

AN INTRODUCTION TO **METABARCODING** DATA ANALYSIS

WORKSHOP OVERVIEW

PART I: METABARCODING BACKGROUND

Considerations to take when designing your study (replicates, mock community, multiplexing, etc.) and general overview of the steps involved in analyzing amplicon data

PART II: INTERACTIVE DATA ANALYSIS

Example data analysis in QIIME2 using Jupyter Notebook in Binder on GitHub

Time left at the end can be used for additional questions

SPEAKER INFO



ERIN BORBEE

Erin is a 4th year PhD candidate in the Lane Lab at URI using metabarcoding to understand drivers of protist biodiversity and biogeography in coastal ecosystems across Indonesia



@ErinBorbee



ELAINE SHEN

Elaine is a 3rd year PhD candidate in the Lane & Humphries Labs at URI studying the drivers of coral reef animal diversity in Indonesia using both visual and metabarcoding approaches

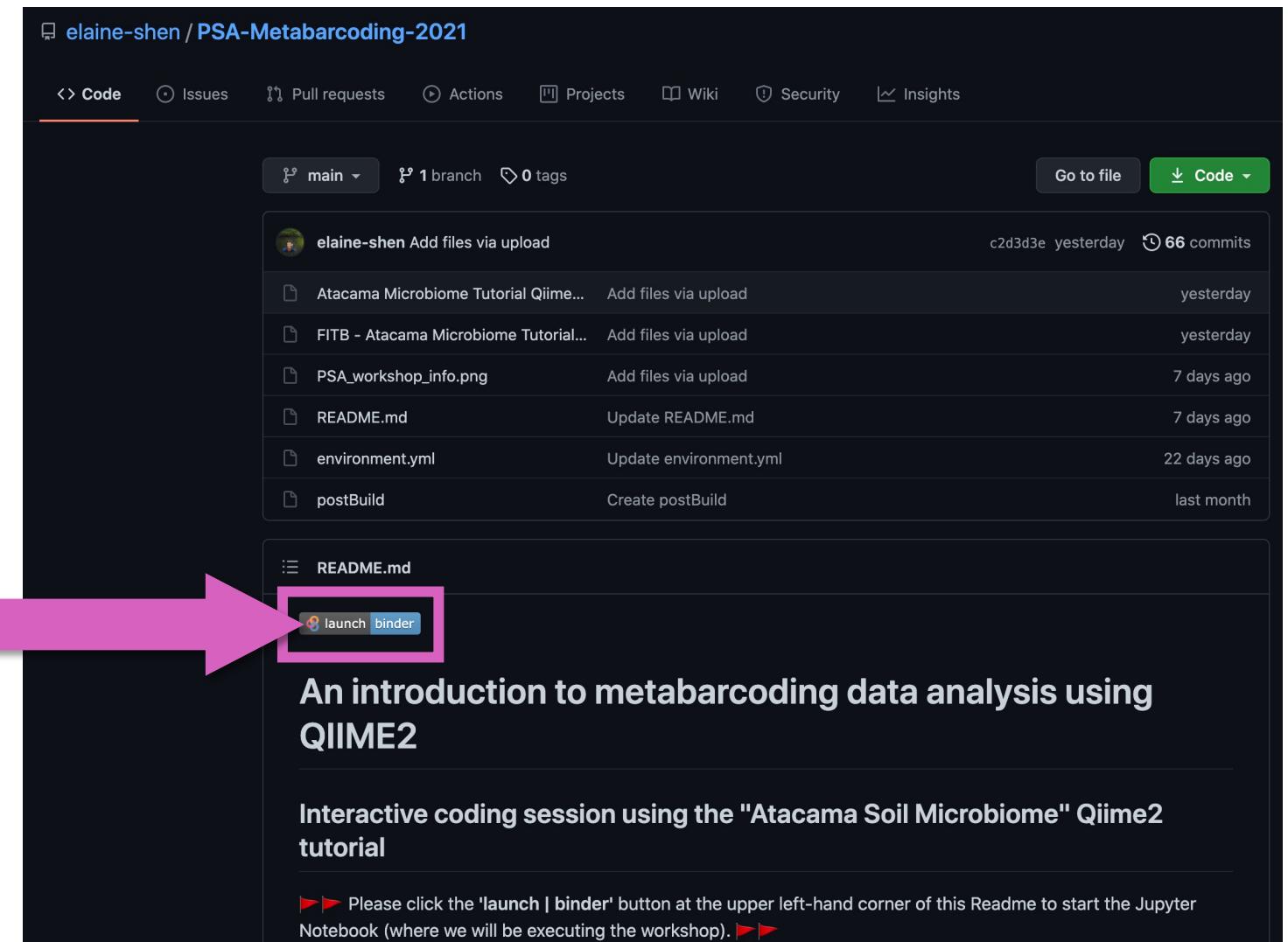


@elaineshen_

BEFORE WE START...

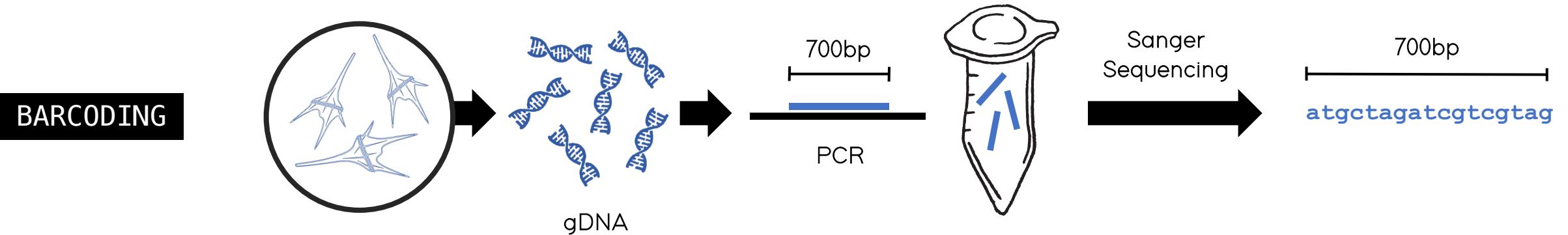
Please launch the binder on GitHub at the following link (also in the chat)

How to launch the binder once in GitHub repository →



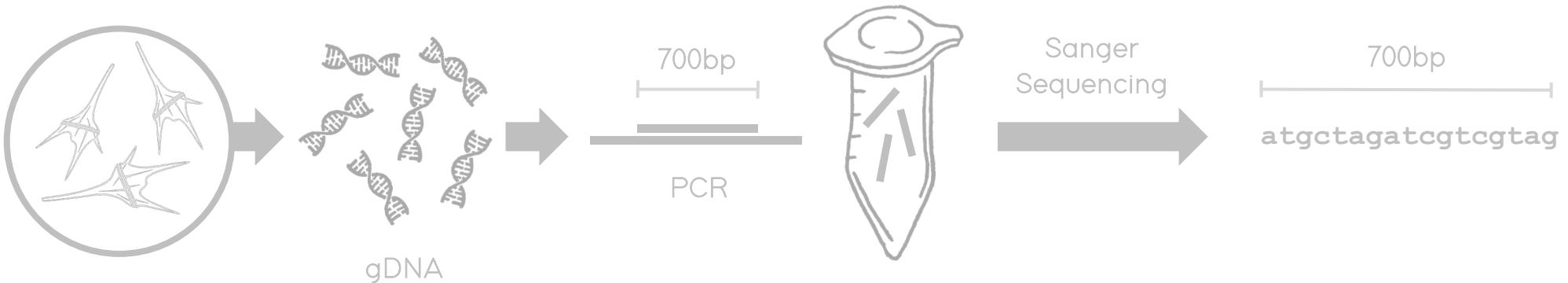
WHAT IS METABARCODING?

WHAT IS METABARCODING?

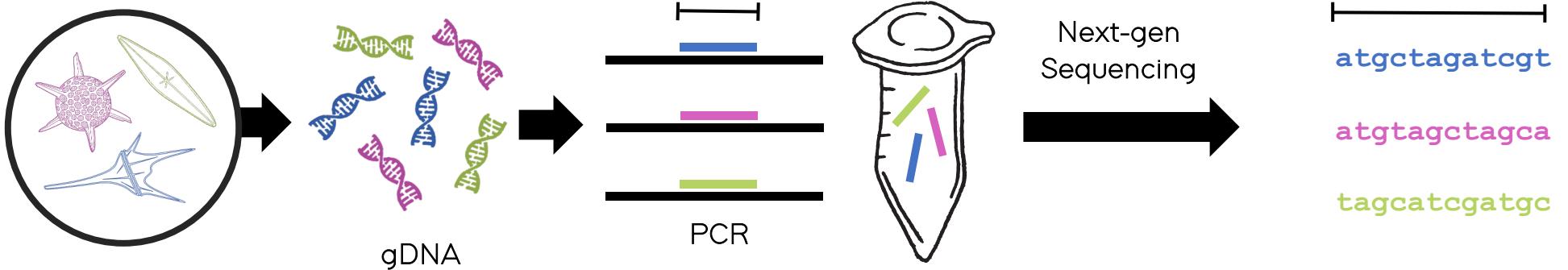


WHAT IS METABARCODING?

BARCODING

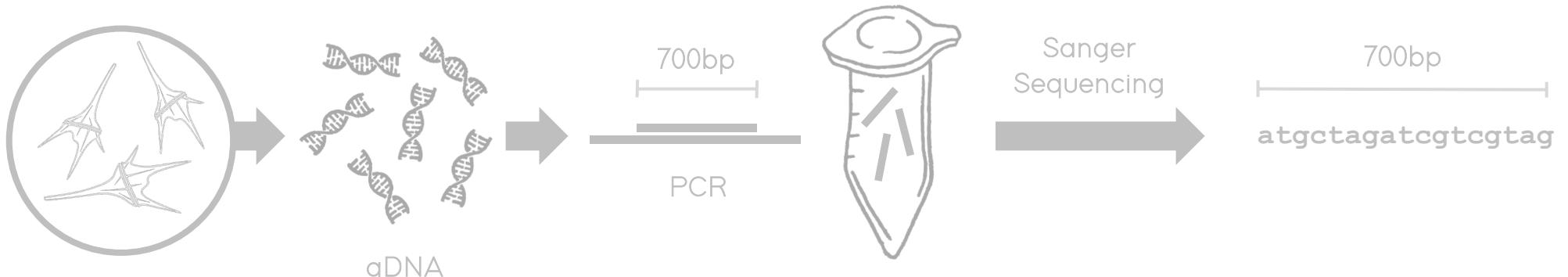


METABARCODING

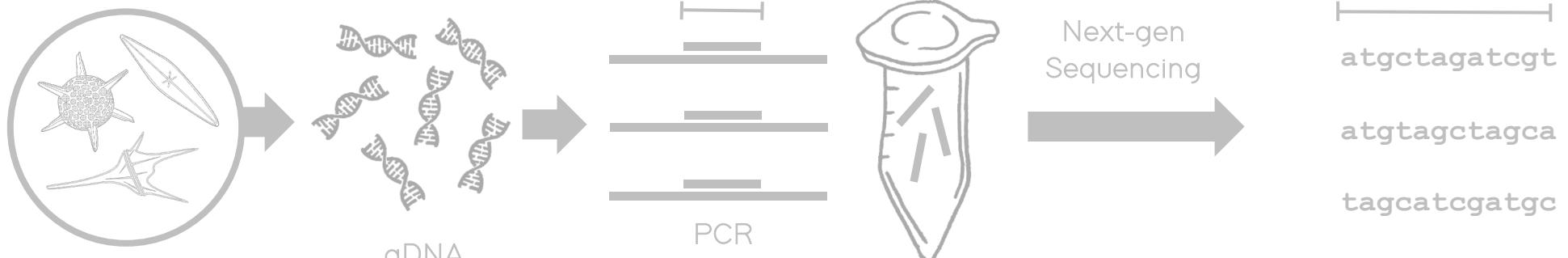


WHAT IS METABARCODING?

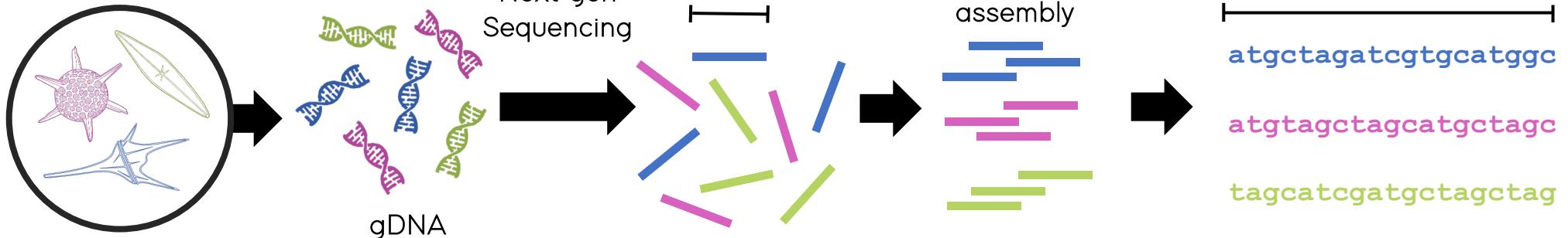
BARCODING



METABARCODING

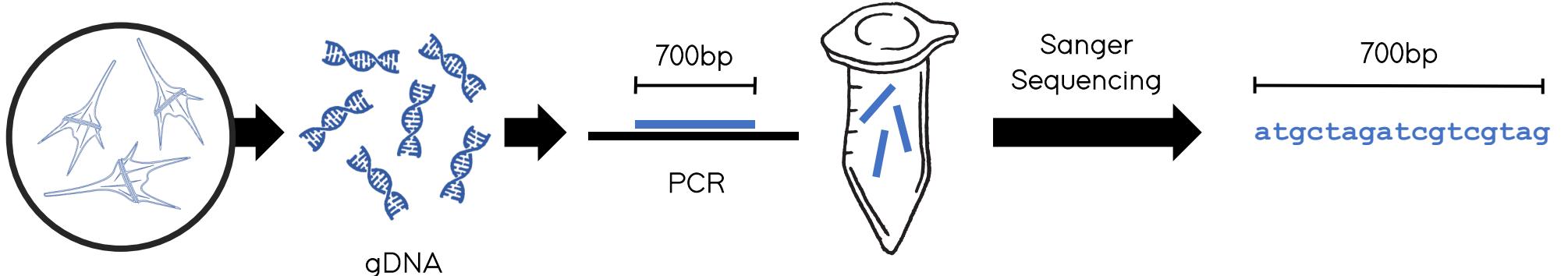


METAGENOMICS

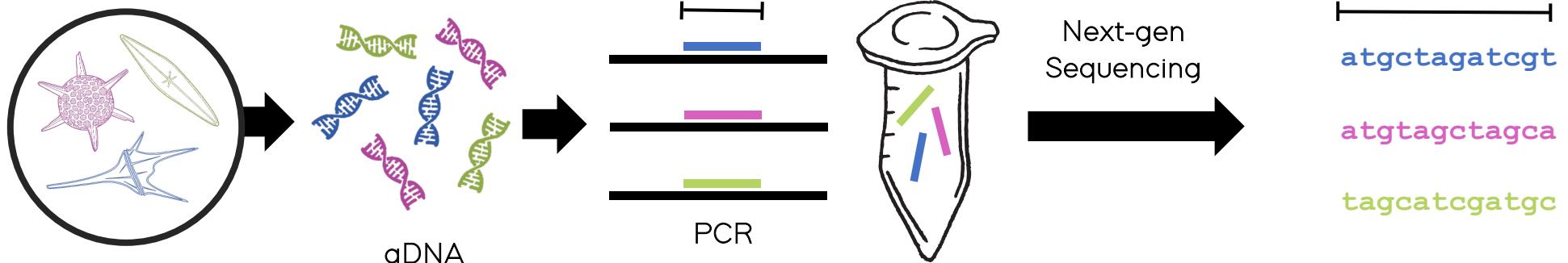


WHAT IS METABARCODING?

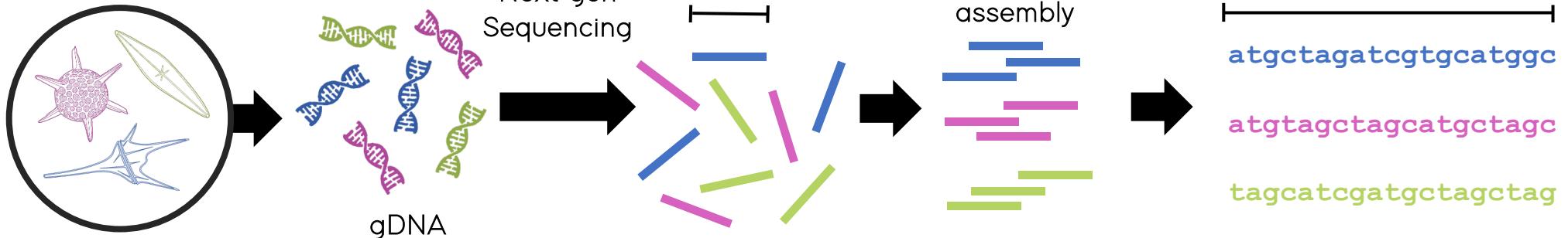
BARCODING



METABARCODING



METAGENOMICS



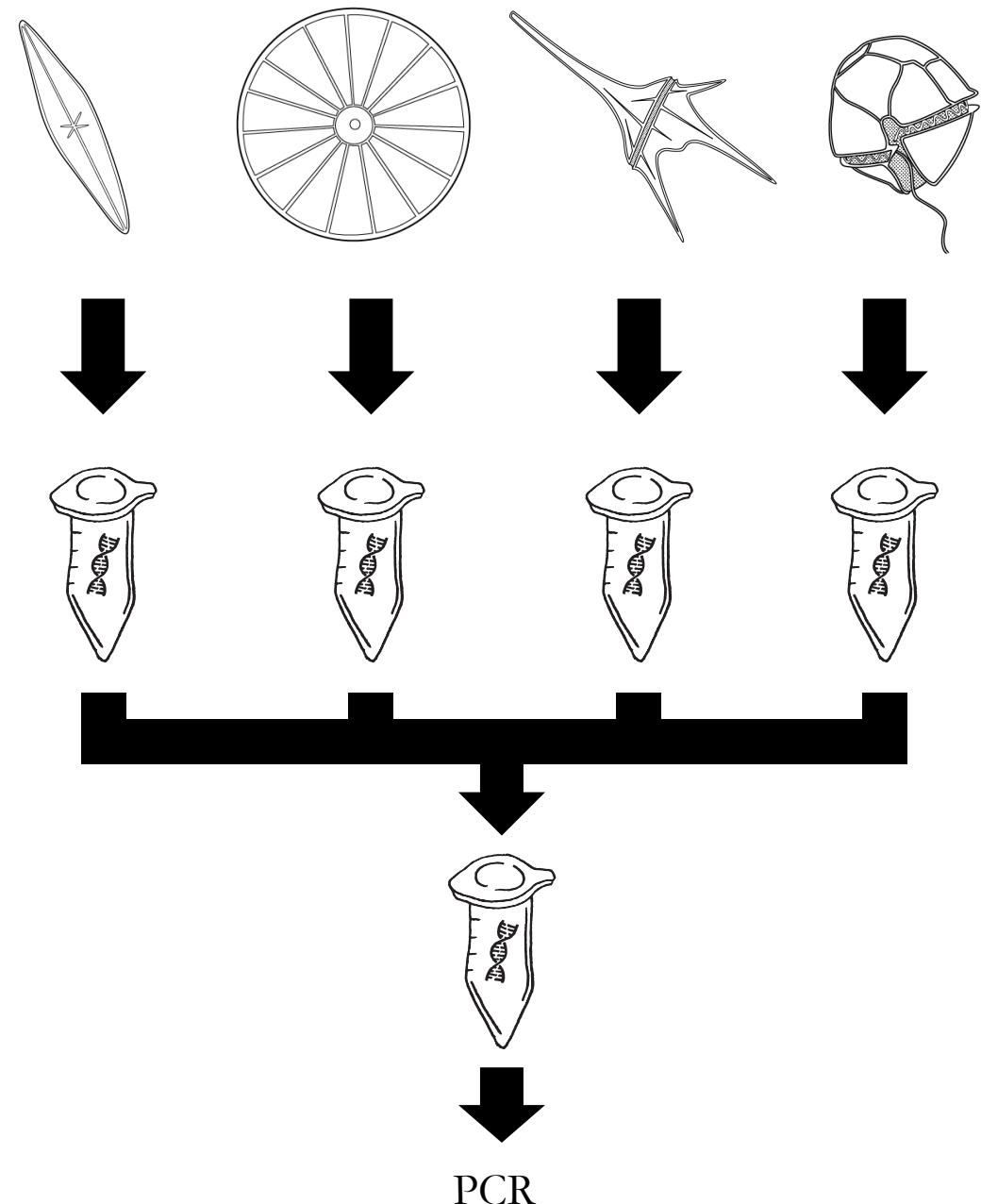
SAMPLING CONSIDERATIONS

MOCK COMMUNITIES

Mock communities are combinations of known concentrations of DNA/cells from various species of interest to simulate a microbiome/community sample

Depending on your target groups you may be able to purchase a mock community, but otherwise it can be easily prepared in the lab

MULTIPLEXING



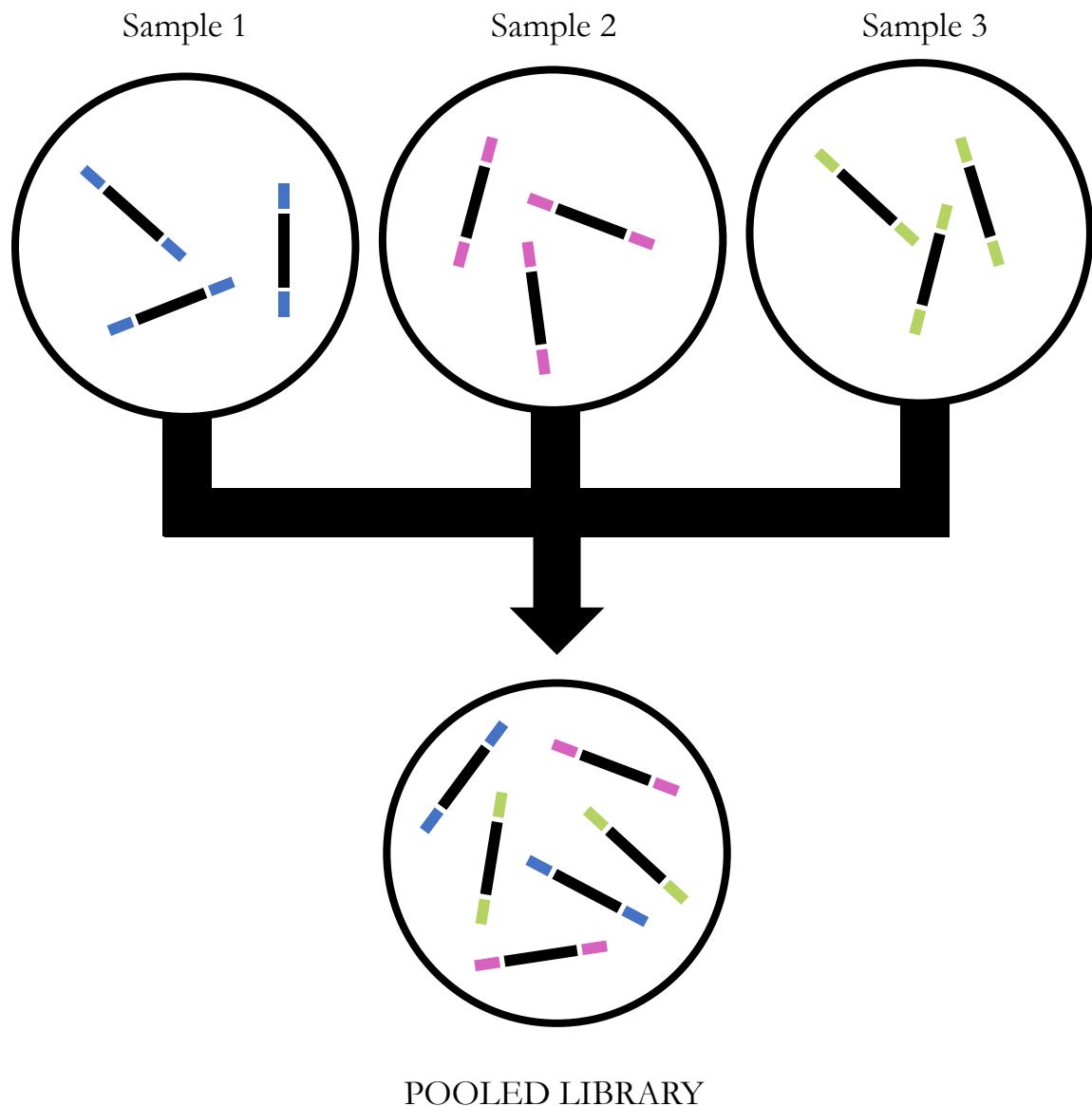
SAMPLING CONSIDERATIONS

MOCK COMMUNITIES

MULTIPLEXING

Multiplexing is the pooling of large numbers of samples to maximize the number of samples sequenced on a single run. The number of samples that can be sequenced on a single run is limited by the number of unique barcode indices available to attach to sequences during library preparation

Another consideration you should take with multiplexing is the diversity of your samples and the coverage you will need per sample to capture the full community



METABARCODING DATA ANALYSIS

- ① QUALITY ASSESSMENT
- ② DEMULTIPLEXING
- ③ PRIMER TRIMMING
- ④ DENOISING & MERGING
- ⑤ TAXONOMIC ASSIGNMENT

METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

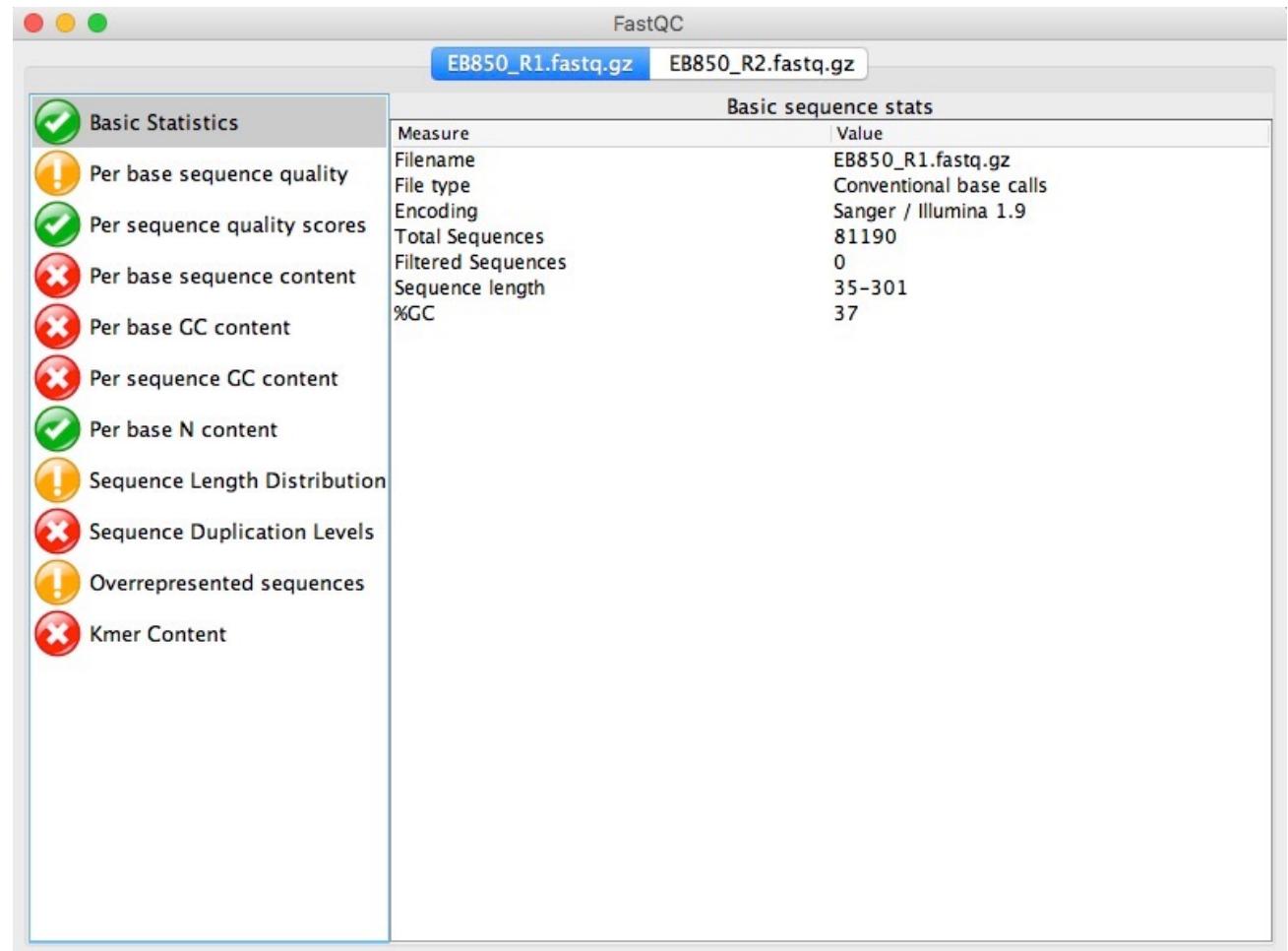
Quality assessment of raw reads and basic sequencing statistics can be evaluated using FastQC or MultiQC, or can also be done in QIIME2

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

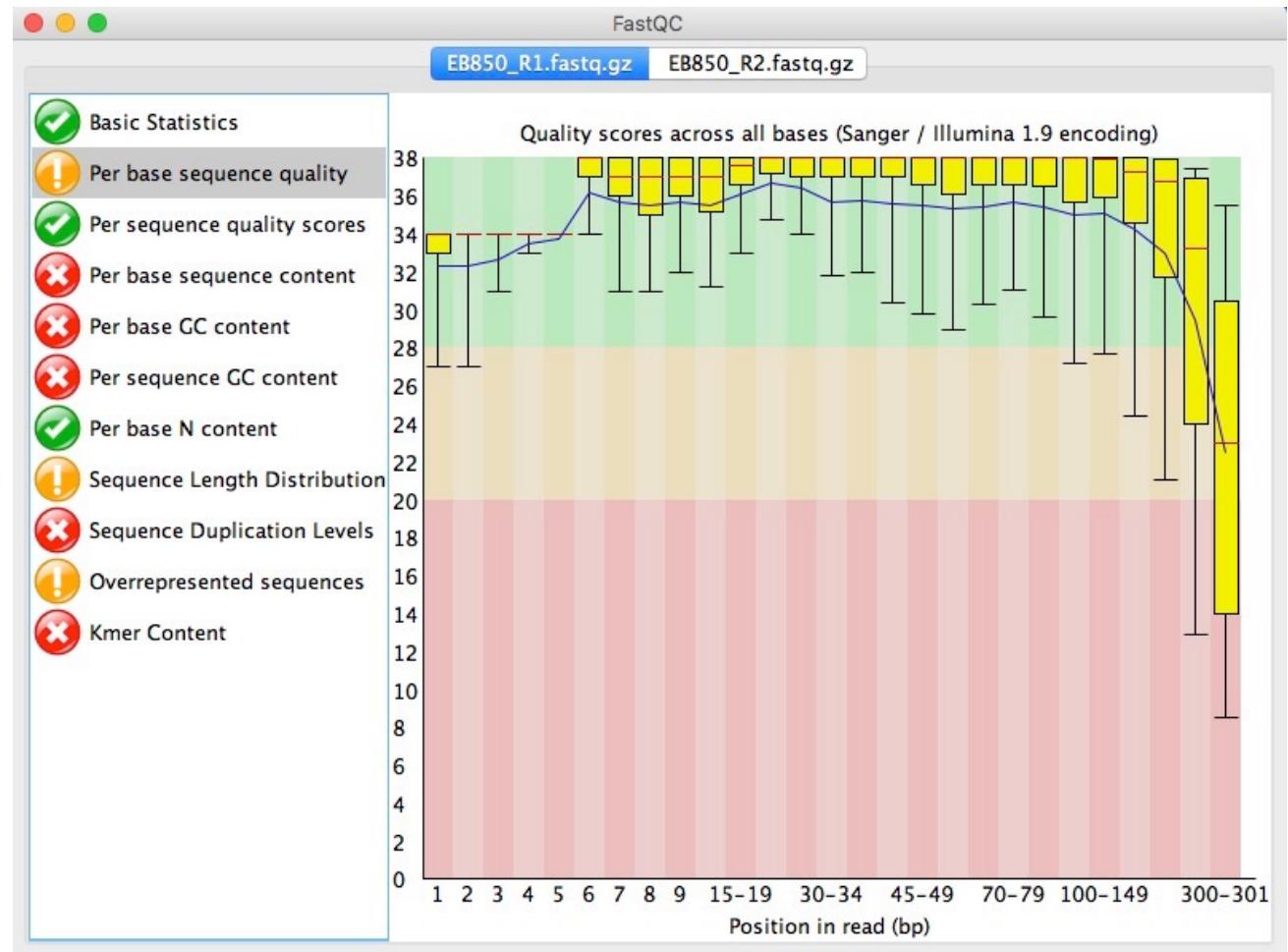
Quality assessment of raw reads and basic sequencing statistics can be evaluated using FastQC or MultiQC, or can also be done in QIIME2

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

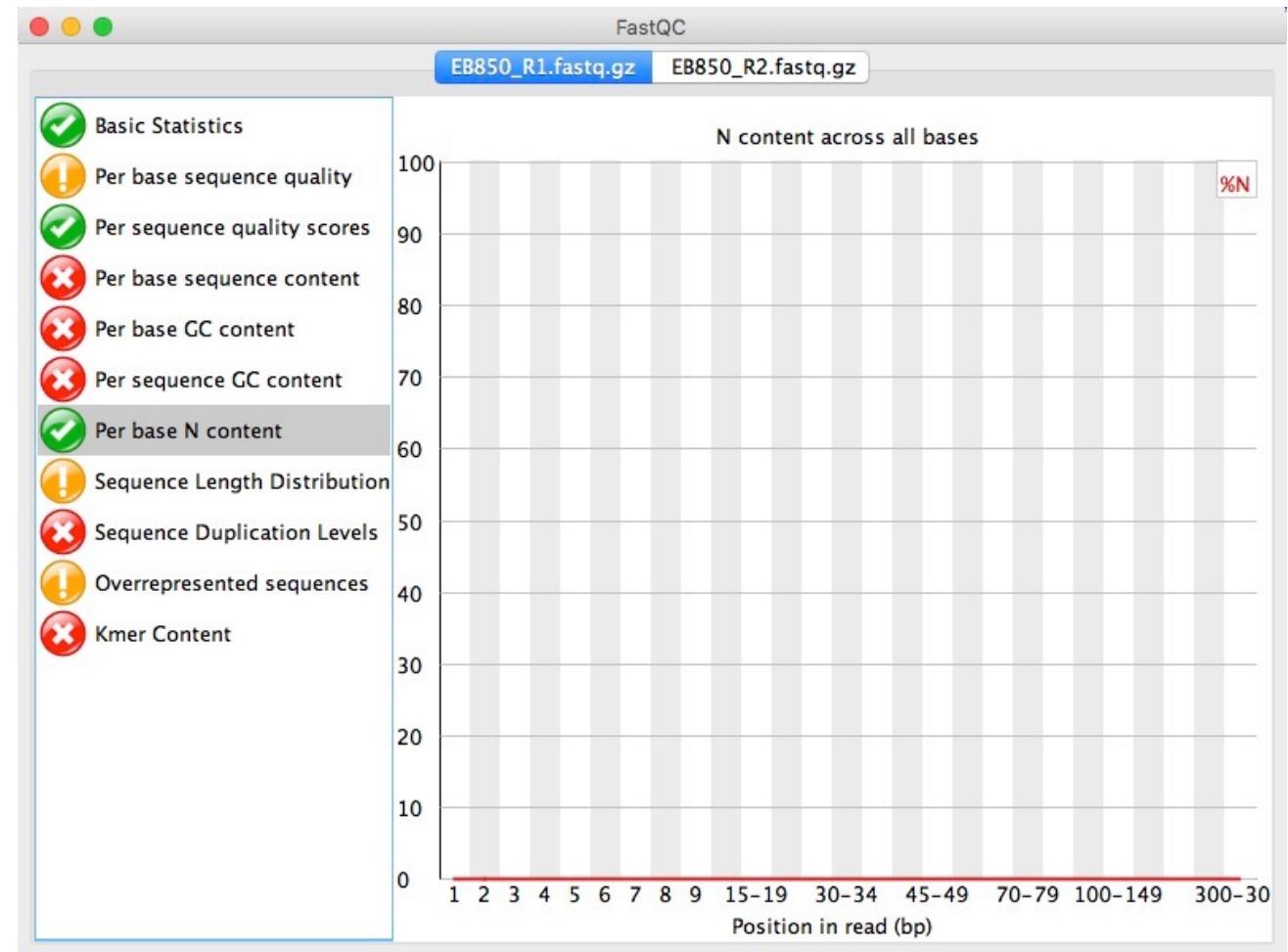
Quality assessment of raw reads and basic sequencing statistics can be evaluated using FastQC or MultiQC, or can also be done in QIIME2

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

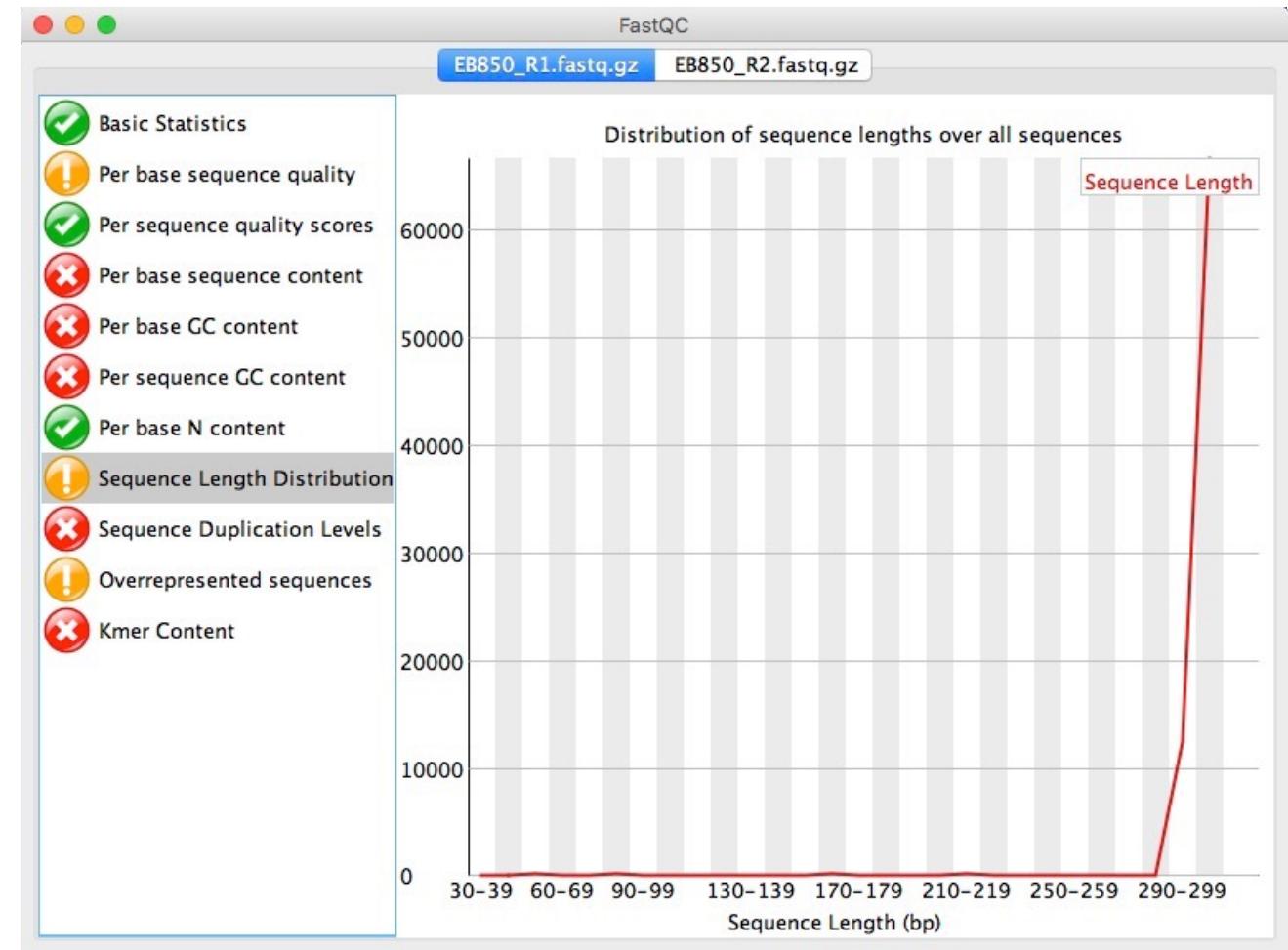
Quality assessment of raw reads and basic sequencing statistics can be evaluated using FastQC or MultiQC, or can also be done in QIIME2

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

Demultiplexing separates pooled libraries back into individual samples using the sequences of the barcode indices

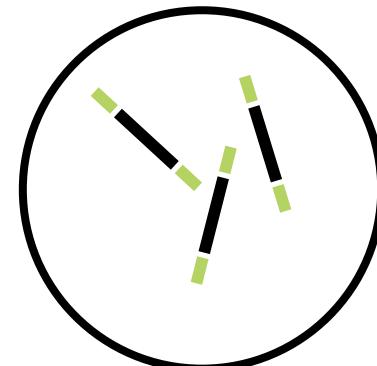
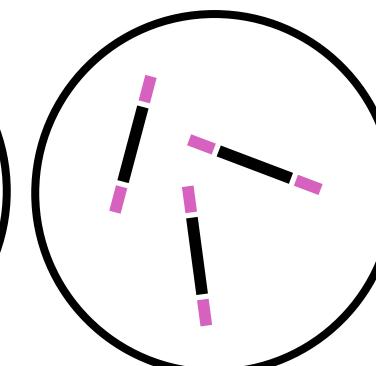
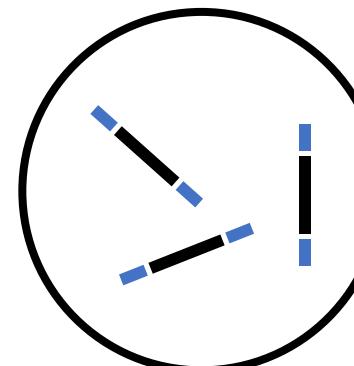
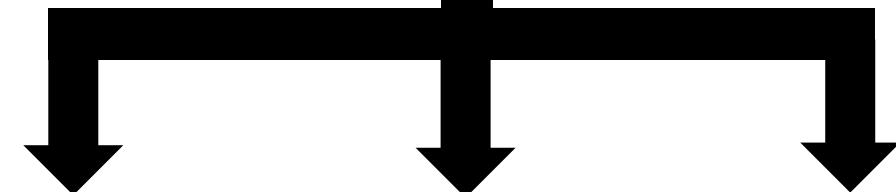
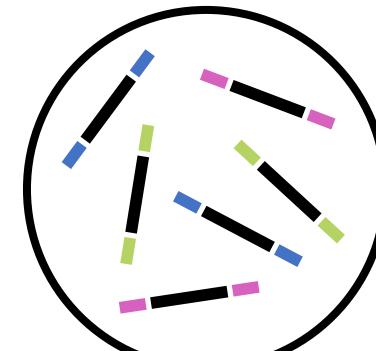
This step may be done by the sequencing center before returning your data to you and therefore might not be necessary

3 PRIMER TRIMMING

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT

POOLED LIBRARY



Sample 1

Sample 2

Sample 3

METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

Demultiplexing separates pooled libraries back into individual samples using the sequences of the barcode indices

This step may be done by the sequencing center before returning your data to you and therefore might not be necessary

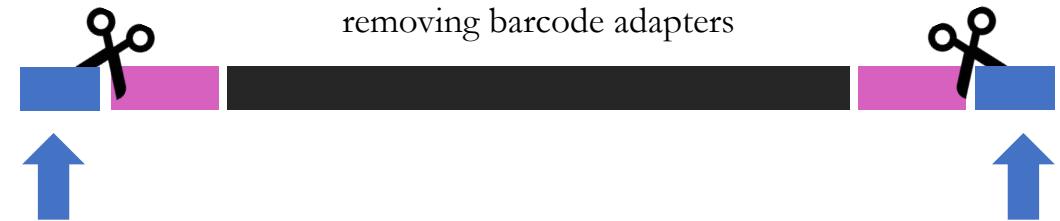
3 PRIMER TRIMMING

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT

DEMULITPLEXING

separating samples and removing barcode adapters



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

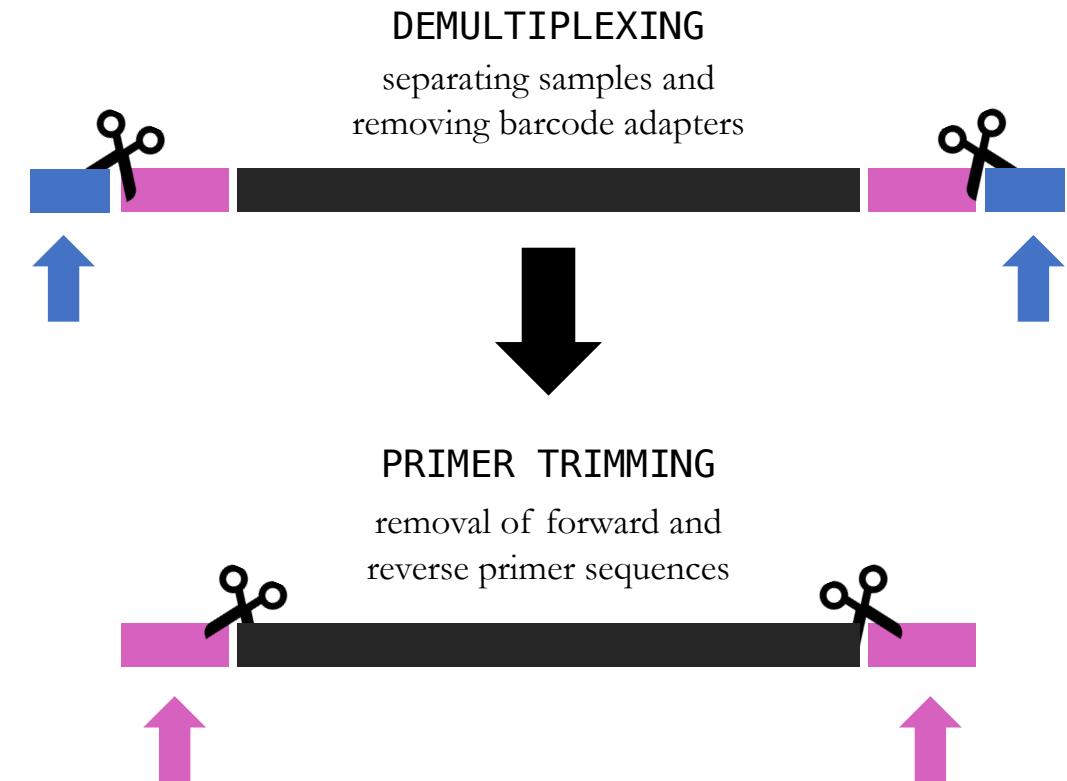
3 PRIMER TRIMMING

Primer trimming can be done using QIIME2 plugins or can be done using programs like Cutadapt or Trimmomatic outside of QIIME2

For paired-end reads, you will need to trim the forward and reverse primers from the front of the reads, and depending on the size of your amplicon and length of your sequencing run, you may need to trim the reverse complements of the primers from the back end of the reads

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

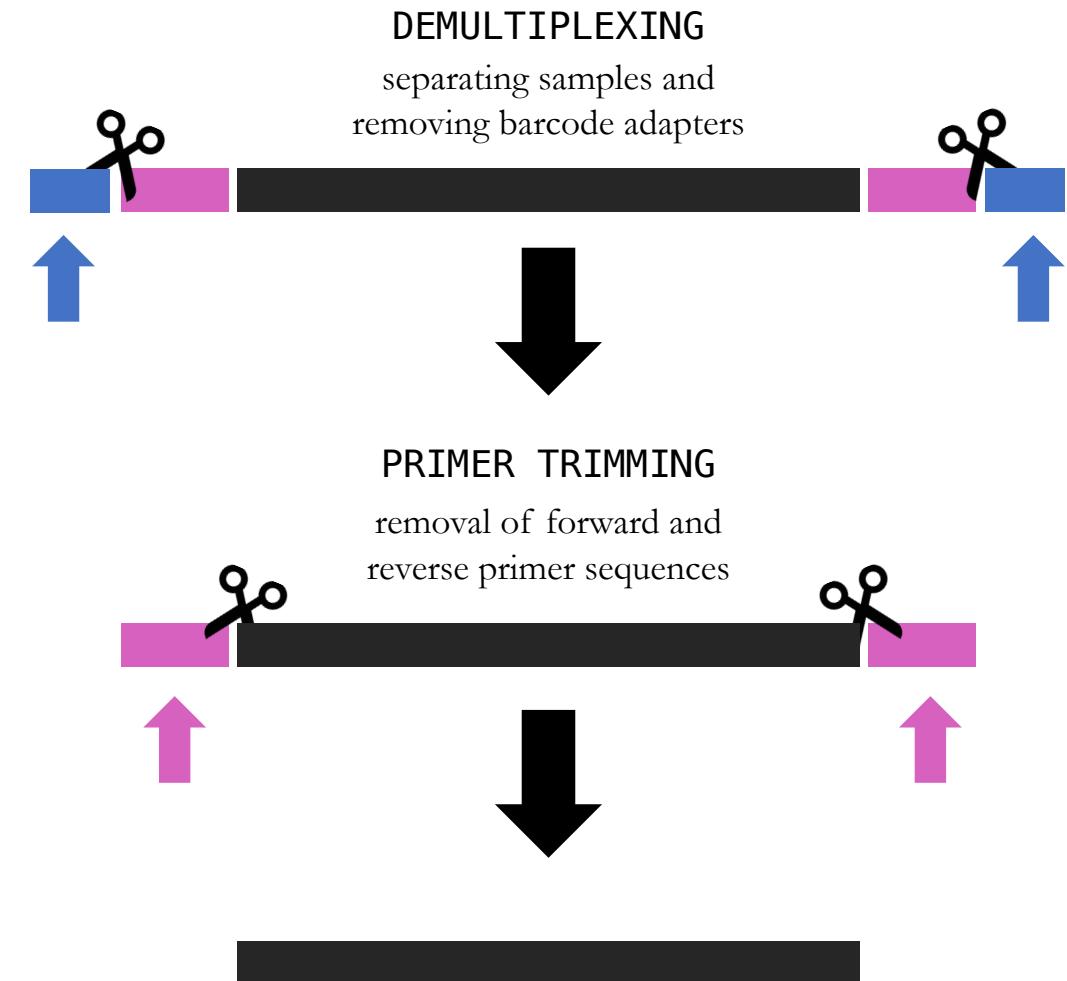
3 PRIMER TRIMMING

Primer trimming can be done using QIIME2 plugins or can be done using programs like Cutadapt or Trimmomatic outside of QIIME2

For paired-end reads you will need to trim the forward and reverse primers from the front of the reads and depending on the size of your amplicon and length of your sequencing run you may need to trim the reverse complements of the primers from the back end of the reads

4 DENOISING & MERGING

5 TAXONOMIC ASSIGNMENT



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras.

This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

MERGING

Goal: Minimize poor quality bases at ends of reads, maximize region of overlap for merging

FORWARD – 150bp



150bp – REVERSE



FINAL MERGED READ – 200bp



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

MERGING

Goal: Minimize poor quality bases at ends of reads, maximize region of overlap for merging

Example:

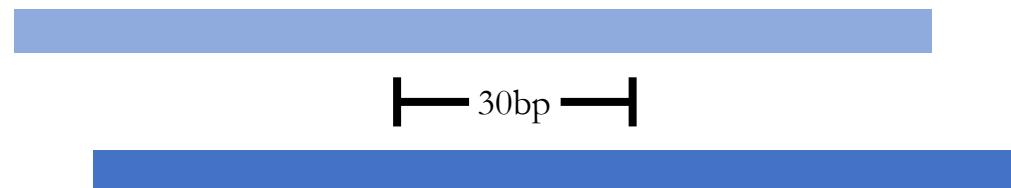
Forward/Reverse read length: 150bp

Expected merged read length: 200bp

Minimum overlap: 30bp

How much can I trim off the end of my reads while still maintaining my minimum overlap of 30bp?

FORWARD – 150bp



FINAL MERGED READ – 200bp

METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

MERGING

Goal: Minimize poor quality bases at ends of reads, maximize region of overlap for merging

Example:

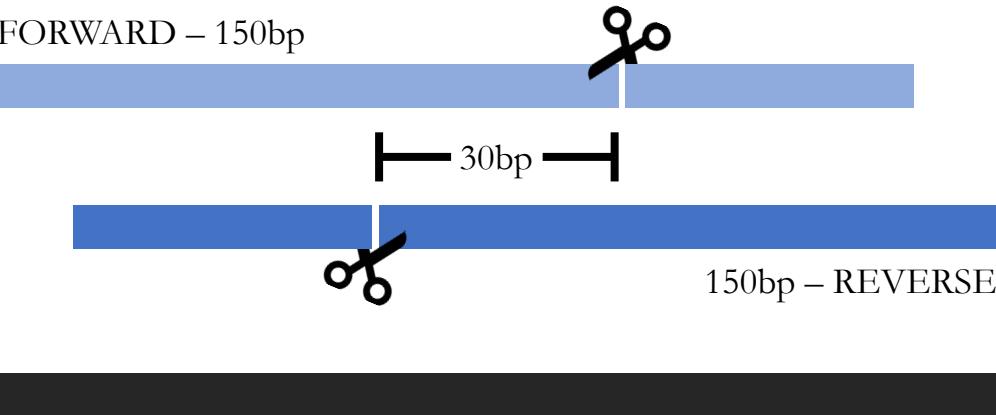
Forward/Reverse read length: 200bp

Expected merged read length: 120bp

Minimum overlap: 30bp

How much can I trim off the end of my reads while still maintaining my minimum overlap of 30bp?

FORWARD – 150bp



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

MERGING

Goal: Minimize poor quality bases at ends of reads, maximize region of overlap for merging

Example:

Forward/Reverse read length: 200bp

Expected merged read length: 120bp

Minimum overlap: 30bp

How much can I trim off the end of my reads while still maintaining my minimum overlap of 30bp?

FORWARD – ?



30bp



? – REVERSE



FINAL MERGED READ – 200bp

METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

MERGING

Goal: Minimize poor quality bases at ends of reads, maximize region of overlap for merging

Example:

Forward/Reverse read length: 200bp

Expected merged read length: 120bp

Minimum overlap: 30bp

How much can I trim off the end of my reads while still maintaining my minimum overlap of 30bp?

FORWARD – ?



30bp



? – REVERSE

Expected length + overlap

$$200 + 30 = 230$$



FINAL MERGED READ – 200bp

METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

MERGING

Goal: Minimize poor quality bases at ends of reads, maximize region of overlap for merging

Example:

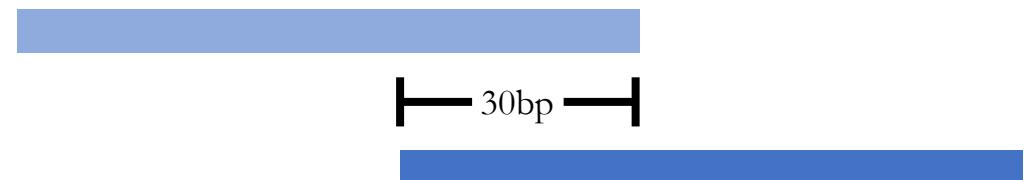
Forward/Reverse read length: 200bp

Expected merged read length: 120bp

Minimum overlap: 30bp

How much can I trim off the end of my reads while still maintaining my minimum overlap of 30bp?

FORWARD – 115bp



Expected length + overlap
 $200 + 30 = 230$

115bp – REVERSE



METABARCODING DATA ANALYSIS

- 1 QUALITY ASSESSMENT
- 2 DEMULTIPLEXING
- 3 PRIMER TRIMMING
- 4 DENOISING & MERGING

Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

- 5 TAXONOMIC ASSIGNMENT

CHIMERAS

Chimeras are artifacts of merging when sequences from two different biological sources are mistakenly merged together



METABARCODING DATA ANALYSIS

1 QUALITY ASSESSMENT

2 DEMULTIPLEXING

3 PRIMER TRIMMING

4 DENOISING & MERGING

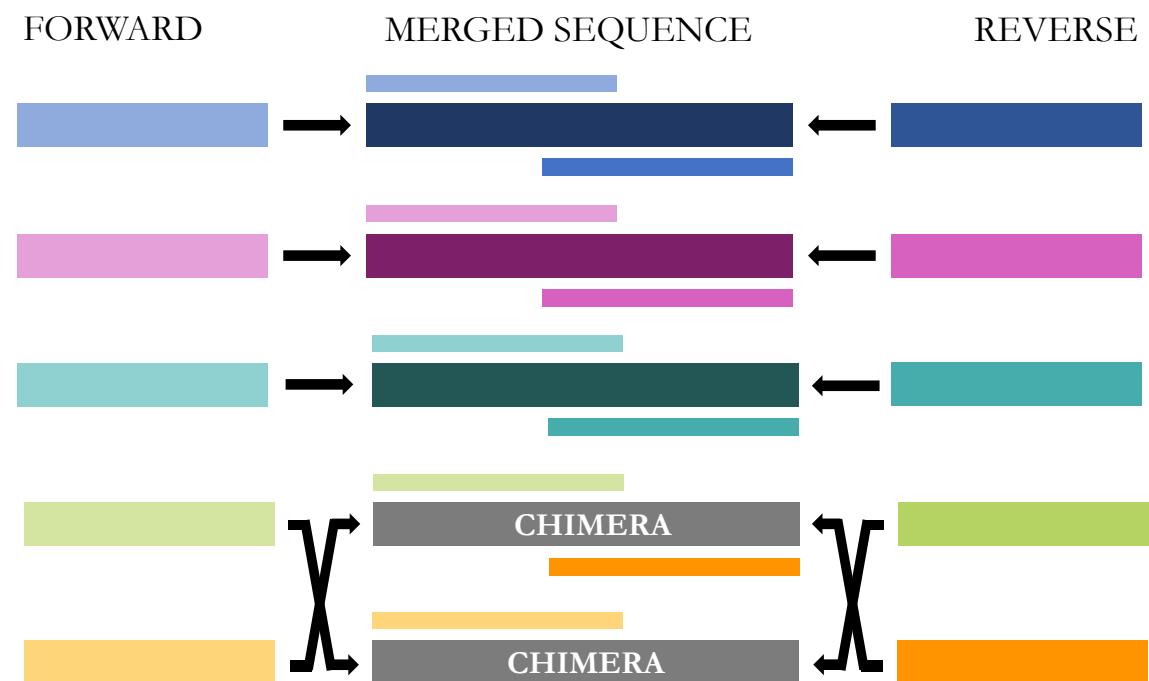
Denoising is the process of detecting and correcting sequencing errors where possible. Following denoising, paired-end reads are then merged and checked for chimeras. This step results in a list of amplicon sequence variants (ASVs) and count table summarizing the number of reads each ASV makes up per sample

QIIME2 offers plugins for both DADA2 and Deblur to accomplish this, but for the purposes of this workshop we will use the DADA2 plugin

5 TAXONOMIC ASSIGNMENT

CHIMERAS

Chimeras are artifacts of merging when sequences from two different biological sources are mistakenly merged together



METABARCODING DATA ANALYSIS

- 1 QUALITY ASSESSMENT
- 2 DEMULTIPLEXING
- 3 PRIMER TRIMMING
- 4 DENOISING & MERGING
- 5 TAXONOMIC ASSIGNMENT

QIIME2 offers a number of methods for classifying ASVs including a Naïve-Bayes classifier, BLAST, and more

Depending on your primer choice, you can use curated databases or you can create a custom database. For 18S, the SILVA and Protist Ribosomal Reference databases are commonly used databases



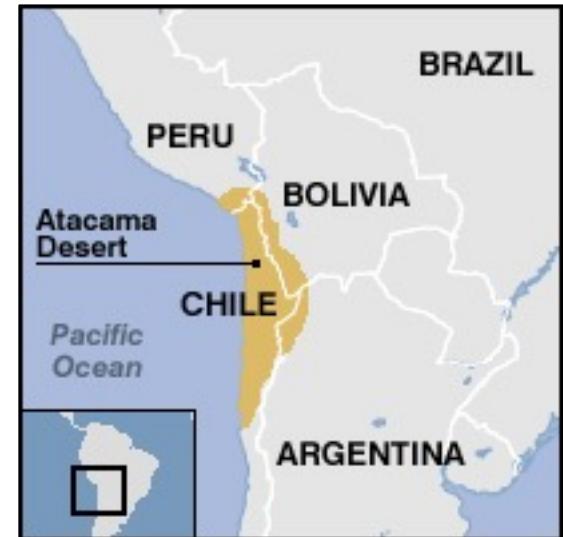
QIIME2 ATACAMA SOIL MICROBIOME TUTORIAL

Soil microbiome data from the Atacama Desert in Northern Chile, one of the most arid regions on the planet

Samples taken across two different east west transects (Baquedano and Yunguay) along which soil relative humidity was positively correlated with elevation

3 pits were dug at each site along the transect to have triplicates of each sample and soil samples were collected at 3 depths within each pit

V4 region of 16S rRNA was amplified from DNA of each soil sample to characterize microbial communities (bacteria & archaea)



QIIME2 ATACAMA SOIL MICROBIOME TUTORIAL

Soil microbiome data from the Atacama Desert in Northern Chile, one of the most arid regions on the planet

Samples taken across two different east west transects (Baquedano and Yunguay) along which soil relative humidity was positively correlated with elevation

3 pits were dug at each site along the transect to have triplicates of each sample and soil samples were collected at 3 depths within each pit

V4 region of 16S rRNA was amplified from DNA of each soil sample to characterize microbial communities (bacteria & archaea)

