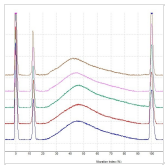

Cost-conscious generation of multiplexed short-read DNA libraries for whole genome sequencing

Ashley Jones[1], Justin Borevitz[1], Norman Warthmann[2]

[1]Research School of Biology, Australian National University. Canberra, ACT Australia, [2]Plant Breeding and Genetics Laboratory of the FAO/IAEA joint division in Seibersdorf, Austria

Oct 19, 2018

Working

dx.doi.org/10.17504/protocols.io.unbevan

EBL_ANU

Ashley Jones
Australian National University

ABSTRACT

Illumina® short-read DNA sequencing has become an integral tool in biology for genome-wide studies. Offering accurate base-pair resolution at the most competitive price, the technology has become widespread. However, the generation of multiplexed DNA libraries remains costly and cumbersome. Here, we present a streamlined cost-conscious protocol for generating multiplexed short read DNA libraries using a transposase from Illumina®. By implementing small volumes that use 1/25th the amount of transposase compared to Illumina® Nextera™ protocols, the cost of library preparation can be significantly reduced, by 1/10th or more. Furthermore, we optimised the protocol to minimise carboxylate bead-based cleanups between steps, further reducing cost, time and DNA input. By developing our own indicies to multiplex nine 96-well plates, up to 864 samples can be placed on a single flow cell. This enables efficient usage of monolithic sequencing platforms that can offer over three terabases of sequencing per flow cell.

PROTOCOL STATUS

Working

Verified across multiple genera.

GUIDELINES

This protocol is designed for Illumina® short-read sequencing and requires previously purified DNA. The DNA will be diluted and later amplified, thus only a small input is needed, as little as 2.24 ng. To aid conducting this protocol, program files for JANUS automated workstations (PerkinElmer®) have been provided in the supplemental with descriptions and a list of compatible hardware. Helpful excel sheets to create standard curves for DNA quantification and dilution strategies are also provided. Lastly, custom indicies are provided to multiplex nine unique 96-well plates, enabling multiplexing up to 864 samples on a single sequencing run.

supplemental.zip

BEFORE STARTING

In this protocol, the transposase being used will be diluted 1/25, therefore up to 25 times more reaction buffer is necessary. Prepare the custom buffer 2x TMgCl_2: 20 mM Tris-HCl (pH 8.4-9), 10 mM $MgCl_2$. Recommended setup:

Component	Stock	Quantity (1 mL)
20 mM Tris-HCl (pH 8.4-9)	500 mM	40 µL
10 mM $MgCl_2$	250 mM	40 µL
Nuclease-free water	N/A	920 µL

Materials

- 96-well half skirt PCR microplate (Axygen Scientific® PCR-96M2-HS-C).
- 96-well PCR microplate white, Roche 480 light cycler (Axygen Scientific® PCR-96-LC480-W-NF) (or an alternative microplate

for fluorescent readings).
- 384-well plate (4titude FrameStar 4ti-0384/B).
- PCR-grade adhesive film for microplates (e.g. Eppendorf storage film 0030127870).
- 2 mL SC micro tube PCR-PT (Starstedt 72694.406).
- 1.5 mL Eppendorf tube.
- 10 or 15 mL Falcon tube.
- MinElute® PCR purification kit (Qiagen 28004).
- Agencourt AMPure XP (Beckman Coulter A63881).
- Quant-iT™ dsDNA assay kit, high sensitivity (Molecular Probes™ Q33120).
- Nextera™ tagment DNA enzyme 1 (TDE1) (Illumina® catalogue # 15027865, also available in Illumina® kits FC-121-1030* or FC-121-1031*).
- Q5® Hot Start High-Fidelity DNA Polymerase (New England BioLabs® M0493).
- 10 mM dNTPs (New England BioLabs® N0447).
- Custom barcode oligos (see supplemental), or Illumina® Nextera™ index kit.
- 70% ethanol (analytical grade or higher).
- Nuclease-free water.

*Illumina® has issued a product obsolescence notification for these kits. However, TDE1 can ordered as a standalone component. This protocol has not tested the bead-bound transposase in the new Illumina® Nextera™ DNA flex kits. However, it is likely minimal to no optimisation would be required. For instance, magnetizing an aliquot of the bead-bound transposase and resuspending in 2x TMgCl$_2$.

Equipment

- JANUS automated workstation, such as NGS express or G3 (PerkinElmer®) (recommended but not essential).
- JANUS automated workstation accessories including; reagent reservoirs, plate adaptor support tiles, 25 and 175 µL filtered tips (PerkinElmer®).
- Infinite® M1000 PRO microplate reader (Tecan).
- PCR machine for 96-well plate.
- PCR machine for 384-well plate or a microplate heat block.
- Magnet rack for PCR tubes or 1.5 mL Eppendorf tubes.
- Qubit Fluorometer (ThermoFisher Scientific).
- High sensitive electrophoresis system; e.g. GXII Labchip (Perkin Elmer®) or Bioanalyzer (Agilent).
- (Optional) Automated agarose gel electrophoresis system, e.g. Pippin Prep (Sage Science).

DNA quantification

1 Ensure each well in the DNA plate is approximately below 20 ng/µL. For instance, quantify a few samples on a NanoDrop (ThermoFisher Scientific) and dilute the plate with nuclease-free water.

2 Create a Quant-iT™ master-mix by mixing dsDNA HS dye 1:200 with dsDNA buffer. For a complete plate, create enough buffer for 96 samples, 32 standards and 12 samples dead volume. Note 97 µL of the buffer will be used per sample. For example:

Number of plates	Samples	Standards	Dead volume	Total samples	dsDNA buffer (µL)	dsDNA HS dye (µL)
1	96	32	12	140	13,512	68
2	192	64	12	268	25,866	130
3	288	96	12	396	38,220	192

NOTE

JANUS automated workstations (PerkinElmer®) generally require at least 1 mL of dead volume in large reservoirs.

NOTE

This reaction setup successfully utilises uses half reactions compared to the official Quant-iT™ protocol.

3 For each DNA plate, transfer 97 µL of working solution into two half-plates (96-well PCR microplate white, Roche 480 light cycler). For instance, filling columns 1-6 on plate 1, 7-12 on plate 2.

4 Transfer an additional 97 µL of working solution into two free columns in each plate (16 wells each plate). This will be used for 8 standards, done in duplicate on each plate.

NOTE

Having standards on each plate is necessary for accuracy when generating standard curves later.

5 Add 3 µL of each DNA sample to a separate well containing working solution, splitting equally across the two plates.

6 Add 3 µL of each standard to the columns dedicated to standards. Perform this in duplicate.

NOTE

Add standards in increasing concentration; 0, 0.5, 1, 2, 4, 6, 8, and 10 ng/µL. Then duplicate in the next column.

7 Seal the plate with adhesive film, vortex and briefly spin down. Incubate for 5 min.

8 Measure fluorescence using microplate reader using excitation/ emission maxima ~502/523 nm.

NOTE

For an Infinite® M1000 PRO (Tecan), use a bandwidth of 5 nm, top mode, flashes: 100 Hz x10, Z-position: 20,000 µm, gain can be optimal or set a manual gain at 96. Ensure the correct plate is entered, in this case Roche Lightcyclyer 480 96 well.

9 Utilising the known concentration of standards included on the plate, create a standard curve to determine DNA concentration of the other samples.

NOTE

For a Infinite® M1000 PRO (Tecan), use files provided in supplemental; helpful excel sheets.

DNA dilution

10 Dilute all DNA samples to 2 ng/µL in a volume of approximately 25 µL. First transfer nuclease-free water to a new 96-well PCR

microplate and then add the appropriate amount of DNA.

NOTE

For JANUS automated workstations (PerkinElmer®), to aid calculating a diluting strategy, files provided in supplemental; helpful excel sheets.

NOTE

When using a JANUS automated workstation, note that the most accurate tip size is 25 µL, however highly concentrated samples may need a larger volume of water to be transferred than 25 µL.

11 Ensure the diluted samples are mixed and briefly spin down if necessary. Can be sealed with adhesive film and stored at -20°C.

12 Quantify the DNA using Quant-iT™ as previously described.

13 From the 2 ng/µL plate, repeat the dilution process, diluting to 0.8 ng/µL in a new 96-well PCR microplate. Store at -20°C until ready for tagmentation.

NOTE

No need to quantify after this second dilution, samples will never all be exactly 0.8 ng/µL, the quantification accuracy has limitations.

Tagmentation

14 Prepare a tagmentation master-mix in a 1.5 mL Eppendorf. For an entire 96-well plate of samples, cater for 102 samples (i.e. a dead volume of 6 samples).

Reagent	µL for 1	µL for 102
2x $TMgCl_2$ (custom buffer)	3	306
Transposase (TDE1)	0.2	20.5
DNA at 0.8 ng/µL (uses 2.24 ng)	2.8	-
TOTAL	**6 µL**	**327 µL (aliquot 3.2 µL)**

NOTE

The transposase, Tagment DNA Enzyme 1 (TDE1), is from Illumina® Nextera™ kit. Older kits call the transposase Amplicon Tagment Mix (ATM) or something similar. Will be a viscous solution that does not freeze.

NOTE

Use ~0.1 µL - 0.125 µL transposase per 1ng DNA input.

15 Pre-heat a PCR machine or heat block at 53°C suitable for 384-well microplates.

NOTE

Ideally, the temperature would be 55°C (for the transposition reaction), however, if there is a chance of overheating, the transposase can denature.

16 Transfer 3.2 µL of the tagmentation master-mix into 96-wells of a 384-well microplate (for example quadrant 1 in each set of four wells).

NOTE

Perform tagmentation in 384 well plate due to the low reaction volume and to enable aspiration with automated pipetting equipment.

NOTE

Left-over tagmentation master-mix can be stored at -20°C and used in 6 months or more. Avoid cycles of freeze-thawing; solution no longer viscous and will freeze solid.

17 Transfer 2.8 µL of DNA (at 0.8 ng/µL) to each of the 96-wells used within the 384-well plate. Gently pipette mix during transfer.

18 Seal plate with adhesive film (PCR-grade), no need to vortex. Centrifuge 800 rcf, 1 min.

19 Incubate plate at 53°C for 30 min. If using a mixing heat block, shake at 300 rpm.

NOTE

Start PCR setup during this time.

20 Briefly cool to room temperature (22-24°C) and proceed without delay. Longer transposition times create shorter fragments.

NOTE

No clean-up of this reaction is necessary.

PCR for enrichment and barcoding

21 Prepare a PCR master-mix without primers and the transposase reaction. For an entire 96-well plate of samples, cater for 102 samples (i.e. a dead volume of 6 samples).

Reagent	µL for 1	µL for 102
Nuclease-free water	5.27	537.5
5x Q5 reaction buffer	2.5	255
5x Q5 high GC enhancer (optional)	3.5	357
10 mM dNTPs	0.5	51
Q5 hot start high-fidelity DNA polymerase	0.23	23.5
2.5 µM forward primer	4	-
2.5 µM reverse primer	4	-

Transposase reaction (template)	5	-
TOTAL	**25 µL**	**1,224 µL (aliquot 11 µL)**

NOTE
PCR master-mix appears to deactivate the transposase.

NOTE
Less than 1x of the Q5 reaction buffer and high GC enhancer will present in each reaction. This has been calculated based on current salt concentrations in the non-cleaned tagmentation reaction.

NOTE
PCR will be performed in a 96-well plate; transposase reaction will be transferred from a 384-well to 96-well microplate.

22 Transfer 11 µL of PCR master-mix to each well of a 96-well PCR microplate.

23 Add 4 µL of forward primer and 4 µL of reverse primer to each well (i5 and i7 oligos). Ensure each well has a unique combination of barcodes.

NOTE
Oligo sequences and recommended combinations are provided in the supplemental.

NOTE
This step can be prone to human error. Ensure tubes are labelled and plate orientations are accounted for (e.g. include blanks in decipherable locations per plate). Using an automated pipetting robot such as JANUS automated workstations (PerkinElmer®) is highly recommended.

24 Transfer 5 µL of transposase reaction (template) from the 384-well microplate to the 96-well microplate containing the PCR master-mix aliquots.

25 Seal plate with adhesive film (PCR-grade), vortex and centrifuge 800 rcf for 1 min.

26 Perform PCR as follows:

72°C for 3 min
98°C for 30 s
10-14 cycles of:
- 98°C for 10 s
- 63°C for 30 s
- 72°C for 60 s

Hold at ≤ 10°C

NOTE
Avoid high number of PCR cycles to limit duplicate fragments and formation of hetero-duplexes. Currently we have been using 13 cycles.

27 Proceed to pooling or store at -20°C.

Pool libraries

28 Add 25 µL of nuclease-free water to all wells (1:1 dilution).

NOTE
Optional, but this helps prevent evaporation and enables more efficient pooling.

29 Add 40 µL of water into 1.5 mL Eppendorf tube.

30 Transfer 10 µL from each well of the PCR microplate (each library) into the same 1.5 mL Eppendorf tube (use robot).

NOTE
10 µL x 96 samples + initial 40 µL = 1,000 µL.

31 Save an aliquot (e.g. 10 µL) of the pool for QC analysis on a high sensitive automated electrophoresis system (e.g. PerkinElmer® GXII Labchip or Agilent Bioanalyzer).

Initial clean: MinElute® columns

32 Clean-up the pooled library with a MinElute® PCR purification kit (Qiagen). First, transfer the pool to a 10 or 15 mL Falcon tube.

NOTE
MinElute® columns have a binding capacity of up to 5 µg and recovers fragments 70 bp – 4 kb.

33 To the pool (1,000 µL), add 5x volume of PB buffer (5,000 µL).

34 Prepare 2 MinElute® columns for the pool (6,000 µL).

Pipette 750 µL into each column, centrifuging for 1 min at 17,900 rcf. Discard the flow-through.

35

36 Repeat the previous step three more times, processing the entire sample.

37 Wash by adding 750 µL of PE, centrifuge for 1 min at 17,900 rcf.

NOTE
Important: without this wash, gel electrophoresis will not work properly (creates a band shift- inaccurate size estimation and elutes too early on automated systems).

38 Discard flow-through and centrifuge while empty for 1 min at 17,900 rcf.

39 Place column in clean 1.5 mL Eppendorf tube.

40 Add 30 µL of nuclease free water to each column, centrifuge 1 min at 17,900 rcf.

41 Aspirate the 30 µL of nuclease-free water flow-through, add it to the column again and centrifuge 1 min at 17,900 rcf.

42 Transfer both elutions to a new Eppendorf tube.

43 Save 2 µL for QC, adding 8 µL of water to create a 1/5 dilution.

44 Measure remaining volume. Volume should be 30 + 30 – 2 = 58 µL, but is generally less.

Optional size selection: Pippin Prep

45 To enrich for longer fragments, agarose gel size selection can be performed. This will remove shorter fragments that will have over-lapping sequencing on paired-end platforms. Using a PippinHT (Sage Science) or similar automated electrophoresis product, gel purify the library, following the manufacturer's instructions. A 1.5% or 2% gel to select 430 -1,000 bp is recommended (adapters combined are ~123 bp and generally 150 bp paired-end sequencing is performed).

NOTE
The PippinHT has 12 lanes, markers can be internal (recommended) or external. 20 µL of DNA goes into each lane and

the manufacturer recommends a maximum 1.5 µg per lane. This can overloaded to 3 µg per lane but the size selection will not be as precise. It is recommended to use 2-3 lanes on the PippinHT depending on DNA quantity.

46 Collect the contents of all elution wells into a 1.5 mL Eppendorf.

47 The library collected can be sent for sequencing in the current buffer. However, it will likely need concentrating; proceed to bead clean-up.

Final clean: AMPure XP beads

48 If the optional size selection was not performed, add 0.6x volume of AMPure XP beads (e.g. 50 µL of MinElute® + 30 µL beads). If size selection was performed, use 1.2x volume of AMPure XP beads.

49 Pipette mix thoroughly but gently.

50 Incubate for 5 min at room temperature.

51 Place on magnet rack for 5 min, or until clear (beads pellet).

52 Keep on the magnet, remove supernatant without disturbing the beads and discard.

53 Add 100 µL of 70% ethanol (or enough to cover the beads), for 30 s.

54 Take off magnet, gently pipette mix and place back on magnet until clear.

55 Remove supernatant and repeat the ethanol wash.

56 Air dry pellet, avoid cracking or over-drying (can remove from magnet).

57 To elute, resuspend the beads in 60 µL water (off magnet), incubate for 5 min.

58 Place on magnet rack for ~5 min, or until clear.

59 Pipette 55 µL of solution into a new PCR tube (ensure no bead carry-over, if necessary place new tube on magnet and repeat the pelleting and transfer).

60 If optional size selection was not performed, repeat the bead clean-up for a second 0.6x AMPure XP clean.

61 Save 2 µL for QC, adding 8 µL of water to create a 1/5 dilution.

Preparation for sequencing

62 Perform a QC analysis on a high sensitive automated electrophoresis system (e.g. PerkinElmer® GXII Labchip or Agilent Bioanalyzer).

> NOTE
> *Include aliquots from the raw (unclean) pool, the MinElute® clean and the final Pippin Prep /AMPure XP clean.*

63 Ensure the library looks as expected, a single broad peak between 250 and 1,550 bp. Estimate the average fragment length.

> NOTE
> *A double peak is indicative of over-amplification; primers have become limited, PCR products have annealed, forming single stranded fragments that are twice as large ('PCR bubble'). As these are annealing products, sequencing can still be performed, but quantification based on dsDNA is not precise.*

> NOTE
> *A stretched peak, similar to a double peak, can form if the electrophoresis system is overloaded. This is referred to as 'bird nesting'; the DNA fragments get intertwined and appear longer. This is not an issue; denaturation during sequencing bridge PCR will eliminate this tangle.*

> NOTE
> *If the library size is very large (e.g. greater than 1 kb), there is not enough transposase fragmentation. Something may be inhibiting the tagmentation if the DNA is highly contaminated. Alternatively, DNA input may have been too high; check the concentration and reduce the DNA input if necessary.*

64 Quantify the library at least twice with a Qubit Floroumeter (e.g. use Quant-iT™ dsDNA assay kit). Determine the average of the readings.

65 Calculate molarity; nM = (ng/µL x 10^6) / (660 x average fragment length).

> **NOTE**
> *Although high sensitive automated electrophoresis systems will provide a molarity calculation, we have found molarity based on Qubit values and an average fragment length the most accurate.*

66 Dilute the library to a suitable molarity for sequencing. For instance, most sequencing facilities require 60 µL at 10 nM for Illumina® platforms. Store at -20°C until ready for sequencing.

67 Send for sequencing (on ice or cold packs). 150 cycle paired end sequencing the Illumina® NovaSeq 6000 is recommended. An S2 flow cell yields up to 1 terabases, an S4 flow cell yields 3 terabases.

> **EXPECTED RESULT**
> Using the protocol described, we have been successfully generating multiplexed short read libraries for multiple plant species (Figure 1). To quantify libraries, Qubit measurements combined with an estimated fragment length to calculate molarity was found to be the best predictor of sequencing yield (Figure 2). As species with smaller genomes give smaller insert sizes and *vice versa*, the ratio of DNA to transposase could be further optimised for each genome for the best results. However, Pippin Prep size selection provides a quick and accurate method to remove shorter fragments that would produce redundant overlapping sequencing on paired-end platforms (Figure 3). Increased insert sizes from Pippin Prep libraries have been verified post-sequencing, however not all short fragments were eliminated and have high affinity for sequencing on Illumina® platforms (Figure 4).

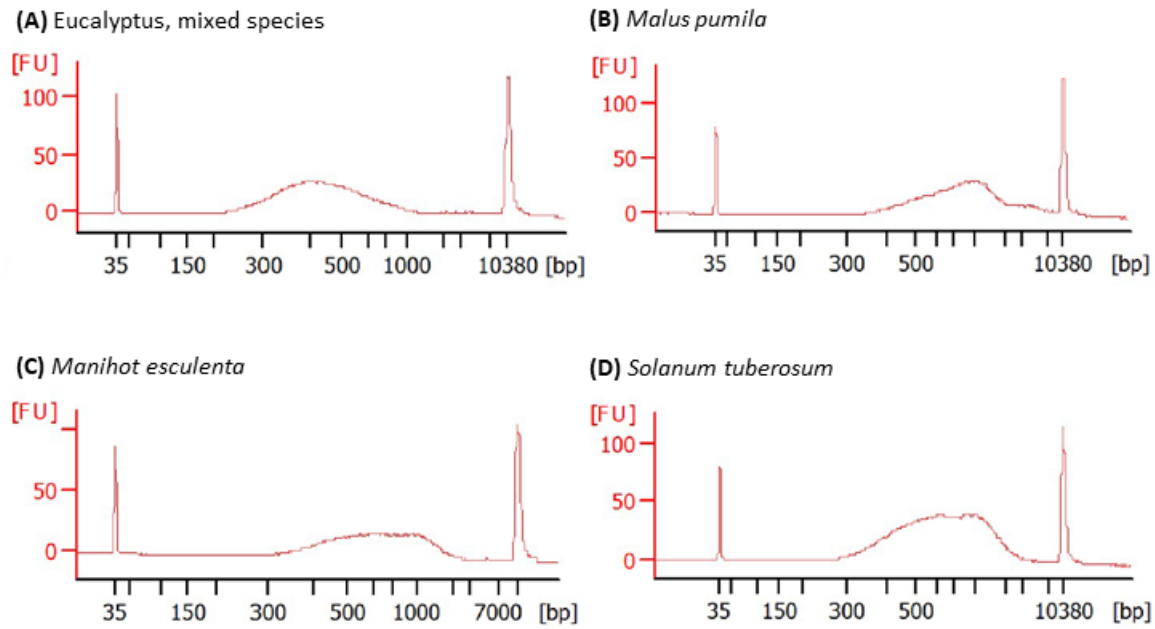

Figure 1: Example DNA libraries verified on a 2100 Bioanalyzer using a high sensitive DNA chip (Agilent Technologies). Electropherograms plot fragment size (base pairs, bp) against fluorescence units (FU). First peak (35 bp) and last peak (10,380 bp) are upper and lower markers respectively. (A) Gum trees, Eucalyptus, mixed species. (B) Apple, *Malus pumila*. (C) Cassava, *Manihot esculenta*. (D) Potato, *Solanum tuberosum*.

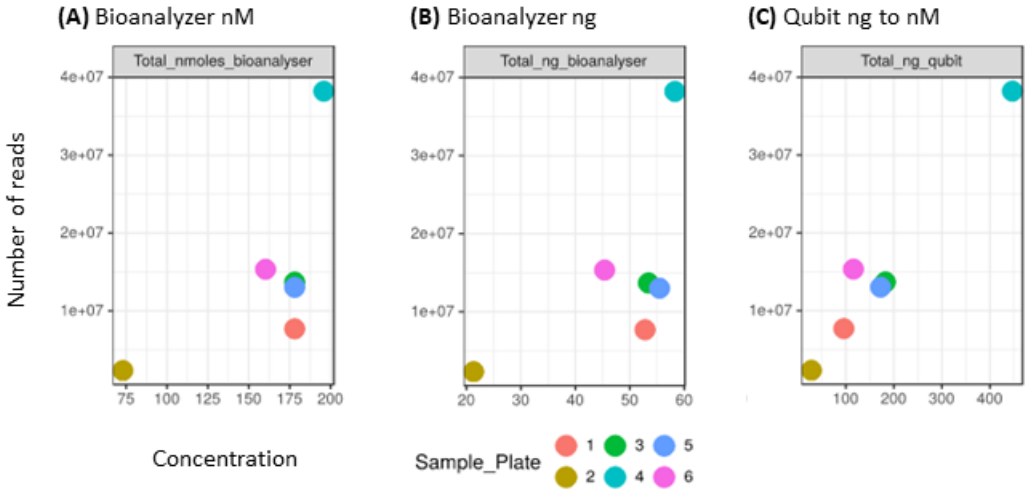

Figure 2: DNA library quantification methods compared to sequencing yield. (A and B) Agilent Bioanalyzer quantifications, calculations based on total nanomoles (nM) and total nanograms (ng) respectively. Qubit fluorometer quantification, based on total ng, which can be converted to nM using an average insert size. Qubit measurements were found to be a more accurate predictor of sequencing yield. This is unexpected, as number of reads is proportional to the number of fragments (molarity) rather than the mass of DNA.

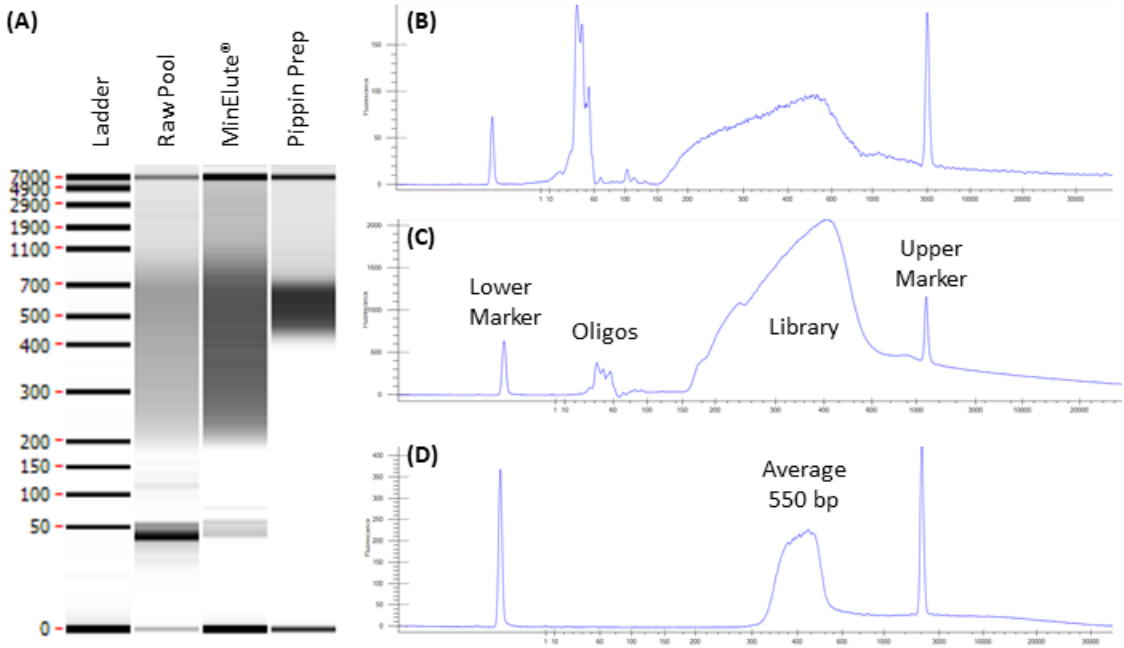

Figure 3: Fragment size selection on a pool of DNA libraries. A pool of 192 Eucalyptus DNA libraries was size selected for 430 to 1,000 bp using a Pippin Prep (Sage Science). This was verified on a LabChip GXII automated electrophoresis system using a high sensitive DNA chip (PerkinElmer). (A) Electrophoresis image. Left to right: ladder (PerkinElmer), raw pool of libraries post-PCR without any cleaning, cleaning with MinElute® columns (Qiagen) and final Pippin Prep. (B to D) Electropherograms, plotting fragment size against fluorescence. Raw pool, MinElute® and pippin prep respectively. First peak and last peak are upper and lower markers. As adapters and PCR oligos collectively add ~123 bp, fragments above 423 bp will have insert sizes of 300 bp, which is ideal for 150 bp paired end sequencing.

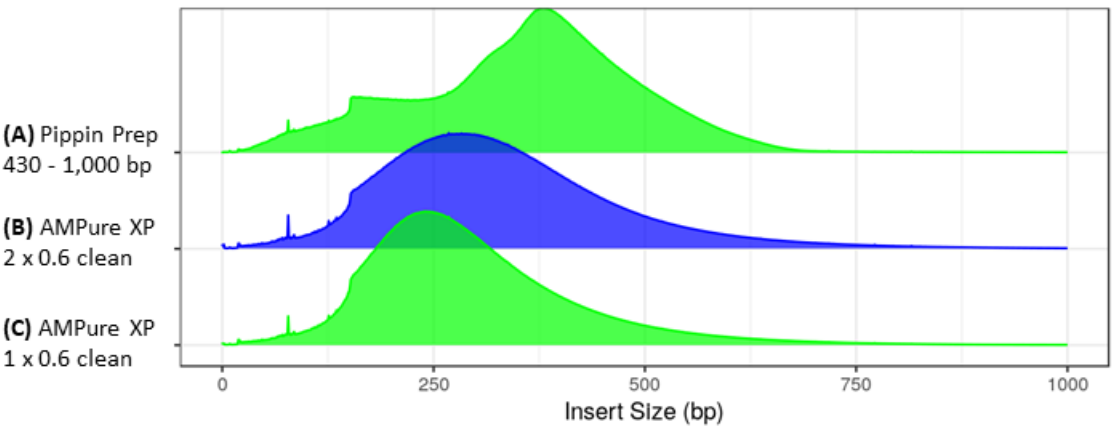

Figure 4: Insert size distribution of three multiplexed library pools after sequencing on the NovaSeq 6000 (Illumina®). Each library pool contains 96-192 Eucalyptus samples and each pool was purified using different strategies. (A) Pippin Prep (Sage Science), size selecting 430 to 1,000 bp. (B) Two AMPure XP bead cleans, both using 0.6x volume of beads compared to the pool volume. (C) A single AMPure XP bead clean using 0.6x volume of beads. Pippin Prep size selection increased the average insert size the most, having an average size of approximately 400 bp. This minimises redundant sequencing, as the 150 bp paired end reads will not overlap. However, small fragments are still present and will preferentially be sequenced.