Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Plant Biology

Development of genomics-based genotyping platforms and their applications in rice breeding

Haodong Chen[1,3], Hang He[1,3], Fasong Zhou[2], Huihui Yu[2] and Xing Wang Deng[1,3,4]

Breeding by design has been an aspiration of researchers in the plant sciences for a decade. With the rapid development of genomics-based genotyping platforms and available of hundreds of functional genes/alleles in related to important traits, however, it may now be possible to turn this enduring ambition into a practical reality. Rice has a relatively simple genome comparing to other crops, and its genome composition and genetic behavior have been extensively investigated. Recently, rice has been taken as a model crop to perform breeding by design. The essential process of breeding by design is to integrate functional genes/alleles in an ideal genetic background, which requires high throughput genotyping platforms to screen for expected genotypes. With large amount of genome resequencing data and high-throughput genotyping technologies available, quite a number of genomics-based genotyping platforms have been developed. These platforms are widely used in genetic mapping, integration of target traits via marker-assisted backcrossing (MABC), pyramiding, recurrent selection (MARS) or genomic selection (GS). Here, we summarize and discuss recent exciting development of rice genomics-based genotyping platforms and their applications in molecular breeding.

Addresses
[1] Peking-Yale Joint Center of Plant Molecular Genetics and Agrobiotechnology, State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China
[2] Life Science and Technology Center, China National Seed Group Co., Ltd., Wuhan 430075, China
[3] Shenzhen Institute of Crop Molecular Design, Shenzhen 518107, China
[4] Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520-8104, USA

Corresponding authors: Chen, Haodong (chenhaodong@pku.edu.cn) and Deng, Xing Wang (xingwang.deng@yale.edu)

Current Opinion in Plant Biology 2013, **16**:247–254

This review comes from a themed issue on **Genome studies and molecular genetics**

Edited by **Qifa Zhang** and **Rod Wing**

For a complete overview see the Issue and the Editorial

Available online 21st May 2013

1369-5266/$ – see front matter,

http://dx.doi.org/10.1016/j.pbi.2013.04.002

Introduction

The concept of 'Breeding by Design' was proposed by Peleman and van der Voort in 2003, aiming to control allelic variations of all important trait-related genes [1]. However, without enough genome information and high throughput genotyping tools, this concept was hardly practiced in empirical breeding. Since the rice genome was sequenced [2–4], significant advancements have been made in the functional genomics of rice (*Oryza sativa* L) that have offered breeders numerous tools and resources to practice breeding by design. The fast accumulation of rice genome resequencing data not only assisted in identification of functional Quantitative Trait Loci (QTL) or genes, but also provided numerous polymorphic genome sequences for molecular marker-development [5••,6,7••,8•]. At the same time, a variety of molecular marker assay platforms with different throughputs have also recently been developed [5••,7••,9••,10]. All these achievements have thus turned the 'design' concept into a practical breeding activity. Here we focus on recent development of genomics-based genotyping platforms and their applications in rice molecular breeding by design.

Development of genotyping platforms

Molecular markers are widely used in genetic studies and new variety development. Various types of genotyping technologies have been developed to meet the requirements of genetic research or breeding programs in rice. Restriction Fragment Length Polymorphism (RFLP) and Simple Sequence Repeats (SSRs) are the representatives of first and second generation markers [11,12]. They played important roles in construction of rice genetic maps and identification of trait-related loci/genes. SSR markers are still frequently used by breeders to assist their genotype screen for target traits. In this review, we will focus on the most recent development of high throughput genotyping platforms, specifically on DNA array platforms and next-generation sequencing (NGS) technologies.

Array-based genotyping

Recently, a number of array-based genotyping technologies have been developed and some of the representatives including Restriction Site-Associated DNA (RAD), Single Feature Polymorphism (SFP) and Single Nucleotide Polymorphism (SNP) have been widely tested. With short DNA tags as probes RAD markers identify genetic polymorphisms linked to particular restriction sites throughout the genome. RAD markers have been successfully used in genotyping both individuals and segregating bulks [13]. SFP markers detect polymorphic signals resulted from differential hybridization of various alleles to DNA probes arrayed on a microchip. Affymetrix

microarrays, composed of 25 nucleotide probes, have been shown to be efficient and convenient in detecting very large number of SFPs in rice [14–17].

SNPs resulted from single-base pair variations are the most abundant DNA markers that are evenly distributed on a whole genome [18]. Almost any gene or locus can be tagged by SNP markers. Extremely high density of a SNP array can assay large number of SNP markers in a high throughput manner. SNP array-based genotyping platform has been considered as one of most favorite options for gene/QTL mapping and marker-assisted crop breeding in the past decade [19]. Furthermore, with the rapid accumulation of rice genome resequencing data, SNP-based markers will continue to be more widely used than any other type of markers [5••,6,7••,20]. Although SNPs can also be detected via PCR or Sanger sequencing, array-based detection techniques are preferable given that they can satisfy different genotyping requirements. Currently, SNPs can be detected in a variety of throughputs, depending on the objectives. Many SNP assay systems have been developed by different companies, such as Illumina's SNP chip platforms (http://www.illumina.com), Affymetrix's SNP array platform (http://www.affymetrix.com), GenomeLab's SNPstream genotyping system (Beckman Coulter, https://www.beckmancoulter.com), and the TaqMan OpenArray genotyping system (Applied Biosystems, http://www.appliedbiosystems.com). Given that Illumina and Affymetrix SNP arrays are more commonly used in the rice community, they will be discussed here in more detail.

A combination of Veracode and GoldenGate technology on Illumina's BeadXpress Reader can be used to genotype 48-SNP, 96-SNP, 192-SNP or 384-SNP per sample. SNPs and their flanking sequences are used to design locus-specific and allele-specific primers for GoldenGate assay. This platform, which can be used to assay thousands of individual samples in a short time period, is both reliable and relatively inexpensive. Recently, several 384-SNP assays on this platform have been developed and used for both variation evaluation and genetic diversity analysis in rice [21,22,23•,24•]. This low-density SNP array platform is useful for the genotyping of early generation breeding materials due to its high-throughput capacity in sample processing.

GoldenGate SNP Beadarrays can provide medium resolution genotyping results. For example, a 1536-SNP GoldenGate array was designed to detect polymorphism within and between the five major subpopulations of *O. sativa* [25]. In another study, a set of 2688 SNPs were used to genotype 151 Japanese rice cultivars that had been released over the last 150 years [26].

The high density Affymetrix SNP chip and the Illumina Infinium SNP chip can be used to perform whole-genome selection. Two high-resolution Affymetrix custom arrays have been designed for rice, one consisting of ~44,000 SNPs (44 K array) and another consisting of ~1 million SNPs (1 M array) [9••,27]. Both have been used to assay genome-wide patterns of genetic variation in worldwide collections of wild and cultivated rice accessions. Recently, we developed another high-resolution SNP array for rice based on Illumina's Infinium platform with 51,478 markers, dubbed RiceSNP50 (unpublished results). SNP probes on this array were selected and designed based on an analysis of more than 10,000,000 SNPs extracted from the sequence data of 801 rice accessions. These SNPs were both preferentially located in genes and evenly distributed across the genome. Thus, it is feasible for the rice community to use this high-resolution SNP array for wide variety of research objectives.

Next-generation sequencing (NGS) technologies

NGS platforms, including Illumina HiSeq2500, ABI 5500xl SOLiD, Roche 454, Ion Torrent and PacBio RS, have been rapidly developed as of late. These platforms make whole-genome sequencing accessible to regular laboratories, especially those seeking to re-sequence species for which there are complete reference genome sequences in existence, such as rice. With advances in NGS technologies, the traditional two-step paradigm of SNP discovery and subsequent assay has been simplified into a single process, in which bioinformatics tools simultaneously analyze the sequence data for both SNP discovery and genotyping [5••,6,7••,8•,28,29•].

Identification of genetic variations controlling rice agronomic traits

For designing ideal rice with superior genotypes, the prerequirement is to understand the genetic basis of agronomically important traits and the allelic variation at those loci. With the development of genomics platforms, more and more loci involved in agronomic traits have been mapped, and their allelic variations have been assessed.

The traditional method for gene isolation or marker–trait association analysis in rice is QTL mapping (also known as linkage mapping), which has been used extensively to identify natural mutations of agriculturally important traits. NGS has been used to construct a genetic map for 150 rice recombinant inbred lines (RILs), which was 35 times more precise in recombination breakpoint determination [28]. Further, 49 QTL with phenotypic effect ranging from 3.2 to 46.0% for 14 agronomic traits were detected in these RILs, which indicated that NGS could provide a powerful solution to map QTL with high resolution [30]. On the basis of low-coverage sequences of RILs, a high-density linkage map was constructed using high-quality SNPs in RILs without genotype data of the parental lines [29•]. The new SNP map detected more QTL with precise map locations, showing advantages in detecting power and resolution compared to the RFLP/SSR map [31].