\_1a\_COI\_OTUs\_shortened\_2020-06-16.txt > 6\_1a\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 6\_3\_COI\_OTUs\_shortened\_2020-06-16.txt > 6\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' H2O-90A\_\_COI\_OTUs\_shortened\_2020-06-16.txt > H2O-90A\_\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' H2O90-B\_\_COI\_OTUs\_shortened\_2020-06-16.txt > H2O90-B\_\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' nega\_ctrl\_1\_COI\_OTUs\_shortened\_2020-06-16.txt > nega\_ctrl\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' neg\_ctrl\_3\_COI\_OTUs\_shortened\_2020-06-16.txt > neg\_ctrl\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' neg\_ctrll\_2\_COI\_OTUs\_shortened\_2020-06-16.txt > neg\_ctrll\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 4\_1\_COI\_OTUs\_shortened\_2020-06-16.txt > 4\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 6\_1b\_COI\_OTUs\_shortened\_2020-06-16.txt > 6\_1b\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

awk -v OFS='\t' '{ seen[$2] += $1 } END { for (i in seen) print i, seen[i] }' 6\_2\_COI\_OTUs\_shortened\_2020-06-16.txt > 6\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt

## **11) add taxonomy information to each sample**

This adds the taxonomy info to each OTU retrieved in each single sample.

**ATTENTION: OTUs without BLAST HIT will be REMOVED from the output**

**the suitable header for the resulting files is:**

**OTU\_number OTU\_count/frequency OTU\_number GB\_number e-value percentage NCBI info TaxID Taxonomy** Species

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 1\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 1\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 1\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 1\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 1\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 1\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 1\_4\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 1\_4\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 1\_5\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 1\_5\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 1\_6\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 1\_6\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 2\_1a\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 2\_1a\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 2\_1b\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 2\_1b\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 2\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 2\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 2\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 2\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 3\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 3\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 3\_2a\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 3\_2a\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 3\_2b\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 3\_2b\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 3\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 3\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 4\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 4\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 4\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 4\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 4\_4\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 4\_4\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 4\_5\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 4\_5\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 4\_6\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 4\_6\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 5\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 5\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 5\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 5\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 5\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 5\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 6\_1a\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 6\_1a\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 6\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 6\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 H2O-90A\_\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > H2O-90A\_\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 H2O90-B\_\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > H2O90-B\_\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 nega\_ctrl\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > nega\_ctrl\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 neg\_ctrl\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > neg\_ctrl\_3\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 neg\_ctrll\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > neg\_ctrll\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 4\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 4\_1\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 6\_1b\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 6\_1b\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

awk -v OFS='\t' 'NR==FNR{a[$1]=$0; next} ($1 in a) {print $0, a[$1]}' COI\_all\_OTUs\_BLAST\_nt\_taxonomy\_2020-07-06 6\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_2020-06-16.txt > 6\_2\_COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt

## **12) create a file that contains all OTUs from all replicates, with original file name added to each line**

awk -v OFS='\t' '{print FILENAME, $0}' \*COI\_OTUs\_shortened\_OTU\_collapsed\_taxonomy\_2020-07-06.txt >> COI\_all\_replicates\_taxonomy\_2020-07-06

## **13) assign and remove species**

done in Excel

* only Metazoa retained
* based on % similarity:

<85% similarity : unkown Metazoa (same for samples classified as Invertebrate environmental sample)

85-97% : higher classification (fish, crustacean etc) retained, but not species

>97% species identification retained

* based on evalue:

>e-20 : unkown Metazoa

e-20 - e-50: higher classification (fish, crustacean etc) retained, but not species

<e-50 species identification retained

(Both threshold applied together; e.g., to retain species names >97% similarity AND evalue < e-50)

* remove all 12S contaminants