A new approach to genotyping: Single Primer Enrichment Technology (SPET), an integrated system for both targeted and de novo genotyping

Introduction
DNA array-based technology is a powerful technique for the detection of targeted SNPs, but it is limited in throughput scalability and prone to ascertainment bias.

Genotyping-by-sequencing (GBS) is a robust approach for enabling large scale, whole-genome studies of genetic variation by the random genotyping of a reduced fraction of the genome.

Single Primer Enrichment Technology (SPET) is an innovative approach that integrates both the targeting of known polymorphisms, allowing to perform targeted genotyping, and a de novo genotyping allowing random SNP discovery. The Allegro Targeted Genotyping (NuGEN Inc.) for SPET analysis relies on a panel of probes targeting selected SNPs and leverages sequencing information for novel allele discovery.

Aim of the study
Two replicates of five Zea mays inbred lines (F7, H99, HP301, Mo17 and W153R) and five F1 crosses (4632 x B73, B73 x B616, B73 x F7, B73 x Mo17 and W135 x HP301) were assayed using the targeted genotyping method.

The aim was to measure the performance of the SPET technology comparing it to the array-based technology, and to validate the method as a genotyping solution suitable for both targeted and de novo genotyping.

Results
- Samples were sequenced on either the Illumina HSeq2500 or NextSeq500 to produce single-end 125 bp reads. Sequencing, alignment and coverage statistics are shown in Table 1.
- The target site coverage distribution (Figure 3) showed an homogeneous performance of the probe panel at a coverage of about 50x.
- The accuracy of the targeted genotyping was about 98.2% considering all positions at a coverage of 20x, and reached a plateau at about 50x. Positions targeted by two probes had a higher percentage of accuracy (98.5% at 20x coverage) with SNP Chip (Figure 4A). The accuracy at homozygous sites was higher with respect to heterozygous ones (Figure 4B).
- The reproducibility of the method, obtained comparing the genotype calls for the two replicates, ranged from 97.87% to 99.73% (Figure 5). Increasing the minimum required coverage led to an increase of reproducibility, with a plateau at about 50x.
- Our results validated 98.2% of polymorphic sites compared with the SNP Chip, with 381,575 true calls over 388,551 total calls at a coverage of 20x.
- The de novo genotyping allowed the discovery of additional polymorphic sites in the regions targeted by the probes. The frequency of the alternative allele in the discovered polymorphic positions was lower than the one in the target site positions (Figure 6).

Sequencing and Alignment Statistics

<table>
<thead>
<tr>
<th>Sample</th>
<th>Read number</th>
<th>Mean Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample A</td>
<td>25,701,144</td>
<td>50</td>
</tr>
<tr>
<td>Sample B</td>
<td>21,975,658</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 1. Sequencing and alignment statistics. The table reports the mean total number of sequenced reads, the mean number of high-quality aligned reads and the mean coverage obtained on the target regions across samples.

Conclusions
The analysis of the whole region captured by the probes designed to target the 27,236 sites allowed the detection of 49,443 additional polymorphic sites, present in at least one of the analyzed samples, thanks to the de novo genotyping approach.

Our experiment showed that SPET technology is a powerful tool for high-throughput, cost-effective genotyping. Further, thanks to the de novo SNP genotyping to circumvent ascertainment bias, which is one limitation of the array technology, the results showed that the SPET approach is promising solution for a wide range of genomic analysis, from fine mapping to GWAS, due to its combination of targeted genotyping and de novo SNP discovery.

References
- Martin, M. (2011). Cotadapt (Martin, 2011) and ERNE filter (same sourceforge.net) were used to remove adaptor sequences and low quality 3’ ends from short reads, respectively.
- Reads were aligned to the Zea mays v4 reference genome using the short read alignt BWA-MEM (Li et Durbin, 2009). SNP calling was performed on uniquely aligned reads using the GATK software (McKenna et al., 2010).

Acknowledgements
Novel variation in plant breeding and the plant pan-genomes (NOAHGRD). Grant agreement n. 29478 - European Research Council.

E-mail: afromasierno@appliedgenomics.org