Editorial   **Open Access**

SNP Array – A Powerful Platform to Accelerate Genetic Studies and Breeding

Yongzhong Xing*

National Key Lab of Crop Genetic Improvement, Huazhong Agricultural University, China

Lots of traits in cereals are quantitatively inherited and controlled by multiple genes. In general, these kinds of traits have wide natural variations. Linkage analysis and Linkage Disequilibrium (LD) analysis are the main approaches to discover and locate target genes underlying the traits. Several kinds of molecular markers such as restriction fragment length polymorphism and simple sequence repeats have been used for genetic mapping. But these markers are not suitable for developing a high-density genetic map. Due to its abundance and uniform distribution throughout a genome, and single nucleotide polymorphisms (SNPs) are considered to be the most desirable molecular markers and have been demonstrated to be efficient markers for developing high-density genome scan [1,2]. Whole-genome sequencing and oligonucleotide microarray are the two main strategies used to create SNP markers. Because of its economic cost, SNP array, a high-throughput genome scan, is an important tool for genetic studies and breeding applications. A SNP array designed with a very huge number of SNPs evenly spaced genetically across genomes has been developed in maize, rice and wheat [3-5]. But currently, they are not widely used for genetic analysis. The most important factor limited the application of SNP array is its comparative high cost to researchers. Meanwhile, the researchers also focus on how many SNPs are positioned on the array and when low-, medium- or high-density SNP arrays are used.

High-throughput and low-cost next generation sequencing is a powerful approach to identify SNPs. Sequencing a SNP detection panel of few genotypes identifies the potential SNPs and then a SNP validation panel of several genotypes is used to filter informative SNPs for SNP array development. For species with a reference genome sequence, it is easy to locate them to genome position. But for species without reference genome sequence, the positions of SNPs could be located to linkage groups by a mapping population. Transcriptome re-sequencing allows rapid and inexpensive SNP discovery within genes and avoids highly repetitive regions of a genome, such as in maize. In addition, SNPs come from the expressed sequence tags that are available in some databases. SNPs discovered from certain genotypes are available to make arrays. For example, an Affymetrix 44K SNP array has been developed and used for GWAS analysis [6,7]. But most SNPs on the array were selected from the Oryza SNP project [8], in which only 20 accessions were included. Thus, a question is raised that these chips may not be suitable for projects utilizing unrelated genotypes because of their inadequate representation in the rice species. Considering the importance of general utility of SNP array within a species, it would be better that SNPs collected in a diverse germplasm are selected to load on the array. Re-sequencing hundreds of diverse landraces would improve the quality and general utility of SNPs. Genome scan platforms in rice have been developed using Illumina Infinium platform based on re-sequencing of several hundreds of varieties [5,9]. For genomic selection, SNP array should contain as many functional markers as possible that involved in well-characterized genes controlling important traits besides the characters mentioned above.

As to the question how many SNPs on the array are optimal, the answer depends on the research purpose and the target plant species. For QTL mapping, several thousands of SNPs are enough for a medium size of mapping population to generate a high-density map either in a small or large size genome or complex polyploidy crops. For high resolution QTL mapping nowadays, the bottleneck is population size rather than marker density. Many markers are co-segregated and redundancy in a small population [10]. Mapping populations of permanent lines such as recombinant inbred lines and double haploids are recommended because these lines can be repeatedly phenol typed in several conditions [11]. To maximize the value of SNP array in QTL mapping, a large population of about 500 lines would be expected. In addition, a medium density SNP array is enough to make phylogenetic analysis, which is helpful to make genetic diversity analysis and select parental lines in breeding programs. While for Genome-Wide Association Studies (GWAS), these numbers are not adequate. The number of SNPs used for GWAS dependents on LD decay. In rice, LD decay is slow, extended to a 200 kb region on average, that indicated the maximum mapping resolution by GWAS is about 200 kb interval [7,12]. Therefore, a reasonable number of SNPs like thousands of hundreds of SNPs are required which ensures 1 SNP per 50 kb. But in maize with a large genome and a rapid LD decay, several millions of SNPs are required to ensure the density of 1 SNP per 10 kb, which is possible to identify a causative mutation at single gene level [13]. MutMap is a novel method to quickly detect mutant based on the polymorphism between two bulked segregation pools detected by re-sequencing [14]. High density SNP array has potential in such analysis [9].

The assay of SNP array commonly includes the three major steps : chip hybridization, chip scanning and raw data analysis. In the process, a set of expensive facility is needed. In general, individual researcher or institute has no ability or no necessary to set up such facility to perform the high throughput genotyping. Thus, most scientists urgently need the service of SNP array assay, but on the other hand, they pay more attentions to the service charge. China National Seed Group Co., Ltd has established SNP arrays for different research purposes in rice and maize. Their costs are dependent on the SNP density of the array and the number of samples to be genotyped. In rice, it costs 150 USD to genotype one individual using RICE6K SNP array in China. But the cost increases up to 300 USD using RICE60K array. On average, developing a high-density genetic linkage map with a population of 200 individual scosts about 30,000 USD, which is acceptable as compared to that using traditional marker. But for genome selection, which is intended to track several complex traits by a high-density marker

***Corresponding author:** Yongzhong Xing, National Key Lab of Crop Genetic Improvement, Huazhong Agricultural University, China, Tel: 86-13720378572; E-mail: yzhxing@mail.hzau.edu.cn

array, only a few promising improved plants would move to further selection in a breeding program, thus 300 USD per genotype could be bear for genotyping with a high throughput SNP array. Especially for deciphering the linkage drag redundancy, high-density SNP array is more efficient.

It is believed that the cost of SNP array assay will be gradually decreased. The reasonable prices and quick high throughput genotyping of SNP array are making the use of SNPs even more attractive and efficient. SNP array will accelerate genetics studies and bridges the gap between genomics and breeding in cereals.

References

1. Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082.

2. Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101: 5-18.

3. Ganal MW, Durstewitz G, Polley A, Be´rard A, Buckler ES, et al. (2011) A large maize (Zea mays L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome.

4. Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploidy wheat with the Illumina Golden Gate assay. Theor Appl Genet 119:507–517.

5. Mc Couch SR, Zhao K, Wright M, Tung CW, Ebana K, et al. (2010) Development of genome-wide SNP assays for rice. Breed Sci. 60:524–535.

6. Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, et al. (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa.* Nat Commun 2: 467.

7. Mc Nally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, et al. (2009) Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106: 12273–12278.

8. Chen HD, Xie WB, He H, Yu HH, Chen W, et al. (2013) A High-Density SNP Genotyping Array for Rice Biology and Molecular Breeding. Mol Plant.

9. Yu H, Xie W, Li J, Zhou F, Zhang Q, et al. (2014) A whole- genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol 12: 28-37.

10. Tan C, Han ZM, Yu HH, Zhan W, Xie WB, et al. (2013) QTL scanning for rice yieldusing a whole genome SNP array. J Genetics and Genomics 40: 629-638.

11. Hu W, Wen M, Han ZM, Tan C, Xing YZ, et al. (2013) Scanning QTLs for Grain Shape Using a Whole Genome SNP Array in Rice. J Plant Biochem Physiol 1: 104.

12. Huang X, Zhao Y, Wei X, Li C, Wang A, et al. (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44: 32–39.

13. Yan JB, Warburton MC rouch J (2011) Association mapping for enhancing maize genetic improvement. Crop Sci 51:433-449.

14. Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, et al. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30: 174–178.