Analytical and Decision Support Tools for Genomics-Assisted Breeding

Citation for published version:

Digital Object Identifier (DOI):
10.1016/j.tplants.2015.10.018

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Trends in plant science

Publisher Rights Statement:
© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Review

Analytical and Decision Support Tools for Genomics-Assisted Breeding

Rajeev K. Varshney,1,2,*, Vikas K. Singh,1 John M. Hickey,3 Xu Xun,4 David F. Marshall,5 Jun Wang,4 David Edwards,2 and Jean-Marcel Ribaut6

To successfully implement genomics-assisted breeding (GAB) in crop improvement programs, efficient and effective analytical and decision support tools (ADSTs) are 'must haves' to evaluate and select plants for developing next-generation crops. Here we review the applications and deployment of appropriate ADSTs for GAB, in the context of next-generation sequencing (NGS), an emerging source of massive genomic information. We discuss suitable software tools and pipelines for marker-based approaches (markers/haplotypes), including large-scale genotypic and phenotypic, data management, and molecular breeding approaches. Although phenotyping remains expensive and time consuming, prediction of allelic effects on phenotypes opens new doors to enhance genetic gain across crop cycles, building on reliable phenotyping approaches and good crop information systems, including pedigree information and target haplotypes.

Breeding for Sustainable Food Production

GAB (see Glossary) has become popular for crop improvement in recent years partly due to availability of low-cost high-throughput genotyping (HTPG) and NGS technologies. Several successful examples of GAB are now available not only in major crop species but also in many so-called 'orphan crops' [1,2]. GAB pipelines involve various steps including: characterization of diverse germplasm collections; development of mapping populations; identification of genomic regions through genetic or association mapping; and application of markers in breeding. Numerous ADSTs are required throughout all four of these steps [3]. Better understanding of the genetic diversity that is present in germplasm collections in gene banks and breeding material helps breeders identify new valuable alleles for breeding. Field evaluation of large germplasm collections is challenging due to, for example, poor genetic background, variation in phenology, the logistics and resources required, and selection of smaller subsets that represent the diversity of the collection. These sets include 'core collections' (10% of the entire collection) [4], 'mini-core collections' (about 10% of the core collection or 1% of the entire collection) [5], and 'reference sets' (usually developed based on the molecular characterization of a composite collection) [6]. Efforts to define these sets should also benefit from the use of ADSTs on germplasm collections.

For trait mapping, two complementary approaches – namely, linkage mapping and association mapping, which in the context of large-scale genotyping and the whole-genome re-sequencing era are now referred to as genome-wide association studies (GWAS) – have been used in crop genetics. Construction of high-quality genetic maps with precise marker orders is critical when undertaking quantitative trait locus (QTL) analysis, which leads to the identification of genomic
regions and markers associated with target traits. Association mapping or GWAS has emerged as a new approach for the identification of causal locigenes for traits of interest and some tools have been developed in recent years. Availability of the re-sequencing data of multiple accessions of the same species or different species has initiated the concept of the pan-genome. The generated hapmap information through pan-genome analysis is useful for the construction of high-density linkage maps. Pan-genomes are useful for the collection of all the genes at clad level. Additionally, with re-sequencing-based mapping of populations now being possible, new approaches for trait mapping using two contrasting bulks for the given traits have also been used.

Once molecular markers linked with traits are identified, they can be used for marker-assisted back crossing (MABC) or marker-assisted selection (MAS) programs [7]. ADSTs can be helpful in selecting superior lines based on foreground and background selection for the next crossing. The other two approaches of GAB also require ADSTs, specifically marker-assisted recurrent selection (MARS), which enables the accumulation of superior alleles from different genetic backgrounds to one background, and genomic selection (GS), which enables enhancing genetic gain in crop breeding. Furthermore, data generated during the course of one GAB program often need to be shared with different partners to better enable future GAB programs in other institutes and countries. Therefore, ADSTs are required for the management, retrieval, and sharing of data.

In view of all of the above, it is evident that appropriate ADSTs and their integrated use at the right time in different steps of GAB is critical for the next generation of genomics and integrated breeding (see Outstanding Questions). This review discusses the need, availability, and future requirements of ADSTs for enhancing the precision and modernizing of crop breeding (Table S1 in the supplemental information online and Figure 1).

Genetic Diversity and Population Genetic Analysis

Genetic diversity estimates help to structure germplasm defining, for example, heterotic pools, and provide useful information to select contrasting parental lines for new breeding populations. Analysis of molecular marker-based estimates of genetic diversity depends on a number of criteria, such as type (dominant or codominant) of markers, number of markers and genotypes, missing data, and proportion of heterozygosity. Similarly, population genetic analysis provides estimates of the allele frequencies that are helpful to breeders because alleles are the raw material for selection in breeding programs [8]. High-density genotyping has revolutionized the identification of favorable alleles in populations, minimizing the risk of recombination between markers and target genes.

To analyze genetic diversity, the numerical taxonomy and multivariate analysis system (NTSYSpc) is one of the most widely used software tools (see Table S1 for a description of all of these tools) [9]. Molecular evolutionary genetic analysis (MEGA) is another widely used program for estimating evolutionary distances and phylogenetic trees from DNA or protein sequence data [10]. Although several other programs/tools are available, DARwin (http://darwin.cirad.fr), a free and easy to use program for diversity and multivariate analysis of datasets that also provides publication-ready figures, has emerged as a popular tool in recent years for genetic diversity analysis. Due to the capacity to generate millions of SNPs in germplasm collections, construction of phylogenetic trees is an increasingly computationally challenging task. In this context, new pipelines such as SNPhylo [11] are being developed. SNPhylo works by selecting one informative SNP from each linkage disequilibrium (LD) block, thereby greatly decreasing the running time without losing much information.

For population genetics analysis, Arlequin is a highly used software package for molecular variance (AMOVA) analysis of datasets that includes several statistics like diversity, genetic
distance, equilibrium analysis, and neutrality tests [12]. DNA Sequence Polymorphism (DnaSP) utilizes DNA sequence data and estimates several measures of DNA sequence variation within and between populations including LD, recombination, gene flow, and gene conversion parameters and can perform several tests of neutrality [13]. GenAIEx, which is based on Microsoft Excel, offers a wide range of population genetic analysis options for the full spectrum of genetic markers with rich graphical outputs for data exploration and publication [14]. Several other software tools have become available in recent years for various applications. For example, Power Marker is useful for simple sequence repeat (SSR) or SNP marker datasets for population genetic analysis, quantitative trait locus (QTL) mapping, molecular breeding, sampling, integrated pipelines, sequencing-based mapping, genetic diversity, and hapmaps. Different ADSTs can be selected based on their suitability for the experiment along with the strength of the tools. The details of individual ADSTs are presented in Table S1 in the supplemental information online.

Based on sequence/marker diversity analysis on a large scale, a core set of germplasm, also called a reference set [16], can be developed. Reference sets seem to be better than core collections (comprising ~10% of the entire collection [4]) and mini-core collections (comprising ~10% of the total core collection or 1% of the total collection [5]) for undertaking GWAS, as discussed below), as they have lower structural components than the full germplasm sets. Furthermore, the concept of selective phenotyping is also increasingly popular for selecting the subsets of mapping populations. This type of mapping is often done using recombination breakpoints to eliminate the need to extensively phenotype large numbers of individuals [17]. This is important in the case of populations like multiparent advanced generation intercross (MAGIC) [a population developed by crossing multiple founder lines (four or eight) to improve the precision and resolution of QTL mapping] where large numbers of lines are available and genotyping can be done in a high-throughput manner, but phenotyping of such large number of lines is challenging.
For a selection of genotypes for core or mini-core formation, PowerCore is a widely used software package. It was developed based on advanced Maximization (M) strategy with a heuristic search for establishing core sets [18]. The M strategy has been used to select specific combinations of accessions that include complete coverage and is useful for selecting entries with the most diverse alleles and eliminating redundancy. It has been suggested that before considering molecular markers datasets for the construction of core sets, the data resolution (DR) needs to be calculated using a jackknife approach to the selection of suitable marker sets [19]. However, for selection of lines (with maximal dissimilarity) from the mapping population for undertaking selective phenotyping, three main methods are available. The minimum moment aberration (MMA) method minimizes the average of all pairwise similarities between the individuals of the population. It can be utilized in selecting F2 recombinants without any missing datasets [20]. maxRec is another statistical tool for selecting lines on the basis of higher numbers of recombination events during the course of the recombination generations [21]. This statistical package is suitable for backcross, double haploid, and recombinant inbred line (RIL) populations. SPCLUST is another program that has been developed for selecting lines from BC, F2 intercross, and complex crosses like four-way MAGIC [19]. The power of QTL detection using selected subsets using SPCLUST was similar to the power that could be achieved by using the entire dataset for analysis for QTL experiments.

Construction of Genetic Maps

Genetic maps serve as the foundation for various genetic applications, such as ordering of genes/markers, QTL mapping, association mapping, and map-based cloning [22]. Genetic maps are useful for anchoring scaffolds to linkage groups as well as assembling (and sometimes correcting) smaller contigs into large contigs [1]. However, construction of high-quality genetic maps depends on the following four parameters: the type of the population (e.g., biparental populations like F2, F2:3, BC, RILs, NILs, DH, multiparental mapping populations); the size of the population (100–500 lines); the number of markers (50–100 000); and the nature of the markers (SSR, DAfT, SNP). Managing all of these parameters requires skills like working on a LINUX platform as well as high-performance computing programs [23].

Construction of linkage maps for small-scale experiments with fewer markers (<500) and smaller population sizes (<200) can still be undertaken with the first-generation and most widely used mapping software tool MAPMAKER [24]. MapDraw is a simple Microsoft Excel-based free software tool that can create attractive linkage maps as well as undertaking various kinds of analysis [25]. JoinMap (https://www.kyazma.nl/index.php/JoinMap/) is a Windows-based software tool that can handle up to 50 000 markers and its key capability is to integrate data from multiple populations. This software generates high-quality publication-ready images. Recombination Counting and Ordering (Record) is a statistical tool that can be utilized for ordering marker loci on genetic maps [26]. Recently, an ultrafast pipeline, namely SimpleMap (http://simplemap-aj.sourceforge.net/), was streamlined for the construction of high-density linkage maps. This pipeline can develop linkage maps with ~1000 loci in <10 min, compared with >8–10 h using other programs.

Currently, genotype data is becoming available for 50 000 to 100 000 marker loci. For such marker densities, MSTMap has been developed and works on a minimum spanning tree (MST)-based method (http://www.mstmap.org/). The MST algorithm uses well-established graph theory and provides an efficient solution to the generation of genetic maps using large numbers of markers and individuals. MSTMap outperforms other mapping programs when the input data are noisy or incomplete. To manage large-scale re-sequencing data on the population, the sequencing enabled genotyping based map (SEG-Map) has also been developed for the construction of linkage maps [27]. This software allows the mapping of short reads generated for progeny into pseudomolecules of the parents of the mapping population, which in turn precisely, GAB is the application of various genetic and genomics tools to develop new breeding lines.

Genomic-estimated breeding values (GEBVs): estimated breeding values generated through genotyping of populations using statistical model(s) and used to select superior individuals in a segregating population.

Genomic selection (GS): is a new method of molecular breeding in which selection of lines is based on GEBVs calculated based on genome-wide markers. GEBVs can be estimated through genotyping and phenotyping of a training population.

High-throughput genotyping (HTPG): a powerful and efficient method for rapid analysis of DNA sequence variations among large number of samples using the most advanced techniques, thereby generating a huge set of datasets that can be analyzed to understand nucleotide variations.

Linkage disequilibrium (LD): nonrandom association between two markers, genes or, QTLs on the same chromosome in a population owing to their tendency to be co-inherited. When variants of two genetic loci are in LD, the variant seen at one locus predicts the variant found at the other.

Linkage drag: the carry-forward of any unwanted genes/loci along with the trait of interest from a donor parent during a backcross breeding program that might reduce the agronomic character of the elite cultivar.

Marker-assisted back crossing (MABC): the breeding method for introgression of major effect loci (two to four) in an elite genetic background through marker-aided foreground selection (selection of plants with the desired alleles from the donor parent) and supplemented with background (selection of plants with higher recurrent parent genome) in a rapid and precise manner.

Marker-assisted recurrent selection (MARS): a marker-based breeding process used to identify and monitor key regions (up to 20 or more) from both of the superior parents for complex traits in consecutive breeding generations.

Mini-core collection: a limited set of accessions (about 10% of a core collection or 1% of the entire collection) without losing much genetic diversity. The smaller size of
enables detection of SNPs. These SNPs can be used to identify recombination breakpoints and for bin map construction. The output data of SEG-Map can be directly used for QTL mapping studies.

For a given species, several genetic maps have sometimes been developed using various mapping populations. As a result, no single genetic map has a marker order for all markers available in that crop, and sometimes maps from different populations are different. Consensus genetic maps based on multiple biparental mapping populations are, therefore, an important resource for providing order for large numbers of marker loci for a given species. These maps are useful for analyzing LD as well as for association analysis and fine mapping of QTLs. Based on the availability of the common markers mapped from different mapping populations, consensus maps have been generated in many crops using the JoinMap program [28–31]. Recently, LPmerge, a new R-based package, has also been developed to construct consensus maps, with a major focus on marker orders to remove and resolve the conflicts in consensus maps [32]. Programs like JoinMap, MSTMap, and SEG-Map are increasingly common for the construction of high-density maps. Similarly, JoinMap or LPmerge will be useful for the development of consensus maps from different mapping populations. At present, medium numbers of markers (200–500) are being used for linkage map development in wheat. However, in the future, with the advent of sequencing-based trait mapping, current programs/methods for the development of high-density linkage maps may become obsolete. With the rapid development of sequencing technologies and the possibility to sequence hundreds/thousands of accessions at species or even genus level, a pan-genome for the species/genus can be developed. Such pan-genomes have already been developed in some crops like maize [33], rice [34], and soybean [35]. The hapmap information coming from these pan-genomes should serve as the foundation for the construction of ‘universal maps’ for given species/genera.

**Linkage-Mapping Based QTL Analysis**

QTL mapping, in general, uses one of following approaches: single marker analysis (SMA), simple interval mapping (SIM), or composite interval mapping (CIM). However, it can be further extended in terms of estimating epistatic and environmental interactions [36,37]. Most QTL mapping tools have been developed for biparental mapping populations (Table S1). However, in recent years some sophisticated tools have been developed for multiparent mapping populations, like MAGIC and nested association mapping (NAM) populations.

Although a range of QTL analysis programs are available, QGene [38], MapManager QTX (http://rubio.bio.indiana.edu/soft/molbio/mac/map-manager-readme.html), and MapManager/QTL (http://www.broadinstitute.org/ftp/distribution/software/mapmaker3/) are the appropriate software tools for SMA. For CIM, WinQTL Cartographer (http://statgen.ncsu.edu/qtlcart/ WQTLCart.htm), MapQTL (https://www.kyazma.nl/index.php/mc.MapQTL), and PLABQTL [39] have been shown to be appropriate software [40]. Inclusive CIM (ICIM) [41] and QTLNetwork [42] are other commonly used programs for QTL mapping.

To analyze marker–trait associations (MTAs) and to finely map genetic regions in multiparent mapping populations of outbred animal stocks, the specialized software package HAPPY was developed [43]. However, in the case of plant species, R-based packages such as R/qtl, R/ricalc, R/mpMap, and R/mpwgaim have been used to map genomic regions [44–46]. For analyzing multiparent mapping populations like NAM populations, an integrated software tool called IciMapping (see detailed description later) has been developed to identify the genomic regions responsible for the trait of interest [41]. As many QTL analysis studies are based on different populations with phenotyping data from different environments, many researchers have started to undertake meta-QTL analysis to understand the genetic determination of complex traits. This approach is also useful for the identification of robust QTLs, which can be subjected
for fine mapping and ultimately useful for the identification of candidate genes. To perform meta-
analysis of QTLs, MetaQTL [47] and BioMercator [48] are promising software packages. 
Additionally, to develop linkage maps and to project QTLs, several other packages have become 
available recently, including MapChart [49], MQ^2 [50], and R/qtlcharts [51].

**GWAS**

In the case of GWAS, understanding population structure and the level and distribution of LD in 
the populations is a prerequisite for using the appropriate approach of association mapping. In 
this context, STRUCTURE [52] is the most extensively used software to detect population 
genetic structure. STRUCTURE generates clusters based on both transient Hardy–Weinberg 
disequilibrium and LD caused by admixture between populations [53,54]. EIGENSOFT is 
another widely used statistical package for the detection and correction of population stratifi-
cation in GWAS using principal component analysis [55]. Similarly, Bayesian analysis of popu-
lation structure (BAPS) is another program for Bayesian inference of the genetic structure, 
especially for analyzing large-scale population genetics data in a population [56]. Furthermore, 
for analysis of re-sequencing data in terms of LD and haplotype block analysis, haplotype 
population frequency estimation, single SNP and haplotype association tests, and permutation 
tests for association significance, SNP analyzer 2.0 has been developed [57]. A detailed list of 
other available software for LD analysis can be found at http://www.genes.org.uk/software/
LD-software.shtml.

For performing association analysis, Trait Analysis by aSSociation, Evolution, and Linkage 
(TASSEL) is the most commonly used and highly cited software in GWAS in plants [58]. This 
software provides several new and powerful statistical approaches for association mapping 
such as the General Linear Model (GLM) and Mixed Linear Model (MLM) [59]. GenABEL 
(http://www.genabel.org/manuals/GenABEL) is a genome-wide SNP-association analysis 
program based on R. PLINK is another highly cited open-source software for whole-genome 
association analysis. This program is designed to perform a range of basic and large-scale 
analyses [60]. PLINK focuses on the analysis of genotype/phenotype data to perform the 
association analysis.

Most association mapping analyses have been conducted based on GLM or MLM, which does 
not seem sufficient for the identification of robust MTAs. Therefore, in the near future, models 
such as multiloci mixed models (MLMMs) and multitrait mixed models (MTMMs) need to 
become more common. To confirm the association of SNPs identified from the GWAS with the 
target traits, nonsynonymous SNP (nsSNP)-based association mapping is one of the most 
promising approaches [61]. Additionally, with the increasing use of small Indels for MTAs, the 
ADSTs available at present need to be modified in such a way that they can accommodate SNPs 
as well as small Indels for performing trait association analysis. Recently, NGS-based trait 
mapping approaches and analysis of re-sequencing data were found promising for the identifi-
cation of target genomic regions. In this context, large numbers of scripts/software were 
developed that could be deployed in NGS-based studies, including haplotype-based GWAS 
(Box 1). For better utilization of GWAS results, the identified MTAs in various crops should be 
made available as open-access databases for the selection and deployment of the most robust 
alleles in crop improvement programs.

**Molecular Breeding**

Significant progress has been achieved in the area of molecular breeding in developing improved 
plant varieties [62,63]. Among various GAB approaches, MAS/MABC has been used extensively 
in public breeding programs. MAS/MABC, in general, do not use any sophisticated tools to 
select plants for advancement or backcrossing. However, open-access visualization tools such 
as Graphical Genotypes (GGT) [64], Flapjack [65], and the molecular breeding design tool
**Box 1. Genomic Tools for Sequencing-Based Mapping and Re-sequencing Analysis**

NGS-based mapping approaches using **bulk segregant analysis (BSA)** have been used for mapping target genomic regions without the construction of linkage maps. However, these approaches require specialized skills and tools. Some approaches have been developed for facilitating trait mapping using NGS approaches. For instance, the ShoreMap approach (simultaneous mapping and mutant identification by deep sequencing) was developed to map the target genes in mutant lines [76]. This software package, available at http://1001genomes.org/software/shoremap.html, is open source and is continuously updated. Similarly, the next-generation mapping (NGM) (http://bar.utoronto.ca/ngm/) pipeline was proposed and developed for trait mapping [77].

The MutMap [78] and QTL-seq [79] approaches facilitate the mapping of targeted genomic regions from EMS-derived mutants and from any desirable genotype, respectively. To perform either of these two analyses, specific bioinformatics pipelines are available at http://genome-e.ibrc.or.jp/home/bioinformatics-team/mutmap. CloudMap (http://usegalaxy.org/cloudmap) is another open-source web-based analytical bioinformatics pipeline for the identification of candidate genes directly from EMS-derived mutants without the development of a mapping population [80].

Re-sequencing of numbers of lines from different crop species opens new avenues and is useful for understanding the evolution of and genetic relationships among individuals. Therefore, for analyzing re-sequencing datasets in terms of haplotypes, construction of hapmaps, haplotype population frequency estimation, single SNP and haplotype association tests, and permutation tests for association significance, Haploviz is a promising tool. Haploviz can analyze thousands of SNPs (tens of thousands in command-line mode) in thousands of individuals [81]. SHAPEIT [82], fastPHASE [83], WHAP [84], and HaploBlockFinder [85] are some other important analytical tools/pipelines for the development of hapmaps/haplotypes and performing GWAS.

---

(MBDT) (https://www.integratedbreeding.net/179/training/bms-user-manual(marker-assisted-backcross-breeding-tool) have become available in recent years for the selection of plants with maximum recurrent parent genome recovery at the global level to eliminate the precise linkage drag on carrier chromosomes. Another data visualization and selection tool called CSSL Finder is useful for developing chromosome segment substitution lines (CSSLs) [66]. This is a useful tool to search a population of advanced backcross lines for a set of lines with the optimized representation of the donor parent genome in the recurrent parent background. This software, in conjunction with a graphical genotype, also displays the phenotypic values of the individual lines. Therefore, this program is useful for the identification of elite/novel CSSLs responsible for a trait of interest.

For MARS, OptiMAS has been developed by the French Agricultural Research Centre for International Development (CIRAD) as a part of the Integrated Breeding Platform (IBP). This software helps in selecting plants possessing superior alleles from elite parents in several cycles of recombination [67]. GS is a new molecular breeding approach that integrates marker data and phenotypic data from a training population to generate a prediction model for predicting genomic-estimated breeding values (GBEVs) for all segregating individuals of a breeding population. Calculation of GEBV requires specific statistical models that treat markers as random effects. The most commonly used GS prediction models are the Random Regression Best Linear Unbiased Predictor (RR-BLUP) [68], BayesA [68], BayesB [68], BayesCπ [68], Bayesian Ridge Regression (RR) [69], Bayesian LASSO [70,71], and Random Forest Regression (RFR) [72]. However, no single statistical model has emerged as being clearly better than the others for all applications. For some applications, it is possible to select the most suitable model after testing several alternative models. In this context, soiGS, a web-based tool for GS based on the RR-BLUP model, has been developed [73]. This software is an easy-to-use analysis platform for performing GS in plant breeding. Similarly, ISMU 2.0 is being developed by ICRISAT, with the close collaboration of several leading institutions. ISMU 2.0, which is an improved version of ISMU 1.0 [74], has several data processing capabilities including several models of GS. This pipeline includes most of the GS models, including RR-BLUP, Kinship Gauss, RR, Bayesian LASSO, BayesA, BayesB, BayesCπ, and RF and works on Windows, CentOS, and Ubuntu platforms [75]. Thus, ISMU 2.0 will be useful for the breeding community to analyze large-scale datasets for GS experiments for enhancing genetic gains (Box 2).
Strategic Outlook on the Future Prospects

While significant advances have been made in the areas of genomics and GAB, further efforts need to be made to develop ADSTs and crop information systems. In the area of data management for crop breeding (storage, curation, analysis, and publication) one size clearly does not fit all, so there is an increasing need to better integrate software tools and develop interoperable application program interfaces (APIs) to facilitate access to diverse tools and databases across different pipelines. In addition to these analytical and bioinformatics needs, we identify here four other areas that could be addressed to improve the efficiency of GAB: (i) further reduction in genotyping costs per line so that genome-wide marker profile data can be generated on large populations in routine breeding activities for enhancing/fixing favorable alleles; (ii) reduction in field-relevant phenotyping costs and implementation of new high-throughput screening methods such as aerial infrared screening; (iii) adoption of best, or at least good, data management practices, starting with adequate resource allocation and implementation of a data management policy at institute level through the joint efforts of management and the donor community; and (iv) a sustainable adoption of ADSTs and associated programs/tools that goes beyond just technological development to include training and suitable support services for breeders.

Deployment of ADSTs and crop information systems must be considered carefully. Some of these issues need to be addressed through an integrated approach and one should not underestimate the difficulty related to technology transfer in the public sector. Local support, such as that provided by the IBP through regional hubs, can be an attractive option to enhance the use of modern breeding tools and services. This would be mainly through capacity building, technical support, and crop-specific expertise.

We are hopeful that the development and deployment of the right ADSTs at the right time, in keeping with the needs, resources, and technical readiness of breeding programs, will usher...
crop improvement programmes into a modern, knowledge-based crop improvement era, leading to sustainable crop production and global food security.

Acknowledgments
R.K.V. thanks the CGIAR Generation Challenge Programme (GCP), the US Agency for International Development (USAID), and the Department of Biotechnology, Government of India for sponsoring research at ICRISAT on the topics mentioned in the review. This study has been undertaken as part of the CGIAR Research Program on Grain Legumes. ICRISAT is a member of the CGIAR Consortium.

Supplemental Information
Supplemental information associated with this article can be found, in the online version, at doi:10.1016/j.tplants.2015.10.018.

References
2. Varshney, R.K. et al. (2014) Marker assisted backcrossing to introgress resistance to Fusarium wilt (Fw) race 1 and Ascochyta blight (Ab) in C 214, an elite cultivar of chickpea. Plant Genome 7, 11
11. Lee, T.H. et al. (2014) SNPpyro: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics 15, 162

Outstanding Questions
What are the major ADSTs with specific features that are available for plant breeding?
Will the construction of genetic maps be obsolete and/or outdated in the context of generating millions of data points on segregating populations?
Can open-source and one-stop integrated platforms facilitate GAB programs in developing countries?
49. de los Campos, G. et al. (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182, 375–385
58. Takagi, H. et al. (2013) QTLnet: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74, 174–183