

Molecular Marker Resources and Their Application

8

Roland Schafleitner

Abstract

Molecular markers are DNA fragments or sequence tags that are associated with a certain location of the genome of an organism. Markers have been used in mungbean to analyze the genetic diversity among germplasm accessions and cultivars and to map important traits, including resistance to pests and diseases. Early studies were performed with isoenzyme and RAPD markers. Microsatellite markers derived from various *Vigna* species were efficiently used for generating genetic maps and map traits such as bruchid resistance in segregating populations. The advent of the mungbean whole genome sequence has strongly improved the access to molecular markers for this crop. Large numbers of single-nucleotide markers have been produced by genotyping by sequencing and whole genome re-sequencing, and the generated information has been used to assess the diversity and population structure of mungbean collections and for mapping traits in segregating populations and germplasm panels. Markers for agro-morphological traits as well as for disease and pest resistance are available for marker-assisted selection in mungbean breeding programs.

8.1 Introduction—The Most Important Molecular Marker Types for Plant Science

Molecular markers can be considered to be landmarks in the genome. A molecular marker is either a DNA fragment or a DNA sequence associated with a specific genomic location. In the past, also protein markers have been used. For being useful, a molecular marker needs to show differences among the genotypes under investigation, by either sequence or fragment length. Subsequently, the most important marker types used in plant sciences are briefly described.

Restriction fragment length polymorphism (RFLP) markers were invented by AJ Jeffries in the late 1970s (Zagorski 2006) and were first applied in human genetics. In the late 1980, this technique was introduced to study genome architecture in major crops such as wheat (Sharp et al. 1989), tomato and potato (Bonierbale et al. 1988). RFLP marker detection involves digestion of genomic DNA with restriction enzymes, labeling of specific DNA fragments—usually with the radioactive isotope ^{32}P —and then using the fragments one by one as a probe in Southern blot analyses (Williams 1989). RFLP markers are usually designed to detect both alleles in a heterozygous sample. These markers were a breakthrough for genetic fingerprinting, but the technique is time consuming, needs large amounts of high-quality DNA, and involves handling of a radioactive substance, making the analysis laborious and expensive. Therefore, this

R. Schafleitner (✉)
World Vegetable Center, Shanhua, Tainan, Taiwan
e-mail: roland.schafleitner@worldveg.org

technology is not used anymore, but many published RFLP markers referring to important pest and disease resistances are still relevant to research and breeding.

The advent of polymerase chain reaction (PCR) opened the path for a number of different marker types. Random amplified polymorphic DNA (RAPD) markers rapidly have become popular, as they are easy and cheap to use on virtually any organism. RAPD applies combinations of short random primers in PCRs to amplify different genomic regions. The obtained DNA fragments are resolved according to their size by agarose or polyacrylamide electrophoresis. The banding patterns can vary among genotypes, resulting in potentially polymorphic markers. The drawbacks of RAPD markers are that they require relatively high-quality DNA, are dominant, and have low reproducibility. Consequently, they are difficult to be compared among experiments and among laboratories. In addition, the multiple bands produced in PCR by RAPD primer pairs make identification of alleles difficult. Converting RAPD markers to more robust sequence characterized amplified regions (SCAR) markers enhances reproducibility and allele identification (Bhagyawant 2016). For developing SCAR markers, polymorphic RAPD fragments are cloned and sequenced, and primers are designed to specifically amplify the polymorphic RAPD fragments. Generally, the primer pairs are made to amplify a single RAPD band (Paran and Michelmoore 1993). Polymorphism of SCAR markers is either scored as presence or absence of the amplified band, or as length polymorphisms in the case of co-dominant SCAR markers.

Amplified fragment length polymorphism (AFLP) uses genomic DNA digested with restriction enzymes that are ligated to adapters with known sequence (Vos et al. 1995). Primers complementary to the ligated adapters are used to amplify DNA fragments. Complexity reduction is achieved by adding one or a few specific bases at the 3' end of the primers to amplify only a subset of the restriction fragments. Presence and absence of specific fragments are scored after separation of the fragments according to their

size on a gel or on a fragment analyzer. DNA bands are visualized either through autoradiography or fluorescence methods. AFLP can produce a large number of markers and is at least partly amenable to automatization. Polymorphic bands can be converted to SCAR markers (see above). Today, AFLP markers are not broadly used any more.

Microsatellite markers (SSR marker) contain simple sequence repeats (SSRs) of 1–8 base pairs (Hamada and Kakunaga 1982). SSR motifs are hot spots for mutations, where DNA polymerase adds or eliminates one or more repeat units during DNA replication. SSR motifs can be present in coding and non-coding sequences. They can be amplified by PCR using flanking sequence-specific primers. Size polymorphisms of specific SSR fragments among different genotypes are scored after size fractionation by electrophoresis or on fragment analyzers. SSR markers are abundant in the genome and are generally co-dominant; however, the degree of their polymorphism varies among species and populations.

Single-nucleotide polymorphism (SNP) markers consist of single-nucleotide changes observed by comparing the DNA of different genotypes. They are very abundant in the genome, but their generation generally requires sequence information. As improvements in sequencing technology made sequence information readily available at low cost, SNP markers have become the marker species of choice.

Strictly speaking, markers like AFLP or RFLP that do not require sequence information also score (beside indels and structural variations) SNPs, but only if they are present at restriction enzyme cutting sites. Likewise, the Diversity Array Technology (DArT) also scores SNPs (and indels and structural variants) by testing for the presence or amount of a specific DNA restriction fragments in a representation derived from the total genomic DNA of different individuals or populations (Jaccoud et al. 2001). The DArT technology is fast and cost-effective, but produces dominant markers.

There are various methods to obtain SNP markers. Comparing genome or transcriptome sequences among individuals is a relatively

simple method to produce SNP information, but cost-effective genotyping of these SNPs in a large population needs specialized technologies based on fluorescence detection in PCR format (Tapp et al. 2000; Semagn et al. 2014), or arrays (Elbasyoni et al. 2018). Smaller SNP sets can be analyzed in large populations by cleaved amplified polymorphic sequence (CAPS) markers (Thiel et al. 2004), high-resolution melting (Liew et al. 2004), mass spectroscopy (Storm and Darnhofer-Patel 2003), or other methods. Dependent of the number of loci to be genotyped, establishing the SNP resource for a population may be a costly investment. Multiplexing SNP assays is an often practiced way to improve cost efficiency (Fan et al. 2003).

SNP genotyping technologies such as genotyping by sequencing (Elshire et al. 2011) and the single-primer enrichment technology (SPET) became very popular. They are based on sequencing a fraction of the genome. Complexity reduction is achieved either by restriction enzyme digestion or by multiplex PCR amplification of target sequences. Both methods can be scaled to obtain a greater or lesser number of SNPs. DNA barcoding allows for pooling of many samples in one sequencing reaction, which leads to dramatic cost reductions (Elshire et al. 2011). GBS and SPET combine SNP discovery and genotyping in one step, but the reproducibility and precision of these methods are inferior to array-based genotyping (Elbasyoni et al. 2018). Both GBS and SPET are patented technologies (patent numbers WO2013009175A1 and US20130231253A1, respectively).

8.2 Molecular Markers in Plant Breeding

Plant breeding consists of crossing the best parents and subsequent identification and recovery of the progeny that outperforms the parents (Moose and Mumm 2008). Genetic gain is defined by (i) the phenotypic variation present in the breeding population, (ii) the probability that a

trait phenotype will be transmitted from parents to offspring (heritability), (iii) the proportion of the population that is selected as parents for the next generation (selection intensity), and (iv) by the time necessary to complete a cycle of selection. All four key factors for genetic gain can be positively impacted by using molecular markers.

One factor affecting genetic gain is the available phenotypic variation in the population. Measuring the phenotypic diversity is affected by the environment. Exotic material may not be adapted to the selection environment and thus not show the real potential for breeding. Phenotypic diversity is positively associated with genetic diversity. Molecular markers are cheap and efficient tools to characterize genetic diversity in populations. They contribute to understand population structure and inform about the presence of heterotic groups in a germplasm set or in breeding populations and facilitate the exploitation of heterosis for producing hybrids and improved populations (Van Inghelandt et al. 2010; Barata and Carena 2006).

Heritability depends on the number of genes affecting a trait, the magnitude of their effects, and the type of gene action associated with the phenotype (Moose and Mumm 2008). Molecular marker technologies facilitate the definition of loci associated with a trait of interest. For traits with low heritability such as yield, molecular markers associated with loci influencing the trait often account for a greater proportion of additive genetic effects than the phenotype alone. Knowledge of the genetic architecture underlying the trait can be exploited to add or eliminate specific alleles that contribute to the breeding value. If linkage drag or epistasis among loci with antagonistic effects on a trait limits the genetic gain, information on loci associated with the traits can be used to break these undesirable allelic relationships (Moose and Mumm 2008).

Selection intensity in conventional plant breeding relies on phenotypic selection. Environmental variability, genotype by environment interaction, and evaluation errors add complexity to phenotypic selection. Multi-environment evaluation improves selection accuracy, but is

time consuming and expensive. Some traits require destructive sampling or exposure of the population to diseases and pests, which affects the recovery of the desired genotypes. Pest and disease resistance screening in natural environments is particularly challenging, as it depends on the presence and activity of the pathogen (and its vector) or the pest. Molecular markers can make selection more precise and increase the selection intensity.

Some traits, including those associated with yield or stress resistances appear at late developmental stages and only can be measured on mature plants. Therefore, large testing populations have to be cultivated up to maturity for selection. Molecular markers associated with traits of interest can allow selecting for these traits at early stages, reducing the time and costs required for plant cultivation and testing. Markers linked to traits of interest make the selection environment independent, allow for selection in off season nurseries, and permit accommodating multiple selection rounds in a year, therefore shortening the time required for completing a selection cycle. The advantages of molecular marker-based selection have been realized by plant breeders, and the technology is now applied on a broad range of crops, including legumes (Varshney et al. 2018).

8.3 Molecular Markers of Mungbean—A Brief History

Before molecular markers became available, genetic studies relied on studying morphological traits, such as flower color in pea (Mendel et al. 1993). Morphological traits that are controlled by a single gene can be used as genetic markers, but their number is limited, and without progeny tests, it is impossible to distinguish heterozygous from homozygous individuals. With the advent of isoenzyme markers, first genetic maps were constructed (Mahmoud et al. 1984). Protein markers were developed based on differences in mobility of certain proteins among different accessions. Mobility differences of seed proteins were used for cultivar detection (Mohanty et al.

2011). Pattnaik and Kole (2002) found protein markers that were polymorphic between MYMV-resistant and susceptible genotypes. The phylogenetic relationship between *Vigna* species was also addressed with protein markers (Kole and Panigrahi 2001). The limited number of polymorphic protein markers favored the development of more polymorphic DNA-based markers. About one decade after the first RFLP marker studies on major crop species, the technology was applied also on “orphaned” crops such as mungbean, and in the early 1990s, the first reports using restriction fragment length polymorphic (RFLP) markers for mapping traits in mungbean were published (Fatokun et al. 1992; Young et al. 1992, 1993). A genetic map based on RFLPs comprising 171 loci on 14 linkage groups was produced (Menancio-Hautea et al. 1993). Reports on molecular marker in mungbean became more frequent with the advent of random amplified polymorphic DNA (RAPD) markers. They were used to assess genetic diversity in germplasm (Santalla et al. 1998) and in cultivars (Lakhanpaul et al. 2000) and to map resistance to the most important diseases and pests such as mungbean yellow mosaic disease (Selvi et al. 2006) and bruchid beetles (Chen et al. 2007). The more reproducible AFLP markers allowed refining diversity studies (Singh et al. 2013) and trait mapping (Chaitieng et al. 2002; Srinives et al. 2010), but none of the markers developed with this technology was reported to be used in breeding. Then, large numbers of SSR markers were assembled for mungbean from sequence data of various *Vigna* species (Somta et al. 2009), or were generated for mungbean genomic sequences (Tangphatsornruang et al. 2009) or transcriptome sequencing data (Gupta et al. 2014; Chen et al. 2015a). A genetic map resolving the 11 linkage groups of mungbean was constructed with 150 SSR markers (Kajonphol et al. 2017), and QTLs for resistances to *Cercospora* leaf spot (Chankaew et al. 2011), powdery mildew (Kasettranon et al. 2010), nutritional traits such as phytic acid content (Sompong et al. 2012), and domestication-related traits (Isemura et al. 2012) were mapped. SSR markers also were used to

assess genetic diversity in a large germplasm collection to establish a mini-core collection (Schafleitner et al. 2015). Large-scale single-nucleotide polymorphic (SNP) marker detection was started in mungbean by transcriptome sequencing (Moe et al. 2011) and by comparing reads obtained by Illumina HiSeq sequencing of the genomes two mungbean cultivars (Van et al. 2013). Soon after, the whole genome sequence of mungbean cultivar VC1973A became available (Kang et al. 2014), paving the path for genotyping by sequencing approaches on this crop (Kang et al. 2014; Schafleitner et al. 2016). Current re-sequencing projects producing huge numbers of markers are likely to provide insight into genome re-arrangements in mungbean.

8.3.1 The First Molecular Markers for Mungbean Breeding: Markers Associated with Bruchid Resistance

The first application of molecular marker in mungbean was a study targeting bruchid-resistant loci. Promising bruchid resistance was discovered in wild mungbean *Vigna radiata* ssp. *sublobata* TC1966 (Fujii et al. 1989). At that time, the markers of choice were RFLPs. It was thought that RFPL may be a suitable marker system especially for crops with relatively small genome size such as mungbean to map genes and guide chromosome walking for gene cloning (Steinmetz et al. 1981). Young et al. (1992) analyzed 58 F₂ progeny derived from *V. radiata* ssp. *sublobata* TC1966 and a susceptible *V. radiata* line with 153 RFLP markers and succeeded to define a RFLP marker 3.6 cM distant from the bruchid resistance locus. One F₂ individual was identified carrying the bruchid resistance gene within a tightly linked double crossover. Such an individual would be highly valuable in developing resistant mungbean lines with reduced linkage drag. Later, Miyagi et al. (2004) succeeded to convert RFLP probes associated with bruchid resistance to PCR-based markers. They screened mungbean BAC libraries

from resistant and susceptible lines with RFLP probes associated with bruchid resistance and identified SSR motives and sequence tagged sites (STS) on these BACs. This experiment yielded PCR-based markers STSbr1 and STSbr2 that co-segregated with an RFLP marker associated with bruchid resistance (Miyagi et al. 2004). STSbr1 was validated on Indian genotypes to be associated with bruchid resistance (Sarkar et al. 2011), while STSbr2 was associated with one of two bruchid-resistant loci in *V. radiata* V2709 (Sun et al. 2008).

In the 1990s, the low-cost and easy-to-use RAPD markers were adopted for mungbean. Conversion of these markers to more robust SCAR markers made this marker type more useful. Chen et al. (2007) identified ten RAPD markers to be associated with bruchid resistance in a bulked segregant analysis in recombinant inbred lines (RILs) derived from a cross between the bruchid-resistant *V. radiata* ssp. *sublobata* line TC1966 and the mungbean yellow mosaic disease-resistant *V. radiata* elite cultivar NM92 bruchid. Three pools were established: One pool consisted of 22 bruchid-resistant F₁₂ RILs (0% infestation), and two pools were made of 20 susceptible RILs with 80–90% damage and 18 susceptible RILs of 90–100% damage. From the ten RAPD markers found to be associated with bruchid resistance in this experiment, the four most closely linked ones were cloned, sequenced, and transformed to SCAR and cleaved amplified polymorphism (CAP) markers. The CAP fragment derived from RAPD marker OPW02a4 was mapped to a location around position 6 mega base on chromosome 5, after the mungbean reference sequence (Kang et al. 2014) became available. QTL analysis using a mix of RAPD, SCAR, CAP, AFLP, and SSR markers (in total 489 markers) in the same population mapped bruchid resistance to linkage groups 7 and 9. Linkage group 9 was tagged with marker DMB158, which later was mapped to chromosome 5 (Schafleitner et al. 2016). QTL mapping using more than 9000 SNPs in population *V. radiata* ssp. *sublobata* TC1966 x *V. radiata* NM94 and more than 6000 SNPs in the cross of the two *V. radiata* lines V2802 x NM94

corroborated the presence of a bruchid resistance locus on chromosome 5 in both resistant lines, TC1966 and V2802. The markers are currently used in pyramiding bruchid resistance with disease resistances and good agronomic performance in mungbean breeding lines (Ramakrishnan Nair, personal communication).

The example bruchid resistance shows that molecular markers are a suitable tool for mungbean crop improvement. However, specifically for the trait bruchid resistance, evolution of marker technologies and access to large numbers of markers did not significantly improve the localization of the major resistance gene. The first RFLP markers associated with this resistance were mapped to a similar location like SSR and SNP markers in more recent experiments. The SSR marker DMB158 mapped nearest to the bruchid resistance locus (Chotechung et al. 2011; Chen et al. 2013). Fine-mapping the bruchid resistance locus confirmed the localization of this marker on chromosome 5 and resulted in two candidate genes VrPGIP1 and VrPGIP2 conferring the resistance (Chotechung et al. 2016; Kaewwongwal et al. 2017). The SNP markers linked to the resistance gene are not better than the SSR marker. This demonstrates that the trait, the biological material, and the assay conditions are by far more important for mapping traits than the marker system. Major resistance genes can be tagged also with simple methods, as long as the phenotyping data are sound. Including more markers and using more modern marker technologies do not necessarily improve the mapping result.

Bruchid resistance markers obtained by genotyping by sequencing were mapped to a reference sequence. The order of these markers on genetic maps was different to the order suggested by mapping the SNPs to the reference sequence (Schafleitner et al. 2016). This may be due to translocations, which caused differences in marker order in experimental populations compared to the reference sequence. Structural variations are important sources for phenotypic diversity. They are defined as genomic variations that involve segments of DNA larger than 1 kb in length and consist of insertions, deletions,

inversions, translocations, and copy number variations (Feuk et al. 2006). Genotyping with markers do not capture all structural variations (Springer et al. 2011). Whole genome re-sequencing is likely to improve the knowledge about structural variations.

8.3.2 Markers for Diversity Analysis in Mungbean

Mungbean is an autogamous (cryptogamous) species. Current cultivars have a narrow genetic base, because only a limited number of genotypes were used for breeding (Kumar et al. 2003). Breeding improved varieties therefore needs access to new genetic diversity. Due to the replacement or disappearance of wild relatives and local cultivars, alleles that could be of high interest for future breeding are continuously lost. Therefore, germplasm collections of mungbean landraces and wild relatives are an important reservoir to source new genetic diversity for breeding. The genetic diversity and population structure of germplasm accessions in gene banks need to be characterized to improve management of the collections by identifying redundant accessions, produce germplasm subsets with certain properties, and identify genotypes of interest for breeding (de Vicente et al. 2006).

Before the advent of molecular markers, diversity analysis depended on geographic information on the site of origin of an accession, pedigree data, morphologic or agronomic traits, or on biochemical data. Geographic origin together with morphological traits, discrete ones like bean color, or continuous ones like plant height or seed size, has been used to classify mungbean germplasm and analyze the diversity of collections (Bisht et al. 1998). The joint analysis of discrete and continuous variables has higher potential than analysis of either discrete or continuous data alone (Gonçalves et al. 2008). However, morphological data are susceptible to environmental variability. Measuring morphological traits in large collections usually is done over several seasons, bearing the risk that environmental variability is causing variation of traits

and subsequent errors in diversity analyses. In addition, morphological differences usually are determined by a small number of genes and may not be representative for the genetic diversity of the entire genome (Carroll 2008). In contrast, DNA markers are likely to reveal most accurately the genetic relationship among genotypes (reviewed by Crawford 1990). Marker genotypes are environment independent, and they are stable over different developmental stages of the plants. Small samples of plant tissue are sufficient for genotyping, and it is not necessary to grow plants to maturity, as it is required for morphological characterization, making genotyping a cheap option for diversity characterization. DNA markers likely provide information on homologous loci among genotypes, while morphological characteristics may be under the control of multiple genes, masking allelic relationships. DNA markers also are by far more abundant than morphological markers, increasing the power to discriminate between genotypes. Finally, scoring DNA markers is generally easier than measuring morphological parameters. Modern marker technologies also are amenable to automatization, further facilitating the approach.

Several marker technologies have been used to characterize mungbean germplasm. Santalla et al. (1998) have used RAPD markers to analyze genetic diversity in a small panel consisting of mungbean germplasm and three individuals of other *Vigna* species. Sixty random decamer primers were tested, and 28 pairs revealed being informative. The resulting phylogenetic tree showed three main clusters, which included *V. radiata* landraces, *Vigna mungo*, and *Vigna luteola*, respectively. Studies on the genetic diversity of Indian mungbean cultivars were also performed with RAPD markers (Lakhanpaul et al. 2000; Datta et al. 2012). These studies had in common that a relatively large number of primer combinations had to be tested and only about the half of the combinations yielded useful RAPD patterns for diversity analysis. RAPD markers can be readily applied on any organism, without previous sequence information. RAPD markers are generally abundant and evenly

distributed over the genome. The main weakness of RAPD markers is their low reproducibility (Schierwater and Ender 1993). Hence, these markers are difficult to be used across laboratories and experiments. The scoring of the bands can be complex and is subject to different interpretation when analyzed by different persons. High-quality DNA is critical for these assays, adding costs to the experiment. All together, these properties make RAPD markers a poor tool to analyze genetic diversity in large genebank collections.

Chattopadhyay et al. (2005) applied a combination of RAPD and inter-simple sequence repeats (ISSR) markers to study genetic diversity in selected mungbean genotypes. ISSRs are regions in the genome flanked by microsatellite sequences. These regions are amplified in PCRs using a primer that contains a microsatellite motif at the 3' end (Gupta et al. 1994). ISSR markers do not need any previous sequence information, are easy to use, and cause low costs. But ISSRs, like RAPD, may be affected by low reproducibility, and the obtained multiple bands may be derived from non-homologous loci and difficult to analyze.

A few reports describe the use of AFLP markers in mungbean diversity analysis (Bhat et al. 2005; Singh et al. 2013). Singh et al. (2013) compared phylograms obtained with ISSR and AFLP markers. They found that AFLP markers were more efficient than the ISSR in assessing genetic diversity, as they yielded more polymorphic markers than ISSR. The comparison of the Jaccard similarity matrices obtained with both marker systems showed only low correlation, and the clustering of genotypes within groups was not similar when AFLP- and ISSR-derived dendrograms were compared. It was hypothesized that the two marker technologies targeted different genomic regions and yielded different numbers of markers, which led to the different phylogenetic clustering of the accessions when the two methods were used. Advice for designing an ISSR experiment and recommendations on using ISSR markers in genetic variation studies has been disclosed in

Ng and Tan (2015). But in general, the easy-to-use SSR and SNP markers have widely replaced other marker systems, including ISSR.

Most diversity studies in mungbean have been accomplished with microsatellite (SSR) markers. These co-dominant markers are abundant in the genome, are easy and cheap to use, and are amenable to multiplexing and automatization (Hayden et al. 2008). Originally, SSR markers were developed from repeat-enriched libraries (Edwards et al. 1996), a labor intense technology. But with readily available sequence information from transcriptomes and genomes, microsatellites became much easier to access (Chen et al. 2015a). Specialized software tools to mine DNA sequences for microsatellite motifs and design primers to amplify microsatellite loci are widely available (da Maia et al. 2008; Wang and Wang 2016). Still, microsatellite markers need to be well chosen to obtain allelic bands in genotyping. Backward and forward mutations (homoplasy) may occur at microsatellite loci and cause underestimation of the genetic diversity (Spooner et al. 2005). In mungbean, SSR markers have been developed using the 5'-anchored polymerase chain reaction technique (Kumar et al. 2002), from genome shotgun sequences (Tangphatsornruang et al. 2009), transcriptome sequences of *V. radiata* (Chen et al. 2015a; Gupta et al. 2014), or have been transferred from other *Vigna* species (Isemura et al. 2012).

The first comprehensive study on mungbean diversity used a set of 19 SSR markers derived from adzuki bean (*Vigna angularis*) on 615 cultivated and wild mungbean accessions (Sangiri et al. 2008). The marker set was selected based on the marker location in the adzuki bean genome to contain at least one marker per linkage group. More alleles were detected in wild than in cultivated accessions, illustrating the lower diversity in the cultivated germplasm set. The study revealed that Australia and New Guinea represent a distinct center of diversity for wild mungbean, while cultivated mungbean has greatest diversity in South Asia. Soon after, the diversity and population structure of mungbean

were analyzed with 15 different SSR markers in 692 mungbean accessions held by the National Agrobiodiversity Center of the Rural Development Administration, Korea. Mungbean germplasm obtained from 27 countries was grouped into seven phylogenetic clades and into two distinct genetic groups (Gwag et al. 2010). In total, 157 mungbean germplasm accessions were genotyped with EST-SSRs (Chen et al. 2015b).

A combination of morphological data and microsatellite markers was used to define a 300 accession mini-core collection that represents a large proportion of the overall diversity of the whole World Vegetable Center mungbean collection of more than 6700 accessions (Schafleitner et al. 2015). In the first step, geographic stratification was performed, and by cluster analysis of eight phenotypic descriptors, a phylogenetic tree was produced. From this tree, 20% of the accessions were randomly selected from each cluster as a core collection containing about 1400 genotypes. The core collection was subsequently genotyped with 20 microsatellite markers, and a mini-core set was selected to represent all detected 122 alleles (Schafleitner et al. 2015). The collection was small enough to be submitted to multilocation trials in various regions in Asia and Africa to discover new traits for mungbean breeding, and it is expected that it is large enough to map traits in genome-wide association studies.

Other marker types such as single-strand confirmation polymorphism, cleaved amplified polymorphic sequence, and SCAR markers that were used for diversity analysis in other crops (Spooner et al. 2005) were not reported for similar works in mungbean, while single-nucleotide polymorphic markers (SNPs) were applied for analyzing several germplasm collections.

The first SNPs for mungbean were reported for pairs of mungbean lines by Moe et al. (2011) from transcriptome sequences, followed by Van et al. (2013) from shotgun Illumina sequences. Availability of the mungbean whole genome sequence (Kang et al. 2014) strongly improved

the access to SNPs for this species. Re-sequencing of selected lines yielded large numbers of SNPs (Liu et al. 2016), and genotyping by sequencing (Elshire et al. 2011) was applied to mungbean populations (Kang et al. 2014; Schafleitner et al. 2016). Genotyping of germplasm accessions of mungbean with SNPs was done on the USDA mungbean collection, the Australian mungbean diversity panel (Noble et al. 2018), on the World Vegetable Center mini-core (Breria et al. 2019).

The SNP-based diversity analysis of 94 cultivated mungbean genotypes from the USDA collection originating from 27 countries was done using a small set of SNP markers (Islam and Blair 2018). From a total of 42 known SNPs (Van et al. 2013), 18 were successfully converted to polymorphic KASP markers. The population could be divided in two subpopulations and one admixture group.

The Australian diversity panel was submitted to GBS. The germplasm set consisted of 466 cultivated and 16 wild accessions. In total, more than 22,000 polymorphic genome-wide SNPs were identified and used to analyze the genetic diversity, population structure, and linkage disequilibrium (Noble et al. 2018). As expected, polymorphism was lower in the cultivated than in the wild accessions. Linkage disequilibrium decay amounted to about 100 kb in cultivated lines and about 60 kb in wild mungbean. Structure analysis identified four distinct subgroups, which broadly corresponded to geographic origin and seed characteristics (Noble et al. 2018).

Genotyping using GBS of the World Vegetable Center mini-core produced more than 24,000 markers for a germplasm panel consisting of *V. radiata* and *V. mungo* and 8000 polymorphic markers for 296 *V. radiata* accessions. From this set, 5447 polymorphic SNPs were used for germplasm characterization and structure analysis, identifying two major populations, one of them falling into three subpopulations, in the World Vegetable Center germplasm set. The mini-core and the genotyping data are currently used to map a number of morpho-agronomic traits.

8.3.3 Molecular Marker for Cultivar Identification and Hybridity Tests

Molecular fingerprinting of varieties and determining purity of seed is a component of quality seed production. Testing seed purity with molecular markers is common for many crops and is considered to be quicker and more cost-effective than grow-out tests (Yashitola et al. 2002). This, however, may not be true for all cases. For example, much of the mungbean seed production and much of its growing area are located in developing countries where wages are low and where there is little access to infrastructure for low-cost genotyping. Therefore, grow-out tests may be still cheaper than genotyping for seed quality assessment. Ali et al. (2010) reported seed quality assessment of Bangladeshi mungbean varieties based on quantifying other seed than mungbean and inert matter in seed lots, seed moisture content, 1000 seed weight and germination tests. Molecular markers have been applied to produce molecular fingerprints of varieties (Tantasawat et al. 2010; Lestari et al. 2014; Reflinur et al. 2017), but reports on systematic use of markers for seed quality monitoring for mungbean are not available.

Monitoring the success of crosses by hybridity tests with molecular markers is a common practice (Solanki et al. 2010). In mungbean, SSR markers are being used to monitor crosses between mungbean germplasm and breeding lines, as well as in wide crosses between cultivated mungbean and wild relatives (Ramakrishnan Nair, personal communication). One or a few polymorphic SSR markers that are generally easy to define and cheap to apply are sufficient for this task.

8.3.4 Developing Markers Linked to Traits of Interest

Disease-resistant cultivars are the cheapest, simplest, and most environmentally safe way to

manage disease. Likewise, improving abiotic stress tolerance of crops can stabilize yields and prevent crop failure. Disease resistance and abiotic stress tolerance are often sourced from landraces and wild relatives. Introgression of biotic and abiotic stress-tolerant traits from unadapted material into elite cultivars is a frequent breeding task. As outlined above, using molecular markers can improve introgression of these traits into elite breeding material.

Developing markers associated with traits of interest include the following steps:

- (1) Establish the genetic resources for trait mapping, for example, a mapping population or a germplasm panel segregating for the trait of interest
- (2) Phenotype the population and generate trait value data, e.g., on resistance or susceptibility to a pathogen or pest, or on tolerance to an abiotic stress factor
- (3) Develop markers and genotype the experimental population
- (4) Associate phenotypes to specific marker genotypes using appropriate statistical methods
- (5) Validate the candidate markers in different genetic backgrounds and produce user-friendly markers for marker-assisted selection.

Once the genetic resources are phenotyped, a marker system has to be chosen to genotype the population or germplasm panel. Today, the most popular markers are SNPs. Genotyping by sequencing has been successfully used in mungbean to generate a large number of SNP markers for bi-parental populations and germplasm panels (Schafleitner et al. 2016; Noble et al. 2018). Ongoing whole genome re-sequencing efforts benefit from the available whole genome reference sequence of mungbean (Kang et al. 2014) and are likely to provide large number of SNPs for this species.

Several methods are available to associate phenotypic traits with genotypes. Bulk segregant analysis uses bulked DNA samples generated from individuals of a segregating population

from a single cross (Michelmore et al. 1991). Each bulk contains DNA from individuals that are identical for a particular trait such as disease and pest resistance or susceptibility, but are arbitrary at all unlinked regions. The two bulks are therefore genetically dissimilar in the selected region but seemingly heterozygous at all other regions. The bulks are screened for genetic differences using suitable markers to identify loci that have contrasting alleles at homozygote state in the two bulks. Bulk segregant analysis is a rapid and simple method to determine association of markers to single gene or oligogenic traits, but is generally not suitable for multigenic traits. On mungbean, bulk segregant analysis has been used to map bruchid resistance (Cheng et al. 2005, 2007; Sun et al. 2008), Mungbean yellow mosaic disease resistance (Selvi et al. 2006; Dhole and Reddy 2013; Karthikeyan et al. 2012), and iron deficiency tolerance (Toojinda et al. 2001). Markers associated with the respective traits were obtained, but no application of these markers in breeding was reported.

Quantitative traits typically are tagged by QTL analysis (Liu 2017). For this task, first, the molecular markers are mapped, on either genetic or physical maps. Then, associations between the trait(s) of interest and the marker genotypes are tested using statistical methods. A large number of reports describe QTL studies on disease resistances, quality, and domestication traits in mungbean. The first reported QTL study on mungbean found associations between seed weight and RFLP marker genotypes (Fatokun et al. 1992). Humphry et al. (2005) investigated the relationships between hard-seededness and seed weight to support breeding of hard- and large-seeded genotypes. A large number of QTL analyses followed, targeting a wide range of morphological, agronomical, and nutritional traits. Disease and pest resistances were probably the traits that were most frequently targeted by such analyses. Results of many of these studies are summarized in the chapter “Genomic Approaches to Biotic Stresses” by Laosatit et al. in this book.

Genome-wide association studies (GWAS) are a quantitative method to test whether a genomic variant (marker genotype) is associated

with a trait of interest using a germplasm panel as experimental population. It assumes that a specific property such as disease resistance, abiotic stress tolerance, or a nutritional trait shared by a subset of the germplasm panel is reflected by a specific marker genotype also shared by these individuals. The markers have to be in linkage disequilibrium with the genes conferring the trait. In comparison to QTL studies on bi-parental populations, GWAS have the advantage to work on germplasm panels and do not need specific mapping populations. Therefore, GWAS can analyze the function of all alleles and haplotypes present in the germplasm set under investigation, while QTL studies on bi-parental populations only take into account the alleles present in the mapping parents. The resolution of GWAS is generally higher than that of QTL analyses in bi-parental populations. Resolution depends on the number of recombination events that separate the investigated genotypes from each other. Bi-parental populations generally have undergone only a low number of recombination events until analysis, while germplasm panels have a long history of evaluation and therefore individuals are usually separated from each other by many recombination events. One of the major drawbacks of GWAS compared to QTL mapping in bi-parental populations is that population structure influences the outcome of the study, but inclusion of population structure into the GWAS model tries to mitigate this effect. Furthermore, GWAS generally requires a larger number of markers than mapping in bi-parental populations. The marker number needed depends on the linkage disequilibrium decay distance in the germplasm panel and is specific for the species and kind of population under investigation. However, modern genotyping technologies provide easy access to large numbers of molecular markers helping to overcome this drawback.

Up to date, only a few genome-wide association studies were undertaken in mungbean. This is probably due to the lack of high-quality phenotyping data for densely genotyped germplasm panels. A pilot GWAS on seed color in the Australian diversity panel revealed five genomic regions associated with this trait (Noble et al.

2018). Ongoing phenotyping efforts for mungbean diversity panels will likely lead to a broader application of GWAS to identify marker trait associations. As the resolution of GWAS often goes down to the gene level (Liu and Yan 2019), these studies not only give markers associated with a trait of interest. They are likely to improve the knowledge on the genetic basis of traits including the causative genes or alleles and their interactions (Hansen 2006).

The bottleneck for successful GWAS are, as mentioned above, high-quality phenotypic data, the complexity of the trait of interest, and the size of the germplasm panel used for determining marker trait associations. Even in associations that are statistically highly significant, false-positive associations may still occur. The large number of statistical inferences, inaccurate genotyping, and too small population sizes make the results prone to errors (Liu and Yan 2019). Like in QTL studies in bi-parental populations, the loci found to be associated in these studies need to be thoroughly validated before drawing conclusions on gene functions or using any of the markers apparently linked to a trait of interest in breeding. Candidate genes and alleles that are found association also in a different population can be assumed to be more likely linked with the trait of interest. In addition, candidate gene knock-out or overexpression studies are suitable methods for validation.

Both QTL analysis and GWAS are appropriate to tag loci conferring a trait with markers, but both approaches are poor tools to analyze complex traits, where a large number of loci contribute to a trait, such as yield and many abiotic stress tolerances (Heffner et al. 2009). Genomic selection became a powerful tool to use molecular markers for selection without associating specific markers to traits. Instead, phenotyping data and high-density genotyping information are used to calculate genomic breeding values for individuals (Heffner et al. 2009). Like that, 1000s of loci along the whole genome are included in the analysis, which reflects the contribution of 1000s of genes to complex traits. The method has been applied in animal breeding with great success (Hayes et al. 2009), is becoming more and

more used in plant breeding (Voss-Fels et al. 2018), and has potential also in legume breeding (Mousavi-Derazmahalleh et al. 2019).

8.3.5 Marker-Assisted Selection

In section “Molecular Markers in Plant Breeding,” some advantages were highlighted for selecting based on molecular markers that are tightly linked to traits instead of using trait values directly. The main advantages of marker-assisted selection are:

- (1) MAS makes selection for traits that are difficult to measure easier
- (2) It allows for selection of traits that are expressed during late developmental stages already at the seedling stage
- (3) It eliminates environmental variability from the selection
- (4) It makes selection of disease-resistant or abiotic stress-tolerant individuals independent of the presence of the biotic or abiotic stimuli (pathogens, pests, vectors, heat, etc.) required for selection
- (5) It helps maintaining recessive alleles during backcrossing
- (6) It facilitates pyramiding multiple traits, especially pyramiding multiple loci for the same trait.

Marker-assisted selection is designed to maintain introgressed loci in the population, while marker-assisted backcrossing (MABC) helps introgressing loci generally from unadapted material into an elite background. The introgressed fragment, in addition to the target gene, may contain genes that reduce the agronomic performance of the line. This effect is called linkage drag. Therefore, fragments that are as small as possible and contain as little genetic material from the donor line in addition to the target gene are preferred. Marker-assisted selection with markers can help to reduce the linkage drag and accelerate the reestablishment of the recurrent parent. Overall, the efficiency of MABC depends on the kind of the

introgressed gene, the recurrent and the donor parent and the population size (Frisch and Melchinger 2005).

MABC combines foreground selection with markers associated with the trait of interest with background selection with markers that pinpoint offspring with maximal recovery of the recurrent parent genotype. The foreground selection monitors presence of the introgressed fragment in the progeny. Marker-assisted foreground selection with co-dominant markers such as SSRs or SNPs that are tightly associated with the trait of interest is particularly practical for traits that are not expressed at the heterozygote stage or are difficult to score. Introgression means that a double recombination occurring on both sides of a target locus has to occur. This event can best be monitored with a marker pair tightly flanking the target gene, and not with a single linked marker. Literature recommends markers to flank the target gene in a distance of maximally 5 cM, but new marker technologies that allow for much greater marker densities enable the choice of more tightly linked marker, to further reduce linkage drag.

Recombinant selection in foreground selection involves identifying backcross progeny with recombination events as near as possible to the target locus, to reduce the size of the donor chromosome segment containing the target locus and reduce linkage drag. As selection is applied on the target locus, there will be less recombination around the donor fragment than for unlinked regions (Hospital et al. Hospital 2001). As double crossover events occurring on both sides of the introgressed fragment are rare, the donor segment can remain very large, even with many backcross generations. The population size for backcrossing has to be adjusted to the crossover probability. The probability of a double crossover can be calculated from the product of the probabilities of a single recombination on both sides of the introgressed fragment. But for close markers, the probability of double crossovers is much lower than the probabilities for single crossover combinations (Young and Tanksley 1989). Producing a very large number of backcross plants would be necessary to

achieve recombination on both sides of the gene in one cross. However, instead of working with a very large population, it is advantageous to select in the first backcross generation a single recombinant on one side and then selecting the recombinant on the other side in a second backcross generation (Young and Tanksley 1989). In summary, the distances between the flanking markers and the target gene, the population size during backcrossing, and the number of backcrosses are critical for reducing linkage drag (Hospital 2001).

Background selection in MABC involves selecting backcross progeny with the greatest proportion of the recovered recurrent parent genome using markers that are unlinked to the target locus and can be used to select against the donor genome. Background selection aims at accelerating the recovery of the recurrent genome. Without markers, the reconstitution of the recurrent phenotype, at least to 97%, can be accomplished within four backcross generations (Frisch et al. 1999), but selection for the introgressed trait affects the recovery and increases the required backcrosses by at least one cycle (Frisch et al. 1999). It was proposed to start with a large backcross population to increase the chance to identify an individual that has recovered the recurrent parent genome to an extent as large as possible, and to reduce the population size for the next generations. Simulation studies estimating the trade-offs between population size, MABC efficiency and costs are available and suggest steps to optimize MABC (Ribaut et al. 2002).

For mungbean, marker-assisted backcrossing efforts were not yet reported, but availability of markers associated with important traits makes it likely that this technology will be used also in mungbean improvement.

8.3.6 Pyramiding Multiple Traits in Breeding Lines

Pyramiding is the process of combining several genes together in a single genotype. Conventional breeding also applies gene pyramiding, but

usually it is laborious and time consuming to check the results of this approach by phenotypic tests. For example, to improve agronomic properties, breeders combine multiple disease resistances in elite lines. Checking resistance to multiple diseases is laborious and requires multiple testing environments. The efficiency of this process can be enhanced by marker-assisted selection. To increase the durability of disease resistances, breeders pyramid various resistance genes from different sources (Hanson et al. 2016). Using conventional phenotypic selection, identification of stacked resistance genes is only possible when pathogen races that can detect specific resistance genes are available. In contrast, molecular markers greatly facilitate gene pyramiding, as they can be designed to be specific for each single resistance gene. Early selection by molecular markers also helps to keep the breeding populations small during gene pyramiding. However, in mungbean, no marker-assisted gene pyramiding efforts were reported so far.

8.3.7 Genomic Selection

Access to marker resources open up new methods for selecting favorable genotypes, if sufficient phenotypic data of the organism under investigation are available. As outlined in section “Developing markers linked to traits of interest,” genomic selection is taking up momentum in crop breeding. It is thought that genomic selection is particularly advantageous for selecting favorable genotypes for complex, multigenic traits. However, the technology requires datasets from different environments and over a number of generations. These sets are not yet available for mungbean, so it will still take some time that this technology can be applied on this crop.

8.3.8 Constraints to Successful Marker-Assisted Selection

Great investments in marker-assisted selection, primary in the private sector, have resulted in

several improved varieties for a range of crops including cereals, oil seed crops, cotton, legumes, and vegetables. Naturally, for minor crops such as mungbean, there has been much less investment in generating breeding tools, including genomic resources for breeding. But the available whole genome sequence for mungbean, germplasm panels displaying the diversity of the crop and coordinated breeding activities such as the Australian Centre for International Agricultural Research (ACIAR)-funded International Mungbean Breeding Network (<https://www.aciar.gov.au/project/CIM/2014/079>) make marker-assisted breeding also accessible for mungbean. Especially, disease and pest resistances are likely to be tackled by marker-assisted breeding in the very near future. Marker-assisted breeding for complex traits such as abiotic stress tolerance in mungbean probably will take longer, as it requires putting in place the phenotypic datasets to make use of molecular breeding for complex traits.

Cost savings compared to classical breeding are often mentioned as advantage of marker-assisted selection. Nevertheless, for some breeding programs, the investment required for marker-assisted selection may still be an issue. There are some early studies reporting several cases where marker-assisted selection was less cost-effective than phenotypic selection (Bohn et al. 2001; Dreher et al. 2003). In the meanwhile, the costs for genotyping have dropped, but the investment for molecular breeding may still be relatively high for small programs working on minor crops. Especially in developing countries, breeder may not have easy access to cost-effective genotyping infrastructure, and low labor costs may make field evaluations cheaper than laboratory work that requires relatively expensive consumables. However, the accelerated release of an improved crop variety may translate into more rapid profits. Therefore, if the additional income generated by improved varieties along the mungbean value chain over time is considered, calculations analyzing the costs

and benefit of marker-assisted selection in plant breeding will most probably show that this technique is meaningful, also on mungbean.

Lack of access to molecular markers does not limit marker-assisted breeding in mungbean anymore, as technologies to obtain large numbers of markers at low cost are available. High-quality phenotypic data are being produced and expertise to combine phenotypic and genotypic datasets is available in mungbean breeding teams. Therefore, the first improved mungbean varieties produced by marker-assisted selection are in sight.

References

- Ali MZ, Khan M, Rahaman AK, Ahmed M, Ahsan AF (2010) Study on seed quality and performance of some mungbean varieties in Bangladesh. *Int J Exp Agric* 2:10–15
- Barata C, Carena MJ (2006) Classification of North Dakota maize inbred lines into heterotic groups based on molecular and testcross data. *Euphytica* 151:339–349
- Bhagyawant SS (2016) RAPD-SCAR markers: an interface tool for authentication of traits. *J Biosci Med* 4:1–9
- Bhat KV, Lakhanpaul S, Chadha S (2005) Amplified fragment length polymorphism (AFLP) analysis of genetic diversity in Indian mungbean [*Vigna radiata* (L.) Wilczek] cultivars. *Indian J Biotechnol*. <http://nopr.niscair.res.in/handle/123456789/5634>
- Bisht IS, Mahajan RK, Patel DP (1998) The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. *Genet Resour Crop Evol* 45:127–133
- Bohn M, Groh S, Khairallah MM, Hoisington DA, Utz HF, Melchinger AE (2001) Re-evaluation of the prospects of marker assisted selection for improving insect resistance against *Diatraea* spp. in tropical maize by cross validation and independent validation. *Theor Appl Genet* 103:1059–1067
- Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120:1095–1103
- Breria CM, Hsieh CH, Yen J-Y, Nair R, Lin C-Y, Huang S-M, Noble TJ, Schafleitner R (2019) Population structure of the world vegetable center mungbean mini core collection and genome-wide association mapping of loci associated with variation of seed coat luster. *Trop Plant Biol*

- Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25–36
- Chaitieng B, Kaga A, Han OK, Wang XW, Wongkaew S, Laosuwan P, Tomooka N, Vaughan DA (2002) Mapping a new source of resistance to powdery mildew in mungbean. *Plant Breed* 121:521–525
- Chankaew S, Somta P, Sorajjapinun W, Srinives P (2011) Quantitative trait loci mapping of *Cercospora* leaf spot resistance in mungbean, *Vigna radiata* (L.) Wilczek. *Mol Breed* 28:255–264
- Chattopadhyay K, Ali MN, Sarkar HK, Mandai N, Bhattacharyya S (2005) Diversity analysis by RAPD and ISSR markers among the selected mungbean [*Vigna radiata* (L.) Wilczek] genotypes. *Indian J Genet Plant Breed* 65:173–175
- Chen HM, Liu CA, Kuo CG, Chien CM, Sun HC, Huang CC, Lin YC, Ku HM (2007) Development of a molecular marker for a bruchid (*Callosobruchus chinensis* L.) resistance gene in mungbean. *Euphytica* 157:113–122
- Chen HM, Ku HM, Schafleitner R, Bains TS, Kuo CG, Liu CA, Nair RM (2013) The major quantitative trait locus for mungbean yellow mosaic Indian virus resistance is tightly linked in repulsion phase to the major bruchid resistance locus in a cross between mungbean [*Vigna radiata* (L.) Wilczek] and its wild relative *Vigna radiata* ssp. *sublobata*. *Euphytica* 192:205–216
- Chen H, Qiao L, Wang L, Wang S, Blair MW, Cheng X (2015a) Assessment of genetic diversity and population structure of mung bean (*Vigna radiata*) germplasm using EST-based and genomic SSR markers. *Gene* 566(2):175–183
- Chen H, Wang L, Wang S, Liu C, Blair MW, Cheng X (2015a). Transcriptome sequencing of mung bean (*Vigna radiata* L.) genes and the identification of EST-SSR markers. *PLoS One* 10(4):e0120273
- Cheng X, Wang S, Wu S, Zhou J, Wang S, Yang Y (2005) Tagging and utilization of bruchid resistance gene using PCR markers in mungbean. *Zhongguo nongye kexue* 38:1534–1539
- Chotechung S, Somta P, Chankaew S, Srinives P, Somta P (2011) Identification of DNA markers associated with bruchid resistance in mungbean. *Khon Khan Agri J* 39:221–226
- Chotechung S, Somta P, Chen J, Yimram T, Chen X, Srinives P (2016) A gene encoding a polygalacturonase-inhibiting protein (PGIP) is a candidate gene for bruchid (*Coleoptera: bruchidae*) resistance in mungbean (*Vigna radiata*). *Theor Appl Genet* 129:1673–1683
- Crawford DJ (1990) *Plant molecular systematics: macro-molecular approaches*. Wiley, New York, USA
- Da Maia LC, Palmieri DA, De Souza VQ, Kopp MM, de Carvalho FI, Costa de Oliveira A (2008) SSR locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int J Plant Genomics* s.2008:412696
- Datta S, Gangwar S, Kumar S, Gupta S, Rai R, Kaashyap M, Singh P, Chaturvedi SK, Singh BB, Nadarajan N (2012) Genetic diversity in selected Indian mungbean [*Vigna radiata* (L.) Wilczek] cultivars using RAPD markers. *Am J Plant Sci* 8:1085–1091
- de Vicente MC, Guzman FA, Engels J, Rao VA (2006) Genetic characterization and its use in decision-making for the conservation of crop germplasm. The role of biotechnology in exploring and protecting agricultural genetic resources, pp 2129–138
- Dhole VJ, Reddy KS (2013) Development of a SCAR marker linked with a MYMV resistance gene in mungbean (*Vigna radiata* L. Wilczek). *Plant Breed* 132:127–132
- Dreher K, Khairallah M, Ribaut JM, Morris M (2003) Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. *Mol Breed* 11:221–234
- Edwards KJ, Barker JH, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques* 20:758–760
- Elbasyoni IS, Lorenz AJ, Guttieri M, Frels K, Baenziger PS, Poland J, Akhunov E (2018) A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Sci* 270:123–130
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):e19379
- Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L (2003) Highly parallel SNP genotyping. In: *Cold spring harbor symposia on quantitative biology*, vol 68. Cold Spring Harbor Laboratory Press, pp 69–78
- Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841–846
- Feuk L, Marshall CR, Wintle RF, Scherer SW (2006) Structural variants: changing the landscape of chromosomes and design of disease studies. *Human Mol Genet* 15(suppl_1):R57–R66
- Frisch M, Bohn M, Melchinger AE (1999) Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci* 39(5):1295–1301
- Frisch M, Melchinger A (2005) Selection theory for marker-assisted backcrossing. *Genetics* 170(2):909–917
- Fujii K, Ishimoto M, Kitamura K (1989) Patterns of resistance to bean weevils (*Bruchidae*) in *Vigna radiata-mungo-sublobata* complex inform the breeding of new resistant varieties. *Appl Entomol Zool* 25:126–132
- Gonçalves LS, Rodrigues R, Amaral Júnior AD, Karasawa M, Sudré CP (2008) Comparison of multivariate

- statistical algorithms to cluster tomato heirloom accessions. *Genet Mol Res* 7:1289–1297
- Gupta SK, Bansal R, Gopalakrishna T (2014) Development and characterization of genic SSR markers for mungbean (*Vigna radiata* (L.) Wilczek). *Euphytica* 195(2):245–258
- Gupta M, Chyi YS, Romero-Severson J, Owen JL (1994) Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor Appl Genet* 89:998–1006
- Gwag JG, Dixit A, Park YJ, Ma KH, Kwon SJ, Cho GT, Lee GA, Lee SY, Kang HK, Lee SH (2010) Assessment of genetic diversity and population structure in mungbean. *Genes Genomics* 32:299–308
- Hamada H, Kakunaga T (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* 298(5872):396–398
- Hansen TF (2006) The evolution of genetic architecture. *Annu Rev Ecol Syst* 37:123–157
- Hanson P, Lu SF, Wang JF, Chen W, Kenyon L, Tan CW, Tee KL, Wang YY, Hsu YC, Schafleitner R, Ledesma D (2016) Conventional and molecular marker-assisted selection and pyramiding of genes for multiple disease resistance in tomato. *Scientia Hort* 201:346–354
- Hayden MJ, Nguyen TM, Waterman A, McMichael GL, Chalmers KJ (2008) Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Mol Breed* 21:271–281
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92:433–443
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. *Crop Sci* 49(1):1–2
- Hospital F (2001) Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics* 158(3):1363
- Humphry ME, Lambrides CJ, Chapman SC, Aitken EA, Imrie BC, Lawn RJ, McIntyre CL, Liu CJ (2005) Relationships between hard-seededness and seed weight in mungbean (*Vigna radiata*) assessed by QTL analysis. *Plant Breed* 124(3):292–298
- Isemura T, Kaga A, Tabata S, Somta P, Srinives P, Shimizu T, Jo U, Vaughan DA, Tomooka N (2012) Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS One* 7:e41304
- Islam AS, Blair MW (2018) Molecular characterization of mungbean germplasm from the USDA core collection using newly developed KASP-based SNP markers. *Crop Sci* 58(4):1659–1670
- Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29:e25
- Kaewwongwal A, Chen J, Somta P, Kongjaimun A, Yimram T, Chen X, Srinives P (2017) Novel alleles of two tightly linked genes encoding polygalacturonase-inhibiting proteins (VrPGIP1 and VrPGIP2) associated with the Br locus that confer bruchid (*Callosobruchus* spp.) resistance to mungbean (*Vigna radiata*) accession V2709. *Front Plant Sci* 8:1692
- Kajonphol T, Sangsiri C, Somta P, Toojinda T, Srinives P (2017) SSR map construction and quantitative trait loci (QTL) identification of major agronomic traits in mungbean (*Vigna radiata* (L.) Wilczek). *SABRAO J Breed Genet* 44(1):71–86
- Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha BK, Jun TH, Hwang WJ, Lee T, Lee J, Shim S (2014) Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat Commun* 11:6443
- Karthikeyan A, Sudha M, Senthil N, Pandiyan M, Raveendran M, Nagrajan P (2012) Screening and identification of random amplified polymorphic DNA (RAPD) markers linked to mungbean yellow mosaic virus (MYMV) resistance in mungbean (*Vigna radiata* (L.) Wilczek). *Arch Phytopath Plant Prot* 45:712–716
- Kasettranon W, Somta P, Srinives P (2010) Mapping of quantitative trait loci controlling powdery mildew resistance in mungbean (*Vigna radiata* (L.) Wilczek). *J Crop Sci Biotechnol* 13(3):155–161
- Kole C, Panigrahi J (2001) *Vigna glabrescens* is a natural allopolyploid of *V. radiata* and *V. umbellata*: evidences from seed protein electrophoresis. *Crop Res* 22:121–128
- Kumar SV, Tan SG, Quah SC, Yusoff K (2002) Isolation of microsatellite markers in mungbean, *Vigna radiata*. *Mol Ecol Notes* 2:96–98
- Kumar S, Gupta S, Chandra S, Singh BB (2003) How wide is the genetic base of pulse crops? In: National symposium on crop diversification and natural resources management, Kanpur, 20–22 Dec 2003
- Lakhanpaul S, Chadha S, Bhat KV (2000) Random amplified polymorphic DNA (RAPD) analysis in Indian mung bean (*Vigna radiata* (L.) Wilczek) cultivars. *Genetica* 109(3):227–34
- Lestari P, Kim SK, Kang YJ, Dewi N, Lee SH (2014) Genetic diversity of mungbean (*Vigna radiata* L.) germplasm in Indonesia. *Plant Genet Res* 12(S1):S91–S94
- Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, Wittwer C (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin Chem* 50(7):1156–1164
- Liu BH (2017) Statistical genomics: linkage, mapping, and QTL analysis. CRC press
- Liu HJ, Yan J (2019) Crop genome-wide association study: a harvest of biological relevance. *Plant J* 97(1):8–18
- Liu MS, Kuo TC, Ko CY, Wu DC, Li KY, Lin WJ, Lin CP, Wang YW, Schafleitner R, Lo HF, Chen CY (2016) Genomic and transcriptomic comparison of nucleotide variations for insights into bruchid

- resistance of mungbean (*Vigna radiata* [L.] R. Wilczek). *BMC Plant Biol* 16:46
- Mahmoud SH, Gatehouse JA, Boulter D (1984) Inheritance and mapping of isoenzymes in pea (*Pisum sativum* L.). *Theor Appl Genet* 68(6):559–66
- Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D, Young ND (1993) Comparative genome analysis of mungbean (*Vigna radiata* L. Wilczek) and cowpea (*V. unguiculata* L. Walpers) using RFLP mapping data. *Theor Appl Genet* 86:797–810
- Mendel G, Corcos AF, Monaghan FV (1993) Gregor Mendel's experiments on plant hybrids: a guided study. Rutgers University Press
- Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci* 88(21):9828–9832
- Miyagi M, Humphry M, Ma ZY, Lambrides CJ, Bateson M, Liu CJ (2004) Construction of bacterial artificial chromosome libraries and their application in developing PCR-based markers closely linked to a major locus conditioning bruchid resistance in mungbean (*Vigna radiata* L. Wilczek). *Theor Appl Genet* 110:151–156
- Moe KT, Chung JW, Cho YI, Moon JK, Ku JH, Jung JK, Lee J, Park YJ (2011) Sequence information on simple sequence repeats and single nucleotide polymorphisms through transcriptome analysis of mungbean. *J Int Plant Biol* 53:63–73
- Mohanty JB, Naik BS, Panigrahi J, Kole, C (2011) Identification of mungbean cultivars based on electrophoretic patterns of seed proteins. *Crop Res* 21(2):134–138
- Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol* 147:969–977
- Mousavi-Derazmahalleh M, Bayer PE, Hane JK, Valliyodan B, Nguyen HT, Nelson MN, Erskine W, Varshney RK, Papa R, Edwards D (2019) Adapting legume crops to climate change using genomic approaches. *Plant, Cell Environ* 42(1):6–19
- Ng WL, Tan SG (2015) Inter-simple sequence repeat (ISSR) markers: are we doing it right. *ASM Sci J* 9(1):30–39
- Noble TJ, Tao Y, Mace ES, Williams B, Jordan DR, Douglas CA, Mundree SG (2018) Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. *Front Plant Sci* 8:2102
- Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor Appl Genet* 85(8):985–993
- Pattnaik J, Kole C (2002) Detection of a protein marker for screening of MYMV resistance mungbean genotype. *Indian J Genet Plant Breed* 62:77–78
- Reffinur R, Lestari P, Lee SH (2017) The potential use of SSR markers to support the morphological identification of Indonesian mungbean varieties. *Indonesian J Agric Sci* 17(2):65–74
- Ribaut JM, Jiang C, Hoisington D (2002) Simulation experiments on efficiencies of gene introgression by backcrossing. *Crop Sci* 42(2):557–565
- Sangiri C, Kaga A, Tomooka N, Vaughan D, Srinives P (2008) Genetic diversity of the mungbean (*Vigna radiata*, Leguminosae) gene pool on the basis of microsatellite analysis. *Aust J Bot* 55:837–847
- Santalla M, Power JB, Davey MR (1998) Genetic diversity in mung bean germplasm revealed by RAPD markers. *Plant Breed* 117:473–478
- Sarkar S, Ghosh S, Chatterjee M, Das P, Lahari T, Maji A, Mondal N, Pradhan KK, Bhattacharyya S (2011) Molecular markers linked with bruchid resistance in *Vigna radiata* var. *sublobata* and their validation. *J Plant Biochem Biotech* 20(2):155–160
- Schafleitner R, Nair RM, Rathore A, Wang YW, Lin CY, Chu SH, Lin PY, Chang JC, Ebert AW (2015) The AVRDC–The World Vegetable Center mungbean (*Vigna radiata*) core and mini core collections. *BMC Genom* 16(1):344
- Schafleitner R, Huang SM, Chu SH, Yen JY, Lin CY, Yan MR, Krishnan B, Liu MS, Lo HF, Chen CY, Long-fang OC (2016) Identification of single nucleotide polymorphism markers associated with resistance to bruchids (*Callosobruchus* spp.) in wild mungbean (*Vigna radiata* var. *sublobata*) and cultivated *V. radiata* through genotyping by sequencing and quantitative trait locus analysis. *BMC Plant Biol* 16(1):159
- Schierwater B, Ender A (1993) Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Res* 21:4647
- Selvi R, Muthiah AR, Manivannan N, Raveendran TS, Manickam A, Samiyappan R (2006) Tagging of RAPD marker for MYMV resistance in mungbean (*Vigna radiata* (L.) Wilczek). *Asian J Plant Sci* 5(2):277–280
- Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol Breed* 33(1):1–4
- Sharp PJ, Chao S, Desai S, Gale MD (1989) The isolation, characterization and application in the *Triticeae* of a set of wheat RFLP probes identifying each homoeologous chromosome arm. *Theor Appl Genet* 78(3):342–348
- Singh R, van Heusden AW, Yadav RC (2013) A comparative genetic diversity analysis in mungbean (*Vigna radiata* L.) using inter-simple sequence repeat (ISSR) and amplified fragment length polymorphism (AFLP). *Afr J Biotechnol* 12:6574–6582
- Solanki RK, Singh S, Kumar J (2010) Molecular marker assisted testing of hybridity of F₁ plants in lentil. *Food Leg* 23:21–24
- Sompong U, Somta P, Raboy V, Srinives P (2012) Mapping of quantitative trait loci for phytic acid and phosphorus contents in seed and seedling of mungbean (*Vigna radiata* (L.) Wilczek). *Breed Sci* 62(1):87–92

- Somta P, Seehalak W, Srinives P (2009) Development, characterization and cross-species amplification of mungbean (*Vigna radiata*) genic microsatellite markers. *Conserv Genet* 10(6):1939
- Spooner D, van Treuren R, de Vicente MC (2005) Molecular markers for genebank management. IPGRI technical Bulletin No. 10. International Plant Genetic Resources Institute, Rome, Italy
- Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL (2011) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet* 5(11): e1000734
- Srinives P, Kitsanachandee R, Chalee T, Sommanas W, Chanprame S (2010) Inheritance of resistance to iron deficiency and identification of AFLP markers associated with the resistance in mungbean (*Vigna radiata* (L.) Wilczek). *Plant Soil* 335:423–437
- Steinmetz M, Minard K, Harvath S, McNicholas J, Srenlinger J, Wake C, Long E, Mach B, Hood L (1981) A molecular map of the immune response region of the major histocompatibility complex of the mouse. *Nature* 300:35–42
- Sun L, Cheng XZ, Wang SH, Wang LX, Liu CY, Li ME, Ning XU (2008) Heredity analysis and gene mapping of bruchid resistance of a mungbean cultivar V2709. *Agric Sci China* 7(6):672–677
- Storm N, Darnhofer-Patel B (2003) MALDI-TOF mass spectrometry-based SNP genotyping. In: *Single nucleotide polymorphisms*. Springer, Totowa, NJ, pp 241–262
- Tangphatsornruang S, Somta P, Uthapaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrun S, Srinives P (2009) Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). *BMC Plant Biol* 9(1):137
- Tantasawat P, Trongchuen J, Prajongjai T, Thongpae T, Petkhum C, Seehalak W, Machikowa T (2010) Variety identification and genetic relationships of mungbean and blackgram in Thailand based on morphological characters and ISSR analysis. *Afr J Biotechnol* 9(27):4452–4464
- Tapp I, Malmberg L, Rennel E, Wik M, Syvanen AC (2000) Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5'-nuclease Taq-Man[®] Assay and molecular beacon probes. *Biotechniques* 28(4):732–739
- Thiel T, Kota R, Grosse I, Stein N, Graner A (2004) SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. *Nucleic Acids Res* 32(1):5
- Toojinda T, Chunwongse J, Tragoonrun S, Vanavichit A, Srinives P, Sommanas W (2001) Identification of AFLP marker linked to genes controlling iron deficiency tolerance in mungbean by bulked segregant analysis. In: *12. Genetics: Gene Revolution Era*, Bangkok (Thailand), 28–30 Mar 2001
- Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet* 120(7):1289–1299
- Van K, Kang YJ, Han KS, Lee YH, Gwag JG, Moon JK, Lee SH (2013) Genome-wide SNP discovery in mungbean by Illumina HiSeq. *Theor Appl Genet* 126(8):2017–2027
- Varshney RK, Pandey MK, Bohra A, Singh VK, Thudi M, Saxena RK (2018) Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor Applied Genet* 17:1–20
- Vos P, Hogers R, Bleeker M, Reijans M, Lee TV, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Res* 23(21):4407–4414
- Voss-Fels KP, Cooper M, Hayes BJ (2018) Accelerating crop genetic gains with genomic selection. *Theor Appl Genet* 19:1–8
- Wang X, Wang L (2016) GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Front Plant Sci* 7:1350
- Williams RC (1989) Restriction fragment length polymorphism (RFLP). *Am J Phys Anthropol* 32(S10):159–184
- Yashitola J, Thirumurugan T, Sundaram RM, Naseerullah MK, Ramesha MS, Sarma NP, Sonti RV (2002) Assessment of purity of rice hybrids using microsatellite and STS markers. *Crop Sci* 42(4):1369–1373
- Young ND, Danesh D, Menancio-Hautea D, Kumar L (1993) Mapping oligogenic resistance to powdery mildew in mungbean with RFLPs. *Theor Appl Genet* 87(1–2):243–249
- Young ND, Kumar L, Menancio-Hautea D, Danesh D, Talekar NS, Shanmugasundaram S, Kim DH (1992) RFLP mapping of a major bruchid resistance gene in mungbean (*Vigna radiata*, L. Wilczek). *Theor Appl Genet* 84(7–8):839–844
- Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the tm-2 locus of tomato during backcross breeding. *Theor Appl Genet* 77:353–359
- Zagorski N (2006) Profile of Alec J. Jeffreys. *Proceedings of the National Academy of Sciences of the United States of America*, vol 103(24): 8918–8920