Plant Molecular Biology (2005) 57:461–485
DOI 10.1007/s11103-005-0257-z

Review

Linkage disequilibrium and association studies in higher plants: Present status and future prospects

Pushpendra K. Gupta*, Sachin Rustgi and Pawan L. Kulwal
*Molecular Biology Laboratory, Department of Genetics & Plant Breeding, Ch. Charan Singh University, Meerut – 250 004 (UP), India (*author for correspondence; e-mail pkgupta36@vsnl.com, pkgupta36@hotmail.com)*

Received 1 June 2004; accepted in revised form 4 January 2005

Key words: association mapping, linkage disequilibrium (LD), plants, population genetics, QTL

Abstract

During the last two decades, DNA-based molecular markers have been extensively utilized for a variety of studies in both plant and animal systems. One of the major uses of these markers is the construction of genome-wide molecular maps and the genetic analysis of simple and complex traits. However, these studies are generally based on linkage analysis in mapping populations, thus placing serious limitations in using molecular markers for genetic analysis in a variety of plant systems. Therefore, alternative approaches have been suggested, and one of these approaches makes use of linkage disequilibrium (LD)-based association analysis. Although this approach of association analysis has already been used for studies on genetics of complex traits (including different diseases) in humans, its use in plants has just started. In the present review, we first define and distinguish between LD and association mapping, and then briefly describe various measures of LD and the two methods of its depiction. We then give a list of different factors that affect LD without discussing them, and also discuss the current issues of LD research in plants. Later, we also describe the various uses of LD in plant genomics research and summarize the present status of LD research in different plant genomes. In the end, we discuss briefly the future prospects of LD research in plants, and give a list of softwares that are useful in LD research, which is available as electronic supplementary material (ESM).

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. This led to major advances in plant genomics research during the last quarter of a century, and made the use of molecular markers a thrust area of research in plant genetics. Two major phenomena involved in the generation of DNA polymorphism detected by molecular markers are mutation and recombination. Therefore, detection of linkage and tracing the history of a DNA polymorphism have been central to the use of molecular markers for a variety of studies (Terwilliger and Weiss, 1998; Nordborg and Tavaré, 2002; Gupta and Rustgi, 2004). However, for the study of linkage, one needs to perform suitably designed crosses, sometimes leading to the development of mapping populations or near-isogenic lines (NILs). This is a serious limitation on the use of molecular markers in some cases, because the desired crosses cannot be made in all cases (e.g. in forest trees), and/or the mapping populations that are examined for this purpose are sometimes too small, with only two alleles at a locus sampled. In view of this, alter-

native methods have been developed and used to study the phenomenon of linkage and recombination on the one hand, and for the study of mutational history of a population on the other. One such method is linkage disequilibrium (LD) -based association analysis that has received increased attention of plant geneticists during the last few years. This approach has the potential not only to identify and map QTLs (Meuwissen and Goddard, 2000), but also to identify causal polymorphism within a gene that is responsible for the difference in two alternative phenotypes (Palaisa *et al*., 2003, 2004). This also allows the identification of haplotype blocks and haplotypes representing different alleles of a gene. In using this approach, an idea of the length of a region over which LD persists is also possible, so that one can plan and design studies for association analysis. The techniques/methods used for estimation of the level of LD, the factors that influence these estimates, and the uses and limitations of this approach have been widely discussed in recent years; some reviews on LD exclusively devoted to the studies in plants also appeared recently (Flint-Garcia *et al*., 2003; Gaut and Long, 2003; Rafalski and Morgante, 2004). In this review, we have tried to summarize LD studies conducted in plants, with major emphasis on newer aspects including LD among multiple loci and among loci with multiple alleles, and the effects of selection (including hitchhiking and epistasis) and gene conversion on LD.

What is linkage disequilibrium/association mapping?

The terms linkage disequilibrium and association mapping have often been used interchangeably in literature. However, we feel that while association mapping refers to significant association of a molecular marker with a phenotypic trait, LD refers to non-random association between two markers or two genes/QTLs or between a gene/QTL and a marker locus. Thus, association mapping is actually one of the several uses of LD. In statistical sense, association refers to covariance of a marker polymorphism and a trait of interest, while LD represents covariance of polymorphisms exhibited by two molecular markers/genes. However, for the above association, the term LD has been considered by some to be inappropriate (Jannink and Walsh, 2002), since LD in the sense discussed above may also be caused due to factors other than linkage (see later). This non-random association is, therefore, more appropriately also termed '*gametic phase disequilibrium*' (GPD), or simply '*gametic disequilibrium*' (Hedrick, 1987) and is examined within populations of unrelated individuals (although they may be related through distant ancestry). However, LD due to linkage is the net result of all the recombination events that occurred since the origin of an allele by mutation, thus providing higher opportunity for recombination to take place between any two closely linked loci.

How to measure LD and test its statistical significance?

The different measures (indices) for estimating the level of LD in plants have largely been described in recent reviews on LD in plants (Flint-Garcia *et al*., 2003; Gaut and Long, 2003). Here, we list and only briefly describe the methods available for the measurement of LD; the statistical tests that are available for testing the significance of these measures are also briefly described. The LD involving multiallelic loci and multilocus conditions will be dealt in a relatively greater detail, since in the past this aspect did not receive the treatment, which it deserved. Details of these methods are available elsewhere in the published literature (Jorde, 2000; Liang *et al*., 2001; Gorelick and Laubichler, 2004).

Two-locus methods

Linkage disequilibrium is often quantified using statistics of association between allelic states at pairs of loci. However, the loci involved may be biallelic (e.g. SNPs, AFLPs) or multiallelic (e.g. SSRs, RFLPs), although sometimes even multiallelic loci are treated as biallelic, since only two alleles are sampled in a mapping population. However, when natural populations or germplasm collections are used for estimation of LD, multiple alleles at each of the two loci can be sampled and used for estimation of LD. Furthermore, although there are measures, which can be used for both

biallelic and multiallelic conditions; the measures that are frequently used for biallelic condition need to be modified for measuring LD under multiallelic condition (Hedrick, 1987; see below). The power of LD mapping under the two conditions may also differ under certain conditions (Czika and Weir, 2004).

Biallelic loci

The different available measures for estimation of LD between any two biallelic loci mainly include D, D', r^2, R, D^2, D^*, Q^*, F', $X(2)$, δ, etc. Some of these measures can also be used for multiallelic situations (see next paragraph). The details of these different indices, the formulae used for their calculation, and the relative merits of each of these indices have been discussed in a number of earlier reviews (Jorde, 2000; Ardlie *et al.*, 2002; Flint-Garcia *et al.*, 2003; Gaut and Long, 2003), so that their description here will be repetitive. However, a caution need to be exercised against an indiscriminate use of any of these measures, because all of these measures except D' are strongly dependent ~~on allele frequencies~~ (Hedrick, 1987); even D' is sometimes dependent on allele frequencies (Lewontin, 1988). Most of these measures are also sensitive to ~~small sample size,~~ and some of them even give negative values of LD under conditions of maximum disequilibrium (Hedrick, 1987).

Of all the above measures of LD, D' and r^2 are the preferred measures of LD, although δ, which is similar to P_{excess} proposed by Lehesjoki *et al.* (1993) and λ proposed by Terwilliger (1995), has been considered by some to be as good as D', because it is also directly proportional to the recombination fraction. Among these two preferred measures (r^2 and D'), while D' measures only recombination differences; r^2 summarizes recombination and mutation history. Also r^2 is indicative of how markers might be correlated with QTL of interest, so that for association studies, often r^2 is preferred (Abdallah *et al.*, 2003). Therefore, the choice between D' and r^2 for a measure of LD may also depend on the objective of the study. In some recent reviews, the differences between different measures of LD have been explained using a figure (Rafalski, 2002 [only D' was calculated in this study]; Flint-Garcia *et al.*, 2003; Gaut and Long, 2003), which has been modified by us incorporating the index δ, and showing the effect of allele frequency on D, D', r^2, and δ (Figure 1).

Multiallelic loci and phase information in heterozygotes

In addition to the biallelic markers like SNPs, multiallelic markers like SSRs, are also often used for association studies. These SSR markers have already been used for a study of population structure in maize and rice (Remington *et al.*, 2001; Garris *et al.*, 2003) and for LD-based association studies in wheat and barley (Kruger *et al.*, 2004; Mather *et al.*, 2004). For LD between two multiallelic loci also, D' (in a modified form) is the most widely used measure of LD for each pair of alleles, or even for overall LD between all the alleles at two loci. It has been shown that the range of D' is largely independent of allele frequencies and other conditions, more often than was previously thought, and that standardization of D' (suggested in the past) is not necessary (Zapata, 2000). D' can also be computed from maximum likelihood (ML) estimates using an expectation-maximization (EM) algorithm (Slatkin and Excoffier, 1996), and strategies have been developed to map quantitative trait loci (QTLs) using D', when the QTL and marker loci are both multiallelic (Abdallah *et al.*, 2003).

The problem of estimating LD among pairs of loci, each with multiple alleles becomes particularly difficult, when individuals are heterozygous at more than one locus and many loci are considered (for multilocus methods, see next section). Under this condition, haplotype phase information is missing, so that s heterozygous loci can be resolved into haplotypes in as many as 2^{s-1} different ways, making inference about haplotype phase difficult. In the past, efforts were made to resolve this haplotype phase problem either through pedigree analysis (Eaves *et al.*, 2000) or through characterization of gametes (Taillon-Miller *et al.*, 2000), through haploid storage tissue of seeds (megagametophyte) (Neale and Savolainen, 2004), through asymmetric PCR or through isolation of single chromosomes for PCR amplification; an improved algorithm based on Hardy-Weinberg equilibrium was also introduced to infer the haplotype phase from PCR-amplified DNA (Clark, 1990). However, there were problems associated with each of

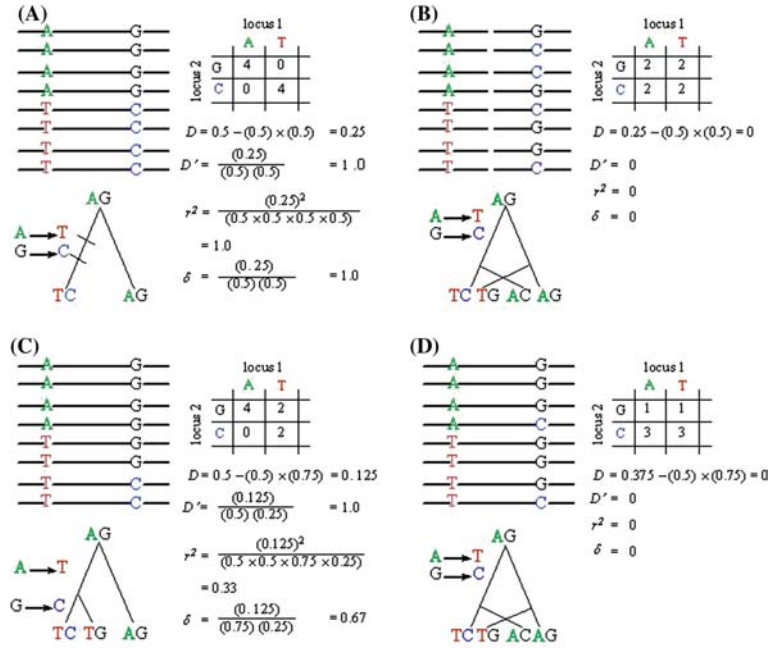

Figure 1. Diagrammatic representation of linkage disequilibrium (LD) between two SNPs showing behavior of D, D', r^2 and δ statistics under following conditions: (A) No recombination (mutations at two linked loci not separated in time); (B) Independent assortment (mutations at two loci not separated in time); (C) No recombination (only mutations separated in time); (D) Low recombination (mutations at two loci not separated in time).

these approaches that were proposed and used in the past. Therefore, more recently, an approach was suggested, where using an EM algorithm, ML estimates of gametic frequencies could be obtained and used for estimation of LD (Kalinowski and Hedrick, 2001). This approach has been successfully applied in some animal systems (e.g. sheep) and a modified form of EM algorithm (optimal step length EM; OSLEM; Zhang *et al*., 2003) has been successfully applied in some plant systems (e.g. tetraploid potato). In future, it will certainly be improved further and then used in other plants also (for details, see Kalinowski and Hedrick, 2001; Simko, 2004; Neale and Savolainen, 2004).

Multilocus methods

In recent years, there has also been emphasis on developing methods for using data from multiple loci for LD mapping, since LD data involving multiple loci will eventually be needed for preparing whole genome LD maps, in the same manner as the whole genome linkage maps were prepared in the past. The approaches for multilocus methods can be broadly classified into (i) 'bottom–up approaches', where we start with individual loci and measure multilocus LD, and (ii) 'top–down approaches', where we start with higher order LD coefficients and then decompose them into lower order LD-terms.

Bottom–up approaches

One of the earliest bottom–up approaches for multilocus measures of LD was λ (Terwilliger, 1995), which is very similar to δ described above. Several other bottom–up multilocus methods proposed later generally make use of covariance structure of marker loci for estimation of LD, and can be classified into the following: (i) *composite likelihood methods* (Devlin *et al*., 1996; Xiong and Guo, 1997); (ii) *least square methods* (Lazzeroni, 1998), and (iii) *haplotype segment sharing methods* (Service *et al*., 1999). These methods and their relative merits and demerits have been discussed by Jorde (2000). More recently, an entropy based method, described as Normalized Entropy Difference (NED), and symbolized by ε has also become

available for multilocus LD (Nothnagel *et al*., 2002, 2004). These multilocus methods for LD estimates as above can be either '*single point methods*' (using information from one marker at a time), or '*multipoint methods*' (using information from multilocus allele frequencies simultaneously). The multipoint methods, in their turn, may be based either on haplotypes (Lou *et al*., 2003) or on frequencies of individual alleles at many marker loci (Johnson, 2004). The haplotype-based multipoint method has also been specifically used for mapping QTL with epistasis (for more about epistasis and LD, see later). These multipoint methods for multilocus LD are still being developed and would be increasingly used in future in both animal and plant systems.

Multilocus methods can also be used for fine mapping of QTLs identified through interval mapping, since several polymorphic loci/genes may be present in an interval carrying a QTL/gene of interest. Similarly at the level of whole genome, a number of biallelic/multiallelic marker loci may occur with a number of biallelic/multiallelic QTLs for a trait. In this case, one can calculate all digenic and higher order (e.g., trigenic and quadrigenic, etc.) LD coefficients and utilize this information for fine mapping of QTLs. However, within the region of high LD, one would like to identify causal polymorphisms, excluding most of the other irrelevant markers/genes that are present. This has been achieved using several multilocus methods, where haplotypes and haplotype blocks, each with a number of loci, are used for study of LD (e.g., Morris *et al*., 2003).

Since multilocus methods would require genotyping data for many marker loci, we may collect data on allele frequencies either by using DNA pools (to reduce cost) or by typing every individual separately for each marker locus. If we use single point method, haplotype phase information is not important, and we can use pooled DNA, even though the method is imprecise, since allele frequencies are determined by peak heights only.

Top–down approaches

We know that co-adapted gene complexes provide a typical example of multilocus LD. Although, an algorithm for computing higher order LD among these gene complexes was provided rather early in a well-cited article (Geiringer, 1944), higher-order LD of these co-adapted complexes could not be quantified and decomposed into lower-order LD till recently (for related references, see Gorelick and Laubichler, 2004). These top–down approaches will also be increasingly used in future for LD studies in both plant and animal systems.

Statistical significance of LD

The association between the allelic states at two different loci can be tested using 2×2 contingency table for χ^2 test. Probability of $< 5\%$ would suggest lack of independence of alleles at two loci, thus indicating association. From a 2×2 contingency table, probability (P) of the independence of alleles at the two loci is generally also calculated through a Fisher's exact test (Fisher, 1935).

Statistical significance (P-value) for LD is also calculated using a multifactorial permutation analysis to compare sites with more than two alleles at either or both the loci (Weir, 1996). One should however, recognize that LD can be found even between unlinked loci, which may be due to the use of a structured population resulting due to selection (including epistasis), genetic drift, migration, mutation, etc. Methods are however, available to deal with this problem (see later for structured populations).

Two ways to visualize or depict the extent of LD

Since D' or r^2 are pair-wise measurements between polymorphic sites, it is difficult to obtain a summary statistics of LD across a region. Following two methods have been suggested to visualize or depict the extent of LD between pairs of loci across a genomic region: (i) *LD decay plots*, and (ii) *Disequilibrium matrices*. These two methods have been widely used, and the readers are referred to earlier reviews for details of these two methods (Flint-Garcia *et al*., 2003; Gaut and Long, 2003). The two methods, when used in specific cases give an idea about the pattern of LD decay in each case, and suggest that variation in LD depends on a variety of factors (see Figures 2, 3 for the two methods, and the next section for factors affecting LD).

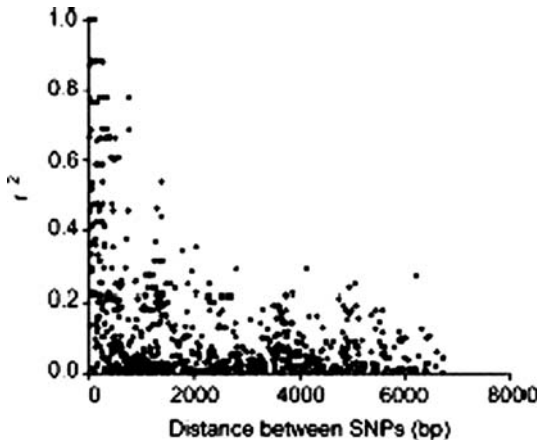

Figure 2. Linkage disequilibrium (LD) decay plot of *shrunken 1* (*sh1*) locus in maize. LD, measured as r^2, between pairs of polymorphic sites is plotted against the distance between the sites (reproduced with permission from Flint-Garcia *et al*., 2003).

Factors affecting LD

There are several factors that influence LD. The factors, which lead to an increase in LD, include inbreeding, small population size, genetic isolation between lineages, population subdivision, low recombination rate, population admixture, natural and artificial selection, balancing selection, etc. Some other factors, which lead to a decrease/disruption in LD, include outcrossing, high recombination rate, high mutation rate, etc. There are other factors, which may lead to either increase or decrease in LD, or may increase LD between some pairs of alleles and decrease LD between other pairs. For instance, mutations will disrupt LD between pairs involving wild alleles, and will promote LD between pairs involving mutant alleles. Similarly, genomic rearrangements may disrupt LD between genes separated due to rearrangement, but LD may increase between new gene combinations in the vicinity of breakpoints due to suppression of local recombination. Other factors affecting LD include population structure, epistasis, gene conversion and ascertainment bias. Since these other factors did not receive the desired attention in earlier reviews, and since they make the current issues of LD research in plants, these are separately discussed later in this review. The study of factors affecting LD as above is particularly relevant, if LD estimates need to be used to study linkage-based association, because one needs to rule out the possibility of factors other than

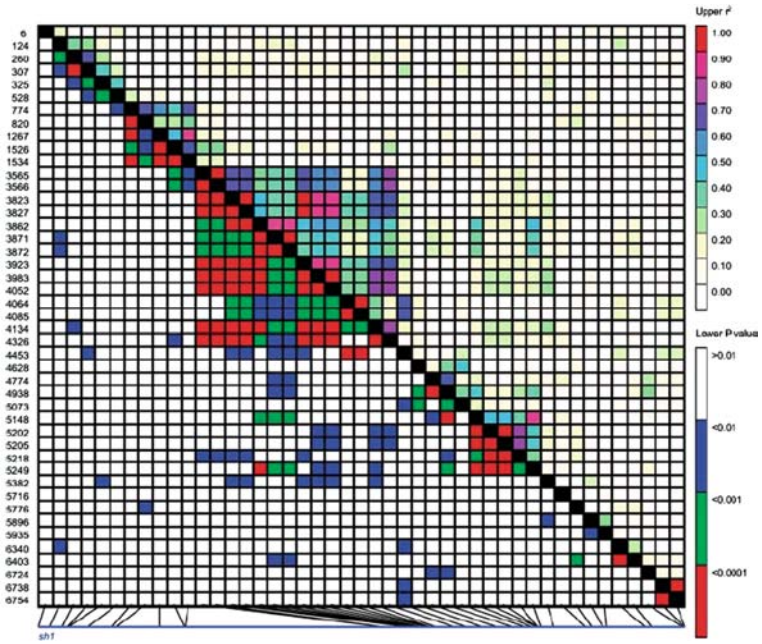

Figure 3. Disequilibrium matrix for polymorphic sites within *shrunken 1* (*sh1*). Polymorphic sites are plotted on both the X–axis and Y-axis. Pair-wise calculations of LD (r^2) are displayed above the diagonal with the corresponding *P*-values for Fisher's exact test displayed below the diagonal (reproduced with permission from Flint-Garcia *et al*., 2003).

linkage causing LD. These factors have been discussed elsewhere in greater detail (Ardlie *et al*., 2002; Jannink and Walsh, 2002; Weiss and Clark, 2002; Flint-Garcia *et al*., 2003; Gaut and Long, 2003) and were also recently listed by Rafalski and Morgante (2004).

Uses of LD in plant genomics

Linkage disequilibrium can be used for a variety of purposes in plant genomics research. One of the major current and future uses of LD in plants would be to study marker-trait association (without the use of a mapping population) followed by marker-assisted selection (MAS). Another important use is the study of genetic diversity in natural populations and germplasm collections and its use in the study of population genetics and in crop improvement programmes respectively. Some of these uses will be briefly discussed in this section.

Marker-trait association and MAS in plants

Marker-trait association in crop plants is generally worked out through linkage analysis, utilizing methods like *t*-test, simple regression analysis and QTL interval mapping, which have been widely discussed (see Melchinger, 1996; Hackett, 2002). Limitations of these methods have also been widely discussed (Darvasi *et al*., 1993; Hästbacka *et al*., 1994; Mackay, 2001; Hackett, 2002). These limitations have largely been overcome in LD-based approach of association mapping, which is going to be used extensively in plant systems, as and when the genome-wide sequences and/or SNP maps become available.

For a study of marker–trait association using LD, the methods may differ for discrete traits and quantitative traits, although sometimes quantitative traits may also be treated as discrete traits. Two methods that have been commonly used for discrete traits in human beings for mapping disease genes are (i) case–control (CC) and (ii) transmission/ disequilibrium test (TDT) (Spielman *et al*., 1993). Similar (but not identical) approaches have also been used in plant systems (Table 1). For instance, one such study involving discrete traits in plants was recently conducted in maize (Palaisa *et al*., 2003), in which 78 out of 81 informative SNP and *InDel* polymorphisms in *Y1* gene were found associated with endosperm color when genotyped over a set of 41 yellow/orange endosperm lines and 34 white endosperm lines. The methodology used in this study is comparable to that used in CC studies in humans. In another study conducted in radiata pine, 200 full sib families were used to study the marker–trait associations. In this study, the parental genotypes were also considered during analyses (Kumar *et al*., 2004), so that the method can be compared with TDT in humans.

The use of LD for mapping of QTLs for a quantitative trait is more problematic, but is also more rewarding, because it allows more precise location of the position of a QTL that controls the trait of interest. When comparing linkage analysis and LD mapping for QTL detection, it has been shown that linkage mapping is more useful for genome-wide scan for QTL, while LD mapping gives more precise location of an individual QTL. One may therefore like to use linkage analysis for preliminary location of QTLs and then use LD for more precise location (Mackay, 2001; Glazier *et al*., 2002; also see later for joint linkage/LD studies). LD between a single marker and a QTL can be measured by regression analysis, where the data on the trait is regressed on the individual marker genotypes, so that significant regressions will identify the markers associated with the phenotype (Remington *et al*., 2001). However, since this association of marker can sometimes be due to reasons other than linkage, we need to conduct further analysis to select markers that are really associated with the trait due to close linkage. This regression of the trait on the marker genotype therefore is sometimes examined by testing two adjacent markers for their association with the trait. In still other cases, we estimate the effect of marker haplotypes on the trait through regression analysis. Haplotypes having similar marker alleles (identical by descent), and associated with similar phenotypic effect should carry a QTL (Meuwissen and Goddard, 2000). Location of such a precise position within a very small chromosome region is possible through LD, but not through linkage analysis, since through linkage analysis, recombination within such a small region may not be available in a finite population that is examined (Mackay, 2001).

Table 1. A list of linkage disequilibrium (LD) studies conducted in plants.

Plant	Study	Salient features	Reference
Gene based studies			
Maize	Association between polymorphisms available in *Dwarf8* (*d8*) gene and flowering time (FT) in 92 inbred lines.	Association of polymorphism in *d8* with plant height and FT suggested that selection at *d8* led to early flowering (up to 7–11 days) in maize. Rapid LD decay was also observed suggesting no association between FT and *tb1* located just 1 cM away from *d8*.	Thornsberry *et al.* (2001)
	Survey of six candidate genes (*id1*, *tb1*, *d8*, *d3*, *sh1*, *su1*) to study the rate of LD decay.	Rapid LD decay was observed ($r^2 < 0.1$) within 1500 bp at *d3*, *id1*, *tb1*, *sh1*; at *su1*, r^2 is >0.4 even after 7000 bp (attributed to selection for kernel sugar and to location of *su1* near centromere); at *d8* LD decay was intermediate.	Remington *et al.* (2001)
	Survey of 18 gene segments to study SNP frequency, haplotype structure and LD in 36 elite maize inbred lines.	Strong LD between SNP loci extending to at least 500 bp was observed in this study, suggesting presence of population structure resulting due to bottlenecks created by selection exercised during plant breeding.	Ching *et al.* (2002)
	Nucleotide diversity (ND), estimation of population-recombination parameter C ($=4Nc$) and its relation with LD.	LD was significant at 5% level in 33 of 528 pair-wise comparisons, but, five associations remained significant after sequential Bonferroni correction. A negative correlation between population-recombination parameter C and LD was also observed.	Tenaillon *et al.* (2002)
	Effect of selection on sequence diversity and LD at two phytoene synthase genes *Y1* and *PSY2*.	At *Y1*, 19-fold difference in ND in white and yellow endosperm lines; association of SNP/*InDel* with the phenotype; at *PSY2*, no association of SNP/*InDel* with phenotype was observed. Positive selection for yellow endosperm was inferred.	Palaisa *et al.* (2003)
	Diversity, LD and association of yellow endosperm with region close to *Y1* gene.	Significantly reduced diversity and high level of LD (from *MLO* gene to *LCBE3*) was observed among SNPs upto 600 kb downstream of *Y1* gene; Significant association of SNPs with yellow endosperm was also observed upto 550 kb upstream and 700 kb downstream of the *Y1*.	Palaisa *et al.* (2004)
	Pattern of ND and LD in the genomic region near the domestication gene *tb1*.	LD only between sites in the region of selective sweep upstream to *tb1*. This study is consistent with the observation that LD typically decays rapidly within individual maize loci.	Clark *et al.* (2004)
	Survey of two *adh1* gene segments to study ND and LD.	High levels of ND and good correlation between LD decay and distance in base pairs were observed at both the gene segments.	Rawat, (2004)
	LD and sequence diversity in a 500 kb region around the *adh1* locus in elite maize germplasm.	Analysis of several loci in the vicinity of *adh1* gene shows that LD as measured by D' and r^2 extends $>$ 500 kb in the germplasm, suggesting either selection of one of the genes in the vicinity of *adh1* or a locally reduced rate of recombination.	Jung *et al.* (2004)

Barley	ND and LD at 3 *adh1–adh3* loci within a species-wide sample of 25 accessions of wild barley.	10-fold difference in ND between *adh1* and *adh3* loci, *adh2* being intermediate. Despite tight linkage between *adh1* and *adh2* loci, selection at *adh1* did not reduce ND (θ) at *adh2* locus; the three *adh* loci were inferred to have been subjected to different evolutionary forces.	Lin *et al.* (2001, 2002)
	LD near the *Ror1* gene.	LD extended ≥ 100 bp. Regions of highly localized recombination events around *Ror1* (close to the 1H centromere) was observed.	Collins *et al.* (2002)
	Structure of LD within genes/EST-derived sequences.	Sequences of 400–900 bp were resequenced from a set of 23 winter, 28 spring and 58 other cultivars. No LD decay was observed within a distance of 450 bp between SNPs.	Stracke *et al.* (2003)
	LD at *Ha* locus on chromosome 5H.	Study is still in progress to throw light on the pattern of ND/LD at *Hardness* locus in barley.	Caldwell *et al.* (unpublished).
Hexaploid wheat	LD analysis of gene sequences from *Ha* locus (group 5) as well as SSRs of group 7 chromosomes.	Study is still in progress to throw light on the pattern of ND/LD at *Hardness* locus in wheat.	Kruger *et al.* (2004)
	Effect of selection on LD in genes close to *VRN1* and *VRN2*.	Study still in progress to throw light on the pattern of ND/LD in *VRN1* and *VRN2* genes on chromosome 5A of wheat.	Dubcovsky (unpublished).
Rice	Population structure and its effect on haplotype diversity and LD surrounding *xa5* locus using genotypic data from 114 accessions for 21 SSRs.	Populations were highly structured. In 75 kb region of *xa5* locus, and additional 45 kb adjoining region, significant LD was observed between sites up to 100 kb apart (r^2 approached 0.1 only after 100 kb) . This is comparable to the level of LD in flowering time gene *FRIGIDA* in *A. thaliana*, but differs markedly from that of *dwarf8* (*d8*) and other genes in maize.	Garris *et al.* (2003)
	LD between *Est-2* and *Amp-3* loci using 2280 Asian rice varieties.	Strong LD; and few allelic combinations predominant. Some combinations also showed group specificity.	Glaszmann, (1986)
Arabidopsis	ND and LD at *Adh* locus; comparison of sequences from ecotypes Columbia and Landsberg.	High degree of LD and some evidence of recombination at *Adh* locus, suggesting role of balancing and directional selection.	Hanfstingl *et al.* (1994)
	LD at genes *AP3* and *PI* based on intraspecific sequence variation.	Highly intragenic LD at both *AP3* and *PI* genes; no intergenic LD	Purugganan and Suddith, (1999)
	Pattern of LD in populations from Michigan, using markers surrounding the disease locus *RPM1*.	Extensive LD was observed; LD on a genome-wide scale decayed over 50-100 cM.	Nordborg *et al.* (2002)
	LD within the locus *FRI*.	In 14 short fragments (0.5-1.0 kb) from a 400 kb region, LD decayed within 250 kb, equivalent to 1 cM; strong LD was observed between sites that were closely linked.	Hagenblad and Nordborg (2002)
	ND and LD at genes *FAH1* and *F3H* in a sample of 21 world wide ecotypes.	Significant LD for both the genes; high ND at the beginning of genes.	Aguade (2001)
	ND and LD in *CLV2* and 10 flanking genes spanning 40 kb region of chromosome I.	Wide variation in ND among 11 genes in the *CLV2* region; strong LD in a 40 kb region.	Shepard and Purugganan (2003)

Table 1. (Continued)

Plant	Study	Salient features	Reference
	LD at *CRY2* (flowering time) gene in a sample of 95 ecotypes.	Haplotype structure in the flanking region (containing six genes) of *CRY2*; high LD in 16 kb region around *CRY2* extending up to ~65 kb.	Olsen *et al.* (2004)
Lettuce	LD decay in *Dm3* resistance-gene cluster scored on 93 diverse lines.	Correlation between LD and genetic distance on the basis of haplotypes, but not on the basis of individual markers; gradual decay of LD between marker haplotypes within the *Dm3* locus, estimated to be ca. 0.2 cM and 200 kb.	van der Voort *et al.* (2004)
Potato	Association between haplotypes identified at *StVe1* locus and resistance to *V. albo-atrum*.	Three common haplotypes identified in a set of 30 potato cultivars at *StVe1* locus; only one haplotype showed significant association with resistance to *V. albo-atrum*.	Simko *et al.* (2004b)
	Association between RGA marker and resistance to potato late blight.	PCR-based RGA-derived markers corresponding to STH (Primer-10), glucanase and lipoxygenase showed association with resistance both in the diploid and tetraploid populations.	Manosalva *et al.* (2001)
	Association between DNA markers and agronomic characters in a collection of 600 potato cultivars.	PCR markers lying within *R1* (a major gene for resistance to late blight) or 0.2 cM from *R1* and introgressed from *S. demissum*, associated with QTL for resistance to late blight & plant maturity.	Gebhardt *et al.* (2004)
	Association between a SSR marker and *Verticillium* resistance in tetraploid potato.	An SSR marker in linkage with *StVe1* associated with QTL for resistance to *V. dahliae* over a set of 137 tetraploid potato cultivars; study led to cloning of QTL for resistance to *V. dahliae*.	Simko *et al.* (2004a)
Grapevine	Level of LD surrounding candidate genes (CGs) for fruit quality traits.	SNPs, identified in 15 CGs and their flanking regions used for genotyping 50 individuals varying in fruit quality, thus helping in studying genetics of traits, MAS, and genome evolution.	Owens (2003a, b; 2004)
Maritime pine	ND, structure and LD in 17 CGs for wood quality traits using samples from 13 provenances of France.	210 polymorphic sites (SNPs and *INDELs*) were detected with one site each per 80 bp and 25 bp in coding and noncoding regions; rapid decrease of LD between sites within most genes.	Garnier-Géré *et al.* (2003)
Scots pine	Study of ND and LD in 12 genes.	High level of haplotype diversity and low LD was observed even between closely linked loci.	Dvornyk *et al.* (2002)
Loblolly pine	Estimation of LD in different genes of *P. taeda*.	LD varies among several genes in *P. taeda*; on an average its values (in terms of r^2) decays to less than 0.20 within ~1500 bp.	Neale and Savolainen (2004)
Genome wide or chromosome wide studies			
Maize	LD between RFLP loci in two synthetic populations of maize.	Two populations were derived from 12 to 16 inbred lines through recurrent selection for 12 generations. LD substantially increased in one population and decreased in the other.	Labate *et al.* (2000)
	LD across the genome measured through LD among 47 SSRs.	High LD was observed, suggesting that SSRs may track recent population structure better than the relatively older SNPs.	Remington *et al.* (2001)
	Pattern of variation at 21 loci; impact of selection, recombination, and LD on sequence diversity.	LD measured as r^2 decreased to <0.25 within 200 bp on an average. Little LD existed between loci, although all the 21 loci were present on chromosome 1.	Tenaillon *et al.* (2001)

	Genetic structure and diversity among 260 maize inbred lines partitioned into five groups.	Diversity was high among tropical lines; significant LD in 66% of SSR pairs was attributed to linkage (structure) within groups.	Liu *et al.* (2003)
	Pattern of ND at 12 loci on chromosome 1; impacts of directional selection and population bottleneck.	Loci under selection identified using multilocus approach; *tb1* and *d8* were confirmed as sites of selection during domestication/breeding; initial domestication, ~500 years ago also confirmed.	Tenaillon *et al.* (2004)
Barley	Extent of LD in barley using 23 winter, 28 spring and 58 other cultivars.	A genome wide survey showed LD decay within 10 cM using SNPs and within 20 cM using SSRs.	Stracke *et al.* (2003)
	LD in 5H of barley using >50 SSRs and SNPs.	Study still in progress and will throw light on the pattern of LD at chromosome 5H of barley.	Ramsay *et al.* (2004)
	SSR marker data for LD and population structure.	Study still in progress and will throw light on the pattern of LD and population structure	Mather *et al.* (2004)
	Use of 33 SSR markers to study association with flowering time and other adaptive traits.	SSRs significantly associated with flowering time under four growing regimes; most associations could be accounted for by close linkage of the SSR loci to earliness *per se* genes.	Ivandic *et al.*(2002)
	Association analysis of a set of RCSLs of *H. vulgare* ssp. *vulgare* with introgressions from *H. vulgare* ssp. *spontaneum.*	Through association analysis (AA), only marker–phenotype relationship with highest correlation was identified as significant, suggesting that AA is more conservative (i.e. generates fewer false positives) than simple linear regression and should be used for marker–phenotype relationships.	Matus *et al.* (2003)
	236 AFLP markers to map yield and yield related characters through LD in 146 spring barley cultivars.	Associations between markers 10 cM apart; many markers showing association with the trait of interest reside in the chromosomal region, where QTLs for the same trait were recorded earlier.	Kraakman *et al.* (2004)
	LD between EST-SNPs and haplotypes in three germplasm groups of barley representing European cultivars, land races and wild accessions.	Lowest ND observed in *H. vulgare* (HV) lines and highest for *H. spontaneum* (HS) lines; significant LD between loci in cultivated HVs, and its absence in landraces of HV and HS ; more haplotypes in HS accessions than in cultivars and landraces of HVs.	Russell *et al.* (2004)
Tetraploid wheat	LD in a durum wheat collection using 70 SSRs.	High level of LD both in syntenic and nonsyntenic pairs of loci; *D'* averaged 0.67 for marker pairs < 10 cM apart and 0.43 for pairs with a 10–20 cM distance.	Maccaferri *et al.* (2004)
	LD among SSRs and its relation with eco-geographical distribution.	Niche-specific LD in two edaphic subpopulations (terra rossa and basalt), suggesting adaptive molecular pattern due to edaphic selection leading to niche-specific LD.	Li *et al.* (2000)
Hexaploid wheat	LD pattern in group-7 chromosomes using SSRs.	Study still in progress and will throw light on LD pattern at group 7 chromosomes of bread wheat.	Kruger *et al.* 2004)
Rice	Genetic diversity and population structure using 169 nuclear loci & 2 chloroplast loci in 236 *O. sativa* and 93 SSR loci in 198 *O. glaberrima* accessions.	Accessions were partitioned into five groups each overlapping the partitions based on phenotype; the study suggested that a higher degree of resolution of population structure is needed to effectively utilize LD for association mapping.	McCouch *et al.* (2004)

Table 1. (Continued)

Plant	Study	Salient features	Reference
Sorghum	LD among 12 loci within a 4 cM/600 kb distal region using RFLPs among 205 accessions.	LD commonly extended beyond 2 cM and decayed within 4 cM in the targeted segment. This will allow localization of favorable variations in the genome.	Deu and Glaszmann (2004)
	Sequence variation and LD in 95 short genomic regions (123–444 bp) aggregating to 29186 bp, sequenced in 27 *S. bicolor* elite inbred lines.	Low level of variation (one-fourth that in maize) and several fold higher level of LD, relative to maize was observed; LD decays within 10 kb or less so that sorghum has a pattern intermediate between maize and *Arabidopsis*; High and low levels of intralocus and interlocus LD were observed.	Hamblin *et al.* (2004)
Ryegrass	Association between phenotype and marker in natural populations.	26 of 589 polymorphic AFLPs significantly associated with heading date; 5 of the 6 most significantly associated markers were mapped near two QTLs for heading date.	Skøt (2004)
Oat	Association between marker and QTs and its relationship with linkage in an oat germplasm pool.	10 of 31 mapped RFLPs showed marker–trait association, suggesting that in most cases associations are due to random drift, parallel natural, or artificial selection, etc. rather than linkage.	Beer *et al.* (1998)
Sugarcane	LD between RFLPs (38 probes) regularly distributed over genome using 59 modern cultivars.	42 cases of bilocus associations involved 33 loci, most separated by >10 cM; 14% of associations had RFLPs from different chromosomes; LD interpreted as the result of foundation bottleneck.	Jannoo *et al.* (1999, 2001)
Arabidopsis	SSR polymorphism and LD using 20 mapped SSRs in a worldwide sample of 42 ecotypes.	23 (12.1%) of 190 pair-wise comparisons were found significant; high level of LD at SSR loci was attributed to allele frequency differences in Japanese ecotypes.	Innan *et al.* (1997)
	LD between AFLPs; recombination vs pattern of DNA polymorphism in 38 ecotypes.	11.4% of 374 polymorphic bands from 472 bands detected), showed significant LD. Effect of recombination events on the pattern of DNA polymorphism was inferred.	Miyashita *et al.* (1999)
	Genome-wide LD using 76 accessions genotyped for 163 SNPs.	LD persisted for up to 250 kb, although, in a specific 500 kb region, LD decayed within 50 kb.	Nordborg *et al.* (2002)
	Genetic diversity and LD at 14 loci, across 170 kb centered on a QTL for resistance to herbivory.	Pair wise LD between polymorphisms were negatively correlated with distance (not for all combinations) suggesting role of gene conversion on genetic diversity throughout the region.	Haubold *et al.* (2002)
Soybean	Haplotype diversity and LD in cultivated and wild soybean.	Similar diversity in N. Am. and Asian *G. max* genotypes; twice the number of haplotypes found in wild soybean (*G. soja*) suggesting LD decay over short distances.	Cregan *et al.* (2002)
	SNP frequency in coding and noncoding sequences and LD in 25 diverse genotypes.	r^2 estimates among haplotypes (each with two or more SNPs) at 54 loci spread over 76.3 kb (coding and non-coding sequences) suggested a low genome-wide LD and limited haplotype diversity.	Zhu *et al.* (2003)
	LD in four different populations: (a) *Glycine soja* (GS); (b) Asian *G. max* (AGM); (c) N. Am. ancestors (NAA); (d) N. Am. Public (NAP) cultivars; multiple fragments throughout 500 kb region were sequenced.	Slow LD decay within 500 kb in NAA cultivars in comparison to NAP population; LD declined rapidly in GS population; intermediate level of LD decay (> 350 kb) in AGM. These would make association analysis feasible in populations similar to AGM and NAP, while fine mapping of genetic factors would be possible in GS population.	Hyten *et al.* (2004)

Sugar beet	Relationship between linkage and LD among 451 mapped AFLP markers in nine sugar beet breeding lines.	Significant but low LD between unlinked markers and significantly higher LD between linked markers (< 3 cM).	Kraft *et al.* (2000)
	LD in organellar genomes (LD between chloroplast and mitochondrial DNA haplotypes).	Strong LD in cpDNA and mtDNA polymorphisms, suggesting that homoplasy in mtDNA was mainly due to migration and that like cpDNA, mtDNA can be used for population genetics studies.	Desplanque *et al.* (2000)
	LD mapping of *B* genes using 137 mapped AFLP markers in 106 sea beets.	Two markers showed significant LD with *B* gene. Results indicated the potential use of LD for gene mapping in natural plant populations.	Hansen *et al.* (2001)
Eucalyptus	Marker–trait association for traits like volume, growth, stem taper and wood quality.	LD between QTL allele and marker allele to select parents for further breeding, and to choose the hybrid families in which QTAs of specific value could be detected and used to identify the best trees.	Verhaegen *et al.* (1998)
Loblolly pine	ND, LD and adaptive variation in natural population using DNA samples sequenced from 32 megagametop- hytes sequenced for 20 drought tolerance related genes	ND high; haplotypes were identified and a set of SNPs distinguishing haplotypes was selected. Screening of SNPs in 435 trees provided relevant information about differential selection and adaptive processes related to drought tolerance.	González-Martínez *et al.* (2004)
Radiata pine	LD at population level and marker–trait associations for MAS using 34 SSRs.	Weak marker–trait LD was observed and could be due to genetic-sampling effects; further validation of markers showing putative association with the trait is needed.	Kumar *et al.* (2004)
	Pattern of LD within specific genes, across a chromosome and across genome.	Intragenic LD extended over short stretches of DNA and between linked markers; no evidence of chromosome and genome-wide LD was obtained.	Wilcox *et al.* (2002, 2004)
Norway spruce	Digenic disequilibrium and the spatial pattern within three *Picea abies* stands using allozyme loci.	More cases of digenic LD were found than were expected; non-random mating was suggested as the possible reason of LD among unlinked or weakly linked loci.	Geburek (1998)
	LD using 20 mapped RAPD loci genotyped over megagametaphytes from 48 trees from Italy.	LD among 5 out of 20 RAPD markers was observed. A weak spatial structure was detected; the study suggested lack of 'family structure' due to isolation-by-distance.	Bucci and Menozzi, (1995)
	Sequence diversity and SNP marker development in Norway spruce.	This study will throw light on the possibilities for the development of SNP markers for practical applications like association studies.	Ivanissevich and Morgante (2001)

Population genetics and evolutionary studies in plants

The neutral theory of evolution holds that majority of polymorphisms observed within and among species are selectively neutral or at least nearly so (Tajima, 1989). Neutrality makes mathematical modeling easy giving a natural null model. Features, like selection, migration and demographic history can then be viewed as perturbation of a standard neutral model.

Natural selection and domestication

In any organism, LD can be used for identifying genomic regions, which have been the targets of natural selection (both directional selection and balancing selection) during evolutionary process. Adaptive selection can leave one of two signatures on a gene region through genetic hitchhiking. Directional selection can reduce levels of polymorphism through the rapid fixation of a new adaptive mutation. Balancing selection can increase levels of polymorphism when two or more alleles are maintained longer than expected under a neutral model. For instance, if a polymorphism maintained by balancing selection is old, it will have enhanced sequence variability in the flanking regions, which may be used as a 'signature of selection'. Due to difficulties inherent in such studies, only very few such studies have been conducted in the past, but more such studies will certainly be conducted in future. One of the difficulties in such studies is due to similar pattern of genetic variation expected due to natural selection on the one hand and population demographic history (size, structure and mating pattern) on the other, although selection affects specific sites, while demography affects the entire genome. Despite these difficulties, the data on human genome sequences and the available SNPs in these sequences made it possible to identify genome-wide signatures of selection in humans (Akey *et al*., 2002; Schlotterer, 2003). In one study in humans, 174 candidate genes were inferred to have been the target of selection (Akey *et al*., 2002). Similar studies have been conducted in *Arabidopsis*, where genomic regions containing the genes *RPM1* and *RPS5* were shown to be the target of selection (Stahl *et al*., 1999; Tian *et al*., 2002; Mauricio *et al*., 2003). More such genomic regions, which have been the targets of selection, are likely to be identified in future, so that we will have a complete set of genes with long lived polymorphisms, each with a region that has been a target of natural selection.

In crop plants, efforts have also been made to identify genomic regions or genes, which were the targets of selection during domestication and subsequent selective breeding. For instance, QTLs for agronomic traits that were selected during domestication were identified through QTL interval mapping (Paterson *et al*., 1995; Peng *et al*., 2003; for a review, see Pozzi *et al*., 2004), even when functions of these genomic regions are unknown. For instance, in a study in maize, as many as 501 genes were screened using 75 EST-SSRs, to obtain signatures of selection. Fifteen of these 75 EST-SSRs gave some evidence of selection (Vigouroux *et al*., 2002). In another study in maize, variability seems to have been reduced in a short regulatory region that lies 5′ upstream of the *teosinte branched1* (*tb1*) locus (Clark *et al*., 2004). Large differences in the pattern of polymorphism between genomic regions are also seen in barley (Lin *et al*., 2001).

Demographic history

It is also possible to infer demographic history of a population from the pattern of DNA polymorphism, if data from a number of independent (unlinked) loci is used and it is assumed that the demographic history affects the entire genome in the same way. Furthermore, it is shown that for a study of demographic history, a large number of loci spread over the whole genome should be used, since it was shown that study involving single locus (or non-recombining genomes represented by mtDNA or cpDNA) may lead to erroneous conclusions. For instance, in *A. thaliana* early studies of recombination at the *Adh* locus indicated extreme population subdivision (Innan *et al*., 1996), but this pattern was not observed in the entire genome in subsequent studies, where a survey of genome-wide AFLPs in *Arabidopsis* suggested a weak isolation by distance and a relatively recent population expansion, indicating ancient subdivision and recent expansion (Miyashita *et al*., 1999; Innan *et al*., 1999).

LD studies conducted in higher plants

Linkage disequilibrium studies have now been conducted in more than a dozen plant systems, both at the individual gene level and at the level of whole genome. In individual species, these studies included (i) estimation of the extent of LD in different plant genomes or in different parts of the genome of an individual species (see Table 2), (ii) measure of nucleotide diversity/haplotype structure, (iii) assessment of the effect of selection/domestication, (iv) identification of marker-trait associations, etc. Results of all these studies are summarized in Table 1 and will not be discussed any further.

Current issues in LD research in plants

As discussed above, there are several limitations in using LD in plant systems despite its demonstrated benefits. These limitations have become the current issues of LD research both in animal and plant systems. Among these limitations, effects of structured populations, epistasis, gene conversion and ascertainment bias on LD estimates have lately become the issues of current interest and are therefore briefly discussed.

Association mapping in structured populations

A population is described as a structured population, when frequencies of a disease or a trait varies across subpopulations, thus increasing the probability of sampling a trait from one subpopulation relative to that of sampling it from another. 'Transmission/disequilibrium test' (TDT) was suggested as one solution to this problem (Spielman *et al.*, 1993; Spielman and Ewens, 1996; Allison, 1997), but one generally prefers to conduct case–control studies, since these have several advantages (Cardon and Bell, 2001) and are cheaper. Therefore methods, employing case–control studies, have been developed for association mapping in structured populations.

Pritchard *et al.* (2000) proposed a population-based method that can detect associations between marker alleles and phenotypes in structured populations. The essential idea of the method is to decompose a sample drawn from a mixed population into several unstructured subpopulations and test the association in the homogeneous subpopulations. The methods have been applied to association analyses in humans (Rosenberg *et al.*, 2002) and crop plants, with modified test statistics being used to deal with quantitative traits (Thornsberry *et al.*, 2001).

Table 2. Linkage disequilibrium (LD) in different plant species.

Species	Mating system	LD range	Reference
Maize	Outcrossing	0.5–7.0 kb	Remington *et al.* (2001), Ching *et al.* (2002), Palaisa *et al.* (2003)
	Outcrossing	0.4–1.0 kb	Tenaillon *et al.* (2001)
Barley	Selfing	10–20 cM	Stracke *et al.* (2003), Kraakman *et al.* (2004)
Tetraploid wheat	Selfing	10–20 cM	Maccaferri *et al.* (2004)
Rice	Selfing	100 kb	Garris *et al.* (2003)
Sorghum	Selfing	<4 cM	Deu and Glaszmann (2004)
	Selfing	$\leq$10 kb	Hamblin *et al.* (2004)
Sugarcane	Outcrossing/Vegetative propagation	10 cM	Jannoo *et al.* (1999)
Arabidopsis	Selfing	250 kb	Nordborg *et al.* (2002)
Soybean	Selfing	>50 kb	Zhu *et al.* (2003)
Sugar beet	Outcrossing	<3 cM	Kraft *et al.* (2000)
Potato	Selfing	0.3–1.0 cM	Gebhardt *et al.* (2004), Simko (2004)
Lettuce	Selfing	$\sim$200 kb	van der Voort *et al.* (2004)
Grape	Vegetative propagation	>500 bp	Rafalski and Morgante (2004)
Norway spruce	Outcrossing	$\sim$100–200 bp	Rafalski and Morgante (2004)
Loblolly pine	Outcrossing	100–150 bp	González-Martínez (2004)
Loblolly pine	Outcrossing	$\sim$1500 bp	Neale and Savolainen (2004)

Epistasis and G × E interactions

Epistasis and G × E interactions are ubiquitous in the genetic control of complex traits and can be studied using a variety of approaches including LD. This aspect has only recently attracted the attention of those using LD for genetic dissection of quantitative traits. It should be recognized that epistatic interactions will lead to LD between loci involved in these interactions, since selection for the trait will allow these loci to stay together even when they are not linked (inter- and intra-chromosomal); these associated loci can be identified through LD and epistatic interactions between them can be identified. In most of the QTL studies involving estimation of epistatic effects, QTLs having main effects are identified first, and then they are tested for interactions. In recent years, methods have been developed and utilized in plants to find such QTLs, which do not have their main effect, but are involved in epistasis (Yu *et al*., 1997; Wang *et al*., 1999; Xing *et al*., 2002; Kulwal *et al*., 2004; for a recent review, see Carlborg and Haley, 2004). In most of these studies involving study of epistasis, however, mapping populations are used, but the approach of LD can be applied for the study of epistasis even in natural populations like those of forest trees, and in diverse germplasm collections relevant to crop improvement.

It was actually shown that in some cases, adjacent genes had low levels of interlocus LD and loosely linked genes had high levels of interlocus LD, suggesting strong epistatic selection. This aspect of the effect of selection of epistatic loci on LD has been particularly studied using inversions heterozygotes in *Drosophila pseudoobscura* (Schaeffer *et al*., 2003), where inversions act as suppressors of recombination to maintain positive epistatic relationships among loci within inverted regions that provided adaptation to a heterogeneous environment. A haplotype-based algorithm has also been proposed for multilocus analysis of LD mapping of QTLs involved in epistasis. The application of this method was validated using a case study, where QTL affecting human body height were successfully detected. In this study, modeling was done for epistatic QTL with gene effects including additive, dominant, additive × additive, additive × dominant and dominant × dominant effects (Lou *et al*., 2003).

Gene conversion and LD

Gene conversion can also be an important factor in shaping fine-scale patterns of LD and haplotype structure of a population. Attempts, therefore, are being made to understand the effects of high rates of gene conversion upon LD maps. Two loci separated by a short distance will generally exhibit low recombination, and therefore, are expected to show almost complete LD and complete linkage. Contrary to this expectation, in several recent studies, a significant fraction of closely placed loci were found to show incomplete LD. This unexpected result has often been found to be due to gene conversion, which can be distinguished from crossing over due to non-availability of all the four possible combinations involving two pairs of alleles. Further, the new combinations obtained due to gene conversion, are not associated with changes in loci downstream of the recombination break point, as is the case with reciprocal recombinants resulting due to crossing over.

It has also been shown that the rate of recombination between nearby loci often increases due to gene conversion, which may often assume alarming levels, its rate in different regions of a genome being as high as 1.5–10 times the rates of crossing over, as exemplified by chromosome 21 in humans (Padhukasahasram *et al*., 2004). This high rate of gene conversion will often lead to breakdown of allelic associations across short distances, thus reducing the magnitude of LD between closely linked adjacent loci. Gene conversion hot spots have also been found to be coincident with the previously identified crossover hot spots, suggesting that in many cases high recombination rates were erroneously attributed to crossing over.

Among plant systems, in order to study the effect of gene conversion (relative to that of crossing over) on LD, Haubold *et al*. (2002) sampled genetic diversity at 14 loci (500 bp each) in chromosome V of *Arabidopsis thaliana*. These 14 loci were distributed across 170 kb of genomic sequence centered on a QTL for resistance to herbivory. It was shown that in this particular genomic region, LD decays with distance (negative correlation). However, when only those pairs of loci were considered, where all the four possible haplotypes (these can result only due to crossing over and not due to gene conversion) were available, this negative correlation between LD and

genetic distance disappeared; and sometimes even became positive. When a test for the relative rates of gene conversion and reciprocal recombination (crossing over) was applied in this study, 90% of the recombination events in the region surveyed were found to have been produced by gene conversion. This strongly suggested that (epistatic) selection together with gene conversion must have produced this pattern.

Ascertainment bias and LD

Ascertainment bias (AB) is the bias introduced by the criteria used to select individuals and/or loci in which genetic variation is assayed, so that it leads to inaccurate estimates of LD. Ascertainment is the way individuals with a trait are selected or found for genetic studies and bias is a difference between the estimated and true value of LD in a statistical sample. Understanding and correcting this AB is essential for a useful quantitative assessment of the landscape of LD across any genome. Specifically, the magnitude of this AB is a function of several factors (Akey *et al*., 2003). A particular problem of AB arises when SNPs identified in small heterogeneous panels are subsequently typed in larger population samples. LD estimates may also be biased depending on the means by which SNPs are first identified to be used in further studies, where genotyping is done using entirely different methods. For instance, SNPs may be first identified by re-sequencing in one population and may be later scored in other populations by genotyping using methods other than re-sequencing. It is important to realize, that SNPs are often identified by *in silico* methods that ascertained SNPs from a small number of chromosomes in a limited number of populations (Taillon-Miller *et al*., 1998; Mullikin *et al*., 2000 for related references, see Akey *et al*., 2003; Kreitman and Rienzo, 2004). Inferences drawn from studies using such SNPs may be influenced by AB.

Rapid progress has been made in quantifying the pattern of LD and haplotypes across entire human genome, and similar efforts are being made in plant systems. The quality and utility of such a proposed LD-based resource could be seriously compromised, if important sampling and analytical factors as above are overlooked. To date, the effect of AB on estimates of background LD has not been rigorously investigated, although Weiss and Clark (2002) pointed out the problem of AB in results of an earlier study aimed towards characterizing the pattern of LD in human genome (Reich *et al*., 2001).

Ascertainment bias can be quantified as the mean absolute fractional error (MAFE), which varies from 0 to 1. Using this measure of AB, it was also shown that the magnitude of AB was higher in the hierarchical approach (when number of chromosomes from a single population are sampled) relative to that in a balanced approach. Therefore, the use of a sample of large number of chromosomes from multiple subpopulations was recommended for future large-scale SNP discovery for estimations of LD (Akey *et al*., 2003).

Future prospects

Association studies and MAS

One of the major uses of LD-based association analysis in future will be the study of marker-trait associations, leading to MAS, which has already been discussed earlier in this review. The approach will be particularly useful in forest trees, where mapping populations can not be easily generated, but MAS will prove extremely useful. For this purpose, LD will also facilitate development of functional markers (FMs), which are the perfect markers for marker-trait association (see Andersen and Lübberstedt, 2003; Gupta and Rustgi, 2004; Simko *et al*., 2004b).

Mapping of QTLs jointly using linkage and LD

In plant genetic studies, different QTL mapping methods, which were developed and extensively used in the past, have been very successful for mapping QTLs within genetic distances that measured only up to 10-30 cM (Alpert and Tanksley, 1996; Stuber *et al*., 1999). However, to utilize QTL in selective breeding or to identify functional genes, a higher level of resolution of position estimates is required. Association studies based on LD may allow mapping at much finer resolution (Remington *et al*., 2001). More recently, it was realized that linkage analysis (LA) and LD mapping both have their own limitations when used alone. The limitations of LA have been

discussed elsewhere (Darvasi *et al*., 1993; Hästbacka *et al*., 1994; Mackay, 2001; Hackett, 2002). However, the major limitation of LD mapping is that it provides little insight into the mechanistic basis of LD detected (e.g., LD may not be due to linkage in all cases), so that, genomic localization and cloning of genes based on LD may not always be successful. This is because a strong LD may sometimes be due to recent occurrence of LD rather than a close physical linkage between the two loci, exhibiting LD. Therefore, a new joint linkage and LD mapping strategy has been devised for genetic mapping, taking advantage of each approach (Wu and Zeng, 2001; Wu *et al*., 2002). The strategy has the power to simultaneously capture the information about the linkage of the markers (as measured by recombination fraction) and the degree of LD created at a historic time. In this approach, a random sample from a natural population and the open pollinated progeny of the sample are analyzed jointly. The approach is based on the principle that during the transmission of genes from parents to progeny, linkage between marker and QTL is broken down due to meiotic recombination. Therefore, it was proposed to have a composite measure of LD involving the following two components: (i) linkage between marker and QTL and (ii) LD created at a historic time. With the measurement of these two components, one can clearly determine the basis of a significant LD between marker and QTL, thus increasing the feasibility of fine mapping and map based cloning of QTLs affecting a QT. Thus, by combining the information about linkage and LD, the joint mapping method displays increased power to detect LD compared to traditional methods of LD analyses. Using this strategy a QTL with major effect on milk fat content was successfully fine-mapped in a 3 cM marker interval on bovine chromosome 14 following multipoint maximum likelihood approach (Farnir *et al*., 2002).

The approach of combined LA and LD for QTL analysis has been extended for multitrait fine mapping of QTLs (Lund *et al*., 2003; Meuwissen and Goddard, 2004). Multitrait QTL mapping has also been recommended for correlated traits, thus increasing the statistical power of detection, and resolving whether the two traits are correlated due to pleiotropic effect of one QTL or due to linkage between QTLs affecting these traits (Jiang and Zeng, 1995; Korol *et al*., 2001). Lou *et al*. (2003) also suggested that if their haplotype-based algorithm for multilocus LD mapping of QTLs is integrated with Wu and Zeng's (2001) model, the relationship between linkage and linkage disequilibrium can be tested, and LD mapping can be made a more predictable and powerful approach. The approach of joint LA and LD has already received considerable attention in studies on animals, and in future the method will certainly be used in plants also.

Haplotype blocks and tagging SNPs

As the number of known SNP markers in a genome increases, genotyping individuals with all the available SNPs will become a formidable task. Several approaches are being suggested to deal with this problem. For instance, patterns of LD (haplotype blocks) are being used for identification of minimum informative subsets of SNPs, also known as tagging SNPs (tSNPs) or haplotype tagging SNPs (htSNPs) (Goldstein *et al*., 2003; Tishkoff and Verrelli, 2003; Halldorsson *et al*., 2004). To identify a minimum set of SNPs (tSNPs/htSNPs), distributed throughout the genome for association testing is one of the major goal of haplotype map (HapMap) project in human (Clark, 2003; The International HapMap Consortium, 2003). Similar studies have been initiated in *A. thaliana* in USA. Other similar studies, at the level of genes were conducted in maize (Ching *et al*., 2002; Palaisa *et al*., 2003; see Rafalski and Morgante, 2004), rice (Garris *et al*., 2003), soybean (Zhu *et al*., 2003), potato (Simko *et al*., 2004b), etc. The efficiency of SNP haplotype analysis may also be increased by DNA pooling and use of microarrays, which can dramatically reduce the number of genotyping assays (Yang *et al*., 2003; Butcher *et al*., 2004).

Linkage disequilibrium maps in plants

Genetic and physical maps of genomes, based on molecular markers have now been constructed in all major crops. The work on the construction of LD maps in humans has already started, but that of the construction of LD maps for plant genomes has yet to start. In humans, LD maps of small regions of the genome or those involving mapping of disease genes relative to molecular markers have been constructed

successfully. In due course of time such mapping will be attempted in plants also. These LD maps will make use of molecular markers that flank marker intervals delimited on the basis of estimations of LD, the distances being represented as LD units (LDU; Zhang *et al*., 2002). LD mapping theory extends the estimation of covariance D for a random sample of haplotypes or diplotypes (disomic genotypes) to the association probability $\rho = D/Q(1-R)$, where D is an estimation of LD (see above), Q is the frequency of the rarest and therefore putatively the youngest allele, and R is the frequency of the associated marker allele (Maniatis *et al*., 2002). The estimates of these three parameters D, Q and R will be utilized for LD mapping. The softwares ALLASS (**all**ele **ass**ociation) and LDMAP VERSION 0.1, March 2002 (both developed by Andrew Collins from University of Southampton, UK) are recommended for use in constructing LD maps.

Appropriate statistical models

Although significant progress has been made in the methods for estimation and interpretation of LD, these methods each suffers from one of the following limitations: (a) They are based on computing some measure of LD defined only for pairs of sites, rather than considering all sites simultaneously. (b) They assume a 'block like' structure for patterns of LD, which may not be appropriate for all loci. (c) They do not directly relate patterns of LD to biological mechanisms of interest, such as recombination rate. Statistical models have also been proposed to overcome these limitations, by relating genetic variation in a population sample to the underlying recombination rate (Li and Stephens, 2003).

Softwares for LD studies

It is apparent from the above discussion that any LD study would be computationally demanding. For this purpose, newer data mining tools and other web resources are being regularly developed. Some of the tools, relevant for estimating LD are listed in electronic supplementary material (ESM, Table 1).

Conclusions

Linkage disequilibrium has been extensively utilized for a variety of purposes including mapping of disease QTLs in humans, but its use in plants has just begun. 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, we are at the threshold of utilizing the approach of LD and association mapping in crop plants in a big way. The approach will be used in several plant genomes for construction of LD maps, for study of marker-trait association both independently and in combination with linkage analysis and for the study of population genetics and evolution both in nature and under domestication. Future studies of LD in crop plants will also elucidate further the structures of plant genomes and will also facilitate the use of MAS and map based cloning of genes for difficult traits.

Acknowledgements

The work was done during the tenure of PKG as Senior Scientist of Indian National Science Academy, New Delhi, and that of PLK as Senior Research Fellow of Council of Scientific and Industrial Research, New Delhi. SR was employed as Junior Research Fellow in an R & D project (BT/AB/03/AR/2002) of the Department of Biotechnology, Government of India. Thanks are also due to Head, Department of Genetics and Plant Breeding, CCS University, Meerut for providing facilities. Careful reading and critical comments from the two anonymous reviewers helped in the improvement of the manuscript.

References

Abdallah, J.M., Goffinet, B., Cierco-Ayrolles, C. and Pérez-Enciso, M. 2003. Linkage disequilibrium fine mapping of quantitative trait loci: A simulation study. Genet. Sel. Evol. 35: 513–532.

Aguade, M. 2001. Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. Mol. Biol. Evol. 18: 1–9.

Akey, J.M., Zhang, K., Xiong, M. and Jin, L. 2003. The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium. Mol. Biol. Evol. 20: 232–242.

Akey, J.M., Zhang, G., Zhang, K., Jin, L. and Shriver, M.D. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Res. 12: 1805–1814.

Allison, D.B. 1997. Transmission disequilibrium tests for quantitative traits. Am. J. Hum. Genet. 60: 676–690.

Alpert, K.B. and Tanksley, S.D. 1996. High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: A major fruit weight quantitative locus in tomato. Proc. Natl. Acad. Sci. USA 93: 15503–15507.

Andersen, J.R. and Lübberstedt, T. 2003. Functional markers in plants. Trends Plant Sci. 8: 1360–1385.

Ardlie, K.G., Kruglyak, L. and Seielstad, M. 2002. Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet. 3: 299–309.

Beer, S.C., Siripoonwiwat, W., O'Donoughue, L.S., Souza, E., Matthews, D. and Sorrells, M.E. 1998. Association between molecular markers and qualitative traits in an oat germplasm pool: Can we infer linkages? J. Agric. Genomics e# 1 http://www.ncgri.org/ag/jag/Index.html.

Bucci, G. and Menozzi, P. 1995. Genetic variation of RAPD markers in a Norway spruce *Picea abies* Karst. population. In: Plant Genome III Conference, January, 1995, Town & Country Conference Center, San Diego, CA.

Butcher, L.M., Meaburn, E., Liu, L., Fernandes, C., Hill, L., Al-Chalabi, A., Plomin, R., Schalkwyk, L. and Craig I.W. 2004. Genotyping pooled DNA on microarrays: a systematic genome screen of thousands of SNPs in large samples to detect QTLs for complex traits. Behavior Genet. 34: 549–555.

Cardon, L.R. and Bell, J.I. 2001. Association study designs for complex diseases. Nature Rev. Genet. 2: 91–99.

Carlborg, O. and Haley, C.S. 2004. Epistasis: too often neglected in complex trait studies. Nature Rev. Genet. 5: 618–625.

Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S., Tingey, S., Morgante, M. and Rafalski, A.J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 3: 19.

Clark, A.G. 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. 7: 111–122.

Clark, A.G. 2003. Finding genes underlying risk of complex disease by linkage disequilibrium. Curr. Opin. Genet. Develop. 13: 296–302.

Clark, R.M., Linton, E., Messing, J. and Doebley, J.F. 2004. Patterns of diversity in the genomic region near the maize domestication gene *tb1*. Proc. Natl. Acad. Sci. USA 101: 700–707.

Collins, N.C., Lahaye, T. and Schulze-Lefert, P. 2002. Linkage disequilibrium and rice synteny near the barley *ROR1* and *ROR2* genes. In: Plant, Animal & Microbe Genome X Conference, 12–16 January, Town & Country Convention Center, San Diego, CA.

Cregan, P., Randall, N. and Youlin, Z. 2002. Sequence variation, haplotype diversity and linkage disequilibrium in cultivated and wild soybean. In: First International Conference on Legume Genomics and Genetics: Translation to crop improvement, 2–6, June. Minneapolis-St. Paul, MN.

Czika, W. and Weir, B.S. 2004. Properties of the Multiallelic Trend Test. Biometrics 60: 69–74.

Darvasi, A., Weintreb, A., Minke, V., Weller, S. and Soller, M. 1993. Detecting marker QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943–951.

Desplanque, B., Viard, F., Bernard, J., Forcioli, D., Saumitou-Laprade, P., Cuguen, J. and Van Dijk, H. 2000. The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in *Beta vulgaris* ssp. *maritime* L.: The usefulness of both genomes for population genetics studies. Mol. Ecol. 9: 141–154.

Deu, M. and Glaszmann, J.C. 2004. Linkage disequilibrium in sorghum. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, W10.

Devlin, B., Risch, N. and Roeder, K. 1996. Disequilibrium mapping: Composite likelihood for pairwise disequilibrium. Genomics 36: 1–16.

Dvornyk, V., Sirviö, A., Mikkonen, M. and Savolainen, O. 2002. Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris*. Mol. Biol. Evol. 19: 179–188.

Eaves, I.A., Merriman, T.R., Barber, R.A., Nutland, S., Tuomilehto-Wolf. E., Tuomilehto, J., Cucca, F. and Todd, J.A. 2000. The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat. Genet. 25: 320–323.

Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisio, S., Simon, P., Wagenaar, D., Vilkki, J. and Georges, M. 2002. Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. Genetics 161: 275–287.

Fisher, R.A. 1935. The logic of inductive inference. J. R. Stat. Soc. Ser. A. 98: 39–54.

Flint-Garcia, S.A., Thornsberry, J.M. and Buckler, E.S. IV. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54: 357–374.

Garnier-Géré, P., Bedon, F., Pot, D., Austerlitz, F., léger, P., Kremer, A. and Plomion, C. 2003. DNA sequence polymorphism, linkage disequilibrium and haplotype structure in candidate genes of wood quality traits in maritime pine *Pinus pinaster* Ait. In: Plant & Animal Genomes XI Conference, 11–15 January, Town & Country Convention Center, San Diego, CA, P241.

Garris, A.J., McCouch, S.R. and Kresovich, S. 2003. Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice *Oryza sativa* L. Genetics 165: 759–769.

Gaut, B.S. and Long, A.D. 2003. The lowdown on linkage disequilibrium. Plant Cell 15: 1502–1506.

Gebhardt, C., Ballvora, A., Walkemeier, B., Oberhagemann, P. and Schüler, K. 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol. Breed. 13: 93–102.

Geburek, T. 1998. Genetic variation of Norway spruce *Picea abies* [l.] Karst. populations in austria I. Digenic disequilibrium and microspatial patterns derived from allozymes. Forest Genet. 5 http://www.tuzvo.sk/~paule/webdoc2.htm.

Geiringer, H. 1944. On the probability theory of linkage in Mendelian heredity. Ann. Math. Stat. 15: 25–57.

Glaszmann, J.C. 1986. Linkage disequilibrium between Est-2 and Amp-3 in rice varieties. Rice Genet. Newslet. 3 .

Glazier, A.M., Nadeau, J.H. and Aitman, T.J. 2002. Finding genes that underlie complex traits. Science 298: 2345–2349.

Goldstein, D.B., Ahmadi, K.I., Weale, M.E. and Wood, N.W. 2003. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. Trends Genet. 19: 615–622.

González-Martínez, S.C., Brown, G.R., Ersoz, E., Gill, G.P., Kuntz, R.J., Beal, J.A., Wheeler, N.C. and Neale, D.B. 2004. Nucleotide diversity, linkage disequilibrium and adaptive variation in natural populations of loblolly pine. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, W3.

Gorelick, R. and Laubichler, M.D. 2004. Decomposing multilocus linkage disequilibrium. Genetics 166: 1581–1583.

Gupta, P.K. and Rustgi, S. 2004. Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct. Integr. Genomics 4: 137–162.

Hackett, C.A. 2002. Statistical methods of QTL mapping in cereals. Plant Mol. Biol. 48: 585–599.

Hagenblad, J. and Nordborg, M. 2002. Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. Genetics 161: 289–298.

Halldorsson, B.V., Bafna, V., Lippert, R., Schwartz, R., De La Vega, F.M., Clark, A.G. and Istrail, S. 2004. Optimal haplotype block-free selection of tagging SNPs for genome-wide association studies. Genome Res. 14: 1633–1640.

Hamblin, M.T., Mitchell, S.E., White, G.M., Gallego, J., Kukatla R., Wing, R.A., Paterson, A.H. and Kresovich, S. 2004. Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics 167: 471–483.

Hanfstingl, U., Berry, A., Kellogg, E., Costa, J.T. III, Rudiger, W. and Ausubel, F.M. 1994. Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for balancing and directional selection. Genetics 138: 811–828.

Hansen, M., Kraft, T., Ganestam, S., Sall, T. and Nilsson, N.-o. 2001. Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. Genet. Res. 77: 61–66.

Hästbacka, J., de la Chapelle, A., Mahtani, M., Clines, G., Reeve-Daly, M.P., Daly, M., Hamilton, B.A., Kusumi, K., Trivedi, B., Weaver, A., Coloma, A., Lovett, M., Buckler, a., Kaitila, I. and Lander, E.S. 1994. The diastrophic dysplasia gene encodes a novel sulfate transporter: Positional cloning by fine-structure linkage disequilibrium mapping. Cell 78: 1078–1087.

Haubold, B., Kroymann, J., Ratzka, A., Mitchell-Olds, T. and Wiehe, T. 2002. Recombination and gene conversion in a 170-kb genomic region of *Arabidopsis thaliana*. Genetics 161: 1269–1278.

Hedrick, P.W. 1987. Gametic disequilibrium measures: proceed with caution. Genetics 117: 331–341.

Hyten, D.L., Song, Q. and Cregan, P.B. 2004. Linkage disequilibrium in four soybean populations. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, P534.

Innan, H., Tajima, F., Terauchi, R. and Miyashita, N.T. 1996. Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. Genetics 143: 1761–1770.

Innan, H., Terauchi, R., Kahl, G. and Tajima, F. 1999. A method for estimating nucleotide diversity from AFLP data. Genetics 151: 1157–1164.

Innan, H., Terauchi, R. and Miyashita, N.T. 1997. Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics 146: 1441–1452.

Ivandic, V.C., Hackett, A., Nevo, E., Keith, R., Thomas, W.T.B. and Foster, B.P. 2002. Analysis of simple sequence repeat (SSRs) in wild barley from the Fertile Crescent: associations with ecology, geography and flowering time. Plant Mol. Biol. 48: 511–527.

Ivanissevich, D. and Morgante, M. 2001. Sequence diversity and SNP marker development in Norway spruce. In: Proceedings of the XLV Italian Society of Agricultural Genetics – SIGA Annual Congress, 26–29 September, Salsomaggiore Terme, Italy.

Jannink, J.-L. and Walsh, B. 2002. Association mapping in plant populations. In: M.S. Kang (Ed.), Quantitative Genetics, Genomics and Plant Breeding, CAB International, pp. 59–68.

Jannoo, N., Grivet, L., Dookun, A., D'Hont, A. and Glaszmann, J.C. 1999. Linkage disequilibrium among modern sugarcane cultivars. Theor. Appl. Genet. 99: 1053–1060.

Jannoo, N., Grivet, L., Dookun, A., D'Hont, A. and Glaszmann, J.C. 2001. Evaluation of the genetic base of sugarcane cultivars and structuration of diversity at the chromosome level using molecular markers. In: Proceedings 4th Annual Meeting of Agril. Scientists, 21–22 October 1999, Réduit, Maurice Réduit, FARGP, pp. 123–135.

Jiang, C. and Zeng, Z.-B. 1995. Multiple trait analysis for genetic mapping of quantitative trait loci. Genetics 140: 1111–1127.

Johnson, T. 2004. Multipoint linkage disequilibrium mapping using multilocus allele frequency data. http://homepages.ed.ac.uk/tobyj/.

Jorde, J.B. 2000. Linkage disequilibrium and the search for complex disease gene. Genome Res. 10: 1435–1444.

Jung, M., Ching, A., Bhattramakki, D., Dolan, M., Tingey, S., Morgante, M. and Rafalski, A. 2004. Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. Theor. Appl. Genet. 109: 681–689.

Kalinowski, S.T. and Hedrick, P.W. 2001. Estimation of linkage disequilibrium for loci with multiple alleles: basic approach and an application using data from bighorn sheep. Heredity 87: 698–708.

Korol, A.B., Ronin, Y.I., Itskovich, A.M., Peng, J. and Nevo, E. 2001. Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. Genetics 157: 1789–1803.

Kraakman, A.T.W., Niks, R.E., Van den Berg, P.M.M.M., Stam, P. and Van Eeuwijk, F.A. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168: 435–446.

Kraft, T., Hansen, M. and Nilsson, N.-O. 2000. Linkage disequilibrium and fingerprinting in sugar beet. Theor. Appl. Genet. 101: 323–326.

Kreitman, M. and Rienzo, A.D. 2004. Balancing claims for balancing selection. Trends Genet. 20: 300–304.

Kruger, S.A., Able, J.A., Chalmers, K.J. and Langridge, P. 2004. Linkage disequilibrium analysis of hexaploid wheat. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, P321.

Kulwal, P.L., Singh, R., Balyan, H.S. and Gupta, P.K. 2004. Genetic basis of pre-harvest sprouting tolerance using single-locus and two-locus QTL analyses in bread wheat. Funct. Integr. Genomics: 4: 94–101.

Kumar, S., Echt, C., Wilcox, P.L. and Richardson, T.E. 2004. Testing for linkage disequilibrium in the New Zealand radiata pine breeding population. Theor. Appl. Genet. 108: 292–298.

Labate, J.A., Lamkey, K.R., Lee, M. and Woodman, W. 2000. Hardy–Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45: 243–255.

Lazzeroni, L.C. 1998. Linkage disequilibrium and gene mapping: An empirical least-squares approach. Am. J. Hum. Genet. 62: 159–170.

Lehesjoki, A.-E., Koskiniemi, M., Norio, R., Tirrito, S., Sistonen, P., Lander, E. and de la Chapelle, A. 1993. Localization of the *EPM1* gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. Hum. Mol. Genet. 2: 1229–1234.

Lewontin, R.C. 1988. On measures of gametic disequilibrium. Genetics 120: 849–852.

Li, N. and Stephens, M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165: 2213–2233.

Li, Y.C., Fahima, T., Peng, J.H., Röder, M.S., Kirzhner, V.M., Beiles, A., Korol, A. and Nevo, E. 2000. Edaphic microsatellite DNA divergence in wild emmer wheat, *Triticum dicoccoides*, at a microsite: Tabigha, Israel. Theor. Appl. Genet. 101: 1029–1038.

Liang, K.-Y., Hsu, F.-C., Beaty, T.H. and Barnes, K.C. 2001. Multipoint linkage-disequilibrium-mapping approach based on the case-parent trio design. Am. J. Hum. Genet. 68: 937–950.

Lin, J.-Z., Brown, A.H.D. and Clegg, M.T. 2001. Heterogeneous geographic patterns of nucleotide diversity between two alcohol dehydrogenase genes in wild barley *Hordeum vulgare* subspecies *spontaneum*. Proc. Natl. Acad. Sci. USA 98: 531–536.

Lin, J.-Z., Morrell, P.L. and Clegg, M.T. 2002. The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild barley *Hordeum vulgare* ssp. s*pontaneum*. Genetics 162: 2007–2015.

Liu, K., Goodman, M., Muse, S., Smith, J.S., Buckler, Ed. and Doebley, J. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165: 2117–2128.

Lou, X.Y., Casella, G., Littell, R.C., Yang, M.C.K., Johnson, J.A. and Wu, R. 2003. A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis. Genetics 163: 1533–1548.

Lund, M.S., Sorensen, P., Guldbrandtsen, B. and Sorensen, D.A. 2003. Multitrait fine mapping of quantitative trait loci using combined linkage disequilibria and linkage analysis. Genetics 163: 405–410.

Maccaferri, M., Sanguineti, M.C., Noli, E. and Tuberosa, R. 2004. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, P416.

Mackay, T.F.C. 2001. The genetic architecture of quantitative traits. Annu. Rev. Genet. 33: 303–39.

Maniatis, N., Collins, A., Xu, C.-F., McCarthy, L.C., Hewett, D.R., Tapper, W., Ennis, S., Ke, X. and Morton, N.E. 2002. The first linkage disequilibrium LD maps: delineation of hot and cold blocks by diplotype analysis. Proc. Natl. Acad. Sci. USA 99: 2228–2233.

Manosala, P., Trognitz, F., Gysin, R., Torres, S., Nino, D., Silmon, R., Trognitz, B., Ghislain, M. and Nelson, R.J. 2001. Plant defense genes associated with quantitative resistance to potato late blight. In: Plant & Animal Genome IX Conference, January 13–17, 2001, Town & Country Hotel, San Diego, CA, P07_30.html.

Mather, D.E., Hayes, P.M., Chalmers, K., Eglinton, J., Matus, I., Richardson, K., von Zitzewitz, J. and Marquez-Cedillo, L. 2004. Use of SSR marker data to study linkage disequilibrium and population structure in *Hordeum vulgare*: Prospects for association mapping in barley. In: Linkage Disequilibrium Workshop, April 4–7, Novotel Barossa Valley Resort, South Australia.

Matus I., Corey A, Filichkin T., Hayes P.M., Vales M.I., Kling J., Riera-Lizarazu O., Sato K., Powell W. and Waugh R. 2003. Development and characterization of recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in *Hordeum vulgare* subsp. *vulgare* background. Genome 46: 1010–1023.

Mauricio, R., Stahl, E.A., Korves, T., Tian, D., Kreitman, M. and Bergelson, J. 2003. Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. Genetics 163: 735–746.

McCouch, S., Garris, A., Semon, M., Lu, H., Coburn, J., Redus, M., Rutger, N., Edwards, J., Kresovich, S., Nielsen, R., Jones, M. and Tai, T. 2004. Genetic diversity and population structure in rice. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, W6.

Melchinger, A.E. 1996. Advances in the analysis of data on quantitative trait loci. In: V.L Chopra, R.B. Singh and A. Verma (Eds.), Proceedings, 2nd International Crop Science Congress, New Delhi, India, pp. 773–791.

Meuwissen, T.H.E. and Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics 155: 421–430.

Meuwissen, T.H.E. and Goddard, M.E. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. Genet. Sel. Evol. 36: 261–279.

Miyashita, N.T., Kawabe, A. and Innan, H. 1999. DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. Genetics 152: 1723–1731.

Morris, A.P., Whittaker, J.C., Xu, C.-F., Hosking, L.K. and Balding, D.J. 2003. Multipoint linkage-disequilibrium mapping narrows location interval and identifies mutation heterogeneity. Proc. Natl. Acad. Sci. USA 100: 13442–13446.

Mullikin, J.C., Hunt, S.E., Cole, C.G., Mortimore, B.J., Rice, C.M., Burton, J., Matthews, L.H., Pavitt, R., Plumb, R.W., Sims, S.K., Ainscough, R.M., Attwood, J., Bailey, J.M., Barlow, K., Bruskiewich, R.M., Butcher, P.N., Carter, N.P., Chen, Y., Clee, C.M., Coggill, P.C., Davies, J., Davies, R.M., Dawson, E., Francis, M.D., Joy, A.A., Lamble, R.G., Langford, C.F., Macarthy, J., Mall, V., Moreland, A., Overton-Larty, E.K., Ross, M.T., Smith, L.C., Steward, C.A., Sulston, J.E., Tinsley, E.J., Turney, K.J., Willey, D.L., Wilson, G.D., McMurray, A.A., Dunham, I., Rogers, J. and Bentley, D.R. 2000. An SNP map of human chromosome 22. Nature 407: 516–520.

Neale, D.B. and Savolainen, O. 2004. Association genetics of complex traits in conifers. Trends Plant Sci. 9: 325–330.

Nordborg, M., Borevitz, J.O., Bergelson, J., Berry, C.C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J.N., Noyes, T., Oefner, P., Stahl, E.A. and Weigel, D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genet. 30: 190–193.

Nordborg, M. and Tavaré, S. 2002. Linkage disequilibrium: what history has to tell us? Trends. Genet. 18: 83–90.

Nothnagel, M., Fürst, R. and Rohde K. 2002. Entropy as a measure for linkage disequilibrium over multilocus haplotype blocks. Hum. Hered. 54: 186–198.

Nothnagel, M., Rohde. K., Yu, F., Willis T., Pasternak, S., Hardenbol, P., Belmont, J., Leal, S.M. and Gibbs, R.A. 2004. Definition of haplotype blocks based on multilocus LD assessment. In: Ninth International Human Genome Meeting, Estrel Convention Centre, Berlin, Germany, 4–7 April, P239.

Olsen, K.M., Halldorsdottir S.S., Stinchcombe, J.R., Weinig, C., Schmitt, J. and Purugganan, M.D. 2004. Linkage disequilibrium mapping of *Arabidopsis CRY2* flowering time alleles. Genetics 167: 1361–1369.

Owens, C.L. 2003a. SNP detection and genotyping in *Vitis*. In: E. Hajdu and E. Borbas (Eds.) VIII International Conference on Grape Genetics and Breeding. .

Owens, C.L. 2003b. http://www.ars.usda.gov/research/projects/projects.htm?ACCN _NO = 407003&showpars = true&fy = 2003.

Owens, C.L. 2004. Combining linkage analysis and linkage disequilibrium mapping to dissect the genetics of fruit color in grapevine. In: Seventh International Symposium on Grapevine Physiology & Biotechnology, 21–25 June, University of California, Davis, California, USA.

Padhukasahasram, B., Marjoram, P. and Nordborg, M. 2004. Estimating the rate of gene conversion on human chromosome 21. Am. J. Hum. Genet. 75: 386–397.

Palaisa, K., Morgante, M., Tingey, S. and Rafalski, A. 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proc. Natl. Acad. Sci. USA. 101: 9885–9890.

Palaisa, K.A., Morgante, M., Williams, M. and Rafalski, A. 2003. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15: 1795–1806.

Paterson, A.H., Lin, Y.-R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S.-C., Stansel, J.W. and Irvine, J.E. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269: 1714–1718.

Peng, J., Ronin, Y., Fahima, T., Röder, M.S., Li, Y., Nevo, E. and Korol, A. 2003. Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. Proc. Natl. Acad. Sci. USA 100: 2489–2494.

Pozzi, C., Rossini, L., Vacchietti, A. and Salamini, F. 2004. Gene and genome changes during domestication of cereals. In: P.K. Gupta and R.K. Varshney (Eds.), Cereal Genomics, Kluwer Academic, The Netherlands pp. 165–198.

Pritchard, J.K., Stephens, M., Rosenberg, N.A. and Donnelly, P. 2000. Association mapping in structured populations. Am. J. Hum. Genet. 37: 170–181.

Purugganan, M.D. and Suddith, J.L. 1999. Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. Genetics 151: 839–848.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol. 5: 94–100.

Rafalski, A. and Morgante, M. 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet. 20: 103–111.

Ramsay, L., Russell J., Macaulay M., Thomas W.T.B., Powell W. and Waugh R. 2004. Linkage Disequilibrium in European Barley. In: 9th International Barley Genetics Symposium (IBGS), 20–26 June, Brnö, Czech Republic.

Rawat, A. 2004. Relationship between ethnic varieties of *Zea mays* from Asia, Africa and Latin America. Ph. D. Dissertation, Jawaharlal Nehru University, New Delhi, India.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R. and Lander, E.S. 2001. Linkage disequilibrium in the human genome. Nature 411: 199–204.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M. and Buckler, E.S. IV. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA 98: 11479–11484.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. and Feldman, M.W. 2002. Genetic structure of human populations. Science 298: 2381–2385.

Russell, J., Booth, A., Fuller, J., Harrower, B., Hedley, P., Machray, G. and Powell, W. 2004. A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47: 389–398.

Schaeffer, S.W., Goetting-Minesky, M.P., Kovacevic, M., Peoples, J.R., Graybill, J.L., Miller, J.M., Kim, K., Nelson, J.G. and Anderson, W.W. 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. Proc. Natl. Acad. Sci. USA 100: 8319–8324.

Schlotterer, C. 2003. Hitchhiking mapping – functional genomics from the population genetics perspective. Trends Genet. 19: 32–38.

Service, S.K., Lang, D.W., Freimer, N.B. and Sandkuijl, L.A. 1999. Linkage-disequilibrium mapping of disease genes by reconstruction of ancestral haplotypes in founder populations. Am. J. Hum. Genet. 64: 1728–1738.

Shepard, K.A. and Purugganan, M.D. 2003. Molecular population genetics of the *Arabidopsis CLAVATA2* region: the genomic scale of variation and selection in a selfing species. Genetics 163: 1083–1095.

Simko I. 2004. One potato, two potato: haplotype association mapping in autotetraploids. Trends Plant Sci. 9: 441–448.

Simko, I., Costanzo, S., Haynes, K.G., Chris, B.J. and Jones, R.W. 2004a. Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato *Solanum tuberosum* through a candidate gene approach. Theor. Appl. Genet. 108: 217–224.

Simko, I., Haynes, K.G., Ewing, E.E., Costanzo, S, Christ, B.J. and Jones, R.W. 2004b. Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic analysis. Mol. Genet. Genomics 271: 522–531.

Skøt, L., Humpherys, M., Heywood, S., Sanderson, R., Armstead, I., Thomas, I., Chorlton, K. and Hamilton, R.S. 2004. The application of genecology to the discovery of associations between phenotypes and molecular markers in natural populations of perennial ryegrass. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, W137.

Slatkin, M. and Excoffier, L. 1996. Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity 76: 377–383.

Spielman, R.S. and Ewens, W.J. 1996. The TDT and other family based tests for linkage disequilibrium and association. Am. J. Hum. Genet. 59: 983–989.

Spielman, R.S., McGinnis, R.E. and Ewens, W.J. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. 52: 506–513.

Stahl, E.A., Dwyer, G., Mauricio, R., Kreitman, M. and Bargelson, J. 1999. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400: 667–671.

Stracke, S., Perovic, D., Stein, N., Thiel, T. and Graner, A. 2003. Linkage disequilibrium in barley. 11th Molecular Markers Symposium of the GPZ: http://meetings.ipk-gatersleben.de/moma2003/index.php.

Stuber, C.W., Polacco, M. and Senior, M.L. 1999. Synergy of empirical breeding, marker assisted selection and genomics to increase crop yield potential. Crop Sci. 39: 1571–1583.

Taillon-Miller, P., Bauer-Sardina, I., Saccone, N.L., Putzel, J., Laitinen, T., Cao, A., Kere, J., Pilia, G., Rice, J.P. and Kwok, P.Y. 2000. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat. Genet. 25: 324–328.

Taillon-Miller, P., Gu, Z., Li, Q., Hillier, L. and Kwok, P.Y. 1998. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. Genome Res. 8: 748–754.

Tajima, F. 1989. Statistical method for testing the neutral mutational hypothesis by DNA polymorphism. Genetics 123: 585–595.

Tenaillon, M.I., Sawkins, M.C., Anderson, L.K., Stack, SM, Doebley, J. and Gaut, B.S. 2002. Patterns of Diversity and Recombination Along Chromosome 1 of Maize (*Zea mays* ssp. *mays* L.). Genetics 162: 1401–1413.

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F. and Gaut, B.S. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA 98: 9161–9166.

Tenaillon, M.I., U'Ren, J., Tenaillon, O. and Gaut, B.S. 2004. Selection versus demography: a multilocous investigation of the domestication process in maize. Mol. Biol. Evol. 21: 1214–1225.

Terwilliger, J.D. 1995. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am. J. Hum. Genet. 56: 777–787.

Terwilliger, J.D. and Weiss, K.M. 1998. Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr. Opin. Biotech. 9: 578–594.

The International HapMap Consortium. 2003. The international HapMap project. Nature 426: 789–796.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E.S. IV. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genet. 28: 286–289.

Tian, D., Araki, H., Stahl, E., Bergelson, J. and Kreitman, M. 2002. Signature of balancing selection in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 99: 11525–11530.

Tishkoff, S.A. and Verrelli, B.C. 2003. Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. Curr. Opin. Genet. Develop. 13: 569–575.

van der Voort, J.R., Sorensen, A., Lensink, D., Van der Meulen, M., Michelmore, R. and Peleman, J. 2004. Decay of linkage disequilibrium in the *dm3* resistance-gene cluster of lettuce. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, P748. pp.

Verhaegen, D., Plomion, C., Poitel, M., Costa' P. and Kremer, A. 1998. Quantitative trait dissection analysis in *eucalyptus* using RAPD markers: 2. linkage disequilibrium in a factorial design between *E. urophylla* and *E. grandis*. Forest Genet. 5: 61–69.

Vigouroux, Y., McMullen, M., Hittinger, C.T., Houchins, K., Schulz, L., Kresovich, S., Matsuoka, Y. and Doebley, J. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidences of selection during domestication. Proc. Natl. Acad. Sci. USA 99: 9650–9655.

Wang, D.L., Zhu, J., Li, Z.K. and Paterson, A.H. 1999. Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. Theor. Appl. Genet. 99: 1255–1264.

Weir, B.S. 1996. Genetic Data Analysis II. Sinauer, Sunderland, MA, USA.

Weiss, K.M. and Clark, A.G. 2002. Linkage disequilibrium and the mapping of complex human traits. Trends Genet. 18: 19–24.

Wilcox, P.L., Cato, S., Ball, R.D., Kumar, S., Lee, J.R., Kent, J., Richardson, T.E., Echt, C.S. 2002. Genetic architecture of juvenile wood density in *Pinus radiata* and implications for design of linkage disequilibrium studies. In: Plant, Animal & Microbe Genomes X Conference, 12–16 January, Town & Country Convention Center, San Diego, CA.

Wilcox, P.L., Cato, S., McMillan, L.K., Power, M., Ball, R.D., Burdon, R.D. and Echt, C.S. 2004. Patterns of linkage disequilibrium in *Pinus radiata*. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, W89.

Wu, R., Ma, C.-X. and Casella, G. 2002. Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. Genetics 160: 779–792.

Wu, R. and Zeng, Z.-B. 2001. Joint linkage and linkage disequilibrium mapping in natural populations. Genetics 157: 899–909.

Xing, Y.Z., Tan, Y.F., Hua, J.P., Sun, X.L., Xu, C.G. and Zhang, Q. 2002. Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor. Appl. Genet. 105: 248–257.

Xiong, M. and Guo, S.-W. 1997. Fine-scale genetic mapping based on linkage disequilibrium: Theory and applications. Am. J. Hum. Genet. 60: 1513–1531.

Yang, Y., Zhang, J., Hoh, J., Matsuda, F., Xu, P., Lathrop, M. and Ott, J. 2003. Efficiency of single-nucleotide polymorphism haplotype estimation from pooled DNA. Proc. Natl. Acad. Sci. USA 100: 7225–7230.

Yu, S.B., Li, J.X., Xu, C.G., Tan, Y.F., Gao, Y.J., Li, X.H., Zhang, Q. and Saghai Maroof, M.A. 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. USA 94: 9226–9231.

Zapata, C. 2000. The D′ measure of overall gametic disequilibrium between pairs of multiallelic loci. Evolution 54: 1809–1812.

Zhang, W., Collins, A., Maniatis, N., Tapper, W. and Morton, N.E. 2002. Properties of linkage disequilibrium LD maps. Proc. Natl. Acad. Sci. USA 99: 17004–17007.

Zhang, P., Sheng, H., Morabia, A. and Gilliam, T.C. 2003. Optimal Step Length EM Algorithm (OSLEM) for the estimation of haplotype frequency and its application in lipoprotein lipase genotyping. BMC Bioinformatics 4: 1–5.

Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassell, C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D. and Cregan, P.B. 2003. Single nucleotide polymorphisms in soybean. Genetics 163: 1123–1134.