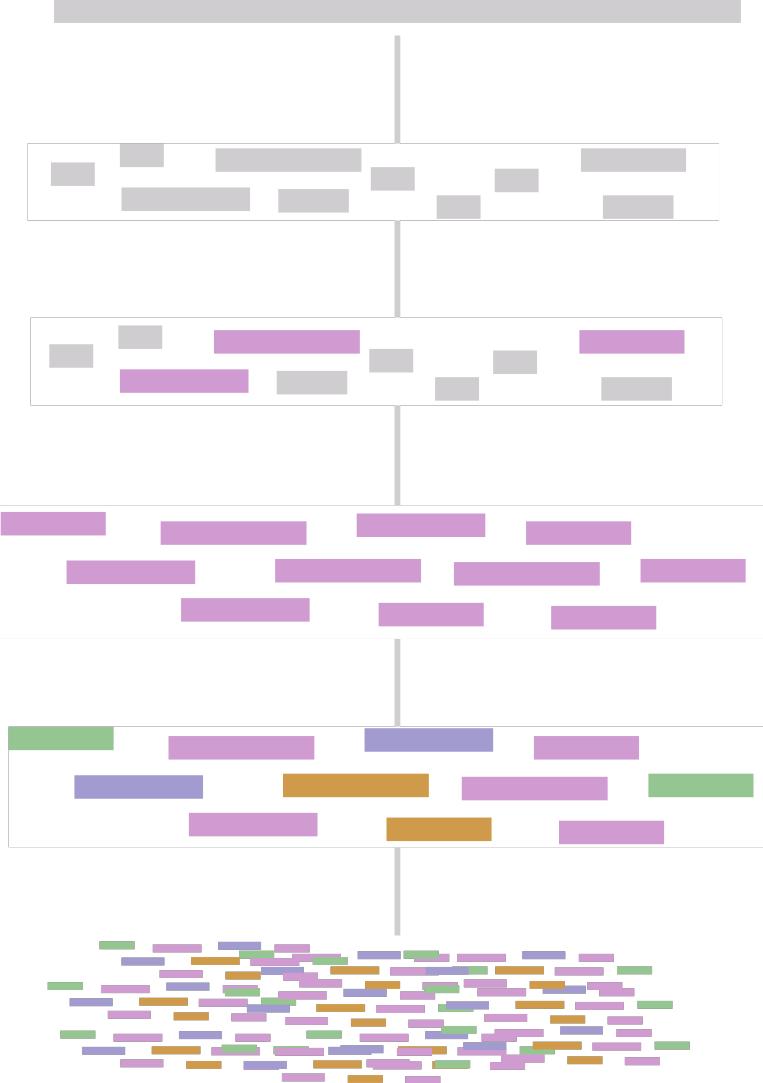


Genomic DNA



# ddRAD: in the laboratory

# Laboratory steps for ddRAD

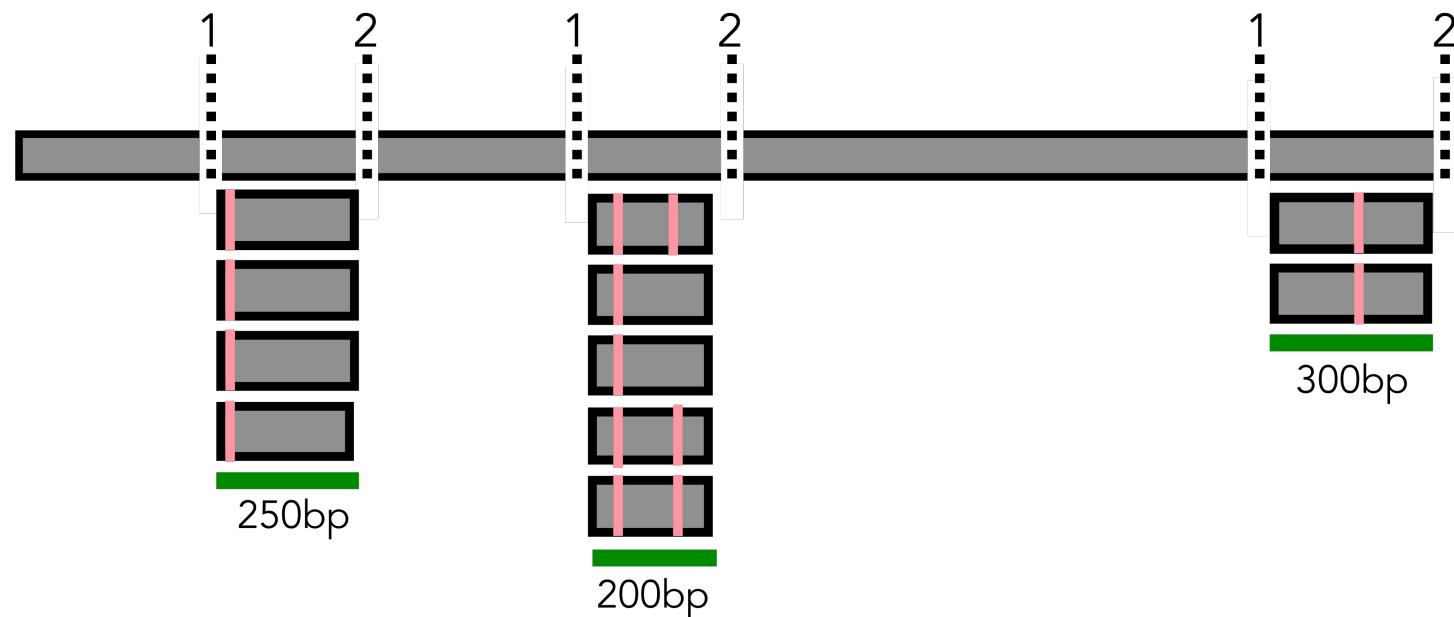
- DNA extraction, cleaning, & quantification (*I won't be providing details about this today*)
- (1) Enzyme selection
- (2) Library preparation

# (1) Enzyme selection

- With ddRAD (not so with other methods like 2bRAD), we're using **two restriction enzymes**
- One is a **frequent** cutter, the other is a **rare** cutter
- This combination will give us customizability of the amount of genome sequenced
- Common enzymes: *SphI*, *MluI*, *EcoRI*, *NlaIII*, *MspI*, *SbfI*

# (1) Enzyme selection: Bioanalyzer check

How many fragments/loci are there?

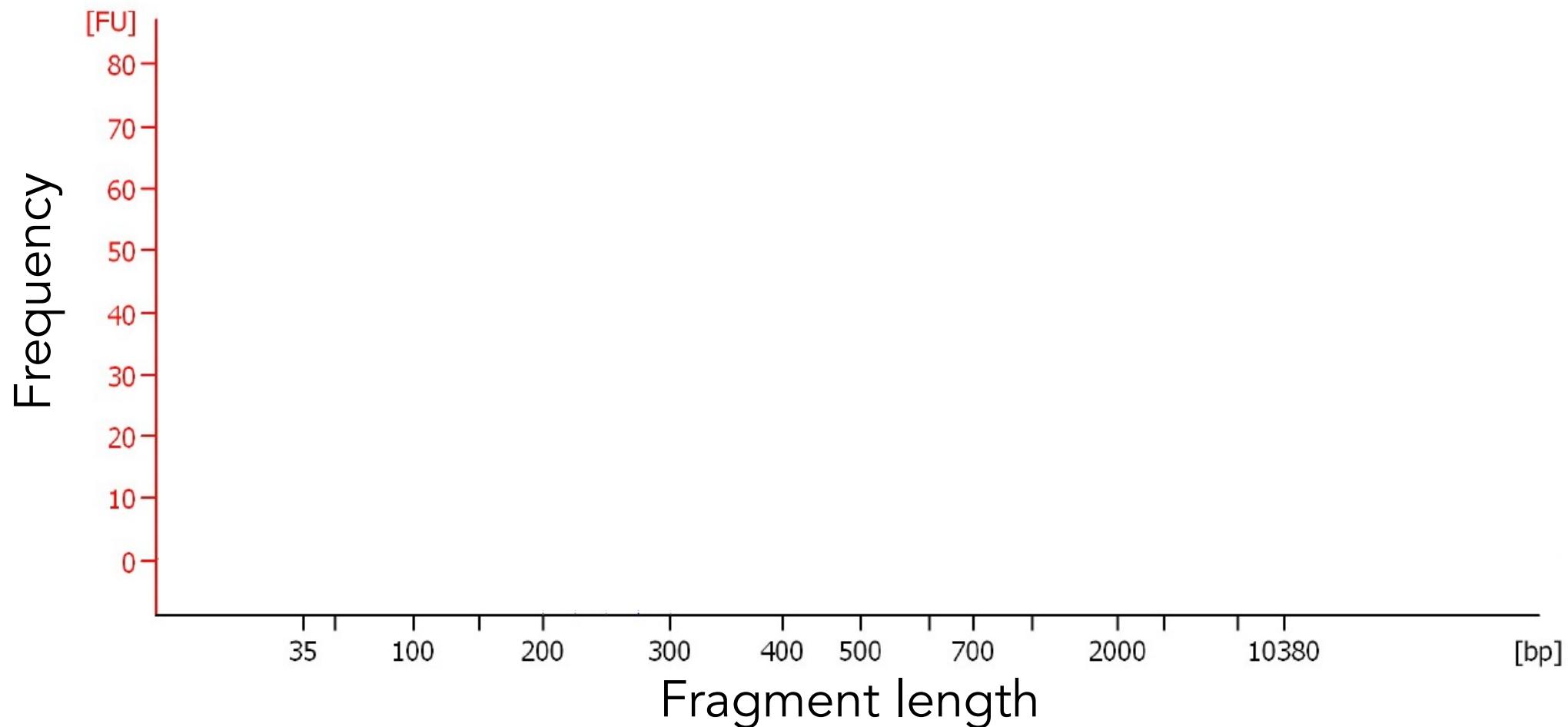


How long are the fragments/loci?

# (1) Enzyme selection: Bioanalyzer check

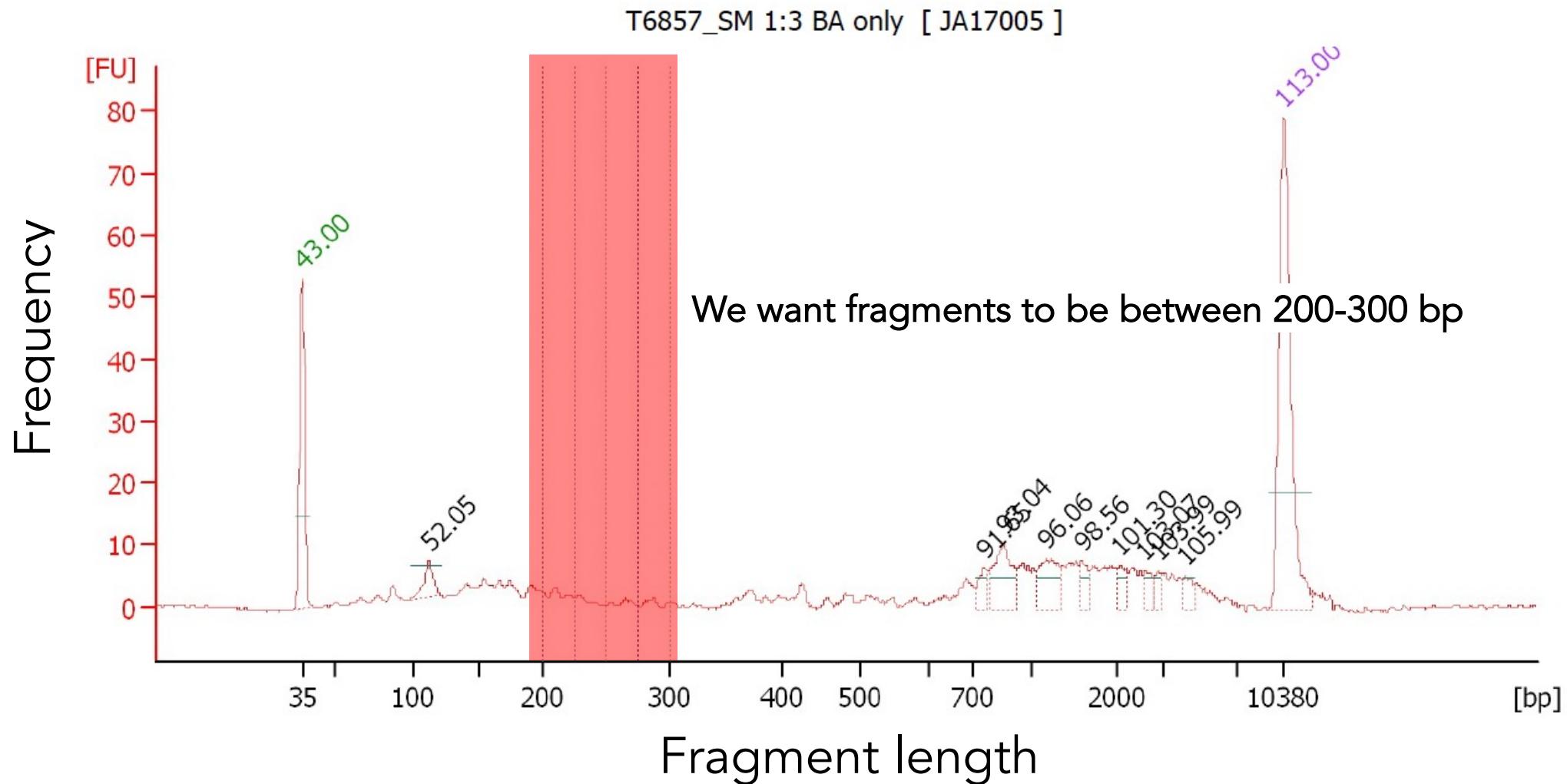
- Digest test samples
  - Straightforward: adding a set volume of RE to DNA
  - Always keep on ice!
  - Thermocycler for reaction & amplification to happen

# (1) Enzyme selection: Bioanalyzer check

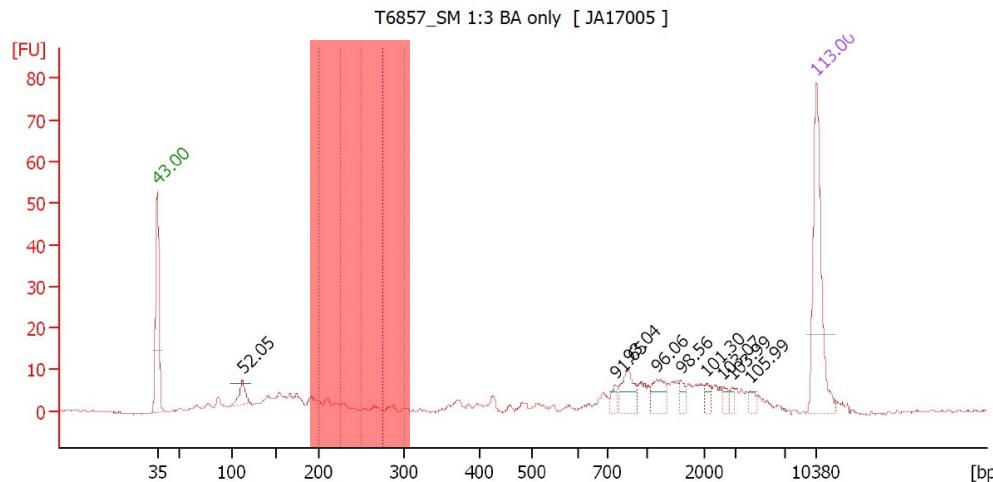


# (1) Enzyme selection: Bioanalyzer check

We are looking to amplify ~2% of the genome



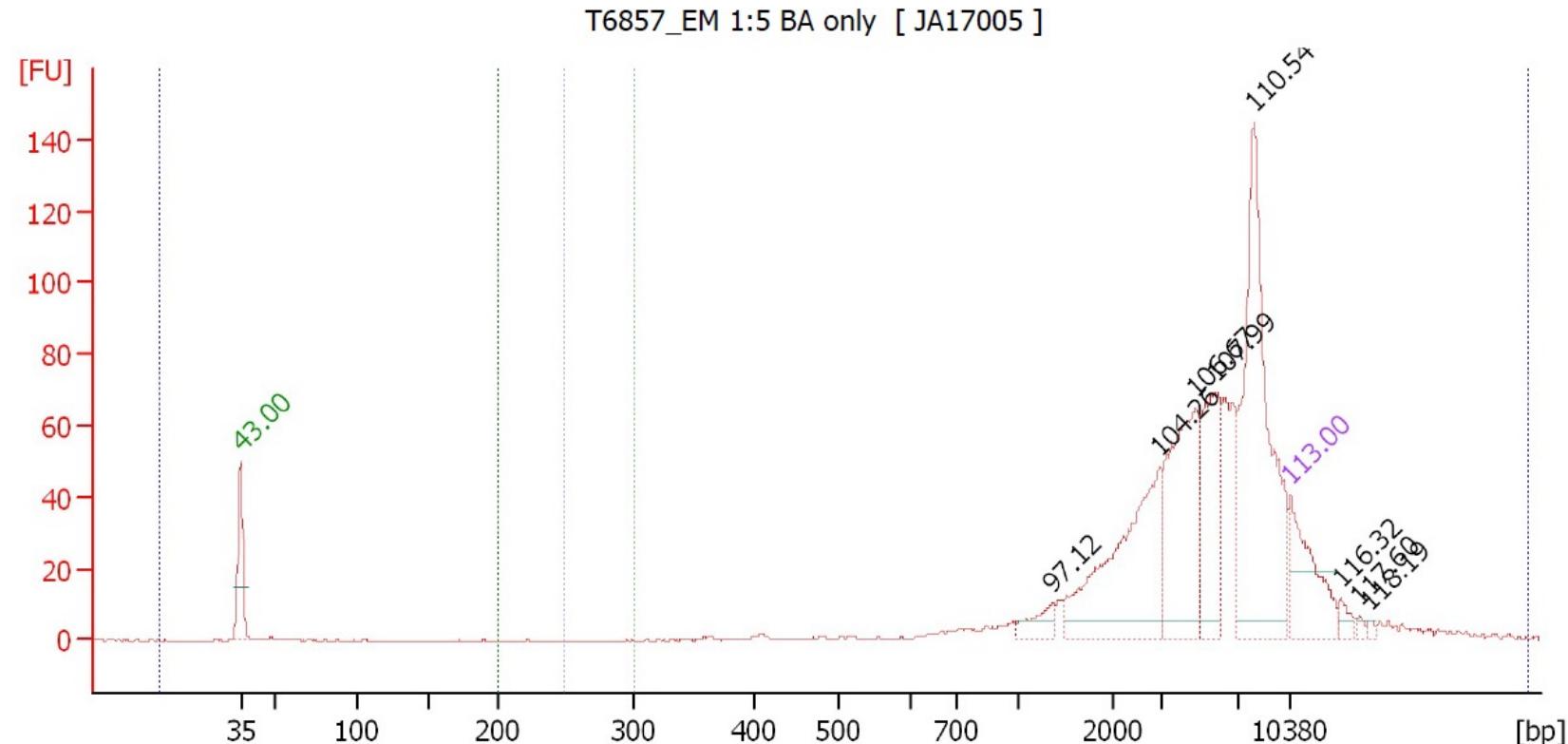
# (1) Enzyme selection: Bioanalyzer check



**Region table for sample 3 :** T6857 SM 1:3 BA only

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
200	250	10.2	5	217	5.5	18.61	130.2	
225	275	4.5	2	244	6.7	7.99	49.8	
250	300	3.2	2	275	4.6	5.54	30.6	

# (1) Enzyme selection: Bioanalyzer check



Region table for sample 4 :

T6857\_EM 1:5 BA only

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
0	25,780	804.1	100	6,130	61.7	959.67	763.9	
200	250	0.0	0	222	0.0	0.00	0.0	
200	300	0.0	0	222	0.0	0.00	0.0	

## (2) Library preparation

Original protocol and other helpful info can be found:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0037135#s5>

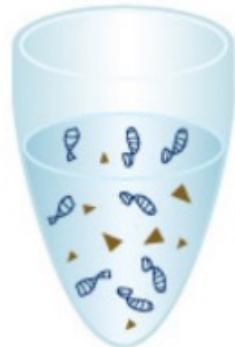
Make your own beads:

[https://ethanomics.files.wordpress.com/2012/08/serapure\\_v2-2.pdf](https://ethanomics.files.wordpress.com/2012/08/serapure_v2-2.pdf)

# Before we start... bead cleanup

- Takes place at the end of nearly every step

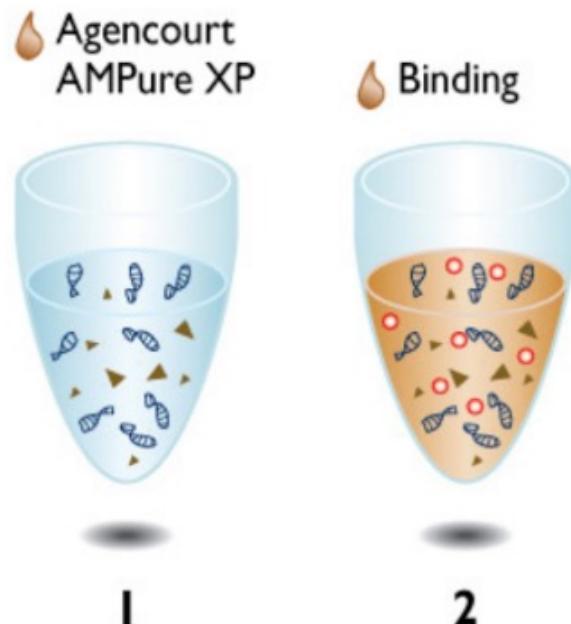
Agencourt  
AMPure XP



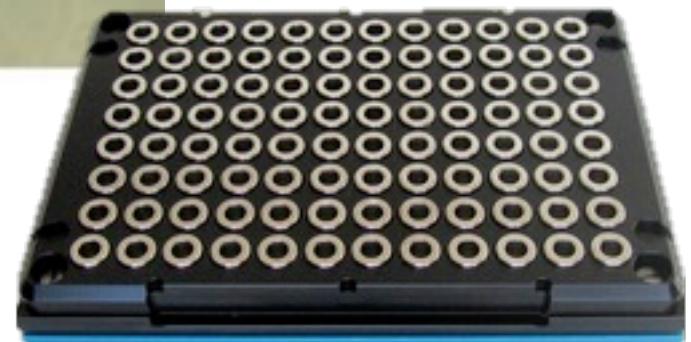
I

# Before we start... bead cleanup

- Takes place at the end of nearly every step



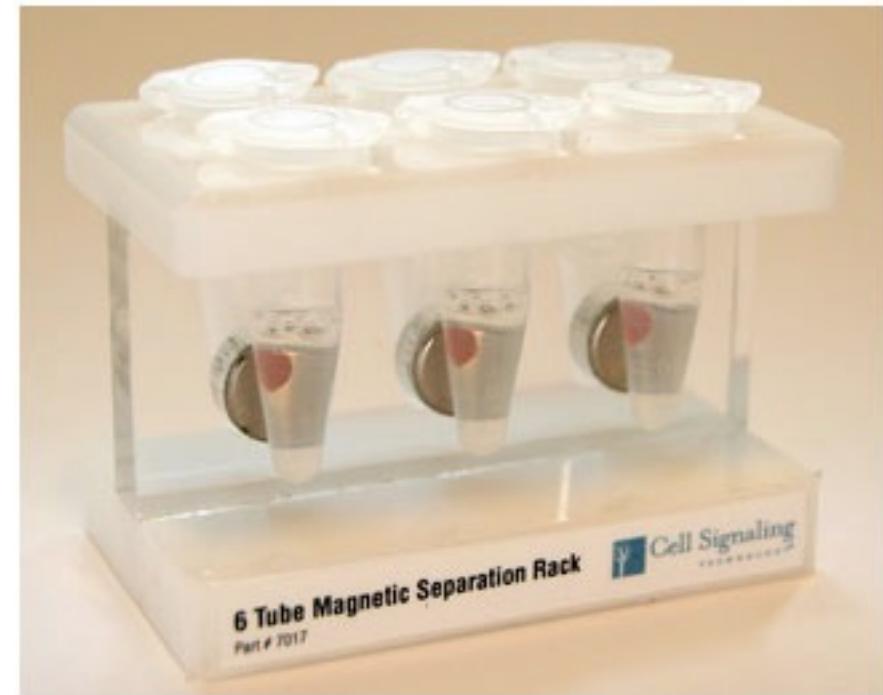
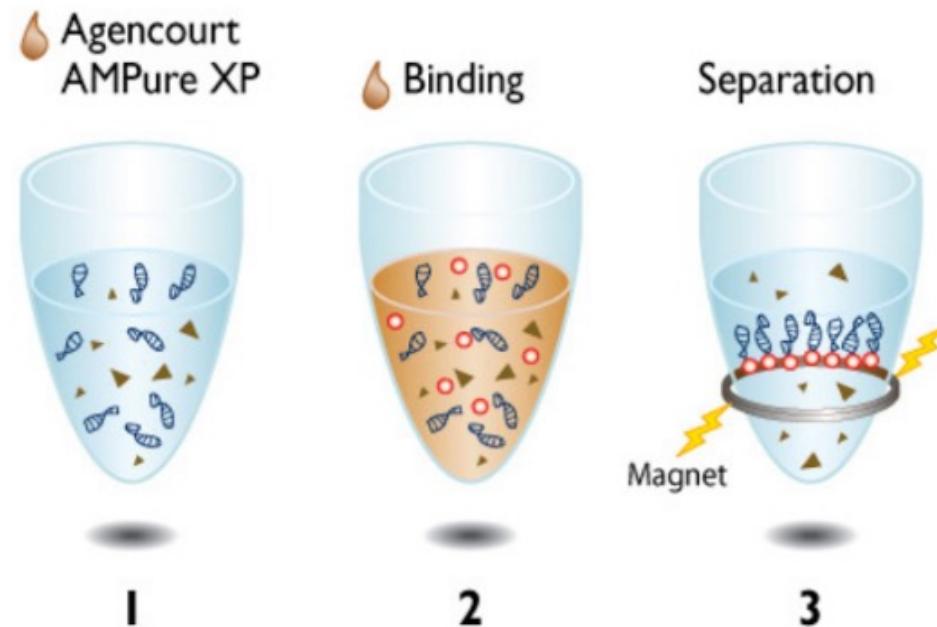
Magnetic bead separator rack



SPRIplate (96 samples)

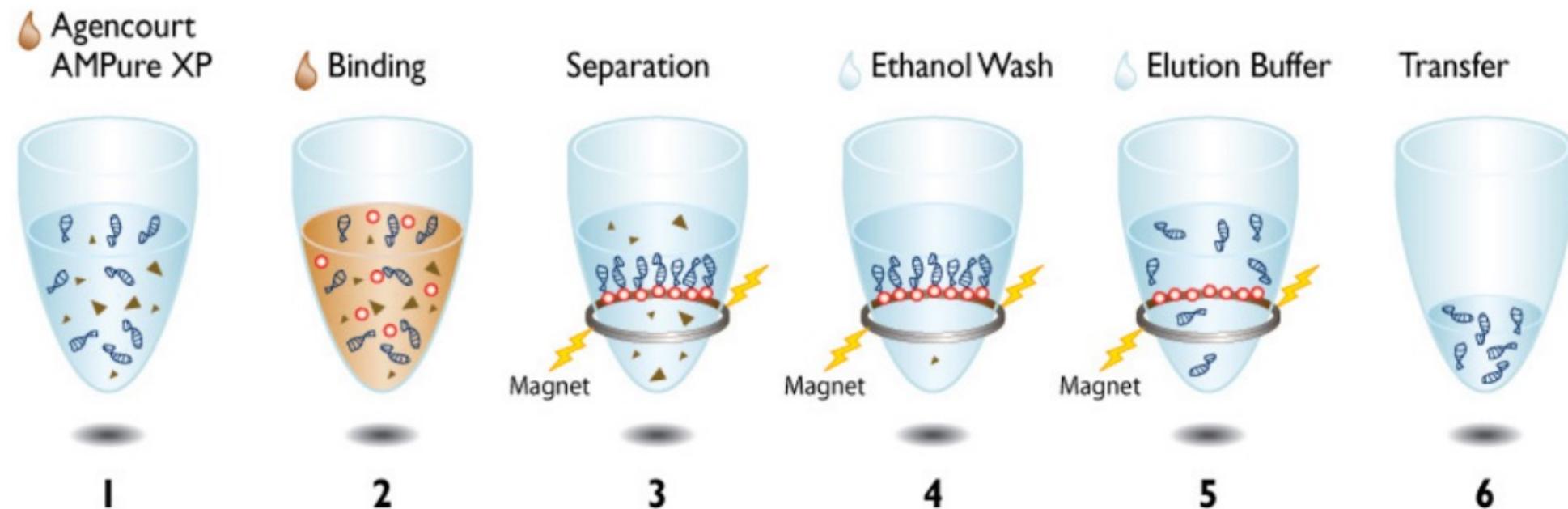
# Before we start... bead cleanup

- Takes place at the end of nearly every step

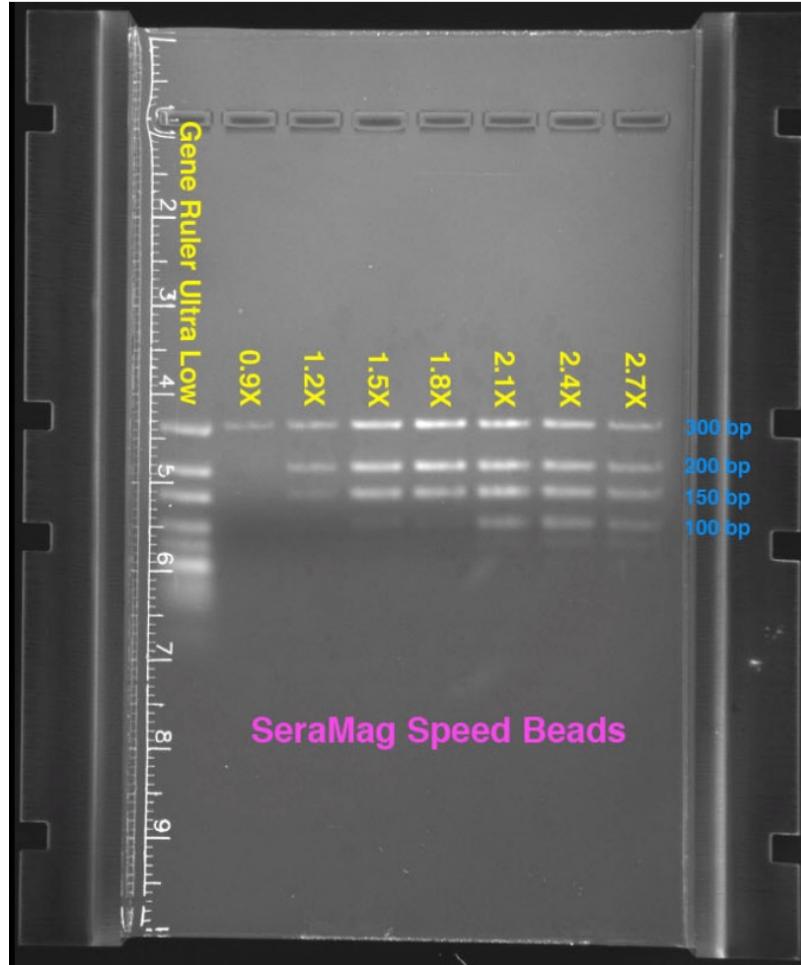


# Before we start... bead cleanup

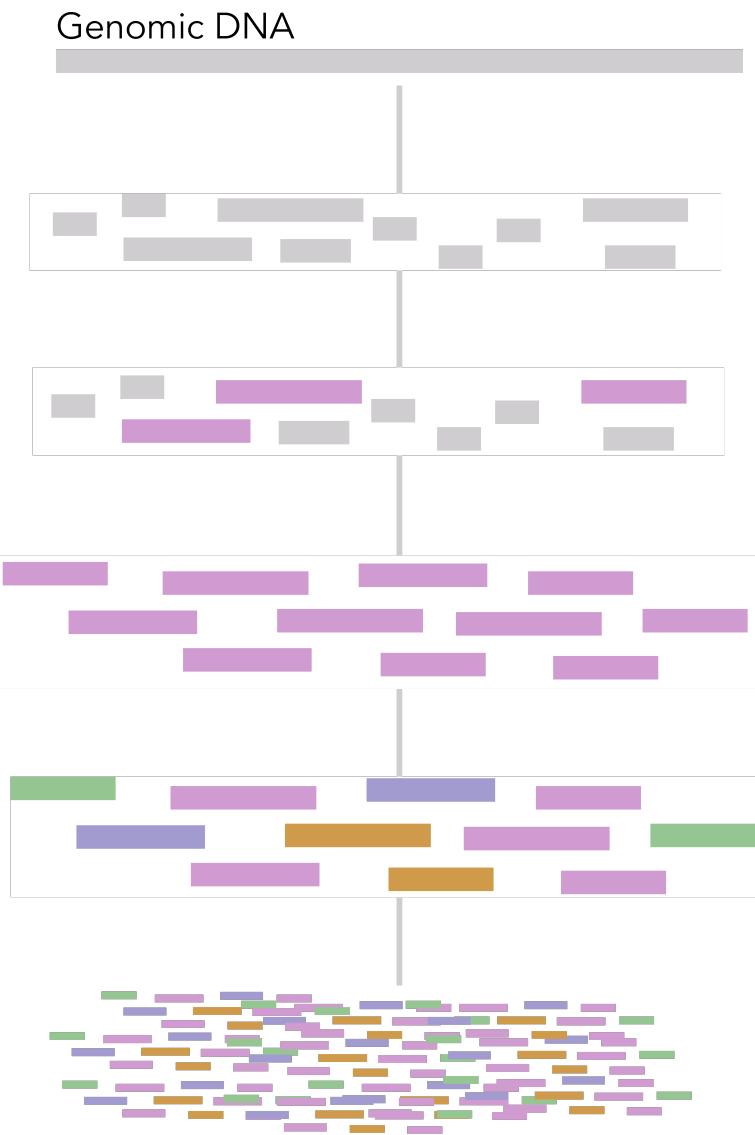
- Takes place at the end of nearly every step



You can actually do some size selection with beads!



## (2) Library preparation: break down into problems



Cut up genome into bits (only ~2% genome)

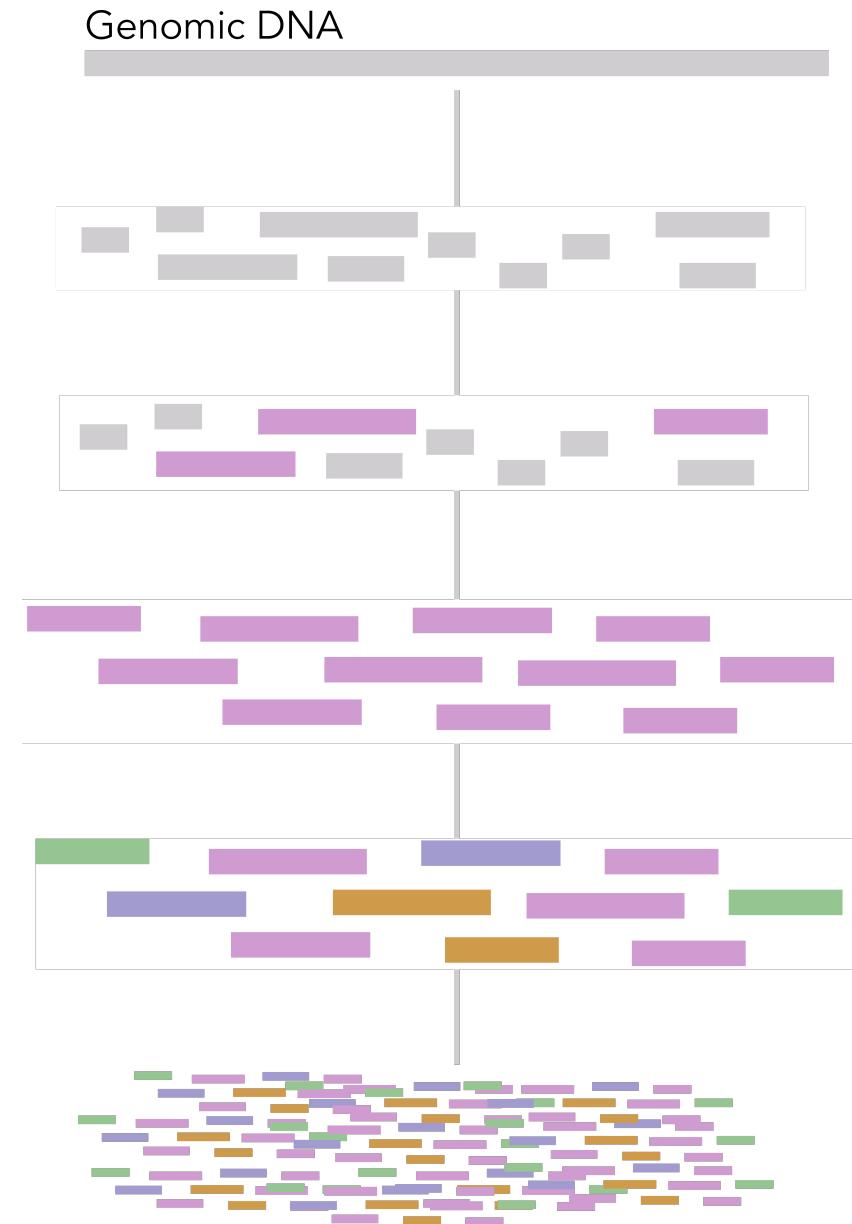
Only want fragments of a certain length

Need unique identifiers for hundreds of samples

Sequencer needs to recognize fragments and assign to individuals

## (2) Library preparation

- (a) Anneal adaptors
- (b) Double digest
- (c) Adaptor ligation
- (d) Pooling
- (e) Size selection
- (f) Streptavidin bead removal
- (g) PCR



## (2) Library prep: annealing adaptors

- Single-stranded oligos must be annealed with their partner
- P1 oligo pairs are unique (48); P2 oligos are universal

Adaptor P1  
(oligos 1.1 & 1.2)

5' ACACCTTTCCCTACACGACGCTTCCGATCT **aatta**  
|||  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA

Adaptor P2  
(oligos 2.1 & 2.2)

CGAGATCGGAAGAGCGAGAACAA 3'  
|||  
TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 5'

**Barcode** on adaptor P1  
helps tell samples apart  
within our bioinformatics  
pipeline

## (2) Library prep: **double digestion**

- We've already gone over how to do this in Bioanalyzer section
- Bead cleanup after digestion
- Quantify DNA with Qubit after

For simplicity's sake, our **fragments** are only a few bp long



These regions are called “**restriction overhangs**”

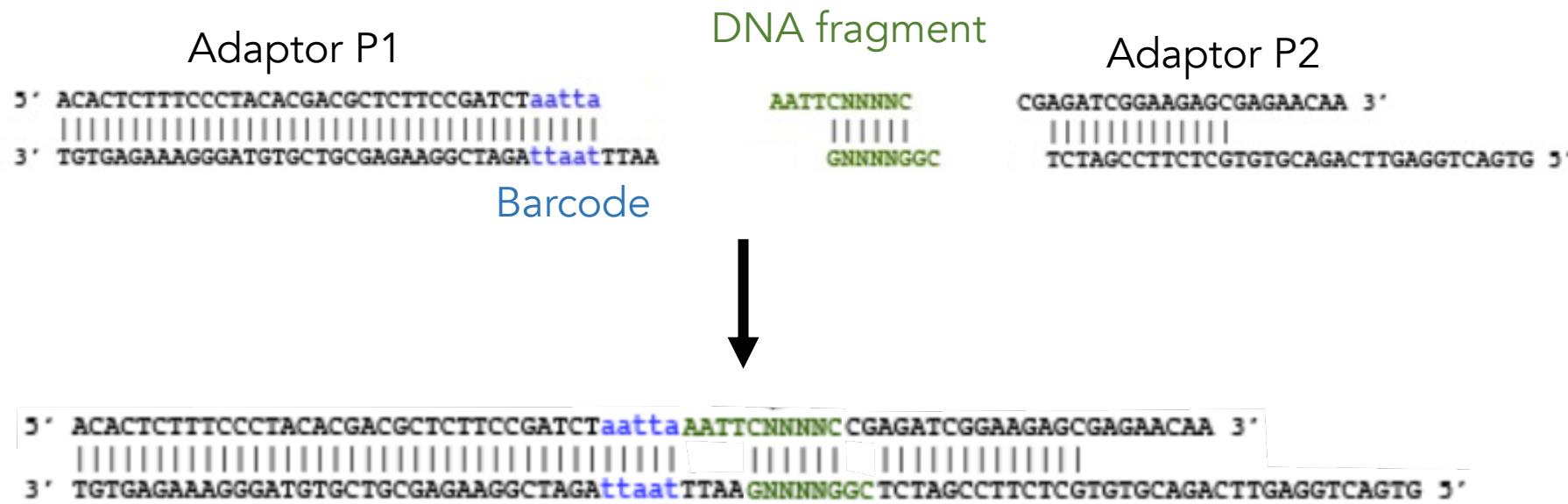
## (2) Library prep: adaptor ligation

- We need to stick the adaptors onto our DNA fragments
- Ratio of adaptor to DNA is *very important*

Input appropriate values in green shaded fields			
Initial DNA mass (ug)	0.05	mass of double-digested genomic DNA in ligation	
	P1	P2	
cut frequency (bp)	3500	1800	average distance between sites for each enzyme (from genome sequence or single digest fragment distribution)
fragment mass (g/mole)	2310000	1188000	
sample mass (g)	5.00E-08	5.00E-08	
fragments/sample (moles)	2.16E-14	4.21E-14	
ends/sample (moles)	4.33E-14	8.42E-14	
ends/sample (pmoles)	0.04	0.08	
target adapter fold excess	5	5	molar excess of adapter to genomic DNA ends
target adapter/sample (pmoles)	0.22	0.42	
target adapter volume/rxn (ul)	2	2	desired amount of working adapter stock to add to ligation rxn
adapter working conc (pmol/ul)	0.11	0.21	
annealed adapter conc (pmol/ul)	4	4	concentration of annealed adapter stock
fold dilution to working stock	36.96	19.01	
volume of working stock to make (ul)	200	200	desired volume of working adapter stock
annealed adapter stock (ul)	5.41	10.52	
1x annealing buffer (ul)	194.59	189.48	
working stock to add to ligation (ul)	2	2	

## (2) Library prep: adaptor ligation

- We need to stick the adaptors onto our DNA fragments



## (2) Library prep: pooling of samples

- Because all your samples are individually barcoded, you can combine them all together into a single tube
- Another bead cleanup after pooling samples

Sample 1

5' AACTCTTCCCTACAGACGCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 5'

Sample 3

5' AACTCTTCCCTACAGACGCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 5'

Sample 2

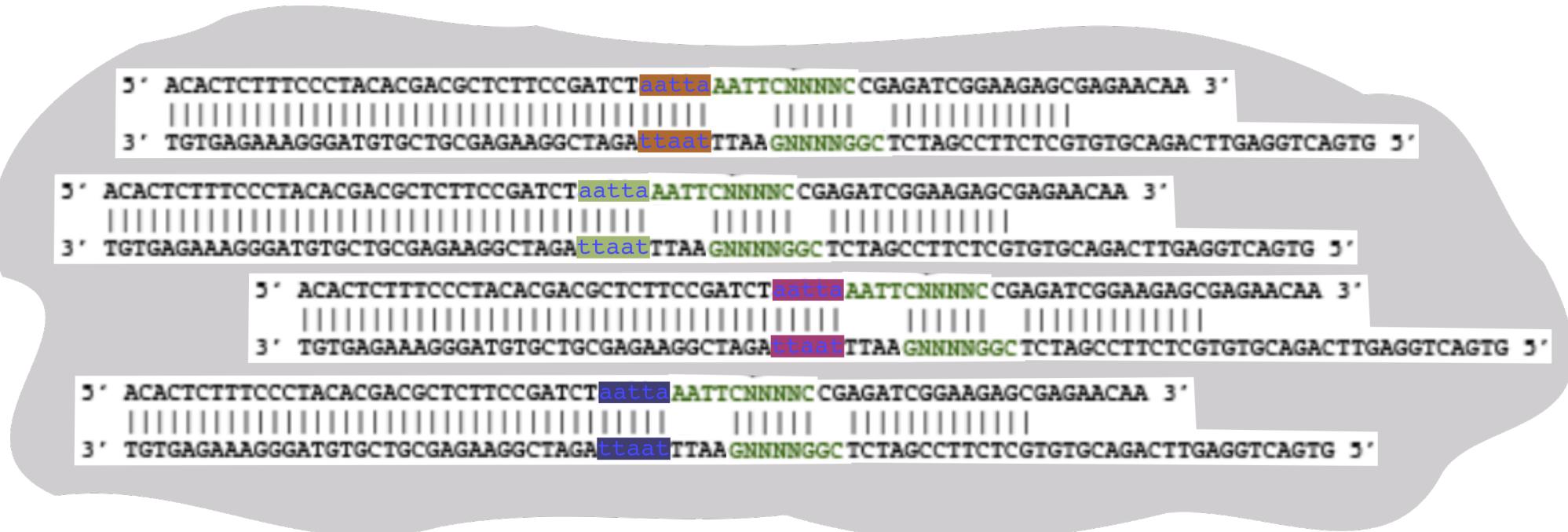
5' AACTCTTCCCTACAGACGCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 5'

Sample 4

5' AACTCTTCCCTACAGACGCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 5'

## (2) Library prep: **pooling of samples**

- Because all your samples are individually barcoded, you can combine them all together into a single tube
- Another bead cleanup after pooling samples



5' ACACCTTTCCCTACACGACGCTCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 3'

5' ACACCTTTCCCTACACGACGCTCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 3'

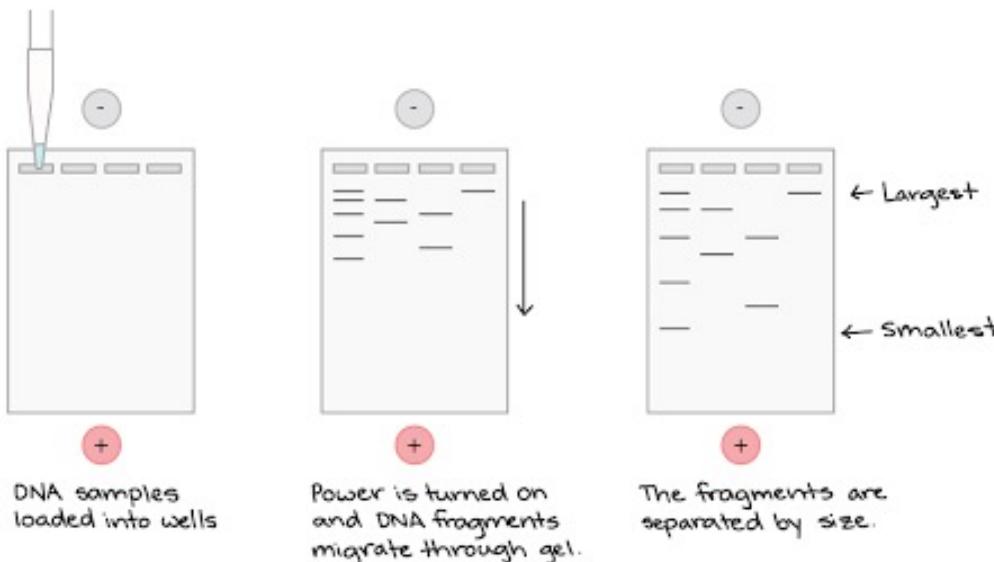
5' ACACCTTTCCCTACACGACGCTCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 3'

5' ACACCTTTCCCTACACGACGCTCTTCCGATCT **aatta** AATTNNNNCGAGATCGGAAGAGCGAGAACAA 3'  
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA **ttaat** TTAA GNNNNNGC TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG 3'

## (2) Library prep: size selection

We do this stage using a specialized tool: Pippin Prep

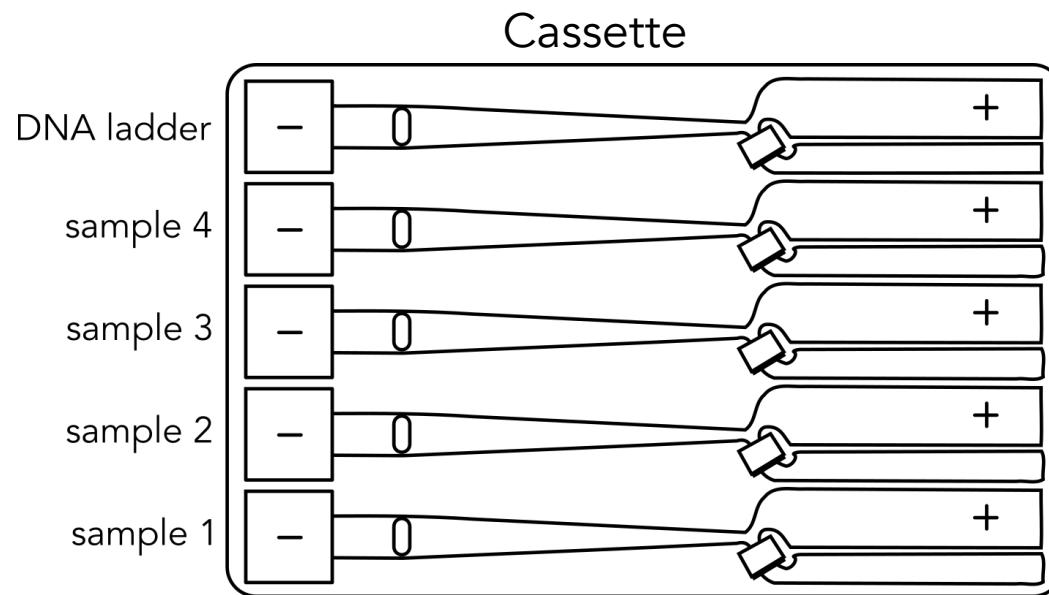
Normal gel electrophoresis



## (2) Library prep: size selection

We do this stage using a specialized tool: Pippin Prep

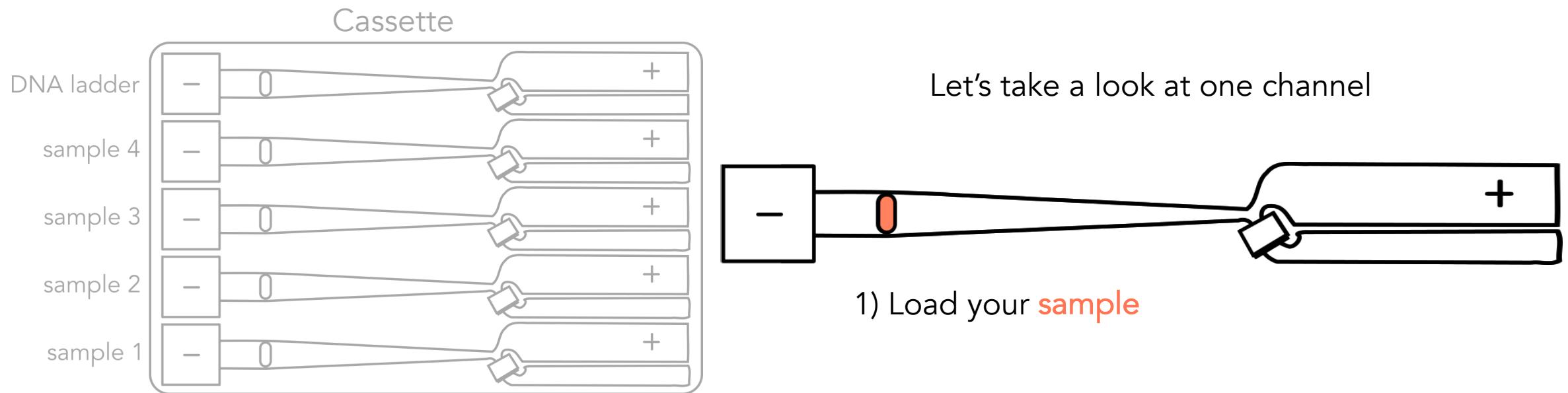
Same underlying principles with the Pippin Prep



## (2) Library prep: size selection

We do this stage using a specialized tool: Pippin Prep

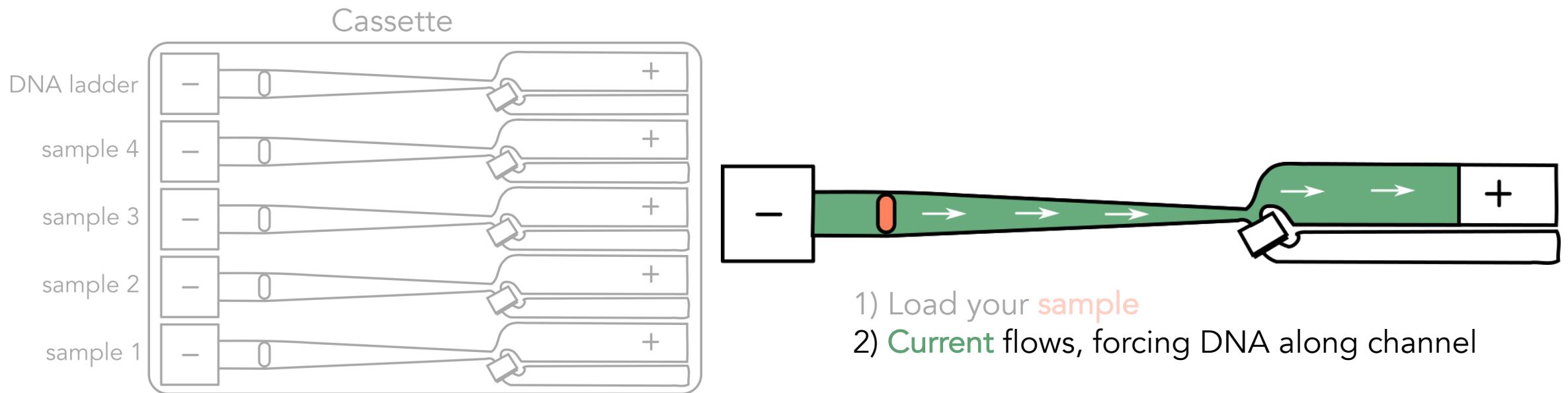
Same underlying principles with the Pippin Prep



## (2) Library prep: size selection

We do this stage using a specialized tool: Pippin Prep

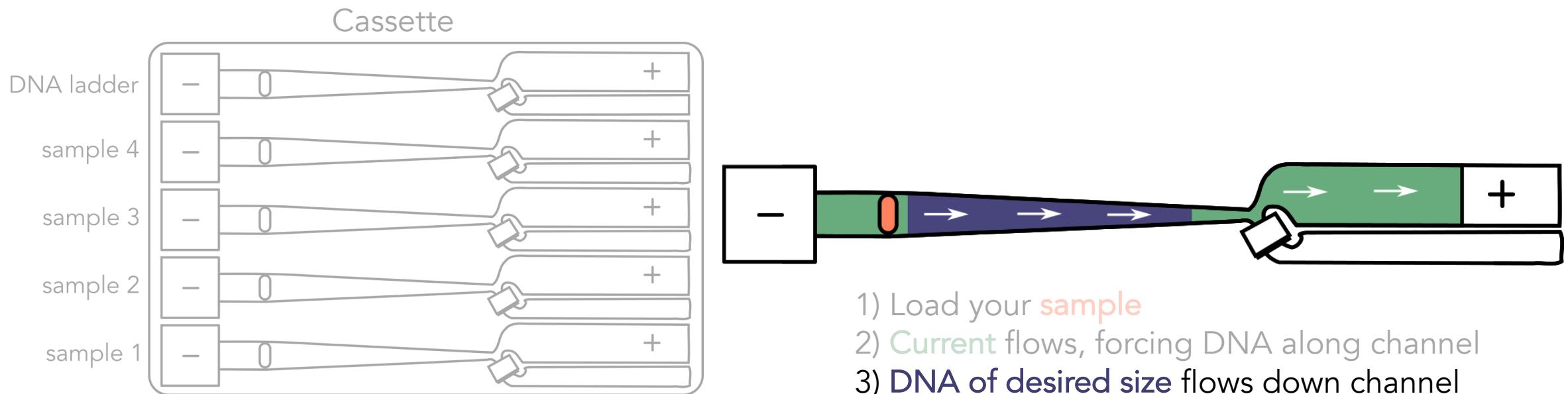
Same underlying principles with the Pippin Prep



## (2) Library prep: size selection

We do this stage using a specialized tool: Pippin Prep

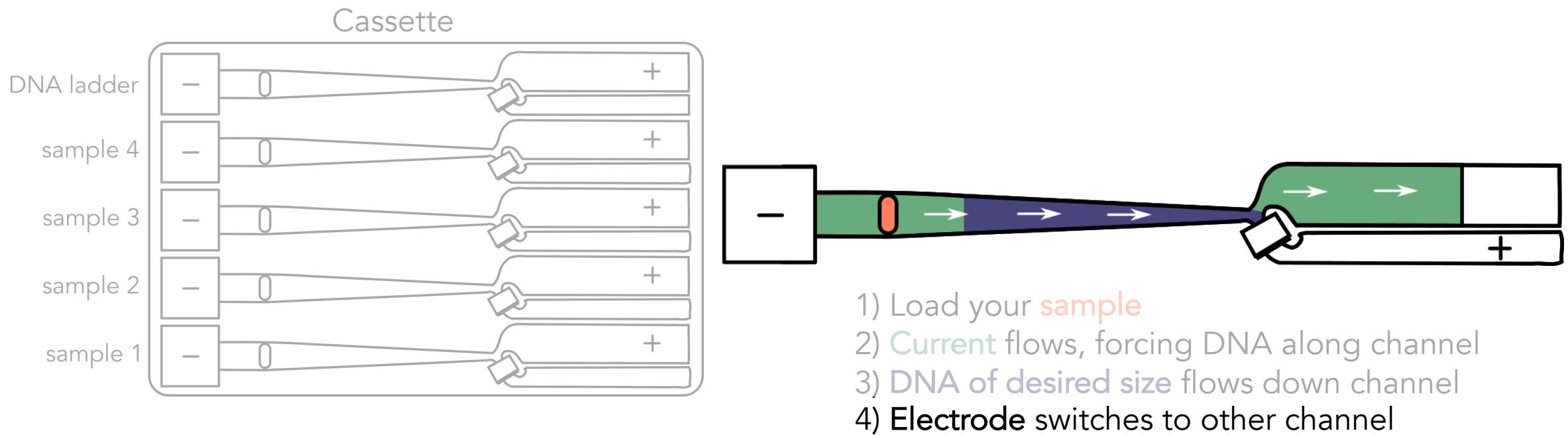
Same underlying principles with the Pippin Prep



## (2) Library prep: size selection

We do this stage using a specialized tool: Pippin Prep

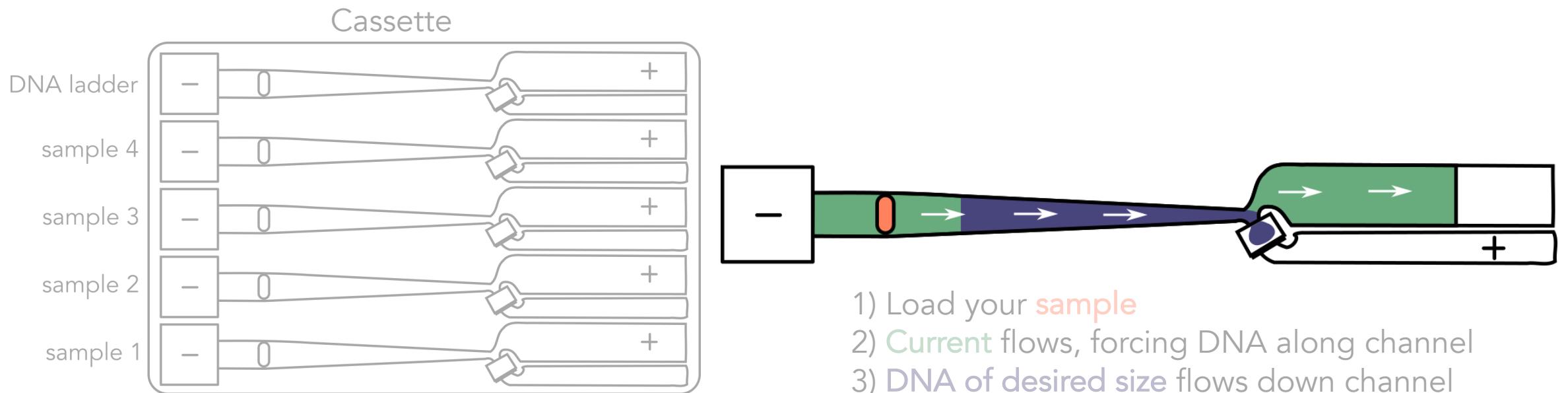
Same underlying principles with the Pippin Prep



## (2) Library prep: size selection

We do this stage using a specialized tool: Pippin Prep

Same underlying principles with the Pippin Prep



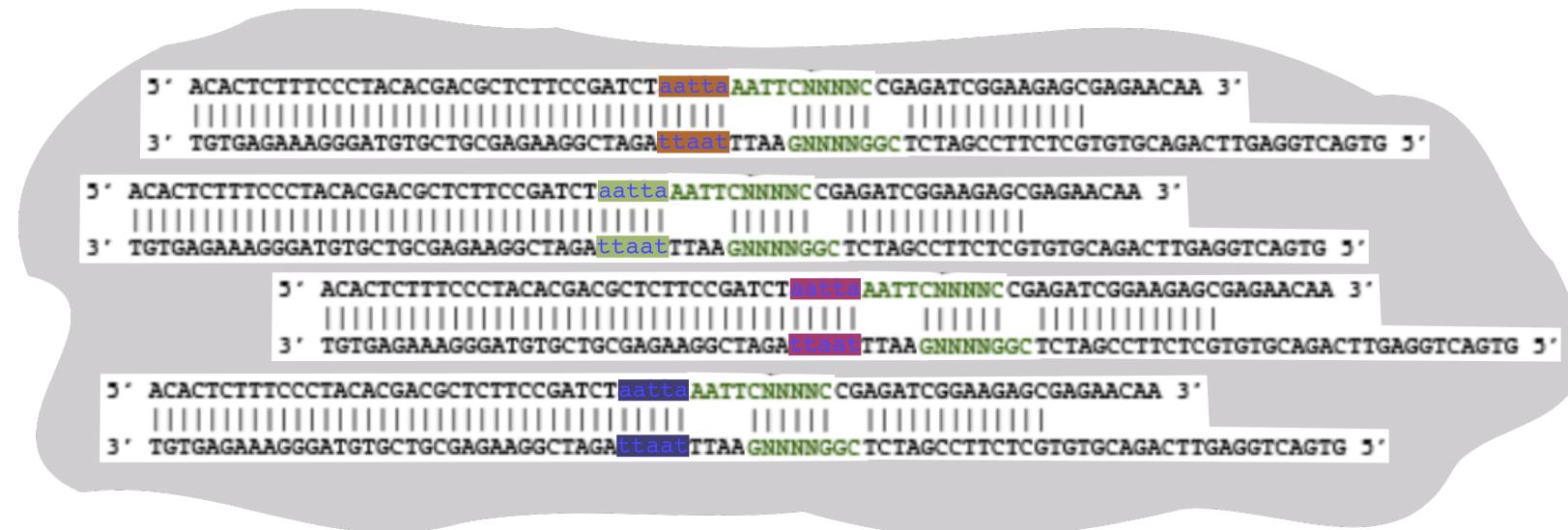
- 1) Load your **sample**
- 2) **Current** flows, forcing DNA along channel
- 3) DNA of desired size flows down channel
- 4) Electrode switches to other channel
- 5) **DNA of desired size caught in collection chamber**

## (2) Library prep: **Streptavidin bead removal**

- Can use special beads (Streptavidin) to remove excess adaptor P1

## (2) Library prep: PCR (and multiplexing)

- You may have noticed that we only have 48 unique adaptor P1s
- What if we have more than 48 samples?



## (2) Library prep: PCR (and multiplexing)

- You may have noticed that we only have 48 unique adaptor P1s
- We can add yet another identifier sequence to our fragments!

PCR primer 1



PCR multiplex primer 2

PCR2 index (1 of 12)

48 adaptor P1 barcodes

## (2) Library prep: PCR (and multiplexing)

- You may have noticed that we only have 48 unique adaptor P1s
- We can add yet another identifier sequence to our fragments!

PCR primer 1



PCR multiplex primer 2

PCR2 index (1 of 12)

48 adaptor P1 barcodes x 12 PCR2 indices

## (2) Library prep: PCR (and multiplexing)

- You may have noticed that we only have 48 unique adaptor P1s
- We can add yet another identifier sequence to our fragments!

PCR primer 1



PCR multiplex primer 2

PCR2 index (1 of 12)

48 adaptor P1 barcodes  $\times$  12 PCR2 indices = 576 individuals uniquely!

# After library prep, what are we left with?

Annealing adaptors

Double digestion



Adaptor ligation

AACACTTTCCCTACACGACGCTCTTCCGATCTaattaAATTCCNNNNCGAGATCGGAAGAGCACACGTCTGAACCTCCAGTCAC  
TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGAttaatTTAAAGNNNNNGCCTCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG

Size selection and pooling

PCR and multiplexing

5' AATGATA CGGC GACC ACCGAG ATCTAC ACTCTT CCCTACACGACGCTCTTCCGATCTaattaAATTCCNNNNCGAGATCGGAAGAGCACACGTCTGAACCTCCAGTCACatcacgATCTCGTATGCCGTCTCTGCTTG 3'  
3' TTACTATGCCGCTGGTGGCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGAttaatTTAAAGNNNNNGCCTCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTGtagtgcTAGAGCATACGGCAGAACAGAAC 5'

— Sequencer uses parts of this sequence as primers

Let's go back to our initial questions

# Annealing adaptors

## Double digestion

Cut up genome into bits (only ~2% genome)

## Adaptor ligation

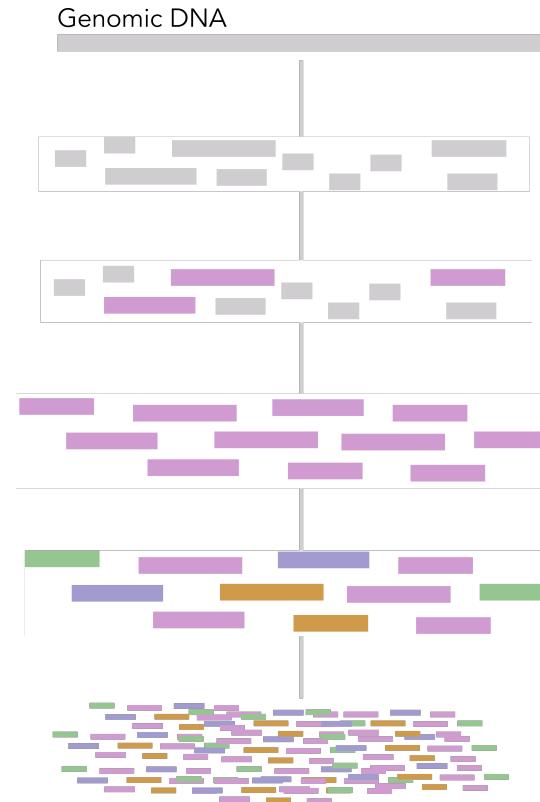
Only want fragments of a certain length

## Size selection and pooling

## PCR and multiplexing

Need unique identifiers for hundreds of samples

Sequencer needs to recognize fragments and assign to individuals



# Let's go back to our initial questions

Annealing adaptors

Cut up genome into bits (only ~2% genome)  
**Double digestion**

Adaptor ligation

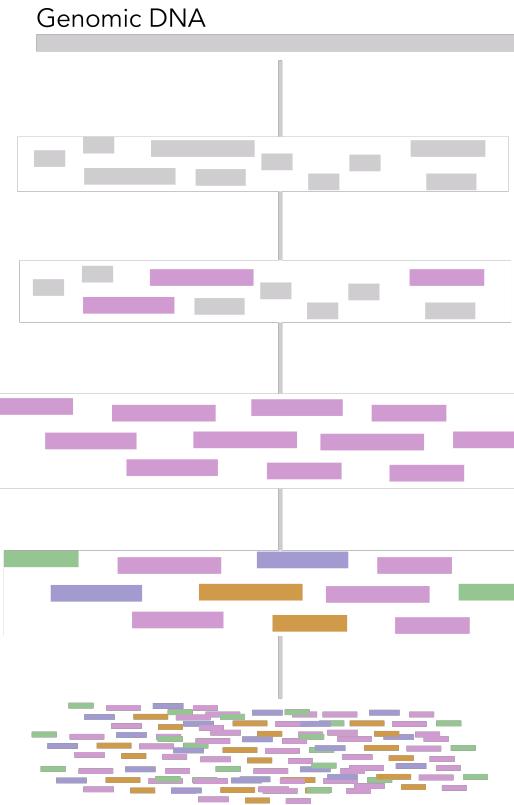
Only want fragments of a certain length

Size selection and pooling

PCR and multiplexing

Need unique identifiers for hundreds of samples

Sequencer needs to recognize fragments and assign to individuals



# Let's go back to our initial questions

Annealing adaptors

Cut up genome into bits (only ~2% genome)  
**Double digestion**

Adaptor ligation

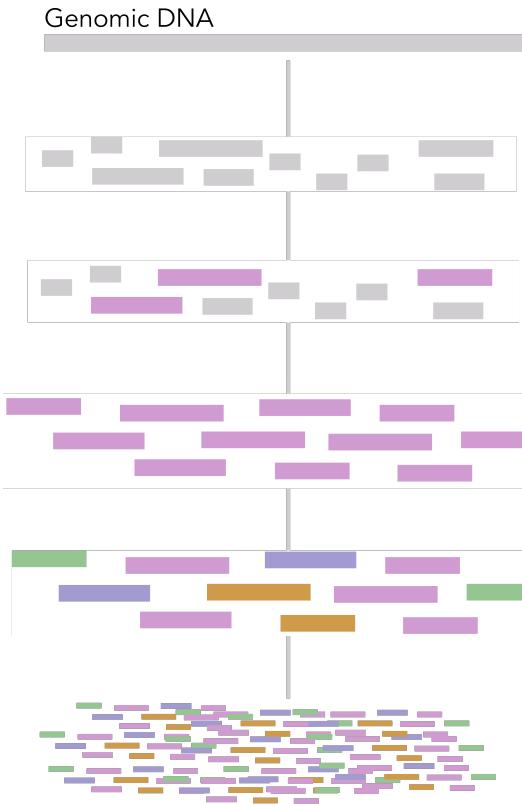
Only want fragments of a certain length  
**Size selection**

*pooling*

PCR and multiplexing

Need unique identifiers for hundreds of samples

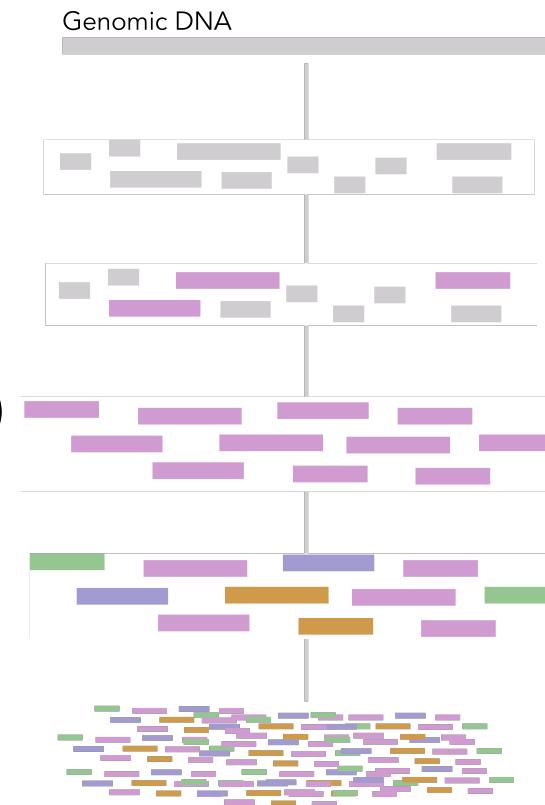
Sequencer needs to recognize fragments and assign to individuals



# Let's go back to our initial questions

Annealing adaptors

Cut up genome into bits (only ~2% genome)  
**Double digestion**



Only want fragments of a certain length  
**Size selection**

*pooling*

multiplexing

Need unique identifiers for hundreds of samples  
**Adaptor ligation & PCR**

Sequencer needs to recognize fragments and assign to individuals

# Let's go back to our initial questions

Annealing adaptors

Cut up genome into bits (only ~2% genome)  
**Double digestion**



Only want fragments of a certain length  
**Size selection**

*pooling*

multiplexing

Need unique identifiers for hundreds of samples  
**Adaptor ligation & PCR**

Sequencer needs to recognize fragments and assign to individuals  
**Adaptor ligation & PCR**

# How to know what to sequence

Genome size x Proportion of genome recovered by enzyme = Total # bases

$$\frac{\text{Total # bases}}{\text{Average fragment size}} = \text{Total number of loci}$$

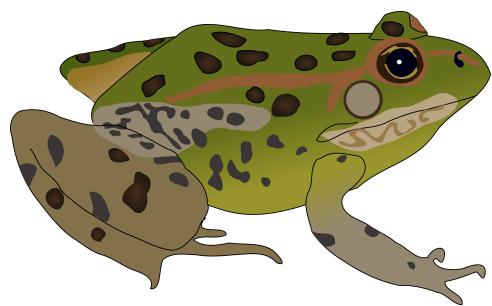
Total number of loci x depth = Requested # reads

# How to know what to sequence

6 Gb x

2%

= 120M bases total



120M bases = 436,363 loci  
275 bp avg.

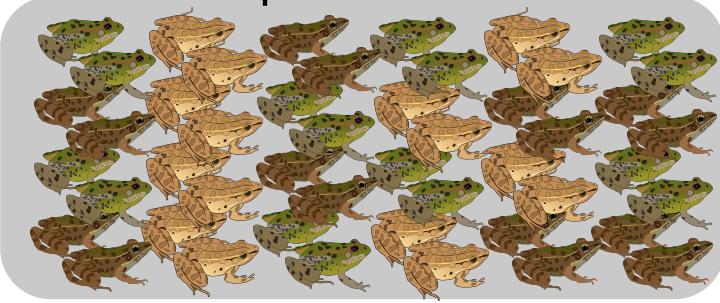
436,363 loci x 12 = 5.23M reads requested (per sample)

**Region table for sample 3 :** T6857 SM 1:3 BA only

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Co lor
200	250	10.2	5	217	5.5	18.61	130.2	
225	275	4.5	2	244	6.7	7.99	49.8	
250	300	3.2	2	275	4.6	5.54	30.6	

# How to know what to sequence

125 samples



5.23M reads requested (per sample)

**5.23M x 125 samples = 654M reads**

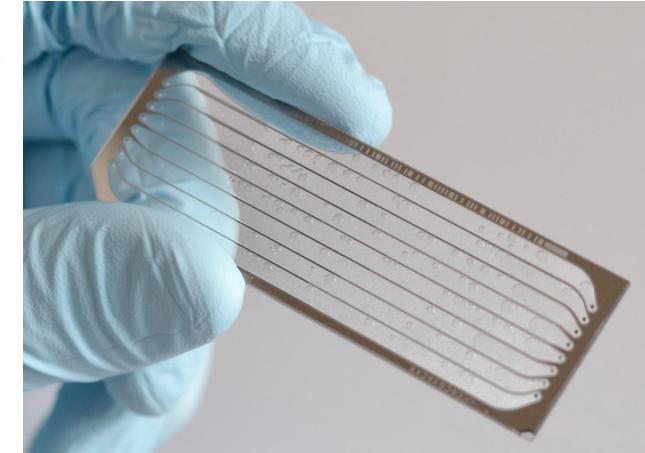
Sequencing facilities may allow you to choose sequencing amount based on one of two things:

**Run Types Available for purchasing only your desired number of reads per sample**

Platform	Run Type	Cost (based on min reads purchased)	External Cost	Reads (min. required per sample)	Minimum Reads Required per Project
NovaSeq S1	PE150	\$120.34 (20M)	\$172.93	2.00E+07	1.00E+08
NovaSeq S1	SR100	\$79.63 (20M)	\$121.43	2.00E+07	1.00E+08
NovaSeq SP	PE150	\$176.78 (20M)	\$223.63	2.00E+07	1.00E+08
NextSeq 500	PE75	\$71.25 (5M)	\$90.14	5.00E+06	4.00E+07
NextSeq 500	SR75	\$45.07 (5M)	\$57.02	5.00E+06	4.00E+07
MiSeq V2	PE250	\$163.25 (1M)	\$206.51	1.00E+06	1.00E+06
MiSeq V3	PE300	\$117.19 (1M)	\$148.25	1.00E+06	1.00E+06

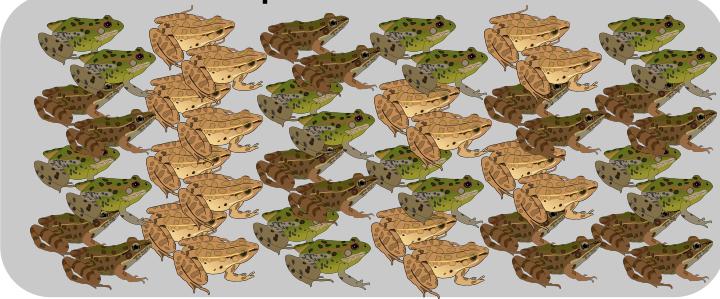
**Run Type Available for purchasing Full Lanes**

Platform	Run Type	Cost	External Cost	Total Reads Produced per Lane
NovaSeq S1 LANE	PE150	\$4,784.50	\$6,052.39	7.00E+08
NovaSeq S1 LANE	SR100	\$3,359.50	\$4,249.77	7.00E+08
NovaSeq SP LANE	PE150	\$3252.57	\$4114.50	3.50E+08



# How to know what to sequence

125 samples



5.23M reads requested (per sample)

**5.23M x 125 samples = 654M reads**

Sequencing facilities may allow you to choose sequencing amount based on one of two things:

Run Types Available for purchasing only your desired number of reads per sample

Platform	Run Type	Cost (based on min reads purchased)	External Cost	Reads (min. required per sample)	Minimum Reads Required per Project
NovaSeq S1	PE150	\$120.34 (20M)	\$172.93	2.00E+07	1.00E+08
NovaSeq S1	SR100	\$79.63 (20M)	\$121.43	2.00E+07	1.00E+08
NovaSeq SP	PE150	\$176.78 (20M)	\$223.63	2.00E+07	1.00E+08
NextSeq 500	PE75	\$71.25 (5M)	\$90.14	5.00E+06	4.00E+07
NextSeq 500	SR75	\$45.07 (5M)	\$57.02	5.00E+06	4.00E+07
MiSeq V2	PE250	\$163.25 (1M)	\$206.51	1.00E+06	1.00E+06
MiSeq V3	PE300	\$117.19 (1M)	\$148.25	1.00E+06	1.00E+06

Request 654M total reads = \$9,300

Request 2 lanes (~700M reads) = \$6,505

Run Type Available for purchasing Full Lanes

Platform	Run Type	Cost	External Cost	Total Reads Produced per Lane
NovaSeq S1 LANE	PE150	\$4,784.50	\$6,052.39	7.00E+08
NovaSeq S1 LANE	SR100	\$3,359.50	\$4,249.77	7.00E+08
NovaSeq SP LANE	PE150	\$3252.57	\$4114.50	3.50E+08