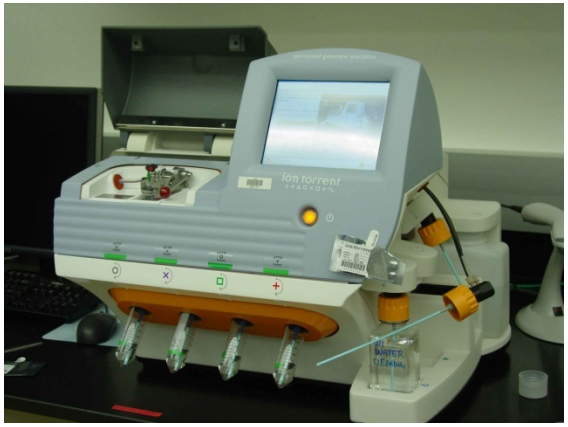
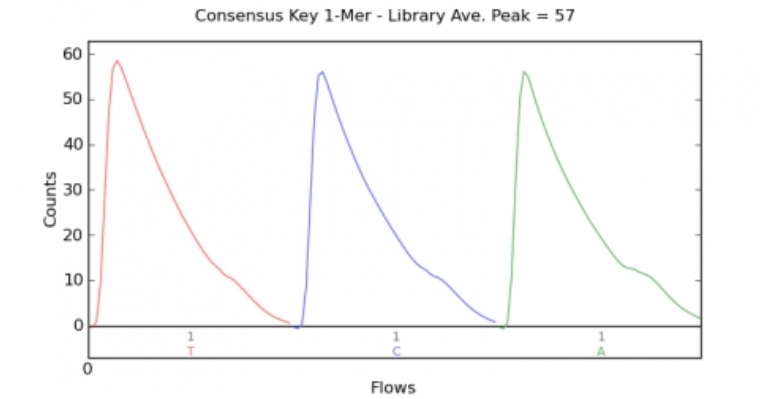
BINF 6350 ITSC 8350
Fall 2011
Biotechnology & Genomics Lab
Creating DNA libraries

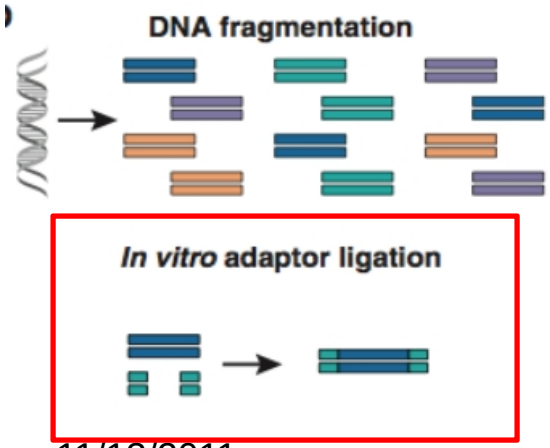
http://webpages.uncc.edu/~jweller2

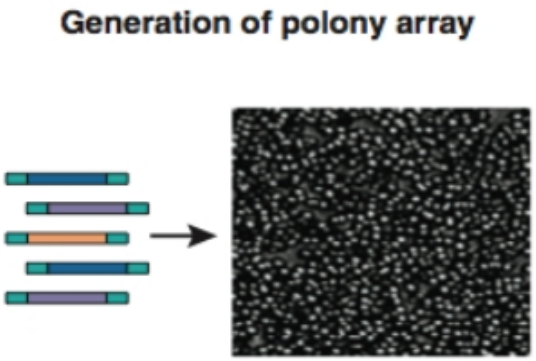


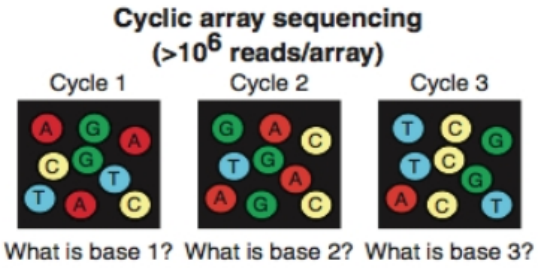


NGS data production steps

- Produce the fragments to be sequenced, clean up
- Modify fragments for processing, clean up
- Separate the modified fragments *from each other*
  - so they don't compete in the PCR process: emulsion reactors, dispersed on a surface
- Amplify each fragment to a base mass
- Sequence-by-synthesis (flowgrams, PPi with reversible terminators, etc.)
- Analyze

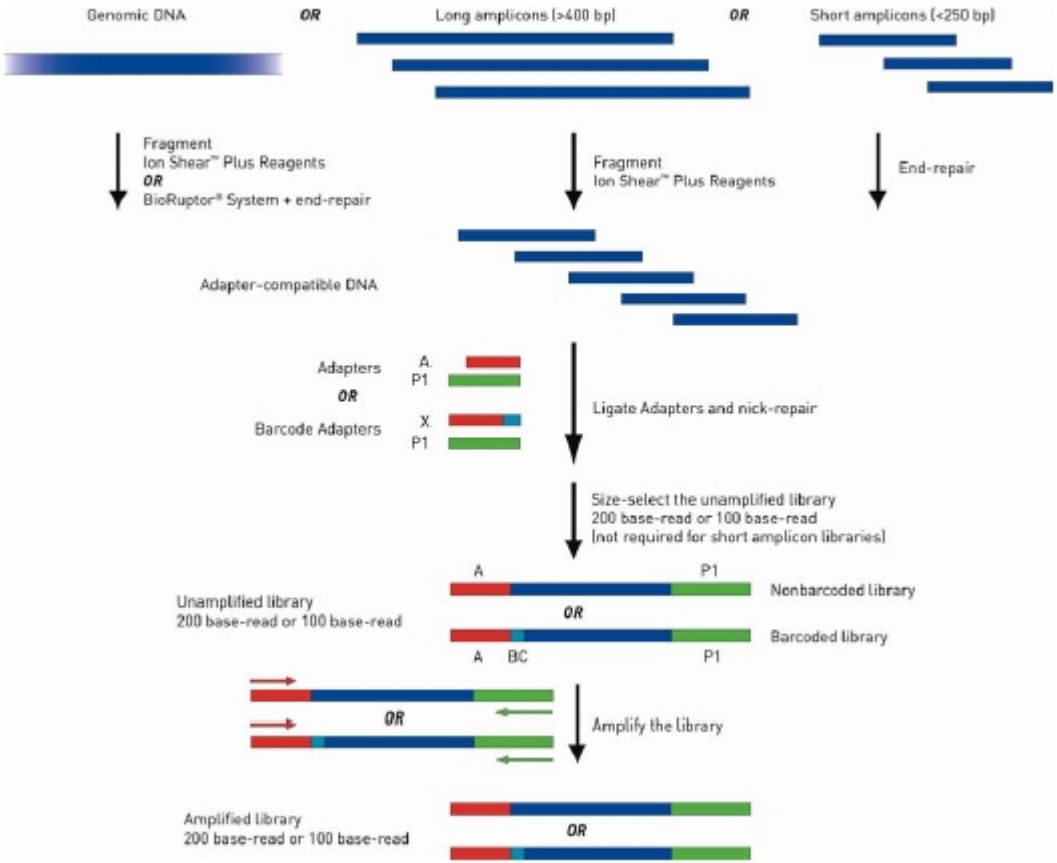






Shendure and Ji, Nat Biotech, 2008

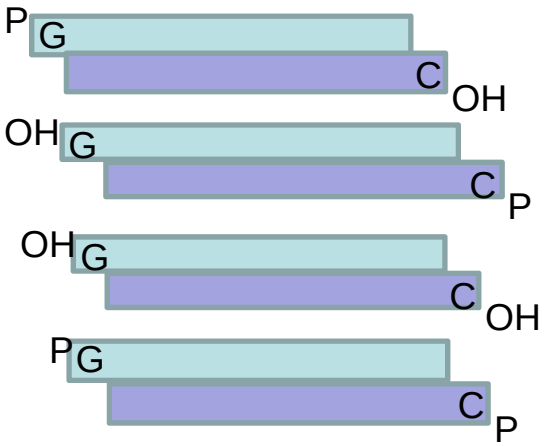
NGS data production steps



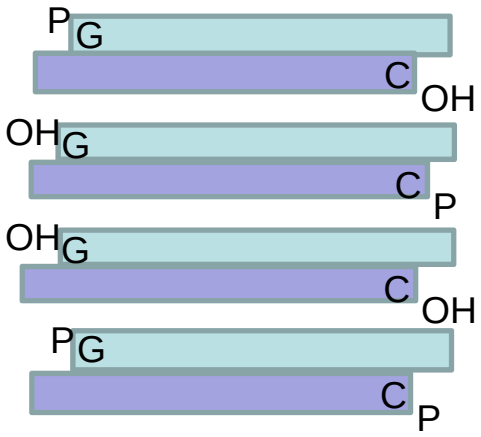
Shendure and Ji, Nat Biotech, 2008

What sorts of fragments results from the various library-generation methods?

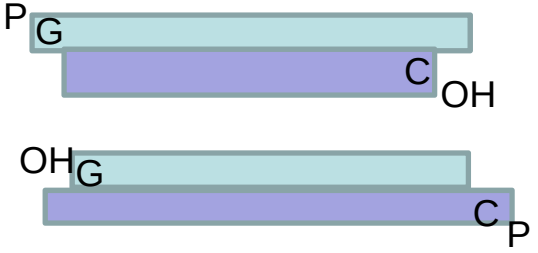
- Mechanical Shearing:
  - Ends: Blunt, 3' and 5' overhangs
  - Chemistry at ends: terminal 5' phosphate or hydroxyl, terminal 3' hydroxyl or phosphate
- PCR products
  - Often have a 3' overhang e.g. +A with Taq polymerase
- Nuclease Products – have specific overhangs



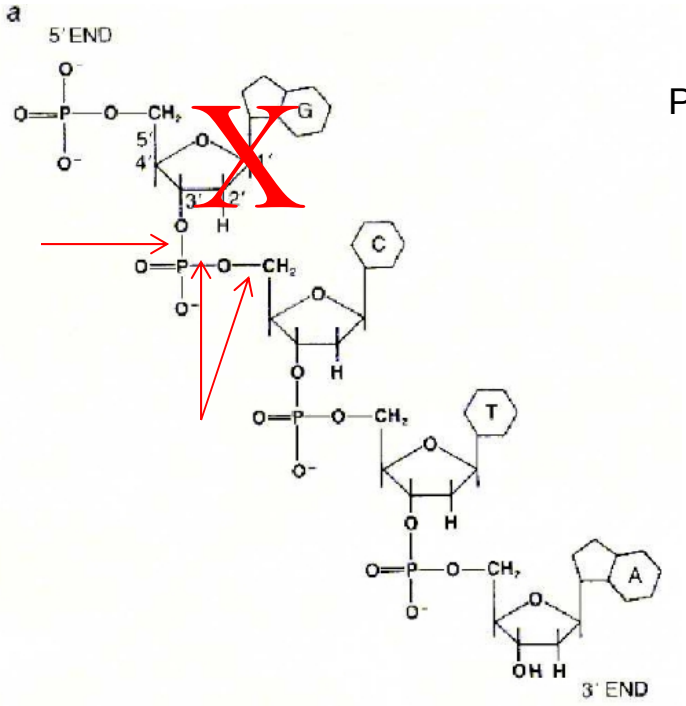
5’ overhangs



3’ overhangs



Assymmetric overhangs



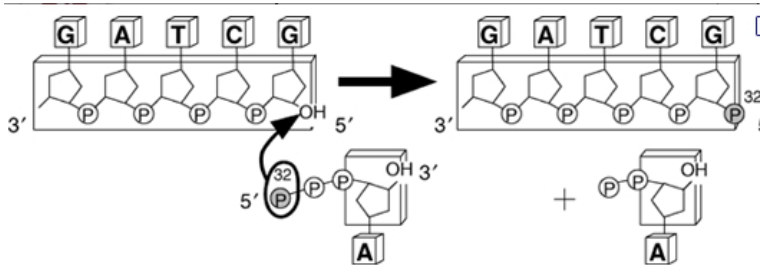
Blunt with assorted ends

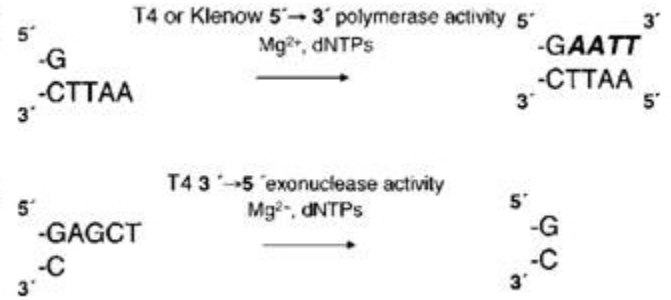
Terminology

- Polishing is to even up staggered ends
  - cut back (excise) with mung bean nuclease, bal31 nuclease
  - fill in with Klenow fragment of E coli DNA polymerase I or by T4 DNA polymerase
    - If there is exonuclease activity then excission will occur
- Replacing chemical groups allows polymerizaton and ligation to occur
  - Polynucleotide kinases will add a 5' phosphate group
  - Removal of a damaged nucleotide by Calf intestinal phosphatase will leave a 3'-OH

End-repair of sheared DNA

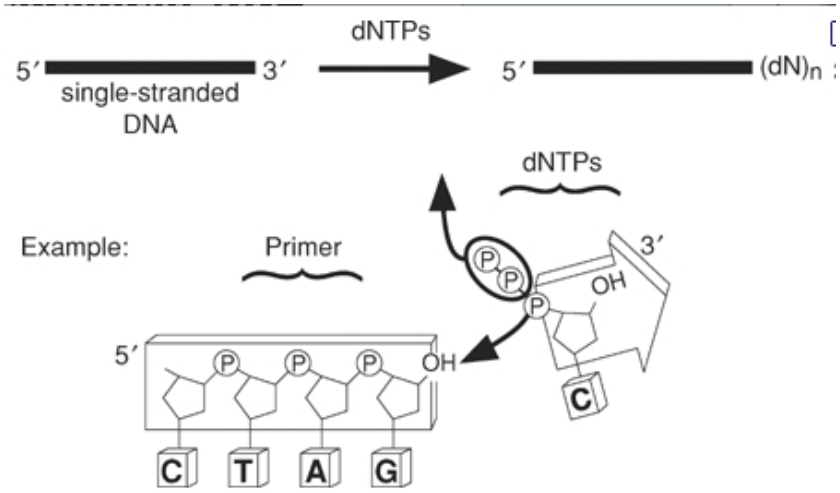
- A frequently used pair of enzymes for end-repair includes PNK and T4 DNA polymerase.
- PNK transfers a P_i from the gamma P_i of ATP to a 5'-OH of polynts (DNA or RNA, ds or ss).
    - It removes the 3'Pi from dNDPs and dNMPs.
        - It is used for end-labeling DNA and RNA
        - It is used to add a 5' Pi to oligonts to allow ligation
        - It is used to remove a 3' Pi
        - It can be bought in a form lacking the 3' phosphatase activity, which allows you to generate pNp to add to the 3' end of DNA or RNA by ligase
- The T4 DNA polymerase is used for gap filling and blunting. It has 3' → 5' but not 5' → 3' exonuclease activity.
    - It removes a 3' overhang
    - It fills in where there is a 5' overhang





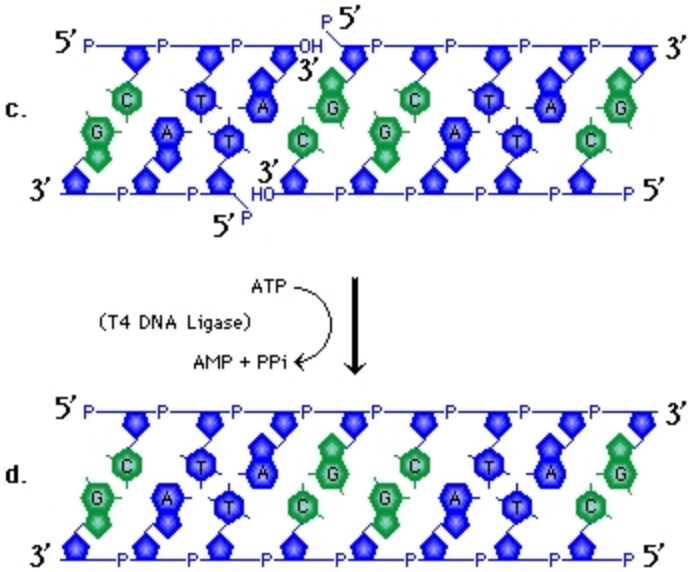
Enzymatic 3'-Phosphorylation

- Calf thymus TdT will add a 3'P_i to the 3'-OH of an oligonucleotide using a dTTP and a Co++ cofactor.
  - It can also add phosphonate groups and ddNTPS (terminators).



Enzymatic 5’-Phosphorylation

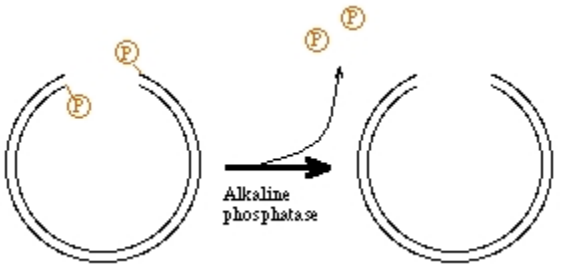
- To covalently link adaptors to target, there must be a 5’Pi and a 3’-OH that the ligase can seal.
- If the oligonucleotide does not have a 5’-Pi then you must add one.
- T4 polynucleotide kinase is most frequently used for this.

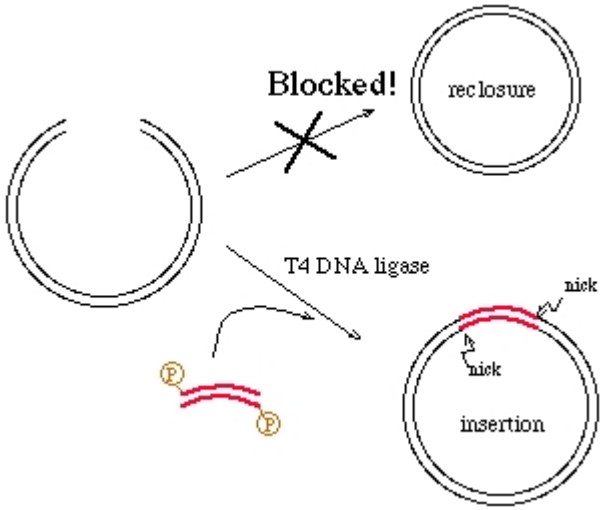


Engineering DNA ends:
http://escience.ws/b572/L6/L6.htm

5’-P_i De-phosphorylation

- If PCR products have 5’-Pi and 3’-OH available, then they can be ligated together – this is called a concatamer.
- To remove the 5’-Pi, you have two enzymatic choices:
  - Calf intestinal Phosphatase (CIP)
    - This is a robust enzyme – it can work in mixtures, but is hard to inactivate
  - Antartic Phosphatase (AP)
    - Is fairly robust to buffers, requires Zn++ cofactor and can be heat inactivated.



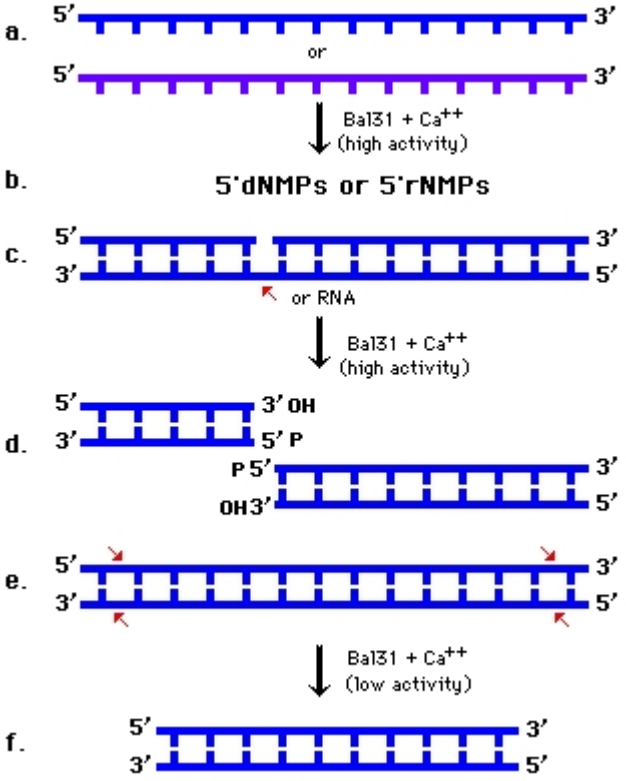


Exo- and Endo-nucleases

- T4 DNA polymerase +dNTPs is used to remove 3’ overhangs
- T4 DNA polymerase +dNTPs or the Llenow fragment and dNTPs or Mung Bean nuclease can be used to fill in a 5’ overhang
- Exonuclease I can be used to remove oligonucleotides post-PCR (also any ss product remaining).

Bal31 Nuclease

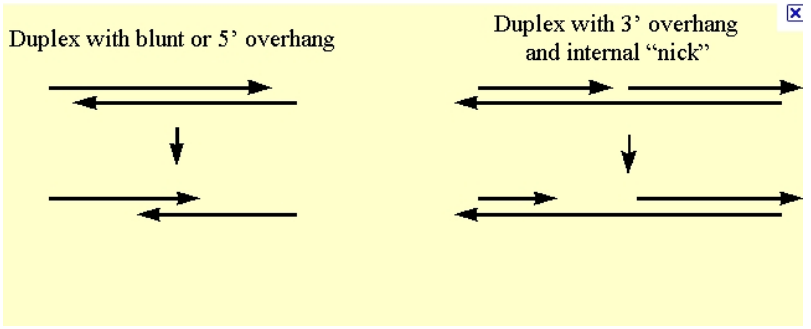
- Has 3' → 5' endonuclease activity
  - works on both dd and dsDNA
  - in the absence of a 5'-P_i
  - can initiate on DNA with 5'- and 3' extensions as well as blunt ends and nicked DNA.
  - It does not introduce nicks but will shorten the DNA product.
  - It will cleave at the opposite of nicks, gaps and ss regions of dsDNA.



Mung Bean nuclease

- A single-strand specific DNA and RNA endonuclease
- Degrade ss extensions from the ends of DNA and RNA molecules, leaving them blunt but ligation-ready.
- Mung Bean nuclease must be used if the goal is to chew back a 5' extension (DNA pol will not work).
- Mung Bean nuclease will also chew back a 3' extension but it is more likely to degrade ends that 'breathe' than Pol I, and also anything that leads to a bubble, or if nicks are present.
- It will not fill in anything.
- Also cleaves hair-pin loops.

http://www.ufrgs.br/depbiot/blaber/section1/section1.htm



dA-Tailing

Adding a dAMP to the 3’ end of a blunt DNA fragment prevents concatamer formation during subsequent ligation steps – and facilitates ligation to adaptors/cloning vectors with complementary dT overhangs.

It contains E coli DNA polymerase I that is 5’→ 3’ exonuclease minus. This is different from the Klenow fragment, which is 3’→ 5’ exonuclease deficient, but you can use a Klenow exo- mutant.

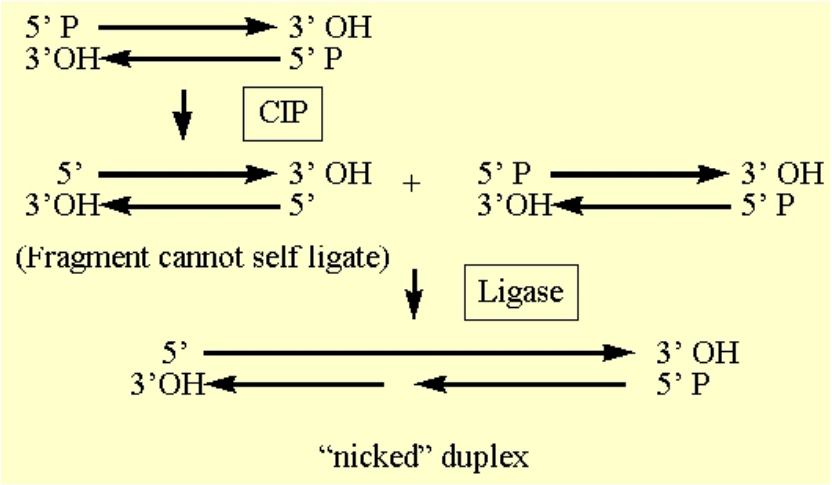
Ligation

For handling ligation of adaptors to target, there are several options

T4 DNA ligase will ligate cohesive and blunt ends and nicks in dsDNA.

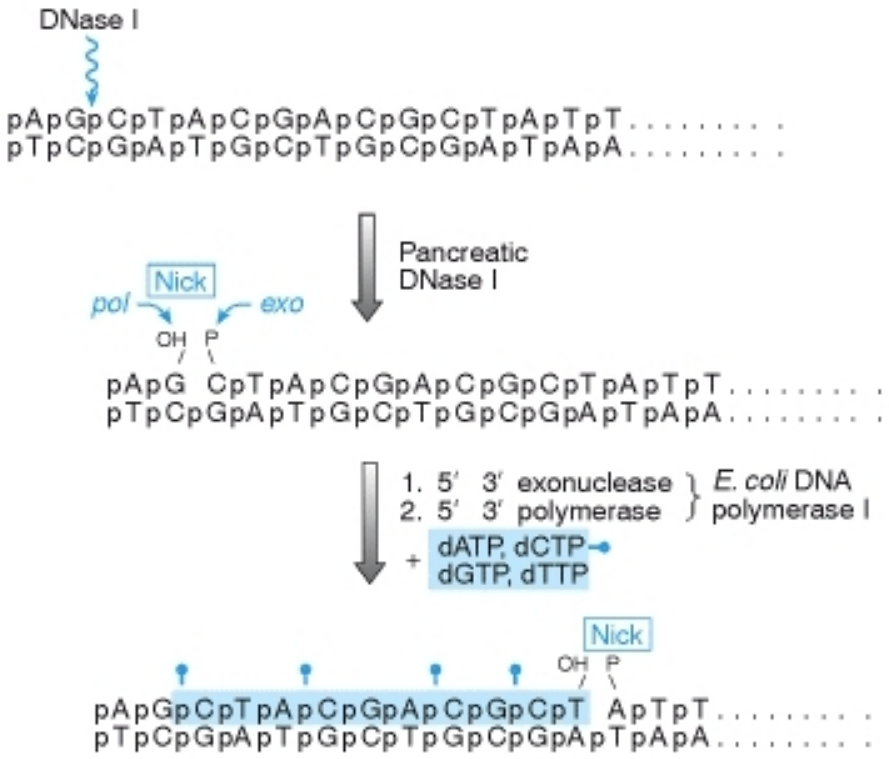
E coli DNA ligase will ligate cohesive ends and nicks in dsDNA.

Taq DNA ligase and 9°N DNA ligase are thermostable and will seal nicks in dsDNA.



What is nick-translation ?

- Nick-translation – the original use of the method was to introduce phosphate backbone cuts (no base or sugar is removed)
- Dnase I (low conc, requires Mg^{++}) leaves exposed 3'-OH and 5'-Pi.
- E coli polymerase synthesizes DNA complementary to the intact strand in a 5' → 3' direction, using the segment containing the 3'-OH of the nick as the primer.
  - This 'moves' the nick along the DNA strand in the 5' → 3' direction.
- The 5'→3' exonuclease activity of the DNA pol I removes the nucleotides in the direction of synthesis.
- If you have modified bases in the 'building block' pool they will be incorporated.



Why polish amplicons?

- PCR-generated fragments often have protruding 3’ ends.
- Each polymerase has its own pattern (e.g. Taq prefers to add ‘dA’ but this is not general to polymerases).
  - For example, the Klenow fragment of E coli DNA pol I has this TdT activity.
  - Pfu DNA polymerase does not have this activity – by adding it post-amplification, the 3’→5’ exonuclease activity will remove 3’ protruding bases.

Other useful enzymes

- T7 DNA polymerase is used to fill gaps but it will nibble in past the blunt-end stage.
- To remove 3' overhangs use Klenow, T4 DNA polymerase, Mung Bean Nuclease
- To remove 5' overhangs use Micrococcal Nuclease
- ExoSAP-IT is exonuclease I and shrimp alkaline phosphatase in combination
  - Exo I degrades ssDNA (so your primers, original template and orphan PCR template that is not ds)
  - SAP hydrolyzes dNTP → dNMP so it cannot be incorporated as a monomer. Since it works in PCR buffer it can be added at the end of a reaction. Inactivates with 80C for 15 min.
- Klenow fragment of E coli DNA polymerase I does not have the exonuclease peptide.
  - It can add <u>one</u> dAMP to the 3' end of a blunt DNA fragment – why is this useful?
- E coli DNA Ligase catalyzes the formation of a phosphodiester bond between a 5' Pi and a 3' –OH in duplex DNA having cohesive (sticky) termini but not blunt ends.
  - It cannot ligate RNA to DNA, or work on ssNA – it is used when you have sticky ends and you want to suppress blunt-end ligation.

Fragment Processing Steps

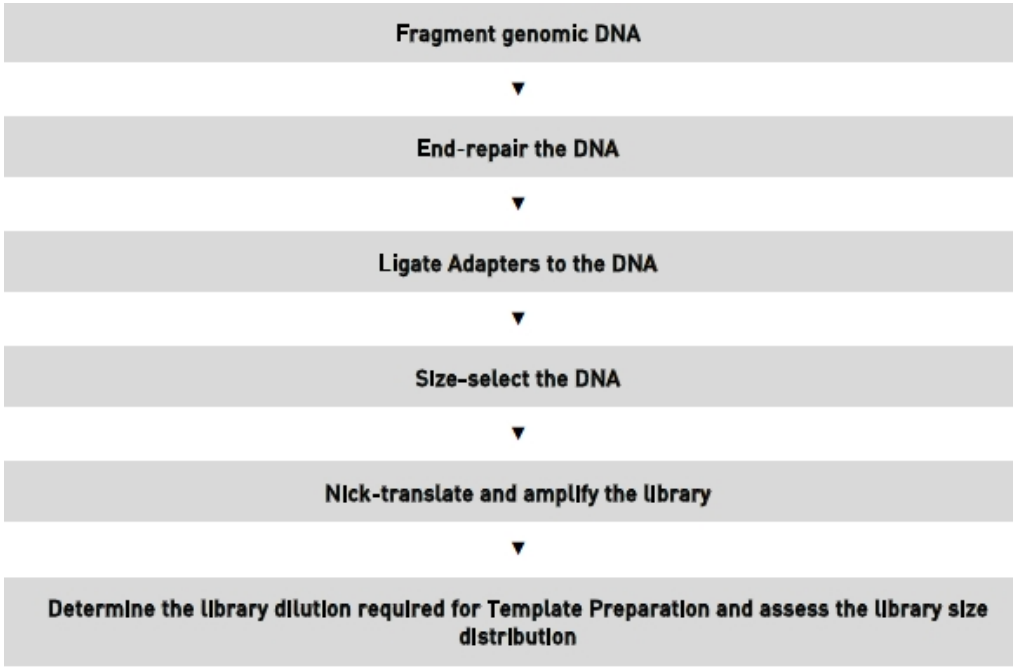
- Expected input: gDNA that has been mechanically sheared.
- We will then
  - end-repair
  - ligate to adaptors
  - nick-translate for completing the linkage.
  - The goal with the Ion Torrent is a library of fragment size 180-210 bp.
- Short amplicons are also suitable (length <150bp).
  - Alternatiely, the amplicons could be ligated together into long chains called concatamers, and then sheared.
    - Why?
  - End-repair is performed (why?)
  - Ligation to the platform-compatible adaptors
  - Nick-translation process as above.

Workflow and Components

- Contents

Component
5X End Repair Buffer
End Repair Enzyme
10X Ligase Buffer
DNA Ligase
Adapters
Platinum® PCR SuperMix High Fidelity
Library Amplification Primer Mix
E. coli DH10B Control DNA (5 µg)*
Low TE

Workflow



Generating Fragments

- Fragment DNA in the BioRuptor system
  - Suspend 1-10ug of DNA in 100-150ul volume in Low-TE buffer (pH 8.0).
  - Seal carefully, keep samples chilled at all times.
  - Use 30sec-On, 30sec OFF pulses, for 15 minutes, check temperature. Repeat 5 times.
    - Note – this protocol says NOT to spin out the samples in between.
    - Confirm a fragment range between 50-500bp by electrophoresis.

Components for end-repair

- End-repair the DNA
- Into a 1.5-mL Lo-bind Tube, pipette
  - 58ul of DNA
  - 100ul of nuclease-free water
  - 40ul of 5X-end-repair Buffer
  - 2ul of End-repair Enzyme Cocktail
  - Mix, quick-spin, incubate at RT for 20 minutes

Ion Fragment Library Kit

- If you are using amplicons in the end-repair process, you have to quantify the fragments so that the pooled amplicon concentration is 8.3nM.

Component	Volume
Pooled amplicon DNA (8.3 nM)	20 µL
Nuclease-free Water	59 µL
5X End Repair Buffer	20 µL
End Repair Enzyme	1 µL
Total	**100 µL**

Agencourt XP Beads

Using Agencourt AMPure XP beads, the initial reaction uses

1.8X bead:sample ratio → remove fragments <100bp

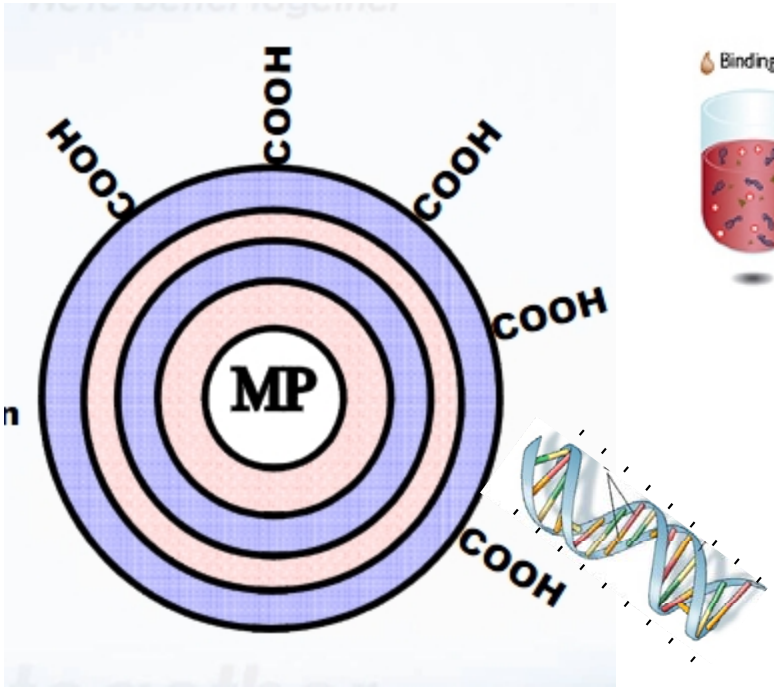
`1.6X bead:sample ratio→ remove fragments <150bp.

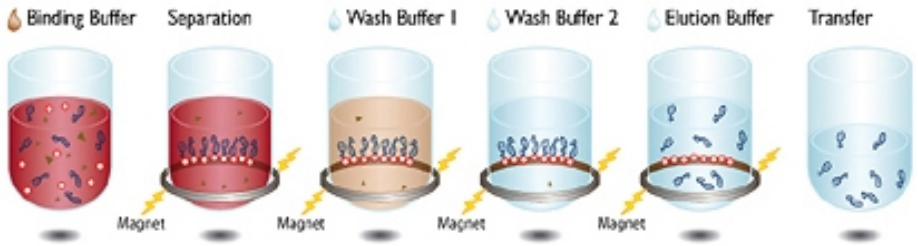
Salts, enzymes, dNTPs, primers are also removed.

AMPure XP beads

- How do AMPure beads work?
- They have a positive net charge (from carboxyl groups), so nucleic acids will stick to them. They are paramagnetic, so they will stick to magnets.
- The buffer used has polyethylene glycol 8000 (PEG8) in it and salt (small polynts need higher salt to bind strongly to the beads).
  - PEG has been used to precipitate DNA since 1975 (Lis and Schleif)
  - Higher mw DNA precipitates at lower conc of PEG.
  - In 6.5% PEG at 0.55M the >700bp DNA was ppt and at 1.1M NaCl DNA> 375bp was ppt
  - At 12% PEG, 0.55M ppt >125bp and at 1.1M NaCl >80bp.
- The Ampure buffer is 1.25M NaCL, 7% PEG8000.
  - A free agent noted that using 10% PEG 8000 with 1.25M NaCl and 10mM MgCl2 gave ppt results similar to the yield with Ampure XP beads. If 7% PEG was used a lot of 50 and 100bp DNA was lost but the 200 and greater was nearly quantitative.
- The beads have a layer of magnetite sealed with a polymer that is coated with carboxylic acid groups.
  - Binding capacity is supposed to be ~2ug/100ug of beads.

Intermediate Clean-up: Paramagnetic Beads





PippenPrep from Sage Science

Target size-selection is a critical aspect of sequencing library preparation.

Some size-selecton modes inlcude use of Agencourt AMPure XP beads, the E-gel Invitrogen platform, gel extraction, and the Pippen Fractionator.

Gel-based Size Selection

- Size select the DNA using the Pippin Prep instrument.
- Select for the 180-210bp range (+/- 20bp around the mean).
- Use the 2% agarose gel cassette.
- Use all 30ul of the ligated DNA to load the gel.
- At the end, when you have removed the eluant from the well, check the volume, and adjust to 60ul with nuclease-free water.

Size Fractionation with the Pippen Prep

- An automated gel extraction device.
- Electrophoresis is used to size-*fractionate* the product
- Each lane splits into 2 termini: elution and waste.
  - Designate a capture range
  - Provide a calibration ladder
  - Targets are eluted from the gel matrix into a buffer chamber.
- Four samples and 1 ladders can be run in parallel.
- For a target length of ~200bp
  - Run time averages 70 min.
  - Use the 2% agarose gel cassette (100-600bp range)
  - Use ladder marker B1.
- Expect ~50% sample loss (for down-stream planning purposes).

Pippin Prep – other features

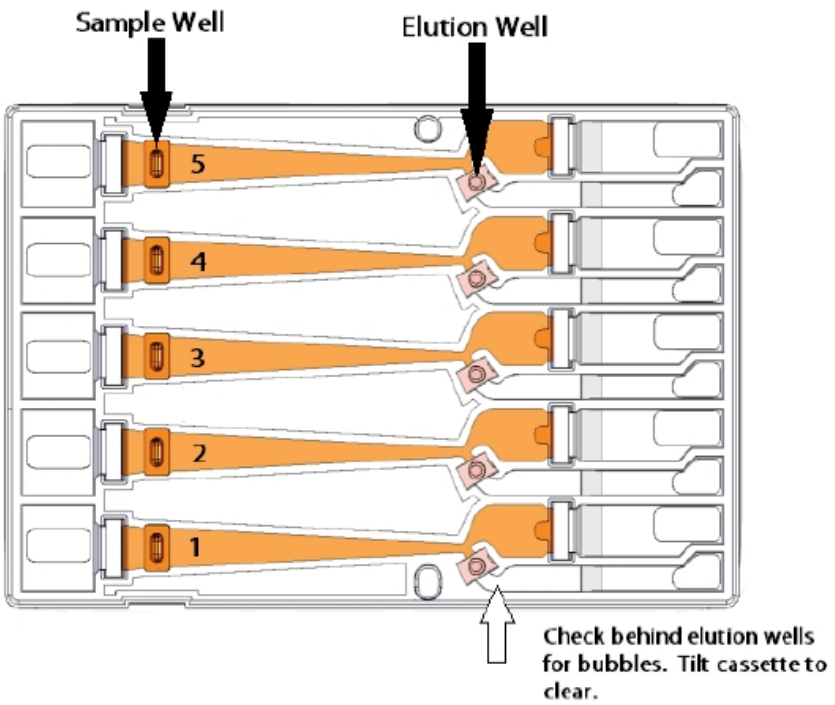
- Each lane is physically separated from the others
  - no cross-contamination.
- Gel is pre-stained
  - Lanes contain EtBr\
  - Since the markers are known by the analysis system you can input the desired size range
  - The system will collect the fraction you request.
- Components:
  - Instrument has epifluorescence detector, electrophoresis requirements, and on-board computer
    - also external monitor, mouse and keyboard.
  - Gel Cassettes include gel and buffer
  - Software

Pippin Prep – sample generalities

- Load per lane:
  - 50ng – 10ug of sheared gDNA
    - volume of 30ul.
  - 5ng-4ug of amplicon
- Prep: add 10ul of thawed loading solution to each sample
- Mix by vortexing
- Quick-spin, load.

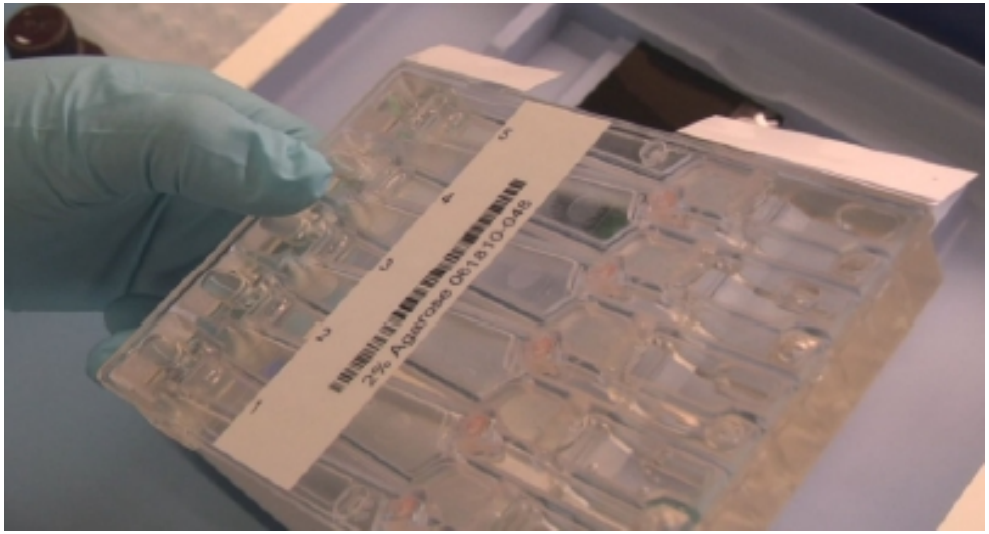
The Cassette

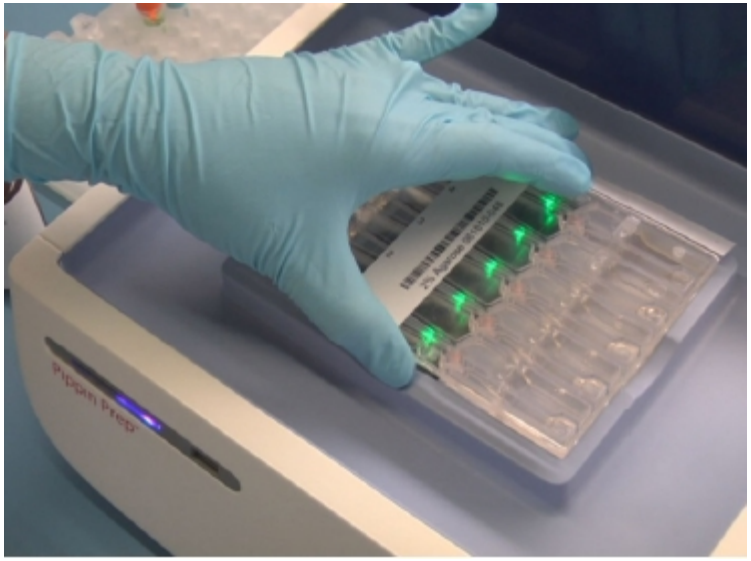
- Prepare the cassette – discard the filter paper.
- Make sure there is buffer in chambers at both ends (50% full).



Inserting the cassette

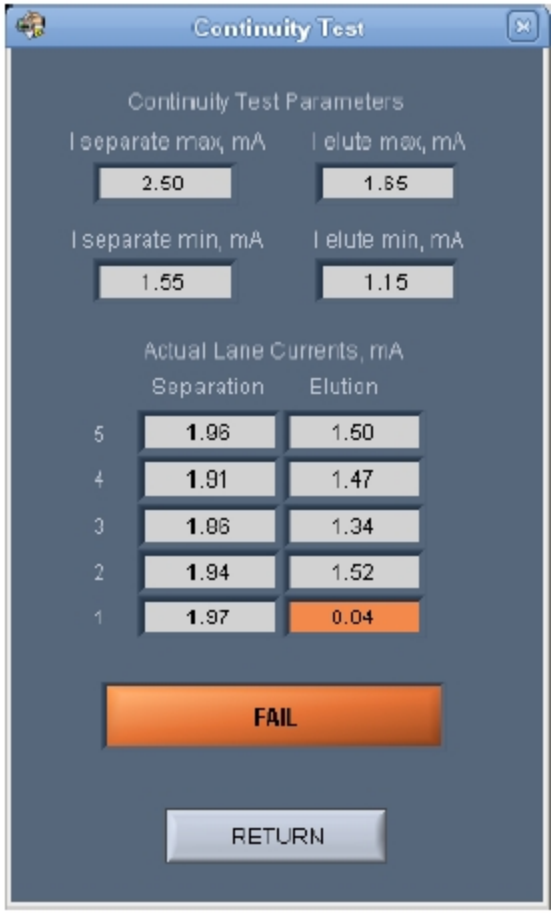
- Prepare the cassette – point the adhesive tabs away and tilt, to check for bubbles behind the elution wells.
- Place cassette in in tray – remove tape seals (2) by pulling slowly towards you.



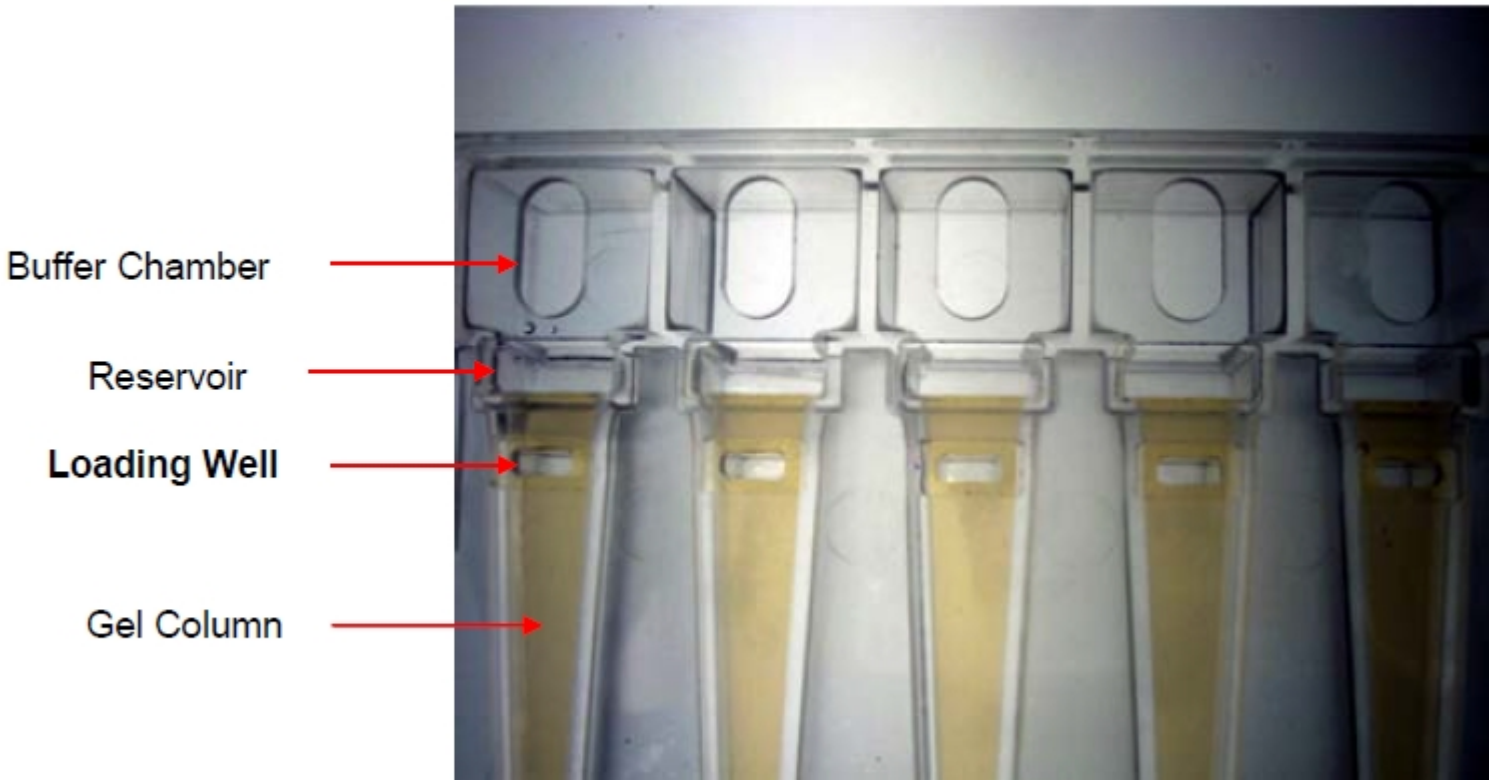


Pippin Prep Interface

- Test the cassette

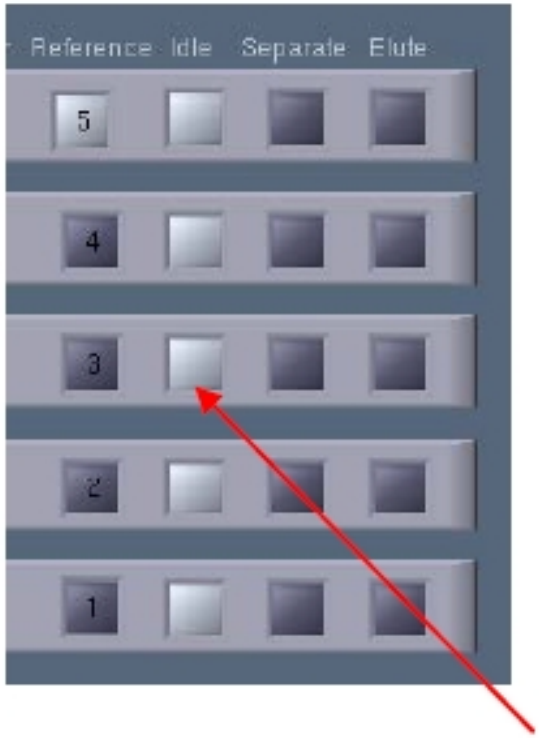


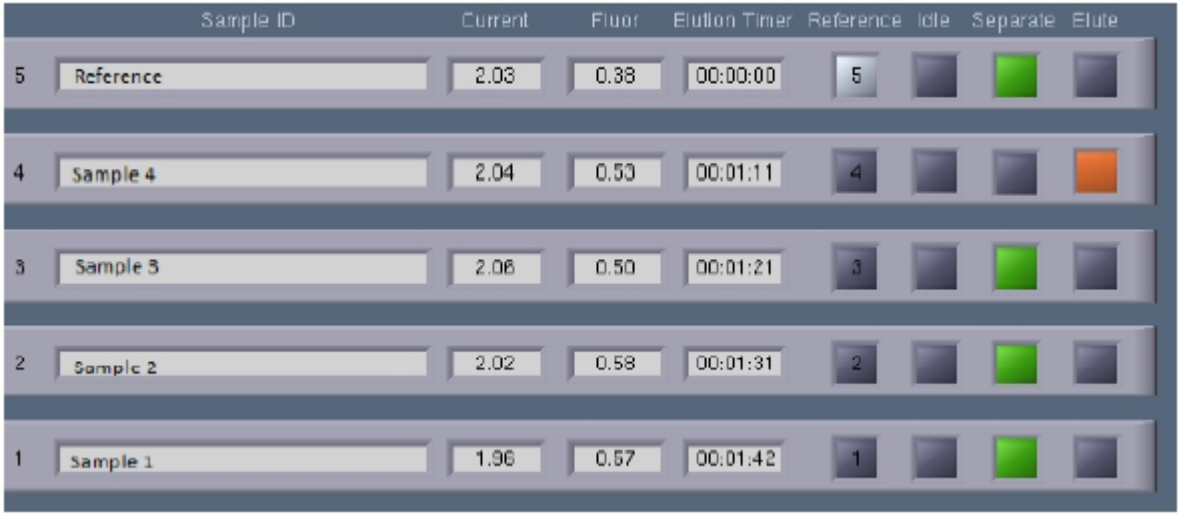
Load the Lanes



- Load the cassette – remove 40ul, (~20 should be left). Add 40 ul of the Marker to the chosen lane, then the samples.

Label the lanes, feedback on status



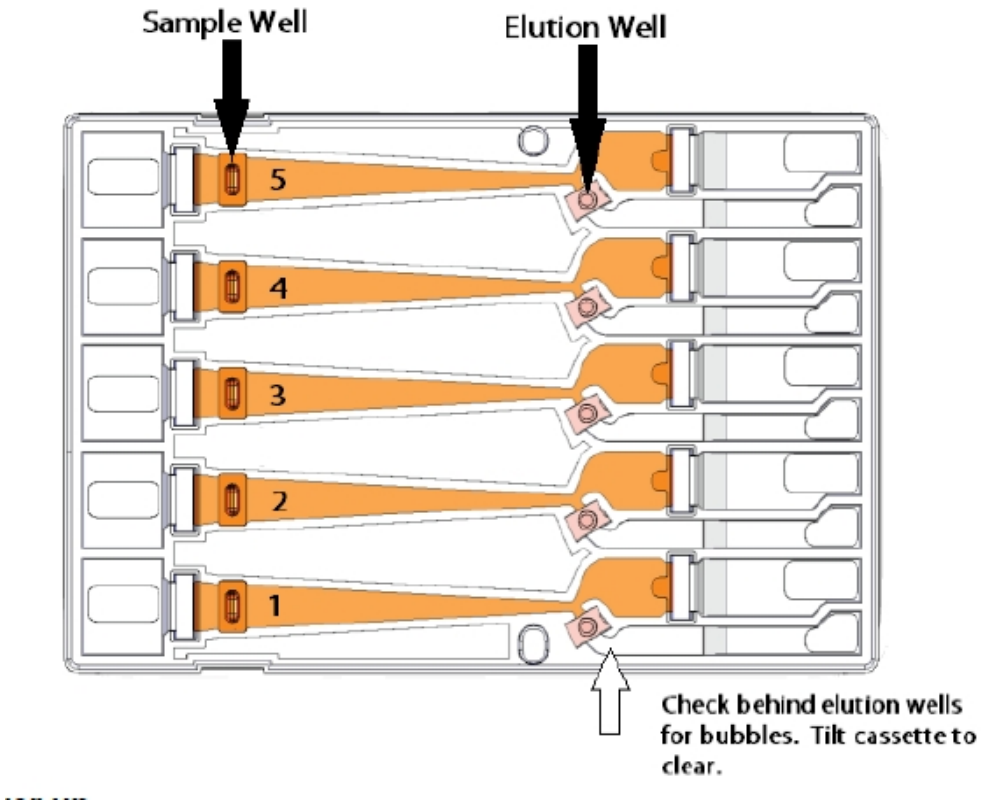


Grey- Marker, Green – running, orange – eluting.

- Load the cassette – remove 40ul, (~20 should be left). Add 40 ul of the Marker to the chosen lane, then the samples.

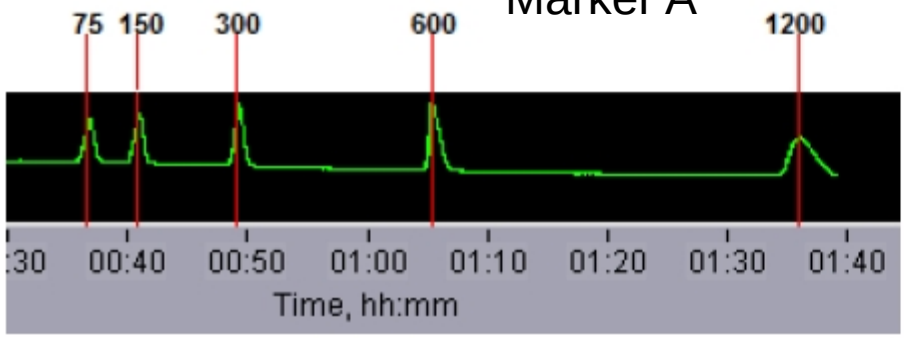
Pippin Prep

- Once you are done eluting the lane returns to separation mode.
- Collect the desired sample from the elution well.



Pippin Prep cassette selection guide

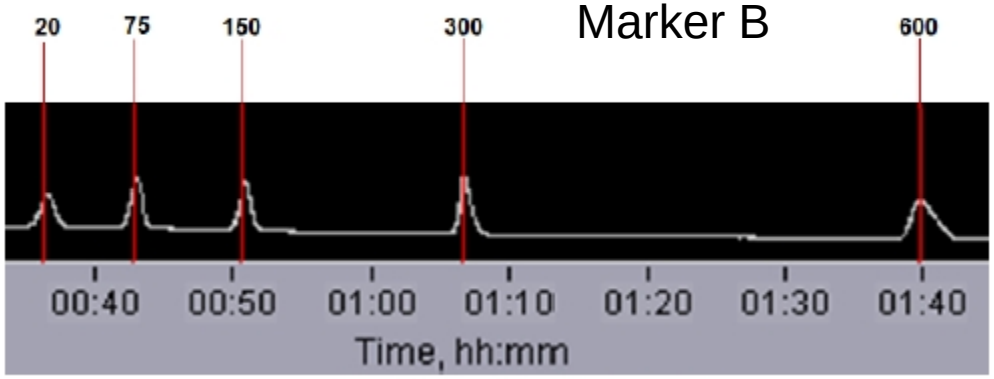
Marker A



1.5% Marker A Reference DNA Markers (bp)

Size, bp	Time to detector (minutes)*	Time to collect min. range (minutes)*
75	35	NA
150	40	NA
300	48	53
600	64	64
1200	95	105
1500	110	130**

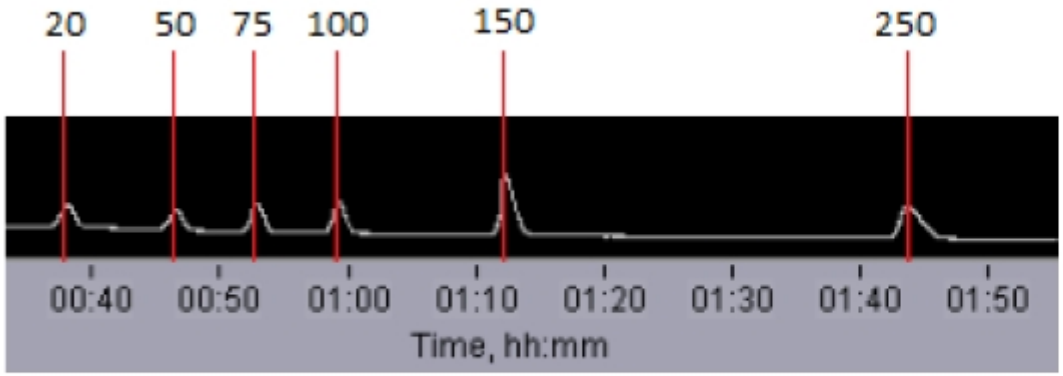
Marker B



Reference DNA, 2% Gel (bp)

Size, bp	Time to detector (minutes)*	Time to collect min. range (minutes)*
20	37	NA
75	43	NA
150	50	57
300	65	73
600	94	111

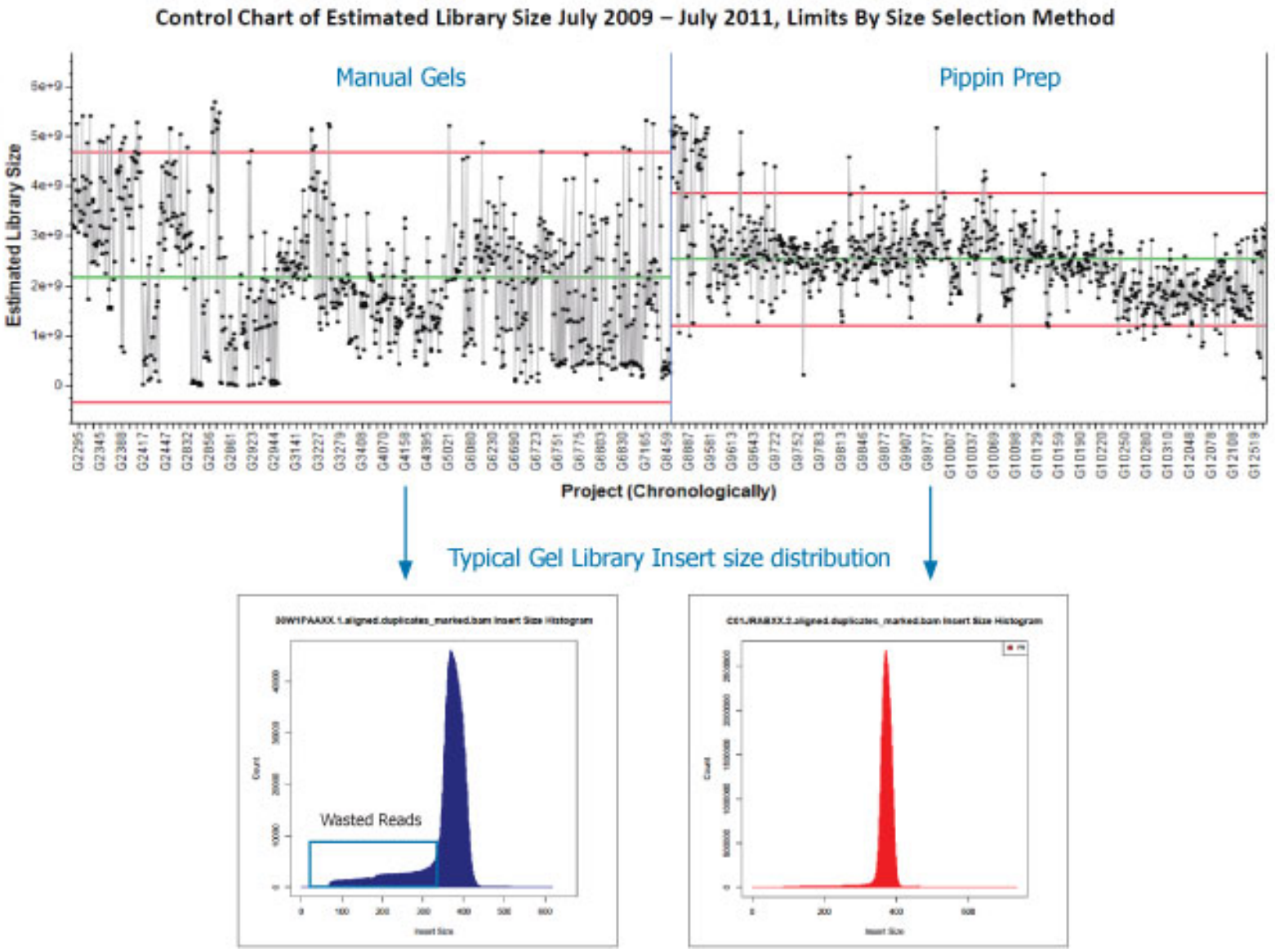
Cassettes cont.



Reference DNA, 3% Gel (bp)

Marker C

Size, bp	Time to detector (minutes)*	Time to collect min. range (minutes)*
20	38	NA
50	47	NA
75	52	59
100	59	67
150	73	82
250	104	118
300	112	131


Control Chart of Estimated Library Size July 2009 – July 2011, Limits By Size Selection Method
Manual Gels
Pippin Prep
Estimated Library Size
Project (Chronologically)
Typical Gel Library Insert size distribution
Wasted Reads

One adaptor is not ‘sealed’ so it would be lost in PCR

- Nick-translate and amplify the library.
- Take 40ul of your size-selected DNA, and use as shown, but split that volume between TWO PCR tubes.
- The cycle as below:

Component	Volume
Platinum® PCR SuperMix High Fidelity	200 µL
Library Amplification Primer Mix	10 µL
Size-selected DNA	40 µL
Total	**250 µL**

Stage	Step	Temperature	Time
Holding	Nick translation	72°C	20 min
Holding	Denature	95°C	5 min
Cycling (2–10 cycles)*	Denature	95°C	15 sec
	Anneal	58°C	15 sec
	Extend	72°C	1 min
Holding	—	4°C	∞

Post-PCR

- The PCR product is the basis of the sequencing library
- QC steps?