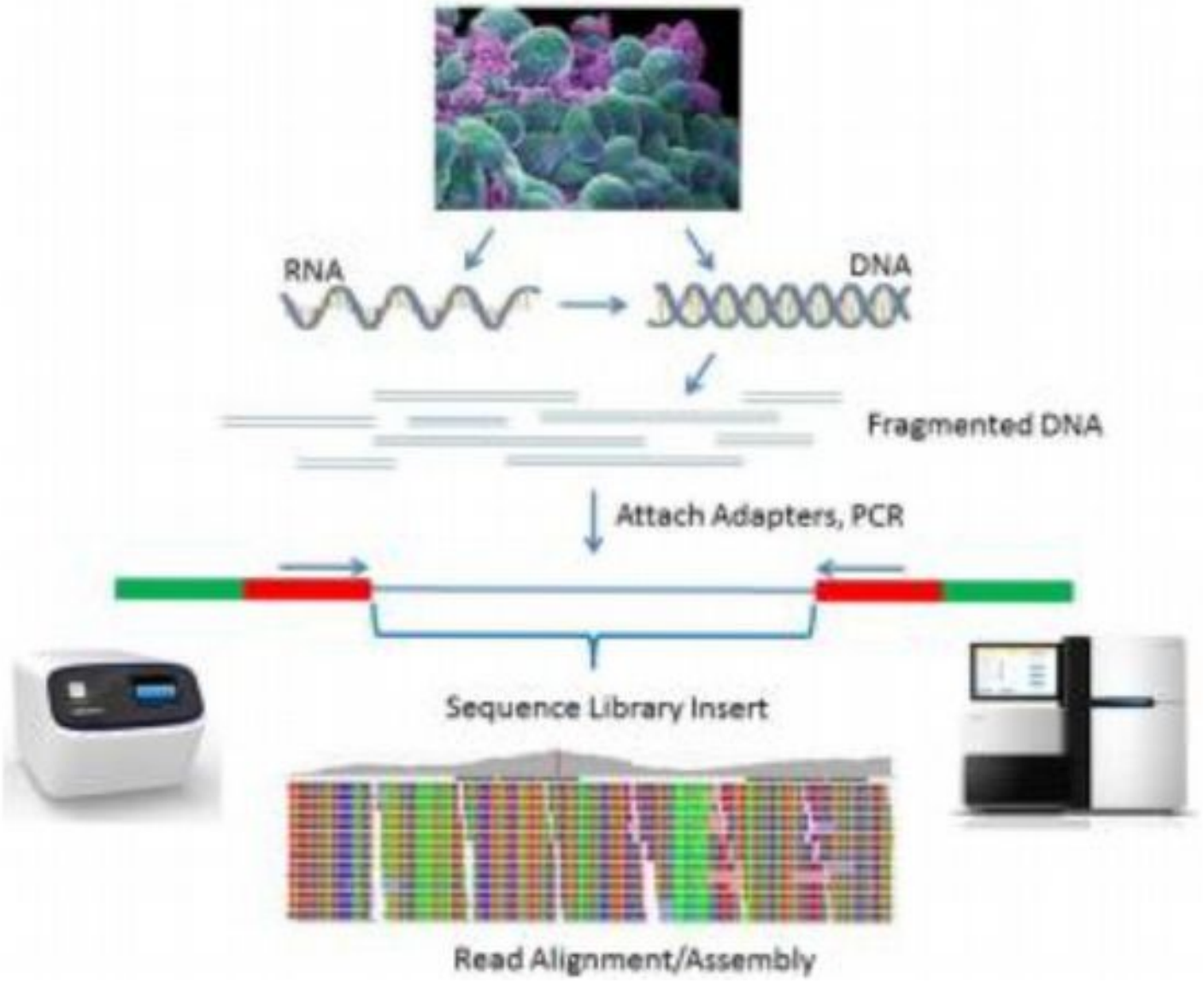
Library construction for next-generation sequencing: Overviews and challenges

- During this time, as sequencing technologies have improved and evolved, so too have methods for preparing nucleic acids for sequencing and constructing NGS libraries.
- For example, NGS library preparation has now been successfully demonstrated for sequencing RNA and DNA from single cells

- Fundamental to NGS library construction is the preparation of the nucleic acid target, RNA or DNA, into a form that is compatible with the sequencing system to be used.
- Now we compare and contrast various library preparation strategies and NGS applications, focusing primarily on those compatible with Illumina sequencing technology.
- However, it should be noted that almost all of the principles discussed in this review can be applied with minimal modification to NGS platforms developed by Life Technologies, Roche, and Pacific Biosciences.


RNA
DNA
Fragmented DNA
Attach Adapters, PCR
Sequence Library Insert
Read Alignment/Assembly

Principal steps for NGS library preparation

- In general, the core steps in preparing RNA or DNA for NGS analysis are:
- (i) fragmenting and/or sizing the target sequences to a desired length,
- (ii) converting target to double-stranded DNA,
- (iii) attaching oligonucleotide adapters to the ends of target fragments, and
- (iv) quantitating the final library product for sequencing.

Fragmentation/Size selection

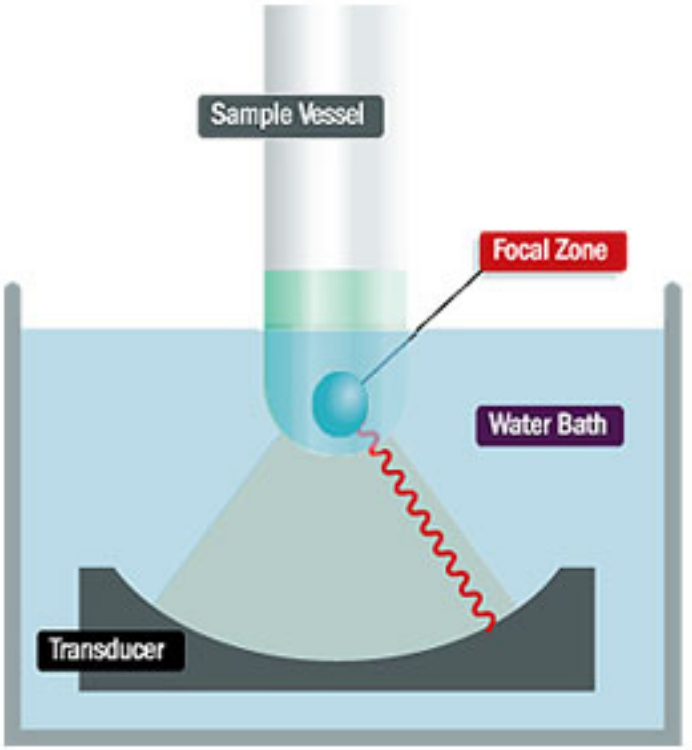
- The size of the target DNA fragments in the final library is a key parameter for NGS library construction.
- Three approaches are available to fragment nucleic acid chains:
  - physical,
  - Enzymatic,
  - chemical.

DNA fragmentation: physical methods

1) Acoustic shearing

The Covaris® instrument (Woburn, MA) is an acoustic device for breaking DNA into 100-5kb bp. Covaris also manufactures tubes (gTubes) which will process samples in the 6-20 kb for Mate-Pair libraries.

http://covarisinc.com/applications/dnarna-shearing



DNA fragmentation: physical methods

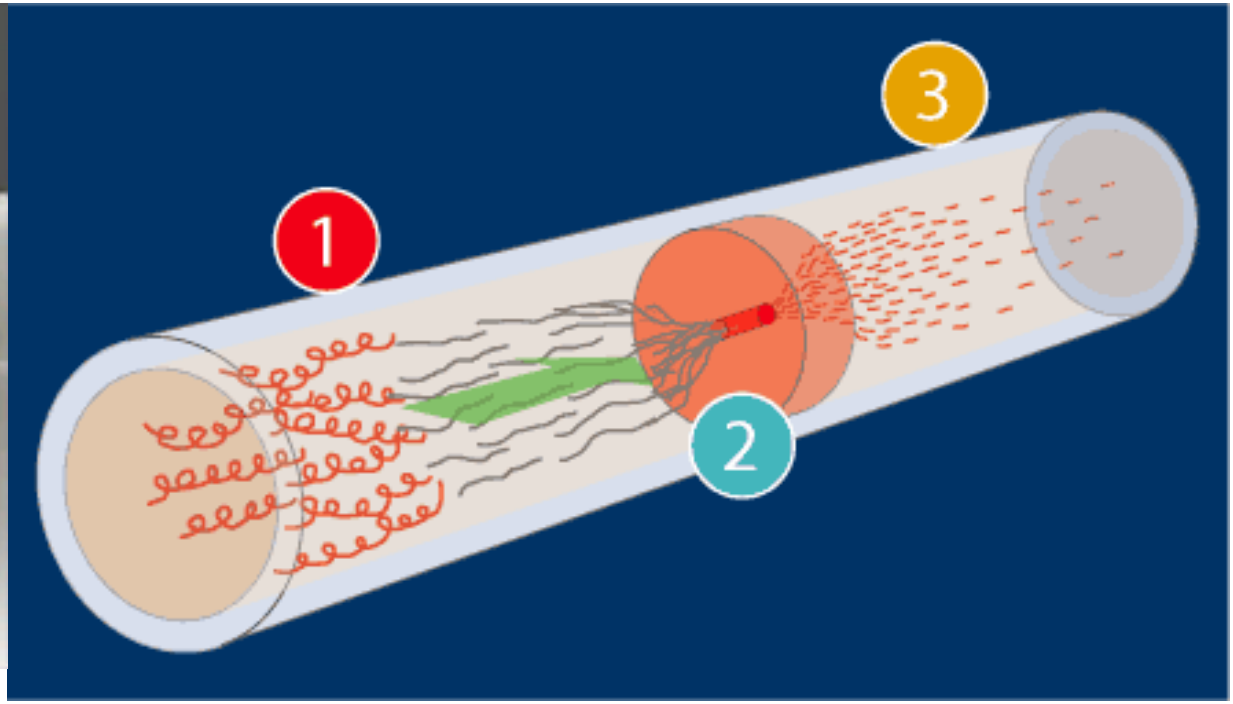
- 2) Sonication

  The Bioruptor® (Denville, NJ) is a sonication device utilized for shearing chromatin, DNA and disrupting tissues. Small volumes of DNA can be sheared to 150-1kb in length.

DNA fragmentation: physical methods

- 3) Hydrodynamic shear

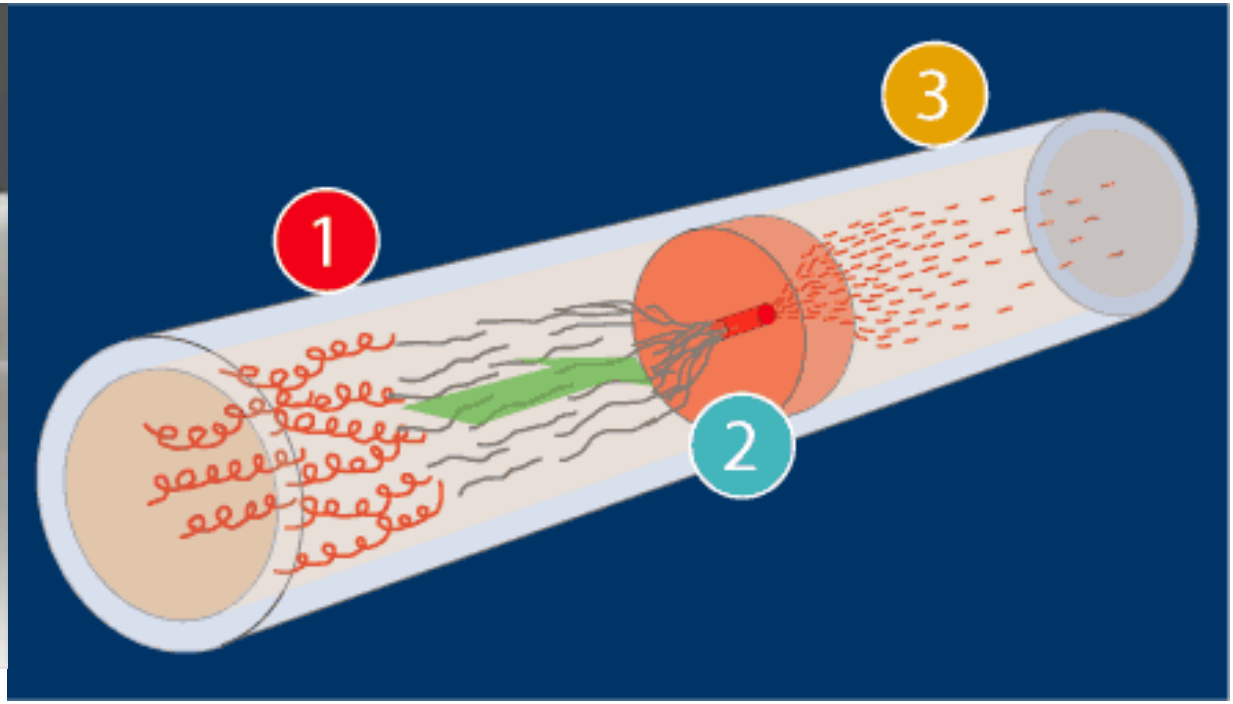
  Hydroshear from Digilab (Marlborough, MA) utilizes hydrodynamic forces to shear DNA.



DNA fragmentation: physical methods

- 3) Hydrodynamic shear

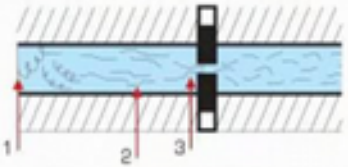
  Hydroshear from Digilab (Marlborough, MA) utilizes hydrodynamic forces to shear DNA.



Data*

The Mechanism

1. DNA in solution is passed through a tube with an abrupt contraction.

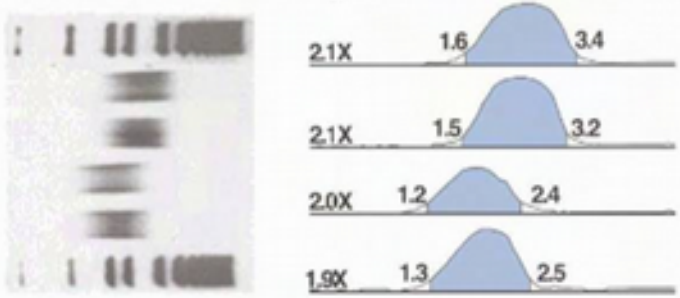


2. As it approaches the contraction, the fluid accelerates to maintain the volumetric flow rate through the smaller area of the contraction.

3. During this acceleration, drag forces stretch the DNA until it snaps. The DNA fragments until the pieces are too short for the shearing forces to break the chemical bonds. The flow rate of the fluid and the size of the contraction determine final DNA sizes.

*Shearing examples shown are pre-cloning data.

Size distribution is tight and consistent [1]

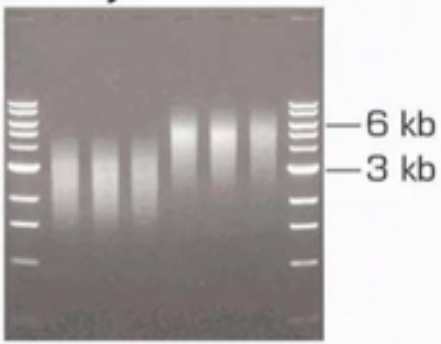


Duplicate DNA samples were sheared at two different flow rates for 20 iterations.

$$\text{size ratio} = \frac{\text{upper size range}}{\text{lower size range}}$$

[1] Thorstenson, Y., Hunicke-Smith, S., Oefner, P., Davis, R. 1998. An Automated Hydrodynamic Process for Controlled, Unbiased DNA Shearing. Genome Research, 8, 848-855.

Consistency of shearing across multiple users and days

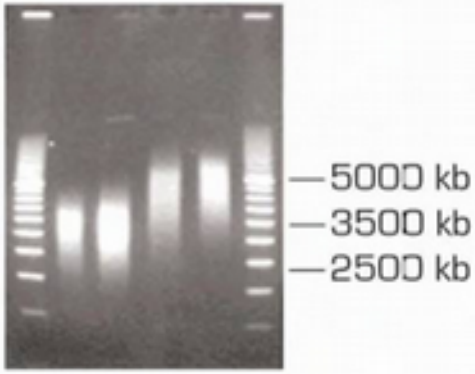


1% agarose gel run at 100V for 1 hour.
All samples taken from same stock of DNA.
Sheared samples: 2 µg/100 µl of Lambda DNA.

Lane	Speed	User	Day
2	10	A	X
3	10	B	Z
4	10	C	X
5	14	A	X
6	14	B	Z
7	14	C	X
1,8 1 kb ladder			

Users
A: Experienced User X: Day 1
B: Intermediate User Z: Day 2
C: First Time User

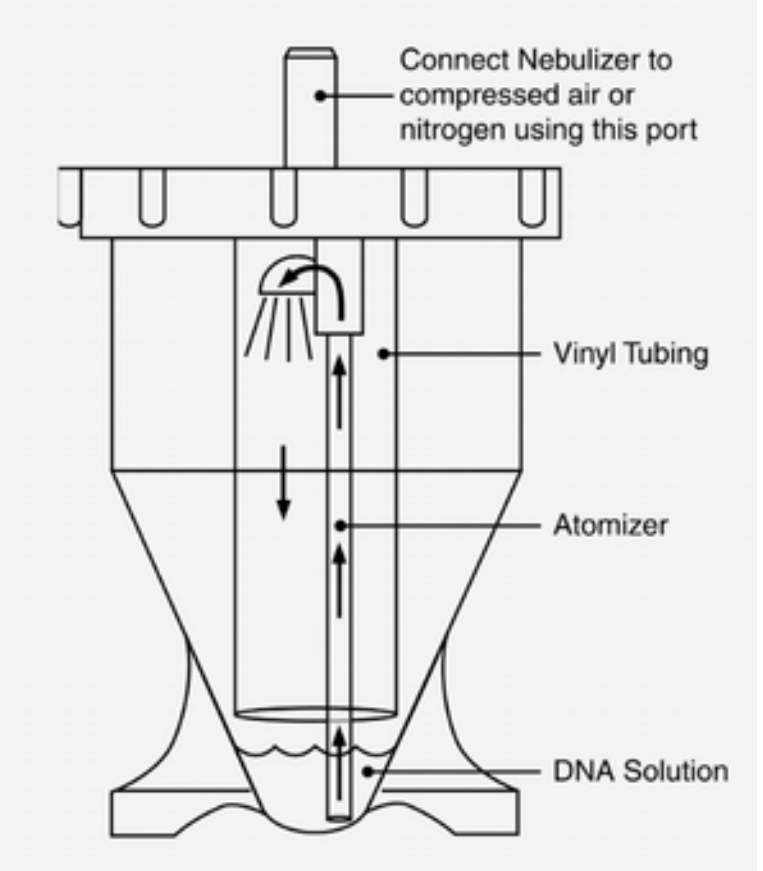
Effect of DNA concentration on fragment size



Lane	Speed	Lambda DNA
2	10	2 µg/200 µl
3	10	50 µg/200 µl
4	14	2 µg/200 µl
5	14	50 µg/200 µl
1,6 500 bp ladder		

1% agarose gel run at 105V for 1 hour.
Loaded 0.125 µg of sample per lane.
Sample source: Lambda DNA.

- Nebulizers (Life Tech, Grand Island, NY) can also be used to atomize liquid using compressed air, shearing DNA into 100-3kb fragments in seconds.



DNA fragmentation: enzymatic Methods

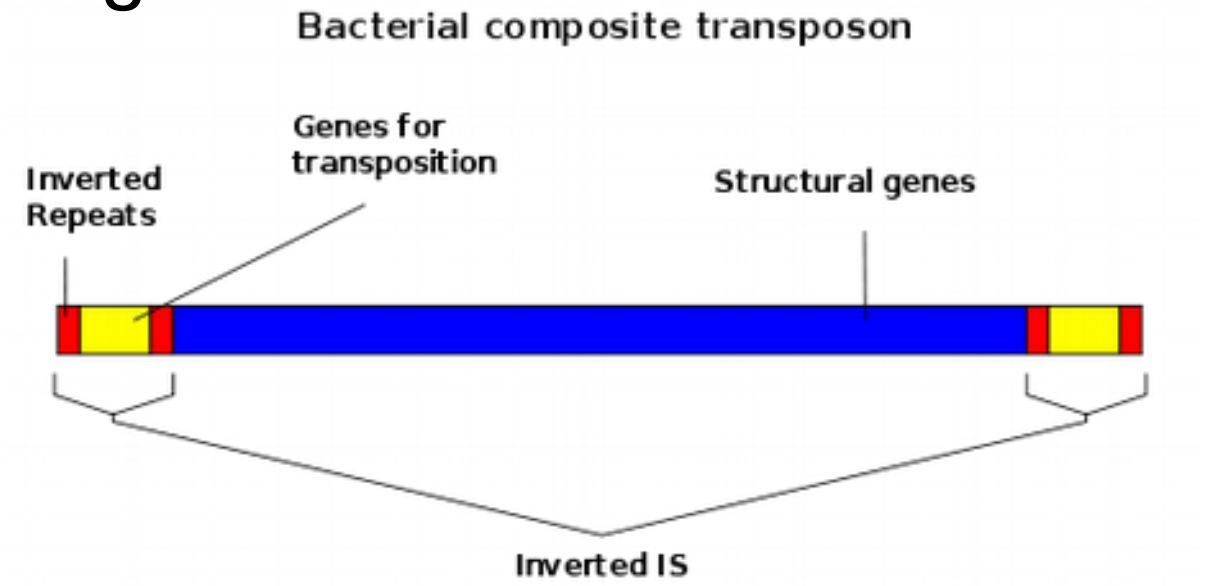
- DNase I or other restriction endonuclease, non-specific nuclease

  maltose binding protein (MBP)-T7 Endo I and a non-specific nuclease Vibrio vulnificus (Vvn), NEB's (Ipswich, MA) Fragmentase

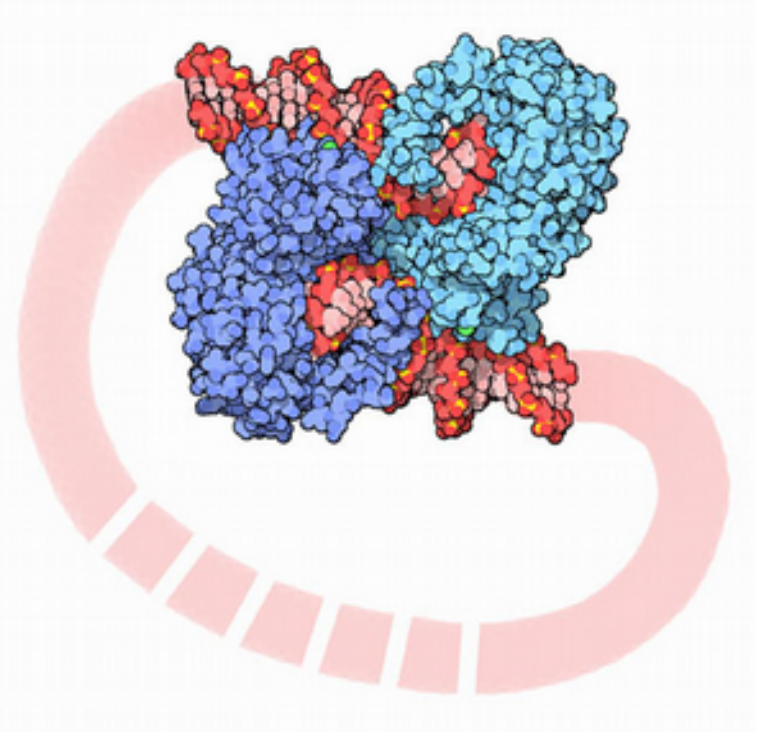
- one randomly generates nicks on dsDNA and the other recognizes the nicked site and cuts the opposite DNA strand across from the nick, producing dsDNA breaks. The resulting DNA fragments contain short overhangs, 5´-phosphates, and 3´-hydroxyl groups.
- https://www.google.com/patents/EP2393940A1?cl=en

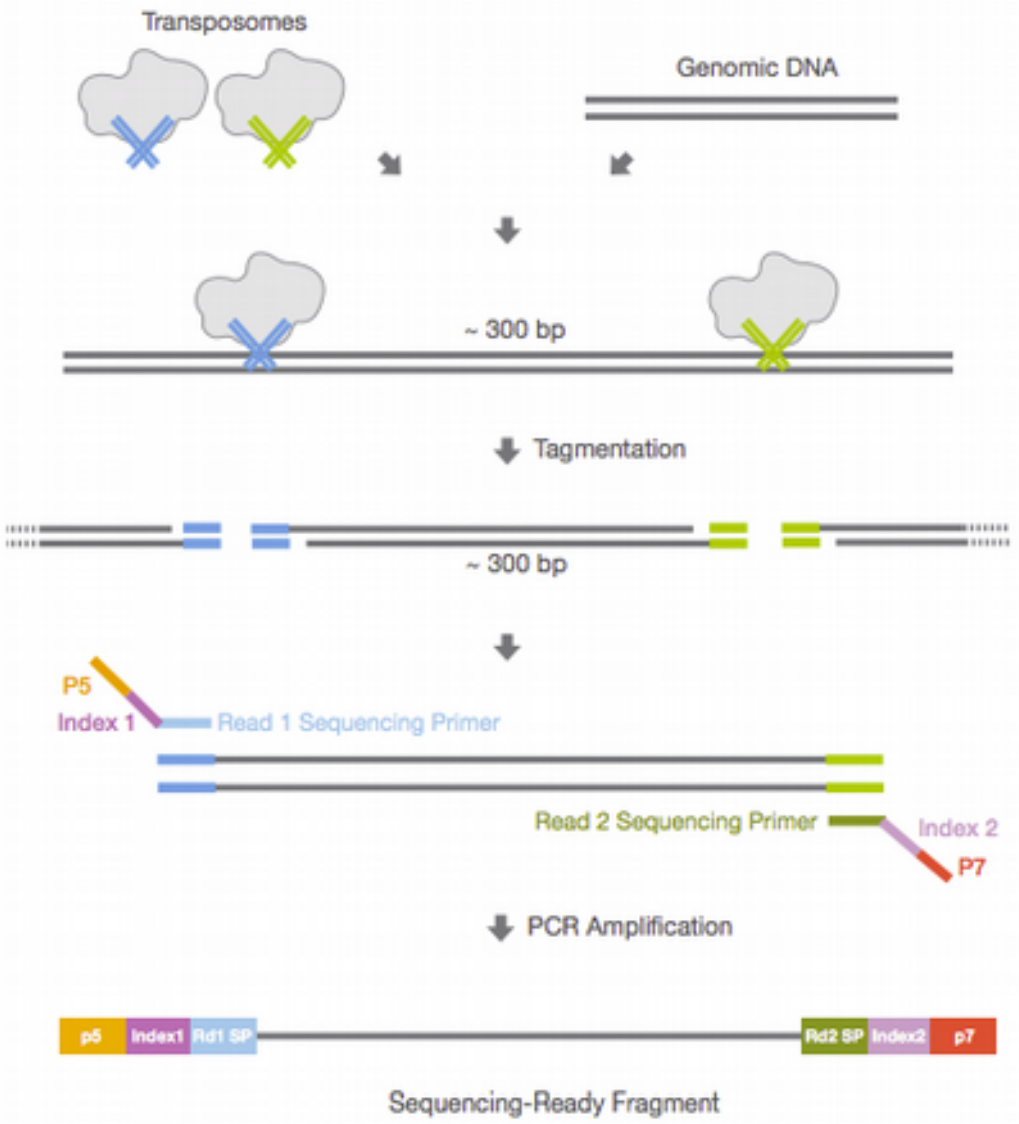
DNA fragmentation: enzymatic Methods

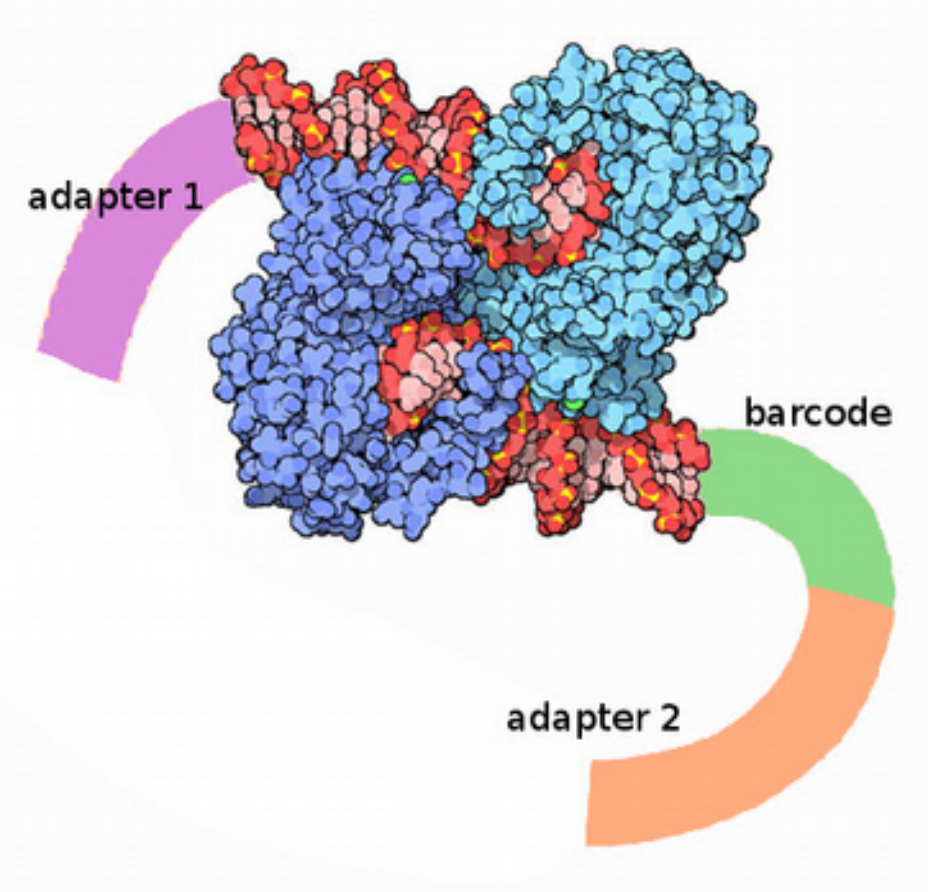
- A transposon is a little piece of DNA that copies itself around inside the genome of an organism, via an enzyme called transposase. Here's what the genetic element looks like :
-



- Transposase binds the element at the inverted repeats on either end, and coils it into a loop. Then it cuts the DNA at the inverted repeats, and the complex floats away. When it’s complexed to DNA, transposase grabs the DNA like this :




Transposomes
Genomic DNA
~ 300 bp
Tagmentation
~ 300 bp
P5
Index 1
Read 1 Sequencing Primer
Read 2 Sequencing Primer
Index 2
P7
PCR Amplification
p5
Index1
Rd1 SP
Rd2 SP
Index2
p7
Sequencing-Ready Fragment


adapter 1
barcode
adapter 2

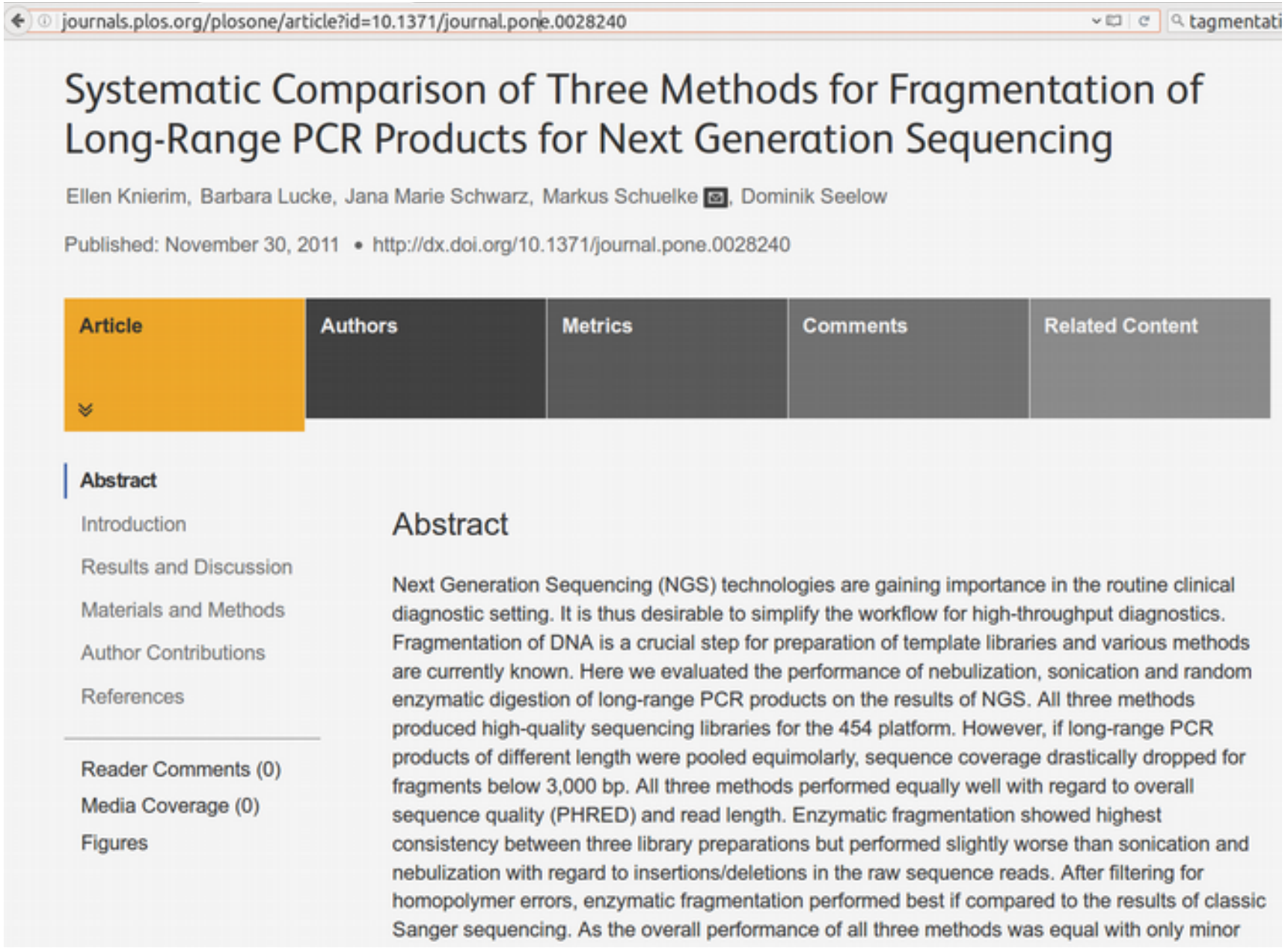
DNA fragmentation: Chemical Fragmentation

Heat and divalent metal cation

Chemical shear is typically reserved for the breakup of long RNA fragments. This is typically performed through the heat digestion of RNA with a divalent metal cation (magnesium or zinc). The length of your RNA (115 bp – 350 nt) can be adjusted by increasing or decreasing the time of incubation.

-

Reference


journals.plos.org/plosone/article?id=10.1371/journal.pone.0028240
tagmentati
Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing
Ellen Knierim, Barbara Lucke, Jana Marie Schwarz, Markus Schuelke, Dominik Seelow
Published: November 30, 2011 • http://dx.doi.org/10.1371/journal.pone.0028240
Article
Authors
Metrics
Comments
Related Content
Abstract
Introduction
Results and Discussion
Materials and Methods
Author Contributions
References
Reader Comments (0)
Media Coverage (0)
Figures
Abstract
Next Generation Sequencing (NGS) technologies are gaining importance in the routine clinical diagnostic setting. It is thus desirable to simplify the workflow for high-throughput diagnostics. Fragmentation of DNA is a crucial step for preparation of template libraries and various methods are currently known. Here we evaluated the performance of nebulization, sonication and random enzymatic digestion of long-range PCR products on the results of NGS. All three methods produced high-quality sequencing libraries for the 454 platform. However, if long-range PCR products of different length were pooled equimolarly, sequence coverage drastically dropped for fragments below 3,000 bp. All three methods performed equally well with regard to overall sequence quality (PHRED) and read length. Enzymatic fragmentation showed highest consistency between three library preparations but performed slightly worse than sonication and nebulization with regard to insertions/deletions in the raw sequence reads. After filtering for homopolymer errors, enzymatic fragmentation performed best if compared to the results of classic Sanger sequencing. As the overall performance of all three methods was equal with only minor

Library size

- Desired library size is determined by the desired insert size (referring to the library portion between the adapter sequences), because the length of the adaptor sequences is a constant.
- In turn, optimal insert size is determined by the limitations of the NGS instrumentation and by the specific sequencing application.

- For example, when using Illumina technology, optimal insert size is impacted by the process of cluster generation in which libraries are denatured, diluted and distributed on the two-dimensional surface of the flow-cell and then amplified.
- While shorter products amplify more efficiently than longer products, longer library inserts generate larger, more diffuse clusters than short inserts.

- Optimal library size is also dictated by the sequencing application.

- For exome sequencing, more than 80% of human exomes are under 200 bases in length

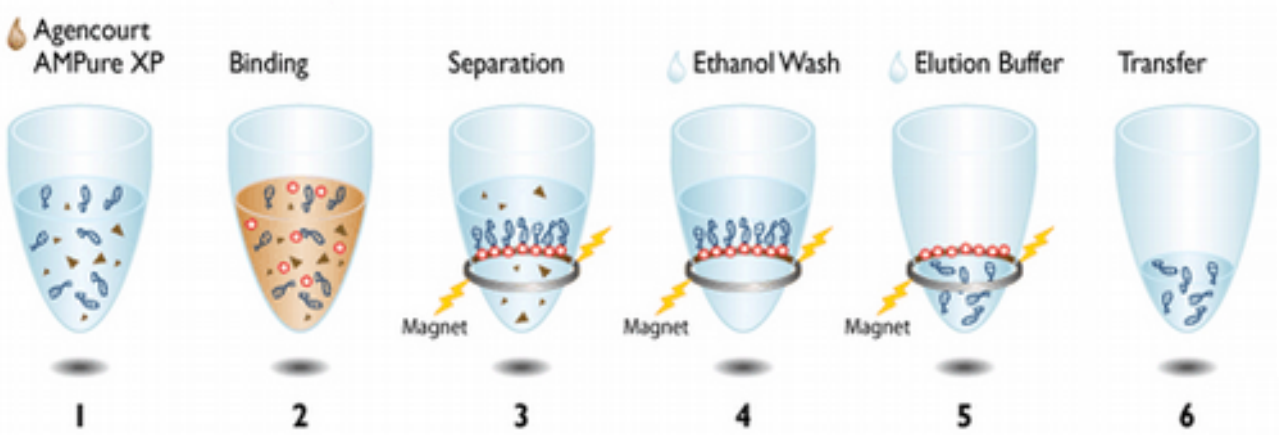
  Sakharkar, M.K., V.T. Chow, and P. Kangueane. 2004. Distributions of exons and introns in the human genome. In Silico Biol. 4:387-393.

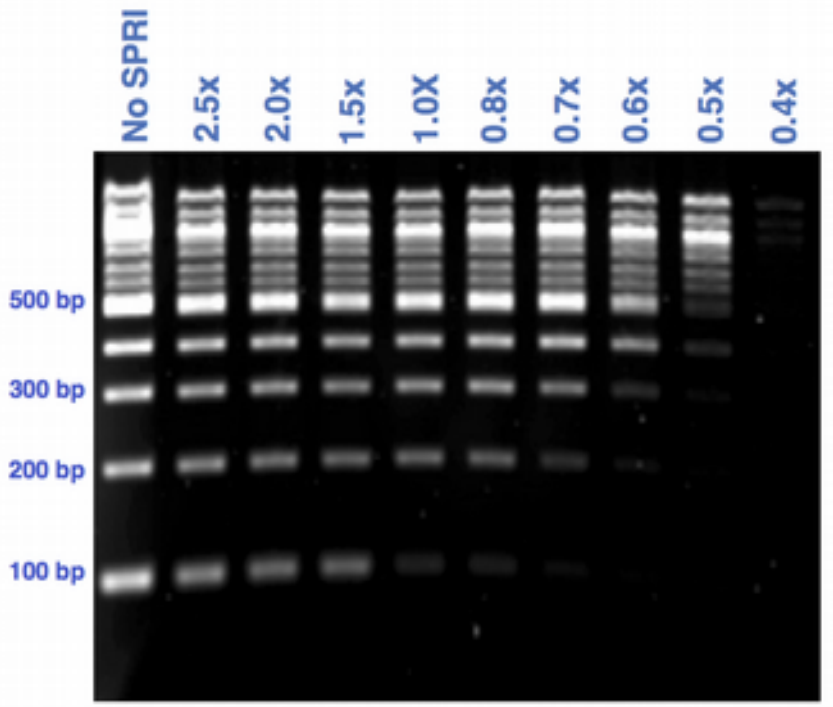
- The size of an RNA-Seq library is also determined by the applications.

- basic gene expression analysis use single-end 100 base reads.
- However, for analysis of alternative splicing or determination of transcription start and stop sites, 2 × 100 base paired-end reads are needed.

- A second post-library construction sizing step is commonly used to refine library size and remove adaptor dimers or other library preparation artifacts.
- Adapter dimers are the result of self-ligation of the adapters without a library insert sequence. These dimers form clusters very efficiently and consume valuable space on the flow cell without generating any useful data.

- magnetic bead-based clean up

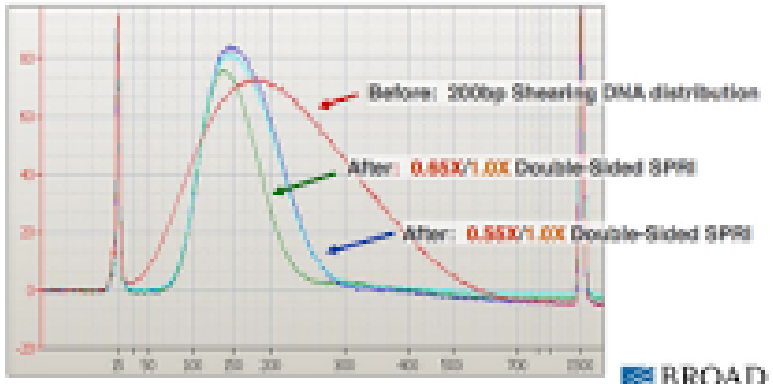




Option 2: Double-Sided SPRI

- By implementing a combination of good shearing with SPRI and "reverse" SPRI, one can select a fairly tight size range *with no gel*:

Results:



BROAD INSTITUTE illumina

- purify the products on agarose gels