Chapter 15

Molecular Plant Breeding: Methodology and Achievements

Rajeev K. Varshney, Dave A. Hoisington, Spurthi N. Nayak, and Andreas Graner

Summary

The progress made in DNA marker technology has been remarkable and exciting in recent years. DNA markers have proved valuable tools in various analyses in plant breeding, for example, early generation selection, enrichment of complex F_1s, choice of donor parent in backcrossing, recovery of recurrent parent genotype in backcrossing, linkage block analysis and selection. Other main areas of applications of molecular markers in plant breeding include germplasm characterization/fingerprinting, determining seed purity, systematic sampling of germplasm, and phylogenetic analysis. Molecular markers, thus, have proved powerful tools in replacing the bioassays and there are now many examples available to show the efficacy of such markers. We have illustrated some basic concepts and methodology of applying molecular markers for enhancing the selection efficiency in plant breeding. Some successful examples of product developments of molecular breeding have also been presented.

Key words: Molecular markers, Marker-assisted selection, Molecular breeding, Polymorphism, Linkage mapping, Association mapping, Marker–trait association.

1. Introduction

The identification of variation and its effective incorporation into germplasm are important components of any crop improvement programme. Such variation can be obtained from either crossing two different parental genotypes or selecting existing variation from the enormously available germplasm in the plant kingdom. Ancient farmers were the first 'plant breeders' by selecting the best plants for their needs. Archaeological evidence indicates that farmers employed selection pressure to meet their demands

Daryl J. Somers et al. (eds.), *Methods in Molecular Biology, Plant Genomics, vol. 513*

DOI: 10.1007/978-1-59745-427-8_15

as early as 12,000 years ago. As knowledge continues to grow, plant breeding has evolved as a major discipline in plant biology. There were many landmarks in plant breeding after the rediscovery of Mendelian genetics. Crossing two morphologically different parental genotypes allowed plant breeders to study the recombination and crossing-over events. Morphological markers played a major role in following genetics of the traits, for example flower colour, shape of the flower, seed size, seed colour, plant height, etc. Morphological markers are not always simple Mendelian-inherited genes, which has reduced their usefulness in plant breeding programmes.

There is enormous diversity (polymorphism) at the DNA level in higher plants, such that no two organisms are likely to be identical in their DNA sequence, including among natural populations of plants *(1)*. Molecular techniques have provided strategies to develop marker systems that detect such DNA variation, which can be used to assist traditional plant breeding *(2, 3)*. Once linkage between a marker locus and the gene for an agronomic trait of interest has been established, DNA-based tests can be used to enable more precise selection in plant breeding *(4, 5)*. This powerful revolution has already demonstrated its impacts in the understanding of, and ability to manipulate, oligogenic and quantitative traits. The development and availability of abundant, naturally occurring, molecular genetic markers during last two decades has generated renewed interest in locating and measuring the effects of genes (polygenes or QTLs – quantitative trait loci) controlling quantitative traits *(6)*.

Molecular markers are now well established as powerful tools in plant breeding and genetics for indirect selection of difficult traits at the seedling stage during plant breeding, thus speeding up the process of conventional plant breeding and facilitating the improvement of difficult traits that can not be improved easily by the conventional methods of plant breeding. In this direction, a large number of genes and QTLs controlling agronomic traits and conferring tolerance to both abiotic and biotic stresses have been identified and tagged using molecular markers in several crop species especially cereals *(7–9)*. In fact, the products of MAS have already been released as varieties in case of some cereal species *(10)*. Some notable examples of the successful deployment of MAS in some species have been listed in **Table 1**.In addition, several programmes and initiatives like molecular breeding programmes in wheat and barley in Australia *(12)* and 'MASWheat' (http://maswheat.ucdavis.edu/index.htm) are under way to conduct MAS in breeding. Molecular breeding strategies are also in use in several crops by several private companies, for example, Monsanto, Pioneer-HiBred, and Syngenta.

In the present genomics or post-genomic era when the sequence data have already become available through genome

Table 1
Some successful examples of molecular breeding in cereals

Achievement	Details	References
Acceleration in varietal development	1. Release of US barley variety Tango that contains two QTL for adult resistance to stripe rust	*(11)*
	2. Advancement of a 'Sloop type' variety with cereal cyst nematode (*CCN*) resistance for commercial release	Comparative Research Centre for molecular plant breeding (CRC-MPB) *(12)*
	3. Release of 'Flagship' variety in Australia in 2004 after following whole genome breeding approach	*(12)*
	4. Release of two Indonesian rice cultivars 'Angke' and 'Conde', in which marker-assisted selection (*MAS*) was used to introduce *xa5* into a background containing *xa4*	*(13)*
	5. Development of quality protein maize (*QPM*) through marker-aided transfer of *opaque2* gene in backcross programmes	*(14)*
	6. Release of an Indian pearl millet hybrid cultivar 'HHB 67-improved' in 2005, which has resistance to downy mildew	C.T. Hash, ICRISAT (personal communication)
	7. Development of an improved version of Pusa Basmati 1 (*PB1*) variety of rice after introgressing the genomic segments, harbouring the bacterial blight resistance namely *xa13* and *Xa21* have been transferred to PB1 from a non-Basmati donor through MAS	T. Mohapatra, NRCPB, IARI, India (personal communication)
Introgression of trait (gene pyramiding)	1. Introgression of *Yd2* gene conferring resistance to barley yellow dwarf virus (*BYDV*) into a BYDV-susceptible barley variety through two cycles of marker-assisted backcrossing	*(15)*
	2. Pyramiding of different resistance genes for barley yellow mosaic virus (*rym4*, *rym5*, *rym9*, and *rym11*) in barley	*(16)*
	3. Use of yield-related QTLs for MAS in maize in private sector	*(17)*
	4. Pyramiding of disease-resistant genes in rice, particularly against blight, blast, and both simultaneously	*(18–21)*

(continued)

Table 1 (continued)

Achievement	Details	References
	5. Pyramiding of insect and blight resistance in rice	*(19)*
	6. Pyramiding of blight resistance with Basmati quality characters in rice	*(22)*
	7. Pyramiding of stay green QTLs in elite but drought sensitive sorghum lines	C.T. Hash, ICRISAT (personal communication)

or EST sequencing projects for some plant species and similar efforts are under way for many other plant species, it has been possible to develop the molecular markers [and novel markers like single-nucleotide polymorphisms (SNPs) and single-feature polymorphisms (SFPs)] directly from genes *(23–25)*. Development of such functional markers may speed up in coming years as these markers will prove promising in marker-assisted breeding and useful resource for assessment of functional diversity in germplasm collection *(26, 27)*.

2. Materials

2.1. Molecular Markers

Molecular markers are specific locations on a chromosome which serve as landmarks for genome analysis (*see* **Note 1**). While selecting the molecular markers for marker–trait association studies, the following points need to be considered.

1. What are the relative costs of the marker assays versus other selection techniques such as phenotypic selection or various bioassay systems?
2. Is the trait dominant versus recessive?
3. What type and size of mapping population and which methodology will be used for marker–trait association studies?

2.2. Mapping Populations

The use of adequate genetic material for marker–trait association studies is another important critical factor (*see* **Note 2**). While doubled haploid (DH) and recombinant inbred line (RIL) populations are most appropriate genetic material for linkage map-based analysis, the F_2 and backcross (BC) populations can be used

for bulked segregant analysis (BSA) *(28)*. For using a particular mapping population for trait mapping, knowledge of the genetics for the trait, as listed below, is also important.

1. What is the nature of the trait? Is it simply inherited or multi-genic?
2. What is the heritability of the trait?
3. How much do the parental lines used to develop the population differ for the target trait?

An alternative methodology to the above-mentioned trait mapping strategies is association or linkage disequilibrium (LD) mapping, based on association between phenotype and allele frequencies (*see* **Note 3**). To effectively employ association mapping, one needs to decide on best population structure.

1. The structure of the population will be related to the trait and purpose.
2. Population structure will differ for (1) self versus outcrossing species, (2) long versus short generation species, and (3) perennial versus annual crop species.

After selecting a suitable mapping population, the population size is another criterion to be considered for trait mapping.

1. For mapping single genes, 50 F_2s/BCs should be adequate *(29)*.
2. For QTL analyses, a minimum of 200 individuals/lines of a population (RILs or DHs) are required *(30, 31)*.
3. For employing LD mapping, a population of at least 300 genotypes is required *(32)*.

2.3. Statistical Analysis and Tools

For conducting marker–trait association by using BSA and linkage maps, three widely used methods have been used: single marker analysis (SMA), simple interval mapping (SIM), and composite interval mapping (CIM) *(33, 34)* (*see* **Note 4**). In the case of LD or association mapping, the estimation of the LD within a species or even within individual genomes and an understanding of the structure of the population are important prerequisites. Subsequently, for conducting the marker–trait association studies, the structure of the population is considered to avoid false positives (*see* **Note 5**).

1. For performing the SMA, QGene *(35)* or MapManagerQTX *(36)* computer programmes can be used.
2. MapMaker/QTL *(37, 38)* and QGene can be used for SIM.
3. QTLCartographer *(39)*, MapManager QTX *(36)*, and PLABQTL *(40)* are the most appropriate computer programmes for conducting CIM.
4. In case of association mapping studies, the most commonly used computer programmes for measuring LD, population structure, and evaluating the trait associations are STRUCTURE *(41, 42)* and TASSEL *(43)*.

3. Methods

3.1. Marker–Trait Association

3.1.1. Selection of Molecular Markers

1. Select and optimize good quality, informative, and high-throughput amenable molecular markers, if possible.
2. Choose the appropriate genotyping platform depending on the size of the population to be studied as well as the number of available molecular markers, thereby per marker per individual experimental cost is minimized.

3.1.2. Polymorphism Survey for BSA and Linkage Map-Based Trait Mapping

1. Screen the parental genotypes with the molecular markers including 'anchor' markers (*see* **Note 6**).
2. Identify the markers that detect polymorphism between parental genotypes (**Fig. 1**).

3.1.3. Measuring LD Decay and Population Structure for Association Mapping

1. Select a *Discovery Panel* comprising of the diverse genotypes from the population to be used for association mapping and isolate the DNA from single plants to avoid heterogeneity.
2. Collect the data on nucleotide sequence from locus samples with genome-wide coverage from the *Discovery Panel*.
3. Measure range of diversity (e.g., decay of LD with physical distance – r^2) to be sampled for association population, marker density required for sufficient coverage of target genomic regions (or the genome) for association, level of population structure that exists within the species, evaluate genome-wide influence of demography, determine the genomic regions targeted by natural selection and domestication, and determine the number and density of the neutral markers required to evaluate background associations.

3.1.4. Genotyping

1. While conducting the BSA, genotype the bulks (two extremes of phenotype, 10–20 individual from each extreme) with the polymorphic markers, identify the putative associated markers with the trait, and subsequently genotype the complete population with the candidate markers (**Fig. 2**).
2. Genotype the mapping population (F_2, BC, RIL, DH) with the polymorphic markers in case of linkage map-based trait mapping strategy. Number of markers to be screened on the population depends on the genome size of the species. However, an average of 100–200 markers spaced less than 15 cM apart are recommended for linkage map-based QTL analysis. In case, the 'anchor' or 'core' markers are available for the species, use some of these markers, representing the arms of each linkage groups, to provide links with other linkage maps and trait information (**Fig. 3**).
3. For association mapping, sequence the target region(s) that are trait dependent across the association mapping population

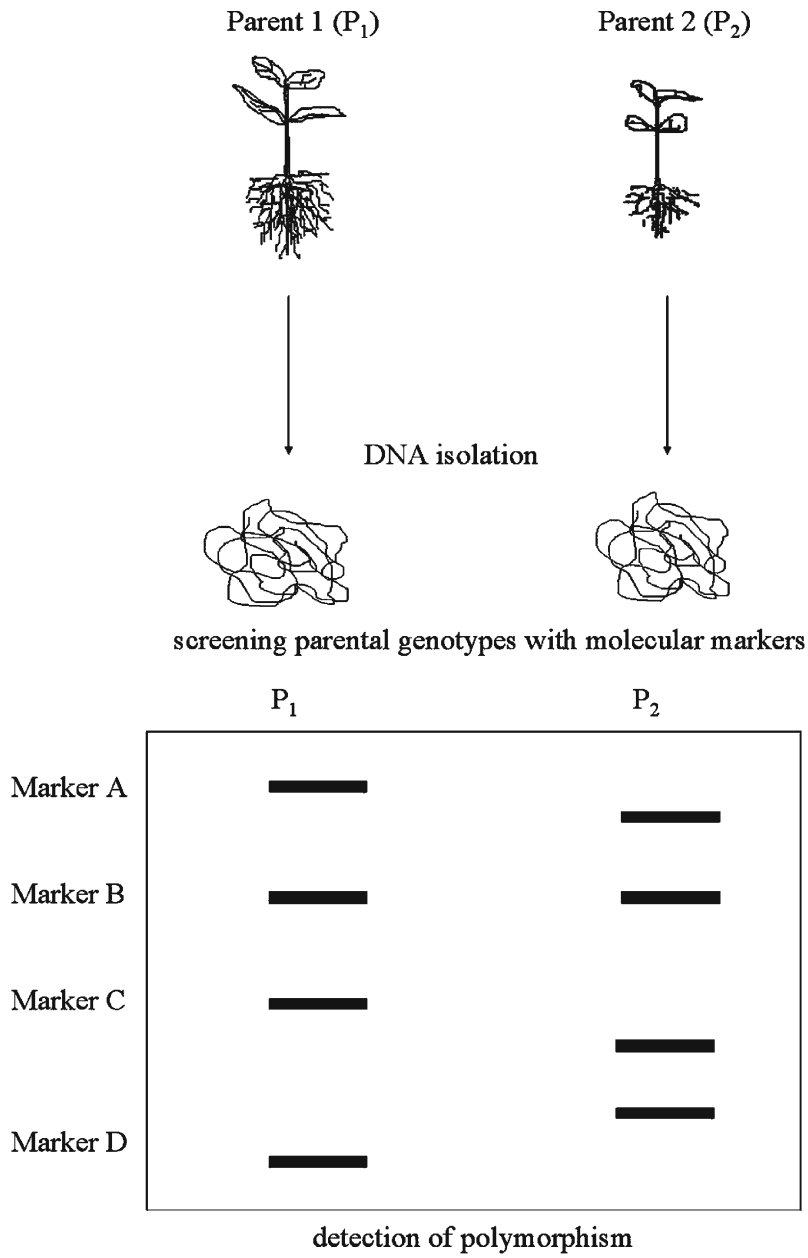

Fig. 1. Detection of DNA polymorphism. This figure shows detection of DNA polymorphisms between two homozygous parental genotypes P_1 and P_2 by using PCR-based microsatellite or SSR markers. Four hypothetical SSR markers (e.g., A, B, C and D in the figure) have been used for amplification of corresponding loci in two genotypes of interest. Separation of PCR products on agarose gel reveals polymorphism (size difference) between P_1 and P_2 for three SSR markers (A, C and D) while the marker B is monomorphic between two genotypes.

(candidate gene sequencing approach) or genotype the population with a suitable number of molecular markers covering the entire genome, depending on the LD decay (whole genome scanning approach). Genotype the population with

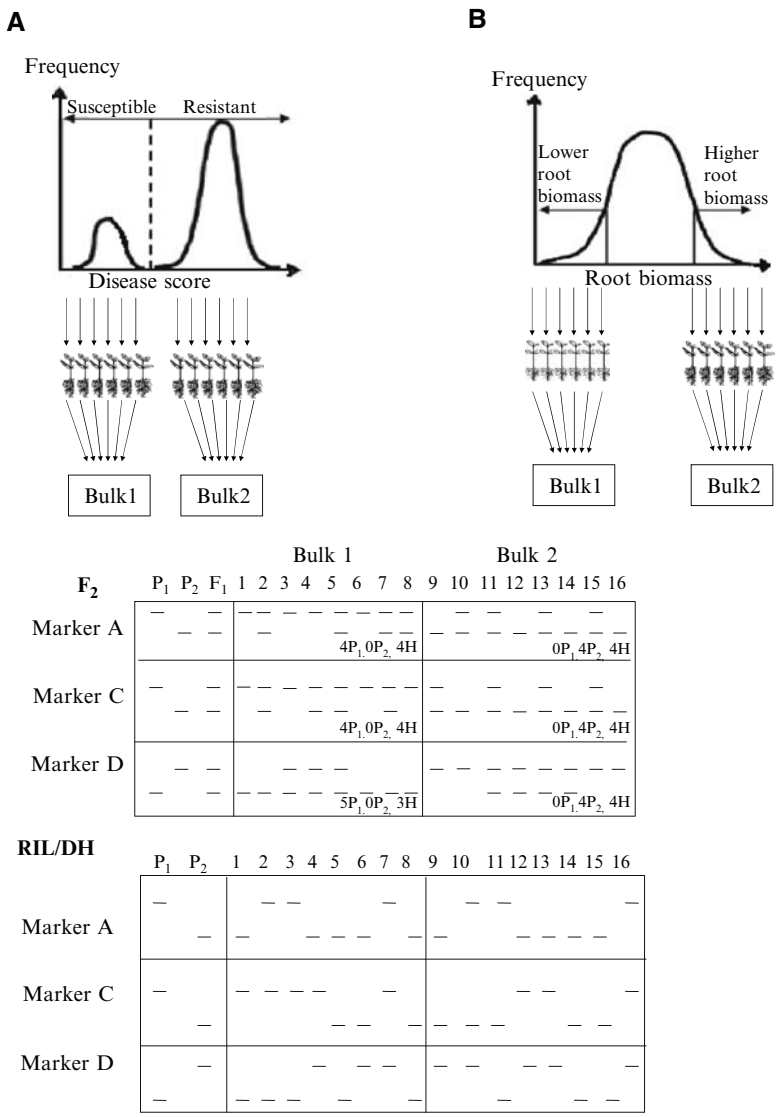

Fig. 2. Bulked segregant analysis (*BSA*) for simple/monogenic and quantitative traits. BSA can be used for both oligogenic trait, for example disease resistance (shown in **A**) and as well as quantitative trait, for example root biomass (shown in **B**). In both cases, two bulks are made from individuals from extreme phenotypes. The pooled DNA from these bulks together with the parental genotypes are screened with the molecular markers for the detection of polymorphism. The putative markers showing a polymorphism between parental genotypes and their characteristic amplification profiles (banding pattern) in corresponding bulks are then selected to screen on the DNA of individual lines of the bulks, and subsequently on the complete set of lines of the mapping population. A typical segregation banding pattern for three hypothetical polymorphic markers A, C and D in case of F_2 population (1:2:1) and recombinant inbred line/doubled haploid (*RIL/DH*) population (1:1) has been shown. The markers A and C in case of F_2 population (as majority of individuals of Bulk 1 and Bulk 2 for these markers reveal the alleles of the respective parents, P1 and P2 respectively) while the markers C and D in RIL/DH in the hypothetical examples, seem to be putative markers associated with trait. Therefore, these markers need to be screened on the complete mapping population or selective lines. Subsequently, the genotyping data obtained on the population together with the phenotyping data may be analyzed using appropriate statistical test (e.g., χ^2-test) for marker–trait association.

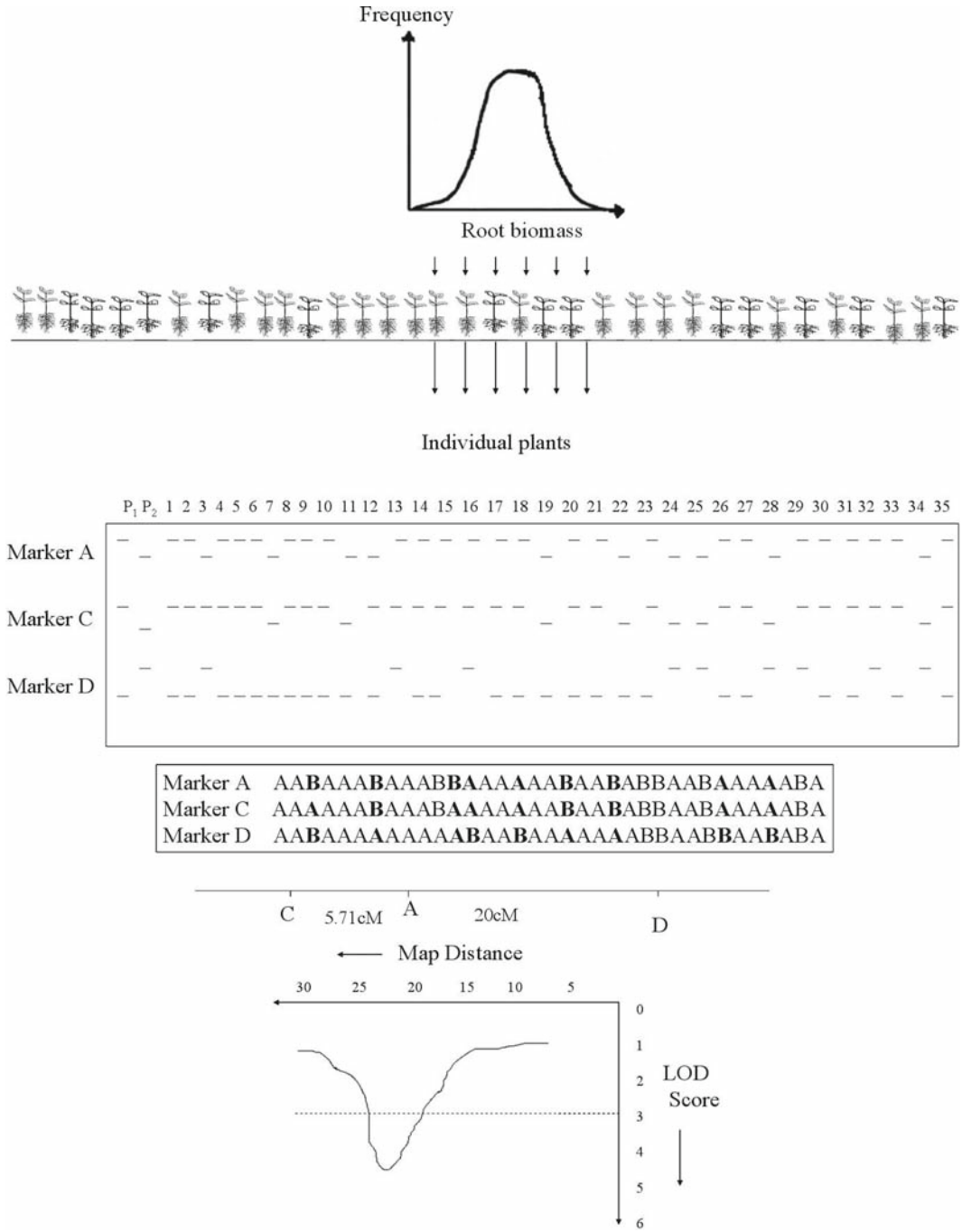

Fig. 3. Linkage mapping-based quantitative trait loci (*QTL*) analysis. For quantitative traits, the most commonly used approach of QTL analyses is based on linkage mapping by using recombinant inbred line (*RIL*) or doubled haploid (*DH*) population. Here polymorphic molecular markers (e.g., A, C and D in the figure) between the parental genotypes have been screened on the complete set of the mapping population. The genotyping data obtained so are used for calculation the recombination frequency between/among the markers and construction of the genetic map. In the figure, a hypothetical molecular linkage map based on three polymorphic markers (A,C and D) is shown. The linkage mapping data together with the phenotyping data on the progeny lines of the population are then used for QTL analysis using appropriate statistical analyses (e.g., regression analysis, interval mapping, and composite interval mapping). A hypothetical QTL peak for the trait on the linkage map, as prepared above, has been shown in the figure.

some neutral markers also in order to test the levels of background stochastic association (**Fig. 4**).

Experimental Details

For analyzing the polymorphism and genotyping the germplasm, several genotyping platforms like agarose gel electrophoresis, polyacrylamide gel electrophoresis, capillary electrophoresis, etc are available. Each of these platforms has some advantages and disadvantages over the others. Nevertheless, there is a need for high throughput, robust, and cost-effective genotyping platforms for

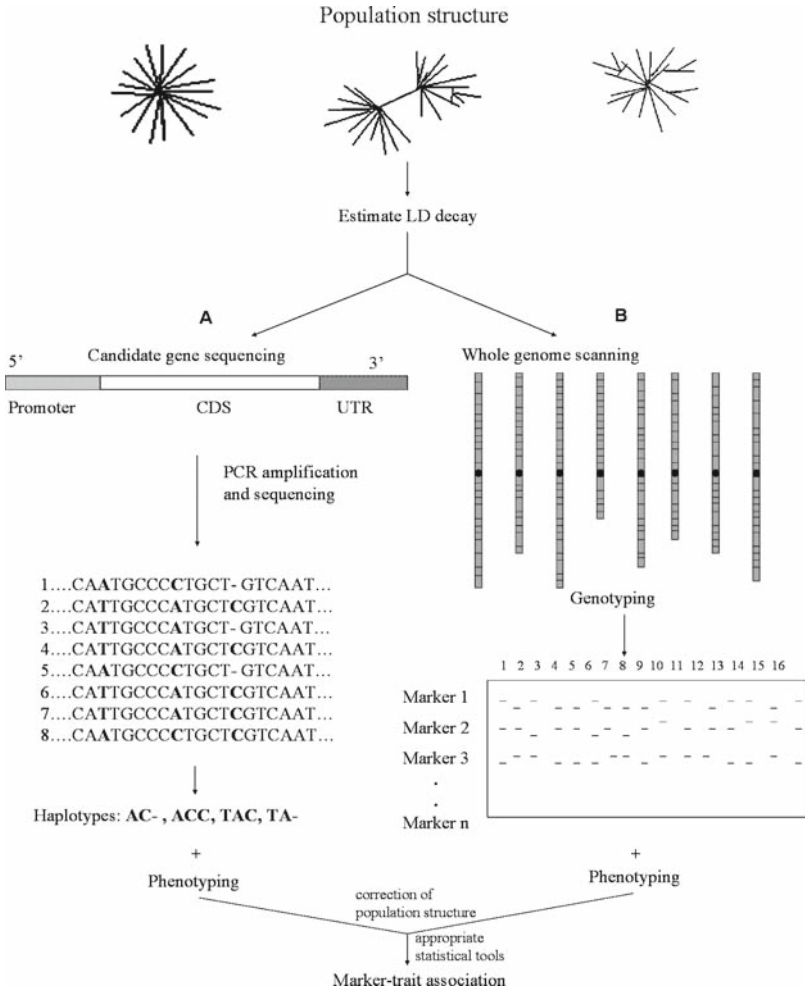

Fig. 4. Linkage disequilibrium (*LD*)-based association mapping for marker–trait association. In the first instance, the germplasm to be used is screened with neutral DNA markers for estimating the population structure and the LD decay in the genome and germplasm. On the basis of these analyses, one of two approaches, that is, candidate gene sequencing (**A**) and whole genome scanning (**B**) is used. In candidate gene sequencing approach, the putative candidate genes contributing the phenotypic variation for the trait are selected and appropriate genic region(s) are sequenced across the germplasm and sequence data are analyzed into haplotypes. In the whole genome scanning approach, the germplasm is screened with the molecular markers, representing the whole genome, based on LD decay study. Subsequently, the haplotype or genotyping data obtained is analyzed together with the phenotyping data with appropriate statistical tools to correct the population structure and marker–trait association.

molecular breeding. Currently, at majority of the places including ICRISAT, capillary electrophoresis platform is used for marker analysis. Therefore, in this section, technical details on analyzing the markers in germplasm using capillary electrophoresis, as we are doing at ICRISAT on ABI3130 and ABI3700, are provided.

1. Set up the PCR reaction as per the optimized conditions to get the amplicons by using fluorescent dye-labelled primers. The forward primers are labelled with one of the fluorescent dyes: Fam (blue), Vic (green), Ned (yellow), and Pet (red). For instance, for setting up the PCR in 10-µl volume, take 1 µl of DNA (2.5 ng/µl) in a PCR tube or a 96-well or 384-well plate and add 1 µl of dNTP (2 mM), 1 µl of 10× Qiagen buffer, 0.25 U of Taq polymerase (e.g., Qiagen), 0.5 µl of labelled forward primer, 0.5 µl of reverse primer, and make volume to 10 µl by adding sterile distilled water. Sometimes, multiplex PCR by using more than one primer pair labelled with different fluorescent dyes can be set up.
2. Put the tube or plate in a thermal cycler and run the PCR using the touchdown PCR profile. For example, for a primer pair with annealing temperature 55°C, set up an initial denaturation at 94°C for 3 min followed by 5 cycles consisting of denaturation at 94°C for 20 s, annealing at 60°C for 30 s with 1°C decrease in every subsequent cycle and extension at 72°C for 30 s and then 30 cycles consisting of denaturation at 94°C for 20 s, annealing at 55°C for 30 s and extension at 72°C for 30 s with the final extension step at 72°C for 20 min.
3. Check the amplification using 2 µl PCR product on 1.2% agarose gel electrophoresis.
4. If multiplexing PCR was not set up in step 1, the PCR products obtained from individual primer pairs labelled with different fluorescent dyes sometimes can be pooled. In such cases, the pooling premix includes 1 µl of each PCR product (s), 7 µl of Hi-Di formamide for denaturing the double-stranded DNA, and 0.2 µl of GeneScan Liz 500 internal lane size standard (Orange) provided by Applied Biosystems (USA). Alternatively, 0.15 µl of GeneScan Rox 500 or Rox HD 400 size standard (red) can be used wherever PET label is not used in the pooled PCR products.
5. Denature the samples at 94°C for 5 min and cool immediately on ice.
6. Put the samples in machine and run capillary electrophoresis.
7. Import the raw allele size data to GeneScan programme for assigning the allele sizes based on the internal size standard. For example, if Liz 500 is used as the internal lane size standard, the PCR products in the range of 35–500 bp can be sized. The sizes for the Liz 500 are 35, 50, 75, 100, 139, 150, 160,

200, 250, 300, 340, 350, 400, 450, 490, and 500 bp. Each of the DNA fragment is labelled with Liz flourophore. Define the analysis parameters based on the amplitude of the PCR product. During the analysis, based on internal size standard, all the peaks of amplified PCR products can be assigned to their proper sizes.

8. Import the GeneScan output file to Genotyper programme and set the preferences according the dyes used. Define the category according to the primer name; select the highest peak and the range of intensity of amplicon to consider for the analysis. The genotyping files appear according to the primer name for all the genotypes screened for polymorphism.
9. Peaks can be labelled as size (bp), peak height, and peak area depending on the user requirement. Inspect some of all the peaks manually to be more confident on allele sizing.
10. Upload the results of genotyping in a table with pre-defined columns. Export this table and save it at computer for application in molecular breeding.

3.1.5. Phenotyping

1. Phenotype the population for trait of interest. This should be replicated temporally and spatially to increase the accuracy and precision of the phenotypic measurements.
2. If possible, measure the trait of interest in quantitative fashion instead of categorically.
3. Evaluate the trait heritability to define the expectation for the genetic component of the phenotypic variance.

3.1.6. Statistical Association

Genotyping data obtained on the population are analyzed together with the replicated phenotyping data using appropriate statistical analysis and tools, as mentioned in **Subheading 2.3**.

Bulked Segregant Analysis

1. Use single-point analysis method involving *t*-test, analysis of variance (ANOVA), and linear regression.
2. Calculate the phenotypic variation arising from the QTL linked to the marker (coefficient of determination, R^2).

Linkage Mapping

1. Use the genotyping data to construct a framework genetic linkage map with a suitable computer programme mentioned in **Subheading 2.3**.
2. Use the anchor markers and their mapping positions to designate the linkage groups (**Fig. 3**).
3. Conduct the SMA, SIM, or CIM for identification of QTLs.
4. Calculate the phenotypic variation (R^2) contributed by the QTL and measure the QTL × QTL, QTL × E, and QTL × QTL × E interaction.

Association Mapping

1. Build statistical model(s) for the expectation of phenotypic correlation with environmental and genetic variability ($V_P = V_G + V_E$).
2. Evaluate the level of co-variance between the phenotypes and combine the highly correlated traits in the same model.
3. Evaluate co-variance between the neutral marker genotypes and candidate gene genotypes, in case candidate gene-based approach is being used.
4. Determine the Type I error thresholds according to the number of tests performed and the level of flexibility in the study.
5. Determine power and false positive rate expectations for the study.
6. Run statistical association tests using appropriate statistical tool/software (e.g., TASSEL).

3.2. Marker Validation

Generally, the markers/alleles associated with the trait should be validated by testing their effectiveness in determining the target phenotype in independent populations and different genetic backgrounds, which is referred to as 'marker validation' *(44, 45*, *see* **Note 7**).

1. Confirm the QTL mapping studies by using independent populations constructed from the same parental genotypes or closely related genotypes used in the primary QTL mapping study. Larger population sizes may be used.
2. If suitable genetic material for the trait is available, for example, near isogenic lines (NILs), genotype the NILs with candidate markers and compare mean trait values of particular NILs with the recurrent parent, the effects of QTLs can be confirmed.
3. Test the presence of the marker, associated with the QTL, on a range of cultivars and other genotypes. There is no guarantee that DNA markers identified in one population will be useful in different population, especially when the populations originate from distantly related germplasm.
4. Markers that reveal polymorphism in different populations derived from a wide range of different parental genotypes will be most useful in breeding programmes *(45)*.
5. If the candidate gene sequencing approach has been used in LD/association mapping, in addition to above, the association of allele with the trait may be verified either through re-evaluation in an independent population sample or through allelic-silencing or knockout studies.

3.3. Marker-Assisted Selection

Once markers that are tightly linked to genes or QTLs of interest have been identified, prior to field evaluation of large number of plants, plant breeders may use specific DNA marker alleles as

a diagnostic tool to identify plants carrying the genes or QTLs. The procedure is called 'marker-assisted selection' or 'marker-aided selection' (commonly referred as MAS) or 'marker-assisted breeding'.

1. Select the markers that are tightly linked with the trait and yield the robust and clear-cut banding pattern. In principle, all such markers may be used in MAS; however, there have been reports of up to five QTLs being introgressed.
2. For early generation selection in typical breeding programme for simple (monogenic) traits, for example, disease resistance, a susceptible parent is crossed with a resistant parent and the F_1 plant is self-pollinated to produce an F_2 population. Use the robust marker, developed for the major gene/QTL controlling trait of interest (e.g., disease resistance) to screen the F_2 population and select only those plants possessing the desirable genotypes (having the alleles conferring resistance) out of the large number (e.g., 2,000) of F_2 plants (**Fig. 5**). It is estimated that up to 75% of plants may be eliminated after one cycle of MAS.
3. In case of marker-assisted backcrossing programme, use the molecular markers associated with the trait for foreground selection, while neutral markers covering the whole genome can be used for background selection (**Fig. 6**). For example, genotype the BC_1F_1s (by crossing the donor genotype for a trait with the F_1s, obtained by crossing the donor and the elite-recipient genotypes) with the molecular markers associated with the trait as well as other neutral markers. The lines possessing the desirable genotype based on the associated markers as well as having higher proportion of genome (fingerprints) of the recipient genotype should be selected and advanced to generate the BC_2F_2s. Similar kind of foreground and background selection with molecular markers can be conducted in next generations until the backcrossing products have the desirable chromosomal segment from the donor genotype into the recipient genotype background.

Notes

1. DNA-based molecular markers can be classified into three categories depending on how the polymorphism is revealed: hybridization-based polymorphisms, PCR-based polymorphisms, and sequence-based polymorphisms. Details about these marker systems are discussed in several reviews and book

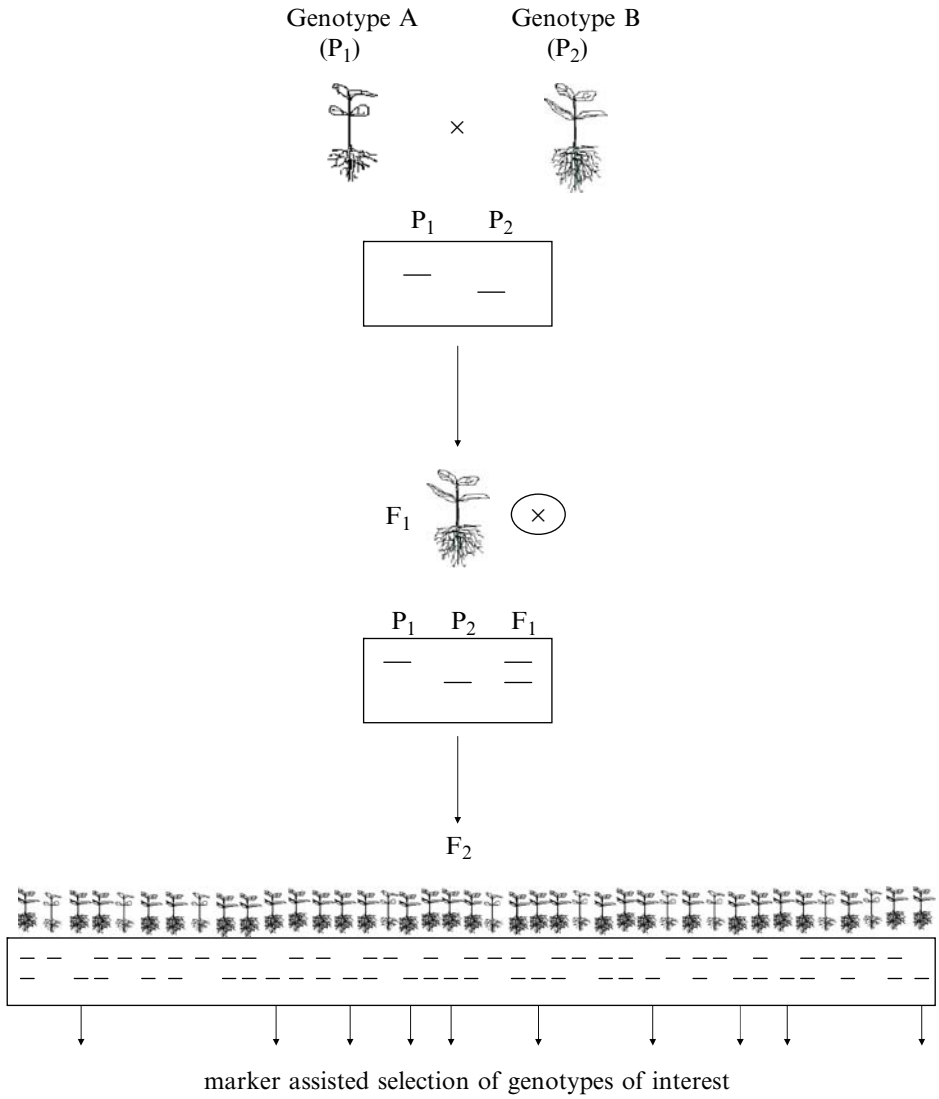

Fig. 5. A hypothetical scheme for early generation selection in a breeding programme. The polymorphic molecular markers between two parental genotypes (P_1 and P_2) can be used to select true F_1 by analyzing co-dominant markers having the alleles (bands) of both parental genotypes. After selfing the F_1 lines, several hundreds (sometimes thousands) F_2 lines are raised. These F_2 lines can be screened with the polymorphic co-dominant markers and based on DNA profiling the progenies of interest, for example the lines having the allele of parental genotype P_2 (with higher root biomass) in the figure have been selected.

chapters *(5, 46, 47)*. The choice of using molecular markers depends on the intended use, the microsatellite or simple sequence repeats (SSRs), however, have been recommended for molecular breeding as they are co-dominant, multi-allelic, and abundant in nature *(48)*. For detection of microsatellite loci in the genome, two PCR-based primer pairs based on flanking regions of the microsatellite are used. For majority of

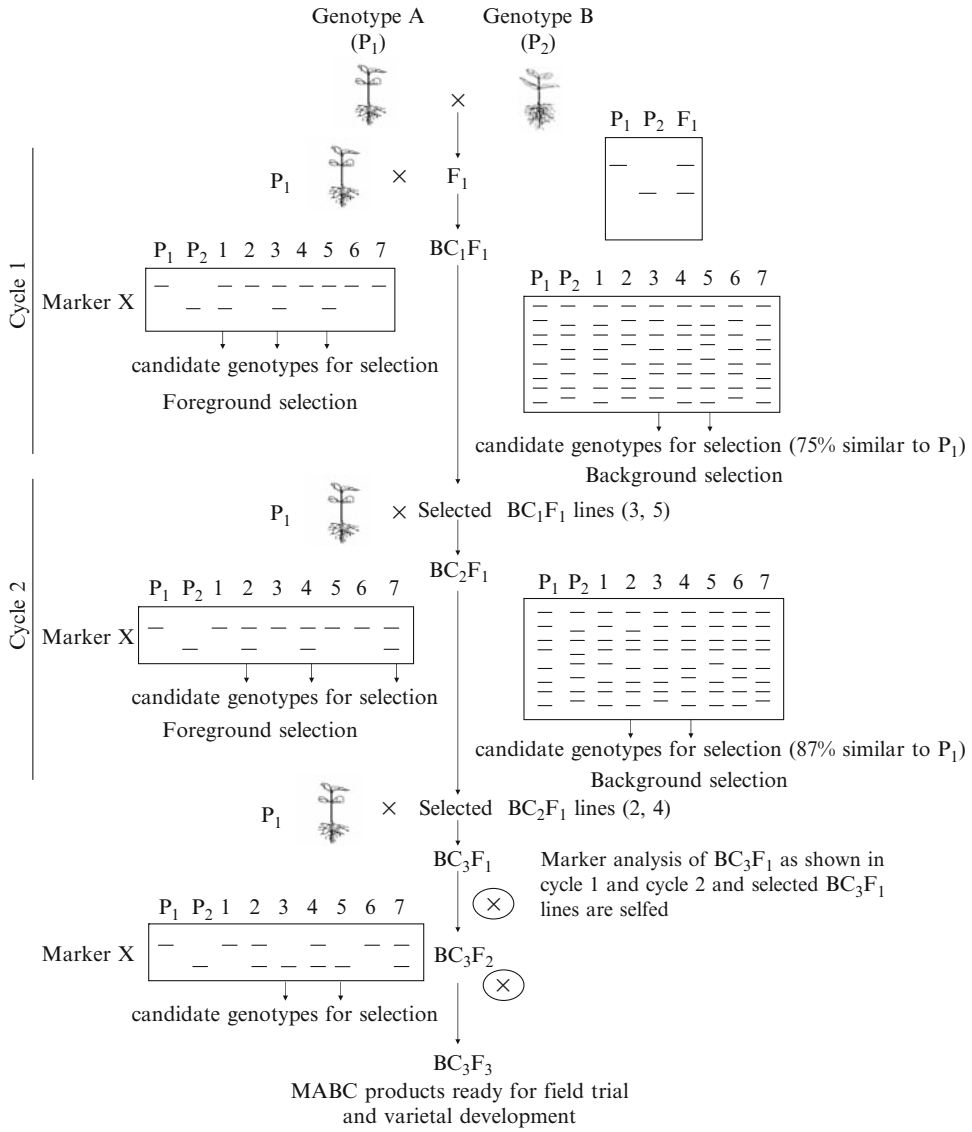

Fig. 6. Marker-assisted backcrossing using foreground and background selection strategies. One of the most important applications of molecular markers to introgress the trait of interest in a genotype of interest has been shown in the figure. For example, marker–trait association studies provide the codominant marker X as a diagnostic or linked molecular marker with a major root trait quantitative trait loci (*QTL*), contributing large phenotypic variation. For this marker, the lower band (allele B) is associated with higher root biomass while the upper band (allele A) is associated with lower root biomass. In case, the genotype A is an elite variety but drought sensitive and the breeder likes to introgress the root trait QTL for higher root biomass in this genotype. The breeder needs to select a genotype (e.g., B) with higher root biomass. The genotype A is crossed with the genotype B and the resulting F_1 lines can be screened with the marker X to confirm the presence of both alleles and then the true F_1 lines can be advanced. These F_1 lines will be backcrossed with the genotype A and the resulting BC_1F_1 lines can be screened with the diagnostic molecular marker X for foreground selection while with the multi-locus marker system(s) for the background selection. The foreground selection (identification of lines with the higher root biomass allele in heterozygous condition), in the figure, suggests to select the line nos. 1, 3, and 5. In parallel, these lines are analyzed for monitoring the genomic background of the recipient (A) genotype. In the figure, out of the three lines, 1, 3, and 5, only two lines, lines no. 3 and 5 fulfil the criteria of foreground (having allele of donor genotype in heterozygous condition) and background (having 75% of genome of the recipient genotype) selection.

the major crop species, these SSR markers are present in large numbers in public domain.

2. In general, for trait mapping based on linkage maps or BSA, several types of mapping populations, derived from crosses involving any two diverse parents, can be used. For instance, an F_2 population or BC population can be derived from F_1 plants through selfing or backcrossing them to one of the parents; RILs can be derived by single seeds descent for at least five or more generations; and DHs can be derived from haploid obtained from F_1 plants through anther/egg cell/ovule culture or distant hybridization. The simplest mapping populations are the F_2 populations or the BC populations; however, these mapping populations are not permanent, while the RILs and DHs are immortal populations and can be stored and shared across the laboratories.
3. Association or LD mapping is the basis for gene mapping in species where large mapping populations can not be readily produced such as mapping in tree species, farm animals, and humans *(49, 50)*. The LD method, unlike the use of biparental mapping populations, uses real breeding populations; the material is diverse and relevant; and the most important genes (e.g., for adaptation) should be co-segregating in such populations *(51, 52)*.
4. Among different statistical analyses for QTL mapping, SMA (or single-point analysis) is the simplest method for detecting QTLs associated with single markers. The statistical methods used for SMA include *t*-tests, ANOVA, and linear regression. Linear regression is most commonly used because the coefficient of determination (R^2) from the marker explains the phenotypic variation arising from the QTL linked to the marker. In fact, this method is generally used in BSA approach for trait mapping. However, the main disadvantages of this method are: (1) the further a QTL is from a marker, the less likely it will be detected as the recombination may occur between the marker and the QTL; (ii) this causes the magnitude of

Therefore, these lines are advanced for further backcrossing. The figure shows the backcrossing of one selected BC_1F_1 line with the A genotype and the resulting BC_2F_1 lines are further screened for foreground selection (with the diagnostic X marker) and the background selection (multi-locus fingerprinting). As a result of the second foreground and background selection, two BC_2F_1 lines (lines no. 2 and 4) are selected that show the allele of higher root biomass (in heterozygous condition) and 87% genome of the donor genotype. Such kind of backcrossing and foreground and background selection are continued upto 3-4 cycles, depending on the nature of the crop species. At the end of MABC, the progeny lines are analyzed using diagnostic marker X, and the plants carrying higher root biomass allele are selected. Subsequently, selfing of selected progeny lines is undertaken till the appropriate marker-assisted back cross (*MABC*) lines with the higher root biomass alleles in the genomic background of the recipient genotype are generated. These lines, eventually, can be taken to the field trials and then other requirements of varietal development can be followed.

the effect of a QTL to be underestimated. The use of a large number of segregating markers covering the entire genome, usually at intervals less than 15 cM, may minimize both problems *(33)*.

Linkage map-based trait mapping approach employs the SIM method that makes use of linkage maps and analyses intervals between adjacent pairs of linked markers along chromosomes simultaneously *(53)*. The use of linked markers for analyses under SIM is considered statistically more powerful compared to single-point analysis as the recombination between the markers and the QTL is taken care of *(34)*.

The CIM approach, however, combines interval mapping with linear regression and includes additional molecular markers in the statistical model in addition to an adjacent pair of linked markers for interval mapping *(54)*. This method is more precise and effective at mapping QTLs as compared to single-point analysis (SMA) and SIM, especially when linked QTLs are involved.

5. While using the association or LD mapping approach, the statistical power of associations is determined by the extent of LD with the causative polymorphism, as well as sample size used for the study *(55, 56)*. The decay of LD over physical distance in the study population determines the marker density required and the level of resolution that may be obtained in an association study. The most commonly used summary statistic for estimation of LD within the association study framework is known as r^2 *(57, 58)*. The r is the Pearson's (product moment) *correlation coefficient* of the correlation that describes the predictive value of the allelic state at one polymorphic locus on the allelic state at another polymorphic locus, where r^2 is the squared value of correlation coefficient that is also called *coefficient of determination* and it explains the proportion of a sample variance of a response variable that is *explained* by the predictor variables when a linear regression is performed *(50)*. *Lewontin's D* is another summary statistic for LD that is commonly used and describes the difference between the coupling gamete frequencies and repulsion gamete frequencies at two loci. From D, a second measure of LD, that is, normalized D' can also be estimated. It is important to estimate the rate of decay of LD with physical distance, to be able to extrapolate information gathered from a small collection of sampled loci to the whole genome investigated. This extrapolation is essential for association mapping study design, since it may be used for determining the marker density required for scanning previously unexplored regions of the genome, as well as determining the maximum resolution that can be achieved for genotype–phenotype associations for the study population.

Another important constraint for the use of association mapping for crop plants is unidentified population sub-structuring and admixture due to factors such as adaptation or domestication *(43, 59)*. Population structure creates genome-wide LD between unlinked loci. When the allele frequencies between sub-populations of a species are significantly different, due to factors such as genetic drift, domestication, or background selection, genetic loci that do not have any effect whatsoever on the trait may demonstrate statistical significance for their co-segregations with a trait of interest (*see* ref. *50* for details). In cases where the population structuring is mostly due to population stratification *(41, 60)*, three methods are often proposed suitable for statistically controlling the effects of population stratification on association tests: (a) genomic control (GC) *(61–63)*, (b) structured association (SA) method including two extensions that are modified for the type of association study as case-control (SA-model) *(42)* or quantitative trait association study (Q-model) *(32, 43)*, and (c) unified mixed model approach (Q + K) *(64)*.

After analyzing the LD decay, population structure, and appropriate genotyping of the population, marker–trait association studies are conducted. Whether the phenotype of interest has a binary or quantitative phenotype is also of interest for the association study design. When a binary trait is being investigated, case-control type populations are required for association analysis, where equivalent sized sub-populations of individuals that display the phenotype of interest (cases) and do not display the phenotype of interest (controls) are queried for allelic association of genetic loci with the case and control phenotypes in a statistically significant manner *(50)*. The statistical test performed is simply a hypothesis test that asks weather the allelic frequency distribution of a locus is the same or different for a given locus between the two sub-populations. Most of the statistical methods aim to detect and correct for the effects of population stratification and ancestry differences between the case and control groups *(50, 65)*.

6. When the same set of molecular markers is used in different mapping populations of the given species to construct the linkage maps, the markers order and the linkage maps can be correlated. Therefore, in order to correlate information from one map to another, common markers are required. Common markers that are highly polymorphic in different mapping populations are called ‘anchor’ or ‘core’ markers. Generally, anchor markers are SSRs or RFLPs *(31)*.
7. The marker validation involves testing the reliability of markers to predict phenotype and indicates whether a marker could be used in routine screening for MAS.

Acknowledgements

Authors (RKV, DAH, SNN) are grateful to Generation Challenge Programme (GCP) of Consultative Group on International Agriculture Research (CGIAR), Indo–US Agricultural Knowledge Initiative (AKI), and National Fund of Indian Council of Agricultural Research (ICAR) and Department of Biotechnology (DBT), Government of India for financial support of their research. SNN is thankful to Council of Scientific and Industrial Research (CSIR), Government of India for sponsoring fellowship.

References

1. Gur, A. and Zamir, D. (2004) Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol.* 2, e245.
2. Tanksley, S.D., Young, N.D., Paterson, A.H., and Bonierbal1e, M.W. (1989) RFLP mapping in plant breeding: new tools for an old science. *BioTechnology* 7, 257–264.
3. Phillips, R.L. and Vasil, I.K. (eds.) (2001) *DNA-Based Markers in Plants (2nd ed)*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
4. Rafalski, J.A. and Tingey, S.V. (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet.* 9, 275–280.
5. Azhaguvel, P., et al. (2006) Methodological advancement in molecular markers to delimit the gene(s) for crop improvement, in *Floriculture*, *Ornamental and Plant Biotechnology: Advances and Topical Issues (Vol. I)* (Teixeira da Silva, J.A., ed.), Global Science Books, London, pp. 460–469.
6. Gupta, P.K. and Varshney, R.K. (2004) Cereal genomics: an overview, in *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 1–18.
7. Jahoor, A., Eriksen, L., and Backes, G. (2004) QTLs and genes for disease resistance in barley and wheat, in *Cereal genomics* (Gupta, P.K. and Varshney, R. K., eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 199–252.
8. Li, W. and Gill, B.S. (2004) Genomics for cereal improvement, in *Cereal genomics* (Gupta, P.K. and Varshney, R.K., eds.). Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 585–634.
9. Tuberosa, R. and Salvi, S. (2004) QTLs and genes for tolerance to abiotic stress in cereals, in *Cereal genomics* (Gupta, P.K. and Varshney, R.K., eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 253–315.
10. Varshney, R.K., Hoisington, D.A., and Tyagi, A.K. (2006) Advances in cereal genomics and applications in crop breeding. *Trends Biotechnol.* 24, 490–499.
11. Toojinda, T., Baird, E., Booth, A., Broers, L., Hayes, P., Powell, W., Thomas, W., Vivar, H., and Young G. (1998) Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker assisted line development. *Theor. Appl. Genet.* 96, 123–131.
12. Langridge, P. (2005) Molecular breeding of wheat and barley, in *In the Wake of Double Helix: From the Green Revolution to the Gene Revolution* (Tuberosa R., Phillips, R.L., and Gale, M., eds.), Avenue Media, Bologna, Italy, pp. 279–286.
13. Toenniessen, G.H., O'Toole,J.C., and DeVries, J. (2003) Advances in plant biotechnology and its adoption in developing countries. *Curr. Opin. Plant Biol.* 6, 191–198.
14. Dreher, K., Morris M., and Khairallah, M. (2000) Is marker assisted selection cost-effective compared to conventional plant breeding methods? The case of quality protein maize, in *Proc 4th Annu Conf Intern Consor on Agricultural Biotechnology Research (ICABR), The Economics of Agricultural Biotechnology*, Ravello, Italy.
15. Jefferies, S.P., King, B.J., Barr, A.R., Warner, P., Logue, S.J., and Langridge, P. (2003) Marker-assisted backcross introgression of the yd2 gene conferring resistance to barley yellow dwarf virus in barley. *Plant Breed.* 122, 52–56.
16. Friedt, W. and Ordon, F. (2007) Molecular markers for gene pyramiding and disease resistance breeding in barley, in *Genomics Assisted Crop Improvement* (Varshney R.K. and Tuberosa, R.T., eds.), Springer, The Netherlands, (Vol. II) pp. 97–120.

17. Koebner, R.M.D. (2004) Marker assisted selection in the cereals: the dream and the reality, in *Cereal Genomics*, (Gupta P.K. and Varshney, R.K., eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 317–329.
18. Sanchez, A.C., Brar, D.S., Huang, N., Li, Z. K., and Khush, G.S. (2000) Sequence-tagged site marker-assisted selection for three bacterial blight resistance genes in rice. *Crop Sci.* 40, 792–797.
19. Singh, S., Sidhu, J.S., Huang, N., Vikal,Y., Li, Z., Brar, D.S., Dhaliwal, H., and Khush, G.S. (2001) Pyramiding three bacterial blight resistance genes (*xa-5*, *xa-13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. *Theor. Appl. Genet.* 102, 1011–1015.
20. He, Y., Li, X., Zhang, J., Jiang, G., Liu, S., Chen, S., Tu, J., Xu, C., and Zhang, Q. (2004) Gene pyramiding to improve hybrid rice by molecular-marker techniques, in *New Directions for a Diverse Planet*: *Proc. 4th Intern. Crop Sci. Cong.* Brisbane, Australia. (http://www.cropscience.org.au/icsc2004/)
21. Narayanan, N.N., Baisakh, N., Vera Cruz, C.M., Gnanamanickam, S.S., Datta, K., and Datta, S.K. (2002) Molecular breeding for the development of blast and bacterial blight resistance in rice cv. IR50. *Crop Sci.* 42, 2072–2079.
22. Joseph, M., Gopalakrishnan, S., Sharma, R.K., Singh, V.P., Singh, A.K., Singh, N.K., and Mohapatra, T. (2004) Combining bacterial blight resistance and Basmati quality characteristics by phenotypic and molecular-marker assisted selection in rice. *Mol. Breed.* 13, 377–387.
23. Rostoks, N., Borevitz, J.O., Hedley, P.E., Russell, J., Mudie, S., Morris, J., Cardle, L., Marshall, D.F., and Waugh, R. (2005) Single-feature polymorphism discovery in the barley transcriptome. *Genome Biol.* 6, R54.
24. Rostoks N., Schmierer, D., Mudie, S., Drader, T., Brueggeman, R., Caldwell, D.G., Waugh, R., and Kleinhofs, A. (2006) Barley necrotic locus nec1 encodes the cyclic nucleotide-gated ion channel 4 homologous to the Arabidopsis HLM1. *Mol. Genet. Genomics* 275, 159–168.
25. Varshney, R.K., Graner, A., and Sorrells, M.E. (2005b) Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10, 621–630.
26. Andersen, J.R. and Lübberstedt, T. (2003) Functional markers in plants. *Trends Plant Sci.* 8, 554–560.
27. Varshney, R.K., Graner, A., and Sorrells, M.E. (2005a) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23, 48–55.
28. Michelmore, R.W., Paran, I., and Kesseli, R.V. (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. USA* 88, 9828–9832.
29. Young, N.D. (1994) Constructing a plant genetic linkage map with DNA markers, in *DNA-Based Markers in Plants* (Vasil, I.K. and Phillips, R.L., eds.), Kluwer, Dordrecht, pp. 39–57.
30. Mohan, M., Suresh, N., Bhagwat, A., Krishna, T.G., Yano, M., Bhatia, C.R., and Sasaki, T. (1997) Genome mapping, molecular and makers and marker-assisted selection in crop plants. *Mol. Breed.* 3, 87–103.
31. Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B., and Pang, E.C.K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–196.
32. Camus-Kulandaivelu, L., Veyrieras, J.B., Madur, D., Combes, V., Fourmann, M. Barraud, S., Dubreuil, P., Gouesnard, B., Manicacci, D., and Charcosset, A. (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the *Dwarf8* gene. *Genetics* 172, 2449–2463.
33. Tanksley, S.D. (1993) Mapping polygenes. *Annu. Rev. Genet.* 27, 205–233.
34. Liu, B. (1998) *Statistical Genomics: Linkage, Mapping and QTL Analysis.* CRC Press, Boca Raton.
35. Nelson, J.C. (1997) Qgene – software for marker-based genomic analysis and breeding. *Mol. Breed.* 3, 239–245.
36. Manly, K.F., Cudmore R.H., and Meer, J.M. (2001) Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome* 12, 930–932.
37. Lincoln, S., Daly, M., and Lander, E. (1993a) Constructing genetic linkage maps with MAPMAKER/EXP. Version 3.0. Whitehead Institute for Biomedical Research Technical Report, 3rd ed.
38. Lincoln, S., Daly M., and Lander, E. (1993b) Mapping genes controlling quantitative traits using MAPMAKER/QTL. Version 1.1. Whitehead Institute for Biomedical Research Technical Report, 2nd ed.
39. Basten, C.J., Weir, B.S., and Zeng, Z.B. (1994) Zmap-a QTL cartographer,. in *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software*, Guelph, Ontario, Canada. (Smith

J.S.G.C., Benkel, B.J., Chesnais, W.F., Gibson, J.P., Kennedy, B.W., and Burnside, E.B., eds.), Published by the Organizing Committee, 5th World Congress on Genetics Applied to Livestock Production.

40. Utz, H. and Melchinger, A. (1996) PLABQTL: A program for composite interval mapping of QTL. *J. Quant. Trait Loci* 2. http://probe.nalusda.gov:8000/otherdocs/jqtl
41. Pritchard, J.K. (2001) Deconstructing maize population structure. *Nat. Genet.* 28, 203–204.
42. Pritchard, J.K. Stephens, M., Rosenberg, N.A., and Donnelly, P. (2000) Association mapping in structured populations. *Am. J. Hum. Genet.* 67, 170–181.
43. Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. et-al. (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* 28, 286–289.
44. Gupta, P.K., Varshney, R., Sharma, P., and Ramesh, B. (1999) Molecular markers and their applications in wheat breeding. *Plant Breed.* 118, 369–390.
45. Langridge, P., Lagudah, E., Holton, T., Appels, R., Sharp, P., and Chalmers, K. (2001) Trends in genetic and genome analyses in wheat: a review. *Aust. J. Agric. Res.* 52, 1043–1077.
46. Gupta, P.K., Varshney, R.K., and Prasad, M. (2002) Molecular markers: principles and methodology, in *Molecular Techniques in Crop Improvement* (Jain, S.M., Brar, D.S., and Ahloowalia, B.S., eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 9–54.
47. Somers, D.J. (2004) Molecular marker systems and their evaluation for cereal genetics, in *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 19–34.
48. Gupta, P.K. and Varshney, R.K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113, 163–185.
49. Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S. (2003) Structure of linkage disequilibrium in plants. *Ann. Rev. Plant Biol.* 54, 357–374.
50. Ersoz, E.S., Yu, J., and Buckler, E.S. (2007) Applications of linkage disequilibrium and association mapping in crop plants, in *Genomics Assisted Crop Improvement* (Varshney R.K. and Tuberosa, R.T., eds.), Springer, The Netherlands (Vol. I). pp. 97–120.
51. Buckler, E.S. and Thornsberry, J. (2002) Plant molecular diversity and applications to genomics. *Curr. Opin. Plant Biol.* 5, 107–111.
52. Yu, J. and Buckler, E.S. (2006) Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160.
53. Lander, E. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.
54. Jansen, R. and Stam, P. (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136, 1447–1455.
55. Long, A.D. and Langley, C.H. (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* 9, 720–731.
56. Wang, Y. and Rannala, B. (2005) In silico analysis of disease-association mapping strategies using the coalescent process and incorporating ascertainment and selection. *Am. J. Hum. Genet.* 76, 1066–1073.
57. Hill, W.G. and Robertson, A. (1968) Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231.
58. Lewontin, R.C. (1988) On measures of gametic disequilibrium. *Genetics* 120, 849–852.
59. Wright, S.I. and Gaut, B.S. (2005) Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* 22, 506–519.
60. Bamshad, M., Wooding, S., Salisbury, B.A., and Stephens, J.C. (2004) Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* 5, 598–609.
61. Devlin, B., Bacanu, S.A., and Röder, K., (2004) Genomic control to the extreme. *Nat. Genet.* 36, 1129–1130.
62. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics* 55, 997–1004.
63. Devlin, B., Röder, K., and Wasserman, L. (2001) Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* 60, 155–166.
64. Yu, J., Pressoir, G., Briggs, W.H., Vroh, Bi, I., Yamasaki, M., Doebley, J., Mcmullen, M., Gaut, B., Holland, J.J., Kresovich, S., and Buckler E. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208.
65. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909.