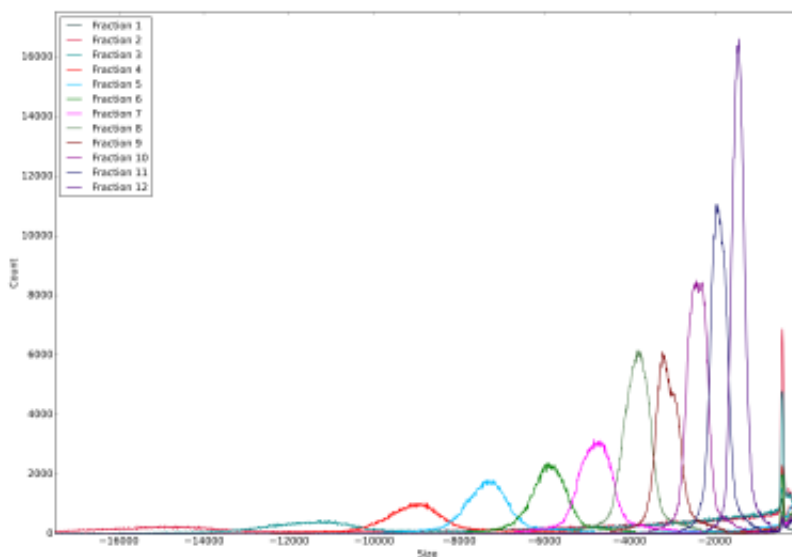


# Faster, better, cheaper: A new method to generate extended data for genome assemblies

July 23 2015

---



For improved accuracy, the wheat genome sequences were mapped back to an existing assembly to determine their actual size as appose to their predicted size. Credit: TGAC

Long range genetic data is an invaluable source for plant, crop and animal genetic research. Sequencing genomes requires breaking them into small manageable pieces and then working out how they go back together - similar to a million or billion piece jigsaw puzzle. To do this, a

combination of short range (a jigsaw piece) and long range (tells you about the nearby pieces) sequence data is needed.

While generating the short range data is relatively straightforward, the long range data is more problematic as quality and quantity of DNA are major factors influencing the outcome. Illumina's Nextera Mate Pair Sample Preparation Kit has helped improve the quality of long range (LMP) data, but they can still be challenging to generate. More complex genomes, typically benefit from accurately size selected LMP data to produce the highest quality genome assemblies.

Although producing a single high-quality size selected LMP library can be difficult, several LMP libraries are often used for larger genome sequencing projects. This new approach allows construction of 12 libraries for less than twice the cost of a single library and reduces the time by 3 to 2 days.

The TGAC team gained early access to a new piece of technology, the SageELF from Sage Scientific, with the aim to develop a more robust, global approach for accurately sized long-range [sequence data](#). The protocol would be more tolerant to DNA quality and quantity and could ensure that the best possible data was generated for any given sample.

"Improving LMP libraries has been a goal of TGAC's for some time, and we have previously published a software tool for processing this data to 'clean it up' before using it in genome assemblies," said Darren Heavens, Lead Author and Team Leader in the Platforms & Pipelines at TGAC.



European Ash Tree and wheat (example organisms this method has already been applied to produce high-quality genome assemblies). Credit: TGAC

"Previous approaches limited us to targeting a single size range at a time, whereas using this new protocol we can simultaneously target up to 12 different size fractions improving the likelihood of achieving the best long range data from a given DNA source, saving both time and money. While most projects wouldn't require sequencing all 12 fractions we can select the best fractions for sequencing".

The scientists hope that this new library construction approach will be widely adopted within the scientific community, having a positive effect in improving genome assemblies. Providing a better understanding of traits of economic interest in crops and animals which are seen as key requirements for breeders such as disease-resistance.

Matt Clark, last author, Plant and Microbial Genomics Group Leader at TGAC, added: "Generating high-quality genome assemblies is important as they form the base upon which we build an understanding of an organism's genetics. Since its development, this approach has already been successfully implemented in a number of our high profile sequencing projects such as bread wheat, durum wheat, *Aegilops sharonensis* and a European Ash Tree more resistant to Ash Dieback Disease, all with very impressive results.

"For all projects requiring long-range sequence information, we are now promoting this method as our Gold Standard approach. Identifying new technologies and their applications is a key aspect of the work we do and we will continue to monitor opportunities in this area."

Provided by The Genome Analysis Centre

Citation: Faster, better, cheaper: A new method to generate extended data for genome assemblies (2015, July 23) retrieved 27 April 2024 from <https://phys.org/news/2015-07-faster-cheaper-method-genome.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.