

More genomic resources for less-studied crops

Rajeev K. Varshney^{1,2,3}, Jean-Christophe Glaszmann^{2,4}, Hei Leung⁵ and Jean-Marcel Ribaut²

¹ International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, A.P., India

² Generation Challenge Programme (GCP), c/o CIMMYT, Int APDO Postal 6-641, 06600 Mexico, DF, Mexico

³ School of Plant Biology (M084), Faculty of Natural and Agricultural Sciences, The University of Western Australia (UWA), 35 Stirling Highway, Crawley, WA 6009, Australia

⁴ UMR DAP, CIRAD, TA A96/03, Avenue Agropolis, 34398 Montpellier, Cedex 5, France

⁵ International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila 1301, The Philippines

Many of the crop species considered to be minor on a global scale, yet are important locally for food security in the developing world, have remained less-studied crops. Recent years have witnessed the development of large-scale genomic and genetic resources, including simple sequence repeat, single nucleotide polymorphism and diversity array technology markers, expressed sequence tags or transcript reads, bacterial artificial chromosome libraries, genetic and physical maps, and genetic stocks with rich genetic diversity, such as core reference sets and introgression lines in these crops. These resources have the potential to accelerate gene discovery and initiate molecular breeding in these crops, thereby enhancing crop productivity to ensure food security in developing countries.

Introduction

Trends in both the growth of the human population and its pattern of consumption in developing countries imply that the global demand for food will continue to grow for at least another 40 years. The focus now is not only on increasing the food supply, but also on improving its quality, particularly with respect to nutritional value [1,2]. Molecular breeding, which combines genotypic and phenotypic information, has emerged as a powerful approach and offers new perspectives to help address these challenges [3,4]. Genetic studies and molecular breeding approaches require basic genomic resources, such as molecular markers, genetic maps and sequence information, which were not available for several less-studied crop species until recently. Depending on the availability of those genomic resources, the main food crops of the world that are relevant to developing countries and are mandate crops of Consultative Group on International Agricultural Research (CGIAR), could be grouped into three tiers (Table 1). Considering the importance of less-studied crops in tiers 2 and 3 for balanced diet or income in developing countries, several international initiatives have contributed significantly to the development of genomic resources in these crops. The objective is twofold: first, to better understand the genetic nature and composition of their genome; and second, to afford molecular breeding

Glossary

Chromosome segment substitution line (CSSL): Each CSSL carries a single or a few defined chromosome segments from a donor line in the genetic background of another line. These lines constitute useful resources for genetic studies, including detection and fine mapping of QTLs for the target traits. In particular, CSSLs can be used for the detection of QTLs with small additive effects that are masked by QTLs with large effects in filial generation 2 (F₂) and recombinant inbred line populations.

Core collection: Generally contains 5–10% of the germplasm collection, and ideally represents at least 70% of the alleles in the whole collection.

Core reference set: A set of genetic stocks that are representative of the genetic resources of the crop and are used by the scientific community as a reference for an integrated characterization of its biological diversity.

Diversity array technology (DART) markers: Represents a low-cost, high-throughput, robust marker system with minimal DNA requirements, which is capable of providing comprehensive genome coverage even for species that do not possess any genomic information. DART is a segment of genomic DNA, the presence of which is polymorphic in a defined genomic representation.

Expressed sequence tag (EST): Short (300–500 bp), single-pass sequence obtained from the 5' or 3' end of mRNA (cDNAs). ESTs represent a snapshot of genes that are expressed at a given developmental stage or in a given tissue.

Gene space: Long, gene-rich regions that contain the vast majority of genes, interrupted by long, gene-poor regions in a genome of a given species. Presence of gene space is a common feature of plants, including cereal species with large genome size owing to the abundance of repetitive DNA (retro and/or transposons) in their genome.

Introgression lines (ILs): A series of lines, each carrying a chromosomal segment from a donor (wild relative) genotype in the background of an elite genotype. Therefore, they constitute an “exotic library” where the entire wild-species genome is partitioned among several lines, each carrying a single homozygous introgressed segment. The nearly isogenic nature of the ILs provides a relative advantage over other segregating populations when rapidly implementing a pyramiding approach via crosses and marker analysis. ILs constitutes an efficient source for the discovery and utilization of genes or QTLs that control traits of agricultural importance.

Map-based cloning (MBC): Also known as positional cloning, MBC is a universal strategy for cloning genes that are fine-mapped or genes whose genetic and physical positions are known.

Marker-assisted selection (MAS): A process of indirect selection for improving the traits of interest by employing markers, such as morphological, biochemical or DNA-based markers. DNA-based molecular markers have been considered the markers of choice for MAS in recent years.

Molecular breeding: A process of deployment of molecular markers in breeding practices for improving selection efficiency and for reducing the breeding cycles during cultivar improvement. MAS, marker-assisted recurrent selection, and genomic selection are different types of molecular breeding approaches.

Molecular marker: DNA-based genetic marker that can detect DNA polymorphisms both at the level of specific loci and at the whole-genome level. In recent years, a variety of molecular markers have become available. Restriction fragment length polymorphisms (RFLPs) were the first to be developed; others include random amplification of polymorphic DNAs (RAPDs), cleaved amplified polymorphic sequence (CAPS), SSRs and amplified fragment length polymorphisms (AFLPs). The most recently developed markers are SNPs and single-feature polymorphisms (SFPs).

Next generation sequencing (NGS) technologies: Modern sequencing technologies, driven by automation, that have surpassed Sanger sequencing

Corresponding author: Varshney, R.K. (r.k.varshney@cgiar.org).

approach in throughput and in cost-effectiveness of sequencing. At present, Roche-454/FLX, Illumina/Solexa and AB SOLiD are the NGS technologies that are in routine use for re-sequencing as well as *de novo* sequencing.

Tentative unique sequence (TUS): A non-redundant transcript sequence that is usually defined based on cluster analysis of short sequence reads generated by NGS technologies.

towards better performance of different cultivars in target environments. The Generation Challenge Programme (GCP), which links basic science with applied research through a broad network of plant scientists from diverse backgrounds, is one of the international initiatives in the 'research for development' arena (www.generationcp.org/gcppartners.php?da=0646141) [5]. As a result of this effort, a significant amount of genomic resources as well as genetic stocks, such as core reference sets of CGIAR and

national genebanks [6] and introgression libraries [7], have been developed in several less-studied crop species.

Studies conducted on plant models like *Arabidopsis thaliana*, *Medicago truncatula* and *Lotus japonicus* have opened the road to the development and application of new technologies and approaches [8]. The current challenge, however, is to deploy genomic information from model crops as well as crop-species-specific genomic and plant-genetic resources in breeding programmes. Here, we attempt to encapsulate progress achieved to date regarding the development of genomic resources and their applications for accessing genetic diversity and identifying favorable alleles for plant breeding. Challenges and opportunities for the deployment of these genomic and genetic resources in breeding programmes, particularly in developing countries, are also discussed.

Table 1. Availability of public domain genomic resources for each of the CGIAR/GCP mandate crops

Crop name	Botanical name	Chromosome number (n), ploidy level (x), (genome size in Mbp)	Marker (mainly SSR and SNP) resource	Genetic map density	Physical map	Transcriptomic resource and gene discovery	Genetic resources and diversity characterization	Trait mapping and molecular breeding
Tier 1 crops								
Barley	<i>Hordeum vulgare</i>	2n=2x=14 (5,500)	++++	+++	+++	+++	+++	++
Maize	<i>Zea mays</i>	2n=2x=20 (2,500)	++++	++++	++++	+++	+++	++++
Rice	<i>Oryza sativa</i>	2n=2x=24 (430)	++++	++++	++++	+++	+++	++++
Sorghum	<i>Sorghum bicolor</i>	2n=2x=20 (750)	+++	+++	+++	+++	+++	+++
Wheat	<i>Triticum aestivum</i>	2n=6x=42 (16,000)	++++	+++	+++	+++	+++	+++
Tier 2 crops								
Chickpea	<i>Cicer arietinum</i>	2n=2x=16 (740)	+++	+++	NA	+++	+++	+++
Common bean	<i>Phaseolus vulgaris</i>	2n=2x=22 (~637)	+++	+++	++	++	+++	+++
Cowpea	<i>Vigna unguiculata</i>	2n=2x=22 (620)	+++	+++	++	++	+++	++
Groundnut	<i>Arachis hypogaea</i>	2n=4x=40 (2,890) 2n=2x=20 (A and B genome: 1,260)	++	++	++	+	+++	++
Potato	<i>Solanum tuberosum</i>	2n=4x=48 (850- 1000)	+++	++	++	++	+++	++
Tier 3 crops								
Banana	<i>Musa acuminata</i>	2n=2x=22 (552-607)	++	++	+	++	++	+
Cassava	<i>Manihot esculenta</i>	2n=2x=36 (760)	++	++	++	++	++	+
Coconut	<i>Cocos nucifera</i>	2n=2x=32 (3,600)	++	++	NA	NA	++	NA
Lentil	<i>Lens culinaris</i>	2n=2x=14 (4,063)	+	+	NA	+	+	+
Millet	<i>Pennisetum glaucum</i>	2n=2x=14 (2,450)	++	++	NA	+++	++	+
Pigeonpea	<i>Cajanus cajan</i>	2n=2x=22 (858)	++	+	NA	++	++	NA
Sweet potato	<i>Ipomoea batatas</i>	2n=6x=90 (2,200- 3,000)	++	++	NA	++	++	NA
Yam	<i>Dioscorea alata</i>	2n=2x=40 (550)	+	+	NA	NA-	+	NA

Abbreviations: NA, not available.

Availability of resources has shown by '+' sign as follows: +, basic; ++, moderate; +++, good; +++++, excellent. '+' colored red represents the contribution of GCP; those in black were already developed and are publicly available.

Current molecular marker technologies

Molecular markers are key tools for assaying genetic variation. Advances in molecular genetic technology have generated a range of marker types, with many reliant on PCR. Some of these widely used marker systems include simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs) and diversity array technologies (DArTs) [9]. SSRs, or microsatellite markers, for various practical reasons, have been the leading marker types for breeding applications [10]. The DArT marker system has proven extremely effective in exposing genetic diversity in germplasm collections, and is well-suited for background selection in molecular breeding programmes [11]. Finally, SNPs, which have been the prime marker platform in human and animal genetics for some time, are now rapidly gaining momentum and are overtaking SSRs as the marker type of choice in plants as well, mainly because of their amenability for high-throughput genotyping and cost-effectiveness [12,13].

SSR and DArT markers

Tier 1 species have benefited from an adequate supply of molecular markers, particularly SSRs (Table 1) [14]. A significant effort has therefore been dedicated recently towards developing SSRs in the tier 2 and 3 crops. Two major approaches have been deployed: the first is based on the construction of SSR-enriched genomic DNA libraries, and the second is centered on an *in silico* analysis of end sequences of clones from bacterial artificial chromosome (BAC) libraries or expressed sequence tag (EST) sequences. As a result, hundreds of SSR markers have become available in several crops, such as chickpea [15,16], common bean [16,17], groundnut [16,18,19] and pigeonpea [16,20] (Table 1). However, the cost of developing SSR markers, especially via the enriched library route, remains rather high.

The need for a whole genome assay, required both for the characterization of diversity and for development of

genetic maps, has led to the application of DArT technology to several tier 2 and 3 crop species (www.diversityarrays.com/genotypingserv.html). Although the first set of DArT arrays were developed for pearl millet, potato, sweet potato, cassava, yam, coconut and *Musa*, the expansion of existing DArT arrays was accomplished for chickpea, groundnut and pigeonpea (Table 2). DArT markers have since proven useful in the areas of germplasm collection diversity assessment [21,22], genome mapping [22–24] and gene tagging [25], and should be useful in marker-assisted selection (MAS), especially for monitoring the recovery of the recurrent parent genome.

BAC libraries

BAC libraries are an important component of physical mapping, map-based cloning, and the analysis of gene and genome structure and function. BAC libraries that provide extensive genome coverage have been established in rice [26], maize [27], wheat [28], barley [29], sorghum [30], soybean [31], common bean [32] and potato [33], among others. Recently, BAC libraries have been constructed in a number of less high profile crop species, such as cowpea, cassava and groundnut [16].

Two independent cowpea BAC libraries, comprising ~74 000 clones and 17× genome coverage, have been developed and high-information-content fingerprinting (HICF) analysis [34] of 60 000 of these clones has allowed their assembly into 790 contigs [35], (<http://phymap.ucdavis.edu/cowpea/>). A similar project in cassava has produced a BAC library of 72 000 clones that represents approximately 10× coverage of the genome. Following the HICF method, ~58 000 clones have been assembled into 2104 contigs [36], (<http://cassava.igs.umaryland.edu/>). Although a 6.5× groundnut library had already been established [37], two new BAC libraries, one derived from the groundnut A genome progenitor *Arachis duranensis*, and the other from its B genome progenitor, *Arachis ipaënsis*, have been constructed recently [38] for finger-

Table 2. CGIAR/GCP crop species for which DArT markers are available

Crop	Genotypes used for developing library	Features on array	Diversity panel surveyed	Polymorphic features ^a	Marker loci on genetic map
Cereal crops					
Barley	>200	>50,000	>700	3,072	~3,000
Wheat	>300	>80,000	>1,000	7,680	~5,000
Rice	>150	>30,000	>300	1,900	~1,000
Sorghum*	>200	>20,000	>1,000	2,500	2,000
Pearl millet*	96	7,680	96	1,500	In progress
Legume crops					
Common bean	150	>22,000	150	2,500	NA
Chickpea*	250	21,500	300	5,400	>700
Groundnut*	164	15,360	300	>5,000	~200
Pigeonpea*	300	29,000	400	>5,000	>400
Root, tuber and other crops					
Potato*	>150	>20,000	>300	9,000	>4,000
Sweet potato*	96	7,680	96	2,000	800
Cassava*	270	18,000	450	2,500	450
Yam*	94	6,890	94	>2,500	NA
Coconut*	130	7,680	130	400	NA
<i>Musa</i> *	100	15,360	300	>5,000	>500

Abbreviation: NA, not available.

Source: DArT Pty Ltd., Australia (www.diversityarrays.com).

*Species for which the research was supported by GCP [65]. A complete list of crop species for which DArT arrays are available is posted at www.diversityarrays.com.

^aThe number of informative markers in chickpea, groundnut and pigeonpea includes related wild species. The number for yam is a projection of current work.

printing. In summary, the physical maps of several tropical crop species are now under development. As physical maps are integrated with genetic maps, they will provide the framework for full genome sequencing.

Sequence-based, next generation genomic resources

Recent years have witnessed the advent of a number of high-throughput sequencing [39–41] as well as genotyping technologies [12,13]. These approaches have great potential to facilitate development of genomic resources even in those crops that have a shortage of genomic resources. Such potential and applications of next generation sequencing (NGS) technologies like Roche/454 and Illumina/Solexa sequencing have been discussed in detail elsewhere [39–41]. In fact, as a result of the use of NGS technologies, whole-genome sequences are becoming available in these plant species, which did not have even basic genomic resources like marker repertoire and genetic maps [42]. Therefore, these approaches have been used for developing sequence-based genomic resources at a fast pace in some tier 2 and 3 crop species, which were considered ‘orphan crops’ in the past (Table 1) [16].

Transcriptomic resources

Several temperate cereal ESTs have been in the public domain for some time (dbEST release 042310, www.ncbi.nlm.nih.gov/dbEST/), but this has not been the case for less-studied crops like cowpea, chickpea, pearl millet or sweet potato. In recent years, however, because of concerted efforts of GCP partners and others, transcriptomic resources in the form of ESTs (generated by Sanger sequencing) and/or short transcript reads (STRs; generated by NGS technologies) have been generated [16]. Large number of ESTs (e.g. 180 000 for cowpea [43], ~20 000 for chickpea [44] and ~11 000 for pigeonpea [45]) have become available recently. On the other hand, Roche/454 has been used to generate >435 000 chickpea and ~495 000 pigeonpea STRs [16,39]. Similarly, ~113 000 tentative unique sequences (TUSs) for pearl millet [46] and ~60 000 TUSs for sweet potato [47] have been defined based on STRs.

SNP resources and high-throughput genotyping platforms

Technological advances in both detection and robotics have seen SNPs largely replacing SSRs as markers for mammalian genetics, and their use in plants is expanding [13]. International collaborations have generated large-scale SNP resources and genotyping platforms in rice (www.ricesnp.org) [48], maize (www.maizegenetics.net) [49,50], wheat (<http://wheat.pw.usda.gov/SNP/new/index.shtml>) [51], barley [52] and soybean [53]. Initiatives like GCP have extended SNP technology to various tropical crop species. Two main approaches have been applied for SNP discovery: the first is based on available EST and STR sequences, and the second on allele re-sequencing of candidate genes or BAC ends. The cowpea SNP programme has delivered ~10 000 SNPs from an *in silico* comparison of ~183 000 ESTs generated from >10 cultivars (<http://harvest.ucr.edu/>) [43]. Re-sequencing of an average of ~600 genes per species among the mapping

parents used in cowpea, chickpea, common bean, pigeonpea, and one of the diploid progenitors of groundnut has identified ~16 000 SNPs in these tropical legume species [16,54]. In the cases of chickpea and pigeonpea, Illumina/Solexa sequencing has also been used to generate STRs for parental genotypes of mapping populations. Alignment of STRs with respective transcriptome assemblies (TUSs) has allowed identification of several thousands SNPs in each chickpea and pigeonpea at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT: www.icrisat.org) [16].

A number of high-throughput SNP genotyping platforms is currently available [13]; among these, the GoldenGate assay has become popular because its unit data point cost is very competitive. A 1536-feature GoldenGate assay has been developed for cowpea [43], as well as a 768-feature GoldenGate assay for cowpea, chickpea, pigeonpea, common bean and one of the diploid progenitors of groundnut [16,54]. Similarly, a SNP-based GoldenGate assay has been developed for cassava. The advent of SNP genotyping has led to the shift from the low-throughput primarily manual world of SSRs to high-throughput platforms that are powered by robotics and automated scoring. As a result, genotyping costs with SNPs have decreased by up to 10-fold, whereas data throughput has increased by the same magnitude, which has enabled a great leap in genotyping efficiency [55].

Genome or gene space sequencing

Rice was the first crop for which genome sequencing was completed [56–58], and genome sequences have recently become available for sorghum [59] and maize [60]. Similar efforts are underway for wheat [61] and barley [62]. Although NGS technologies were originally meant for re-sequencing, ever-increasing sequence read length and advances in bioinformatics have made NGS possible for undertaking *de novo* genome sequencing [41,42]. Although gene space was sequenced for cowpea [63], these novel approaches have enabled complete genome sequencing for species like cassava (www.phytozome.net/cassava) and potato (www.potatogenome.net), and some efforts have been initiated in pigeonpea (www.icrisat.org/gt-bt/IIPG/Home.html). Sequencing of the AA genome of *Musa* has also been announced by Genoscope and the French Centre for International Agricultural Research (CIRAD) (www.musagenomics.org). In the coming years, sequences of either the whole genome or at least the gene space should become available for a majority of crop species, including the less-studied species.

Novel and better-characterized sources of genetic diversity

The use and creation of genetic diversity to stack favorable alleles into germplasm for enhanced agronomic performance in target environments is a primary task of the plant breeder. Worldwide, ~1400 *ex situ* genebanks curate six million accessions, which cover most of the common crop species [6]. Direct identification of favorable alleles in genebank accessions, however, is often masked due to poor adaptation or poor agronomical performance of the genebank material. Therefore, several approaches have been

Table 3. Development of core reference set for CGIAR mandate crop species

Crop ^a	Composite or core collection for molecular characterization	SSR markers	Size of reference set defined	Lead institution and partners ^b
Tier 1 crops				
Barley	2,676	14	300	ICARDA, CAAS
Maize	1,775	50	234	CIMMYT, CAAS, IITA, Agropolis
Rice	2,757	50	Analysis in progress	IRRI, CAAS, CIAT, WARDA, Agropolis
Sorghum	3,393	39	345	ICRISAT, Agropolis, CAAS
Wheat	3,000	50	372	CIMMYT, Agropolis-CAAS-ICARDA
Tier 2 crops				
Chickpea	3,024	50	300	ICRISAT, ICARDA
Common bean	625	36	192	CIAT, EMBRAPA
Cowpea	1,871	16	345	IITA
Groundnut	991	21	300	ICRISAT, EMBRAPA
Potato	1,000	50	Analysis in progress	CIP
Tier 3 crops				
Banana	327	48	96	Bioversity, IITA, Agropolis
Cassava	3,000	36 + DArTs	250	CIAT, EMBRAPA, IITA
Coconut	1,014	30	359	Agropolis
Faba bean	1,000	20	152	ICARDA
Finger millet	1,000	20	300	ICRISAT
Foxtail millet	500	20	200	ICRISAT
Lentil	1,000	24	150	ICARDA
Pearl millet	1,000	20	300	ICRISAT
Pigeonpea	1,000	20	300	ICRISAT
Sweet potato	500	50	Analysis in progress	CIP
Yam	350	20	342	IITA

^aCrop species in which molecular characterization has been supported by GCP.

^bLead institute is listed first, and collaborating institutes follow: Agropolis: Agropolis International, France (www.agropolis.fr); Bioversity: Bioversity International (www.bioversityinternational.org); CAAS: Chinese Academy of Agricultural Sciences, China (www.caas.net.cn); CIAT: International Center for Tropical Agriculture, Colombia (www.ciat.cgiar.org); CIMMYT: International Maize and Wheat Improvement Center, Mexico (www.cimmyt.org); CIP: International Potato Center, Peru (www.cipotato.org); EMBRAPA: Brazilian Agricultural Research Corporation, Brazil (www.embrapa.br); ICARDA: International Center for Agricultural Research in the Dry Area, Syria (www.icarda.org); IITA: International Institute of Tropical Agriculture, Nigeria (www.iita.org); IRRI: International Rice Research Institute, Philippines (www.irri.org); WARDA: The Africa Rice Center, Benin (www.warda.org).

suggested to create, characterize and tap genetic diversity in the breeding programme; two cases in which significant efforts have been made for assessing and capturing favorable alleles for breeding are discussed in the following sections.

Core reference sets

Most likely, a very small portion of the *ex situ* accessions has been used in plant breeding programmes. With an objective to enhance use of germplasm in breeding, various formulations for selecting representative sub-collections have been derived from the original concept of the core collection [64]. For each crop listed in Table 3, a systematic molecular characterization of germplasm from various collections has been conducted to define the core reference set to enable rationale sub-sampling. The central idea is to promote the use of a common set of representative material by the scientific community. This concept of core reference set to help R.E.A.D. (represent existing diversity - enter the whole collection - assess phenotype variation - dissect trait-gene associations) germplasm has been described [6].

As a result of concerted efforts of scientists from CGIAR centres and national programmes coordinated and supported by GCP, genotyping of ~32 000 accessions, which represents global composite collections of 21 species from 12 CGIAR/advanced research institute (ARI)/national agricultural research system (NARS) genebanks has been undertaken at 14–50 SSR loci or with DArT arrays (Table 3) [65]. The genotypic data were initially used to design a core reference set for each species, which consisted of 96–359 accessions. The aim of the core reference set was to

capture ~80% of the molecular diversity present in the sampled germplasm [66]. The next step is to gather robust phenotypic data sets, assembled from trials conducted under a range of agro-climatic conditions and genome-wide high-density marker data on these core reference sets. These data, although challenging, are essential for the identification of phenotype/genotype associations that might deliver marker tags for agronomic traits. In addition, multi-location field trials could also succeed in identifying lines that might be included in breeding efforts because they perform well with respect to a specific trait of interest.

Introgression libraries

Depending on the presence of a superior allele for a trait, such as disease resistance, in a crop relative or wild species, the identification of favorable alleles can be complicated because of often-poor agronomic performance of such non-adaptive germplasm. In such cases, production of a complete set of chromosome segment substitution lines (CSSLs) [e.g. recombinant inbred introgression lines (ILs)] provides avenues to explore the largely untapped reservoir of useful alleles of interest that remain in the wild species. Such a case is shown in Figure 1, in which the ILs of rice have been developed by using chromosome segments of *Oryza meridionalis*. Such lines enable assessment of the phenotypic impact of the wild species genome in parts while keeping the genetic background of the cultivated recipient genotype. This strategy has been widely utilized for introgression of favourable alleles from wild relatives in tomato, rice, wheat and barley [4].

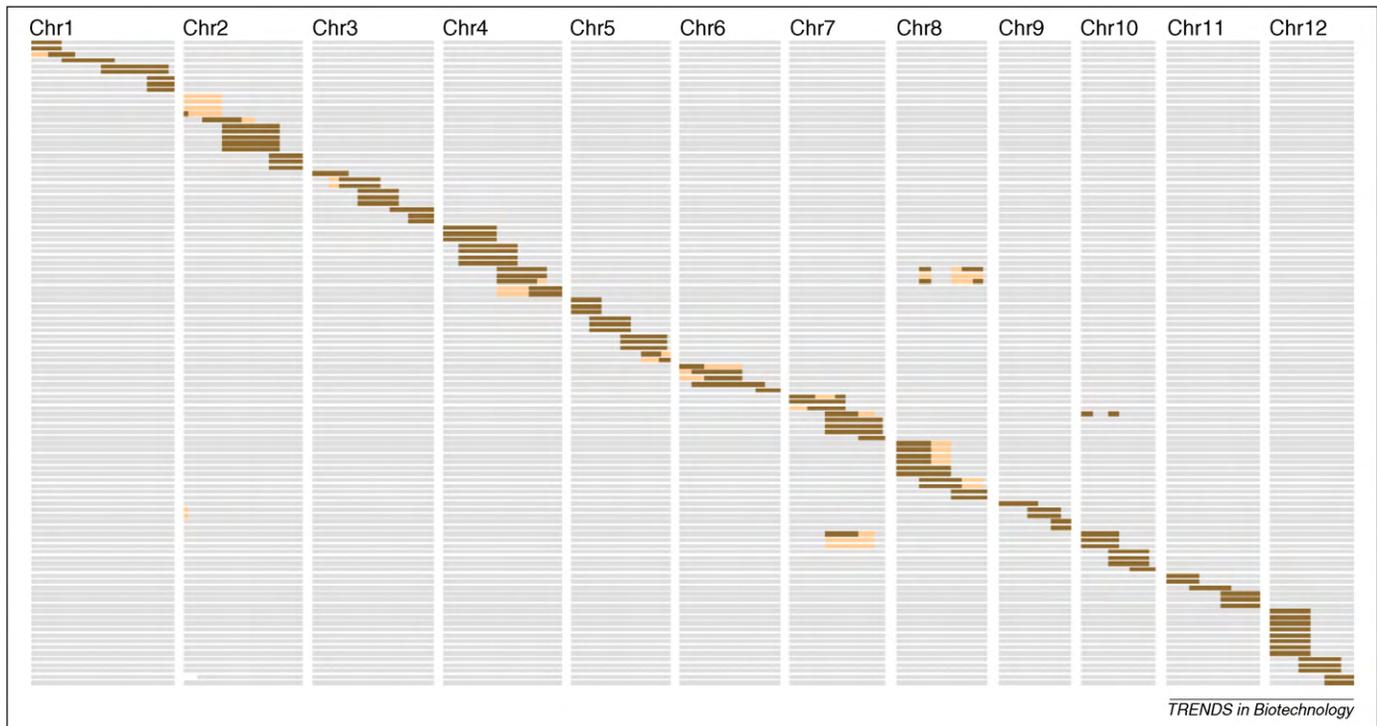


Figure 1. Graphical genotypes of a set of rice CSSLs, selected in the BC₃ generation of a cross between cultivated rice and *O. meridionalis* by researchers at CIAT (Source: Mathias Lorieux and colleagues). Background selection was carried out using a genome-wide set of SSR and miniature inverted repeat transposable elements markers. The filled brown bars represent *O. meridionalis* segments, and the orange bars represent fragment reduction or background recovery. Abbreviation: Chr, chromosome.

Inspired by these success stories, the International Center for Tropical Agriculture (CIAT) in Colombia in collaboration with the Institut de Recherche pour le Développement (IRD), and CIRAD in collaboration with the Institut Sénégalais de Recherches Agricoles (ISRA) have developed ILs to transfer exotic alleles into elite lines of rice and groundnut, respectively. The donors of the various sets of *Oryza* spp. ILs involve *Oryza glaberrima*, *O. meridionalis*, *O. rufipogon*, *O. glumaepatula* and *O. barthii*, and the ILs have featured in a variety of genetic analyses and strategic breeding crosses. An example of the former is the use of ILs derived from the cross *O. sativa* with *O. glaberrima* to identify quantitative trait loci (QTLs) for plant height, tiller number, panicle length, male fertility, grain weight and grain yield. Of particular note is the genetic characterization and mapping of a major QTL that underlies resistance to the stripe necrosis virus [67]. Introgression in groundnut has been based on a cross between a cultivated type and synthetic tetraploid derived from a hybrid between the diploid progenitor species *A. ipaënsis* and *A. duranensis* [68]. The level of marker information between the mapping population parents has allowed for the construction of a coarse-scale genetic map (298 loci spread over 21 linkage groups, mean inter-marker distance of 6.1 cM). A back-cross strategy has been applied to the F₁ hybrid to reduce the proportion of the progenitor genome present, and these materials are seen as having substantial potential to extend the genetic base of cultivated groundnut.

The impact of genomic resources on breeding

The productivity of most tropical crops is limited by a combination of biotic and abiotic constraints. For most of these, plants have evolved variation for the level of

tolerance or resistance. A crucial pre-requisite for the deployment of markers to support stress tolerance or resistance breeding is the development of a genetic map, followed by identification of gene-based or gene-linked markers to be used in a predictive mode for plant breeding, such as MAS [69]. However, until a few years ago, maps in cowpea, chickpea, pigeonpea, pearl millet and sweet potato were either non-existent or displayed an inadequate level of marker density. The far-reaching improvement in molecular marker technology witnessed over the past few years has helped to put in place serviceable genetic maps in all these species (Table 1). As an example, an AFLP-based cowpea map has recently been replaced by a 928-SNP marker map that consists of 11 linkage groups with a mean inter-marker distance of <1 cM [43]. The situation in groundnut is even more extreme: marker polymorphism in the species had been thought to be too poor to construct a genetic map, but now SSR-based linkage maps have been constructed for both diploid progenitors [18,70] and tetraploid species [71]. The reference chickpea genetic map developed at ICRISAT comprises >1500 data points [15,16], and mapping projects in pearl millet and sweet potato are being conducted at ICRISAT and the International Potato Center, Peru (CIP), respectively.

Based on the availability of genomic resources such as markers, genetic maps and transcriptomic resources, QTL identification or gene cloning projects have been approached in several less-studied crops for both biotic and abiotic stress (Table 1). These include aluminium toxicity tolerance in sorghum [72], virus resistance in cassava [73], and drought tolerance in cowpea, groundnut, chickpea and common bean (www.generationcp.org/gcptli/) [16]. While cloned genes (e.g. for aluminium tolerance in sorghum [72]) are being used for mining the superior

alleles in African germplasm, the major QTLs identified for root traits in chickpea are being used in marker-assisted backcrossing (MABC) programmes in elite cultivars of India and Africa. By using the ICC 4958 genotype as a donor genotype, and JG 11 as a recurrent genotype, MABC has led to development of BC₃F₃ lines at ICRISAT. After multiplication of seeds of these lines, agronomic performance will be evaluated in targeted environments. In the case of cassava, the MAS approach has already been successfully utilized to develop 14 genotypes that combine cassava mosaic disease (CMD) resistance and high yields by using eight markers associated with the dominant *CMD2* gene for CMD [73]. Similarly, MAS efforts in pearl millet have led to the development of an “HHB 67 Improved” hybrid for downy mildew (DM)-resistance [74]; this study has demonstrated that, while conventional backcross transfer of DM resistance to improve parental lines took nearly nine years, a marker-based approach facilitated it in just three years.

A bright future for genomics-assisted breeding in the developing world

Most of the key developing country crops now find themselves in a situation in which genomic resources are sufficient to support meaningful genetic studies and molecular breeding [55]. Several successful genomics-assisted breeding programmes have been built through collaborations between CGIAR institutes and NARS partners [73–76]. Various bottlenecks still impede adoption in these countries; limited human resources, inadequate field infrastructure and limited capacity in information management remain major challenges. The magnitude of these challenges is exacerbated where it is important to breed for biotic (pests or diseases) and abiotic (drought, heat, cold or salinity) stresses, thus making accurate phenotyping challenging. On a positive note, the history of formal breeding in developing countries has been short, thus allowing for greater potential for improvement relative to genetic gains obtained in developed countries. However, through virtual platforms, such as the Integrated Breeding Platform (<http://mbp.generationcp.org/>) initiated in 2009 and coordinated by GCP in collaboration with the Bill & Melinda Gates Foundation, further aided by the information and communication technology revolution, breeders in developing countries will soon have better access to genomic resources, advanced laboratory services, and robust analytical and data management tools. The task remains challenging as the Platform is intended to serve a broad range of users who are working on different crops for different environments [55].

Although few economic analyses have been conducted to assess the potential impacts of molecular breeding in developing countries, there are some well-known examples of success. One of them is the development of submergence-tolerant rice cultivars through MABC, which has improved yields in >15 million hectares of rain-fed lowland rice in South and Southeast Asia [77]. A recent *ex ante* economic study of molecular breeding of rice for tolerance to salty and low-phosphorus soils in selected Asian countries has estimated that the method saves a minimum of 2–3 years, which results in significant incremental benefits in the range of \$300–800 million USD [78].

Large-scale genomic resources and specialized genetic stocks that have become available in tier 2 and 3 crops, which have been less-studied crops until recently, are expected to enhance molecular breeding such as MABC and marker-assisted recurrent selection (MARS) [55]. This will lead to enhanced crop productivity and, in turn, increased food security in developing countries. It is, however, unrealistic to project that large-scale molecular breeding activities will be conducted in the near-term in a broad set of developing countries, and the impact of this approach will be hardly homogeneous in its implementation [55]. The prospects are bright for breeders in some developing countries, particularly the newly industrialized ones with sustainable breeding programmes that benefit from institutional support. These countries should take advantage of large international initiatives by accessing germplasm data, genetic resources, genomic tools, and methodologies that will allow them to conduct efficient genetic studies and molecular breeding.

Acknowledgements

The authors are thankful to the entire GCP community for their hard work that led to the development of genomic resources and specialized genetic stocks mentioned in this article. Special thanks are due to Matthew Blair (CIAT, Columbia), Tim Close (University of California-Riverside, USA), Doug Cook (University of California-Davis, USA), Pooran Gaur (ICRISAT, India), Marc Ghislain (CIP, Peru), Tom Hash (ICRISAT, India), Sarah Hearne (International Institute of Tropical Agriculture, Nigeria), Andrzej Kilian (DART Pty Ltd, Australia), Jean-Francois Rami (CIRAD, France), Mathias Lorieux (CIAT, Colombia), Andrew Paterson (University of Georgia, USA), Pablo Rabinowicz (University of Maryland, USA), Roland Schaffeitner (CIP, Peru) and Hari Upadhyaya (ICRISAT, India) to share useful information presented in this review. Thanks are also due to Robert Koebner (smartEnglish, UK), and Abhishek Bohra, Anuja Dubey and Priyanka Puvvada from ICRISAT for their help while preparing and editing the MS.

References

- Godfray, H.C.J. *et al.* (2010) Food security, The challenge of feeding 9 billion people. *Science* 327, 812–818
- Alston, J.M. *et al.* (2009) Agricultural research, productivity, and food prices in the long run. *Science* 325, 1209–1210
- Tester, M. and Langridge, P. (2010) Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822
- Varshney, R.K. and Dubey, A. (2009) Novel genomic tools and modern genetic and breeding approaches for crop improvement. *J. Plant Biochem. Biotechnol.* 18, 127–138
- Ribaut, J.-M. *et al.* (2008) International programs and the use of modern biotechnologies for crop improvement. In *Genomics of Tropical Crop Plants* (Moore, P. and Ming, R., eds), pp. 21–63, Springer
- Glazmann, J.C. *et al.* (2010) Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Biol.* 13, 167–173
- Zamir, D. (2001) Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* 2, 983–989
- Koebner, R.M.D. and Varshney, R.K. (2006) Development and application of genomic models for large crop plant genomes. In *Model Plants and Crop Improvement* (Varshney, R.K. and Koebner, R.M.D., eds), pp. 1–9, CRC Press
- Jones, N. *et al.* (2009) Markers and mapping revisited, finding your gene. *New Phytol.* 183, 935–966
- Gupta, P.K. and Varshney, R.K. (2000) The development and use of microsatellite markers for genetics and plant breeding with emphasis on bread wheat. *Euphytica* 113, 163–185
- Kilian, A. *et al.* (2005) The fast and the cheap, SNP and DArT-based whole genome profiling for crop improvement. In *Proceedings of the International Congress: In the Wake of the Double Helix, from the Green Revolution to the Gene Revolution* (Tuberosa, R., *et al.*, eds) pp. 443–461, Avenue Media

- 12 Gupta, P.K. *et al.* (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101, 5–18
- 13 Varshney, R.K. (2010) Gene based marker systems in plants, high throughput approaches for marker discovery and genotyping. In *Molecular Techniques in Crop Improvement* (Jain, S.M. and Brar, D.S., eds), pp. 119–140, Springer
- 14 Varshney, R.K. *et al.* (2006) Advances in cereal genomics and applications in crop breeding. *Trends Biotechnol.* 24, 490–499
- 15 Nayak, S.N. *et al.* (2010) Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor. Appl. Genet.* 120, 1415–1441
- 16 Varshney, R.K. *et al.* (2009) Orphan legume crops enter the genomics era! *Curr. Opin. Plant Biol.* 12, 202–210
- 17 Blair, M.W. *et al.* (2009) Development of microsatellite markers for common bean (*Phaseolus vulgaris* L.) based on screening of non-enriched, small-insert genomic libraries. *Genome* 52, 772–782
- 18 Moretzsohn, M.C. *et al.* (2005) A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). *Theor. Appl. Genet.* 111, 1060–1071
- 19 Cuc, L.M. *et al.* (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). *BMC Plant Biol.* 8, 55
- 20 Varshney, R.K. *et al.* (2010) Pigeonpea genomics initiative (PGI), an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Mol. Breed.* DOI: 10.1007/s11032-009-r-r9327-2
- 21 Yang, S. *et al.* (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. *Theor. Appl. Genet.* 113, 585–595
- 22 Wenzl, P. *et al.* (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci. USA* 101, 9915–9920
- 23 Mace, E.S. *et al.* (2008) DArT markers, diversity analyses and mapping in *Sorghum bicolor*. *BMC Genomics* 9, 26
- 24 Hippolyte, I. *et al.* (2010) A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangement among bananas. *BMC Plant Biol.* 10, 65
- 25 Grewal, T.S. *et al.* (2008) Mapping quantitative trait loci associated with barley net blotch resistance. *Theor. Appl. Genet.* 116, 529–539
- 26 Matsumoto, T. *et al.* (2008) Development in rice genome research based on accurate genome sequence. *Int. J. Plant Genomics*, Article ID 348621
- 27 Cone, K.C. *et al.* (2002) Genetic, physical, and informatics resources for maize. On the road to an integrated map. *Plant Physiol.* 130, 1598–1605
- 28 Gupta, P.K. *et al.* (2008) Wheat genomics, present status and future prospects. *Int. J. Plant Genomics*, Article ID 896451
- 29 Sreenivasulu, N. *et al.* (2008) Barley genomics, an overview. *Int. J. Plant Genomics*, Article ID 486258
- 30 Paterson, A.H. *et al.* (2008) Genomics of sorghum. *Int. J. Plant Genomics*, Article ID 362451
- 31 Lightfoot, D. (2008) Soybean genomics, developments through the use of cultivar “Forrest”. *Int. J. Plant Genomics*, Article ID, 793158
- 32 Schlueter, J.A. *et al.* (2008) BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. *Trop. Plant Biol.* 1, 40–48
- 33 Bryan, G.J. and Hein, I. (2008) Genomics resources and tools for gene function analysis in potato. *Int. J. Plant Genomics*, Article ID 216513
- 34 Luo, M.-C. *et al.* (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the SNAPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82, 378–389
- 35 Ma, Y. *et al.* (2009) Genomic resources of cowpea (*Vigna unguiculata*). In *Plant and Animal Genome Conference XVII*, P393. (www.intl-pag.org/17/abstracts/P05f_PAGXVII_393.html)
- 36 Luo, M.C. *et al.* (2009) A BAC based fingerprint map of the cassava genome. In *Plant and Animal Genome Conference XVII*, P158. (www.intl-pag.org/17/abstracts/W21_PAGXVII_158.html)
- 37 Yuksel, B. and Paterson, A.H. (2005) Construction and characterization of peanut HindIII BAC library. *Theor. Appl. Genet.* 111, 630–639
- 38 Guimaraes, P. *et al.* (2008) BAC libraries construction from the ancestral diploid genomes of the allotetraploid cultivated peanut. *BMC Plant Biol.* 8, 14
- 39 Varshney, R.K. *et al.* (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530
- 40 Deschamps, S. *et al.* (2010) Utilization of next-generation sequence platforms in plant genomics and genetic variant discovery. *Mol. Breed.* 25, 553–570
- 41 Rounsley, S. *et al.* (2009) *De Novo* next generation sequencing of plant genomes. *Rice* 2, 35–43
- 42 Huang, S. (2009) The genome of the cucumber, *Cucumis sativus* L. *Nature Genet.* 41, 1275–1281
- 43 Muchero, W. *et al.* (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. *Proc. Natl. Acad. Sci. U. S. A.* 106, 18159–21816
- 44 Varshney, R.K. *et al.* (2009) A comprehensive resource of drought- and salinity responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). *BMC Genomics* 10, 523
- 45 Raju, N.L. *et al.* (2010) The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). *BMC Plant Biol.* 10, 45
- 46 Rajaram, V. *et al.* (2010) Development of EST resources in pearl millet and their use in development and mapping of EST-SSRs in four RIL populations. In *Plant and Animal Genome Conference XVIII*, P. 373. (www.intl-pag.org/18/abstracts/P05d_PAGXVIII_373.html)
- 47 Schafleitner, R. *et al.* (2010) 454 pyrosequencing of sweetpotato (*Ipomoea batatas*) transcriptome. In *Plant and Animal Genome Conference XVIII*, P. 23. (www.intl-pag.org/18/abstracts/P01_PAGXVIII_023.html)
- 48 McNally, K.L. *et al.* (2009) Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12273–12278
- 49 Yan, J. *et al.* (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* 4, e8451
- 50 Gore, M.A. *et al.* (2009) A first-generation haplotype map of maize. *Science* 326, 1115–1117
- 51 Akhunov, E. *et al.* (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina Golden Gate assay. *Theor. Appl. Genet.* 119, 507–517
- 52 Close, T.J. *et al.* (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10, 582
- 53 Hyten, D.L. *et al.* (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theor. Appl. Genet.* 116, 945–952
- 54 Penmetta, R.V. *et al.* (2010) Studies of genome evolution in papilionoid legumes. In *Plant and Animal Genome Conference XVIII*, W178. (www.intl-pag.org/18/abstracts/W24_PAGXVIII_178.html)
- 55 Ribaut, J.M. *et al.* (2010) Molecular breeding in developing countries, challenges and perspectives. *Curr. Opin. Plant Biol.* 13, 213–218
- 56 Goff, S.A. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100
- 57 Yu, J. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92
- 58 International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800
- 59 Paterson, A.H. *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556
- 60 Schnable, P.S. *et al.* (2009) The B73 maize genome, complexity, diversity, and dynamics. *Science* 326, 1112–1115
- 61 Paux, E. *et al.* (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322, 101–104
- 62 Mayer, K.F.X. *et al.* (2009) Gene content and virtual gene order of barley chromosome 1H. *Plant Physiol.* 151, 496–505
- 63 Timko, M.P. *et al.* (2008) Sequencing and analysis of the gene-rich space of cowpea. *BMC Genomics* 9, 103
- 64 Brown, A.H.D. (1989) Core collection, a practical approach to genetic resources management. *Genome* 31, 818–824
- 65 The CGIAR Generation Challenge Programme at 6 (2010) From discovery to application and impact in crop improvement. Brochure of Generation Challenge Programme, Mexico. (www.generationcp.org/UserFiles2/File/Research/Transit-strategy/Impacts-Phase-I-GCP.pdf)
- 66 Upadhyaya, H.D. *et al.* (2008) Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* 8, 106–138

- 67 Gutierrez, A.G. *et al.* (2010) Identification of a Rice stripe necrosis virus resistance locus and yield component QTLs using *Oryza sativa* × *O. glaberrima* introgression lines. *BMC Plant Biol.* 10, 6
- 68 Fonceka, D. *et al.* (2009) Genetic mapping of wild introgressions into cultivated peanut, a way toward enlarging the genetic basis of a recent allotetraploid. *BMC Plant Biol.* 9, 103
- 69 Ribaut, J.-M. and Hoisington, D.A. (1998) Marker-assisted selection, new tools and strategies. *Trends Plant Sci.* 3, 236–239
- 70 Moretzsohn, M.C. *et al.* (2009) A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. *BMC Plant Biol.* 9, 40
- 71 Varshney, R.K. *et al.* (2009) The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 118, 729–739
- 72 Magalhaes, J.V. *et al.* (2007) A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nature Genet.* 39, 1156–1161
- 73 Okogbenin, E. *et al.* (2007) Marker-assisted introgression of resistance to cassava mosaic disease into latin American germplasm for the genetic improvement of cassava in Africa. *Crop. Sci.* 47, 1895–1904
- 74 Hash, C.T. *et al.* (2006) Teamwork delivers biotechnology products to Indian small-holder crop-livestock producers, pearl millet hybrid “HHB 67 improved” enters seed delivery pipeline. *SAT eJournal, ejournal.icrisat.org*, 2, p 3
- 75 William, H.M. *et al.* (2007) Wheat breeding assisted by markers, CIMMYT’s experience. *Euphytica* 157, 307–319
- 76 Thomson, M.J. *et al.* (2010) Marker assisted breeding. In *Abiotic Stress Adaptation in Plants: Physiological, Molecular and Genomic Foundation* (Pareek, A. *et al.*, eds), pp. 451–469, Springer
- 77 Septiningsih, E.M. *et al.* (2009) Development of submergence-tolerant rice cultivars: the *Sub1* locus and beyond. *Ann. Bot.* 103, 151–160
- 78 Alpuerto, V.E.B. *et al.* (2009) Economic impact analysis of marker-assisted breeding for tolerance to salinity and phosphorous deficiency in rice. *Rev. Agr. Econ.* 31, 779–792