Automating PacBio library construction: Scaling the Darwin Tree of Life Project at the Wellcome Sanger Institute

Ben Farr[1], Shelly-Ann Coutts[1], Lesley Shirley[1]*, James Watts[1], Craig Corton[1], James Uphill[1], Paul Lomax[2], Maryia Karpiyevich[2], Gregory Young[3], Camille Conner[3]

1. The Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom; 2. SPT Labtech, Melbourne Science Park, Melbourne, United Kingdom; 3. Pacific Biosciences, Menlo Park, California, USA

Background

One of The Wellcome Sanger Institutes' (WSI) flagship programmes is the Darwin Tree of Life (DToL), which aims to generate platinum standard reference genomes of the estimated 70,000 eukaryotic species in Britain and Ireland. This initiative aims to advance our understanding of life on Earth by unravelling the genetic blueprints of a wide range of organisms.

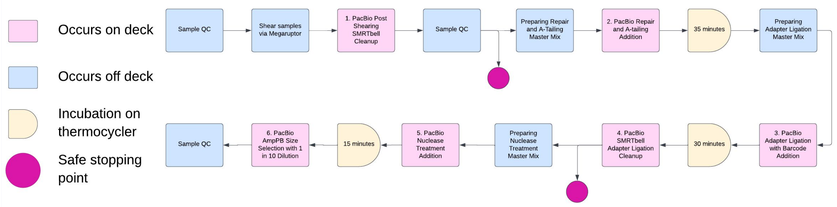
To address the scale of this challenge, the WSI automation team is evaluating robust automated solutions to enable the generation of long-read next-generation libraries. Here we describe the development and validation of a novel automated method on SPT Labtech's firefly® liquid handling platform to construct whole genome sequencing (WGS) libraries from genomic DNA using the PacBio SMRTbell prep kit 3.0 (SPK3).

PacBio SMRTbell prep kit 3.0 (SPK3) workflow overview

PacBio long read sequencing can produce high quality data (>Q30) with read lengths of 15-20kb. This enables the resolution of repetitive sequences, segmental duplications, and centromeres to produce more contiguous and phased assemblies.

The WSI's SPK3 pipeline has a capacity to process 250 samples a month. These samples are typically processed manually in batches of 24 and each batch takes ~3 days to process including sample QC. At the current rate WSI will not be able to deliver DToL's ambitions.

Diagram key



PacBio

firefly®

The SPT Labtech firefly® platform combines an air displacement multichannel pipetting head with 6 positive displacement heads.

Key features include:

- A 384-channel air displacement head which can load:
  - Arrays of 96 or 384 tips
  - Multiples of 8 or 16 channel pipette strips
- 6 positive displacement heads which can dispense 200 nL to 4 mL volumes
- 2 moving decks with a total of 16 positions
- Temperature controlled and shaker modules
- A compact footprint of 66 cm x 56 cm x 78 cm

These features and the unique combination of dispensing technologies make this platform a viable candidate for automating the SPK3 library preparation.

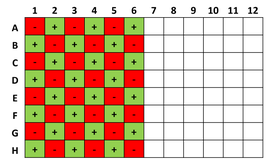
sptlabtech

Initial protocol validation

6 x automated methods were written on SPT Labtech's firefly®, these were split into:

- 3 x DNA purification assays
- 3 x reagent addition assays

The methods were validated using a homogenous stock of cricket DNA that was extracted using a KingFisher Apex and the Qiagen MagAttract HMW (High Molecular Weight) DNA extraction kit.



Upon successful completion of 8 test libraries, a subsequent chequerboard validation was performed. The chequerboard plate was comprised of the previously used stock of homogenous cricket material for the positive control samples and elution buffer for the negative control samples (image right). This validation is carried out to ensure that there is uniformity across the plate and to check for the presence of any potential intraplate contamination in the negative controls.

An equivolume pool of all 48 libraries on the Chequerboard plate was created and sequenced on PacBio's Sequel IIe platform. The box and whisker plot shows the positive cricket controls had an average of ~93,000 HiFi reads each with a CV of 13.6%.

Darwin Tree of Life test panel validation

Due to the wide variety of species processed at the WSI for the DToL programme, it was important to test how the automated library preparation would work on a range of phylogenetically diverse species. A test panel of 8 different species was selected:

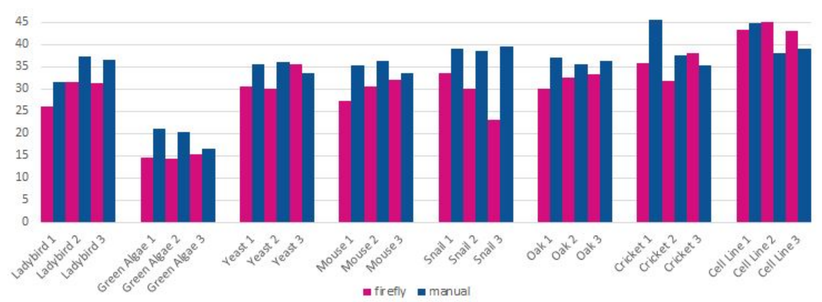
- *Adalia bipunctata* / Ladybird
- *Dunaliella primolecta* / Green Algae
- *Metschnikowia zobellii* / Marine Yeast
- *Mus musculus* / Mouse
- *Physella acuta* / Freshwater Snail
- *Quercus robur* / Oak
- *Teleogryllus oceanicus* / Cricket
- *Homo sapiens* / Human (Cell line)

2 sets of libraries were prepared from each species in triplicate:

- Set 1 prepared manually following PacBio's recommended procedure.
- Set 2 prepared following the novel automated method.

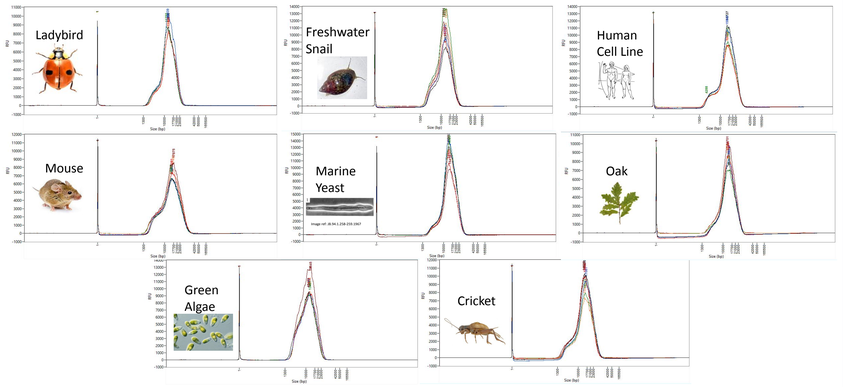
Where possible, variables were controlled e.g. the master mixes for the library preparation were split and shared between all samples.

As shown in the graph below the libraries prepared on the SPT Labtech firefly® have a comparable yield and CV% to the manually prepared libraries of the same species.

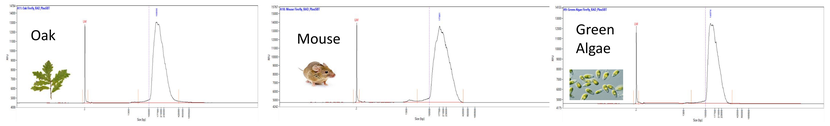


Darwin Tree of Life test panel validation continued

The Agilent Femto Pulse was used to validate the sizing of the libraries, as shown in the traces below. When traces of the same species are overlaid, the samples prepared using the firefly are analogous to those prepared with PacBio's manual procedure.



The Oak, Mouse and Green Algae pools were selected for sequencing. To optimise the sequencing results a size selection was performed using a sage science PippinHT instrument. Traces from the Agilent Femto Pulse showing the size distribution of the pools after size selection are shown below.



The 3 pools were sequenced on a PacBio Revio. Each pool contained 3 libraries prepared on the firefly and 3 libraries produced manually. The table below shows the key mean sequencing metrics for each set of triplicates. All 3 pools gave acceptable yields and produced high quality sequencing data. All libraries had a mean HiFi quality score of Q35 or higher.

Species	HiFi yield (Gb)	Replicates	HiFi yield (Gb)	Mean HiFi read length (bp)
Green Algae	77.3	firefly	12.13	14556
		manual	13.53	14534
Mouse	98.6	firefly	17.31	16562
		manual	15.47	16459
Oak	90.9	firefly	15.34	14889
		manual	14.87	14843

Summary

PacBio SMRTbell Preparation Kit 3 libraries were successfully prepared on the SPT Labtech firefly®. Key performance metrics were shown to be in keeping with the libraries prepared following the recommended manual procedure.

Testing of the SPT Labtech firefly® has demonstrated that it is a viable option for automating long read library preparation. Its use could support increased throughput, ensure consistent performance and reduce ergonomic risk.