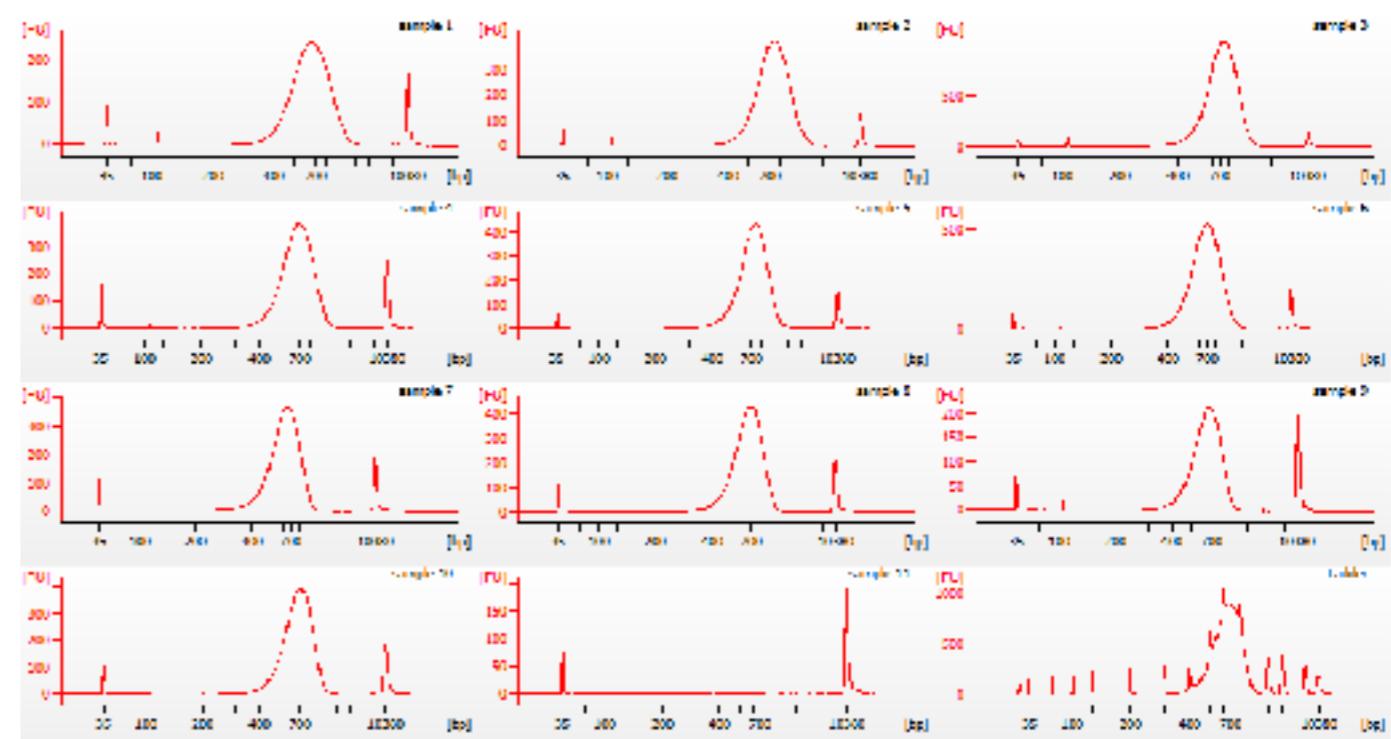
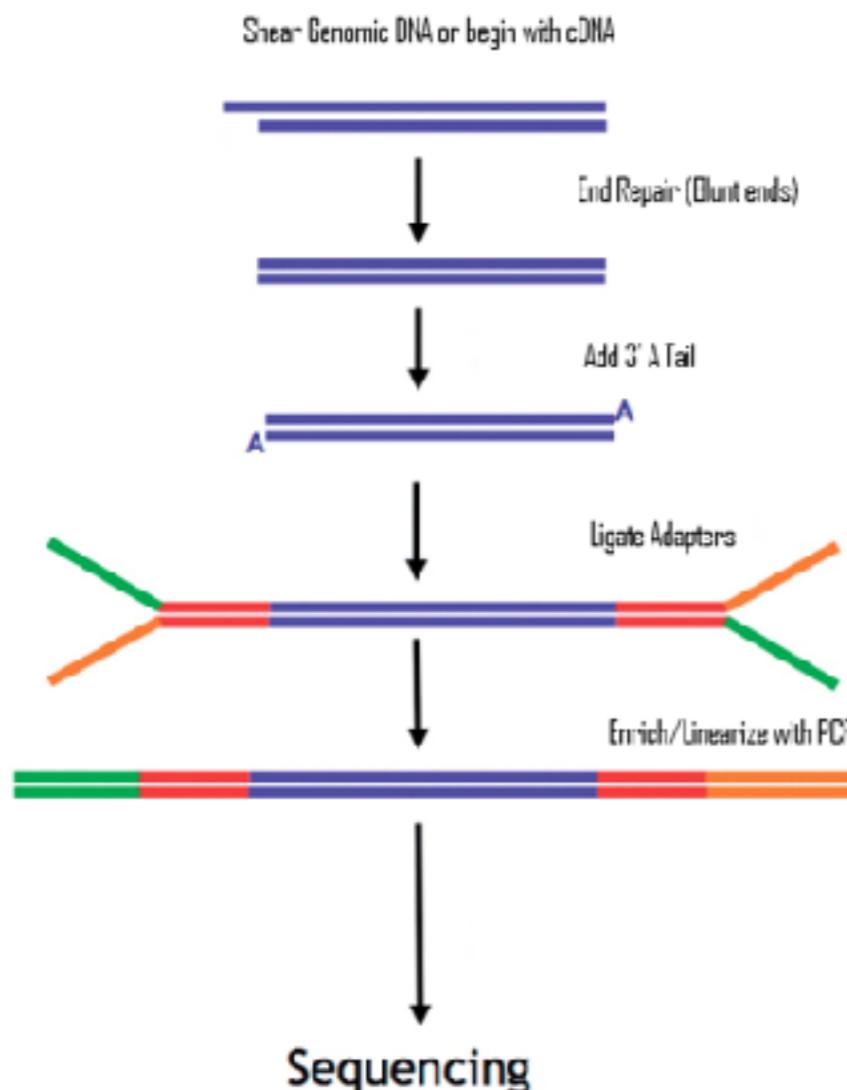


GENOMIC LIBRARY PRODUCTION AND QUALITY CONTROL



Elliot Gardner
PhD Candidate
Plant Biology and Conservation
Northwestern University and Chicago Botanic Garden

OBJECTIVES

1. Describe the process of generating TruSeq-style libraries for Illumina sequencing.
2. Understand potential cost saving measure such as reduced-volume reactions, homebrewing reagents such as SPRI beads, and custom index sequences to increase multiplexing.
3. Evaluate a sample and decide on an appropriate library preparation strategy for it.
4. Weight the relative advantages of the major commercially available library preparation kits and determine the most cost-effective kit for a given HybSeq study.
5. Develop and apply appropriate protocol modifications for preparing libraries from degraded herbarium DNA.
6. Describe how SPRI beads work.
7. Conduct and interpret appropriate library quality control.

ROAD MAP

1. Workflow overview (Illumina TruSeq Nano)
2. Recommendations for working with degraded specimens
3. Other kits and indexing options
4. Quality control

WORKFLOW OVERVIEW

A. Sonication

(Bead cleanup)

B. End repair



TruSeq® Nano DNA Library Prep Kit

C. Size-selection

(Bead cleanup)

D. A-tailing

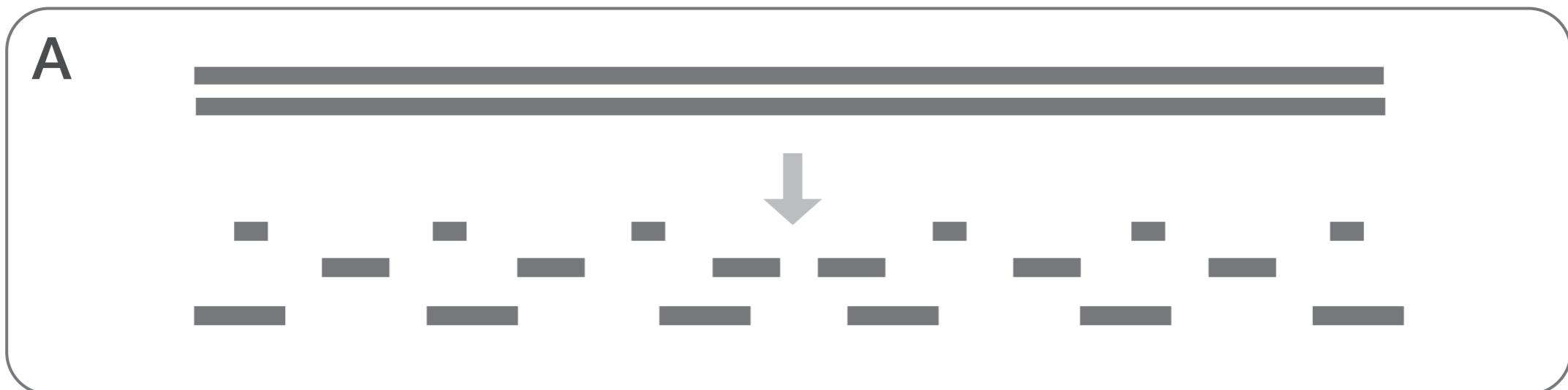
E. Adapter ligation

(Bead cleanup x 2)

F. PCR amplification

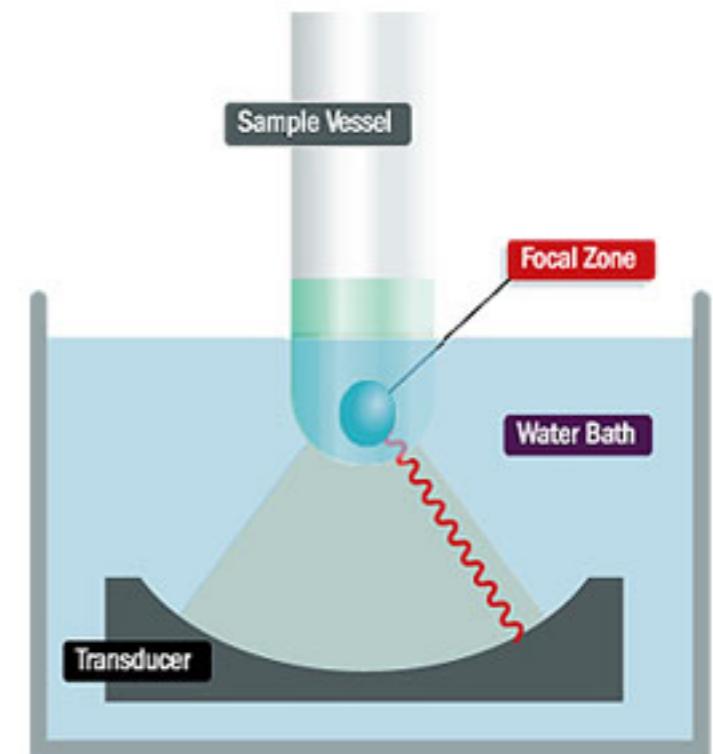
(Bead cleanup)

SONICATION



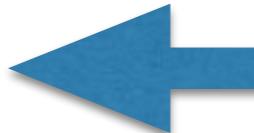
Library construction begins with genomic DNA that is subsequently fragmented.

Covaris®



SONICATION - SAMPLE VOLUME

Illumina recommends 50 µL
25 µL also works fine



microTUBE AFA Fiber Pre-Slit
Snap-Cap 6x16mm (25)

£4 each, but reusable (unofficially):
Wash with:

1. 10% bleach
2. Sterile water
3. 100% ethanol

Covaris microTUBE-15 supports 15 µL



SONICATION

Table 1 350 bp Insert Settings

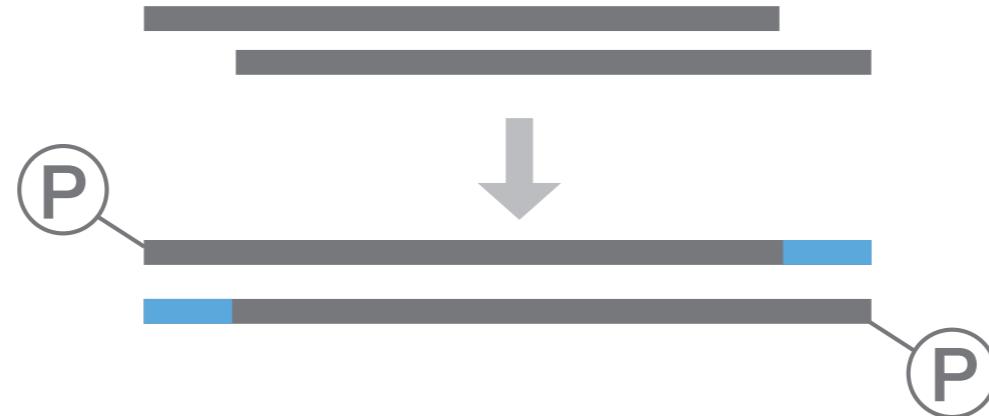
Covaris Setting	M220	S220	S2	E210
Duty Factor (%)	20	5	10	
Intensity	—	—	5.0	
Peak/Displayed Power (W)	50	175	23	14
Cycles/Burst		200		
Duration (seconds)	65	50	45	
Mode	—	Frequency sweeping		
Temperature (°C)	20		5.5–6	

Table 2 550 bp Insert Settings

Covaris Setting	M220	S220	S2	E210
Duty Factor (%)	20	5	10	
Intensity	—	—	2.0	
Peak/Displayed Power (W)	50	175	9	7
Cycles/Burst		200		
Duration (seconds)	45	25	45	
Mode	—	Frequency sweeping		
Temperature (°C)	20		5.5–6	

END REPAIR

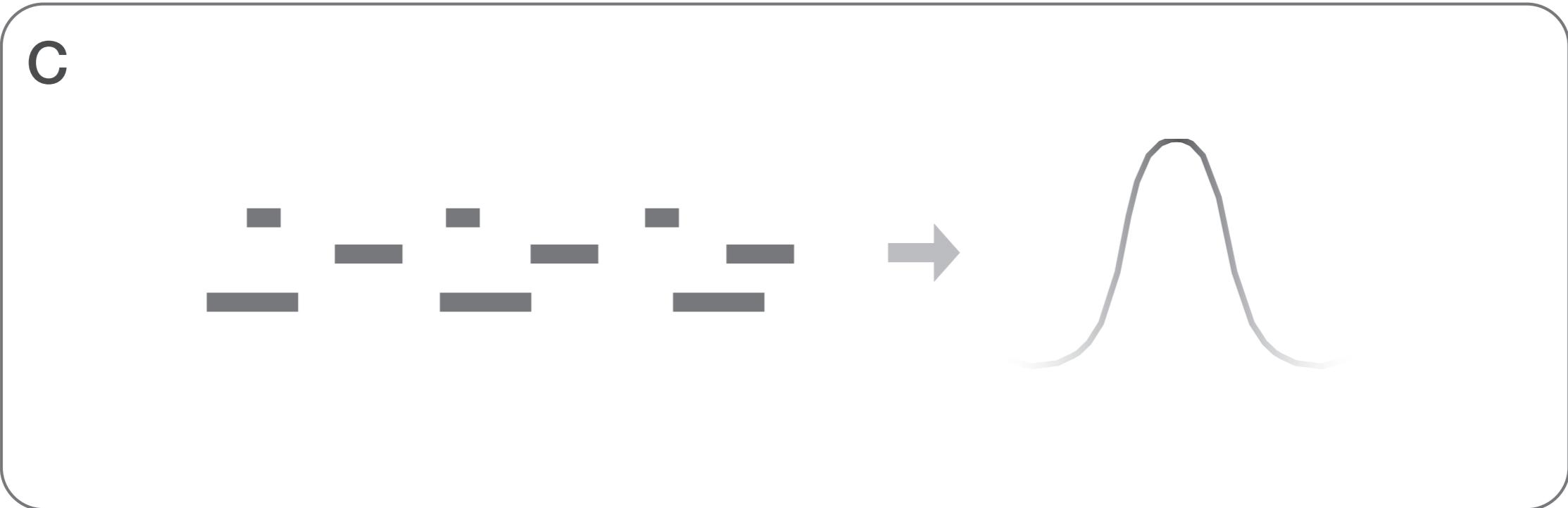
B



Blunt-end fragments are created.

Can be performed in 1/4 volume.

SIZE SELECTION



Fragments are narrowly size selected with sample purification beads.

Can be performed in 1/4 volume
but may be less accurate.

SIZE SELECTION



Left Side Size Selection Process Overview

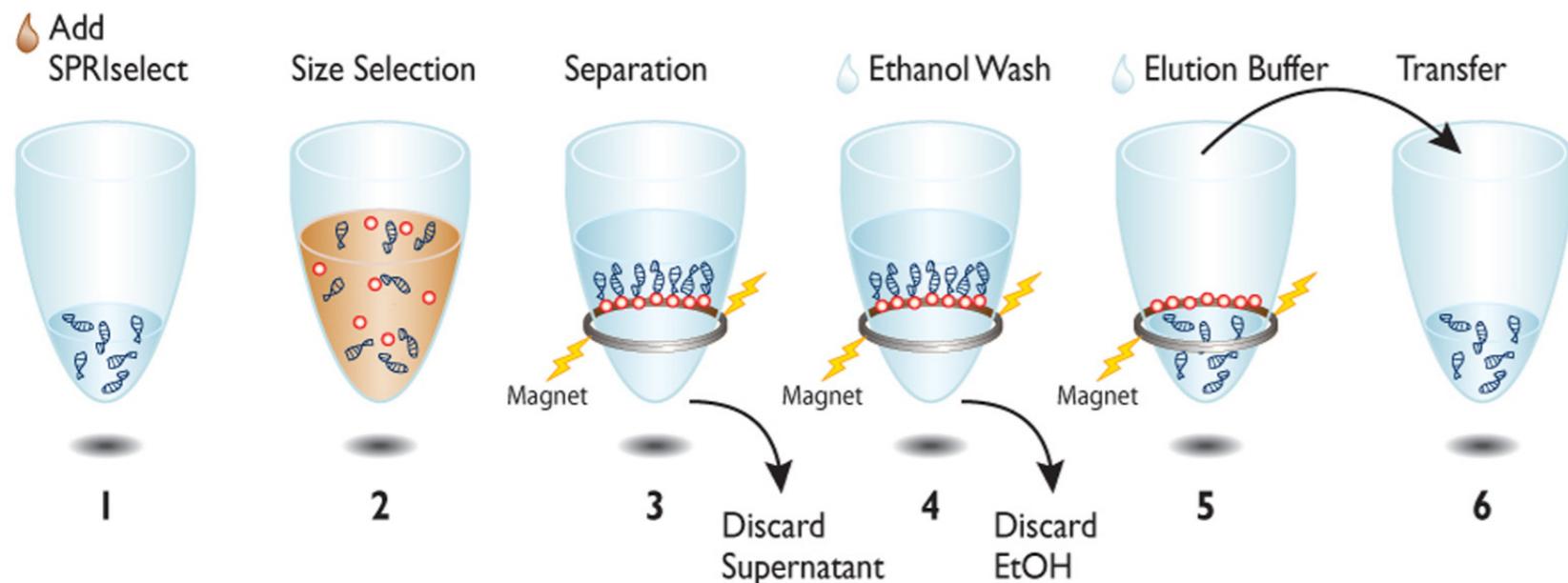
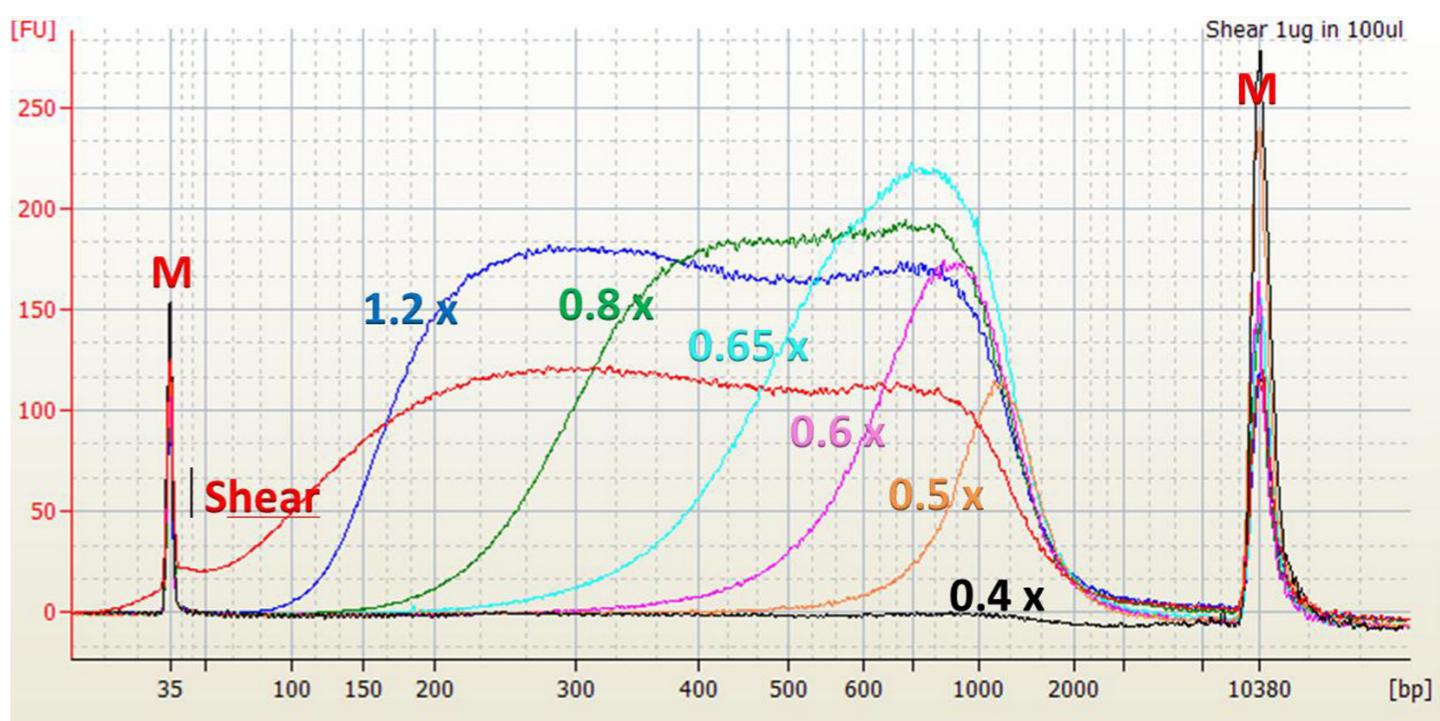
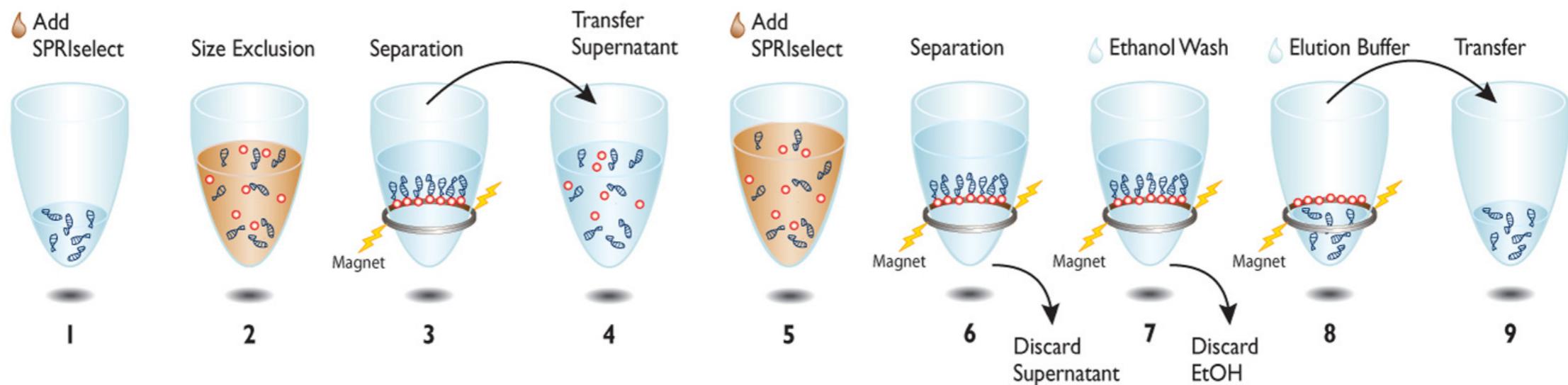


Figure 1 Agilent High Sensitivity DNA chip Electropherogram.



SIZE SELECTION

Right Side Size Selection Process Overview

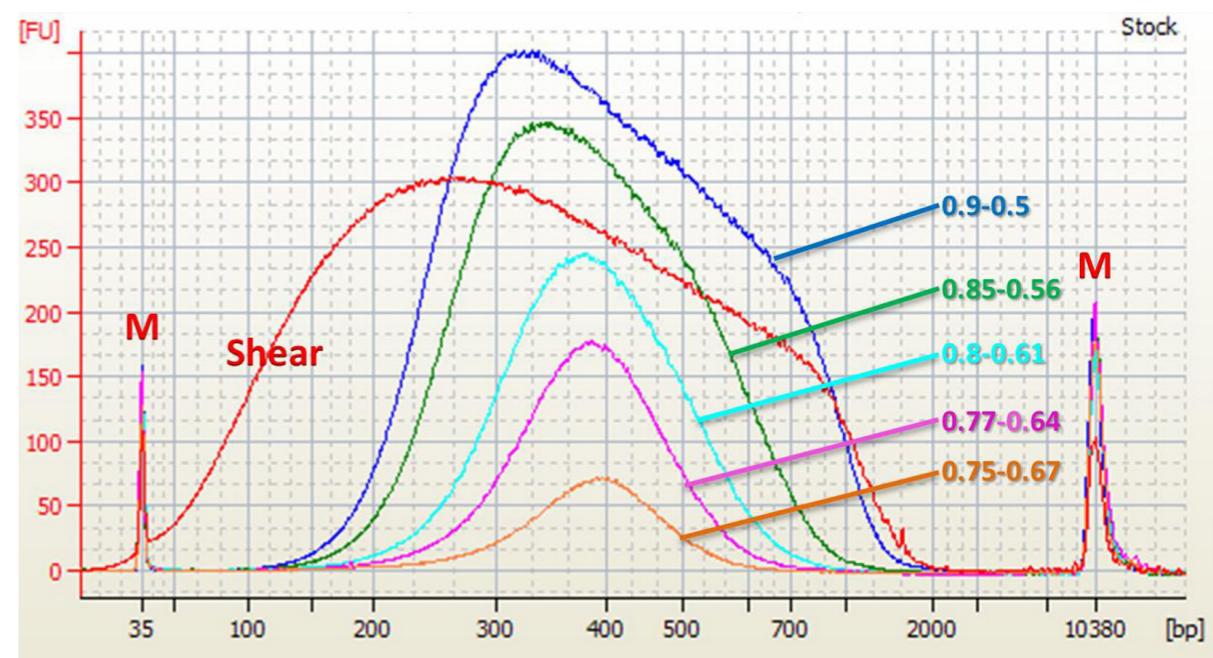
**Table 1** Typical Recovery (determined by experiment)

Ratios (Left-Right)	bp Region	Selection Delta (bp)	bp Region's % of Shear	Recovered % of bp Region	Recovered Region's % of Shear
Shear	40-3000	2960	100.0%	100.0%	100.0%
0.9-0.5	175-1300	1125	72.7%	60.4%	43.9%
0.85-0.56	200-700	500	61.8%	49.6%	30.6%
0.8-0.61	230-660	430	52.1%	33.4%	17.4%
0.77-0.64	260-575	315	40.8%	21.1%	8.6%
0.75-0.67	280-540	260	33.7%	10.1%	3.4%

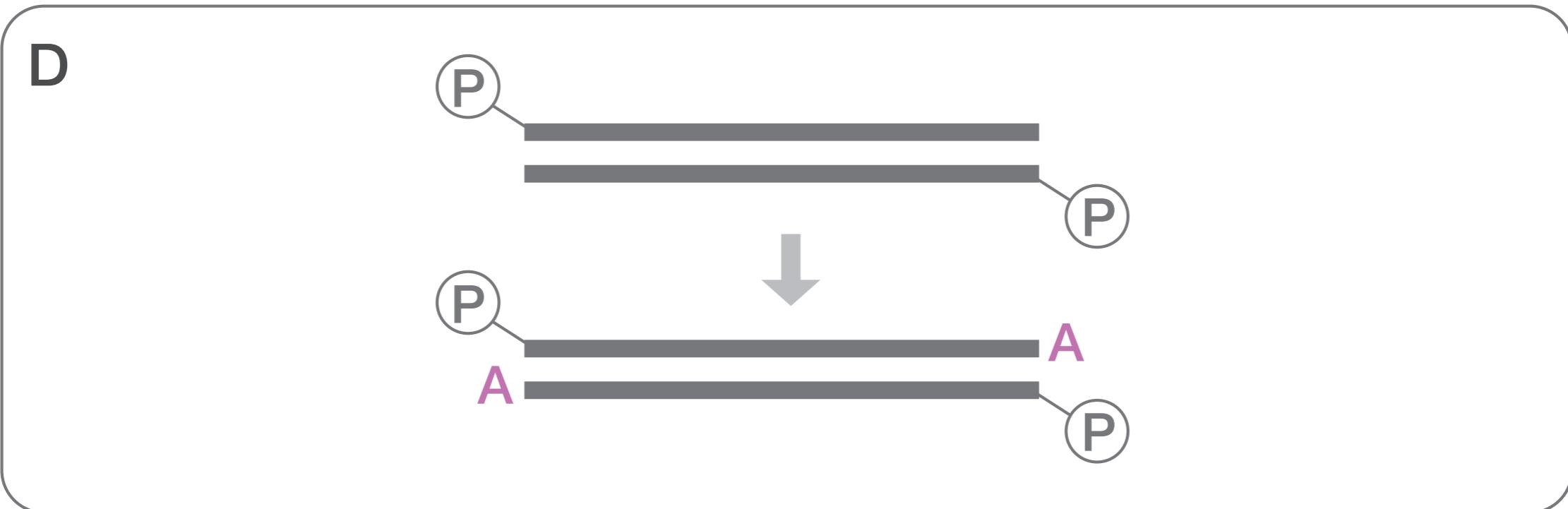
Percentages calculated from Agilent 2100 Expert Smear Region analysis Conc. [pg/ μ L]



B24965AA
October 2012

Figure 5 Agilent High Sensitivity DNA chip Electropherogram.

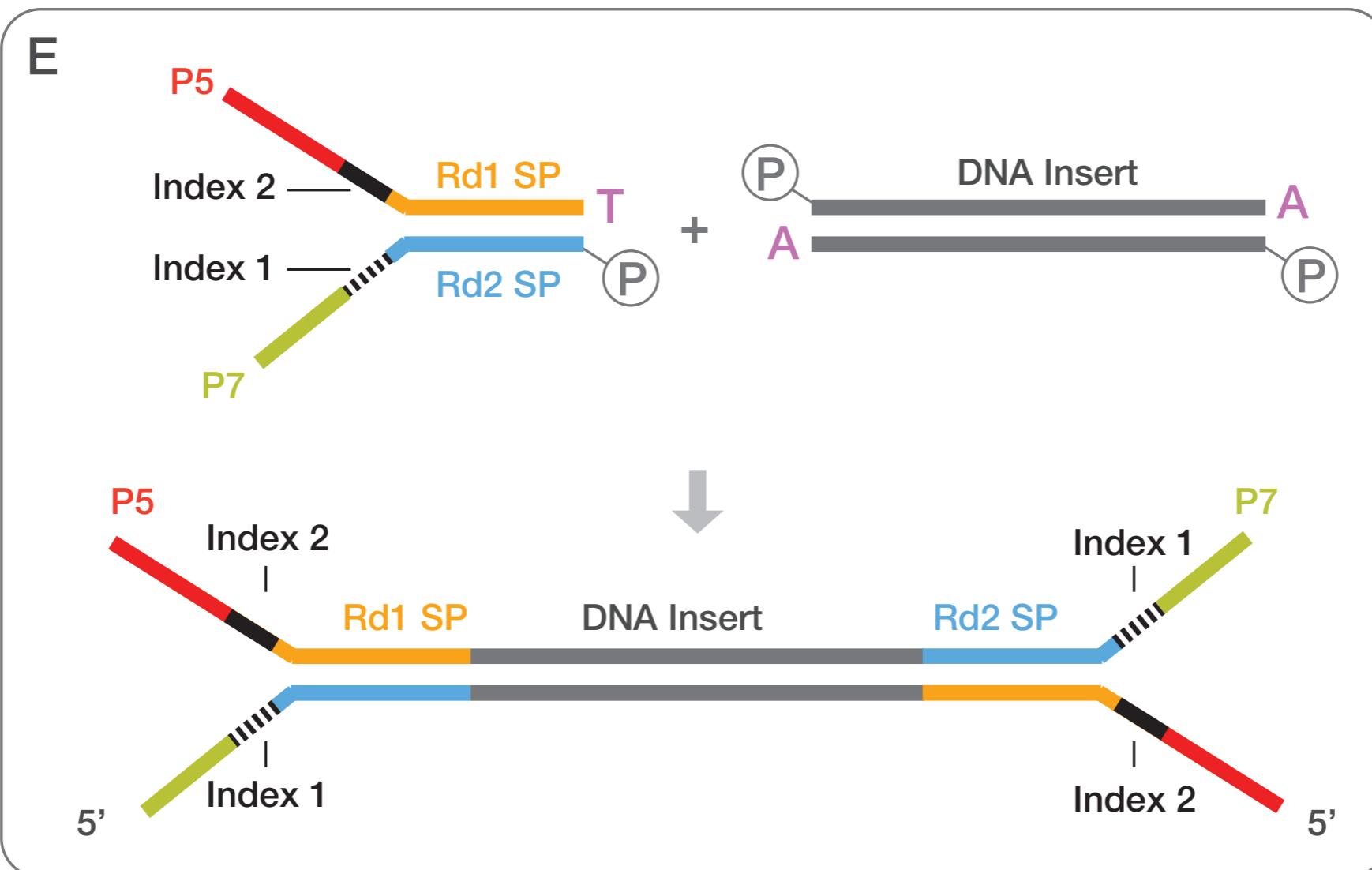
A-TAILING



A-base is added.

Can be performed in 1/4 volume.

ADAPTER LIGATION



Dual-index adapters are ligated to the fragments.

96 unique index (barcode) combinations
Can be performed in 1/4 volume.

PCR AMPLIFICATION



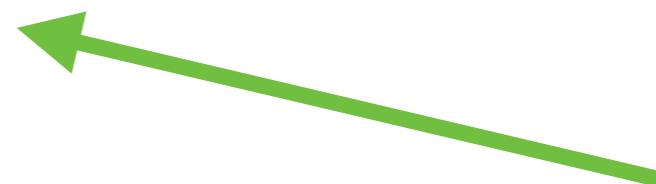
Ligated product is amplified and ready for cluster generation.

- **IMPORTANT:** use at most half of your adapter-ligated template on the first go so that you can do more PCR if necessary!
- Can be performed in 1/4 volume but at least 1/2 volume is preferable to ensure enough product.
- Default is 8 cycles, but up to 12 may be required, especially if prior steps were performed in reduced volumes.

DEGRADED SPECIMENS: MODIFICATIONS TO CONSIDER

A. ~~Sonication~~

(Bead cleanup)



skip

B. End repair

C. ~~Size-selection~~

(Bead cleanup)

Increase beads and add isopropanol to save small fragments

D. A-tailing

E. Adapter ligation

(Bead cleanup ~~x 2~~)

**Dilute adapter /
increase ligation time**

Reduce to one for low-input samples

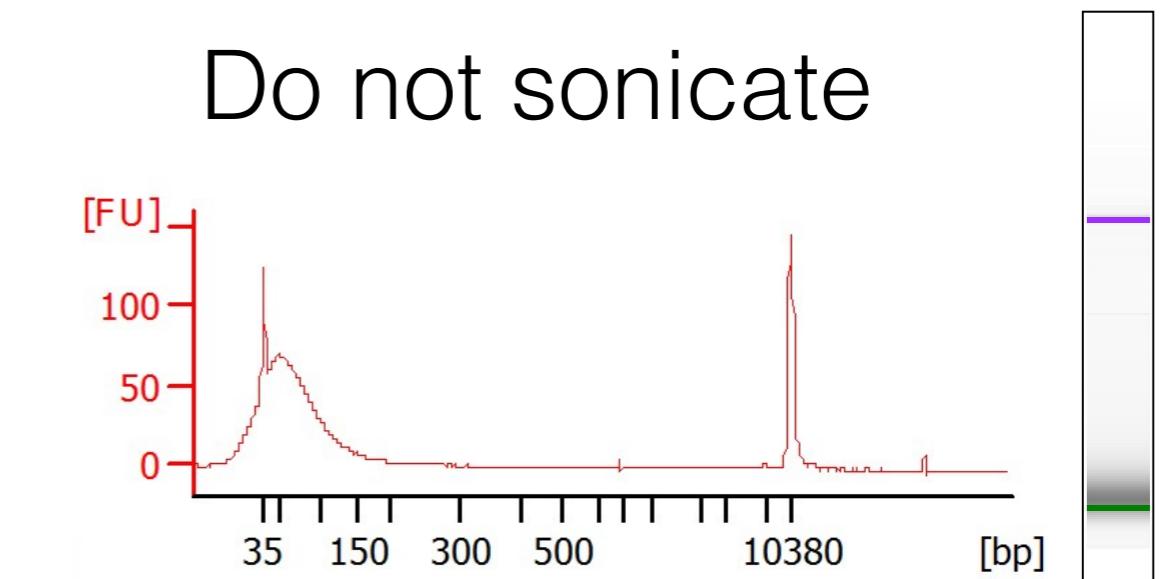
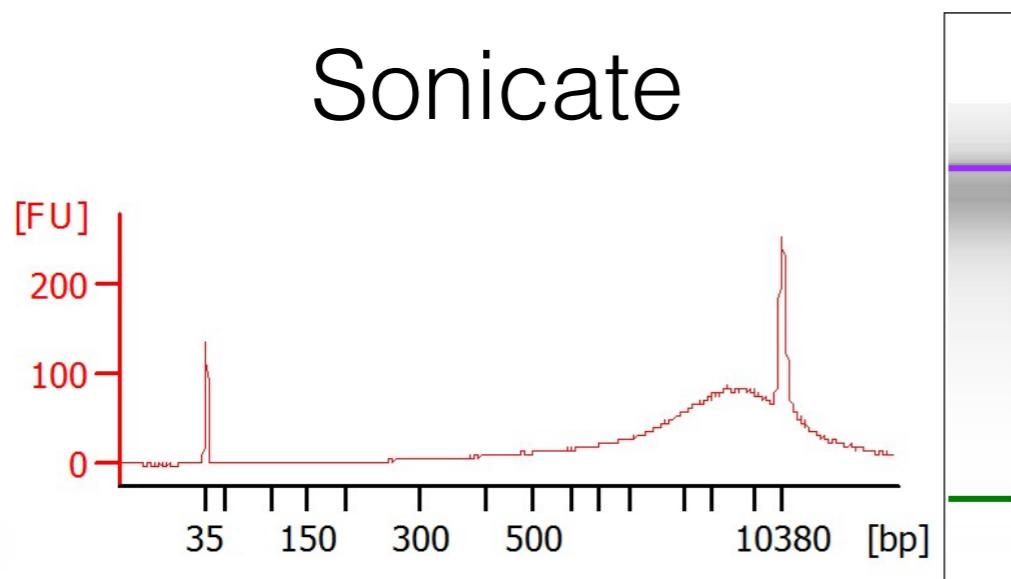
F. PCR amplication

(Bead cleanup)

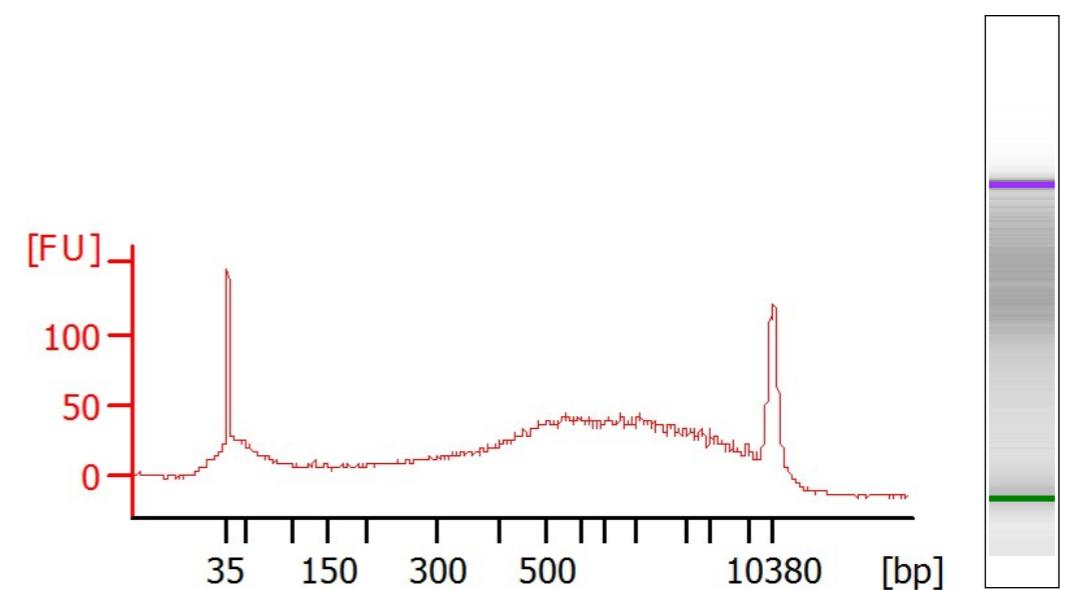
Increase cycles

Additional 0.7x cleanup to remove adapter dimer

SONICATION

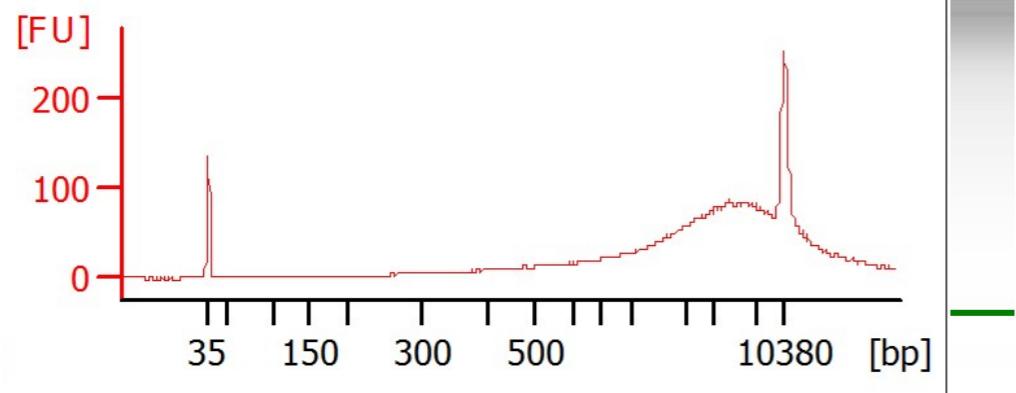


Would probably work either way (I would sonicate)

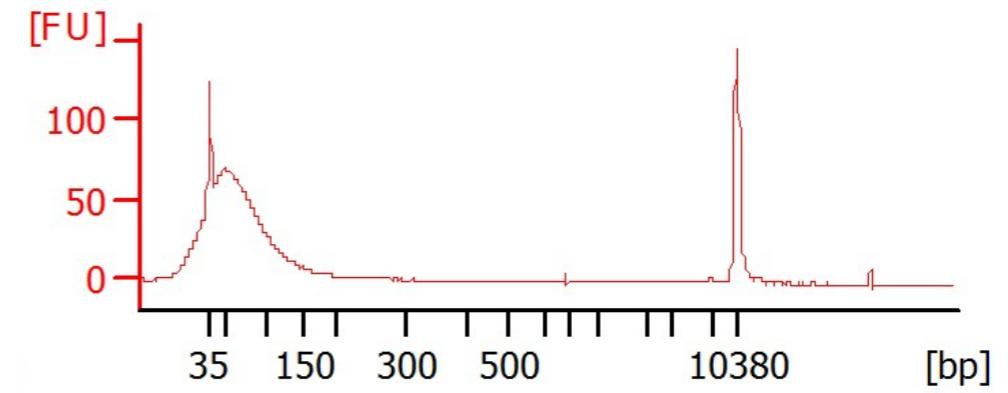


SIZE SELECTION

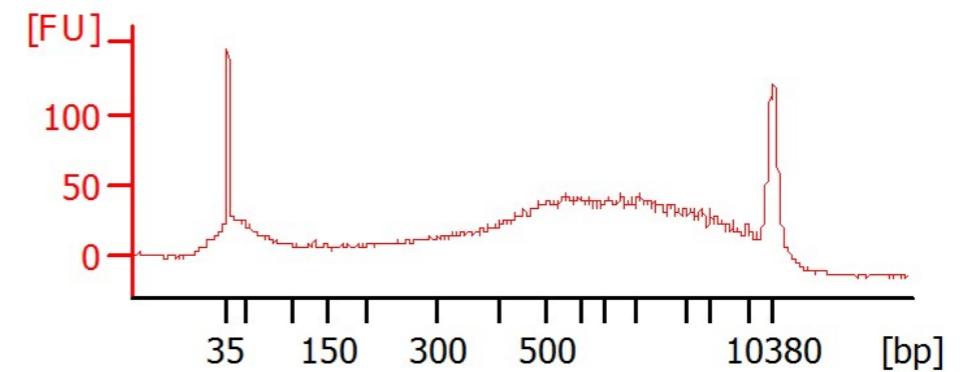
Size select



Do not size select



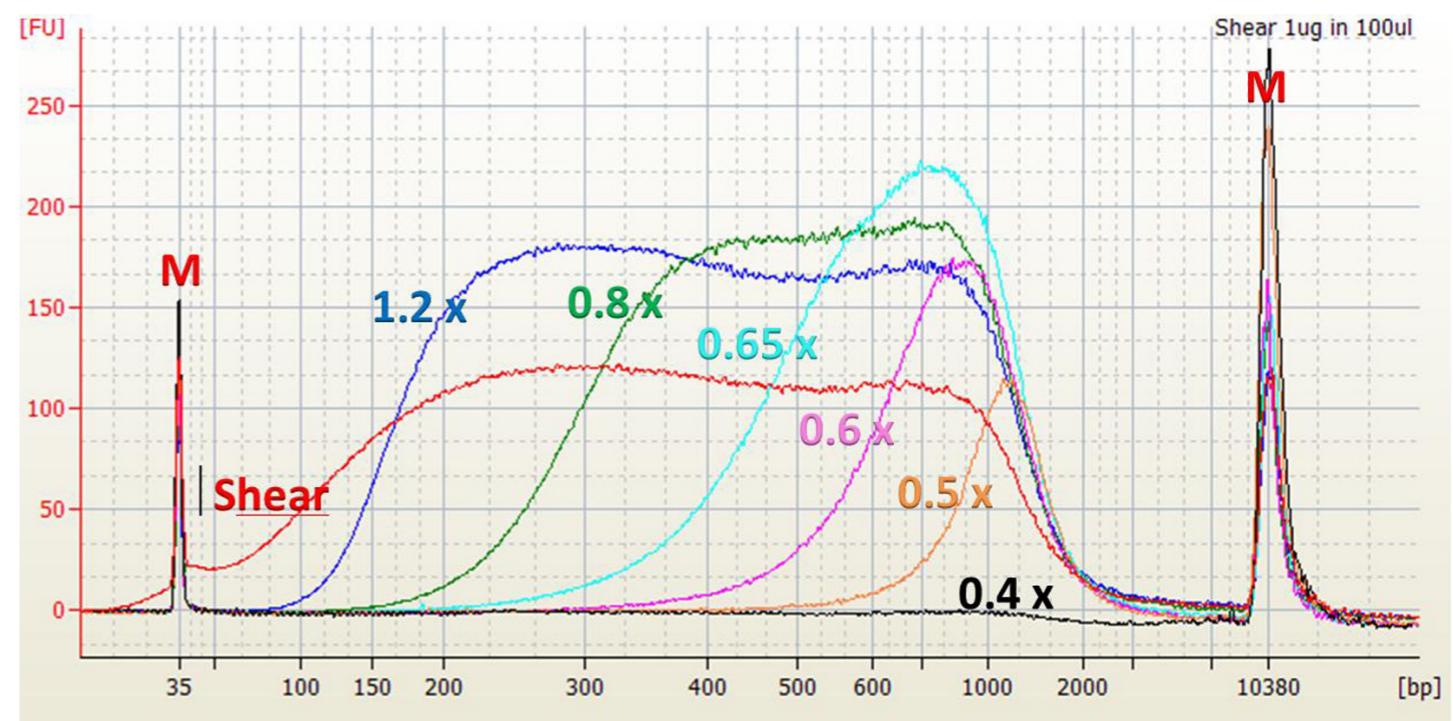
Size select



BEAD CLEANUP AFTER END REPAIR/SIZE SELECTION

To avoid losing small fragments, clean with 1.8x beads + 5.4x isopropanol.

Figure 1 Agilent High Sensitivity DNA chip Electropherogram.



SPRIselect User Guide

B24965AA
October 2012

ADAPTER LIGATION

Dilute adapter with resuspension buffer:

1:1 for 25ng DNA post-A-tailing

4:1 for 10ng DNA post-A-tailing

9:1 for 5ng DNA post-A-tailing



KAPA Hyper Prep Kit

Technical Data Sheet

Table 2. Recommended adapter concentrations for libraries constructed from 1 ng – 1 µg input DNA*

Fragmented DNA per 50 µL ER & AT reaction	Adapter stock concentration	Adapter:insert molar ratio	Fragmented DNA per 50 µL ER & AT reaction	Adapter stock concentration	Adapter:insert molar ratio
1 µg	15 µM	10:1	25 ng	7.5 µM	200:1
500 ng	15 µM	20:1	10 ng	3 µM	200:1
250 ng	15 µM	40:1	5 ng	1.5 µM	200:1
100 ng	15 µM	100:1	2.5 ng	750 nM	200:1
50 ng	15 µM	200:1	1 ng	300 nM	200:1

*Adapter:insert molar ratio calculations are based on a mode DNA fragment length of 200 bp, and will be higher for longer DNA fragments, or slightly lower for DNA fragmented to a mode size <200 bp. The lower adapter:insert molar ratios recommended for inputs >100 ng represent a fair compromise between library construction efficiency and cost; higher library yields will be achieved if a higher adapter concentration is used.

ADAPTER LIGATION

Consider increasing the ligation time and/or only a single cleanup post-adapter ligation for very low-input samples.

But note that both of these options may increase adapter dimer!



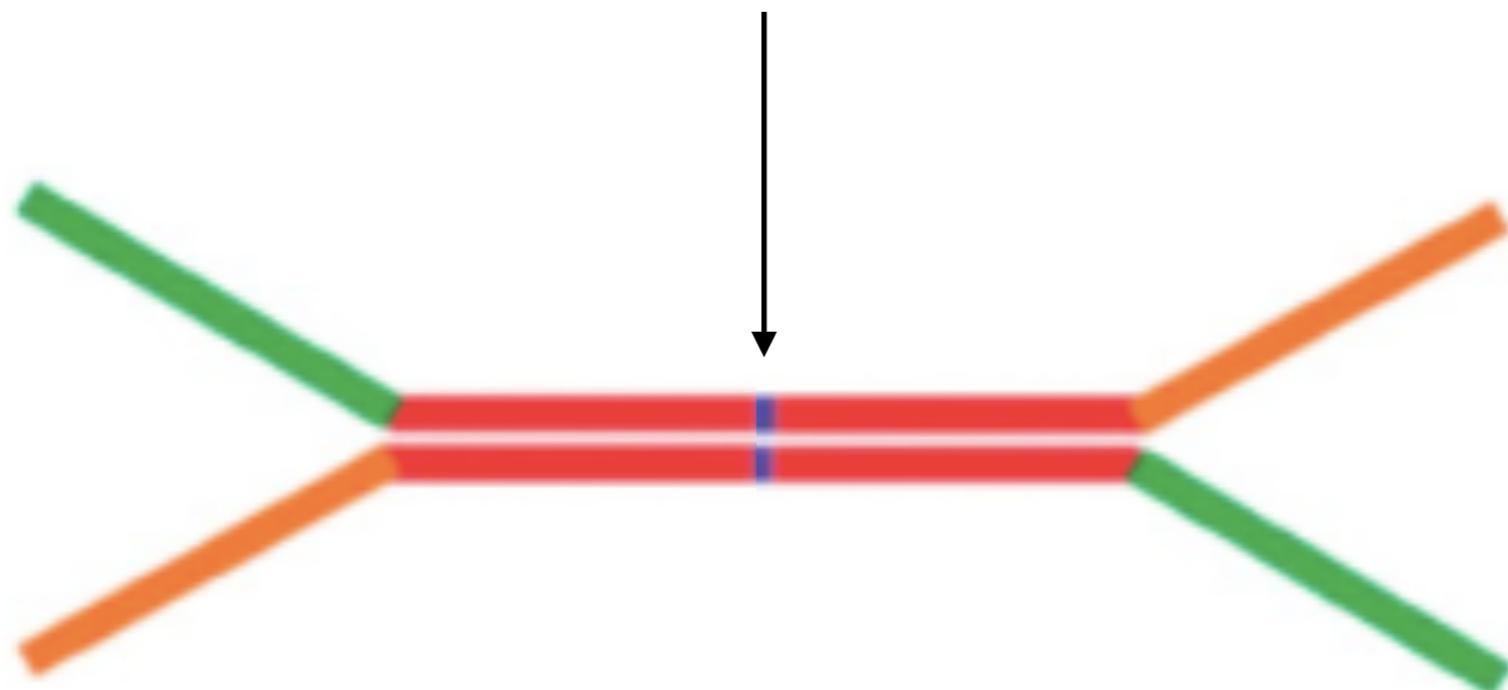
<http://www.grammar.zone/out-of-the-frying-pan-into-the-fire/>

PCR AMPLIFICATION

- For degraded / low-input samples, increase cycles.
- 12 is usually a good number.
- If you have to go past 16, it will almost certainly not sequence well.
- REMINDER: use at most half of your adapter-ligated template on the first go so that you can do more PCR if necessary without redoing the library!

ADAPTER DIMER IS YOUR ENEMY

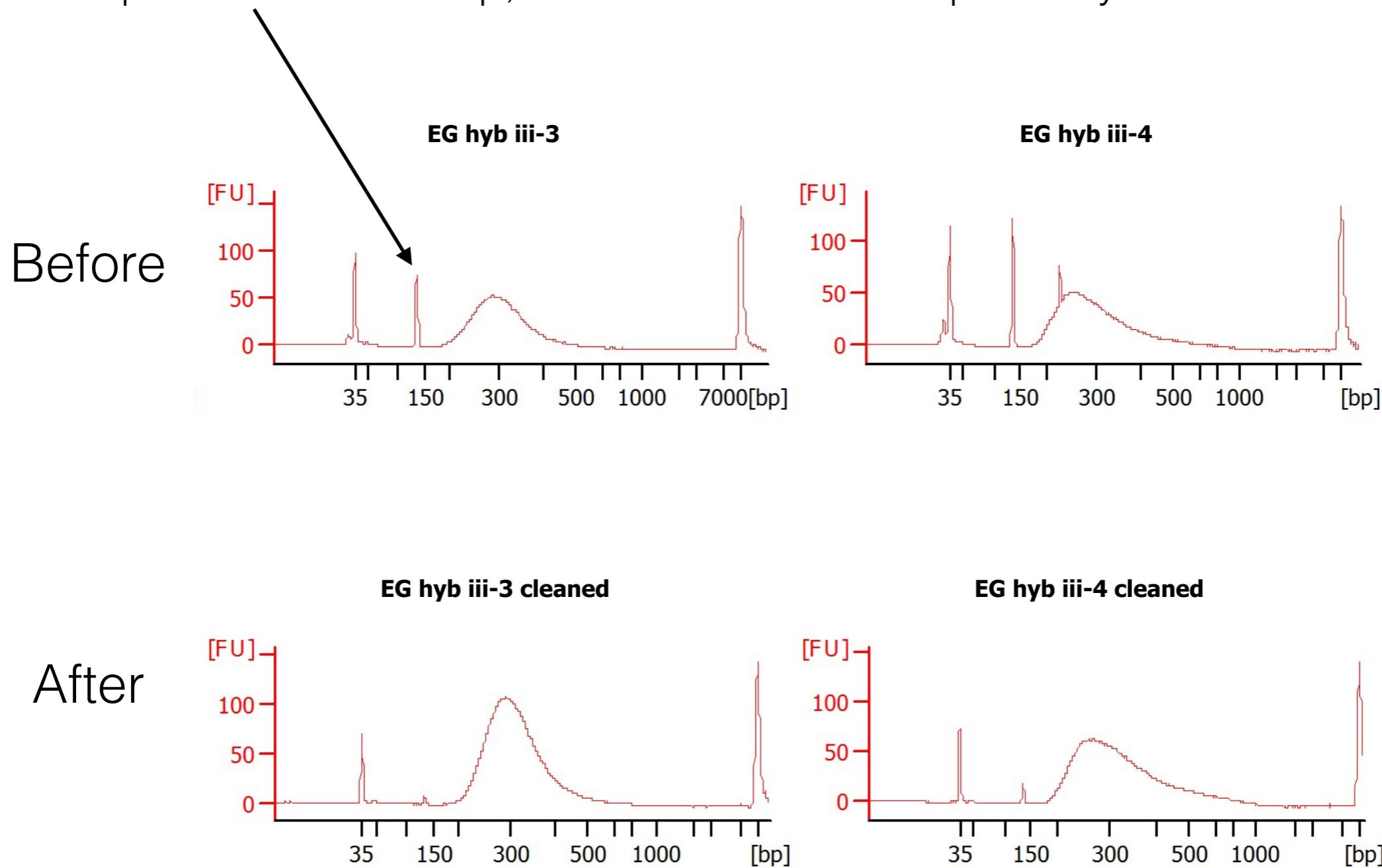
No insert - just two adapters stuck together.



If these get into the flow cell, you will sequence them instead of your inserts!

GETTING RID OF DIMER

Adapter dimer is 144 bp, so a 0.7x bead cleanup usually does the trick.

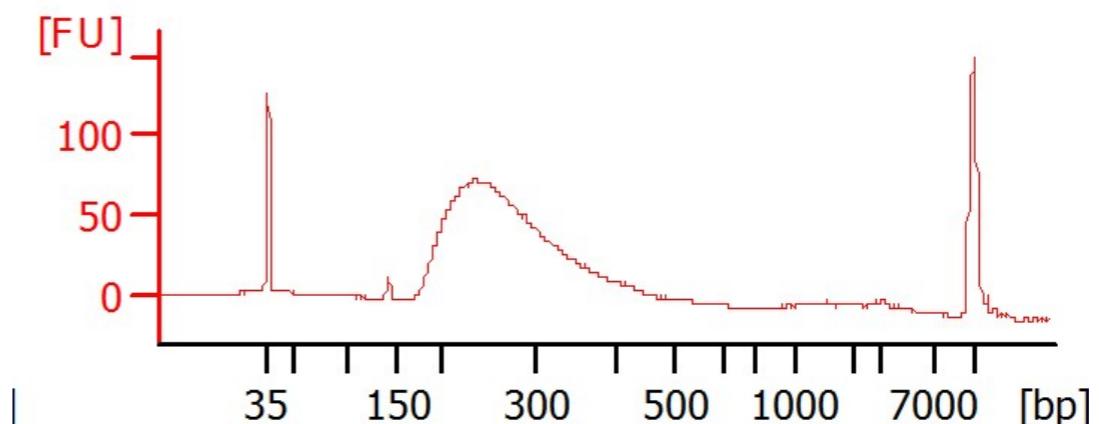


GETTING RID OF DIMER

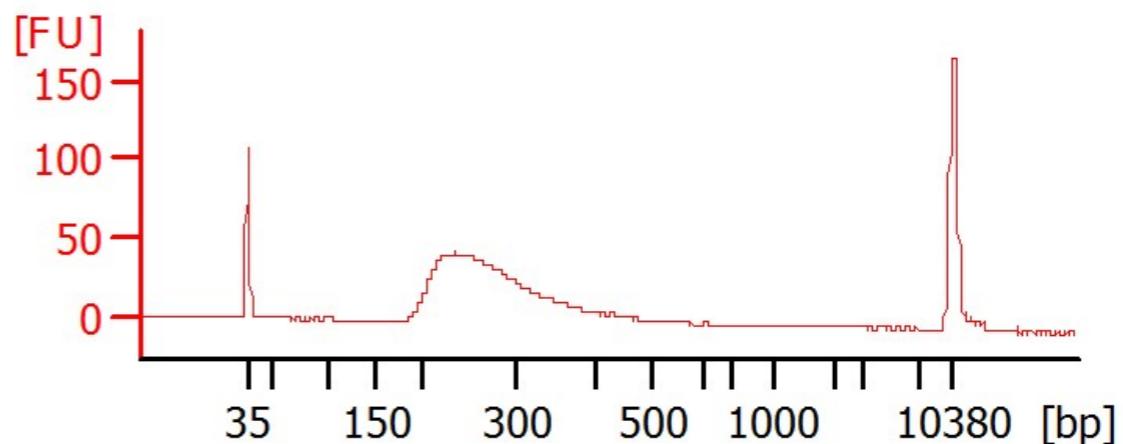
The BluePippin or PippinPrep can size-select very accurately to eliminate dimer without cutting into your library.



Hyb 7.2



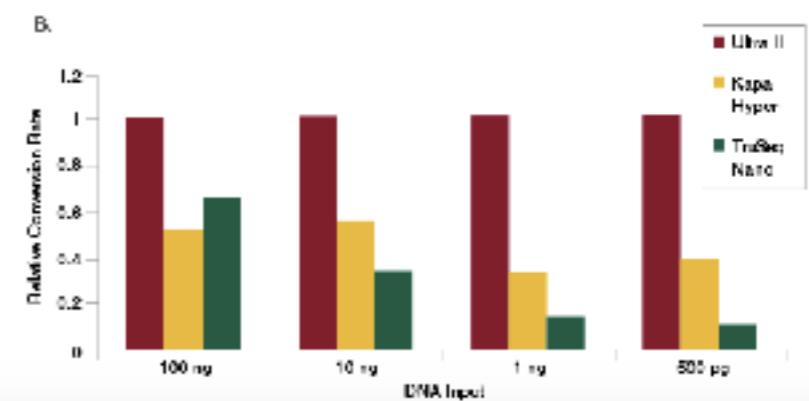
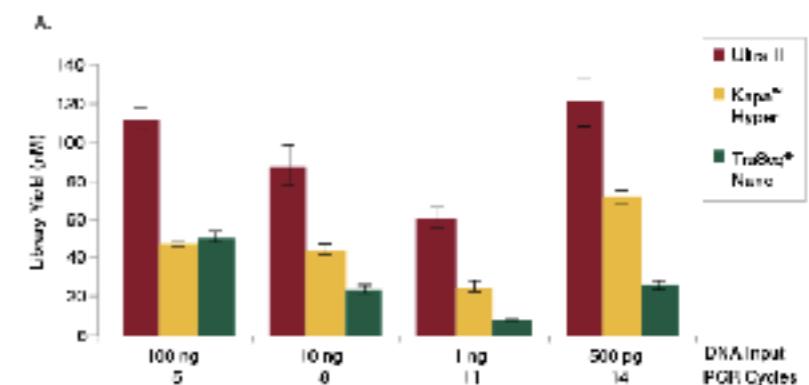
7.2 bp



OTHER KITS

- NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®
- KAPA Hyper Prep Kit
- Faster than the Illumina TruSeq Nano, but do not come with beads or adapters.
- Can be used at 1/4 vol (possibly even less).
- Fewer cleanups, so less potential for loss of material (helpful for herbarium samples)
- NEB claims the highest yield.

NEBNext Ultra II produces the highest library yields and conversion rates from a broad range of input amounts.



REDUCED VOLUMES TO SAVE COSTS

- Illumina TruSeq Nano kit: (£25 per sample at full volume)
 - At 1/2 volume £12.5
 - At 1/3 volume £8.33
 - At 1/4 volume £6.25
- Small volumes require:
 - Very accurate pipetting
 - Similar DNA input, but can be reduced for non-degraded samples
- Impact of smaller volumes:
 - Less accurate size selection (more subject to pipetting errors)
 - More loss of material during bead cleanup steps
 - Can be reduced by bead carryover and reactivation with PEG/NaCl
 - Usually require more PCR cycles
 - Not ideal for whole-genome sequencing
 - Not really a problem for HybSeq

OPTIONS FOR ADAPTERS

- NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Set 1) - 96 unique combinations
- NEXTflex-HT Barcodes | Bioo Scientific - 96 dual-index combinations or 384 unique single-index barcodes.
- Custom TruSeq-style adapters (from IDT or other oligo manufacturer)
- Adapterama system (Glenn et al.): thousands of unique combinations
 - Only one adapter - a universal stub
 - Index sequences are added using PCR primers (as with NEBNext oligos)

Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext)

 Travis C. Glenn, Roger Nilsen, Troy J. Kieran, John W. Finger, Todd W. Pierson, Kerin E. Bentley, Sandra Hoffberg, Swarnali Louha, Francisco J. Garcia-De-Leon, Miguel Angel del Rio Portilla, Kurt Reed, Jennifer L. Anderson, Jennifer K. Meece, Sammy Aggery, Romdhane Rekaya, Magdy Alabady, Myriam Belanger, Kevin Winker,  Brant C. Faircloth

doi: <https://doi.org/10.1101/049114>

QUALITY CONTROL

- Required: Qubit for quantification
- Recommended for a subset of samples: BioAnalyzer or TapeStation.



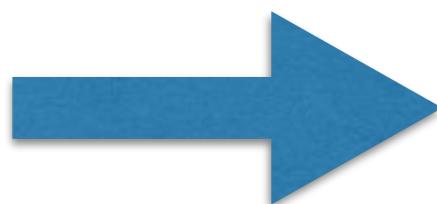
QUALITY CONTROL - QUANTIFICATION

How much library do you need?

Minimum: 100 ng / number of samples per hyb pool

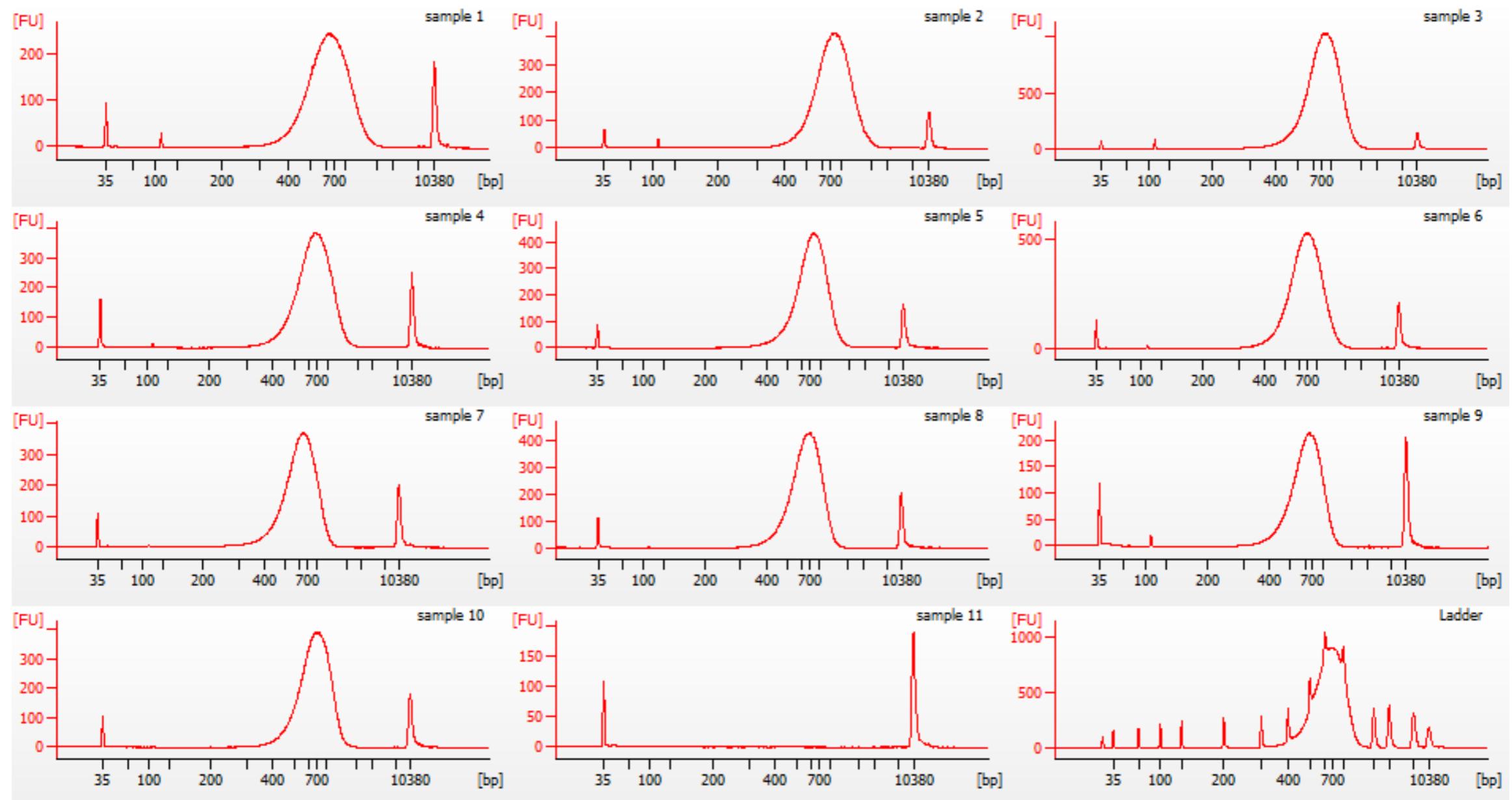
Better: 1000 ng / number of samples per hyb pool

Maximum: 2000 ng / number of samples per hyb pool

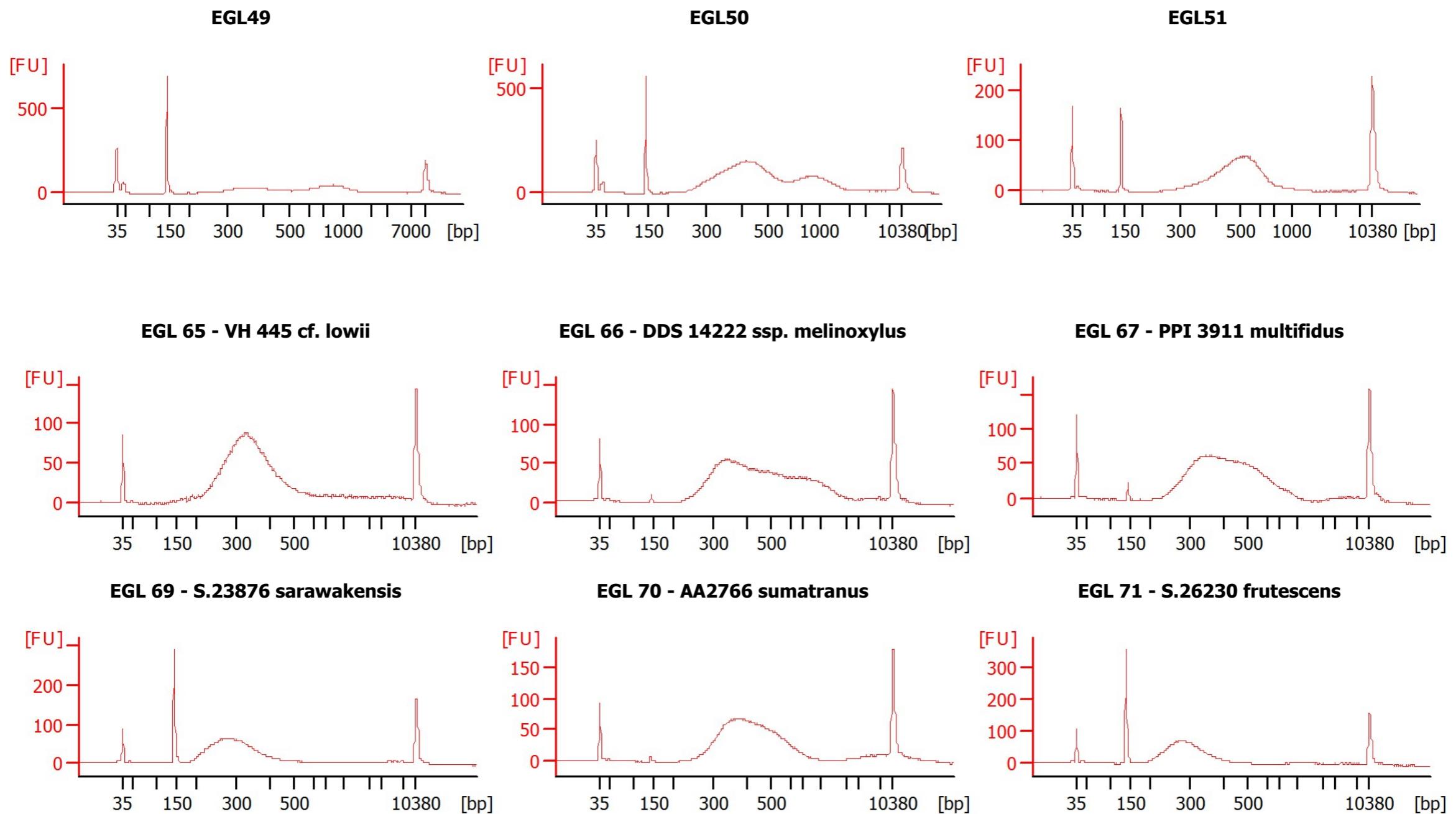


For pools of 6: 17–333 ng
For pools of 20: 5-100 ng

QUALITY CONTROL - BEAUTIFUL LIBRARIES

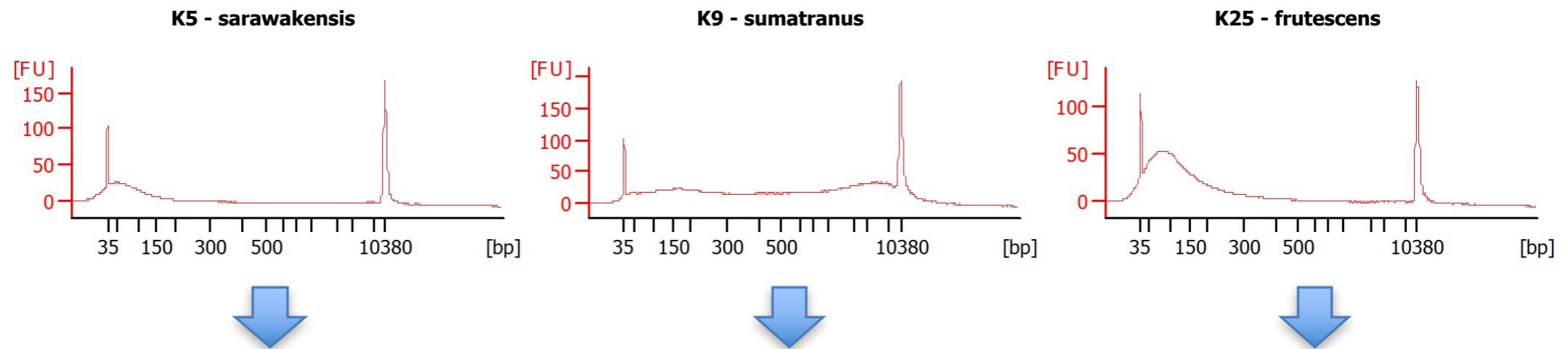


QUALITY CONTROL - UGLY LIBRARIES (USUALLY STILL WORK)

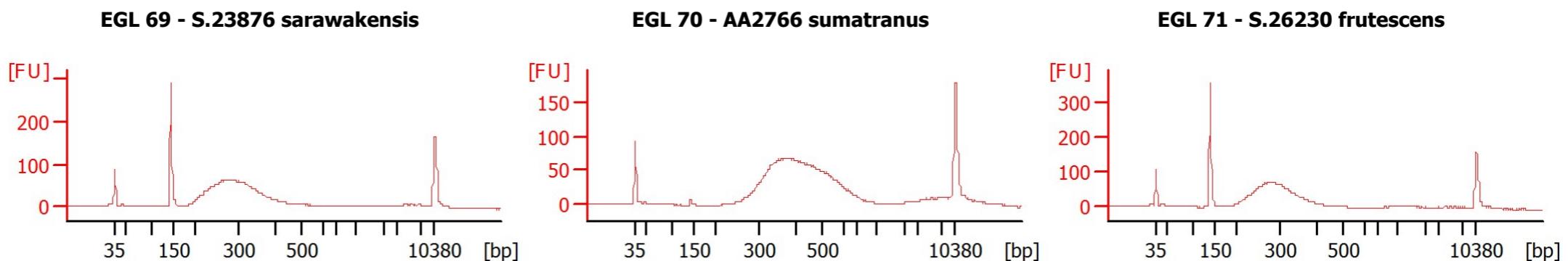


QUALITY CONTROL - HERBARIUM LIBRARIES BEFORE AND AFTER

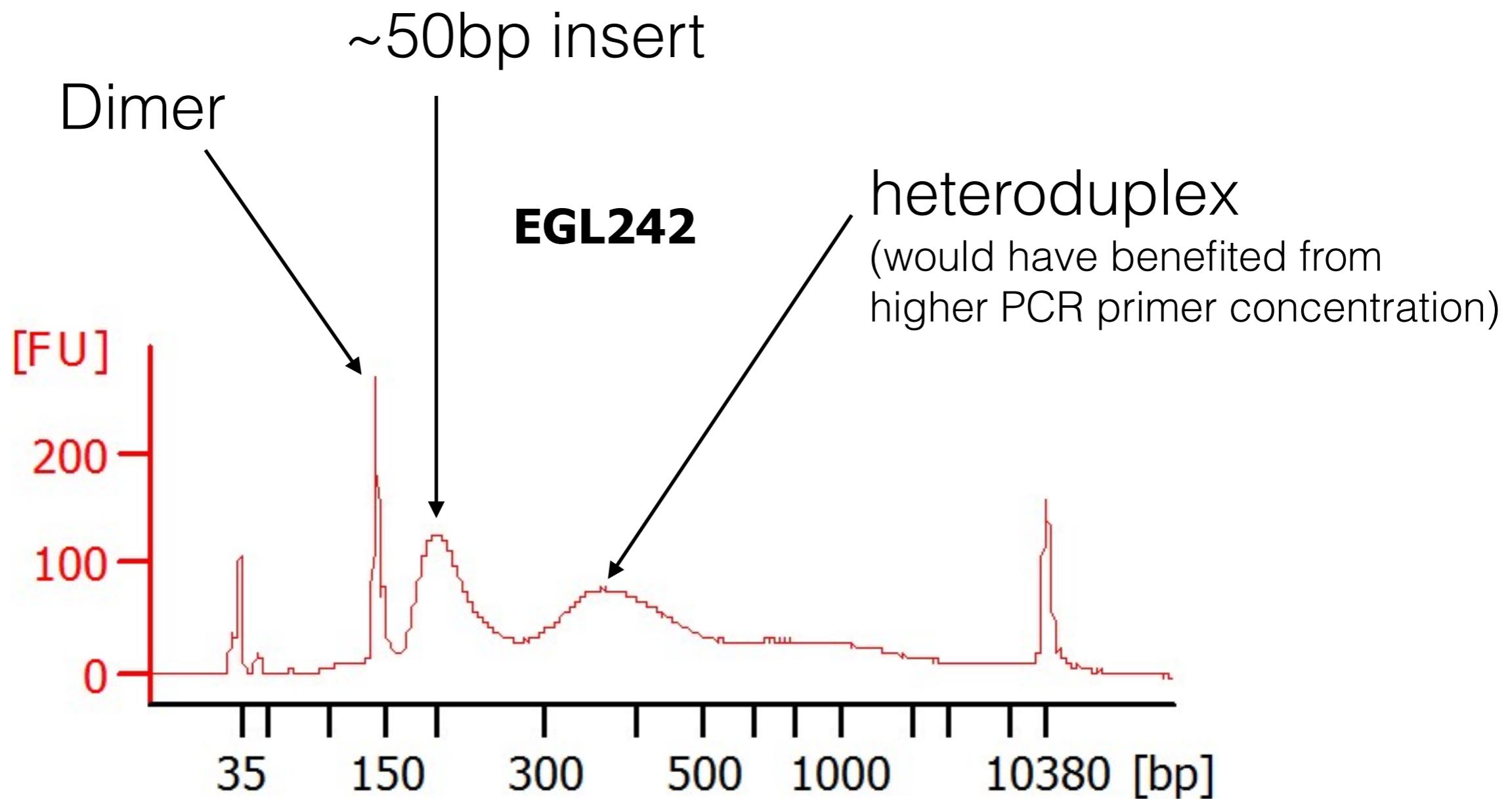
DNA



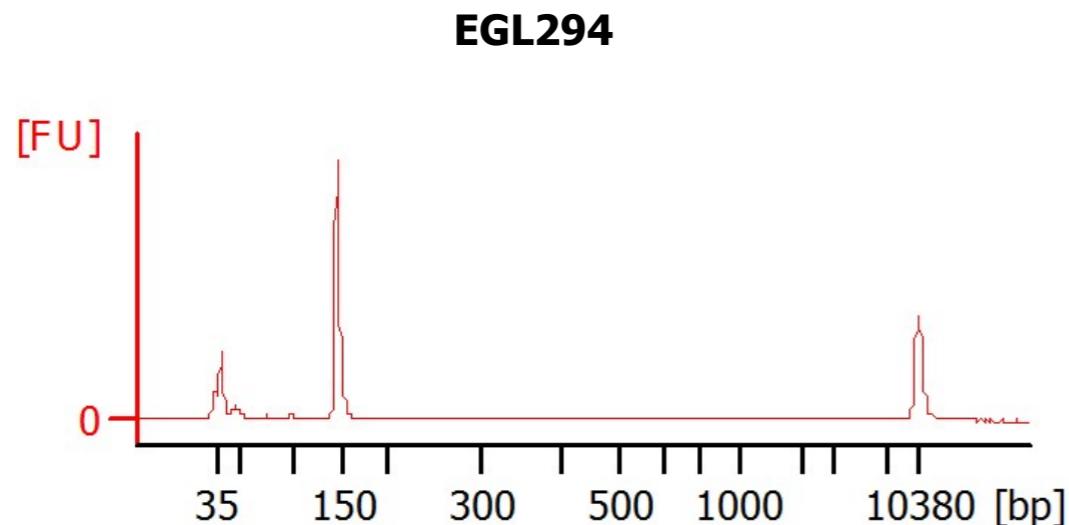
Library



QUALITY CONTROL - VERY SHORT INSERTS

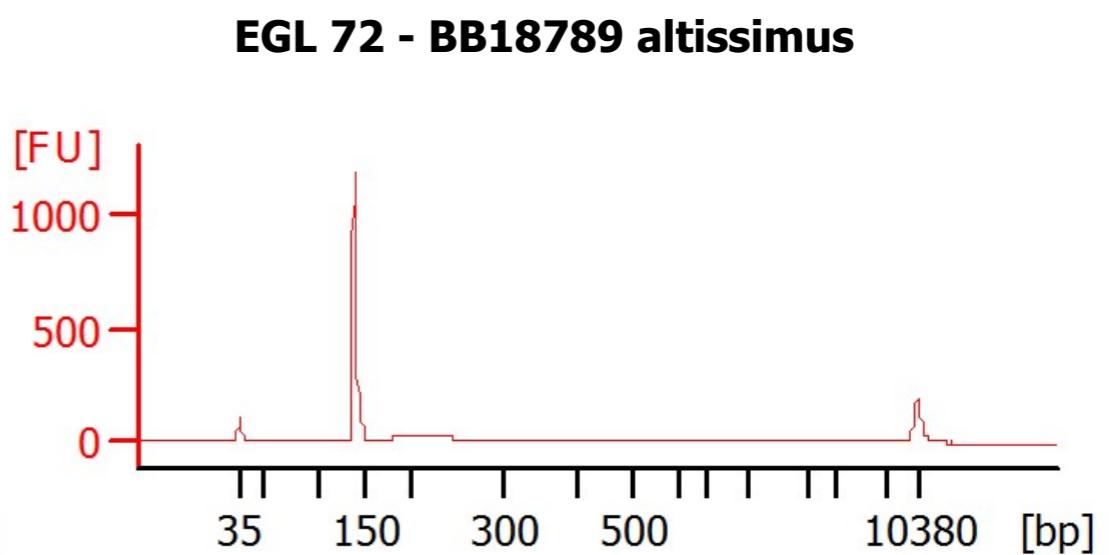


QUALITY CONTROL - TRULY AWFUL LIBRARIES



~60 year-old specimen from US. Did not work.

~80 year-old specimen from BO.
Recovered partial sequences for
59% of targets: usable in
supermatrix phylogeny.



BUILDING CONFIDENCE

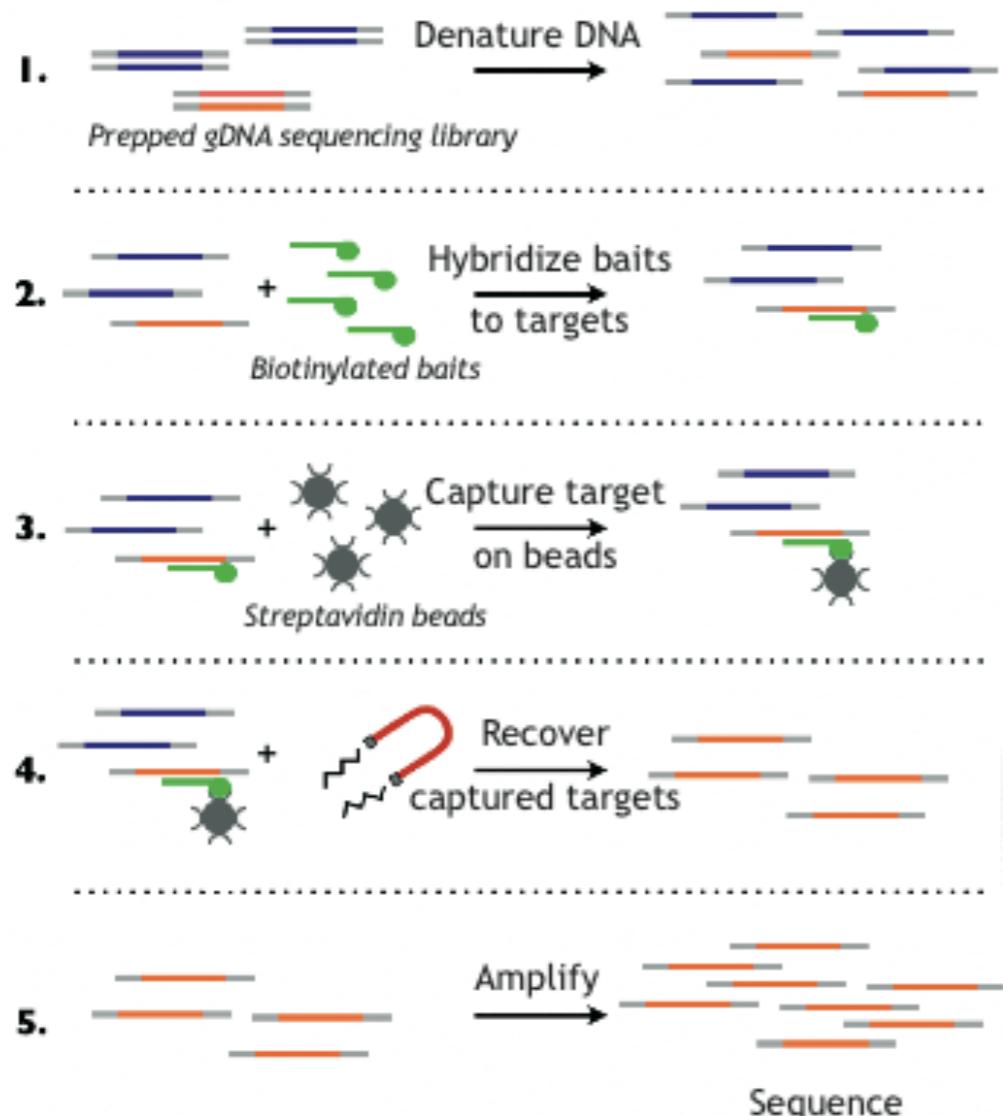
Age of material is not that important!



ACTIVITY

- Return to your case study groups and estimate a timeline for the project from sampling to sequencing.
- You may decide on the number of samples and the number of personnel working on the project.
- As we have not gotten to it yet, allocate two weeks for each round of pooling+hybridization+sequencing.

POOLING STRATEGIES, TARGET ENRICHMENT, AND SEQUENCING



Elliot Gardner
PhD Candidate
Plant Biology and Conservation
Northwestern University and Chicago Botanic Garden

OBJECTIVES

1. Describe how hybridization-based target enrichment works.
2. Apply existing phylogenetic knowledge to develop an appropriate hybridization strategy.
3. Balance cost and predictability to develop an appropriate pooling strategy.
4. Conduct and interpret quality-control assays on enriched libraries.
5. Calculate expected sequencing depth given a specified insert size, multiplexing level, and sequencing platform.
6. Choose an appropriate sequencing strategy for a HybSeq study.

ROAD MAP

1. Workflow overview (MYbaits)
2. Pooling samples
3. Final quality control
4. Sequencing

HYBRIDIZATION WORKFLOW OVERVIEW

Required consumables and equipment:

- MYbaits kit (£2,000 for the smallest custom kit: baits for ca. 500 samples plus reagents for samples — or £800 for the PAFTOL Angiosperm-wide kit)
- Dynabeads® MyOne™ Streptavidin C1, 2 mL (£310.00)
- KAPA HiFi PCR Kit x 100 reactions (£486)
 - Or other PCR master mix
- MYbaits reagent kit (£300, 48 reactions)
- PCR tubes (2 per reaction)
- LoBind tubes x 100 (£20) - need 2 per reaction
- PCR-grade water
- Thermocycler
- Heat block or water bath
- Magnet (tube or plate)



HYBRIDIZATION WORKFLOW OVERVIEW

Basic HybSeq hygiene:

- Remember that baits are RNA
- Clean surfaces with RNase Killer or something similar
- Use filter tips
- Use RNase-free tubes and water

HYBRIDIZATION WORKFLOW OVERVIEW

- 1) DNA sequencing library is heat-denatured in the presence of adapter-specific blocking oligonucleotides



- 2) Library and blockers are dropped to the hybridization temperature, allowing blockers to hybridize to the library adapters



- 3) Biotinylated RNA baits are introduced and allowed to hybridize to targets for several hours



- 4) Bait-target hybrids are pulled out of the solution with streptavidin-coated magnetic beads



- 5) Beads are stringently washed several times to remove non-hybridized and nonspecifically-hybridized molecules



- 6) Captured DNA library is released from the beads and amplified



POOLING

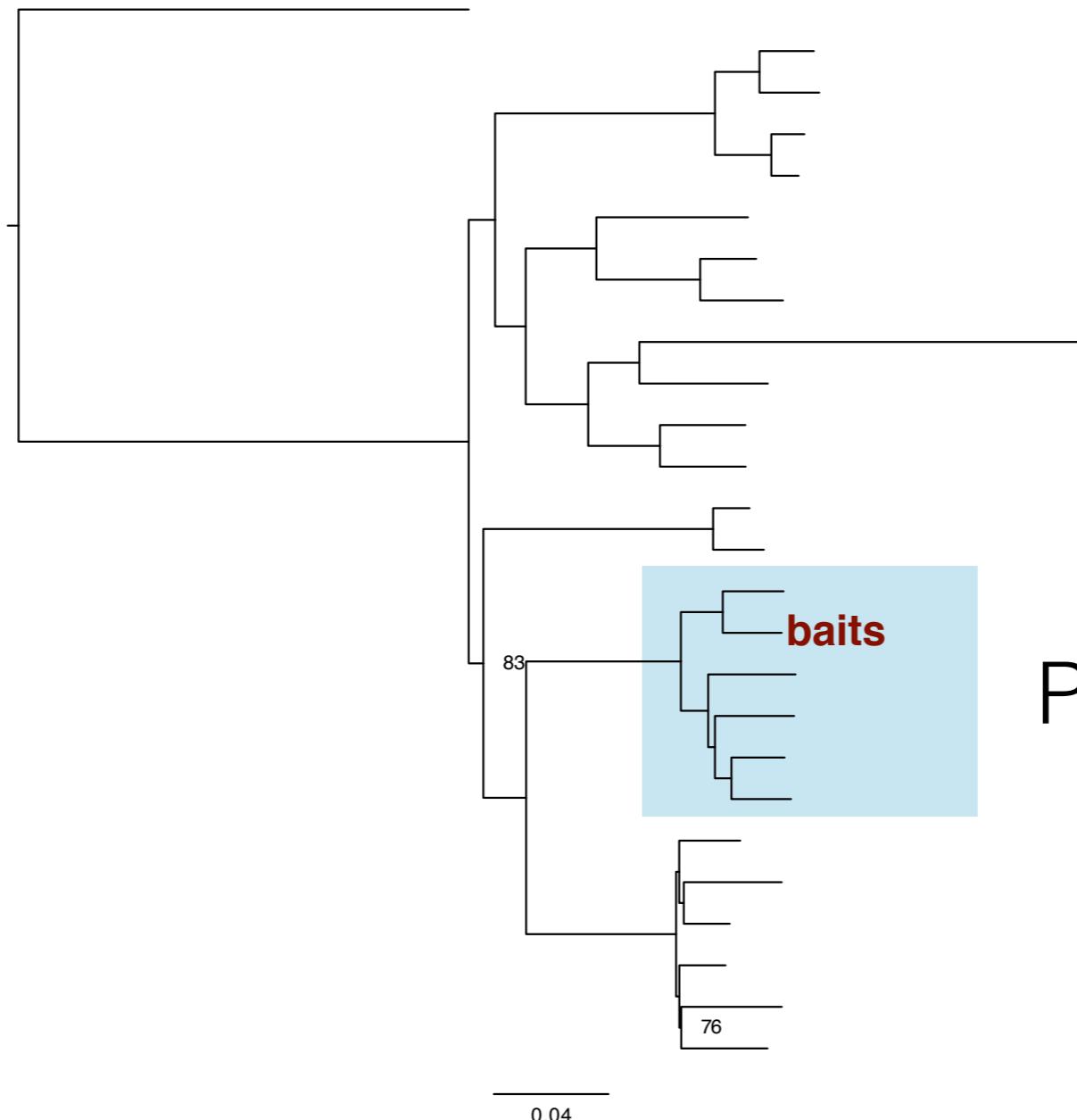
Combine desired libraries, dry down, and resuspend in 7 uL of water.

Pooling = opportunities for libraries to out-compete each other

Various possible reasons:

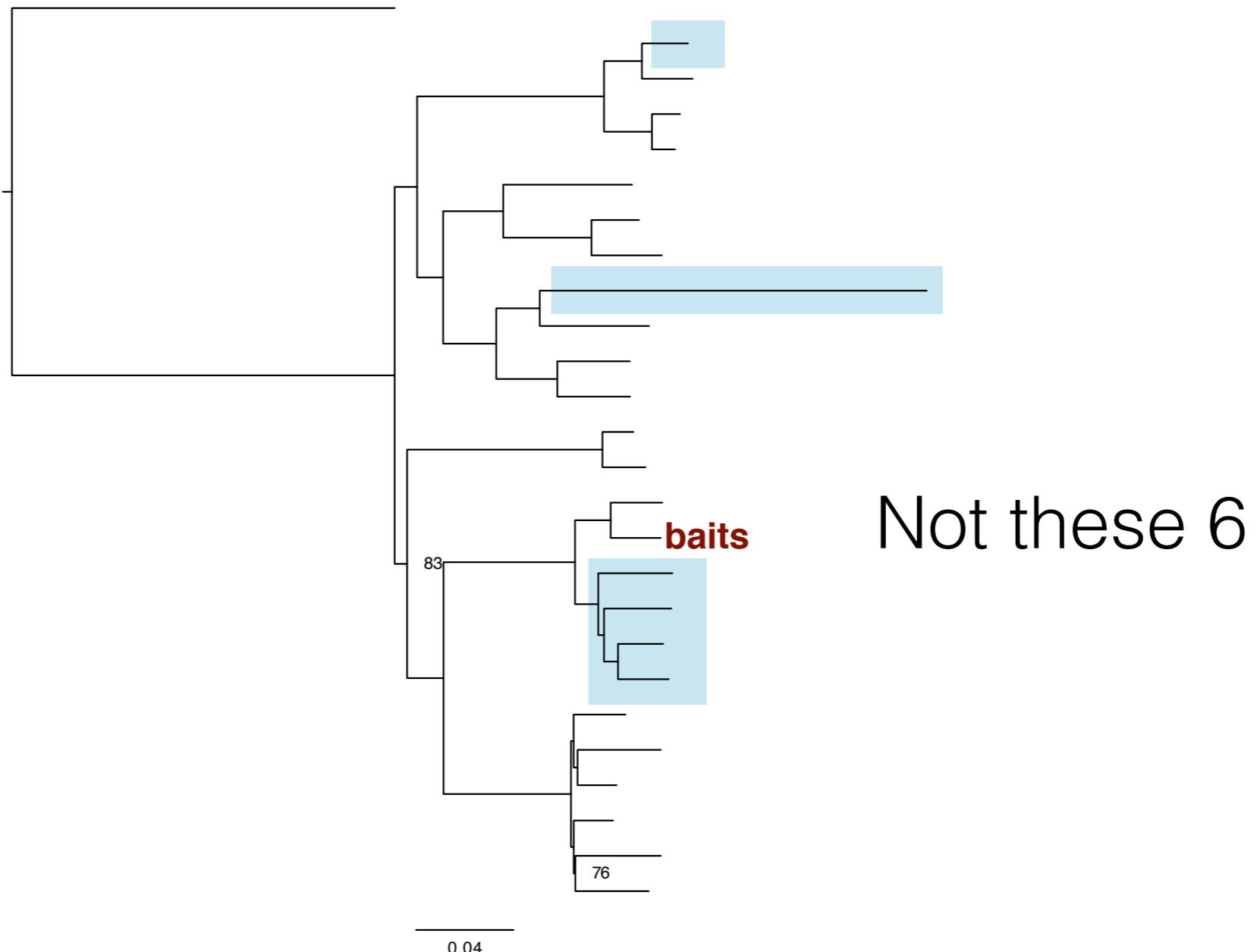
- Differing affinities to the baits (due to phylogenetic distance)
- Differing library qualities
- Inaccurate library quantification (e.g. caused by dimer)

POOLING



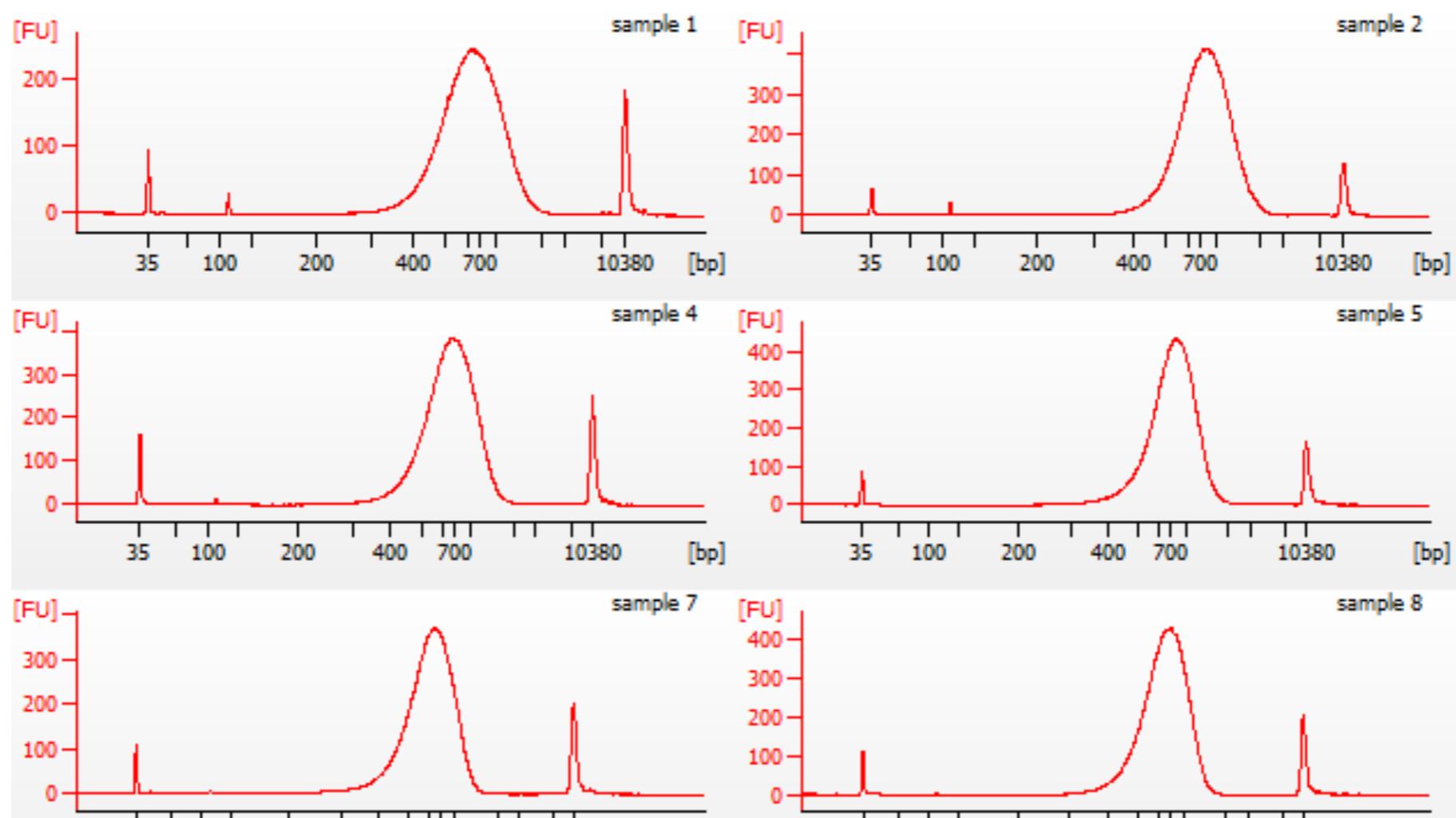
Pool these 6

POOLING



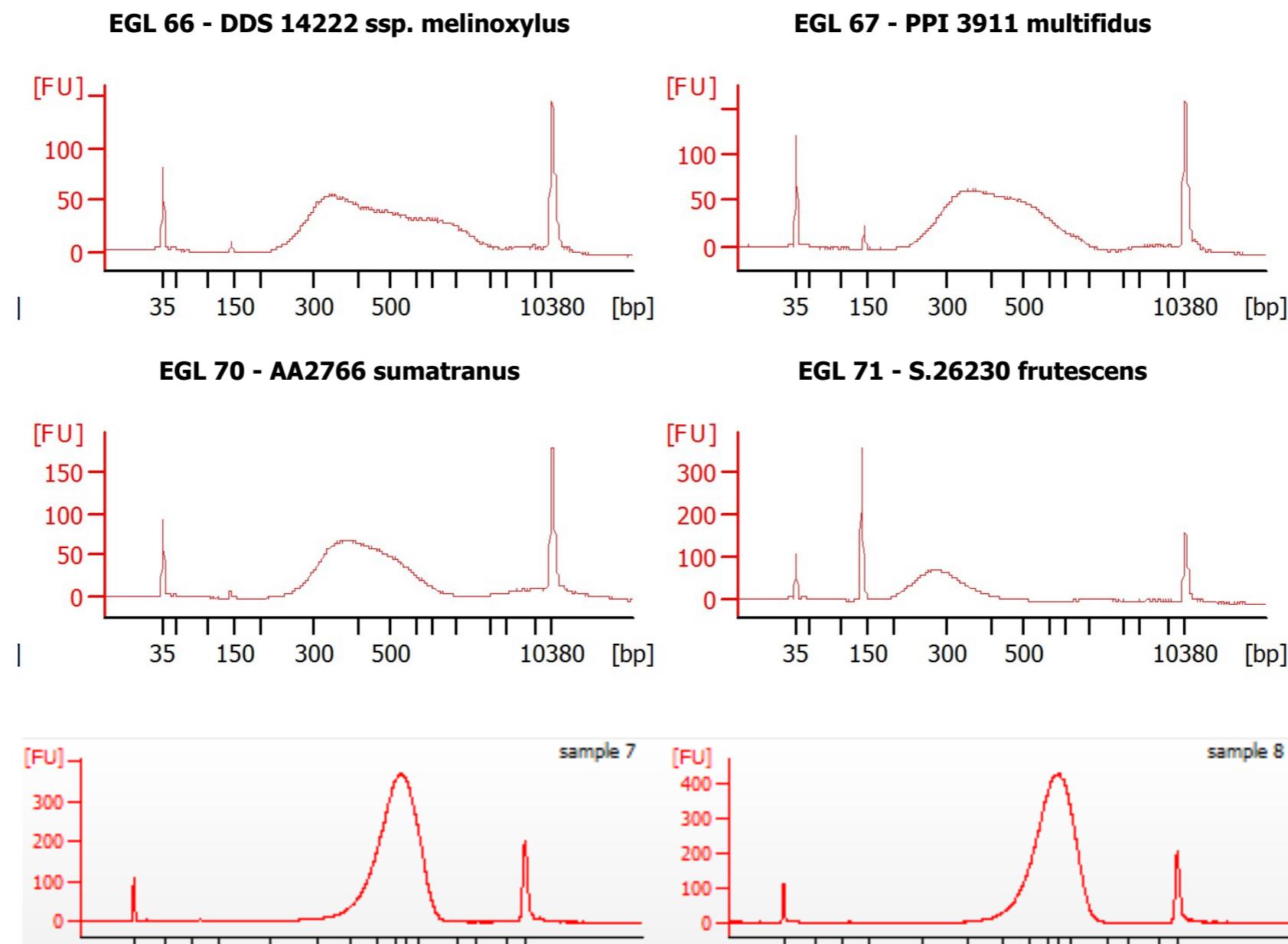
POOLING

Pool these 6



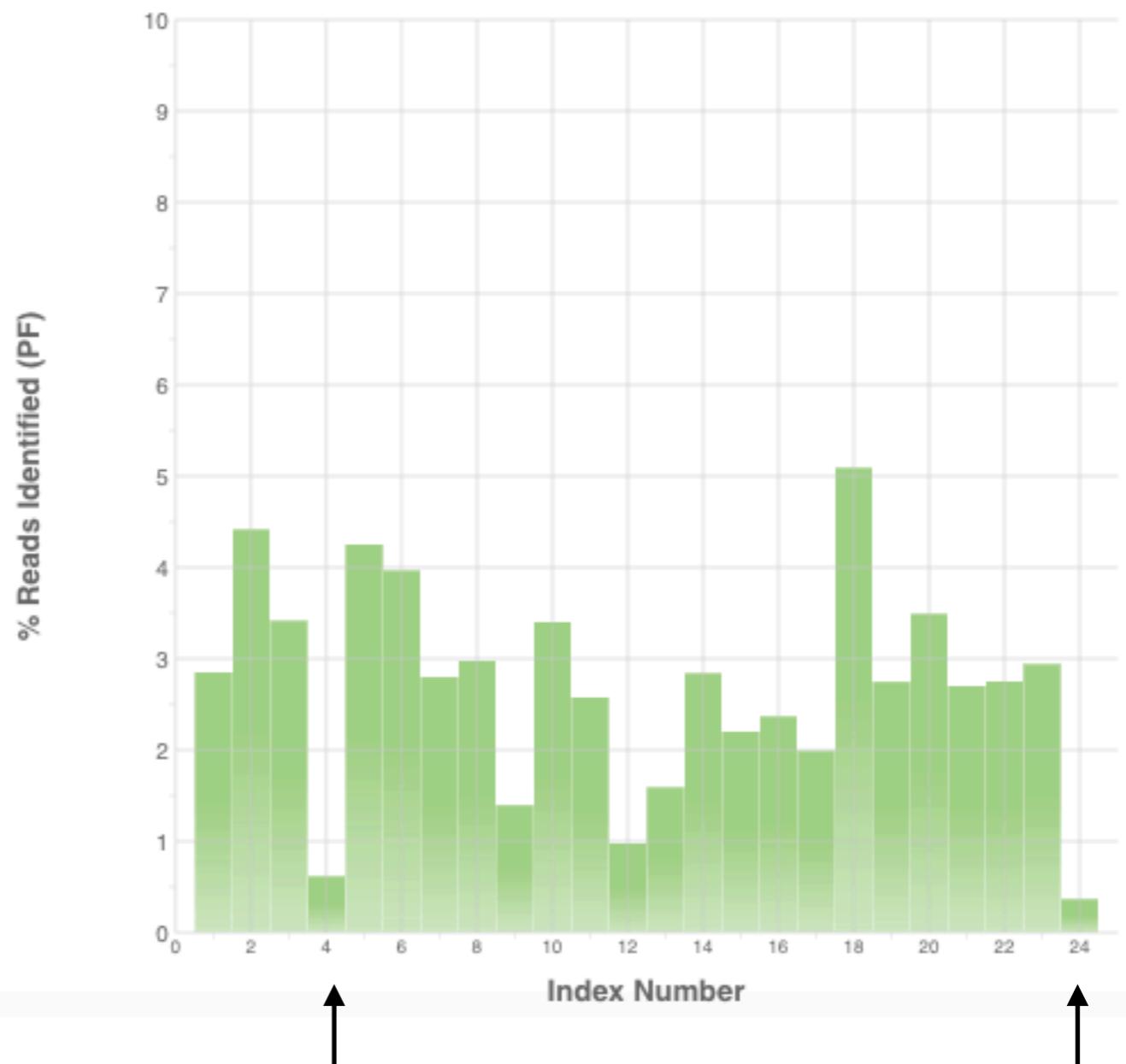
POOLING

Not these 6



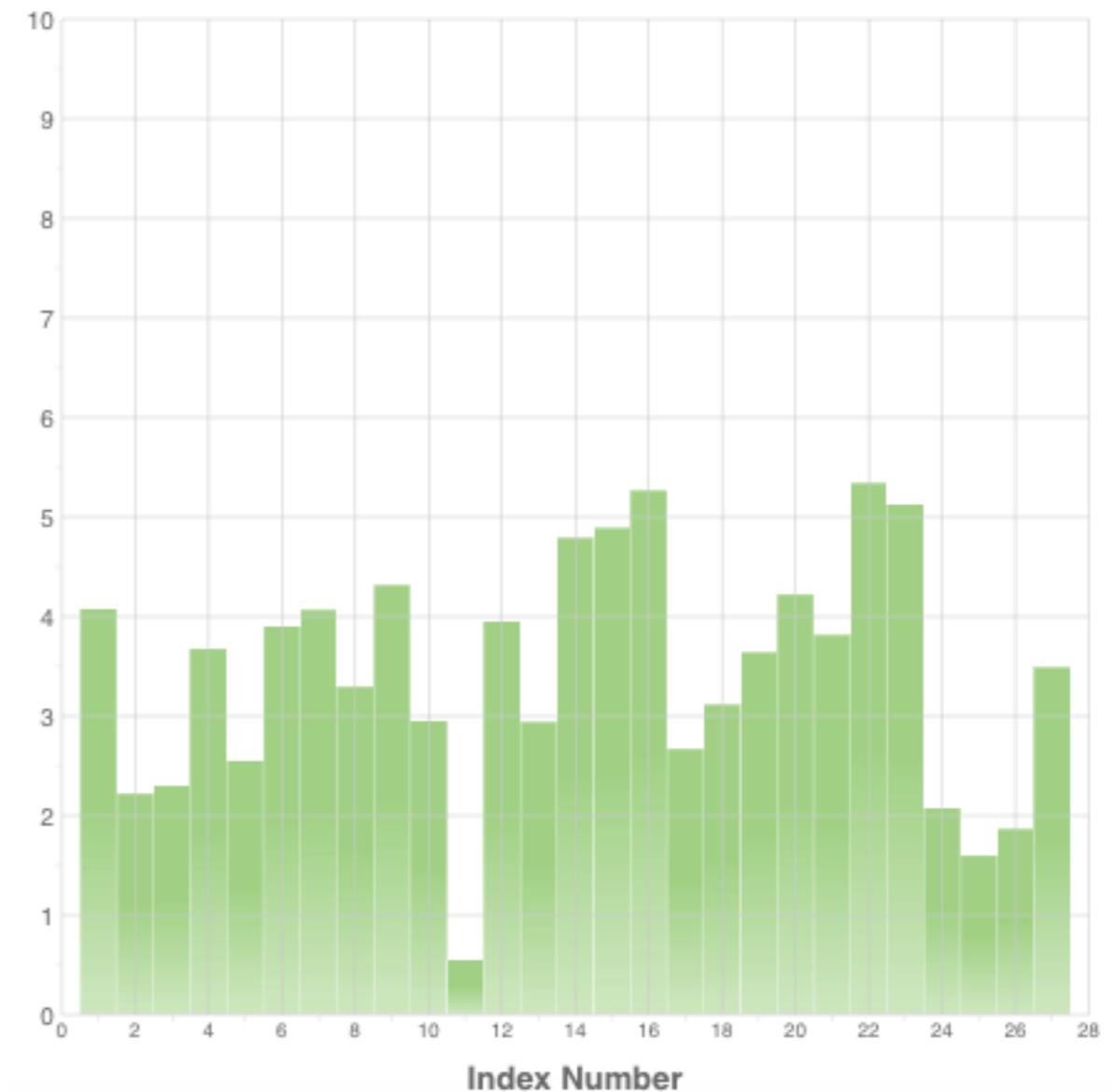
POOLING FOR HYBRIDIZATION

Improper pooling of taxa phylogenetically distant from baits with taxa close to baits (pools of 6).



POOLING FOR HYBRIDIZATION

Two pools of 24 libraries,
all in the same species
complex, all from silica-
dried material.



POOLING FOR HYBRIDIZATION

How many libraries to pool?

Bigger pools = cost savings but less predictability.

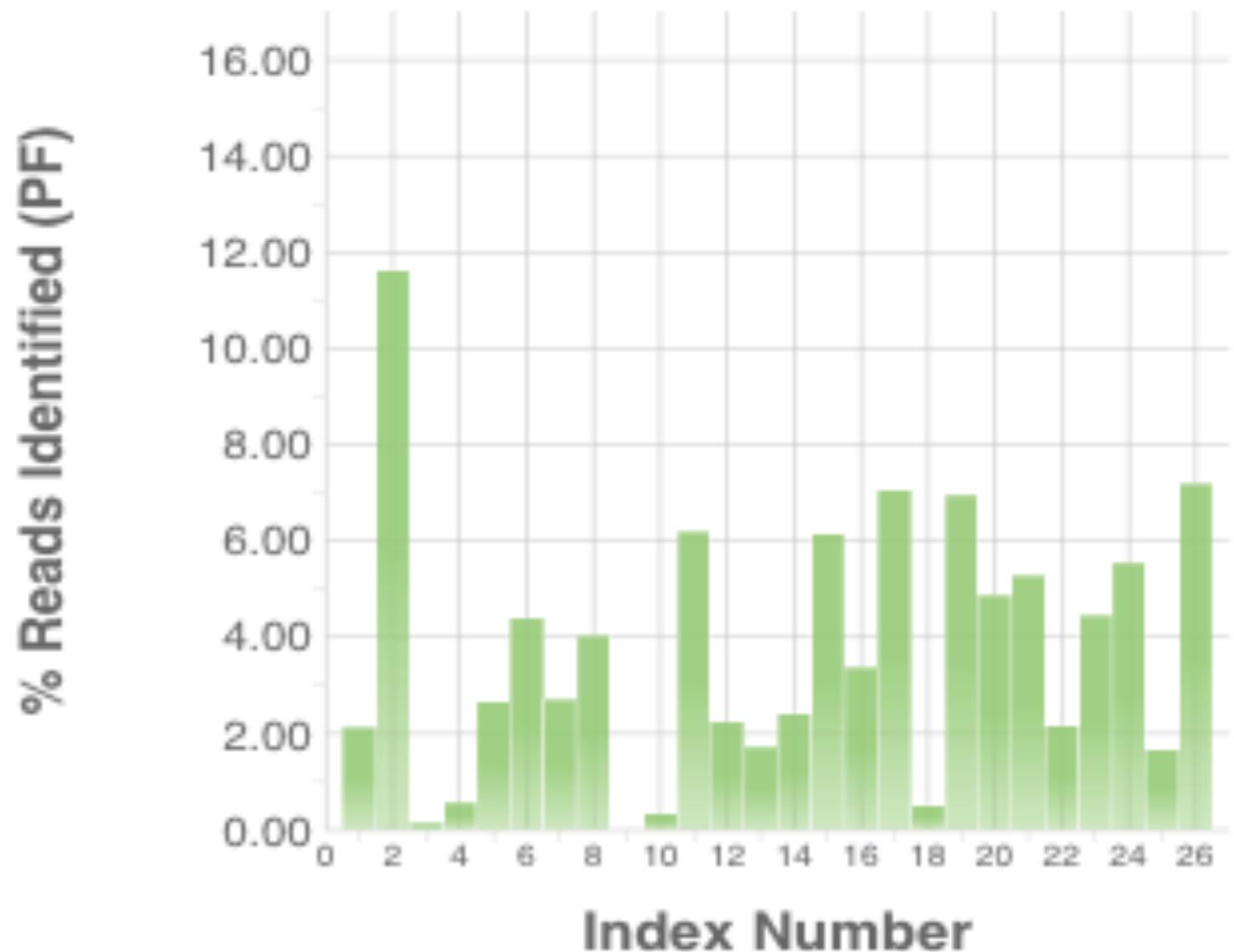
Smaller pools allow more control.

Consider 0.25x bait reactions to reduce pool size.

- Dilute baits 3:1 in RNase-free water and proceed according to the standard protocol.
- Extra reagents available from MYcroarray for ca. GBP 250 (48 reactions).

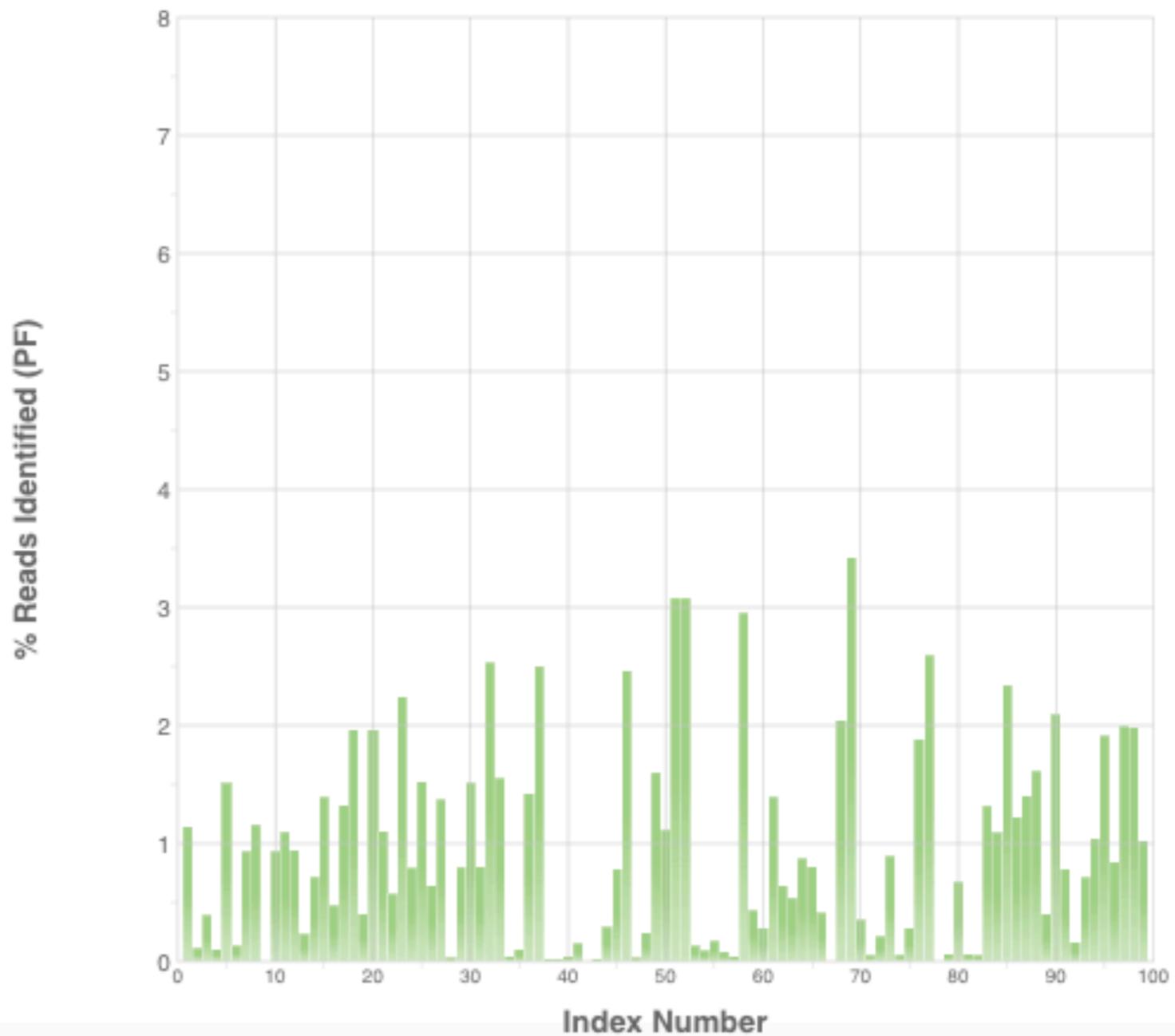
POOLING FOR HYBRIDIZATION

Four pools of 6-8, all herbarium specimens.



POOLING FOR HYBRIDIZATION

Six pools of 16–18,
mostly herbarium
specimens.



TIPS FOR HYBRIDIZATION

Hybridize for at least 16 hours. Use default temperature (65C) unless there is a good reason not to.

Use no more than 1/3 of your enriched product for your first PCR attempt.

Use as few PCR cycles as possible, but most labs we've spoken to have to use at least 14.

TIPS FOR HYBRIDIZATION

If the material is very old and degraded and/or contains a lot of non-endogenous DNA, consider hybridizing twice, perhaps at a cooler temperature.

- Low input leads to low enrichment
- *Artocarpus* samples enriched twice:
 - 20 hours at 60C —> 11% reads on target
 - Repeated —> ca. 70% reads on target

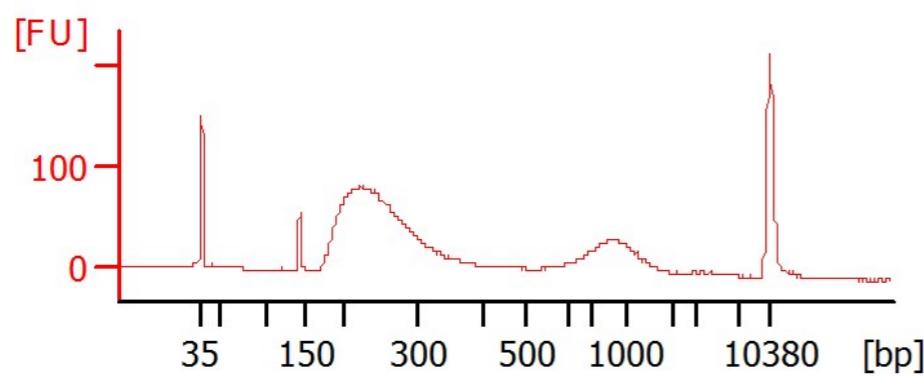


QUALITY CONTROL POST-PCR

Enriched product must be assessed for concentration and quality before sequencing.

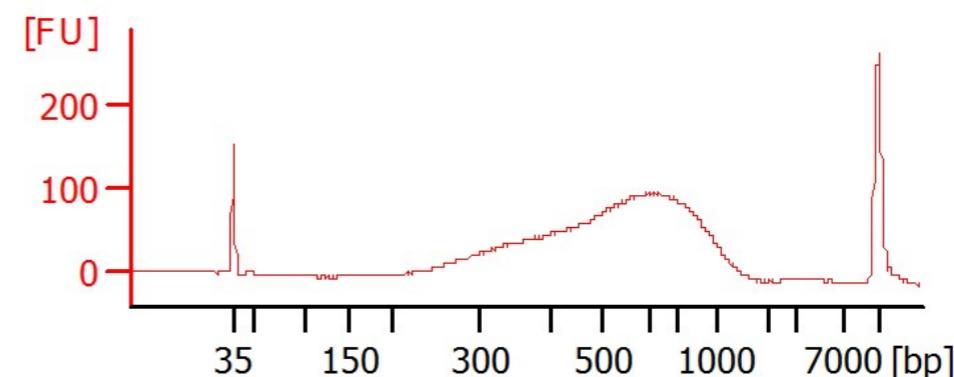
Any dimer at all must be cleaned out (beads or BluePippin)

Hyb 7.4



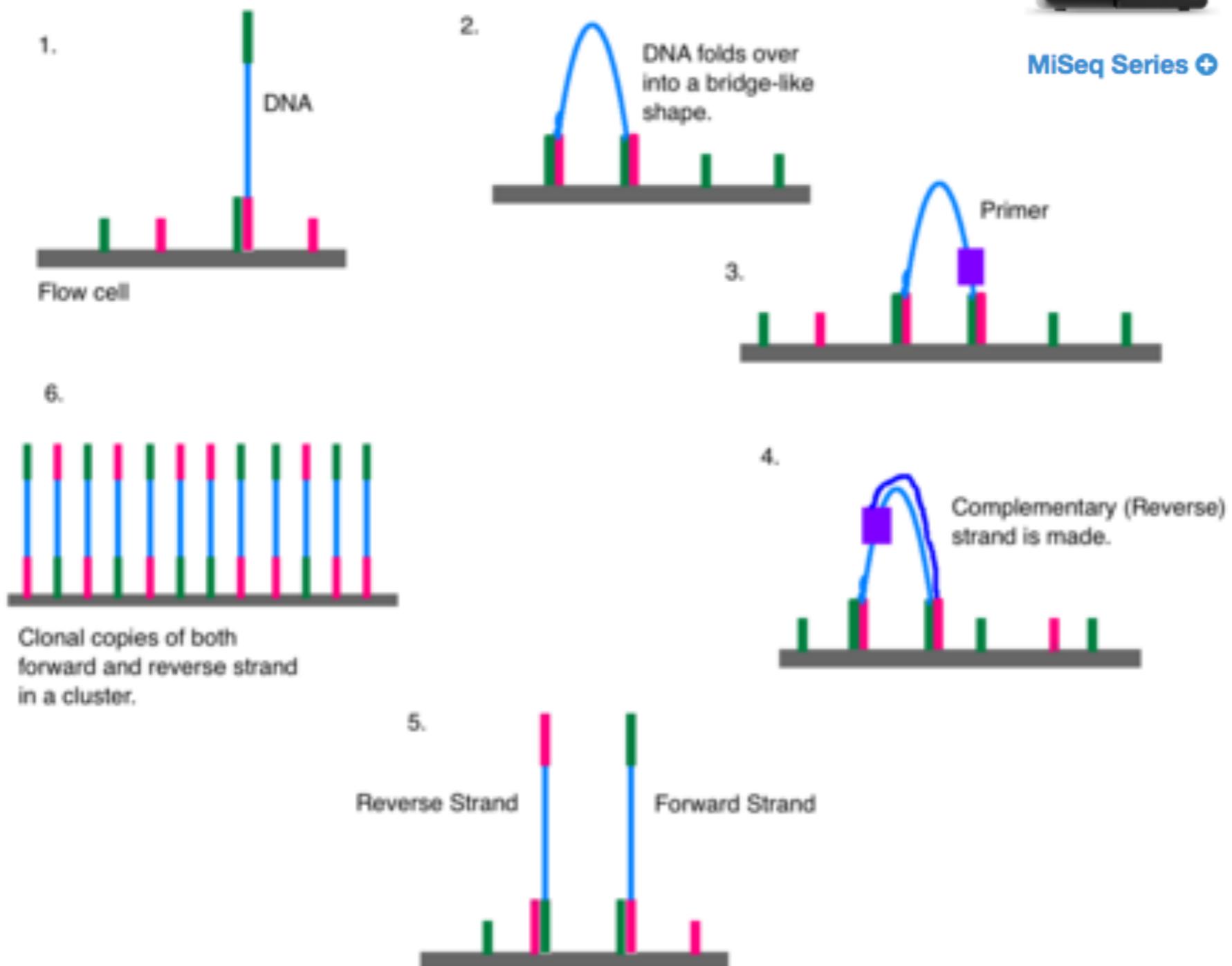
Must be cleaned and run again on BA or TapeStation

Hyb 7.1



Okay to run

ILLUMINA SEQUENCING



MiSeq Series +



NextSeq Series +

MISEQ SPECIFICATIONS



MiSeq Series

MiSeq Reagent Kit v3		
Read Length	2 × 75 bp	2 × 300 bp
Total Time*	~21 hrs	~56 hrs
Output	3.3–3.8 Gb	13.2–15 Gb

CALCULATING SEQUENCING DEPTH

Ideal depth calculation:

Reads * read length * enrichment / samples / target size



(or insert size, whichever is smaller)

e.g.: 50M reads * 300bp * 0.7 enrichment / 96 samples / 1,500,000 bp targets
= 72x

or if inserts are small:

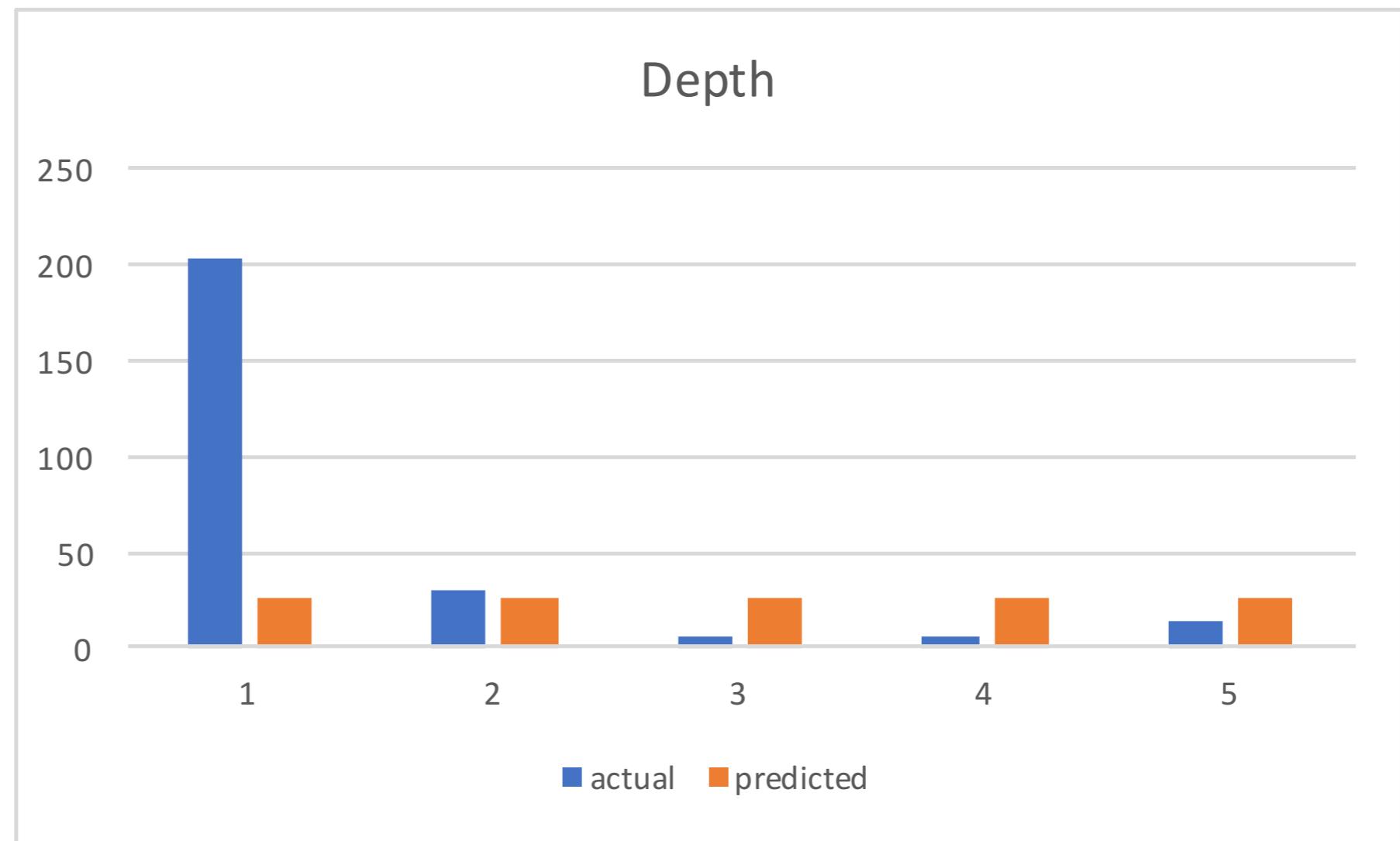
50M reads * 100bp * 0.7 enrichment / 96 samples / 1,500,000 bp targets
= 24x

CALCULATING SEQUENCING DEPTH

However, pooling during hybridization allows some samples to outcompete others.

So always aim high:
e.g., 50–100x

Examples from my last run

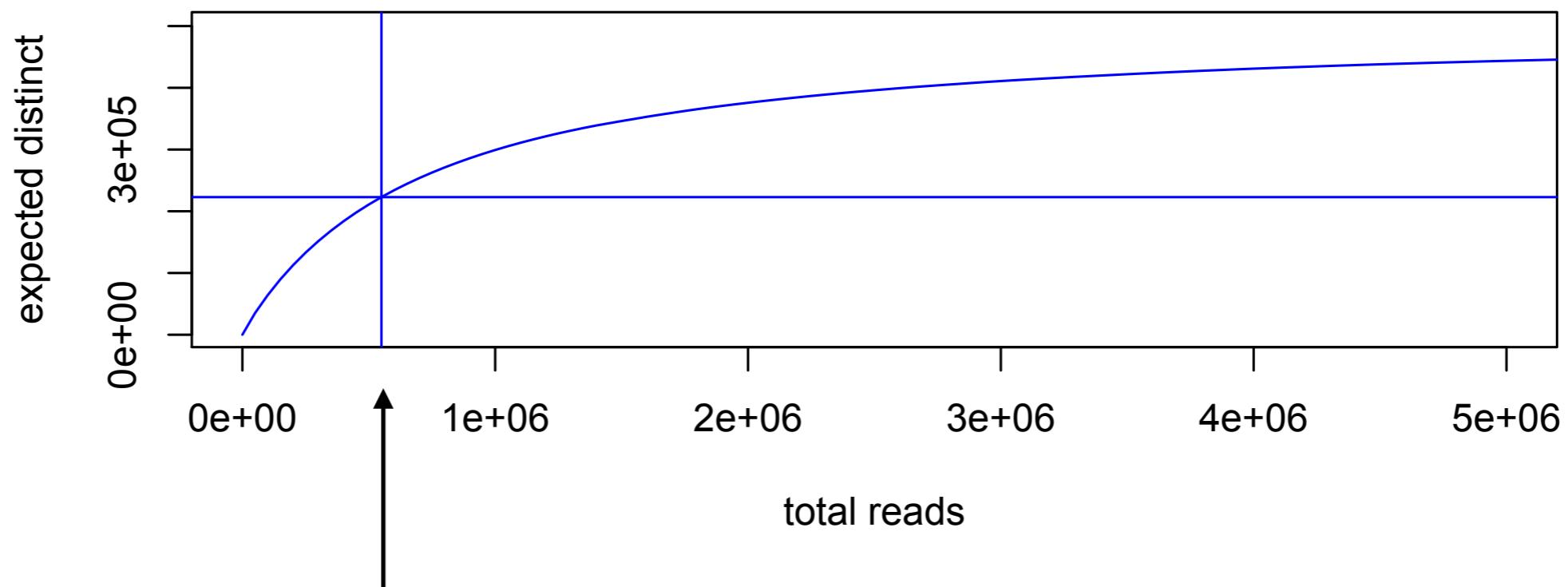


DECIDING WHETHER TO SEQUENCE MORE DEEPLY (OR HYBRIDIZE AGAIN)

Preseq - simulate deeper sequencing to see if it's worth it.

Daley & Smith 2013. Nature Methods 10, 325–327

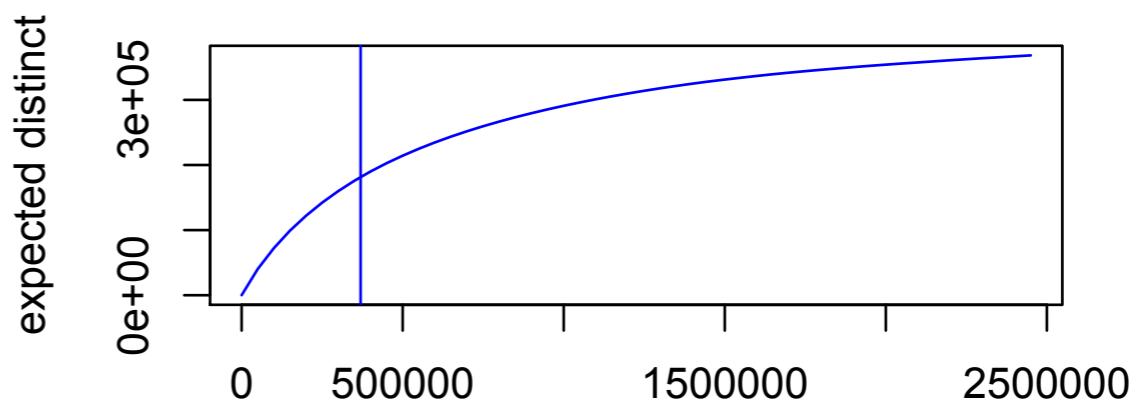
Good sample: all targets recovered



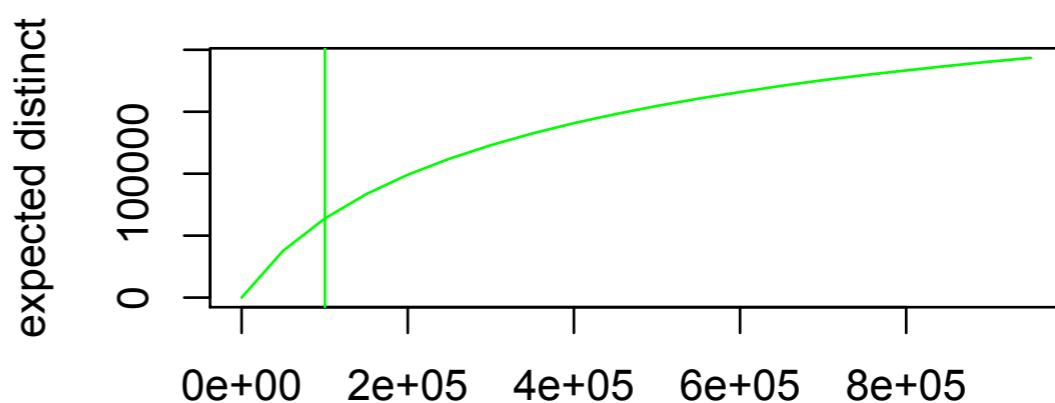
Here's where we are.

DECIDING WHETHER TO SEQUENCE MORE DEEPLY (OR HYBRIDIZE AGAIN)

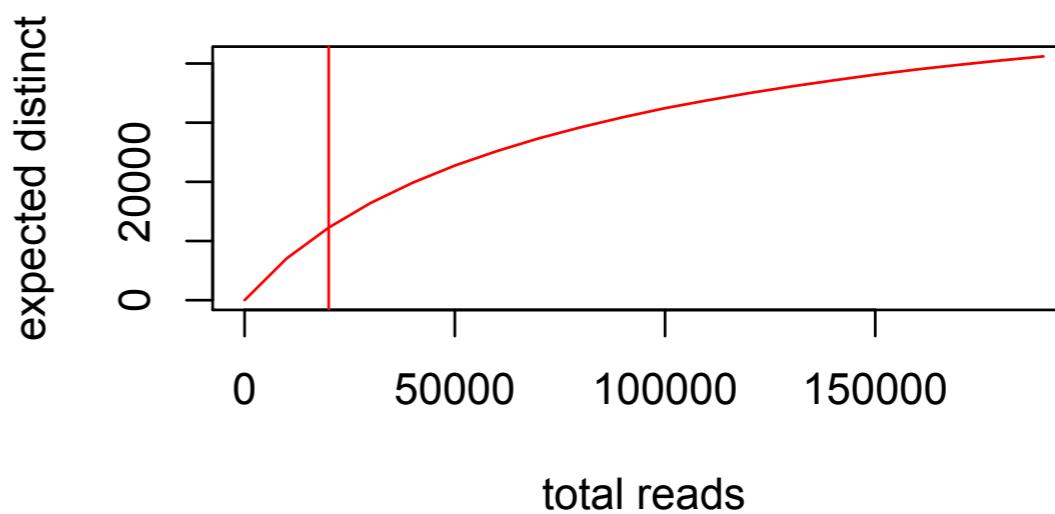
high recovery sample (86% of targets)



medium recovery sample (64% of targets)



low recovery sample (19% of targets)



ACTIVITY

- Follow the link on GitHub to the table of libraries and concentrations.
- Develop a pooling strategy based upon concentrations and taxonomy (using the provided paper - Zerega et al. 2010)
- Decide on a sequencing strategy and calculate expected depth.

Sequencing options:

- MiSeq: 2 x 25M reads, length 75 or 300bp
- NextSeq: 2 x 100M reads, length 75 or 150bp
- HiSeq: 2 x 300M reads, length 100