On-flow cell weighted low pass genome sequencing simplifies the workflow and accelerates breeding programs in agriculture

Junhua Zhao[1], Adeline Mah[1], Yirong Yang[2], Laure Edoli[1], Connor Thompson[1], Xin Wu[2], Jianxiang Peng[2], Bryan Lajoie[1], Shawn Levy[1], Benjamin Song[1], Bingbing Wang[2], Michael Previte[1]

[1]Element Biosciences, San Diego, CA, 92121

[2]Biobin Data Sciences, Changsha, Hunan, China

Element Biosciences

Introduction

Low pass genome sequencing along with genotype imputation has been shown as a cost-effective alternative to genotyping arrays for trait mapping in the agriculture field. While research has shown the feasibility of driving down the sequencing depth to 1x or 0.5x, coverage uniformity throughout the genome and low coverage of important trait sites usually make it challenging to increase sample numbers in each sequencing run to further drive down sequencing depth. To resolve this, additional target capture library preparations to enrich the interest sites can provide valuable information when combined with low pass sequencing. However, these additional library preparation processes are usually costly and complex.

Recently, we explored the application of a probe-specific flow cell to enable low pass sequencing and selected enrichment simultaneously, using a standard whole genome sequencing library.

We collected tens to hundreds of important single nucleotide polymorphisms (SNP) / Quantitative Trait Nucleotides (QTN) in maize and assigned ranking scores to them. As an initial test, we designed the corresponding probes targeting 11 SNP/QTN sites and produced probe-specific enriched flow cells for this study. The whole genome sequencing library preparation was performed using gDNA extracted from the seeds of maize inbred and hybrid strains (B73, B73xMo17). The libraries were sequenced on an AVITI™ instrument with probe-specific flow cells, without further enrichment process. The whole genome coverage CV below 0.05-0.06. Comparing to the coverage of the genome, we observed 40-50x fold-enrichment across the 11 probe sites, providing enough coverage to confidently call SNP/QTN.

This study showed this novel technology of probe-specific enriched flow cell achieved uniform whole genome coverage and simultaneously captured interesting trait SNP/QTN sites with higher coverage to confidently provide genotyping information at much lower low cost and with a simplified workflow. The described workflow bridges the gap between the completeness of sequencing with the cost-effectiveness of microarray approaches and will facilitate the breeding process in many agricultural species.

Results

Figure 1 – Element Weighted Low Pass Sequencing using Trinity Workflow

Trinity workflow enables target enrichment without hands-on process of hybridization. Pre-enriched library prepared was loaded to AVITI Trinity flow cell. Targeted regions were enriched on the flow cell and sequenced with Avidite Based Chemistry.

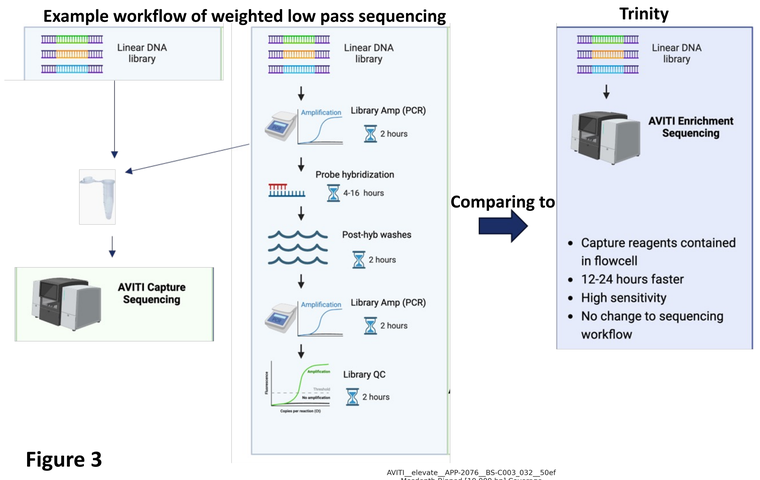


Figure 2 – Maize Elevate PCR Plus Library Prep and Sequencing

Elevate PCR Plus Library Prep workflow generated 8 indexed libraries with gDNA extracted from maize seeds of B73 and B73xMo17 strains. The enrichment was performed on flow cells with designed probes. Sequenced at 2x150 and achieved >90% Q30.

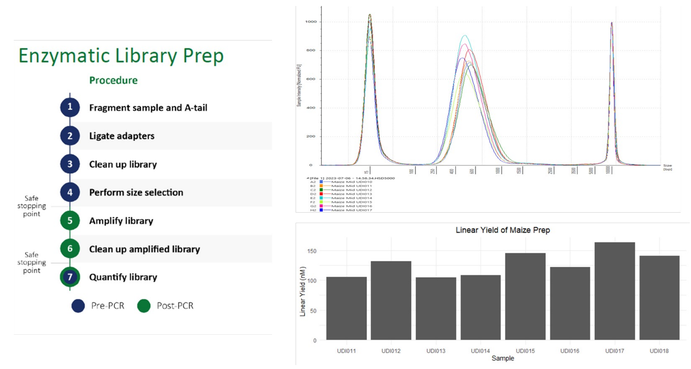


Figure 3

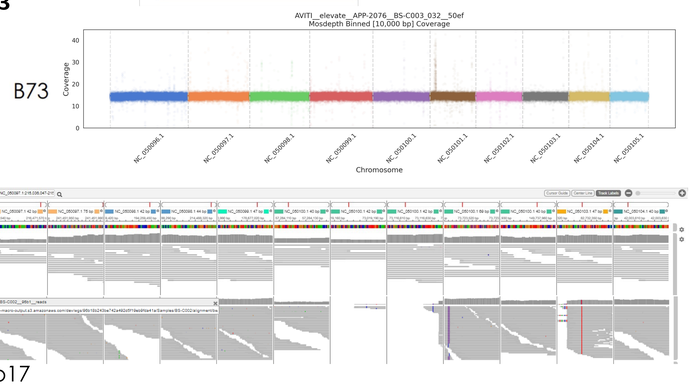


Figure 4a

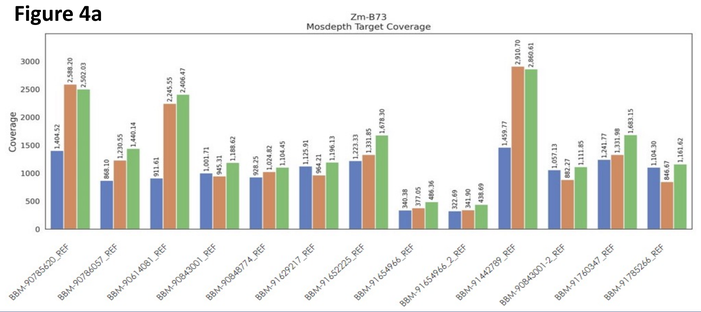


Figure 3 – Genome Coverage

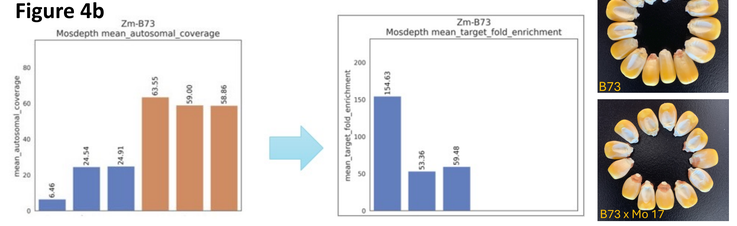
Even coverage throughout the genome. All regions with designed probes were captured and enriched in final sequencing data. Example of genome coverage from B73 whole genome sequencing coverage and B73xMo17 strain with Trinity workflow respectively.

Figure 4 – Target Enrichment

a. Comparing to the whole genome coverage, 13 interesting trait SNP/QTN sites were simultaneously captured by designed probes on flow cell and showed 300-1000x coverage.

b. Overall target regions achieved 50x to 150x fold-enrichment comparing to the whole genome background in 3 different workflow conditions.

Figure 4b



Conclusions

- This study showed this novel technology to combine probe-specific enrichment and uniform whole genome sequencing in one workflow, taking out the burden of manual target enrichment, saving cost and time.
- Using this technology, Element Trinity workflow can use pre-enrichment library as input, and achieve tunable 50x to150x fold-enrichment of interested target region with designed probe.
- Combining low pass whole genome sequencing and tunable coverage of interesting trait SNP/QTN sites will facilitate the breeding process in many agricultural species.

Probe Design

Table 1 - List of Maize Probe Information

probe_id	Gene_ID	Best-hit-arabi-name	Best-hit-arabi-defline	Best-hit-rice-name	Best-hit-rice-defline	trait
BBM-90785620	Zm00001eb108060	AT2G21470	SUMO-activating enzyme 2	LOC_Os07g39780	SUMO-activating enzyme subunit 2, putative, expressed	KMC10|KDR15
BBM-86291963	Zm00001eb210140	AT5G60640	PDI-like 1-4	LOC_Os02g01010	OsPDIL1-4 protein disulfide isomerase PDIL1-4 expressed	KMC20|KMC25|KMC30|KDR15
BBM-90786057	Zm00001eb108570	AT5G61060	histone deacetylase 5	LOC_Os07g41090	histone deacetylase, putative, expressed	KMC25|KMC30|KMC35|KDR35
BBM-91760347	Zm00001eb345900	AT1G06890	nodulin MtN21 /EamA-like transporter family protein	LOC_Os05g07670	solute carrier family 35 member E3, putative, expressed	KMC40|KDR40
BBM-90848774	NA	NA	NA	NA	NA	KDR15
BBM-91442789	Zm00001eb230510	AT5G37475	Translation initiation factor eIF3 subunit	LOC_Os02g02990	translation initiation factor eIF3 subunit putative expressed	KDR30
BBM-91654966	NA	NA	NA	NA	NA	KDR40
BBM-91785266	NA	NA	NA	NA	NA	KMC10|KMC15
BBM-91629217	Zm00001eb192250	NA	NA	LOC_Os02g55520	zinc finger C3HC4 type domain containing protein expressed	KMC30
BBM-90614081	Zm00001eb117740	NA	NA	LOC_Os07g04550	pleckstrin homology domain-containing protein putative expressed	KMC30|KDR30

1. The SNP postion is based on maize B73 RefGen_v5 genome
2. The best hit in arabi and rice comes from Phytozome database
3. related traits are based on article Xu C, Ren Y, Jian Y, Guo Z, Zhang Y, Xie C, Fu J, Wang H, Wang G, Xu Y, Li P, Zou C. Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. Mol Breed. 2017;37(3):20. doi: 10.1007/s11032-017-0622-z. Epub 2017 Feb 16. PMID: 28255264; PMCID: PMC5311085.

© 2024 Element Biosciences, Inc. All rights reserved. Element Biosciences, Avidity Sequencing, AVITI, and the Element Biosciences logo are trademarks of Element Biosciences, Inc. Other names mentioned herein may be trademarks of their respective companies.
Visit elementbiosciences.com for more information.

For research use only

Visit **elementbiosciences.com** for more information.