# 9 Making Rice Genomics Work for Africa

**Susan McCouch,[1]* Rod A. Wing,[2] Mandè Semon,[3] Ramaiah Venuprasad,[3] Gary Atlin,[4] Mark E. Sorrells[1] and Jean-Luc Jannink[5]**

[1]*Plant Breeding and Genetics Department, Cornell University, Ithaca, New York;* [2]*Arizona Genomics Institute, School of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, Arizona;* [3]*Africa Rice Center (AfricaRice), Ibadan, Nigeria;* [4]*International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico;* [5]*United States Department of Agriculture Agricultural Research Service, Cornell University, Ithaca, New York*

## Introduction

In this chapter we explore opportunities and challenges associated with making rice genomics work for Africa. In the first section, we discuss how sequencing technology is being used to describe the extent, distribution and evolution of genetic variation in African rice. The sequence information is used to discover single nucleotide polymorphisms (SNPs), to inform germplasm conservation efforts, and to provide a framework for dissecting phenotype–genotype relationships. In the second section, we discuss strategies for identifying genotype–phenotype associations and using that information to enhance the efficiency of applied plant breeding programmes. The strategies include quantitative trait locus (QTL) mapping, genome-wide association mapping, marker-assisted selection (MAS) and genomic selection. We take note of some of the biological, logistical and institutional challenges that must be addressed, and envision the development of a functioning breeding pipeline that makes effective use of new tools.

Many of the tools and resources needed to develop genomics-based rice breeding programmes in Africa are already in place. These include diverse germplasm resources, reference genome sequences (RefSeq) for both wild and cultivated *Oryza* species, high- and low-density SNP arrays and sequence-based genotyping platforms, co-ordinated phenotyping networks, and models and strategies for using genomics information to predict phenotypic performance in a breeding programme. Given this foundation, the feasibility of implementing genomics-based breeding in African rice programmes ultimately depends on the ability to integrate large genotypic and phenotypic data sets in real time and the ability to practise rapid-cycle genotype-based selection. To meet this challenge, new types of data management, decision-support systems, and timely nursery management will be required, as well as new institutional partnerships, training programmes and coordinated research networks.

Finally, genomics-facilitated selection can only improve breeding efficiency if useful models are developed linking DNA polymorphisms with plant performance in appropriate field environments. The challenge facing the next generation of rice researchers and breeders is to implement

---

\* Corresponding author: srm4@cornell.edu

cost-effective strategies that take advantage of high-throughput genotyping and increasingly efficient phenotyping strategies to make reliable predictions about plant performance in field environments that are relevant to African farmers.

## What Do We Mean by Genomics?

Genomics is the study of the genomes of organisms, or the entirety of their hereditary information. It includes intensive efforts to determine the complete DNA sequence of individuals, to annotate that sequence by identifying structural features (such as chromosomes, genes, regulatory sequences, repetitive elements, and polymorphisms), and to interpret the functional significance or breeding value of these features. The development of increasingly high-throughput and low-cost sequencing methodologies makes it possible to identify and study genome-wide DNA variation in hundreds, thousands or millions of individuals in real time (Craig *et al.*, 2008; Schuster, 2008; Edwards and Batley, 2010).

These new sequencing technologies have dramatically changed the landscape for detecting and monitoring genetic variation in crop plants (Rafalski, 2002; Duran *et al.*, 2009; Edwards and Batley, 2010). SNPs are the most abundant form of genetic variation in plants and other eukaryotic genomes (Kwok *et al.*, 1996). As genetic markers, SNPs represent sites in the genome where the DNA sequence differs between individuals. SNPs are rapidly replacing simple sequence repeats (SSRs) as the DNA marker of choice for applications in plant breeding and genetics because they are more abundant, stable, amenable to automation, efficient and cost-effective to detect, and straightforward to manage in a database (Schuster, 2008; McCouch *et al.*, 2010). SNPs occur in both coding and non-coding regions of nuclear and plastid DNA, however not all SNPs are equally informative. Informative SNPs must first be polymorphic in the germplasm of interest. If they are polymorphic, they may be informative because they are in or near genes that contribute to a phenotype of interest or they may be useful simply because they provide a fingerprint that uniquely identifies a line or lineage. SNPs that are directly responsible for a phenotype (because they cause a change in the protein product of a gene or alter the way the gene is expressed) are called causative or functional SNPs, while SNPs that do not directly affect the way a gene functions are called indicative or linked SNPs. All types of SNPs are of interest to plant researchers; functional SNPs directly affect plant biology or plant performance, while linked SNPs can be used to track alleles of interest or to trace the evolutionary history of a gene, a region of the genome or the genome as a whole, and collections of genome-wide SNPs, even without knowing if they are linked to a gene or QTL of interest, are useful in genomics-assisted breeding applications (Gupta *et al.*, 2001; Rafalski, 2002).

SNPs are generally discovered by sequencing two or more genomes, aligning the sequences and comparing them. In 2005, rice became the first crop species to have its genome completely sequenced, and this accomplishment rapidly catalysed international interest in rice as a model genome for crop genetic research (IRGSP, 2005). For many years, the *Oryza sativa* subsp. *japonica* (temperate) variety Nipponbare was the only rice genome that had been sequenced to high accuracy via the sequencing of physically aligned bacterial artificial clones (BACs). Additional rice varieties were 're-sequenced' using a shotgun approach, where short sequence reads were aligned to the Nipponbare reference genome. Re-sequencing is a strategy that generates millions of short-sequence reads (50–150 base pairs [bp]) at random throughout the genome. These short sequences are aligned to the higher-quality, fully assembled Nipponbare sequence in order to identify SNPs, insertion/deletions (indels) and other forms of DNA variation that distinguish the re-sequenced genome from Nipponbare. This approach has been used to sequence dozens of diverse rice genomes from different subpopulations and species (Ammiraju *et al.*, 2010a; Huang *et al.*, 2010; McCouch *et al.* 2010; Xu *et al.*, 2011). Once the short reads are aligned to Nipponbare, DNA polymorphisms (SNPs, indels, copy number variants or CNVs) are individually identified based on their nucleotide position in the Nipponbare reference genome. Alignment to a reference genome that has been annotated to identify genes makes it possible to predict whether a SNP falls within or near a gene of interest, and whether a genic SNP is expected to cause a functional change in the protein product (i.e. non-synonymous change) that might alter the

expression of the gene (Ondov *et al.*, 2008). This information can be very useful in predicting whether a particular SNP or indel is responsible for a phenotype of interest.

## Sequencing of *Oryza* Species

### Sequencing of *O. glaberrima*

*Oyrza glaberrima* cv. CG14 was the first African rice genome to be completely sequenced to high accuracy using a pooled BAC approach (Wing *et al.*, 2012, unpublished results). CG14 now serves as an independent reference genome for domesticated African rice, alongside Nipponbare for Asian rice. The *O. glaberrima* RefSeq is being annotated to identify genes, transposons and other functionally relevant motifs. It is also being compared with the Nipponbare sequence to identify structural variation, novel genes and selective sweeps. This analysis will allow us to identify specific genes and QTLs that distinguish *O. glaberrima* from *O. sativa*, and will provide insights about the traits and genomic regions that were under selection by early agriculturalists in West Africa.

### Reference sequences for wild relatives

In addition to the *O. glaberrima* RefSeq, an international effort is under way to generate high-quality de-novo reference sequences from representative accessions of all 21 wild *Oryza* species under the organizational umbrella of the International *Oryza* Map Alignment Project (IOMAP) (Goicoechea *et al.*, 2010) (Fig. 9.1). The wild relatives of rice, although agronomically inferior, contain a virtually untapped reservoir of agriculturally important genes that could be used to improve the cultivated rice species. Once developed, these RefSeqs will be available for SNP calling to facilitate diversity analysis of rice and its wild relatives (Ammiraju *et al.*, 2010a). The IOMAP consortium is emphasizing the generation of high-quality de-novo RefSeqs rather than re-sequences, because of the significant amount of structural variation detected in targeted gene regions. So far, this has been demonstrated both for specific regions, such as *Adh*1, *Moc*1 and *Hd*1 (Ammiraju *et al.*, 2008,

2010b; Lu *et al.*, 2009; Sanyal *et al.*, 2010) and genome wide (Hurwitz *et al.*, 2010). Additionally, efforts are under way to generate de-novo RefSeqs for other AA genome wild relatives, as well as for each of the major subpopulations of *O. sativa*, because it is acknowledged that a single RefSeq is insufficient to comprehensively understand or capture genome diversity within *Oryza*.

### Characterization of genes on chromosome 3, short arm

In addition to these full genome *Oryza* reference sequences, the IOMAP project aims to functionally characterize the majority of genes on the short arm of chromosome 3 as part of an international effort to functionally characterize all rice genes by 2020 (Zhang *et al.*, 2008). Interest in this region of the genome was related to the fact that the Wing laboratory led the US effort to sequence the short arm of chromosome 3 as part of the International Rice Genome Sequencing Project (IRGSP, 2005). A nearly complete set of chromosome 3 short arm RefSeqs is available for 12 *Oryza* species (indicated with an asterisk in Fig. 9.1) and the *Oryza* outgroup species *Leersia perrieri*.

Once these sequences have been annotated to identify the genes and regulatory sequences they contain, it will be possible to determine how much useful variation is found in distantly related wild species and to begin to introgress useful variation (from the AA genome species). It should also be possible to take advantage of the new knowledge contributed by the IOMAP initiative to modify domesticated genomes via transgenics in cases where novel genes are found in genomes that are not sexually compatible with *O. sativa* and *O. glaberrima*.

### Diversity of *Oryza* in Africa

Africa contains the largest diversity of *Oryza* species in the world, with three AA genome species (*O. glaberrima, O. barthii* and *O. longistaminata*), the BB genome species (*O. punctata*), the CC genome species (*O. eichingeri*), the FF genome species (*O. brachyantha*), and finally the BBCC tetraploid species (*O. punctata*). These species span 10–15 million years of evolutionary history
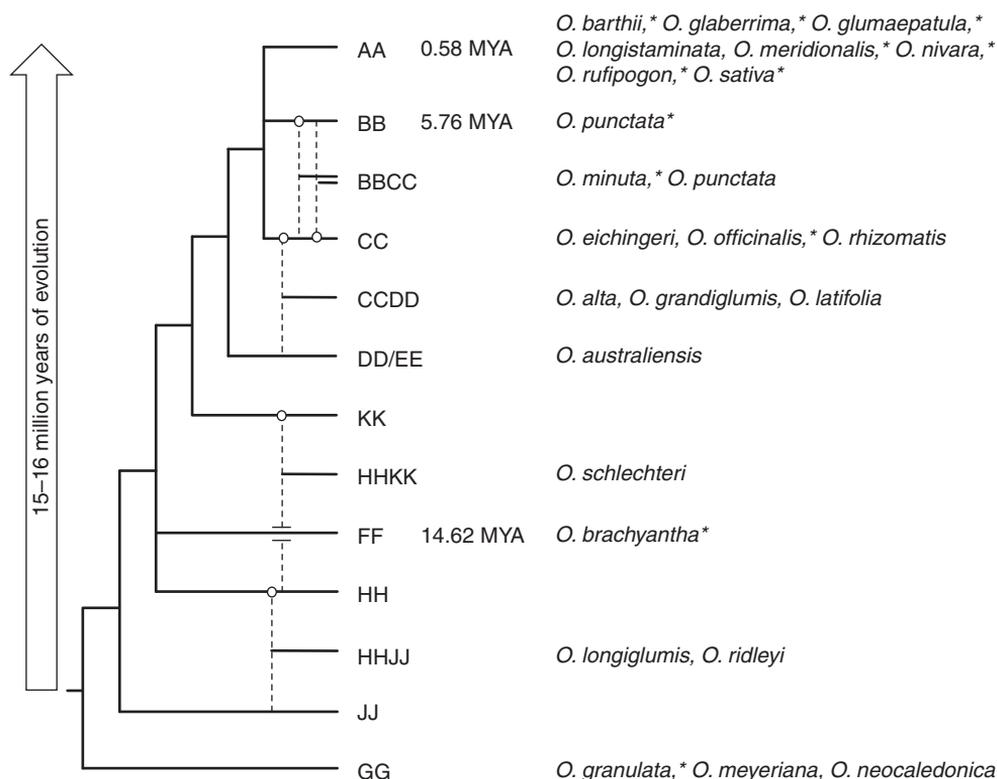
**Fig. 9.1.** *Oryza*: 24 species (two cultivated) – ten genome types. Asterisks indicate species for which a nearly complete set of chromosome 3 short arm reference sequences is available. (After Ge *et al.*, 1999; Ammiraju *et al.*, 2008; www.knowledgebank.irri.org/extension/index.php/wild-rice-taxonomy.) MYA, million years ago.

(Vaughan, 1994), making Africa a critical player in understanding and preserving the genomic diversity of the *Oryza* genus.

### Brief descriptions of the African Oryza species and their sequencing status

*Oryza glaberrima* was domesticated approximately 3500 years ago from its wild ancestor, *O. barthi* in West Africa (Porteres, 1962; Klee *et al.*, 2000, 2004). There is some contention about whether the domestication process was truly independent of Asian rice (Nayar, 2010), but it is evident that *O. glaberrima* experienced a drastic domestication bottleneck and, compared to *O. barthii,* has a very narrow genetic base. This suggests that there is ample opportunity for African breeders to exploit the genetic diversity of its closely related wild ancestor.

The other closely related AA genome species, *O. longistaminata*, is a highly diverse species capable of reproducing both clonally, via rhizomes, and sexually, where it behaves primarily as an outcrossing species. To agriculturalists, *O. longistaminata* is a noxious weed because of the invasive nature of its rhizomatous habit which makes it very difficult to eradicate. The first disease-resistance gene cloned in rice was *Xa21*, which was derived from a backcross introgression line between *O. sativa* and *O. longistaminata* (Ikeda *et al.*, 1990; Song *et al.* 1995) (Table 9.1). RefSeqs for both *O. barthii* and *O. longistaminata* are in progress.

The diploid BB genome species, *O. punctata*, is a serious noxious weed in East Africa (Vaughan, 1994; Brink and Belay, 2006). The IOMAP consortium completed a RefSeq for this genome, which is slightly larger than the AA genome species at 425 megabases (Mb) (Ammiraju *et al.*, 2006). This genome will serve as an important evolutionary outgroup for all the AA genome species comparisons, as it is estimated to have

**Table 9.1.** African rice species as sources of resistance and tolerance to abiotic and biotic stresses.

| Stress | *O. barthii* AA | *O. glaberrima* AA | *O. longistaminata* AA | *O. punctata* BB | *O. eichingeri* CC | *O. brachyantha* FF |
|---|---|---|---|---|---|---|
| Drought | Tolerance | Tolerance | Tolerance | | | |
| High mineral content in soil | | Tolerance | | | | Tolerance |
| Bacterial blight | Resistance | Resistance | Resistance | | | Resistance |
| Bacterial leaf streak | Resistance | | | | | |
| Stem rot | | | | Resistance | Resistance | |
| Tungro virus | | | | | | Resistance |
| RYMV | | Resistance | | | Resistance | |
| Nematodes | | Resistance | Resistance | | | |
| African rice gall midge | | Resistance | | | | |
| Brown planthopper | | | Resistance | Resistance | Resistance | |
| Green leafhopper | | | | | Resistance | |
| Stem borer | | | | | | Resistance |
| Whorl maggot | | | | Resistance | | Resistance |

diverged from the AA genome lineage about 2–3 million years ago. The classification of *O. punctata*'s tetraploid BBCC genome is still in question. Several important traits have been identified in *O. punctata*, including resistance to bacterial leaf blight and brown planthopper (Brink and Belay, 2006), which could potentially be used to improve cultivated rice.

*Oryza eichingeri* is one of three CC genome species and the only one endemic to Africa. Potentially useful traits that have been identified in specific accessions include resistances to rice stem rot (Figoni *et al.*, 1983), *Rice yellow mottle virus* (RYMV) (Brar and Khush, 1997), brown planthopper (Brar and Khush, 2003) and green leafhopper (Brar and Khush, 2003) (Table 9.1). The genome of *O. eichingeri* is about 650 Mb, twice that of *O. glaberrima* (Ammiraju *et al.*, 2006). Development of a full genome sequence for this species is still in the planning stage.

*Oryza brachyantha* (FF genome) is one of the most distantly related species of cultivated rice, having diverged from a common ancestor approximately 10 million years ago. It contains alleles for disease and insect resistance, as well as tolerance to laterite soils (Figoni *et al.*, 1983; Heinrichs *et al.*, 1985; Chaudhary and Khush, 1990; Khush and Brar, 2002; Brar and Khush, 2003) (Table 9.1). It has the smallest genome of any *Oryza* species at 362 Mb (Ammiraju *et al.*, 2006) and is being sequenced under the leadership of Mingsheng Chen (Chinese Academy of Sciences, Beijing).

The rapid accumulation of sequence information on diverse species opens the door to a new age of germplasm conservation, evaluation and utilization (McCouch *et al.*, 2012). It also challenges gene banks and the breeding community to join forces with geneticists, molecular biologists, computational biologists and evolutionists to better understand the relationship between DNA, RNA, phenotypic and environmental variation and to better utilize the wealth of natural variation for plant improvement.

## Breeding Using Diverse Gene Pools of *Oryza*

Despite global awareness about the value of genetic variation and efforts to preserve genetic resources in the world's gene banks, only a small fraction of the naturally occurring wild and cultivated variation in rice has been explored to date. This is changing as genomics and sequencing-based activities begin to provide descriptions of germplasm resources at the molecular level (Sakai *et al.*, 2011). Since the mid-1990s,

marker-based breeding strategies at Africa Rice Center (AfricaRice) have focused primarily on introgressing single genes/QTLs or simply inherited traits from diverse donors into elite recurrent parents (see Ndjiondjop *et al.*, Chapter 12, this volume). Marker-assisted selection (MAS) has been very productive in rice, where many genes and QTLs of large effect are known to confer agronomically useful phenotypes, and where SNPs, SSRs and other molecular markers are available for efficient backcross conversion (Fig. 9.2) (Collard and Mackill, 2008).
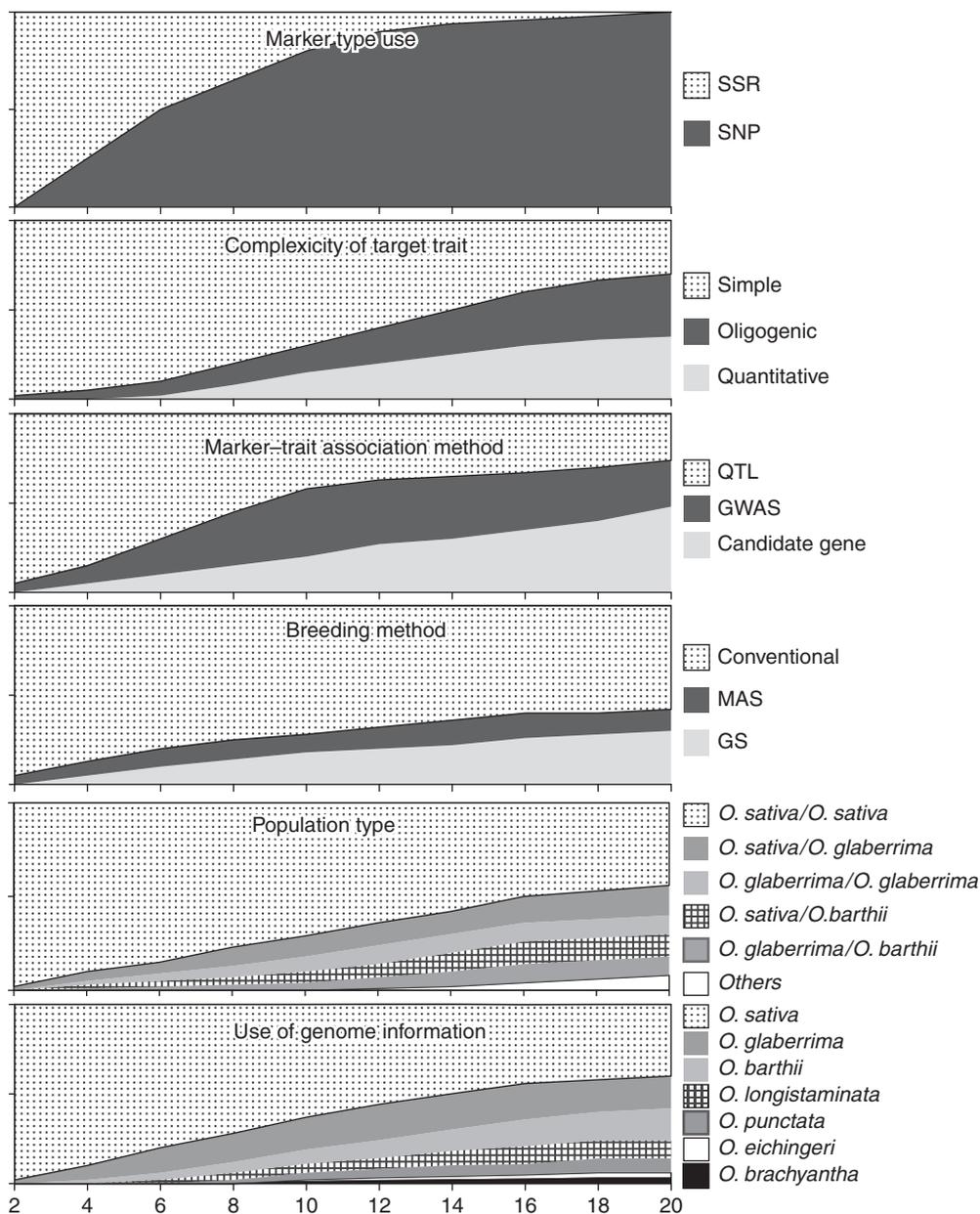


**Fig. 9.2.** Projected relative contribution of different factors to genomics-assisted rice breeding over a 20-year time period (2000–2020) in Africa. GS, genomic selection; GWAS, genome-wide association studies; MAS, marker-assisted selection; QTL, quantitative trait locus; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

Significant progress has been made using marker-assisted backcrossing to move genes and QTLs across species (e.g. *O. sativa* × *O. glaberrima*) or subspecies (*O. sativa* subsp. *indica* × *O. sativa* subsp. *japonica*) boundaries. In these cases, there are usually sterility barriers to contend with, and these must be overcome on a case-by-case basis. To address the sterility barrier, several forms of interspecific (*O. sativa* × *O. glaberrima*) bridging materials have been created (AfricaRice, 2010). These bridge materials make it easier to exploit the range of variation available in *O. glaberrima* and *O. barthii* and to recombine that variation with diverse forms of Asian rice (AfricaRice, 2010).

With the availability of new, high-resolution genomics platforms, there is growing interest in a molecular breeding strategy known as genomic selection. Genomic selection involves the use of genome-wide SNP scans to model phenotypic performance, and makes it possible for a breeder to select on multiple, complex (quantitatively inherited) traits simultaneously, without prior knowledge of the genes or QTLs involved. Genomic selection is most efficient when the objective is to make rapid genetic gain within a well-adapted, well-defined breeding pool, and where selection is being imposed on additive genetic variation (see 'Genomic selection' below).

To take advantage of the diverse gene pools available in rice, African rice breeders are likely to handle multiple streams of germplasm in their breeding programmes. One stream focuses on introgressing novel genetic variation from diverse and often unadapted donors, and the other focuses on recombining favourable alleles within adapted gene pools. The first is a form of pre-breeding and requires the development of inter-specific or inter-subspecific populations, where the objective is to enhance the performance of adapted, elite lines by introgressing selected novel alleles from divergent sources (Tanksley and McCouch, 1997; Xiao *et al.*, 1998; Moncada *et al.*, 2001; Li *et al.*, 2004; Sarla and Mallikarjuna Swamy, 2005; McCouch *et al.*, 2007; Venuprasad *et al.*, 2009, 2011a). The second breeding stream involves only elite, adapted materials that are inter-mated in an effort to improve their overall performance. This may involve the improvement of defensive traits (e.g. resistance to diseases, insects, weeds, abiotic stress), grain quality characteristics or the improvement of yield per se (Lamkey and Lee, 2006).

While the quantitative genetic models underlying these two streams are different (the first involves additive, epistatic and genotype-by-genotype interaction [G×G] effects and the second targets mostly additive genetic variation), in both cases genomics can help drive the efficiency of breeding programmes and enhance genetic gain per cycle of selection. In the case of genomic selection, information about the effects of single genes or specific introgressions that are known to be of value can be readily incorporated into the quantitative models used to predict performance. For complex traits like yield under drought, several potentially useful large-effect QTLs have been detected whose expression varies depending on the genetic background (e.g. Bernier *et al.*, 2007; Venuprasad *et al.*, 2009, 2011b,c). It is particularly important to address the question of genetic background effects in rice because of the number of deeply differentiated subpopulations and species that are used in most breeding programmes. Prediction of intro-gressed allele effects across a range of genetic backgrounds will be necessary to facilitate the efficient utilization of variation from wild and unadapted donor materials. Detailed sequence information can be used to model G×G interactions and to facilitate the identification of allelic series where different donor accessions may contribute subtle trait variation based on finely tuned differences among alleles at the same loci.

## Understanding Genotype–Phenotype Relationships

Understanding the relationship between geno-typic (SNP) variation and phenotypic variation is key to the application of genomics in plant improvement, and it is also an important area of basic study in the biological sciences. Using both 'forward genetics' (phenotype to genotype) and 'reverse genetics' (genotype to phenotype) approaches, researchers are investigating how variation at the DNA level contributes to an organism's phenotype in the context of normal growth and development, response to environment, and ability to withstand biotic or abiotic stress. Our understanding of genotype–phenotype

relationships for rice in Africa will expand and deepen with the accumulation of data and information related to germplasm and growing environments of interest to African rice breeders. This will enable scientists to develop predictive models where genotypic information is used to predict phenotypic performance and to test these predictions in the field. As participants in this process, the new generation of plant breeders in Africa will collectively begin to transform plant breeding from a largely black-box activity into an increasingly predictive and hypothesis-driven science.

Genomics-based studies that examine genotype–phenotype relationships can be used to enhance the productivity and sustainability of rice production in Africa. Here we review four approaches: (i) quantitative trait locus (QTL) mapping (Tanksley, 1993; Mauricio, 2001); (ii) genome-wide association studies (GWAS) (Zhu *et al.*, 2008); (iii) marker-assisted selection (MAS) (Lande and Thompson, 1990; Collard and Mackill, 2008); and (iv) genomic selection (GS) (Meuwissen *et al.*, 2001; Jannink *et al.*, 2010). Some of these approaches aim to identify markers that uniquely tag genes or regions of the genome that condition traits or phenotypes of interest for MAS, while others utilize genome-wide polymorphisms to identify parents with complementary forms of variation for use in crossing, or to predict which offspring are most likely to outperform others in a breeding programme (as in GS).

These methodologies are of interest because the resulting associations or correlations between genotype and phenotype provide the basis for predicting the performance of new lines from DNA evidence. In turn, these predictions enable much more rapid and inexpensive selection of the most valuable lines, accelerating the breeding cycle and increasing selection intensity. All four of the genomics-based studies reviewed below require the same essential ingredients: a relevant population of plants, a set of genotypes for each member of the population, data on trait values or phenotypic performance for each member of the population, and appropriate analysis tools. The essential difference between studies focusing on genetic discovery and those focusing on plant improvement have to do with the finite nature of genetic studies, which are conducted as discrete projects, while genomic selection is integrated into a continuous product-development pipeline that has to be managed by the breeders. None the less, because both types of studies share the same ingredients and require similar attention to the quality of their execution, we begin by reviewing strategies that maximize their efficiency.

## Genotyping and phenotyping strategies

To achieve the highest level of statistical power for identifying QTLs, experimental and environmental variation associated with both the genotypic and the phenotypic assays must be minimized. Genotyping with SNPs is currently the most reliable method for achieving this on the genotypic side, because SNP technology is highly automated and readily differentiates the four nucleotides (ACTG), eliminating a large degree of the experimental error and subjectivity involved in interpreting SSR polymorphisms and restriction fragment length polymorphisms (RFLPs), and greatly facilitating data integration across laboratories. In rice, numerous different types of SNP detection platform are available, including several low-density assays designed for specific types of populations (e.g. 384-SNPs; Thomson *et al.*, 2011), a medium-density fixed array (44,000 SNPs; Zhao *et al.*, 2011), a high-density fixed array (1M SNPs; McCouch *et al.*, 2010) and the genotyping by sequencing approach (Huang *et al.*, 2010; Elshire *et al.*, 2011; Wang *et al.*, 2011). In addition, re-sequencing data for hundreds or thousands of different rice genomes, including several African species (as outlined above), is being generated and will soon be publicly available as the basis for designing virtually any SNP assay of interest (Ammiraju *et al.*, 2010a; McCouch *et al.*, 2010; Xu *et al.*, 2011).

Fixed-array or uniplex SNP assays are well-suited to linkage and association mapping or single-gene introgression projects where the key cost concern is the price per marker data point. However, they remain too expensive for routine use in breeding where the key cost is the genotyping cost per selection unit (the line, family or individual under selection). Breeders are likely to begin to routinely acquire and use genome-wide marker data when the cost of genotyping is

equivalent to the cost of phenotyping a new selection candidate in a single replicated field trial (about US$20–30 in many species and environments). Genotyping based on highly multiplexed next-generation sequencing of restriction-site associated DNA (RAD) (Baird *et al.*, 2008) – also referred to as genotyping by sequencing (GBS) (Elshire *et al.*, 2011) – is rapidly reducing genotyping costs to this level, and will shift the phenotyping:genotyping cost balance on a per-line basis strongly in favour of genotyping. At this cost, genotyping a line or population is a fraction of the cost of acquiring highly precise phenotypic information for low-heritability agronomic traits, which requires replicated field testing over several locations and years.

With the rapid evolution of high-throughput genotyping platforms and technologies, it is difficult for small laboratories and institutions to keep up with the pace of change. Thus, routine genotyping is almost always out-sourced to professional genotyping centres that offer competitive pricing based on economies of scale. Collectively, these centres can ensure ready access to the newest technologies, which are almost always cheaper, faster and technically more straightforward than older ones. Advanced laboratories in the EU, USA and Asia routinely out-source their genotyping to commercial centres for these reasons. In contrast, expertise on the phenotyping side, which is in critically short supply, is much more deeply rooted in local scientific communities that are well positioned to evaluate materials directly in environments of interest. This suggests that the competitive advantage of international agricultural research centres, such as AfricaRice, and national agricultural research systems (NARS) located in rice-growing environments, will be to invest in new phenotyping capabilities and to enhance the efficiency of phenotyping strategies, while out-sourcing most of their genotyping activities. Phenotyping capabilities are in great demand and a serious investment in sophisticated and efficient phenotyping and analytical capacity would drive new forms of collaborative research internationally, bringing renewed attention to the advantages of the international agricultural research centres.

In any phenotyping endeavour, variation due to genetics must be distinguished from variation due to environment. When phenotyping is done in the field, this has traditionally been addressed by replicating experiments over years and locations, as well as by controlling as many environmental variables at each site as possible. When experiments are conducted in a growth chamber or greenhouse, environmental variation may be further reduced and this may enhance a researcher's ability to identify meaningful genotype–phenotype associations. For example, traits such as disease or insect resistance are amenable to evaluation under controlled conditions because they often require that specific strains of a pathogen or pest be used for inoculation and that plants be protected from other stresses that may interfere with the evaluation of the disease response. Similarly, for some abiotic stresses, a specific amount, timing or type of stress may be critical to the evaluation, making it preferable to evaluate plants under controlled conditions. While controlled conditions can greatly accelerate the identification of genes and QTLs underlying complex phenotypes, these associations must ultimately be tested in the field to determine their relevance and reliability for breeding applications. For composite traits, such as yield, drought resistance and flowering time, where the genotype × environment interaction component is known to be high, and the complexity of environmental signals over the course of the growing season cannot be modelled in a greenhouse or growth chamber, phenotypic evaluation is best performed directly in the field. New tools and screening methodologies are being used to enhance the precision and efficiency of phenotyping, in concert with the use of genomics to enhance the power and efficiency of genetic characterization. For example, geographic information systems (GIS) are widely used to help identify and characterize target populations of environments (TPEs) selected to be representative of the larger spectrum of production environments, and remote-sensing technologies make it feasible to evaluate the performance of large populations of plants under diverse field conditions. Where possible, simultaneous phenotypic evaluation of a QTL-mapping population or GWAS diversity panel under both controlled and field conditions provides valuable data for determining whether the same or different SNPs are significantly associated with a phenotype in different environments. Ultimately, advances in phenotyping, genotyping and environmental

characterization together make it possible for breeders to begin to tailor varieties to specific environments.

It is critical that when phenotyping a mapping population or association mapping panel, a high level of broad-sense heritability (H) or repeatability be achieved to ensure reliable estimation of effects (this H should be estimated on a marker or haplotype basis, as discussed below). This is especially important when alleles are at low frequency in the population, with the minor allele occurring in only a few lines. Broad-sense heritability is rarely reported for GWAS phenotypic information, but is the key parameter to monitor in any genetic analysis of quantitative traits that are affected by the environment. In association or QTL-mapping experiments, investments aimed at increasing phenotyping precision, in order to achieve a high level of H, provide high returns because the effects estimated in these experiments are the basis for subsequent marker-assisted breeding and introgression.

Existing phenotypic data can be leveraged at virtually any phase of a breeding programme to begin to identify markers or QTLs that are associated with a major effect on plant performance. Selective genotyping, including bulked segregant analysis (BSA) (Lander and Botstein, 1989; Darvasi and Soller, 1992; Navabi *et al.*, 2009; Sun *et al.*, 2010) may be performed on a population that has been evaluated for phenotypic performance in the field, greenhouse or growth chamber to identify marker alleles that exist at different frequencies in individuals with contrasting phenotypes. The breeder harvests leaf tissue (and ideally seed) from one or both phenotypic extremes of the population, extracts DNA and performs a genotyping assay using one of the many SNP assays available to the rice community (the density of SNPs required at this point depends on the population structure of the lines being evaluated) and compares SNP allele frequencies among the two phenotypic groups. This approach was successfully used by Bernier *et al.* (2007) and Venuprasad *et al.* (2009) to identify large-effect QTLs for grain yield under drought stress in rice, and by Nandi *et al.* (1997) to identify a submergence-tolerance QTL in lowland rice.

Finally, we note that to maximize the benefit of applying genomics to plant breeding, researchers need to evaluate the effects of **alleles** more so than those of lines. Traditional plant breeding research has developed effective tools for the identification and crossing of favourable lines. Genomics research, on the other hand, is providing tools to increase the efficiency of manipulating alleles, by identifying desirable alleles and pyramiding them or by using their estimated effects to predict the value of new lines. For the purpose of allele evaluation, the replication of **lines** is much less, or not at all, necessary: the alleles themselves are replicated across lines regardless of whether the lines themselves are replicated. Experiments with unreplicated lines can therefore have the greatest QTL detection power (Knapp and Bridges, 1990) and provide the best predictive ability (Zhong *et al.*, 2009). The same holds true for GWAS studies. Daetwyler *et al.* (2007) and Hayes *et al.* (2009) have described the relationship between trait heritability on an evaluation-unit basis (lines, families, etc.), the number of loci affecting the trait of interest, and the size of the GWAS population. Their analyses indicate that the precision of haplotype effect estimation increases with the size of the population, and that optimal designs consider the marker or haplotype as the unit of evaluation, rather than the line.

Earlier mapping experiments emphasized achieving a high repeatability of effect estimates for individual lines because genotyping was much more expensive than phenotyping. Now that the balance has shifted, optimal designs – i.e. those that maximize the repeatability of haplotype rather than line effects – are those that test as many lines as possible with little or no replication of individual lines. Designs maximizing the number of unreplicated lines used in the genetic analysis of populations, although theoretically superior, were impractical until genotyping costs dropped to the point where genotyping became inexpensive relative to phenotyping. We have now reached this point, which has far-reaching consequences for the implementation of field trials for phenotyping complex traits in the context of genomics-based crop improvement.

## QTL mapping

A QTL is a region of the genome defined by molecular markers that is predicted to contain a gene or genes associated with a specific trait

(Mauricio, 2001; Bernardo, 2008). QTL mapping involves the analysis of a population (or populations) derived from a cross between two parents, where individual plants, lines or families within the population have been characterized for a set of well-distributed molecular-marker polymorphisms (RFLP, SSR, SNPs, etc.), as well as for one or more quantitative traits. A QTL is declared when there is a statistical association between the segregation of molecular polymorphism and a measurable phenotype, using the individual segregants within the population as replicates (Tanksley, 1993). The phenotype of interest may be a feature of the whole plant, an organ or a tissue, or it may be characterized as a feature associated with the DNA, RNA or protein. It may be evaluated under field conditions or experiments may be conducted under controlled conditions in a growth chamber or greenhouse. The objective of QTL analysis is to identify the position and relative importance of genetic factors that collectively determine a trait or phenotype of interest, and to identify sources of favourable and unfavourable alleles at each QTL (Bernardo, 2008).

QTL mapping is relevant to the agricultural community because it provides a way of genetically dissecting quantitative variation found in naturally occurring germplasm resources and offers insight into the linkage and epistatic relationships among genes and QTLs controlling diverse traits of interest. The QTL database in Gramene offers one of the largest repositories of QTL information for rice in the world (Ni *et al.*, 2009). Plant breeders are able to make direct use of QTL results for MAS as long as the genetic materials in their breeding programme are identical or closely related to the parents used to detect the QTLs. Molecular geneticists use QTLs as a first step in map-based cloning to identify genes underlying the QTL and to examine their molecular function and role in a biochemical or regulatory pathway. QTL information is valuable in both cases because it narrows the genomic search space for identifying genes underlying complex phenotypes, and it provides global information about the genetic architecture of those phenotypes, including the location and relative importance of each locus contributing to the quantitative trait.

It should be noted that QTL mapping works well when the genetic architecture of a trait is such that one or a few loci account for a large proportion of the genetic variance. All QTL effects are estimated with error, but when effects are very large, the 'signal-to-noise' ratio is large and effects are estimated with reasonable accuracy. When this is not the case, i.e. when the trait is actually controlled by many genes with small effects that are similar in size, QTL mapping experiments can provide misleading results due to the selection bias that results from the process of ranking QTLs by their effect sizes, and then selecting as significant only those that exceed a given threshold. If the true sizes of effects of genes contributing to a polygenic trait are small and similar, as is likely to be the case for many complex traits, then much of the difference in estimates will result from experimental 'noise' that will not recur in subsequent experiments. The situation is analogous to a cultivar trial in which differences are small, and error variance is large relative to genotypic differences; in such trials, when cultivars are ranked on the basis of their mean yields, and then the high tail of the distribution is selected, the effect sizes of the selected genotypes always shrink back towards the mean in subsequent trials, in rough proportion to their repeatability in the estimation experiment. This type of selection bias in QTL mapping leads to a more or less random subset of QTLs being declared significant in any particular experiment (Lande and Thompson, 1990; Beavis, 1994; Bernardo, 2008; Heffner *et al.*, 2009). MAS schemes based on such QTL effects are likely to give disappointing results. When the genetic control of traits is truly polygenic, approaches based on genome-wide association of phenotypes with haplotypes at many loci are likely to be superior (Heffner *et al.*, 2009).

### Genome-wide association studies (GWAS)

GWAS, like QTL mapping, are based on the analysis of phenotypes and genotypes, but in GWAS, the genotyping and phenotyping are performed on a diverse collection of unrelated strains (referred to as a diversity panel) rather than on the progeny of a bi-parental cross (referred to as a QTL-mapping population) (Yu *et al.*, 2005; Zhu *et al.*, 2008). Genetic relationships among individuals that comprise association mapping panels vary widely. The use of very diverse panels may make it difficult to phenotype certain traits

in the field, if some genotypes are so poorly adapted that trait expression is affected; in practice, association mapping panels should be composed of the most diverse set of adapted lines possible. GWAS analysis, like QTL mapping, relies on the strength of linkage disequilibrium (LD) between markers and phenotypes in the materials under study, and both aim to identify genes or regions of the genome that underlie complex phenotypes. While both approaches depend on identifying association between markers (i.e. SNPs) and traits of interest, GWAS does so in the context of evolutionary biology and population genetics, while QTL mapping does so in the context of inheritance genetics (Bernardo, 2008). GWAS generally provides greater resolution for the same population size, due to the fact that LD generally decays more quickly in a set of unrelated lines compared to bi-parental mapping populations (Flint-Garcia *et al.*, 2003; Famoso *et al.*, 2011). This is because there have been more generations of effective recombination separating lines in a diversity panel from their last common ancestor, than among segregants derived from a recent bi-parental cross. Despite these differences in resolution, the two approaches are complementary and are often pursued jointly (Yu *et al.*, 2008; Legarra and Fernando, 2009; McMullen *et al.*, 2009; Famoso *et al.*, 2011). They can be used to dissect complex traits in virtually any segregating population of interest as long as genotypes and phenotypes can be reliably assayed on the same individuals or families.

### Marker-assisted selection (MAS)

MAS represents an early stage in the development of genomic-assisted breeding strategies (Lande and Thompson, 1990) and can be used as soon as marker-alleles are identified that can reliably predict a trait or a phenotype of interest in populations that are relevant to the breeder. The reliability of the association between a marker-allele and a phenotype of interest is key because MAS substitutes selection on phenotype with selection on specific marker-alleles. Most MAS programmes use small numbers of markers to select for genes or QTLs of relatively large effect and use two schemes: (i) backcrossing to introduce favourable alleles from a donor variety into one or more elite genetic backgrounds; and

(ii) recurrent selection to cumulate alleles from several loci in a breeding population, using markers to enrich favourable alleles in each generation. This latter practice is called marker-assisted recurrent selection (MARS; Charmet *et al.*, 1999). A requirement of using markers identified in prior QTL studies is that these markers can be traced as identical by descent (IBD) to the parents carrying the favourable allele in the QTL analysis. Thus, there must be a clear line of descent from the QTL study to the breeding material, with loci being followed by marker genotyping throughout. In addition, when breeding populations are generated by crossing highly divergent materials, as in the case of *O. sativa* × *O. glaberrima* or *O. sativa* × *O. longistaminata* populations, breeders must ensure that the marker–trait relationship is not disrupted by dramatic changes in the genetic background as the favourable alleles are transferred to an elite agronomic background.

As reviewed by Collard and Mackill (2008), MAS has several advantages over conventional phenotypic selection in a breeding programme, as long as the marker genotypes can be inexpensively generated within the context of the normal breeding cycle and the information can be communicated to the breeder in a timely and cost-effective way. The practice saves the breeder's time, resources and effort because it allows selection to be carried out at the seedling stage (even for traits expressed late in the life of the plant), it permits the selection of heterozygous individuals without the need for progeny testing, it facilitates gene pyramiding for durable disease and insect resistance, and it greatly enhances the efficiency of backcross conversion by facilitating both foreground and background selection, thereby helping to break linkage between favourable and unfavourable alleles (known as linkage drag).

As the cost of genotyping continues to fall and the number of markers available for rice continues to rise, large populations of lines can now be genotyped at much lower cost than it would take to evaluate them in multi-location trials for key phenotypes associated with performance (Jannink, 2005; Heffner *et al.*, 2010). In rice, MAS has been used successfully to rapidly introgress genes and QTLs into elite breeding lines to enhance biotic and abiotic stress tolerance, grain quality and yield (Xiao *et al.*, 1998;

Ashikari and Matsuoka, 2006; McCouch *et al.*, 2007; Neeraja *et al.*, 2007; Septiningsih *et al.*, 2008; Shanti *et al.*, 2010; Lorieux *et al.*, Chapter 10, this volume; Ndjiondjop *et al.*, Chapter 12, this volume). MAS strategies require high confidence in marker–QTL allele association and therefore do not reliably handle quantitative traits with complex genetic architecture involving hundreds of genes or QTLs, where this confidence cannot be obtained. Even for traits with such architectures, however, a few large-effect loci may segregate that are amenable to these strategies.

### Genomic selection (GS)

Genomic selection (GS) extends the use of markers to breeding highly polygenic traits when there can be little confidence in specific marker–QTL associations, and is complementary to conventional MAS. Genomic selection couples the relevant phenotyping of large plant breeding populations with high-density marker technologies to deal with quantitative traits such as yield, drought tolerance or flowering time (Meuwissen *et al.*, 2001; Buckler *et al.*, 2009; Jannink *et al.*, 2010). This coupling requires effective database solutions to handle the large amounts of phenotypic and genotypic (marker) data, as well as new statistical methods. The requisite genotyping capacity is already available in rice, as inexpensive genome-wide marker-detection platforms permit parallel scoring of a few hundred up to hundreds of thousands of polymorphisms across the rice genome (Huang *et al.*, 2010; McCouch *et al.*, 2010; Wang *et al.*, 2011). At the same time, algorithms have been developed that can simultaneously estimate the effects of all markers on phenotype, providing unbiased predictions of the performance of newly genotyped lines and capturing the influence of even the many small effects determining quantitative traits (Meuwissen *et al.*, 2001; Jannink *et al.*, 2010).

To generate accurate predictions, genomic selection proceeds as follows. A training population must be constituted and characterized both genotypically and phenotypically, generating a joint genotype–phenotype data set. This population should be closely related to the breeding lines on which genomic selection will be practised. For a breeding programme, a logical choice for the training population are experimental lines that have previously passed through the programme and are therefore already well characterized in the environments that the programme targets (Heffner *et al.*, 2009; Jannink *et al.*, 2010). In practice, this is likely to be the early-generation testing component of the breeding programme, with the model updated every time a test of new breeding materials is conducted. This joint genotype–phenotype data set is analysed using one of the statistical methods that have been proposed to train a prediction model (a comprehensive review is given by Lorenz *et al.*, 2011).

Models can be optimally designed to predict either the performance of the lines themselves or that of their progeny. In the latter case, the prediction is called the genomic estimated breeding value (GEBV) and it is the optimal predictor for selecting parents to initiate a new round of crosses and breeding. Prediction models can be developed for any trait for which there is sufficient phenotypic data, and these multivariate predictions can then be combined into a selection index, much like phenotypic measurements. A great value of GS methods is that, through evaluation of the effects of alleles that occur in many lines across a breeding programme, they leverage investment made in all phenotyping efforts for the purpose of improving the prediction for any given line. This leveraging is an important advancement over conventional phenotypic selection where a phenotypic measurement really provides information only on the line itself.

In conventional breeding systems, selection is based almost exclusively on the phenotype of a candidate. In GS schemes, selection (at least in the preliminary stages) is based on the candidate's genotype, which is quickly and inexpensively determined, rather than on its phenotype, which is very expensive to measure reliably, particularly for low-heritability traits. This 'decoupling' of selection from phenotyping can potentially increase breeding progress in three main ways:

**1.** It permits estimates of haplotype effects from previous testing to be used in predicting line performance, potentially increasing accuracy.
**2.** It allows selection intensity to be increased by permitting selection of lines that have been inexpensively genotyped, but not expensively phenotyped.

**3.** It permits recurrent selection to be conducted without intervening cycles of line extraction and phenotyping, allowing a radical shortening of the breeding cycle.

Developing the capacity for GS represents a significant investment of time, labour and organizational resources for a breeding programme (Eathington *et al.*, 2007), but it is an essential step forward in an era where genotyping hundreds or thousands of accessions can be accomplished more quickly and economically than evaluating those same materials for phenotypic performance in the field. We are witnessing a dramatic shift in how we value and invest in phenotypic evaluation. Previously, when molecular evaluation was slower and more expensive than phenotypic evaluation, genotyping of a line made it precious and, in order to leverage value from the genotype, candidate lines were extensively evaluated over replications, locations and years. The reverse is now true: the phenotyping of a line makes it precious, and inexpensive genotyping serves to extract value from that investment, whether the measurements have been highly replicated or not.

These changes require non-trivial improvements in bioinformatic and logistical support to jointly manage and analyse, on time, thousands of marker data points in hundreds or thousands of individuals every season. If appropriate coordination can be achieved between the genotyping and phenotyping activities, nursery management and decision support for selection, GS can serve as a valuable proxy for field-based evaluation, increasing genetic gains per unit time (Heffner *et al.*, 2010). Taking advantage of historical data available to a breeding programme, the accuracy of GS predictions can be validated relatively rapidly. Fully realizing the benefits of GS in breeding schemes across a programme is far easier said than done, however, and can be expected to take years. African rice programmes would be well advised to begin to integrate the use of genome-wide SNP panels to document the genotypic variation that is being utilized by breeding programmes throughout the continent. The information will provide the basis for all future efforts to assess genotype–phenotype associations and to assign GEBVs for materials of interest to breeders and farmers in Africa.

## Genomics and adaptive breeding

Just as genomics tools are facilitating the mobilization of genetic resources from wild and cultivated relatives for use in breeding, they will also open the door to the utilization of elite germplasm resources coming from outside of Africa. Many elite lines have hitherto been inaccessible because of lack of adaptation (disease and insect resistance, tolerance to problem soils, etc.) or grain quality. In breeding terms, a mega-environment may be defined as a set of environments in which lines perform similarly. In rice, mega-environments are defined by crop management (e.g. transplanting versus direct seeding; irrigated versus rainfed), hydrology (upland versus lowland), soils and climate (e.g. tropical, subtropical or temperate), among other delineations. Particular combinations of hydrology, crop management and climate may recur across vast regions, and even on different continents; there are many rice-growing areas in Africa where conditions are similar to those in parts of South or South-East Asia, or South America.

In the past, movement of elite germplasm between mega-environments has been hampered by the occurrence of region-specific diseases or local grain-quality preferences. However, with over a decade of investment in identifying markers closely linked to or mapping within major genes controlling diseases important in Africa, such as blast, bacterial leaf blight and *Rice yellow mottle virus*, as well as grain quality traits such as grain shape and aroma, the use of MAS allows breeders to rapidly introgress specific traits of interest. This helps to 'adapt' elite but exotic lines coming from the same mega-environment(s) but different geographic regions, to the local biotic or socio-economic environment. GS approaches would also facilitate the movement of new, potentially elite germplasm among regions within the same mega-environment. If mega-environments are well characterized, breeders will be able to confidently identify multiple locations at which the agronomic performance of breeding lines is expected to be highly correlated with their own, with obvious implications

for identifying new sources of useful elite materials. Investment in breeding-site characterization could connect distant but environmentally similar locations into breeding networks that could exchange promising new lines on the basis of GEBVs.

## Summary of Genomic-assisted Breeding Potential

The key conceptual difference between conventional breeding and either MAS or GS approaches is that in the former, the line or family is the unit of evaluation, whereas in the latter, marker or haplotype alleles are evaluated (Meuwissen *et al.*, 2001). Because marker alleles recur across lines, additive effects may be estimated over many environments regardless of the identity of lines that contain them. Effects estimated previously in relevant training sets can be used to predict the performance of new lines containing those same alleles, despite the fact that the new lines may not have been phenotyped (Eathington *et al.*, 2007; Heffner *et al.*, 2009; Lorenzana and Bernardo, 2009). This concept is particularly important in the application of GS to breeding for stress-prone environments. It means that not every line needs to be evaluated in the target environment (e.g. drought-affected or N-deficient) to predict its performance in that environment. Rather, through allele effect estimates, lines evaluated in specific mega-environments 'share information' on a target environment with all members of a cohort.

This information sharing has exciting ramifications beyond breeding for improved adaptation to optimal and stress environments. It means that different breeding programmes may contribute to each other's success at picking broad-adaptation winners through data pooling – in effect, allowing a number of small but cooperating breeding programmes to leverage each other's efforts to increase their breeding efficiency.

Both scientific research and institutional innovations are needed to take advantage of these ideas and bring them to fruition. On the scientific side, a programme needs to be established to examine the relationships between allelic effects for environments targeted by different

breeding programmes. When considering possible sources of difference between allelic effects important to national rice improvement programmes in Africa, Asia and America, two sources must be distinguished. First, the effect of the allele may be modulated by specifics of the environments targeted by the national programmes – for example, the allele may be beneficial under one rainfall pattern but not another. Second, less obviously, the effect of the allele may be modulated by genetic background: average allele frequencies across loci might differ sufficiently among national programmes to generate gene-interaction effects. Divergence in the average genetic background of African national programmes is an empirical issue of utmost importance that is, at present, poorly quantified. It is easily addressed by developing genome-wide SNP profiles for representative breeding lines and populations from each of the major rice-breeding programmes in Africa and assembling this information in a central diversity database. A central research body such as AfricaRice would be well positioned to address this question and provide a firm empirical foundation for future work seeking to enhance mutual benefit among African national programmes through cooperative genotype–phenotype analyses.

The mutual benefits that could be derived from cooperation through genomics will require central coordination and coordinated research. A far-reaching and transformative use of the potential to predict performance is referred to as 'open source' collaborative breeding. In this model, a central research body like AfricaRice collects and analyses genotype–phenotype information from a number of national programme partners or small seed and breeding companies in Africa, Asia and America. In exchange for the information delivered by the breeding programmes, AfricaRice would deliver analyses, providing predictions for new lines that combine both genome-wide estimates of value for highly polygenic traits and genotypes for large-effect QTL affecting oligogenic traits. These predictions contribute decision support for the selection of parents and specific crosses. AfricaRice's own breeding efforts would focus on validating those predictions, as well as developing pre-breeding lines and improving source populations to address deficiencies identified broadly across its 'client' breeding programmes. Deliverables would include

genotyped lines tailored to specific environments and markets, as well as predictions of their performance. The sustainability and scaling up of this approach would be driven by continued decreases in genotyping cost per data point and per DNA sample, and by the needs for constant updating and validating of training sets, developing breeding plans and breeder-friendly software for implementation of GS, and trained personnel in applied breeding programmes to manage rapid-cycle genotype-based selection.

## Institutional Implications for International Research Centres

The insight that alleles, rather than the lines themselves, need to be evaluated across environments opens opportunities for research programmes in Africa to benefit fully from the global drop in genotyping costs by increasing selection intensity. This can be achieved by investing early in the genotyping of large numbers of African rice accessions and breeding lines, using the genotyping information as the basis for selecting informative subsets of lines for phenotyping in target environments, and achieving a high level of precision in estimating GEBVs. It is important that the training population(s) be representative of the breeding population(s) (selection candidates) and that the training population(s) be adequately evaluated in the target environments.

International research centres, like AfricaRice, may position themselves in a key niche between upstream technology developers and downstream producers of finished varieties, serving an essential role as generators of pre-breeding materials, training populations and elite candidate lines with associated GEBVs, as well as serving a vital role in educating the next generation of African plant breeders. Generating reliable GEBVs will require ongoing collaboration with both public- and private-sector programmes on the ground in Africa to develop and evaluate appropriate training sets in target environments. AfricaRice's ability to support these activities (not to take exclusive responsibility for them) will be synergistic in the African context, enabling the system as a whole to deliver improved rice varieties to meet the diverse needs of African farmers.

It is important that researchers at AfricaRice stay abreast of the latest developments in genotyping technology, statistical analysis and information management through collaborations with universities and research institutions or companies worldwide. Collaboration is under way to re-sequence the contents of the AfricaRice gene bank, including both *O. glaberrima* and *O. sativa,* as well as related wild *Oryza* species, and to develop a public database of information about genetic variation in *Oryza* species in Africa. A similar investment is needed to genotype breeding populations that provide the foundation for selection by each of the national programmes as well as the international centre(s) on the African continent. This information can be used to systematically explore these materials and develop novel pre-breeding populations, develop SNP assays tailored to the needs of the African rice community, and to expand the rice gene pool available to breeders and researchers worldwide. This is likely to expand the rational use of *O. glaberrima, O. barthii* and other African rice germplasm, as well as to accelerate the adaptation of stress-tolerant *O. sativa* lines adapted to African conditions. Links with breeding programmes and seed companies will be essential to ensure that the generation of training sets and GEBV information is relevant to local needs; and iterative cycles of genotyping and phenotyping, along with consistent exchange of information and feedback, will be essential to fine-tune the process.

Finally, making rice genomics work for Africa will inevitably involve the evolution of a new institutional culture. Incorporating high-throughput genomics information into a breeding programme requires new management structures and reporting protocols. Vast amounts of genomic information are generated and processed with every cycle of selection, and this requires rapid and transparent communication and exchange of information among all members of a team. Successful integration of genomics can improve breeding efficiency by accelerating the selection of parents, reducing breeding cycle times, enhancing genetic gain per cycle of selection (due to improvements in the accuracy with which offspring with high breeding value are selected), and helping to match varieties/alleles to specific environments. However, it also imposes new requirements on the people involved, including

formulation of explicit goals, objectives and time frames for each activity, transparency of work plans and transparent monitoring of both progress and impediments, open access to information at all levels of transaction, formalized reporting and communication pipelines, and streamlined teamwork among collaborating scientists. In many ways, these changes mimic the organization of many private-sector breeding organizations. For this reason, internships for young African plant breeders in the private sector may be an efficient way of helping to prepare them for a productive career in public service where they are expected to be able to integrate genomics into breeding programmes.

As discussed by Reece and Haribabu (2007), the sociological changes required to enable multi-disciplinary teams to work together with a shared vision are often the most difficult to achieve and require visionary and highly skilled research managers to guide the process. At every level, innovation will be required and adjustments will have to be made to match expectations with local realities, but the process of making genomics work for Africa has already begun, and a great deal will be learned in the years ahead as we explore the opportunities that unfold before us.

## Acknowledgements

## Note

[1] QTL analysis can detect many significant loci (QTLs), but attention is focused only on those QTLs with highly significant LOD scores (logarithm [base 10] of odds). This is to avoid type 1 error when using small (biased) sample/population sizes. However, it leads to an inflated estimate of the percentage variance explained ($R^2$ value) in the second step of the analysis where a model is built based only on the highly significant QTLs, ignoring the many QTLs of small effect. Thus, the model overestimates the effect ($R^2$ value) of the loci included in the model, and fails to acknowledge the 'minor' effect loci that collectively also contribute to the phenotype. For this reason, QTLs that are detected using small mapping populations (such as are normally used in QTL-mapping studies) appear to explain a greater portion of the phenotypic variance than they really do.

## References

AfricaRice (2010) *New Breeding Directions at AfricaRice: Beyond NERICA*. Africa Rice Center, Cotonou, Benin.

Ammiraju, J.S., Luo, M., Goicoechea, J.L., Wang, W., Kudrna, D., Mueller, C., Talag, J., Kim, H.R., Sisneros, N.B., Blackmon, B., Fang, E., Tomkins, J.B., Brar, D., MacKill, D., McCouch, S., Kurate, N., Lambert, G., Galbraith, D.W., Arumuganathan, K., Rao, K., Walling, J.G., Gill, N., Yu, Y., SanMiguel, P., Soderlund, C., Jackson, S. and Wing, R.A. (2006) The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza. Genome Research* 16, 140–147.

Ammiraju, J.S., Lu, F., Sanyal, A., Yu, Y., Song, X., Jiang, N., Pontaroli, A.C., Rambo, T., Currie, J., Collura, K., Talag, J., Fan, C., Goicoechea, J.L., Zuccolo, A., Chen, J., Bennetzen, J.L., Chen, M., Jackson, S. and Wing, R.A. (2008) Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *The Plant Cell* 20, 3191–3209.

Ammiraju, J.S., Song, X., Luo, M., Sisneros, N., Angelova, A., Kudrna, D., Kim, H.R., Yu, Y., Goicoechea, L.J., Lorieux, M., Kurata, N., Brar, D., Ware, D., Jackson, S. and Wing, R.A. (2010a) The OryzaBAC

resource: a genus-wide and genome scale tool for exploring rice genome evolution and leveraging useful genetic diversity from wild relatives. *Breeding Science* 60, 536–545.

Ammiraju, J.S., Fan, C., Yu, Y., Song, X., Cranston, K., Pontaroli, A.C., Lu, F., Sanyal, A., Jiang, N., Rambo, T., Currie, J., Collura, K., Talag, J., Bennetzen, J.L., Chen, M., Jackson, S. and Wing, R. (2010b) Spatio-temporal patterns of genome evolution in allotetraploid species of the genus *Oryza. The Plant Journal* 63(3), 430–442.

Ashikari, M. and Matsuoka, M. (2006) Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends in Plant Science* 11, 344–350.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376.

Beavis, W.D. (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Wilkinson D.B. (ed.) *Proceedings of the 49th Annual Corn and Sorghum Research Conference*. American Seed Trade Association, Washington, DC, pp. 250–265.

Bernardo, R. (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science* 48, 1649–1664.

Bernier, J., Kumar, A., Venuprasad, R., Spaner, D. and Atlin, G. (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Science* 47, 507–516.

Brar, D.S. and Khush, G.S. (1997) Alien introgression in rice. *Plant Molecular Biology* 35, 35–47.

Brar, D.S. and Khush, G.S. (2003) Utilization of wild species of genus *Oryza* in rice improvement. In: Nanda, J.S. and Sharma, S.D. (eds) Monograph on genus *Oryza*. Science Publishers, Enfield, New Hampshire and Plymouth, UK, pp. 283–309.

Brink, M. and Belay, G. (Eds) (2006) *Plant Resources of Tropical Africa 1. Cereals and Pulses*. PROTA Foundation, Wageningen, Netherlands; Backhuys Publishers, Leiden, Netherlands; and Technical Centre for Agricultural and Rural Cooperation, Wageningen, Netherlands.

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J.C., Ma. Goodman, M., Harjes, C., Guill, K., Kroon, D.E., Larsson, S., Lepak, N.K., Li, H., Mitchell, S.E., Pressoir, G., Peiffer, J.A., Oropeza Rosas, M., Rocheford, T.R., Cinta Romay, M., Romero, S., Slavo, S., Sanchez Villeda, H., Sofia da Silva, H., Sun, Q., Tian, F., Upadyayula, N., Ware, D., Yates, H., Yu, J., Zhang, Z., Kresovich, S. and McMullen, M.D. (2009) The genetics architecture of maize flowering time. *Science* 325(5941), 714–718.

Charmet, G., Robert, N., Perretant, M.R., Gay, G., Sourdille, P., Groos, C., Bernard, S. and Bernard, M. (1999) Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. *Theoretical and Applied Genetics* 99, 1143–1148.

Chaudhary, R.C. and Khush, G.S. (1990) Breeding rice varieties for resistance against *Chilo* spp. of stem borers in Asia and Africa. *Insect Science and Its Application* 11, 659–669.

Collard, B.C.Y. and Mackill, D.J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B* 363, 557–572.

Craig, D.W., Pearson, J.V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J.J., Pawlowski, T.L., Laub, T., Nunn, G., Stephan, D.A., Homer, N. and Huentelman, M.J. (2008) Identification of genetic variants using bar-coded multiplexed sequencing. *Nature Methods* 5, 887–893.

Daetwyler, H.D., Villanueva, B., Bijma, P. and Woolliams, J.A. (2007) Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics* 124, 369–376.

Darvasi, A. and Soller, M. (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics* 85, 353–359.

Duran, C., Appleby, N., Clark, T., Wood, D., Imelfort, M., Batley, J. and Edwards, D. (2009) AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. *Nucleic Acids Research* 37, D951–D953.

Eathington, S.R., Crosbie, T.M., Edwards, M.D., Reiter, R.S. and Bull, J.K. (2007) Molecular markers in a commercial breeding program. *Crop Science* 47, S154–S163.

Edwards, D. and Batley, J. (2010) Plant genome sequencing: applications for crop improvement. *Plant Biotechnology Journal* 8, 2–9.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.D. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5), e19379.

Famoso, A.N., Zhao, K., Clark, R.T., Tung, C.-W., Wright, M.H., Bustamante, C., Kochian, L.V. and McCouch, S.R. (2011) Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through

genome-wide association analysis and QTL mapping. *PLoS Genetics* 7(8): e1002221. DOI: 10.1371/journal.pgen.1002221 (accessed 24 February 2012).

Figoni, R.A., Rutger, J.N. and Webster, R.K. (1983) Evaluation of wild *Oryza* species for stem rot (*Sclerotium oryzae*) resistance. *Plant Disease* 67, 998–1000.

Flint-Garcia, S.A., Thornsberry, J.M. and Buckler, E.S. IV (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54, 357–374.

Ge, S., Sang, T., Lu, B.R. and Hong, D.Y. (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proceedings of the National Academy of Sciences USA* 96(25), 14400–14405.

Goicoechea, J.L., Ammiraju, J.S.S., Marri, P.R., Chen, M.S., Jackson, S., Yu, Y., Rounsley, S. and Wing, R.A. (2010) The future of rice genomics: sequencing the collective *Oryza* genome. *Rice* 3(2–3), 89–97.

Gupta, P.K., Roy, J.K. and Prasad, M. (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science* 80, 524–535.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. and Goddard, M.E. (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92, 433–443.

Heffner, E.L., Sorrells, M.E. and Jannink, J.L. (2009) Genomic selection for crop improvement. *Crop Science* 49, 1–12.

Heffner, E.L., Lorenz, A.J., Jannink, J.L. and Sorrells, J.E. (2010) Plant breeding with genomic selection: gain per unit time and cost. *Crop Science* 50, 1681.

Heinrichs, E.A., Viajante, V.D. and Romena, A.M. (1985) Resistance of wild rices, *Oryza* spp., to the whorl maggot *Hydrellia philippina ferino* (Diptera, Ephydridae). *Environmental Entomology* 14, 404–407.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T. and Zhang, Z. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* 42, 961–967.

Hurwitz, B.L., Kudrna, D., Yu, Y., Sebastian, A., Zuccolo, A., Jackson, S.A., Ware, D., Wing, R.A. and Stein, L. (2010) Rice structural variation: a comparative analysis of structural variation between rice and three of its closest relatives in the genus *Oryza. Plant Journal* 63, 990–1003.

Ikeda, R., Khush, G.S. and Tabien, R.E. (1990) A new resistance gene to bacterial blight derived from *O. longistaminata. Japanese Journal of Breeding* 40 (Suppl. 1), 280–281.

IRGSP (International Rice Genome Sequencing Project) (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800.

Jannink, J.L. (2005) Selective phenotyping to accurately map quantitative trait loci. *Crop Science* 45, 901–908.

Jannink, J.L., Lorenz, A.J. and Iwata, H. (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* 9, 166–177.

Khush, G.S. and Brar, D.S. (2002) Biotechnology for rice breeding: progress and potential impact. Paper presented at the 20th Session of the International Rice Commission, Bangkok, Thailand, 23–26 July.

Klee, M., Zach, B. and Neumann, K. (2000) Four thousand years of plant exploitation in the Chad Basin of northeast Nigeria I: the archaeobotany of Kursakata. *Vegetation History and Archaeobotany* 9(4), 223–237.

Klee, M., Zach, B. and Stika, H.-P. (2004) Four thousand years of plant exploitation in the Lake Chad Basin (Nigeria), part III: plant impressions in potsherds from the final stone age Gajiganna culture. *Vegetation History and Archaeobotany* 13(2), 131–142.

Knapp, S.J. and Bridges, W.C. (1990) Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* 126, 769–777.

Kwok, P.Y., Deng, Q., Zakeri, H., Taylor, S.L. and Nickerson, D.A. (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* 31, 123–126.

Lamkey, K.R. and Lee, M. (Eds) (2006) *Plant Breeding, The Arnel R. Hallauer International Symposium*. Blackwell, Ames, Iowa.

Lande, R. and Thompson, T. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756.

Lander, E.S. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.

Legarra, A. and Fernando, R.L. (2009) Linear models for joint association and linkage QTL mapping. *Genetics Selection Evolution* 41, 43–60.

Li, J., Xiao, J., Grandillo, S., Jiang, L., Wan, Y., Deng, Q., Yuan, L. and McCouch, S. (2004) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. *Genome* 47, 697–704.

Lorenz, A.J, Chao, S., Asoro, F.G., Heffner, E.L., Hayashi, T., Iwata, H., Smith, K.P., Sorrells, M.E. and Jannink, J.L. (2011) Genomic selection in plant breeding: knowledge and prospects. *Advances in Agronomy* 110, 77–123.

Lorenzana, R. and Bernardo, R. (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics* 120, 151–161.

Lu, F., Ammiraju, J.S.S., Sanyal, A., Zhang, S., Song, R., Chen, J., Li, G., Sui, Y., Song, X., Cheng, Z., de Oliveira, A.C., Bennetzen, J.L., Jackson, S.A., Wing, R.A. and Chen, M. (2009) Comparative sequence analysis of *MONOCULM1*-orthologous regions in 14 *Oryza* genomes. *Proceedings of the National Academy of Sciences USA* 106(6), 2071–2076.

Mauricio, R. (2001) Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Reviews Genetics* 2, 370–381.

McCouch, S.R., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septinginsih, E., Edwards, J., Moncada, P., Xiao, J., Garris, A., Tai, T., Martinez, C., Tohme, J., Sugiono, M., McClung, A., Yuan, L.P. and Ahn, S.N. (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa. Euphytica* 154, 317–339.

McCouch, S.R., Zhao, K., Wright, M., Tung, C.-W., Ebana, K., Thomson, M., Kimball, J., Reynolds, A., Tyagi, W., Wang, D., DeClerck, G., Ali, M.L., McClung, A., Eizenga, G. and Bustamante, C. (2010) Development of genome-wide SNP assays for rice. *Breeding Science* 60, 524–535.

McCouch, S.R., McNally, K.L., Wang, W. and Sackville-Hamilton, R. (2011) Genomics of gene banks: a case study in rice. *American Journal of Botany* 99(2), 407–423. DOI:_10.3732/ajb.1100385.

McMullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., Brown, P., Browne, C., Eller, M., Guill, K., Harjes, C., Kroon, D., Lepak, N., Mitchell, S.E., Peterson, B., Pressoir, G., Romero, S., Rosas, M.O., Salvo, S., Yates, H., Hanson, M., Jones, E., Smith, S., Glaubitz, J.C., Goodman, M., Ware, D., Holland, J.B. and Buckler, E.S. (2009) Genetic properties of the maize nested association mapping population. *Science* 325, 737–740.

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Moncada, M., Martinez, C., Tohme, J., Chatel, M., Borrero, J., Gauch, H. and McCouch, S. (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC$_2$F$_2$ population evaluated in an upland environment. *Theoretical and Applied Genetics* 102, 41–52.

Nandi, S., Subudhi, P.K., Senadhira, D., Manigbas, N.L., Sen-Mandi, S. and Huang, N. (1997) Mapping QTLs for submergence tolerance in rice by ADLP analysis and selective genotyping. *Molecular and General Genetics* 255(1), 1–8.

Navabi, A., Mather, D.E, Bernier, J., Spaner, D.M. and Atlin, G.N. (2009) QTL detection with bidirectional and unidirectional selective genotyping: marker-based and trait-based analyses. *Theoretical and Applied Genetics* 118(2), 347–358.

Nayar, M.N. (2010) Origin of African rice from Asian rice. In: *Innovation and Partnerships to Realize Africa's Rice Potential*. Proceedings of the Second Africa Rice Congress. Africa Rice Center, Cotonou, Benin, pp. 1.18.1–1.18.13. Available at: www.africarice.org/workshop/ARC/1.18%20Nayar%20ed2.pdf (accessed 23 February 2012).

Neeraja, C., Maghirang-Rodriguez, R., Pamplona, A., Heuer, S., Collard, B., Septiningsih, E., Vergara, G., Sanchez, D., Xu, K. and Ismail, A.M. (2007) A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theoretical and Applied Genetics* 115, 767–776.

Ni, J., Pujar, A., Youens-Clark, K., Yap, I., Jaiswal, P., Tecle, I., Tung, C.-W., Ren, L., Spooner, W., Wei, X., Avraham, S., Ware, D., Stein, L. and McCouch. S.R. (2009) Gramene QTL database: development, content and applications. *Database* 2009, bap005. DOI:10.1093/database/bap005.

Ondov, B.D., Varadarajan, A., Passalacqua, K.D. and Bergman, N.H. (2008) Efficient mapping of Applied Biosystems SOLiD sequence data to a reference genome for functional genomic applications. *Bioinformatics* 24, 2776.

Porteres, R. (1962) Berceaux agricoles primaires sur le continent Africain. *Journal of African History* 3(2), 195–210.

Rafalski, A. (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* 5, 94–100.

Reece, J.D. and Haribabu E. (2007) Genes to feed the world: the weakest link? *Food Policy* 32(4), 459–479.

Sakai, H., Ikawa, H., Tanaka, T., Numa, H., Minami, H., Fujisawa, M., Shibata, M., Kurita, K., Hamada, M., Kanamori, H., Namiki, N., Wu, J., Itoh, T., Matsumoto, T. and Sasaki, T. (2011) Distinct evolutionary patterns of *Oryza glaberrima* deciphered by genome sequencing and comparative analysis. *The Plant Journal* 66(5), 796–805.

Sanyal, A., Ammiraju, J.S., Lu, F., Yu, Y., Rambo, T., Currie, J., Kollura, K., Kim, H.R., Chen, J., Ma, J., San Miguel, P., Mingsheng, C., Wing, R.A. and Jackson, S.A. (2010) Orthologous comparisons of the *Hd1* region across genera reveal *Hd1* gene lability within diploid *Oryza* species and disruptions to microsynteny in sorghum. *Molecular Biology and Evolution* 27(11), 2487–2506.

Sarla, N. and Mallikarjuna Swamy, B.P. (2005) *Oryza glaberrima*: a source for the improvement of *Oryza sativa. Current Science* 89, 955.

Schuster, S.C. (2008) Next-generation sequencing transforms today's biology. *Nature Methods* 5, 16–18.

Septiningsih, E.M., Pamplona, A.M., Sanchez, D.L., Neeraja, C.N., Vergara, G.V., Heuer, S., Ismail, A.M. and Mackill, D.J. (2008) Development of submergence-tolerant rice cultivars: the *Sub1* locus and beyond. *Annals of Botany* 103, 151–160.

Shanti, M.L., Devi, G.L., Kumar, G.N. and Shashidhar, H.E. (2010) Molecular marker-assisted selection: a tool for insulating parental lines of hybrid rice against bacterial leaf blight. *International Journal of Plant Pathology* 1, 114–123.

Song, W.-Y., Wang, G.-L., Chen, L.-L., Pi, L.-Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.-X., Zhu, L.-H., Fauquet, C. and Ronald, P. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21. Science* 270(5243), 1804–1806.

Sun, Y., Wang, J., Crouch, J.H. and Xu, Y. (2010) Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. *Molecular Breeding* 26, 493–511.

Tanksley, S.D. (1993) Mapping polygenes. *Annual Review of Genetics* 27, 205–233.

Tanksley, S.D. and McCouch, S.R. (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066.

Thomson, M.J., Zhao, K., Wright, M., McNally, K.L., Rey, J., Tung, C.-W., Reynolds, A., Scheffler, B., Eizenga, G., McClung, A., Ismail, A., de Ocampo, M., Mojica, C., Reveche, M.Y., Dilla, C.J., Mauleon, R., Leung, H., Bustamante, C. and McCouch, S.R. (2011) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding*. DOI: 10.1007/s11032-011-9663-x (accessed 23 February 2012).

Vaughan, D.A. (1994) *The Wild Relatives of Rice: A genetic resources handbook*. International Rice Research Institute, Manila, Philippines.

Venuprasad, R., Dalid, C.O., Del Valle, M., Zhao, D., Espiritu, M., Sta Cruz, M.T., Amante, M., Kumar, A. and Atlin, G.N. (2009) Identification and characterization of large-effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. *Theoretical and Applied Genetics* 120, 177–190.

Venuprasad, R., Impa, S., Veeresh Gowda, R.P., Atlin, G.N. and Serraj, R. (2011a) Rice near-isogenic-lines (NILs) contrasting for grain yield under lowland drought stress. *Field Crops Research* 123(1), 38–46.

Venuprasad, R., Bool, M.E., Quiatchon, L. and Atlin, G.N. (2011b) A QTL for rice grain yield in aerobic environments with large effects in three genetic backgrounds. *Theoretical and Applied Genetics*. DOI: 10.1007/s00122-011-1707-4 (accessed 23 February 2012).

Venuprasad, R., Bool, M.E., Quiatchon, L., Sta Cruz, M.T., Amante, M. and Atlin G.N. (2011c) A large-effect QTL for rice grain yield under upland drought stress on chromosome 1. *Molecular Breeding*. DOI: 10.1007/s11032-011-9642-2 (accessed 23 February 2012).

Wang, L., Wang, A., Huang, X., Zhao, Q., Dong, G., Qian, Q., Sang, T. and Han, B. (2011) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. *Theoretical and Applied Genetics* 122(2), 327–340.

Xiao, J., Li, J., Grandillo, S., Ahn, S., Yuan, L., Tanksley, S. and McCouch, S. (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon. Genetics* 150, 899–909.

Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., Li, J., He, W., Zhang, G., Zheng, X., Zhang, F., Li, Y., Yu, C., Kristiansen, K., Zhang, X., Wang, J., Wright, M., McCouch, S., Nielsen, R., Wang, J. and Wang, W. (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology* 30(1), 105–111.

Yu, J., Presspoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S. and Buckler, E.S. (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38(2), 203–208.

Yu, J., Holland, J.B., McMullen, M.S. and Buckler, E.S. (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178, 539–551.

Zhang, Q., Li, J., Xue, Y., Han, B. and Deng, X.W. (2008) Rice 2020: a call for an international coordinated effort in rice functional genomics. *Molecular Plant* 1, 715–719.

Zhao, K., Tung, C.-W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., McClung, A.M., Bustamante, C.D. and McCouch, S.R. (2011) Genome-wide association mapping reveals rich genetic architecture of complex traits in *Oryza sativa. Nature Communications*. DOI: 10.1038/ncomms1467 (accessed 23 February 2012).

Zhong, S., Dekkers, J.C.M., Fernando, R.L. and Jannink, J.L. (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182, 355–364.

Zhu, C., Gore, M., Buckler, E.S. and Yu, J. (2008) Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20.