

# Allele Mining and Deployment for Rice Improvement Through Sequencing and SNP Genotyping: Status and Potential

Michael J. Thomson

International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines. Email: M.Thomson@cgiar.org

## Abstract

*One of the main objectives of building up genomic resources in a crop such as rice is to provide a deeper understanding of the important genes and alleles for crop improvement. In the end, the most valuable contribution of genomics will be to accelerate molecular breeding efforts to meet the challenges of raising rice production in a sustainable way under increasingly difficult conditions. Thus, if one looks at genomics research with this ultimate goal in mind, focusing on the practical needs of breeding applications will help to identify the most relevant discoveries from upstream research. Recent advances in QTL mapping, molecular genetics and functional genomics have provided a wealth of information about the key genes and pathways involved in many traits in rice. At the same time, rapid developments in sequencing and single nucleotide polymorphism (SNP) genotyping technologies are leading to a massive amount of data on allelic variation across wild and cultivated rice accessions. The convergence of these two areas of research is quickly leading to unprecedented opportunities to identify essential genes controlling traits of interest, characterize and select the most promising alleles at each genetic locus, and use those alleles in precise molecular breeding strategies for rice improvement. This review will briefly examine the recent advances in SNP discovery, high-throughput SNP genotyping, and allele mining in rice with a focus on innovative methods and future prospects for applying these discoveries to enable more powerful and precise molecular breeding approaches.*

**Keywords:** Rice, allele mining, sequencing, SNP genotyping

## Review of recent advances in allele mining

The basis for developing more precise molecular markers for breeding applications is allele mining, which can be defined as the process of finding the best alleles from available germplasm at the key genes controlling the traits of interest. The process of allele mining requires two essential components: (1) identifying the genes controlling traits of interest, and (2) finding the best alleles providing the desired trait characters. Subsequently, gene-based molecular markers can be developed that are diagnostic or predictive of the trait characters. The ideal scenario is when the causal variant, also referred to as the functional nucleotide polymorphism (FNP) or quantitative trait nucleotide (QTN), can be identified, since that polymorphism can be used to design a functional marker. Functional markers are also referred to as perfect markers, as they will always co-segregate with the trait phenotype since no recombination event can separate this marker from the associated phenotype. The foundation for allele mining began to form in the 1990's with the first successful map-based cloning of

important genes and QTLs in plants. These cloned genes provided the basis for identifying causal variants, even for quantitative traits (Flint and Mott 2001). QTL cloning and the molecular dissection of complex traits have since provided many insights into the genes that control agronomically important traits (Collins et al 2008; Yamamoto et al 2009). These genes then become targets for more detailed characterization of the alleles and specific polymorphisms leading to the desired trait phenotypes. A landmark review has described the process of developing functional markers, whether by association studies or direct proof of allele function (Andersen and Lubberstedt 2003). Since then, several studies have provided excellent examples of functional markers, such as the ones for fragrance and *xa5* resistance in rice (Bradbury et al 2005; Iyer-Pascuzzi and McCouch 2007). A number of reviews have also described efforts to develop functional markers in maize and wheat (Lubberstedt et al. 2005; Bagge et al 2007; Liu et al 2012). A particularly impressive study performed in-depth allele mining at the *Pm3* locus for powdery mildew in wheat: 1,320 landraces were

screened for resistance, with 111 identified as having the *Pm3* gene but with potentially novel alleles, of which 45 were completely sequenced—leading to the identification of 16 previously unknown *Pm3* alleles leading to tolerance (Bhullar et al 2009). However, the total number of functional markers in rice remains quite low, due to the time and resources needed to functionally characterize polymorphisms at important genes. In the future, these constraints on allele mining and marker development will be reduced as more genomics resources become integrated into efforts for molecular breeding (Varshney et al 2005; Glaszmann et al 2010; Fukuoka et al 2010). Once targeted functional markers become available, they can be used for precise marker-assisted backcrossing (MABC) to transfer QTLs (Collard et al 2008), as well as marker-assisted recurrent selection and genomic selection to include many loci of small effect in the selection model (Bernardo 2008).

New advances in SNP genotyping and sequencing technologies also promise to accelerate progress in allele mining. Early SNP discovery efforts included a survey of 160,000 SNPs across 20 varieties in the *Oryza* SNP project (McNally et al 2009), which led to the development of a 44K SNP chip that has enabled genome-wide association studies to identify key loci controlling important traits (Zhao et al 2011; Famoso et al 2011). High-throughput genotyping sets using Illumina 384-plex SNP scans are also providing more efficient workflows for mapping, diversity analysis and molecular breeding (Chen et al 2011; Thomson et al 2012). Although multiplex SNP genotyping technologies can provide high quality data with low levels of missing data, the current trend is favoring next-generation sequencing (NGS) as the technique of choice for most allele mining and crop improvement applications (Varshney et al 2009). Resequencing cultivated and wild rice accessions, has led to the identification of regions under selection during domestication (He et al 2011; Xu et al 2012a). Low coverage next-gen sequencing has also enabled a high-resolution genome-wide association study of 950 rice varieties to identify important loci (Huang et al 2012). Moreover, multiplexed sequencing employing barcoded adaptors at restriction enzyme cut sites has led to a robust genotyping by sequencing (GBS) approach for low-cost genome-wide SNP scans (Elshire et al 2011). Integrated approaches employing the latest in sequencing technology are providing a greater opportunity for improving the entire range of molecular breeding techniques (Xu et al 2012b). Perhaps the most promising of the NGS applications involves rapid fine-mapping of genes and mutations by combining bulk segregant analysis with whole genome sequencing using recently described techniques such as SHOREmap, next-generation mapping, MutMap and related methods

(Schneeberger et al 2009; Austin et al 2011; Schneeberger and Weigel 2011; Mokry et al 2011; Abe et al 2012). These advances have now set the stage for large scale allele mining efforts that will provide an excellent opportunity to design the next generation of functional markers for molecular breeding in rice.

### **Innovative studies underway for genome-wide association studies**

There are several innovative studies currently underway that will enable more powerful allele mining through association analysis and SNP discovery. The first is the development of a Rice Diversity Research Platform led by Cornell University and supported by IRRI based on a high resolution SNP chip that should provide over 700,000 SNPs on approximately 2,000 rice accessions (McCouch et al 2010; Tung et al 2010). These accessions will be characterized in depth by a Global Phenotyping Network through the Global Rice Science Partnership (GRiSP) to provide a wealth of data for the largest genome-wide association study (GWAS) in any plant species to date. By combining high quality genome-wide SNP data with precise phenotyping, valuable marker-trait associations and beneficial alleles will be identified across important traits. A complementary approach using next-generation sequencing will provide greater ability to identify all potential polymorphisms, including insertion-deletion (INDEL) sites and rare SNPs. Over 100 rice accessions have been resequenced through the Rice SNP Consortium ([www.ricesnp.org](http://www.ricesnp.org)), providing information on over 27 million SNPs across these diverse accessions (Susan McCouch, pers. comm.). A larger project based on collaboration between scientists at the Chinese Academy of Agricultural Sciences, BGI-Shenzen, and IRRI is currently sequencing several thousand rice accessions as the first phase of IRRI's vision to sequence over 100,000 accessions in the TT Chang Genetic Resources Center at IRRI (Ken McNally, pers. comm.). Once complete, the vast amount of data on sequence variation across global rice accessions will provide a rich pool of DNA polymorphism data from which to mine useful alleles for breeding applications for many years to come.

### **Prospects for developing functional markers**

Although the recent progress in developing new genomics resources in rice has been staggering, there remain several challenges that need to be addressed before the full potential of allele mining can be realized. The primary challenge is the great need for many more bioinformatics specialists to focus their efforts on rice sequence analysis—not only to analyze the data, but also to develop the tools required to translate the tsunami of sequence data into user-friendly forms to enable other scientists to interact with the data and

extract useful information. It will take a massive investment to increase the bioinformatics capacity of the rice community, and a concerted effort for informatics specialists to work more closely with geneticists and breeders to better understand the deeper issues involved in identifying useful alleles from the germplasm collections. There is a need for software that will allow unique SNP haplotypes across specific regions to be clearly defined, visualized and correlated with phenotypes associated with the underlying germplasm. Such a tool could also highlight SNPs and INDELs that have potential implications for changing gene function, such as those that introduce premature stop codons, alternate splice sites, and non-synonymous amino acid changes. These tools would allow geneticists to focus into regions of the genome containing known genes, fine-mapped QTLs and candidate genes for traits of interest, and closely examine patterns of sequence variation that might affect gene function. By including high resolution SNP and sequence data from thousands of wild and cultivated rice accessions, the entire range of genetic diversity can be screened even for rare alleles that affect key traits. The integration of genome-wide sequence polymorphism data with detailed phenotypic characterization, combined with new tools for SNP haplotype analysis and visualization, will provide the resources needed for rice scientists to further investigate and dissect the molecular basis of trait variation in rice. This will enable the identification of functional markers and diagnostic SNP haplotypes that are predictive of the desired trait characters—thus ushering in a new era of precise allele mining and functional marker development to accelerate molecular breeding efforts. Once the causal variants are identified, these markers need to be tested on high throughput SNP genotyping platforms, such as the Fluidigm EP1 system, and validated across larger germplasm panels to verify their ability to predict the desired phenotypes. Successfully validated markers can then be deployed across the breeding programs to enable more efficient selection methods and more rapid gains in rice improvement to meet future challenges.

#### REFERENCES

- Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S and Terauchi R.** 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotech* 30: 174-178.
- Andersen JR and Lübberstedt T.** 2003. Functional markers in plants. *Trends Plant Sci* 8: 554-560.
- Austin RS, Vidaurre D, Stamatou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, McCourt P and Guttman DS.** 2011. Next-generation mapping of *Arabidopsis* genes. *Plant J* 67: 715-725.
- Bagge M, Xia X and Lübberstedt T.** 2007. Functional markers in wheat. *Curr Opin Plant Biol* 10(2):211-216.
- Bernardo R.** 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci* 48: 1649-1664.
- Bertrand C, Collard Y and Mackill DJ.** 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil Trans R Soc B* 363: 557-572.
- Bhullar NK, Street K, Mackay M, Yahiaoui N and Keller B.** 2009. Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus. *Proc Natl Acad Sci USA* 106: 9519-9524.
- Bradbury LMT, Henry RJ, Jin Q, Reinke RF and Waters DLE.** 2005. A perfect marker for fragrance genotyping in rice. *Mol Breed* 16: 279-283.
- Chen H, He H, Zou Y, Chen W, Yu R, Liu X, Yang Y, Gao Y, Xu J, Fan L et al.** 2011. Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L). *Theor Appl Genet* 123:869-879
- Collins NC, Tardieu F and Tuberosa R.** 2008. Quantitative trait loci and crop performance under abiotic stress: where do we stand? *Plant Physiol* 147: 469-486.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES and Mitchell SE.** 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.
- Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, Bustamante C, Kochian LV and McCouch SR.** 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genetics* 7: e1002221.
- Flint J and Mott R.** 2001. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat Rev Genet* 2: 437-445.
- Fukuoka S, Ebana K, Yamamoto T and Yano M.** 2010. Integration of genomics into rice breeding. *Rice* 3: 131-137.
- Glaszmann JC, Kilian B, Upadhyaya HD and Varshney RK.** 2010. Accessing genetic diversity for crop improvement. *Curr Opin Plant Biol* 13: 167-173.
- He Z, Zhai W, Wen H, Tang T, Wang Y, Xuemei Lu, AJ, Hudson RR, Wu C-I and Shi S.** 2011. Two evolutionary histories in the genome of rice: the

- roles of domestication genes. *PLoS Genet* 7: e1002100.
- Huang Y, Zhao X, Wei C, Li A, Wang Q, Zhao W, Li Y, Guo L, Deng C, Zhu 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.
- Iyer-Pascuzzi AS and McCouch SR.** 2007. Functional markers for *xa5*-mediated resistance in rice (*Oryza sativa*, L.). *Mol Breed* 19: 291-296.
- Liu Y, He Z, Appels R and Xia X.** 2012. Functional markers in wheat: current status and future prospects. *Theor Appl Genet* 125: 1-10.
- Lübberstedt T, Zein I, Andersen JR, Wenzel G, Krützfeldt B, Eder J, Ouzunova M and Chun S.** 2005. Development and application of functional markers in maize. *Euphytica* 146: 101-108.
- McCouch SR, Zhao K, Wright M, Tung C-W, Ebana K, Thomson M, Reynolds A, Wang D, DeClerck G, Ali MDL, McClung A, Eizenga G and Bustamante C.** 2010. Development of genome-wide SNP assays for rice. *Breed Sci* 60: 524-535.
- McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, et al.** 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci USA* 106: 12273-12278.
- Mokry M, Nijman IJ, Dijken AV, Benjamins R, Heidstra R, Scheres B and Cuppen E.** 2011. Identification of factors required for meristem function in *Arabidopsis* using a novel next generation sequencing fast forward genetics approach. *BMC Genomics* 12: 256.
- Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D and Andersen SU.** 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* 6: 550 – 551.
- Schneeberger K and Weigel D.** 2011. Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci* 16: 282-288.
- Thomson MJ, Zhao K, Wright M, McNally KKL, Rey J, Tung C-W, Reynolds A, Scheffler B, Eizenga G and McClung A, et al.** 2012. High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Mol Breed* 29:875–886.
- Tung C-W, Zhao K, Wright MH, Ali ML, Jung J, Kimball J, Tyagi W, Thomson MJ, McNally K and Leung H, et al.** 2010. Development of a research platform for dissecting phenotype–genotype associations in rice (*Oryza* spp.). *Rice* 3: 205-217.
- Varshney RK, Graner A and Sorrells ME.** 2005. Genomics-assisted breeding for crop improvement. *Trends Plant Sci* 10: 621-630.
- Varshney RK, Nayak SN, May GD and Jackson SA.** 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotech* 27: 622-630.
- Xu X, Liu X, Ge S, Jensen JD, Hu F, et al.** 2012a. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotech* 30: 105-111.
- Xu Y, Lu Y, Xie C, Gao S, Wan J and Prasanna BM.** 2012b. Whole-genome strategies for marker-assisted plant breeding. *Mol Breed* 29: 833-854.
- Yamamoto, T, Yonemaru J and Yano M.** 2009. Towards the understanding of complex traits in rice: substantially or superficially? *DNA Res* 16: 141–154.
- Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam RM, Reynolds A, Mezey J, McClung AM, Bustamante CD and McCouch SR.** 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Comms* 2:467.

---

**Citation:** Thomson MJ. 2013. Allele mining and deployment for rice improvement through sequencing and SNP genotyping: status and potential. In: Muralidharan K and Siddiq EA, eds. 2013. *International Dialogue on Perception and Prospects of Designer Rice*. Society for Advancement of Rice Research, Directorate of Rice Research, Hyderabad 500030, India, pp 25-28.