Table of contents for BUGSS folder

Raw sequence data from MiSeq

archive-GSF2129-corrected-20190404.tar

Miseq data spreadsheet has values for number of sequences from instrument and those that made it into the merged files below.

Series of files these are merged and trimmed sequences by sample.

Not yet compatible completely with usearch (sample names are NOT part of sequence name)

But what I uploaded as fasta for the arb-NGS site

GSF2129-Aug-1-3\_merged.fastq

GSF2129-Dec-17-1-2\_merged.fastq

GSF2129-Dec17-2-1\_merged.fastq

GSF2129-Dec17-2-2\_merged.fastq

GSF2129-July17-1-3\_merged.fastq

GSF2129-June-2-2\_merged.fastq

GSF2129-May-1-2\_merged.fastq

GSF2129-May-1-3\_merged.fastq

GSF2129-May-2-1\_merged.fastq

GSF2129-Nov-17-1-3\_merged.fastq

GSF2129-Nov17-1-2\_merged.fastq

GSF2129-Nov17-S2-2\_merged.fastq

GSF2129-Oct17-S1-2\_merged.fastq

GSF2129-Oct17-S2-1\_merged.fastq

GSF2129-Oct17-S2-2\_merged.fastq

GSF2129-Oct17-S2-3\_merged.fastq

GSF2129-Sept-1-1\_merged.fastq

GSF2129-Sept-1-2\_merged.fastq

GSF2129-Sept-2-2\_merged.fastq

GSF2129-Sept-2-3\_merged.fastq

--------------------------------------------------------------------------------------------------------------------------

This is for the sample data downloaded from Tara Oceans

Otutab\_out.txt This has a list of otu (really just sequence ids) followed by two columns, one for each of the two samples.

Otus\_sintaxtable This file has a column with the otu name and then the taxonomy. The numbers in parentheses are k-mer based bootstrap numbers.

Bugslec2.pdf What I was going to show you but which I only did part of for the ‘lecture’

BUGS commands week 2 is the ‘how to’ part plus some of the class outline for the second session.

18S\_BUGSS\_commands is the week 1 version of the commands file.

ERR562560\_1.fastq.gz. “Forward” or “1” reads for the sample ERR562560.

ERR562560\_2.fastq.gz Corresponding reverse reads

ERR562383\_1.fastq.gz “Forward” or “1” reads for the sample ERR562383

ERR562383\_2.fastq.gz Corresponding reverse reads

pr2\_version\_4.11.1\_UTAX.fasta This is reference database plus taxonomy.

Oriented\_primers.stripped This is the input file for the arb-NGS website. Basically 100,000 reads taken from each sample, then ‘oriented’ or turned the right way, then looking for primers with a text search, then trimming off the primers.

<http://taraoceans.sb-roscoff.fr/EukDiv/>

Where the first week tiny dataset came from:

Station TARA\_048

ERS489336 an identifier

You can plug this id into like so...

<https://www.ebi.ac.uk/ena/data/search?query=ERS490503>

More metadata TARA\_20101005T1828Z\_072\_EVENT\_NET\_N1\_D\_(100m)\_PROT\_NUC-RNA(1L)\_N5-20\_TARA\_N000000837

ERR1718062 a batch of sequences

ERR1718062.fastq is the file I put in the drive

These are the files for the second week -- the bigger more realistic ones.

http://taraoceans.sb-roscoff.fr/EukDiv/

Search results for ERS490473 DCM 180-2000 microns TARA\_845

More metadata: TARA\_20101005T1406Z\_072\_EVENT\_NET\_N1+N2\_D\_(100 m)\_PROT\_NUC-RNA(1L)\_N180-2000\_TARA\_N000000845

ERX521720 Illumina Genome Analyzer IIx paired end sequencing; Marine... Library Library preparations from V9-18S rDNA amplicons

ERR562383\_1.fastq.gz

ERR562383\_2.fastq.gz

And the other

<https://www.ebi.ac.uk/ena/data/view/ERX521429>

Links you to

ERR562383\_1