

REVIEW ARTICLE

Association Mapping for Improvement of Quantitative Traits in Plant Breeding Populations

Umesh R. Rosyara¹ and Bal K. Joshi²

¹ South Dakota State University, Plant Science Department, Brookings, South Dakota, USA

² Biotechnology Unit, NARC, Khumaltar, PO Box 1135 Kathmandu, Nepal

*Correspondence Author recent address: joshibalak@yahoo.com

ABSTRACT

DNA-based molecular markers have been extensively utilized for mapping of genes and quantitative trait loci (QTL) of interest based on linkage analysis in mapping populations. This is in contrast to human genetics that use of linkage disequilibrium (LD)-based mapping for fine mapping of QTLs using single nucleotide polymorphisms. LD based association mapping (AM) has promise to be used in plants. Possible use of such approach may be for fine mapping of genes / QTLs, identifying favorable alleles for marker aided selection and cross validation of results from linkage mapping for precise location of genes / QTLs of interest. In the present review, we discuss different mapping populations, approaches, prospects and limitations of using association mapping in plant breeding populations. This is expected to create awareness in plant breeders in use of AM in crop improvement activities.

Kew words: Association mapping, plant breeding, DNA marker, quantitative trait loci

INTRODUCTION

The development and use of molecular markers for the detection and exploitation of DNA polymorphism in plant and animal systems is one of the most significant developments in the field of molecular biology and biotechnology. Of mapping techniques, linkage based mapping is popular in mapping genes in self and cross pollinated crop species. The objective of such genetic mapping is to identify simply inherited markers in close proximity to genetic factors affecting

quantitative traits (quantitative trait loci, or QTL). This localization relies on processes that create a statistical association between marker and QTL alleles and processes that selectively reduce that association as a function of the marker distance from the QTL. When using crosses between inbred parents to map QTL, we create in the F₁ hybrid complete association between all marker and QTL alleles that derive from the same parent.

Recombination in the meioses that lead to doubled haploid, F_2 , or recombinant inbred lines reduces the association between a given QTL and markers distant from it. Unfortunately, arriving at these generations of progeny requires relatively few meioses such that even markers that are far from the QTL (e.g. 10 cM) remain strongly associated with it. Such long-distance associations hamper precise localization of the QTL. One approach for fine mapping is to expand the genetic map, for example through the use of advanced intercross lines, such as F_6 or higher generational lines derived by continual generations of outcrossing the F_2 [1]. In such lines, sufficient meioses have occurred to reduce disequilibrium between moderately linked markers.

When these advance generation lines are created by selfing, the reduction in disequilibrium is not nearly as great as that under random mating. The central problem with any of the above approaches for fine mapping is the limited number of meioses that have occurred and (in the case of advanced intercross lines) the cost of propagating lines to allow for a sufficient number of meioses. An alternative approach is association mapping (AM), taking advantage of events that created association in the relatively distant past. Assuming many generations, and therefore meioses, have elapsed since these events, recombination will have removed association between a QTL and any marker not tightly linked to it.

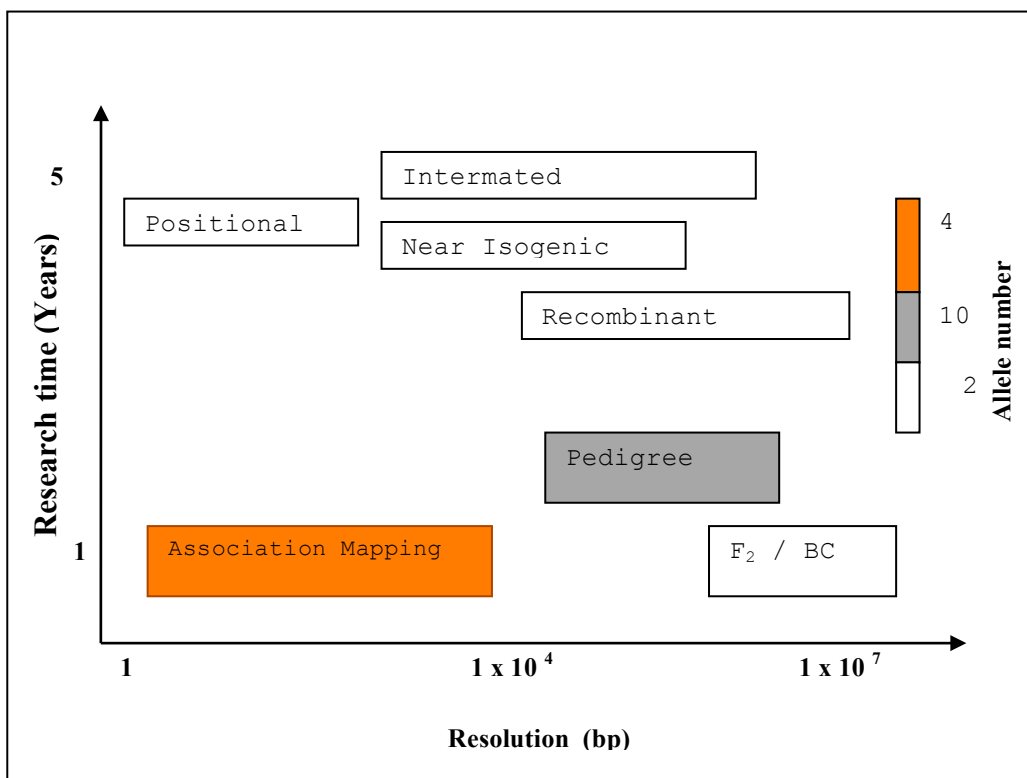


Fig.1: Schematic comparison of various methods for identifying nucleotide polymorphism trait association in terms of resolution, research time and allele number. BC, backcross.

AM, also known as association analysis (AA) or linkage disequilibrium mapping is a method that relies on linkage disequilibrium to study the relationship between phenotypic variation and genetic polymorphisms [2]. Thus Linkage mapping counts recombination between markers and the unknown genes whereas association mapping measure correlation between marker alleles and trait allele in a population (linkage disequilibrium). Association mapping allows for much finer mapping than standard bi-parental cross approaches. Time requirement and resolution of association mapping is compared with other types of mapping approaches (Figure 1).

Linkage Disequilibrium

Linkage disequilibrium (LD) is the nonrandom combination of alleles at two genetic loci, which in random mating populations is mostly generated by mutation and genetic drift, and decays by recombination. The trend of LD decay is shown in graphs with different recombination fractions (Figure 2). Therefore, LD will be observed between two loci if they are in tight linkage or if the haplotype is recent (Hedrick, 2005). Mutations are rare events hence, it is expected that most mutations happened many generations ago and should be in linkage equilibrium with other loci, unless they are very closely linked. While significant LD in random mating populations is evidence of tight linkage, population perturbations like

migration, inbreeding, and selection can build up LD among loosely linked or even unlinked loci. Therefore, the characteristics of the population under study must be recognized when conducting AA or AM and interpreting its results.

Studies have shown that LD levels vary both within and between species (for detail, [2]. For example, LD extends less than 1000 bp [3] for maize landraces and roughly 2000 bp for diverse maize inbred lines [4], but can be as high as 100 kb for commercial elite inbred lines [5]. LD decay can also vary considerably from locus to locus. For example, significant LD was observed up to 4 kb for the Y1 locus (encoding phytonene synthase), but was seen at only 1 kb for PSY2 (a putative phytonene synthase) in the same maize population [6]. Beside the outbred maize, many LD studies have also been carried out in other plant species [7-12].

A variety of mechanisms generate linkage disequilibrium, and several of these can operate simultaneously. The two most common mechanisms include populations expanding from a small number of founders and through admixture. The haplotypes present in the founders will be more frequent than expected under equilibrium. Three special cases are noteworthy. First, genetic drift affects gametic phase disequilibrium (GPD) by this mechanism in that a population

experiencing drift derives from fewer individuals than its present size. Second, by considering an individual with a new mutation as a founder, we see that its descendants will predominantly receive the mutation and loci linked to it in the same phase. Linked marker alleles will therefore be in GPD with the mutant allele. Finally, an extreme case arises in the F_2 population derived from the cross of two inbred lines. Here, all individuals derive from a single F_1 founder genotype and association between loci can be predicted based on their mapping distance. Second, gametic phase disequilibrium arises in structured populations when allelic frequencies differ at two loci across subpopulations, irrespective of the linkage status of the loci. Admixed populations, formed by the union of previously separate populations into a single panmictic one, can be considered a case of a structured population where sub-structuring has recently ceased.

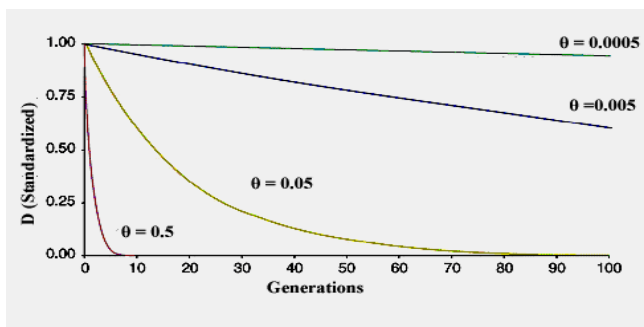


Fig. 2. Decay of linkage disequilibrium with time for four different recombination fractions (θ).

For unlinked loci, $\theta = 0.5$ and LD decays rapidly within a small number of generations. For closely linked loci, the decay in LD is extremely slow. D, Coefficient of linkage disequilibrium (Source: [13]).

METHODS FOR ASSOCIATION MAPPING

Multi-parent Advanced Generation Intercross

In the advanced intercross [1], F_2 individuals are intermated for several generations before mapping. The successive rounds of recombination cause LD to decay and the precision of QTL location to increase. This approach has now been extended to include populations with multiple parents, to take into account information from multiple linked markers [14, 15] and to prioritize candidate polymorphisms [16, 17]. The multiparent advanced generation intercross (MAGIC) was first proposed and applied to mice [14] and is described as heterogenous stock. Recently more successes are described [18]. In both crops and animals, an advantage of the method is that a population can be established containing lines that capture the majority of the variation available in the gene pool. Although it might take several years before these populations are suitable for fine mapping, they are cheap to set up and their value as mapping resources increases with each generation. In plants, MAGIC can be used to combine coarse mapping with low marker densities on lines derived from an early generation, with fine mapping using lines derived from a more advanced generation of crossing and a higher marker density. If such

populations were established now, they would be well placed to exploit the advances in genomics technology and reduction in genotyping and sequencing costs predicted to occur in the next few years [16, 19, 20].

The Transmission Disequilibrium Test

The ability to map QTL in collections of breeders' lines, old landraces or samples from natural populations has great potential. In these populations, LD often decays more rapidly than in controlled crosses. Furthermore, phenotypic data often already exist, saving time and money. The challenge is to distinguish QTL–marker associations arising from LD between closely linked markers from spurious background associations. The first and most robust method of achieving this was the transmission disequilibrium test (TDT) introduced by Richard Spielman and colleagues in 1993 [21]. The TDT provides a way of detecting linkage in the presence of disequilibrium [21]. Neither linkage alone nor disequilibrium alone (i.e. between unlinked markers) will generate a positive result so the TDT is an extremely robust way of controlling for false positives. At its simplest, multiple families consisting of two parents and a single progeny are collected, as shown in Figure 3. Starting from such trios, different models have been evolved since then and some new models allow nuclear families [22] to extended family [23] for quantitative trait analysis in addition

to qualitative traits. The test of association for extended families allows use of available genotypic and phenotypic data from family of any size and structure. Different possible types of families that can be analyzed are shown in Figure 4.

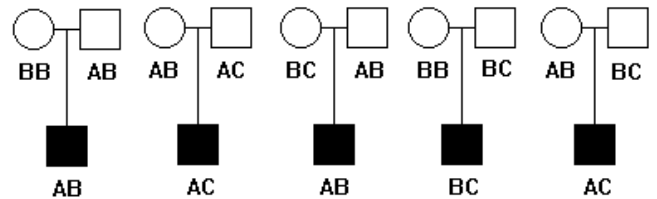


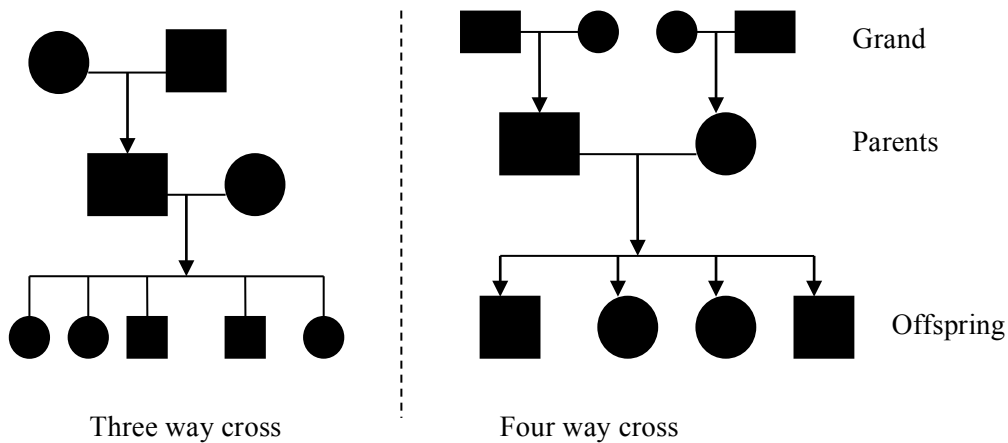
Figure 3. The transmission disequilibrium test

In the simplest case, progeny are selected for an extreme phenotype and transmissions to the progeny from heterozygous parents counted. In the case shown, The A allele is transmitted to affected offspring four times out of five

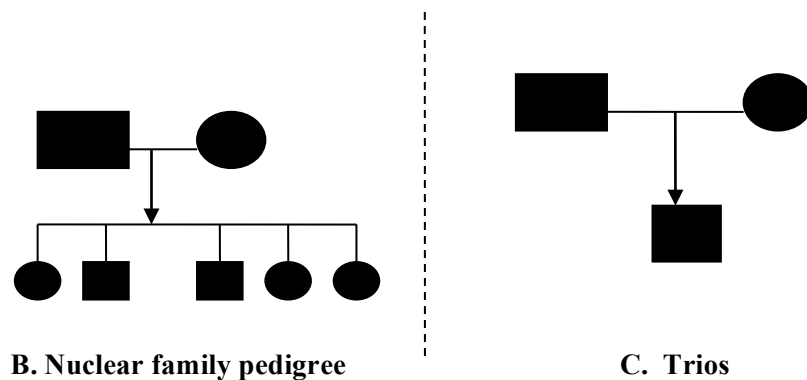
The single progeny in each family is usually selected for an extreme phenotype. In human genetics this typically means they are affected by the disease under study. Parents and progeny are genotyped, but only parents heterozygous at the marker locus are included in the analysis. From each parent, one allele must be transmitted to the progeny and one is not transmitted. Over all families, a count is made of the number of transmissions and non-transmissions. In the absence of linkage between QTL and marker, the expected ratio of transmission to nontransmission is 1:1. In the presence of linkage it is distorted to an extent that depends on the strength of LD

between the marker and QTL. The distortion is tested in a chi-squared test. Power depends on the strength of LD and on the effectiveness of selection of extreme progeny in driving segregation away from expectation. This elegant test is extremely robust to the effects of population structure, but is susceptible to an increase in false positive results generated by genotype error and biased allele calling [24]. This risk can be reduced by modeling genotype errors and missing data in the

analysis [25-27] or by comparing the transmission ratio for extreme phenotypes with that for control individuals or for the opposite extreme. The TDT has been extended to study haplotype transmissions, quantitative traits, the use of sib pairs rather than parents and progeny, and information from extended pedigrees. TDT and other family-based association tests are reviewed elsewhere [28].



A. Extended pedigree



B. Nuclear family pedigree **C. Trios**

Fig. 4. Different type of families for association mapping elite inbred lines

In crops, parental and progeny lines are usually separated by several generations of gametogenesis rather than by one. In this case, the TDT is still valid, but might no longer be so robust, the process of breeding might itself distort segregation patterns. A family-based association test that is applicable to plant breeding programs has recently been proposed [29]. The authors point out that for candidate gene studies, this method is more cost effective than the alternative methods described below given that no additional control markers are required. However, some power will be lost because only progeny derived from F_1 s known to have a heterozygous marker genotype are informative.

Genomic Control

Population structure arising from recent migration and population admixture will generate LD between a trait and markers distributed over the whole genome. This can be detected by studying whether the distribution of the test statistic for association, estimated empirically from a set of genome-wide distributed markers, differs from the expected null distribution. This is the basis of genomic control (GC) [30, 31]. To estimate the empirical distribution accurately would require many markers. However, all that is required is to estimate the mean test statistic and compare it with its expected value (1.0 for a 1 degree of freedom chi-squared test) for which only ~50

markers are needed [32]. If the average chi-squared at a set of 50 control markers is much greater than 1.0, population structure is indicated. For any candidate marker, the null-hypothesis is now no longer absence of association between it and the trait. Rather, it is that there is no association above the background level resulting from population structure. To test for this, we simply divide the observed chi-squared between the candidate and trait by the average chi-squared at the control markers and look up the p-value of the adjusted chi-squared in the usual manner.

GC is valid for any single degree of freedom test. Preferably, the control markers should loosely match the test marker in allele frequency, but this is not crucial [31]. For quantitative traits, the difference between trait means for each marker class is usually tested in a t test. Provided the number of observations is reasonably large, t^2 is distributed as a 1 degree of freedom chi-squared and GC can still be carried out. More recent work has suggested that greater accuracy is achieved by treating the test statistic as an F test with one degree of freedom (df) in the numerator and degrees of freedom in the denominator equal to the number of control loci [33]. More sophisticated versions of GC are available. With large numbers of candidate polymorphisms to test, the majority are not expected to be genuinely associated with the trait. In this case, procedures and software are

available in which, in effect, the candidate markers act as their own controls. GC has also been extended to control for bias in accuracy of genotyping between DNA samples from different origins [34] and to tests with >1 df [35]. GC also corrects for unknown kinship among collections of lines [30]. The presence of related lines can greatly increase the frequency of false positives. For many crop datasets this will be the greatest source of bias. The correction of the false positive rate using GC comes at a cost: power is always decreased. This loss of power can be great in cases of extreme population subdivision. Furthermore, because loci can vary in their differentiation between populations, the uniform adjustment of GC might be insufficient for some candidate polymorphisms and overcorrect at others.

Structured Association

Structured association (SA) provides a sophisticated approach to detecting and controlling population structure [36-38]. Again, additional markers are required, randomly distributed across the genome. Just as for GC, recent migration and population admixture are assumed to generate LD among unlinked and loosely linked markers that have yet to decay fully. However, we expect the parental populations themselves to be in linkage equilibrium. By trial and error one could allocate the individuals in our sample to

parental populations such that disequilibrium within populations was minimized. One could then include information on population membership in the test of association. This is the approach taken for SA. First individuals are allocated to populations, then this information is used to control for population membership in the test of association [36-38]. To allocate individuals to populations we need to know in advance how many populations there are. If unknown, this can be estimated: the allocation process is repeated for different possible numbers and the best fitting selected. Nevertheless, deciding on population number can be problematic. The computer program STRUCTURE [37] uses computationally intensive methods to partition individuals into populations. Many individuals or lines will not belong uniquely to one, but will be the descendents of crosses between two or more ancestral populations. STRUCTURE also estimates the proportion of ancestry attributable to each population. Following allocation of individuals to populations, the test for association is carried out in a model fitting exercise. Here, the principle is that variation attributable to population membership is accounted for first, using estimates of population membership from STRUCTURE, and then the presence of any residual association between the marker and phenotype is tested. For example, to test for association between a quantitative trait and a microsatellite, the trait is first regressed on the

estimated coefficients of population membership and then on the marker – coded as a factor as if in an analysis of variance [39]. SA is effective in detecting and adjusting for the presence of population structure, but does not deal with consanguinity within populations. Recently, Ed Buckler's group introduced a method in which population membership is estimated using STRUCTURE and kinship among varieties is estimated empirically from a second set of control markers [40]. The analysis takes into account both population structure and the correlation between individuals that results from their relationships. This method is implemented in the software TASSEL.

ASSOCIATION MAPPING IN PLANT BREEDING POPULATIONS

Scientific plant breeding is a recent activity that normally involves a narrow genetic pool, such that breeding populations can be traced back to relatively few original parents, normally landraces, within a relatively small number of generations (e.g. [41, 42]). Under this scenario, mutations play a minor role and most of the observed LD is expected to reflect the haplotypes of the original parents. Moreover, because there were few opportunities for recombination between the time of introduction of a parent and the present, LD in some plant breeding populations may not reliably indicate tight

linkage. Between unlinked loci, LD can be caused by simultaneous selection of combinations of alleles at different genes, including epistasis, and by population structure [43]. Both phenomena should be common in plant breeding populations.

Selection should affect LD in parts of the genome related to traits that are relevant for the breeding program. This source of distortion should be taken into consideration in the interpretation of results of AA in a case-specific manner. In contrast, population structure is expected to affect the pattern of LD over the whole genome and must be controlled a priori for correct association analysis [38]. Most of the literature on AA refers to human populations or theoretical panmictic populations. There is limited information and discussion about applications of this technique to plant breeding. As the information generated by QTL studies accumulates, a method is needed to convert efficiently that information into practical tools for plant selection. Association analysis can be an effective approach for closing the gap between QTL analysis and marker-assisted selection.

The objective of this review paper is to discuss potential applications of association mapping for plant breeding populations. Plant breeding populations include basically three types - germplasm bank collections, synthetic populations, and elite lines.

Choice of Populations for AM in Plant Breeding Programs

In a plant breeding program, three main types of populations could be considered for implementation of AM: germplasm bank collections, elite breeding materials, and synthetic populations. The application of AA differs among those populations in several aspects (Table 1). For efficient integration of AM with other methods currently in use, material that is routinely generated and evaluated should be used for both purposes. In the case of germplasm banks, core collections are expected to represent most of the genetic variability with a manageable number of accessions [44], and thus are suitable for genetic studies. In the case of elite materials, the sample could be composed by lines and checks evaluated in regional trials. For synthetic populations, the evaluation unit should be also the association unit (or closely related to it), whether it is an individual or a family.

Germplasm Bank Core Collections

Samples representing the genetic diversity of a species are attractive for AM because of the wide allele diversity encompassed. Methods of selection of core collections often involve genotyping unlinked markers to compute genetic distances, thus providing information about population structure. The process of selection of a minimum sample with maximum

variation has a normalizing effect that is expected to reduce population structure and LD between unlinked loci, thus creating a situation favorable for association analysis [45]. A difficulty likely to occur in this type of material is related to genetic heterogeneity within samples. Landraces and natural populations often consist of open-pollinated varieties or mixtures of genotypes, and the DNA extraction, genotyping, and phenotyping schemes must account for this variability.

Core collections are useful materials for AM of qualitative traits, such as disease resistance or special quality characteristics (color, aroma, etc). Studies focusing on domestication-related traits such as seed dormancy, shattering, or inflorescence type also could require wide phenotypic variation, beyond the limits of cultivated germplasm [46]. Conversely, the broad genetic variability of those collections normally make them unsuitable for analysis of quantitative traits because part of the accessions would be unadapted to growing conditions and prevalent diseases, resulting in poor precision of trait measurement.

Common ancestors of distantly related individuals occurred many generations ago; therefore, LD is expected to have decayed to short genetic distances. For this reason, AM in core collections will probably require candidate genes or major QTL mapped within narrow confidence intervals [47]. Compared with linkage-based fine mapping and

Table 1. Comparison of different types of populations for association analysis. Depends on the collection, conservation and sampling schemes

<i>Aspects of association mapping</i>	<i>Germplasm bank</i>	<i>Elite material</i>	<i>Synthetic populations</i>
Samples	Entries of a core collection	Inbred lines and cultivars	Individual plants or progenies
Sample turnover	Static	Gradually substituted	Ephemeral
Source of phenotypic data	Phenotypic screenings	Replicated yield trials	Evaluation for recurrent selection
Type of traits	High heritability and domestication traits	Low heritability, yield	Depends on the evaluation scheme
Level of LD	Low	High	Intermediate
Population structuring	Medium	High	Low
Allele diversity among samples	High	Low	Intermediate
Allele diversity within samples	Variable [†]	1 allele	1 or 2 alleles [‡]
Resolution of AM	High	Low	Intermediate and increasing
Power of AM	Low	High	Intermediate and decreasing
Application of significant markers	Marker-assisted backcross	Marker-assisted selection	Incorporation in selection index

[†]Depends on the collection, conservation and sampling schemes. [‡]For diploid species. (Source: [45].)

positional cloning [48] the AM approach would offer the advantage of simultaneously detecting the effect and screening the germplasm for useful alleles. Significant markers would be useful for introgression of the new variation into elite germplasm through marker-assisted backcrossing [49], while markers used for population structure inference could be used to speed up the recovery of the recurrent parent genome. Theoretical projections indicate that the use of two markers per chromosome for selection against the donor genotype could shorten the transfer by about two generations [50]

Elite Lines and Cultivars

Maximum relative efficiency of marker-assisted selection compared with phenotypic selection is expected when heritability is low and markers capture a significant portion of the variation for the trait [51].

Elite lines are desirable materials for AM of low heritability traits, including yield, yield components, and tolerance to abiotic stresses because elite lines are genetically stable and are well adapted to normal growing conditions.

In plant breeding programs, there is normally a large body of phenotypic data accumulated for

elite lines and cultivars from replicated field experiments over locations and years. Use of those data for AM requires statistical models accounting for covariances introduced both by experimental design (years, locations, replicates) and polygenic effects. Moreover, those data are often unbalanced because new lines are included in field trials each year, while other lines are discarded. Maximum likelihood solutions of mixed-effects models yield minimum-variance unbiased estimates of allele effects from unbalanced data, taking into account the correlation structure of the data [52]. Mixed-effects models were used to analyze plant height, disease resistance, and grain moisture in maize [53] and grain size and milling quality in wheat [45].

Population structure can be prominent in elite material because it is common for closely related lines to be admitted to advanced trials. If pedigrees are known, the relationships among the lines can be determined [41] and used to control for polygenic effects [54]. In this case, it is not essential to estimate population structure through unlinked markers, although there may still be interest in marker data as a genetic fingerprint for variety protection [55] and for purity control of seed production.

A typical elite plant breeding pool is derived from few founders in the recent past, and is submitted to intense selection. For those

reasons, LD is expected to be high in this material, and the first experimental results confirm this expectation [3, 5]. Although AM in elite lines may not offer much improved resolution compared with QTL analysis in biparental mapping populations, there are at least two important advantages: a substantially higher level of polymorphism and detection of favorable alleles directly in the target population. Elite lines are natural candidates for crossing to generate the next round of breeding, and significant markers could be used for marker-assisted selection in the progeny.

Synthetic Populations

Although the potential of synthetic populations for AM is largely unknown, they might be the plant breeding materials that best approximate the assumption of random mating because synthetics are normally designed and maintained to minimize inbreeding. Population structure is expected to be mild or absent, which is an important advantage of synthetics for AM. If the experimental material represents a single intermating population, the power of AM is maximized and the risk of false associations is minimized [56]. Nevertheless, population structure can still occur because of differences in flowering time, plant height, and other traits that may lead to assortative mating.

Genotypic information could be useful in all phases of population breeding. In the choice of

parents to form the population, knowledge of the genetic distance among lines would be useful to achieve a compromise between high means for agronomic traits and high allelic variability. By genotyping samples of subsequent cycles with unlinked markers, breeders can monitor changes in allele diversity, effective population size, and population structure [57, 58]. The allele diversity of synthetic populations depends on the number and divergence of parents and the intensity of selection applied. The level of LD in synthetic populations is expected to be high in the initial generations, such that a genome scan could detect large chromosome segments associated with traits, and trace them back to parental haplotypes. In subsequent generations, the decay of LD by recombination would favor increasingly refined mapping. However, synthetic populations are often submitted to recurrent selection, a breeding scheme consisting of successive cycles of evaluation, selection, and recombination [59]. Intense selection could build up LD by favoring allelic combinations or by promoting genetic drift [6]. For this reason, populations subjected to mild or no selection would be preferred for AM. [60] developed a population for association analysis from the Illinois high/low oil populations, with 10 generations of recombination without selection.

A short time frame is a fundamental characteristic of plant breeding populations for

AM, compared with natural populations. Therefore, in plant breeding populations, the most significant association does not necessarily indicate the position of the gene [45]. In the long term, linkage becomes the major factor defining the association between QTL or gene and marker, and only closely linked markers remain in high LD; however, the time required to achieve this situation is longer than most breeding programs have been in existence. For this reason, AM in plant breeding programs should be considered a method of detection of markers for indirect selection, rather than a method for fine-mapping QTL [45]. To alleviate this problem, the breeder should use methods like recurrent selection, which maximizes the heterozygosity and the opportunities for recombination.

The resolving power of LD mapping depends on how rapidly LD decays with genetic distance. This varies between populations of landraces, wild progenitors and modern cultivars as a result of the diverse history to which crop plants have been subjected since their domestication [61]. In some populations, LD will decay so rapidly that they are best suited for fine mapping, whereas in others the decay might be so slow that whole genome scans are practical. In crops where collections of contemporary, historical and wild material exist, selection of different sets of lines might permit both fine and coarse mapping [61]. However, in most crops, marker density is

currently too low for genome scans. Before attempting these, power calculations should demonstrate that, given the rate of decay of LD in the population to be studied, the density of markers and their allele frequency distribution are adequate to detect linked QTL accounting for specified proportions of the phenotypic variation. Population size is also important. An LD mapping experiment will almost always have lower power than a family based linkage mapping experiment of equivalent size: if 100 lines are just sufficient for a family based linkage mapping study, they will be too few for LD mapping. For these reasons this is believed that the best use of LD mapping is to refine the location of QTL identified in family based linkage mapping and candidate gene studies. While linkage mapping methods offer a high power to detect QTL in genome-wide approaches, association mapping methods have the merit of a high resolution to detect QTL [4]. Linkage and association analysis are thought to be complementary to each other in terms of providing prior knowledge, cross-validation, and statistical power [62]. So if both analyses are done this is expected to help in proper location of QTLs. Longer term, prospects for high-throughput genotyping and sequencing might make whole-genome scans by LD mapping more feasible. The challenge is to identify and create the appropriate populations so that computational, analytical and profiling advances can be rapidly harnessed by the crop science community. For

plant breeding application, at current situation AM could be useful for validating location of QTL of interest and identifying favorable allele for marker aided selection. Once a genetic marker has been demonstrated to be associated with a phenotypic trait of interest, it can be used as a selection target to obtain an indirect response in the trait. In recurrent selection, markers could be used to store information acquired from phenotypic evaluations, which can be used for selection in later cycles. Likewise, in pedigree breeding, markers could carry information about yield potential from the phase of replicated field trials to the phase of single-plant selection, when evaluation of yield cannot be made with reasonable precision.

CONCLUSION

With the availability of high density maps in a number of crop plants, the whole genome sequences in model plants like Arabidopsis and rice, and the sequences of gene rich regions in crops like sorghum, maize and wheat the association mapping tool have future for increasing applications. Even though most of plant breeders' populations could not be used for fine mapping as such the association mapping could be helpful in identification of favorable alleles for marker aided selection and cross validation of results of linkage based mapping.

REFERENCES

1. Darvasi A, Soller M: **Advanced intercross lines, an experimental population for fine genetic mapping.** *Genetics* 1995, **141**:1199-1207.
2. Flint-Garcia SA, Thornsberry JM, Buckler ES: **Structure of linkage disequilibrium in plants.** *Annu. Rev. Plant Biol.* 2003, **54**:357-374.
3. Tenailon MI, Sawkins MC, Long AD, et al. **Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.).** *Proc. Natl. Acad. Sci. USA* 2001, **98**:9161-9166.
4. Remington DL, Thornsberry JM, Matsuoka Y, et al. **Structure of linkage disequilibrium and phenotypic associations in the maize genome.** *Proc. Natl. Acad. Sci. USA* 2001, **98**:11479-11484.
5. Ching A, Caldwell KS, Jung M, et al. **SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines.** *BMC Genet.* 2002, **3**:19.
6. Palaisa KA, Morgante M, Williams M, Rafalski A: **Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci.** *Plant Cell* 2003, **15**:1795-1806.
7. Ingvarsson PK: **Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae).** *Genetics* 2005, **169**:945-953.
8. Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB: **Nucleotide diversity and linkage disequilibrium in loblolly pine.** *Proc. Natl. Acad. Sci. USA* 2004, **101**:15255-15260.
9. Morrell PL, Toleno DM, Lundy KE, Clegg MT: **Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. spontaneum) despite high rates of self-fertilization.** *Proc. Natl. Acad. Sci. USA* 2005, **102**:2442-2447.
10. Garris AJ, McCouch SR, Kresovich S: **Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.).** *Genetics* 2003, **165**:759-769.
11. Hamblin MT, Mitchell SE, White GM, et al. **Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of sorghum bicolor.** *Genetics* 2004, **167**:471-483.
12. Nordborg M, Borevitz JO, Bergelson J, et al. **The extent of linkage disequilibrium in *Arabidopsis thaliana*.** *Nat. Genet.* 2002, **30**:190-193.
13. Mackay I, Powell W: **Methods for linkage disequilibrium mapping in crops.** *TRENDS in Plant Science* 2007, **12**:doi:10.1016/j.tplants.2006.12.001.
14. Mott R, Talbot CJ, Turri MG, Collins AC, Flint J: **A method for fine mapping quantitative trait loci in outbred animal stocks.** *Proc. Natl. Acad. Sci. USA* 2000, **97**:12649-12654.
15. Mott R, Flint J: **Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks.** *Genetics* 2002, **160**:1609-1618.
16. Yalcin B, Flint J, Mott R: **Using progenitor strain information to identify quantitative trait nucleotides in outbred mice.** *Genetics* 2005, **171**:673-681.
17. Valdar W, Solberg LC, Gauguier D, et al. **Genome-wide genetic association of complex traits in heterogeneous stock mice.** *Nat. Genet.* 2006, **38**:879-887.

18. Valdar W, Flint J, Mott R: **Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice.** *Genetics* 2006, **172**:1783-1797.
19. Syvanen AC: **Toward genome-wide SNP genotyping.** *Nat. Genet.* 2005, **57**:S5-S10.
20. Macdonald SJ, Pastinen T, Genissel A, Cornforth TW, Long AD: **A low-cost open-source SNP genotyping platform for association mapping applications.** *Genome Biol.* 2005, **6**:R105.
21. Spielman R, McGinnis S, Ewens WJ: **Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).** *Am. J. Hum. Genet.* 1993, **52**:506-516.
22. Abecasis GR, Cookson WO, Cardon LR: **Pedigree tests of transmission disequilibrium.** *Eur. J. Hum. Genet.* 2000, **8**:545-551.
23. Abecasis GR, Cardon LR, Cookson WOC: **A general test of association for quantitative traits in nuclear families.** *Am. J. Hum. Genet.* 2000, **66**:279-292.
24. Mitchell AA, Chakravarti A: **Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test.** *Am. J. Hum. Genet.* 2003, **72**:598-610.
25. Allen AS, Rathouz PJ, Satten GA: **Informative missingness in genetic association studies: Case-parent designs.** *Am. J. Hum. Genet.* 2003, **72**:671-680.
26. Gordon DS, Heath C, Liu X, Ott J: **A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data.** *Am. J. Hum. Genet.* 2001, **69**:371-380.
27. Gordon D, Haynes C, Johnnidis C, et al. **A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents.** *Eur. J. Hum. Genet.* 2004, **12**:752-761.
28. Laird NM, Lange C: **Family-based designs in the age of large-scale gene-association studies.** *Nat. Rev. Genet.* 2006, **7**:385-394.
29. Stich B, Melchinger AE, Piepho H, et al. **A new test for family-based association mapping with inbred lines from plant breeding programs.** *Theor. Appl. Genet.* 2006, **113**:1121-1130.
30. Devlin B, Roeder K: **Genomic control for association studies.** *Biometrics* 1999, **55**:997-1004.
31. Reich DE, Goldstein DB: **Detecting association in a casecontrol study while correcting for population stratification.** *Genet. Epidemiol.* 2001, **20**:4-16.
32. Bacanu SA, Devlin B, Roeder K: **Association studies for quantitative traits in structured populations.** *Genet. Epidemiol.* 2002, **22**:78-93.
33. Devlin B, Bacanu SA, Roeder K: **Genomic control in the extreme.** *Nat. Genet.* 2004, **36**:1129-1130.
34. Clayton DG, Walker NM, Smyth DJ, et al. **Population structure, differential bias and genomic control in a large-scale, case-control association study.** *Nat. Genet.* 2005, **37**:1243-1246.
35. Zheng G, Freidlin B, Gastwirth JL: **Robust genomic control for association studies.** *Am. J. Hum. Genet.* 2006, **78**:350-356.
36. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and**

- correlated allele frequencies.** *Genetics* 2003, **164**:1567-1587.
37. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945-959.
38. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** *Am. J. Hum. Genet.* 2000, **67**:170-181.
39. Aranzana MJ, Kim S, Zhao K, et al. **Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes.** *PLoS Genet.* 2005, **1**:60.
40. Yu J, Pressoir G, Briggs WH, et al. **A unified mixed-model method for association mapping accounting for multiple levels of relatedness.** *Nat. Genet.* 2006, **38**:203-208.
41. Bered F, Barbosa-Neto JF, Carvalho FIF: **Genetic variability in common wheat germplasm based on coefficients of parentage.** *Genet. Mol. Biol.* 2002, **25**:211-215.
42. Lu H, Redus MA, Coburn JR, et al. **Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis.** *Crop Sci.* 2005, **45**:66-76.
43. Hartl D, Clark A: *Principles of population genetics.* Sunderland, MA: Sinauer; 1997.
44. Zhang XR, Zhao YZ, Cheng Y, et al. **Establishment of sesame germplasm core collection in China.** *Genet. Resour. Crop Evol.* 2000, **47**:273-279.
45. Breseghello F, Sorrells ME: **Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars.** *Genetics* 2005, (doi:10.15).
46. Clark RM, Linton E, Messing J, Doebley JF: **Pattern of diversity in the genomic region near the maize domestication gene *tb1*.** *Proc. Natl. Acad. Sci. USA* 2004, **101**:700-707.
47. Thornsberry JM, Goodman MM, Doebley J, et al. **Dwarf polymorphisms associate with variation in flowering time.** *Nat. Genet.* 2001, **28**:286-289.
48. Yan L, Loukoianov A, Tranquilli G, et al. **Positional cloning of the wheat vernalization gene *VRN1*.** *Proc. Natl. Acad. Sci. USA* 2003, **100**:6263-6268.
49. Frisch M, Melchinger AE: **Selection theory for marker-assisted backcrossing.** *Genetics* 2005, **170**:909-917.
50. Hospital F, Chevalet C, Mulsant P: **Using markers in gene introgression breeding programs.** *Genetics* 1992, **132**:1199-1210.
51. Lande R, Thompson R: **Efficiency of marker-assisted selection in the improvement of quantitative traits.** *Genetics* 1990, **124**:743-756.
52. Pinheiro JC, Bates DM: *Mixed-Effects Models in S and S-PLUS.* New York: Springer; 2000.
53. Parisseaux B, Bernardo R: **In-silico mapping of quantitative trait loci in maize.** *Theor. Appl. Genet.* 2004, **109**:508-514.
54. Zhang YM, Mao Y, Xie C, et al. **Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.).** *Genetics* 2005, **169**:2267-2275.
55. Röder MS, Wendehake K, Korzun V, et al. **Construction and analysis of a microsatellite-based database of European wheat varieties.** *Theor. Appl. Genet.* 2002, **106**:67-73.
56. Cardon LR, Palmer LJ: **Population stratification and spurious allelic association.** *Lancet* 2003, **361**:598-604.

57. Courtois B, Filloux D, Ahmadi N, et al. **Using molecular markers in rice population improvement through recurrent selection.** In *Population improvement: A way of exploiting the rice genetic resources of Latin America.* edited by Guimarães EP Rome: FAO; 2005:52-74.
58. Ramis CA, Badan CDC, Guimarães EP, Díaz A, Gamboa CE: **Molecular markers as tools for rice population improvement.** In *Population improvement: A way of exploiting the rice genetic resources of Latin America.* edited by Guimarães EP Rome: FAO; 2005:75-94.
59. Fehr WR: *Principles of cultivar development: Theory and technique.* Ames, IA: Macmillan; 1987.
60. Laurie CC, Chasalow SD, LeDeaux JR, et al. **The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel.** *Genetics* 2004, **168**:2141-2155.
61. Caldwell KS, Russell J, Langridge P, Powell W: **Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*.** *Genetics* 2006, **172**:557-567.
62. Wilson LM, Whitt SR, Ibanez AM, et al. **Dissection of maize kernel composition and starch production by candidate gene association.** *Plant Cell* 2004, **16**:2719-2733.