What samples should be processed together? What do we expect to contain unique information?

Separate files by size fractions

Need to be able to assess biological information from file name

Diet 50 and 77 expected to be similar

Store as separate directory metadata with raw files

Never combine biological replicates

Directory XL

Directory AVG

Directory TINY

Sequences (PE1) 77A_2-18_XL.R1.fastq.gz

Sequences (PE1) 77A_2-19_XL.R1.fastq.gz

Sequences (PE2) 77A_2-18_XL.R2.fastq.gz.R2.fastq.gz

Sequences (PE2) 77A_2-19_XL.R2.fastq.gz.R2.fastq.gz

Run-associated Metadata 77A_2-18_XL.csv

Run-associated Metadata 77A_2-19_XL.fcsv

Sequences (PE1) 50A_2-18_XL.R1.fastq.gz

Sequences (PE2) 50A_2-18_XL.R2.fastq.gz.R2.fastq.gz

Run-associated Metadata 50A_2-18_XL.csv

Metadata

Lane information

Wetlab metadata

Mapping IDs (e.g., 77A_2-18_XL = DR1)

Analysis: Assemble together

Output files: One assembly; abundance files for each dataset

XL_assembly.fasta (sequence IDs, sequences)

For each sample ID, sequence IDs (linked to assembly) and count for each sequence ID)

Assembled Reference Metagenome

Gene counts for each sample

Annotation of Assembly

Sequence IDs and taxonomy and function for each ID's closest hit to a reference database

Metadata for each sample

Sample IDs and treatment information