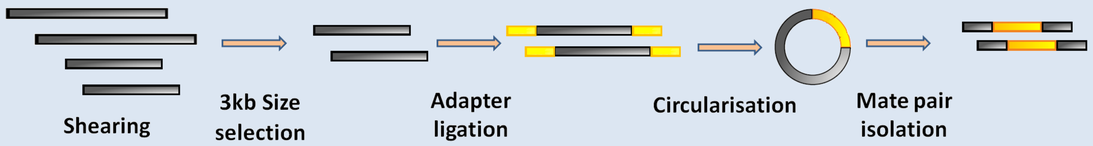
Improvements to 3kb Long Insert Size Paired-End Library Preparation

Naomi Park, Lesley Shirley, Michael Quail, Harold Swerdlow
Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

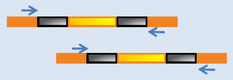
Introduction

Long insert size paired-end libraries, also known as mate-pair libraries, are highly valuable for determining the orientation and relative positions of contigs produced by *de novo* sequencing and assembly. Construction of these libraries can be problematic in the case of low sample availability and/or quality. Additionally, the number of libraries generated in parallel is limited by the complex nature of the library preparation steps.

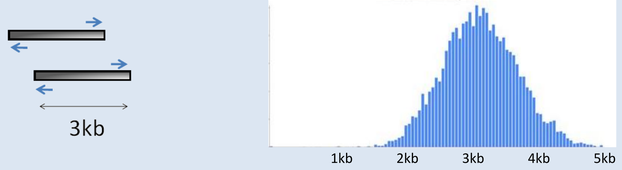
Preparation of a 3kb long insert size paired-end library can broadly be described in the following steps:



At Sanger the Isolated mate-pair undergoes Illumina library preparation (end repair, A-tail, Illumina adapter ligation and amplification), before sequencing on the Miseq or Hiseq.



The forward and reverse reads map to the reference genome, ~3kb apart, in an outward configuration:



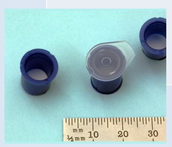
This poster summarises how adjustments made to steps within the library construction can improve the quality of the final mate-pair library, and/or enable library construction from as little as 800ng of gDNA of varying integrity.

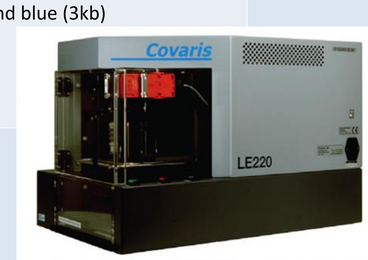
Shearing

We have used the Hydroshear and Covaris to shear gDNA to 3kb with comparable size distributions. The same gDNA sample, sheared using Covaris blue miniTUBEs (manufacturer recommended settings) and the Hydroshear (std assembly, speed code 13, 20 cycles):

The Hydroshear and Covaris both have unique advantages and disadvantages:

	Advantage	Disadvantage
Hydroshear	• Flexibility - Parameters can be altered to generate desired fragment size	• Low throughput - frequent user intervention required • Assemblies may become blocked by particulates
Covaris miniTUBE	• Higher throughput – upto 24 samples can be sheared in 1 hour using LE220 without user intervention • High recovery – samples are contained within the miniTUBE, minimising sample loss	• The ability to adjust parameters to generate desired fragment size is limited by miniTUBE type; red (5kb) and blue (3kb)



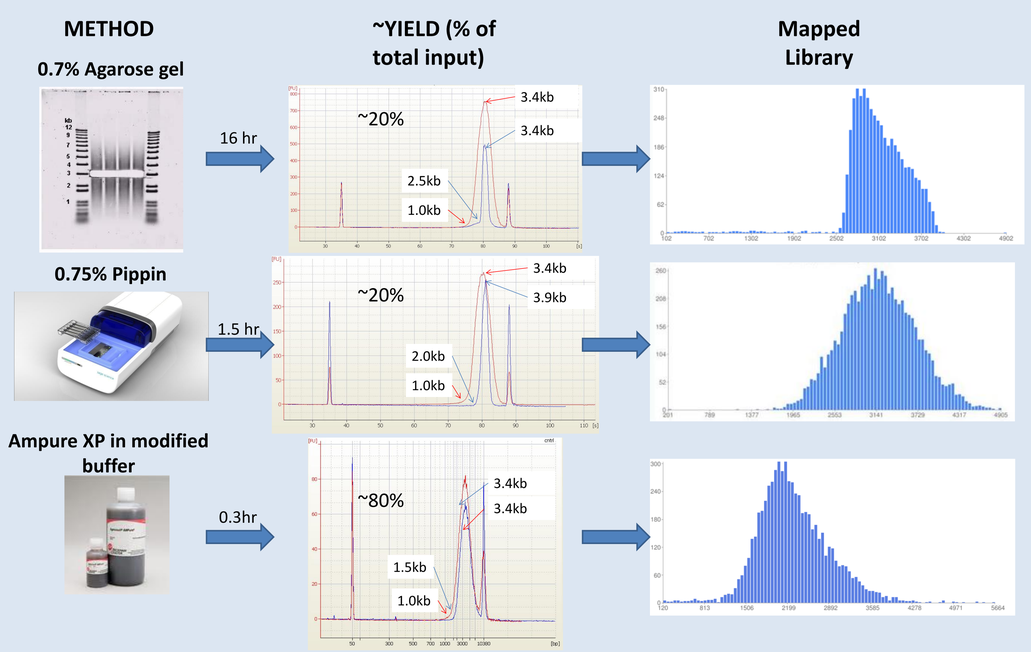


Due to the advantages of the Covaris miniTUBE, we now use this method for 3kb mate-pair projects with limited sample availability and/or high sample number. For mate-pair projects in which the desired insert size cannot be generated using Covaris miniTUBEs, the Hydroshear is used.

Size-Selection

The length and range of the material which undergoes circularisation during mate-pair construction determines the gap size of the paired reads of the final library. The method of size-selection used at Sanger is determined by the final library gap size required and the availability of input material.

We have compared three different methods of 3kb mate-pair library size-selection, manual 0.7% agarose gel, 0.75% Pippin (Sage Science), and Ampure XP bead (resuspended in an in-house modified buffer). Pre size-selected, 3kb sheared fragments (red) are overlaid with post size selected fragments (blue) in the bioanalyser traces below:

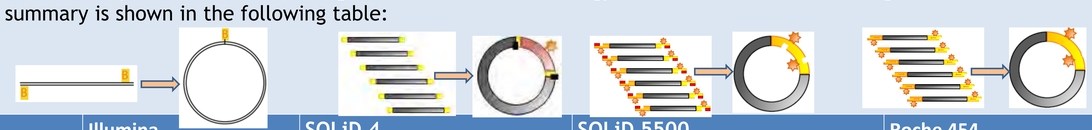


In the case of ≥5 µg starting material, the Pippin 0.75% cassette is our method of choice. In contrast to a manual gel, it is complete in less than 2 hours and successfully removes smaller fragments. However, as the 0.75% Pippin cassette settings are dependant upon input quantity and size, these must be established first.

In the case of 0.8-5 µg starting material, inexpensive size-selection with Ampure XP in modified buffer removes the majority of fragments <1.5 kb. This enables a degree of size-selection when samples are limited, in addition to enabling an automated, high throughput size-selection method.

Circularisation

Illumina's mate-pair sequencing protocol enables the production of 2-5kb libraries from a single kit, which are ready for sequencing on the Illumina platform. Other sequencing platforms employ alternative methods of circularisation within their mate-pair protocols, which can be adapted to generate libraries for Illumina sequencing. We have sequenced Illumina adapted 3kb libraries, using circularisation methodology from Illumina, Life Technologies and Roche. A summary is shown in the following table:



	Illumina	SOLiD 4	SOLiD 5500	Roche 454
Circularisation Method	• 3' Biotinylation • Blunt Ended Ligation • Random Shearing	• Adapter biotinylated • 2bp hybridisation and ligation • Nick translation and digestion	• Adapter biotinylated • Longer bp hybridisation • Nick translation and digestion	• Adapter biotinylated • Cre/lox recombination • Random shearing
Positives	• Single kit	• High number of outward facing reads • Sequence even either side of adapter	• High number of outward facing reads • Greater efficiency of circularisation than SOLiD 4 • Sequence even either side of adapter	• Cre/lox recombination is highly efficient for circularisation
Negatives	• Biotin can insert into nicked DNA, causing inward facing reads • Circularisation inefficient • Sequence uneven surrounding junction	• Low yield from circularisation	• Requires both left and right adapter ligation, 50% of adapter ligated fragments will be unable to circularise • Circularisation reaction involves heating to 70°C	• Requires both left and right adapter ligation, 50% of ligations will be unable to circularise • Sequence uneven surrounding junction

Due to the positives and negatives of each commercial circularisation method, we have designed our own method to address these.

Future Work

There is demand for mate-pair libraries on a larger scale. To address this we are looking to optimise a mate-pair workflow suitable for automation. Additionally, we are looking to optimise a method to improve the generation of larger mate-pair libraries.