A method to simultaneously construct up to 12 different sized Illumina Nextera long mate pair libraries with reduced DNA, time and costs.

Darren Heavens, Gonzalo Garcia Accinelli, Bernardo Clavijo and Matthew Clark

The Genome Analysis Centre (TGAC), Norwich Research Park, Norwich, UK, NR4 7UH

Introduction

Determining insert size for Long Mate Pair libraries and confirming they span known repeat sizes in genomes of interest is important to ensure that the best possible genome assemblies are achieved. We present a method to simultaneously size select and construct up to 12 libraries at a time and then map the generated reads back to the available assembled sequences to accurately calculate insert sizes. These can then be used to determine which libraries to sequence to greater depth and to use the accurate insert size information in *de novo* genome assemblies to improve outputs.

TGAC LMP Library Workflow using SageELF

Two Nextera LMP reactions (inputs 3ug and 6ug) are pooled to produce a single library with a broad range of insert sizes (1.5kb to >17kb, see Figure 1).

Size-select 5ug of pool on **SageELF** (3.5 hour separation time). QC on Bioanalyzer 12000 chip (see Figure 2).

Circularize o/n, exonuclease, shear to 450bp (Covaris).

Capture fragments containing biotinylated junction adapter on mag beads.

End-repair, and ligate adapters (each size fraction receives a unique barcode).

Enrichment amplification (Kapa HiFi polymerase).

CleanPCR bead clean-up. Quantify all size fractions (Bioanalyzer HS, Qubit HS).

Library Validation

Pool all size fractions at equimolar amounts.

BluePippin size selection (370-470bp). Quantify using Kapa qPCR Quant kit.

MiSeq 2X300bp reads.

Process data with NextClip and map to reference genome with BWA-mem to measure insert size for each fraction.

Data

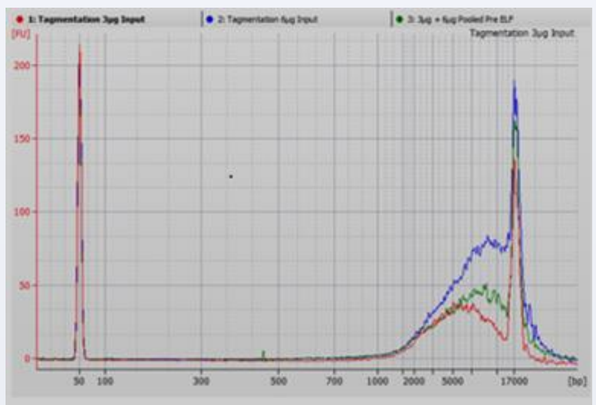

Figure 1. BioAnalyser Images of Tagmented DNA. 3µg input (red) and 6µg input (blue) tagmented DNA were pooled post strand displacement (green).

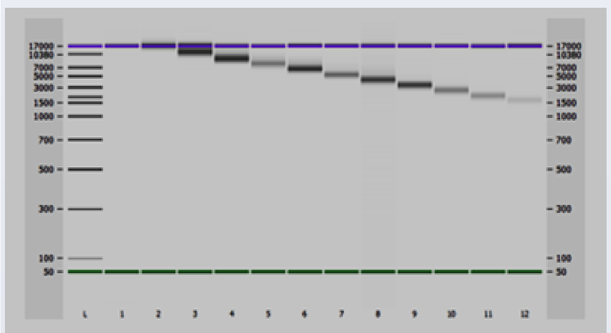

Figure 2. BioAnalyser Gel Images of DNA post size selection. ELF size selected fractions were analysed to estimate fragment length prior to circularisation.

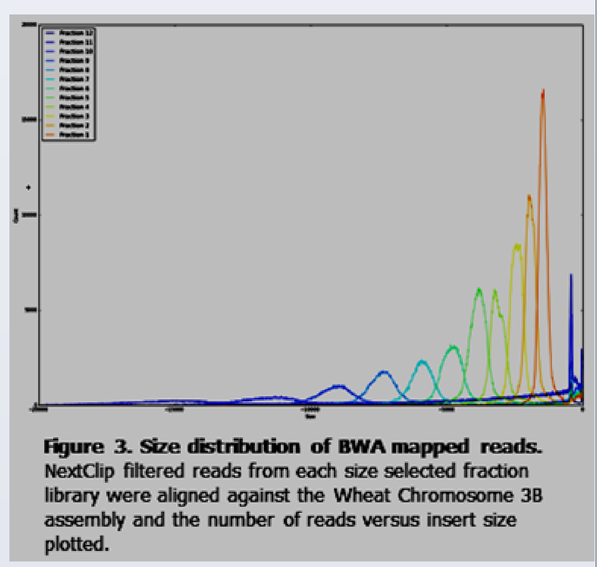

Figure 3. Size distribution of BWA mapped reads. NextClip filtered reads from each size selected fraction library were aligned against the Wheat Chromosome 3B assembly and the number of reads versus insert size plotted.

Key Benefits of TGAC ELF workflow:

Broad size range of Tagmentation reaction becomes an advantage (one reaction gives many insert size fractions).

Much improved DNA recovery in ELF system (>40%) compared with BluePippin (25%) allows up to 12 LMP libraries to be generated from a single 5 ug sample.

ELF workflow + MiSeq validation run gives accurate measurement of insert size.

ELF workflow avoids the problem of inaccurate targeting of a single specific size range sometimes encountered using BluePippin or manual gels.

Possible to produce 60 LMP libraries from 5 samples using a single 10-rxn Nextera kit.

Workflow is being used successfully at TGAC for WGS of wheat.

Contact

Darren Heavens: darren.heavens@tgac.ac.uk