

NATIONAL
MUSEUM *of*
**NATURAL
HISTORY**

Smithsonian



Laboratories of Analytical Biology

Genome Skimming: wet lab

Katie Murphy

Laboratories of Analytical Biology

NMNH

March 2024

Code of Conduct (Carpentries)



We are following the Carpentries organization's Code of Conduct



https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

Code of Conduct (excerpts)

https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

We are dedicated to providing a welcoming and supportive environment for all people, regardless of background or identity.

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

Our Plan for Today

1. Introductions
2. Plans for the week
3. Input DNA Quantification
4. Start Library Prep
5. Lecture: Genome Skimming, Library Prep
6. Continue Library Preparation

Introductions

Organizers for workshop:

Katie Murphy,

Tripp Macdonald,

Matt Kweskin,

Vanessa González

YOU:

Name, Unit,

Ice breaker: favorite
taxon... (species or higher
level)



Our Workflow for the Week:

1. Quantify your DNA w/ Qubit
2. Library Preparation: NEBNext Ultra II FS with iTru
3. Quantify Libraries on the iD3 Plate Reader with Quant-iT
4. Assess library size on the Tapestation 4200
5. Library Pooling & QC
6. qPCR
7. Sequencing Form Submission to LAB

Before we head to lab...



Take a look at the protocols



Working in pairs at lab benches
(share reagents/equipment)

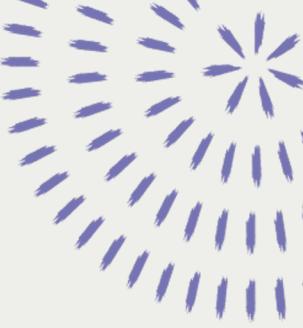


Calculate reaction mixes for
Library Steps 1 and 2



To the Lab!

Genome Skimming



“Low-pass, shallow sequencing of WGS libraries”

Low-cost high-value genetic data

Useful for recovering high-copy fraction of the genome

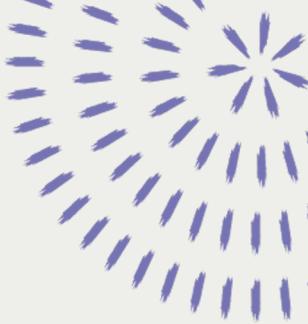
- Organellar genomes (Mito, plastid)
- Ribosomal DNA
- Nuclear repeats (microsats, transposable elements)

No specialized lab workflow needed (*outside of standard NGS techniques*)

Recommend enzymatic shearing to speed up workflows, limit sample handling



Why use it?



Straightforward wet lab methods



Robust method that works across a wide range of DNA input quality and quantity - great for museum specimens

- successful assembly of targets is not guaranteed -



More informative than traditional barcoding studies



Build a better reference database!

Key terms:

Library

DNA Insert

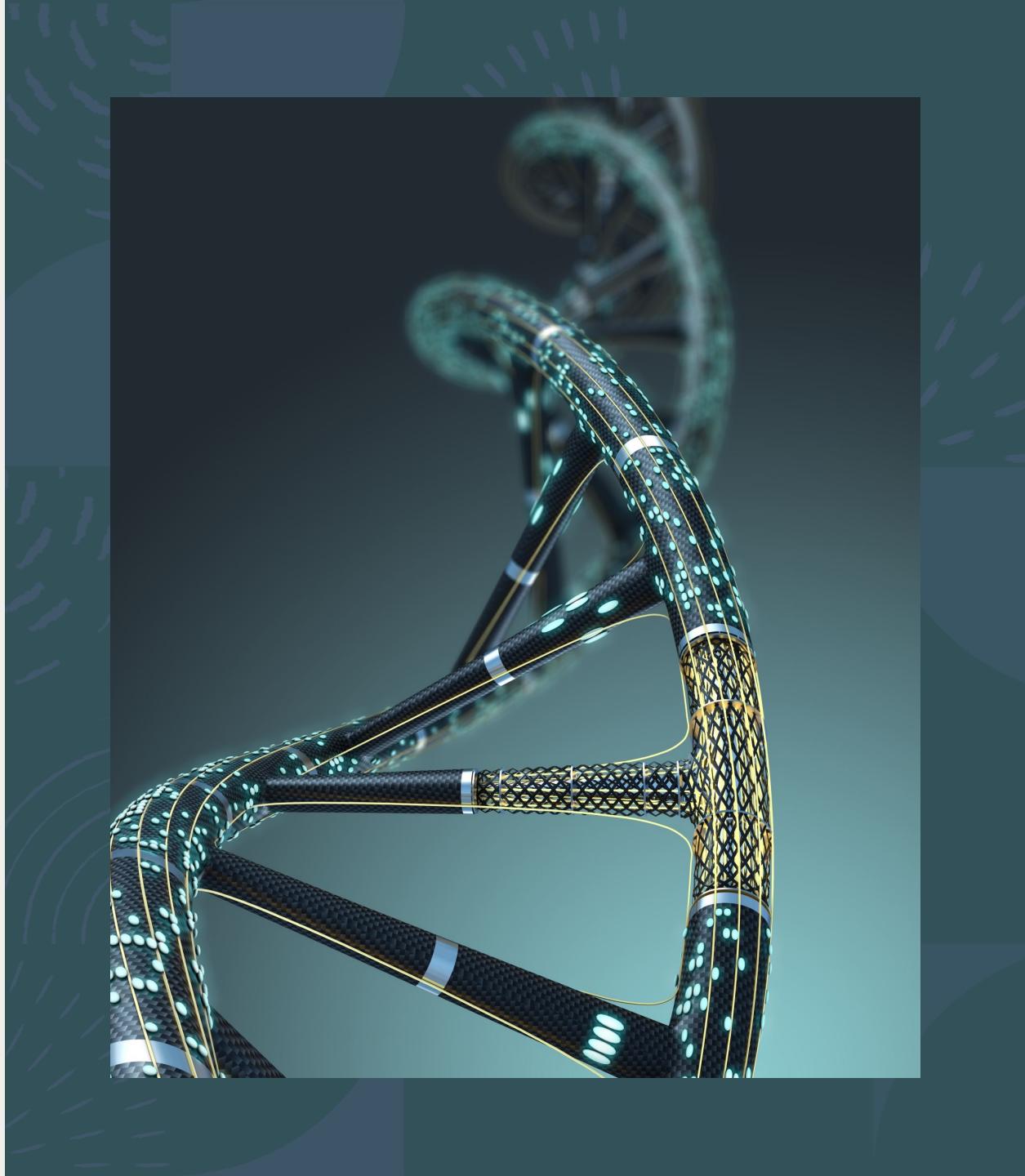
Pool

Adapter

Index/indices

Read

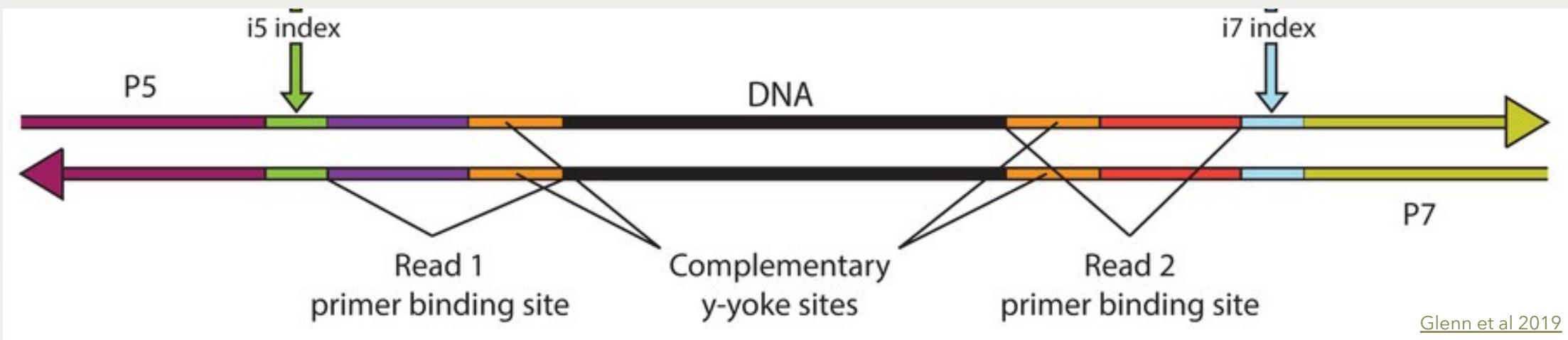
Read length



Key terms

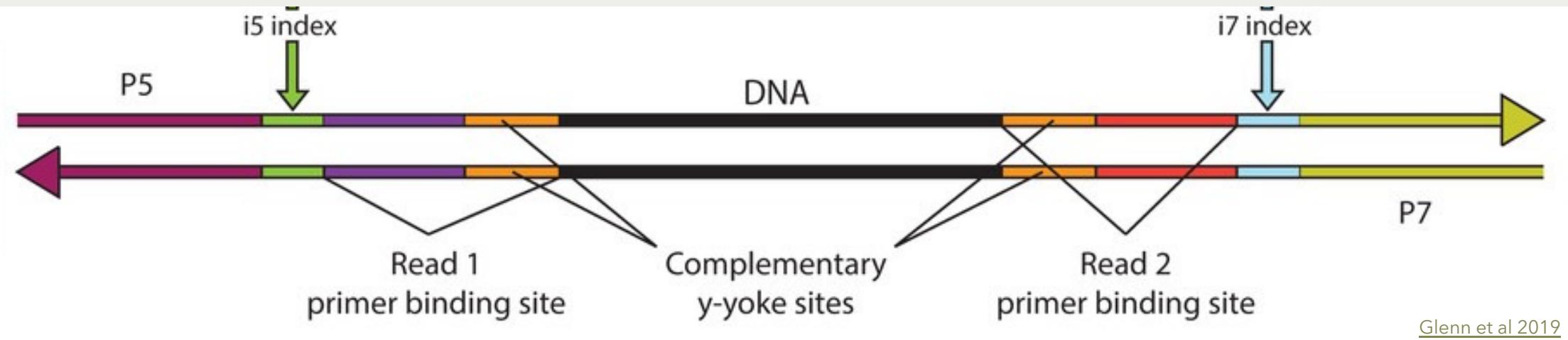
- **Library**

- a collection of DNA fragments originating from a single sample, that have been prepared for Next Generation Sequencing (NGS).
 - Each library fragment contains a DNA **insert** from your sample, plus special sequences called “indices” and “adapters”.



Key terms

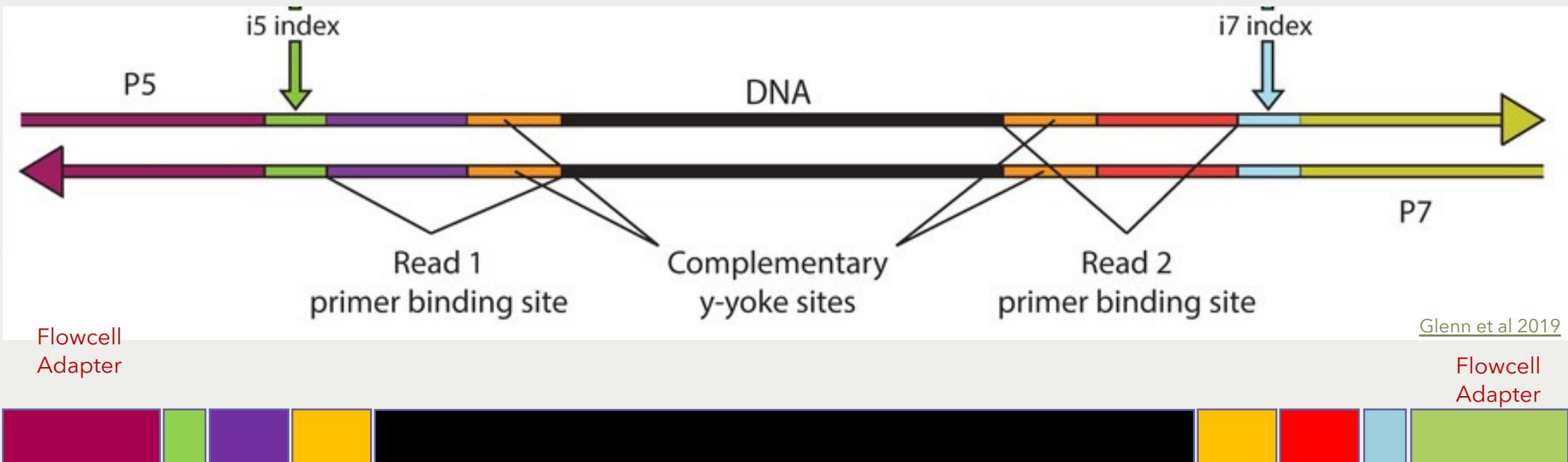
- **Index** (plural = indices)
 - a unique DNA “tag”, or sequence, usually 6-12 bp in length, that is added onto the ends of each DNA library. This unique ID allows you to pool (“multiplex”) multiple samples together onto a single sequencing run, and then bioinformatically separate (“demultiplex”) the sequencing data back into individual samples afterwards.
 - Our libraries will have two indices – **i5** & **i7**



Key terms

- **Adapter**

- specific sequences of DNA that are added to all DNA libraries in order to make them "sequenceable"
- One adapter is the iTru y-yoke adapter, which contain the read binding sites and is located between our inserts and our indices
- The second adapter type is the flowcell adapter, the outermost region of our library fragments



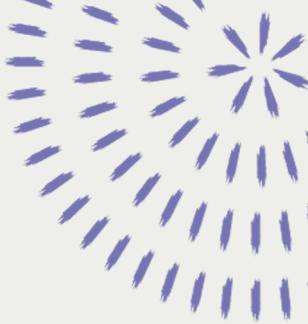
Key terms

- **Pool**

- a group of individual libraries that have been multiplexed (pooled), which can be loaded onto a sequencer



Key terms



Read

- The sequence of bases (nucleotides) corresponding to one DNA fragment

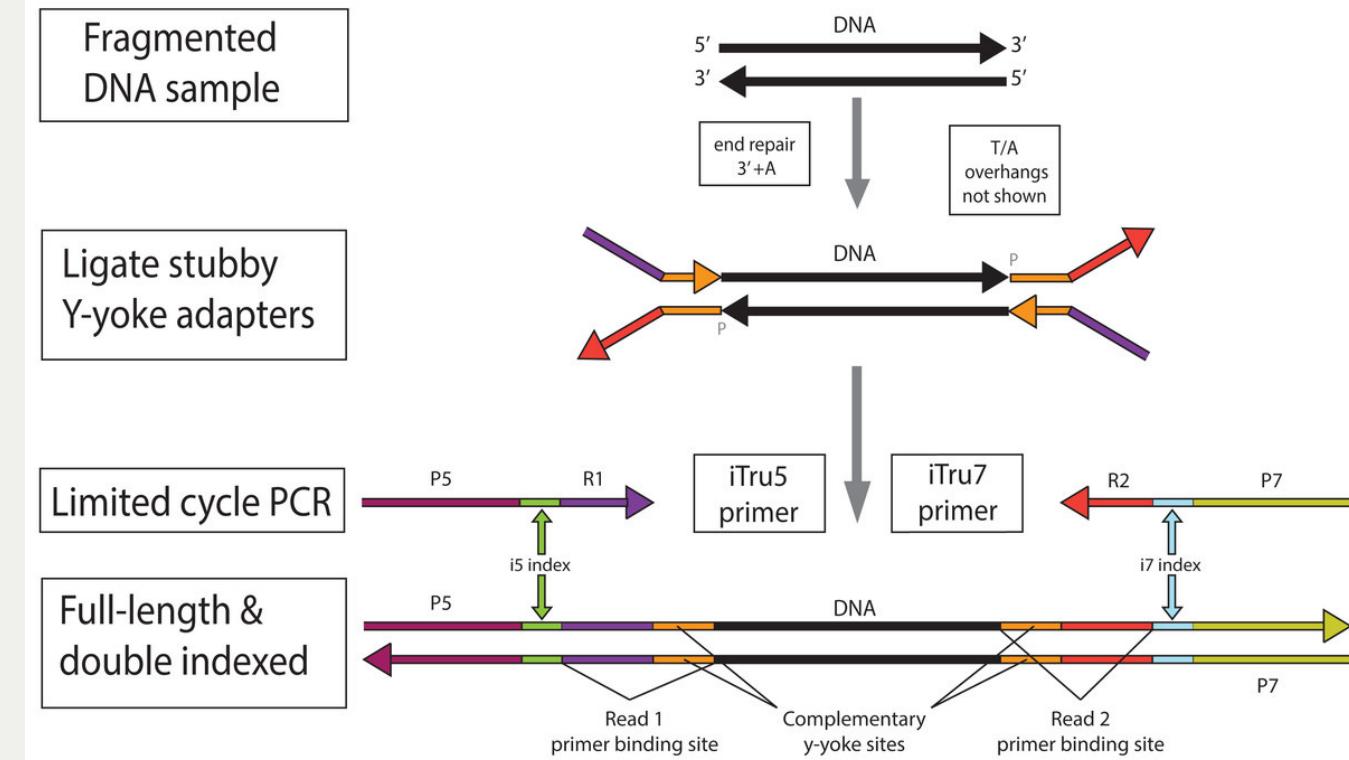
Read length

- Depends on the sequencing platform and type of input material
- For Illumina, read length is usually 150-300 bp (defined when the sequencing run is set up, and limited by run chemistry).
- For long read sequencers such as PacBio or Oxford Nanopore, read length can be > 100 kB, and is often limited by your fragment size rather than by the sequencing chemistry.

Illumina Library Preparation with iTru adapter & indices: the general workflow

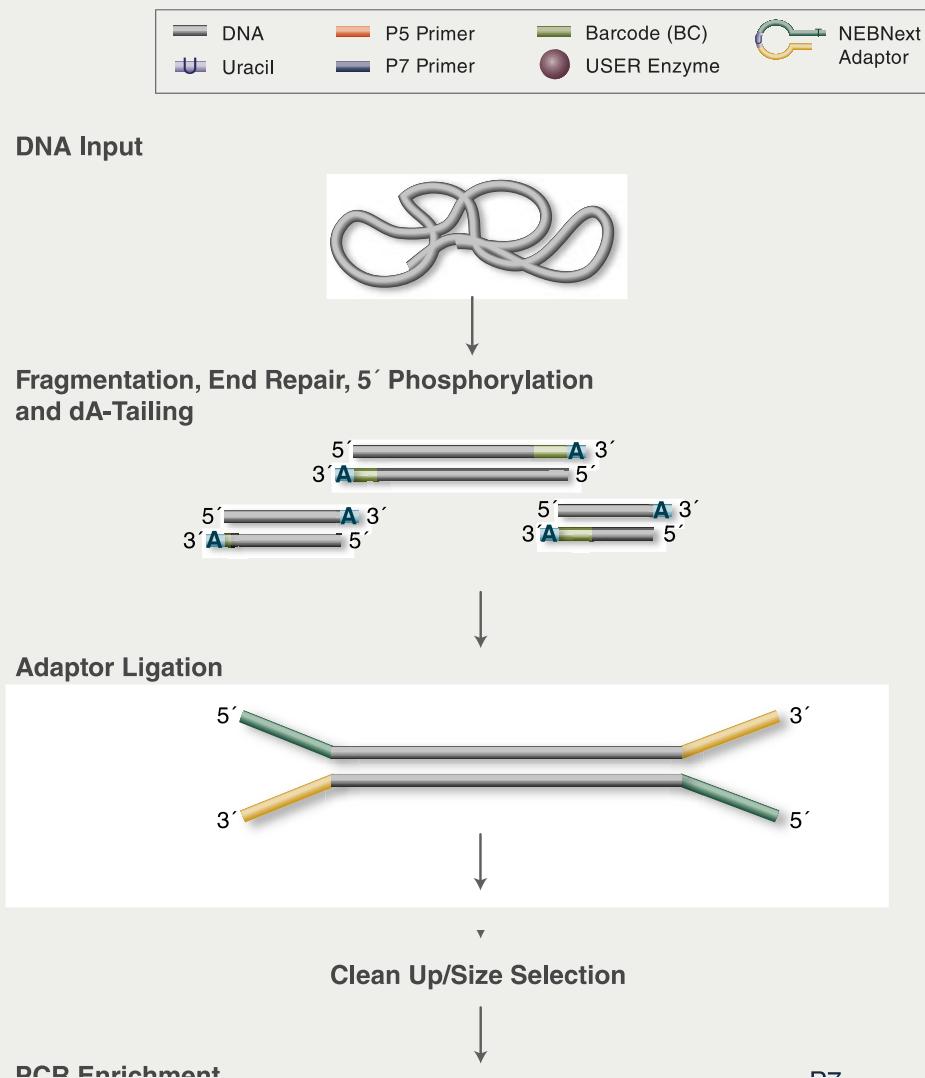
Glenn et al 2019

iTru Library Method

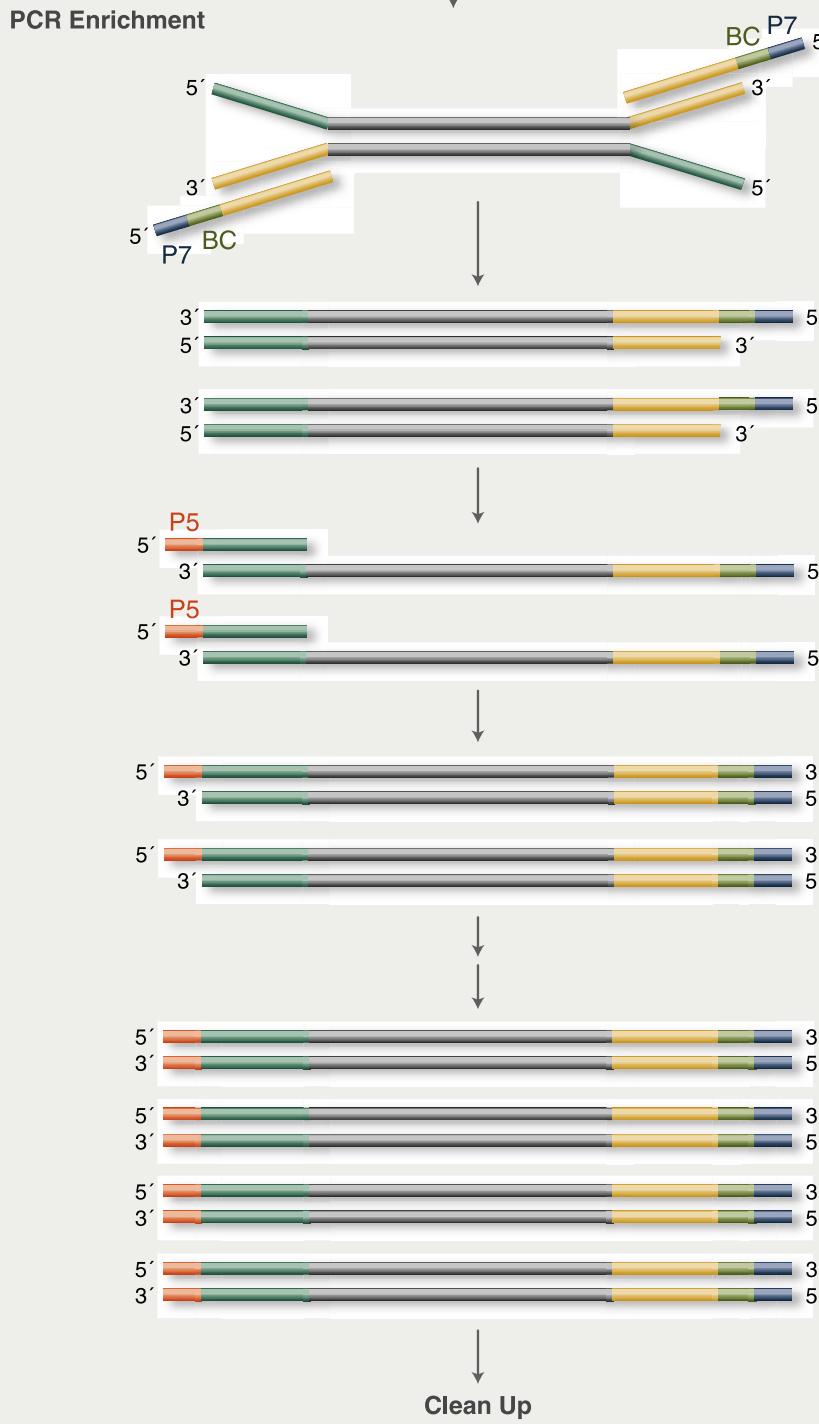


Library Preparation: NEBNext Ultra II FS

Figure 1. Workflow demonstrating the use of NEBNext Ultra II DNA Library Prep Kit Illumina.



Library Preparation: NEBNext Ultra II FS



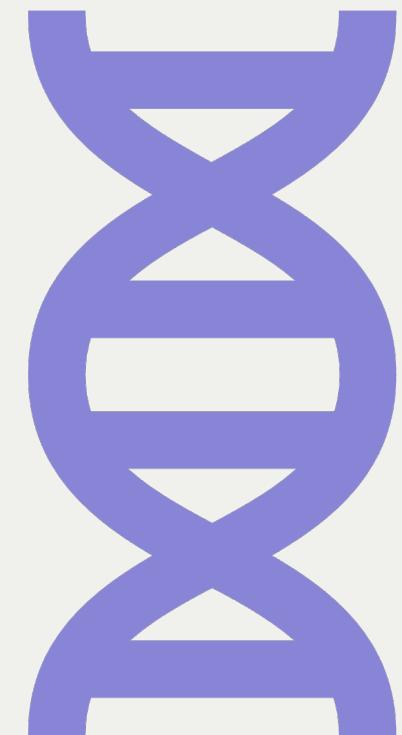
Input DNA considerations:

- Quantity: 500 pg - 500 ng
 - The more DNA you have, the greater DNA diversity will be present in your sample
- Quality - this will impact what fragmentation you need to do.
 - A wide range of quality works; no set relationship between DNA quality and success of assembling the full mito genome.



DNA Fragmentation Method

- **Enzymatic:**
 - Much quicker when working with numerous samples; reduces sample handling
 - May require optimization; has some bias in terms of fragmentation pattern
- **Sonication:**
 - the “gold standard”; less biased fragmentation than enzymatic
 - Slower process with more sample handling than enzymatic
- **Can you omit DNA fragmentation? Yes, but...**
 - Make sure you adjust your library prep protocol accordingly!
→ If following LAB’s protocol w/ NEB, you’ll need to purchase a separate end repair & dA tailing module from NEB in order to prep libraries without fragmentation.

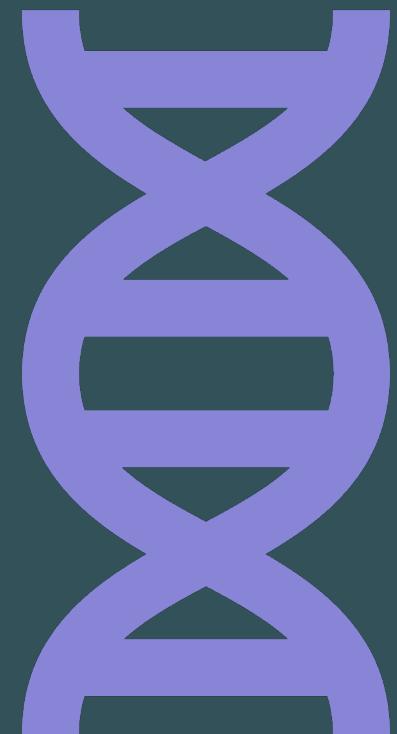


Insert size: What is best?

You want to maximize your sequencing info!

Ideal insert size is roughly the size of 2x your single read length.

If your sequencing run will be paired end 150 bp reads, that means you'll get 300 bp worth of sequencing info (2x150bp). So you'll get the most sequencing information if your insert size is 300 bp or longer.



Library Kit Selection

- Our protocol uses NEB as it is cost effective and many users are familiar with it, but any commercial brand should work.
 - The other most common kit around here is KAPA HyperPrep or HyperPlus Library prep kits. The cost is fairly similar to NEB, and should have similar success. If you're already used to using this brand, feel free to stick with it!
 - Make sure to follow the kit's protocol if you're using an alternative from NEB.

Index Selection

- Highly recommended: Unique dual indices
 - No i5 or i7 index is re-used between samples
 - Minimizes risk of contamination in sequencing data due to index hopping (where one index “hops” to the wrong sample)

Our Plan for Today

1. Introductions

2. Plans for the week

3. Input DNA Quantification

4. Start Library Prep

5. Lecture: Genome Skimming, Library Prep

6. Continue Library Preparation