Genet. Sel. Evol. 35 (2003) 513–532

DOI: 10.1051/gse:2003037

Original article

Linkage disequilibrium fine mapping of quantitative trait loci: A simulation study

Jihad M. ABDALLAH[a*], Bruno GOFFINET[b], Christine CIERCO-AYROLLES[b], Miguel PÉREZ-ENCISO[a]

[a] Station d'amélioration génétique des animaux, Institut national de la recherche agronomique, Auzeville BP 27, 31326 Castanet-Tolosan Cedex, France
[b] Unité de biométrie et intelligence artificielle, Institut national de la recherche agronomique, Auzeville BP 27, 31326 Castanet-Tolosan Cedex, France

(Received 27 June 2002; accepted 17 January 2003)

Abstract – Recently, the use of linkage disequilibrium (LD) to locate genes which affect quantitative traits (QTL) has received an increasing interest, but the plausibility of fine mapping using linkage disequilibrium techniques for QTL has not been well studied. The main objectives of this work were to (1) measure the extent and pattern of LD between a putative QTL and nearby markers in finite populations and (2) investigate the usefulness of LD in fine mapping QTL in simulated populations using a dense map of multiallelic or biallelic marker loci. The test of association between a marker and QTL and the power of the test were calculated based on single-marker regression analysis. The results show the presence of substantial linkage disequilibrium with closely linked marker loci after 100 to 200 generations of random mating. Although the power to test the association with a frequent QTL of large effect was satisfactory, the power was low for the QTL with a small effect and/or low frequency. More powerful, multi-locus methods may be required to map low frequent QTL with small genetic effects, as well as combining both linkage and linkage disequilibrium information. The results also showed that multiallelic markers are more useful than biallelic markers to detect linkage disequilibrium and association at an equal distance.

linkage disequilibrium / quantitative trait locus / fine mapping

1. INTRODUCTION

Linkage disequilibrium (LD), or nonrandom allelic association between loci, has been used to locate simply-inherited Mendelian disease genes in human

* Correspondence and reprints
E-mail: abdallah@germinal.toulouse.inra.fr

populations [8,16,17]. More recently, there has also been an increasing interest in using LD for fine mapping of complex disease genes [15,28,33,36] and quantitative trait loci (QTL) [4,24,27,31]. For details on the use of LD in mapping disease genes the reader is referred to the reviews by Pritchard and Przeworski [26] or by Jorde [15].

Linkage disequilibrium can be potentially useful but has been less studied for quantitative traits. It is problematic for quantitative traits because they are influenced by environmental factors. As for most putative genes, QTL genotypes are not known. Therefore, information on the QTL has to be inferred using phenotypic data and marker genotypes. In addition, genetic heterogeneity, *i.e.*, multiple mutations at the functional locus, has not been widely considered in usual LD mapping methods.

Linkage analysis, which is based on following the cosegregation of marker and phenotypic data through a pedigree, is often used to localise genes within several centimorgans. The main advantage of LD mapping over linkage analysis is that it makes use, in principle, of all historical recombinations in populations of unrelated individuals, giving more precise estimates of gene location. For the purpose of gene mapping, an ideal measure of LD is one that is a monotone decreasing function of recombination distance and that is robust to departures due to random drift. It is well known, however, that the pattern of LD may be extremely variable due to the history of recombination and to the history of mutations and it is the variability due only to recombination history that is useful for mapping purposes [25]. Spurious LD can also occur due to population admixture and, more importantly, the region of highest association with the trait may not necessarily be the one that contains the causal mutation.

The main objectives of this study were to measure the extent and pattern of LD and assess its usefulness for fine-scale mapping of quantitative trait loci in simulated populations. We used single-marker regression analysis to detect the association between marker loci and the QTL.

2. MATERIALS AND METHODS

2.1. Simulations

Simulations were carried out based on two extreme scenarios. The first (LD scenario) assumes that at some point in the history of the population a mutation in a quantitative trait locus occurred in one haplotype of a single individual. This results in a complete initial linkage disequilibrium between the QTL locus and other loci in the region. The second scenario assumes initial linkage equilibrium in the base population (LE scenario) between the QTL and markers as well as between markers. Note that the LE scenario is equivalent to having many origins of the same mutation and then differential amplification due to genetic drift. The first model is the simplest and most frequently used in human genetics epidemiological studies. Real situations most likely will fall in between these two extreme models.

Initially, we considered an 18 cM chromosomal region with 40 markers and a biallelic QTL in a base population (G0) of 500 individuals. We then confined our analyses to a 3 cM region with the QTL and 30 equally spaced markers because the regression *P*-value was very rarely significant beyond 3 cM. We also considered populations of 100 and 200 individuals. The case of 100 individuals resulted in high rates of allele fixation, and therefore it was not possible to get enough replicates in a reasonable computing time.

Two types of markers were considered: biallelic (SNP) and multiallelic (MST) markers. In G0, each MST marker had five alleles at equal frequencies. An initial allele frequency of 0.5 was used for SNP markers. Two hundred generations of intermating were simulated. Haplotypes for offspring were simulated by choosing parents at random and allowing individuals to inherit recombinant or non-recombinant haplotypes based on Mendel's laws and recombination probabilities. The MST marker alleles were allowed to mutate at a rate of 10^{-4} per generation using a stepwise mutation model, *i.e.*, an allele increased or decreased its count by one. Mutation was assumed negligible for SNP.

In the LD scenario, a single allele was simulated for the QTL in G0 and 100 generations later (G100) a QTL mutation with a positive effect on the trait was introduced in one haplotype of a single individual. A slight selective advantage was conferred to the mutated haplotype for a few generations such that the expected QTL frequency was 0.02 in the first 10 generations after the mutation. This was done to ensure that not many simulations are lost due to the rapid loss of QTL alleles as a result of genetic drift. In later generations (G111–G200), haplotypes of progeny were inherited at random from the population. In the LE scenario, QTL and marker loci in G0 were simulated, assuming a linkage equilibrium with a QTL frequency of 0.20 for the allele with a positive effect on the trait.

The simulation procedure used here is based on the gene dropping method [22]. Other simulation methods based on the coalescent theory can be used [5, 14, 37, 40]. However, these methods are complex especially for multiple markers. Importantly, simulations, as herein, allow us to assess the variability of LD across different conceptually repeated populations.

The simulations were discarded when in any generation fixation occurred for QTL alleles or any of the markers. Simulations were classified based on QTL frequency in the last generation (G200). The lowest number of replicates for any class was 700 with a total of 10 000 simulations required in G200 for the LD scenario and 6000 for the LE scenario.

The phenotype, y, of the quantitative trait for an individual was simulated as $y = g + e$, where g is the additive genetic value of the QTL genotype of the individual and e is an environmental value drawn from a normal distribution with a mean of 0 and variance of 1.0. Following Falconer and Mackay [3], for a QTL locus with two alleles, Q and q, and an additive QTL effect equal to a (in standard deviations), the genetic values of the genotypes QQ, Qq, and qq are a, d, and $-a$, respectively. The additive genetic variance explained by the trait locus is $2p(1-p)a^2$ where p is the frequency of the QTL. We evaluated three values for a, namely 1.0, 0.5, and 0.25 sd with $d = 0$ (assuming no dominance effects on the trait) in all cases.

2.2. Linkage disequilibrium measures

Measures for the estimation of linkage disequilibrium were the standardised disequilibrium coefficient D' [9], and the squared correlation of allele frequencies, r^2 [11,12,34]. These two measures of LD are widely used in the literature *e.g.* [2,25]. According to Hill and Weir [12], r^2 is the most often used measure of LD. Furthermore, D' and r^2 are easily calculated for multiallelic loci.

For two multiallelic loci A and B, D' and r^2 are obtained as:

$$D' = \sum_i \sum_j p_i q_j |D'_{ij}|,$$

where p_i and q_j are the population allele frequencies of the ith allele on locus A and the jth allele on locus B. $D'_{ij} = \dfrac{D_{ij}}{D_{\max}}$ is the Lewontin normalised LD measure [19], where $D_{ij} = x_{ij} - p_i q_j$, and x_{ij} is the frequency of the haplotype with alleles i and j on loci A and B, respectively. $D_{\max} = \min[p_i q_j, (1-p_i)(1-q_j)]$ when $D_{ij} < 0$, and $\min[p_i(1-q_j), (1-p_i)q_j]$ when $D_{ij} > 0$. The squared correlation of allele frequencies is calculated as:

$$r^2 = \sum_i \sum_j \frac{D_{ij}^2}{p_i q_j}.$$

2.3. Regression analysis

The measure of LD for quantitative traits is a measure of association. Here we used regression analysis to test for association because it is simple and has well characterised statistical properties. The phenotypic trait value y_i of individual i was regressed on the number of copies x_{ij} of allele j of marker M according to the regression model:

$$y_i = \mu + \sum_j b_j x_{ij} + e_i,$$

where μ = the population mean of the quantitative trait, b_j = the regression coefficient on allele j of marker M, and e_i = the residual error for $i = 1$ to the number of individuals and $j = 1$ to the number of alleles. The F-statistic to test the significant association of marker M with the QTL was obtained by testing the above model against the model $y_i = \mu + e_i$, *i.e.*, we tested the overall association of marker alleles on the trait. The corresponding P-values (the probability of an F-value as large or larger than the observed F-statistic given the null hypothesis of no association, [35]) were obtained using the appropriate degrees of freedom.

The power to map the QTL within a given interval (0.5, 1.0, and 1.5 cM) was calculated as the proportion of replicates where at least one single-marker analysis showed a significant (P-value < 0.05) association with the trait locus.

3. RESULTS AND DISCUSSION

3.1. LD Pattern

Average values of D' and r ($\sqrt{r^2}$) between the QTL and microsatellite markers are plotted as a function of genetic distance and class of QTL frequency (Fig. 1). The decay of linkage disequilibrium by recombination distance is evident. This decay is slower for classes with a low QTL frequency. The results were similar for biallelic markers (data not shown). However, mean values of r were smaller for biallelic markers than for MST due to differences in allele frequencies.

Both D' and r depend on QTL frequency but the behaviour of D' as a function of QTL frequency was opposite to that of r. This occurs because D' and r weigh allele frequencies inversely. It has been indicated by other researchers [2, 6,9] that, unlike D', other measures of LD, including r^2, depend on allele frequency. Lewontin [20], however, argued that even D' is not independent of gene frequency and that there are generally no gene frequency independent measures of association between the loci. Nordborg and Tavaré [25] showed that measures of LD including D' depend on the frequencies of the markers and the disease gene. They further argued that this frequency-dependence is best viewed as age-dependence; the more frequent an allele, the older it is.

Table I shows the percentage of replicates where maximum LD between QTL and markers was within 0.5, 1.0 and 1.5 cM. Frequency of maximum LD increased as QTL frequency increased for both D' and r^2, indicating that the accuracy of LD mapping is very sensitive to QTL frequency, the more extreme the QTL frequencies the less accuracy is to be expected. For MST, the maximum disequilibrium was higher when measured by D' compared to r^2 but the opposite was generally true for SNP. This may be irrelevant for QTL mapping because information on QTL alleles is usually absent, but it

Table I. The frequency (%) that the maximum disequilibrium was with the marker within a specified distance in the linkage disequilibrium (LD) scenario. The base population (G0) was simulated assuming linkage equilibrium. A QTL mutation was introduced in G100. The results are from generation 100 after the introduction of a QTL mutation (G200).

Distance (cM)	Marker type		QTL frequency (P)				
			$0 < P < 0.05$	$0.05 < P < 0.1$	$0.1 < P < 0.15$	$0.15 < P < 0.2$	$P > 0.2$
0.5	MST[1]	D'	49.6	78.0	87.5	90.6	92.3
		r^2	44.4	60.5	72.5	79.6	85.3
	SNP[2]	D'	30.7	54.1	61.1	66.9	72.2
		r^2	37.0	52.6	60.6	64.7	80.1
1.0	MST	D'	69.3	90.2	95.3	97.7	97.7
		r^2	63.4	80.6	88.7	92.9	97.7
	SNP	D'	51.6	73.1	78.9	85.0	86.7
		r^2	59.3	72.9	81.3	86.0	95.2
1.5	MST	D'	80.4	94.9	97.4	98.6	98.9
		r^2	75.8	89.2	94.3	97.4	99.2
	SNP	D'	66.1	83.8	88.7	92.9	92.2
		r^2	73.6	86.2	90.8	92.7	97.7

[1] multiallelic markers; [2] biallelic markers.

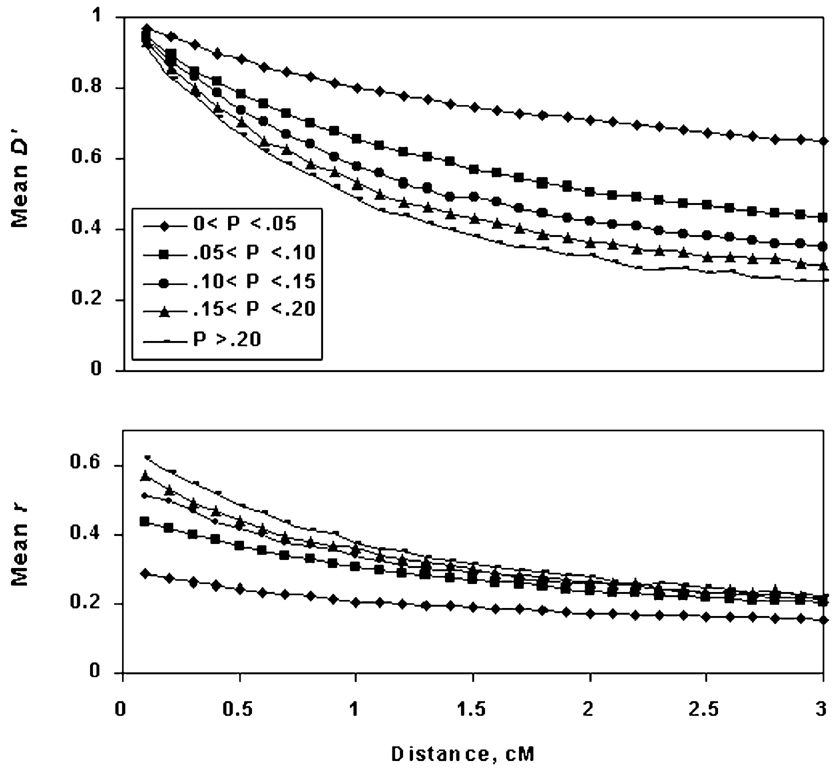

Figure 1. Mean values of D' and $r = \sqrt{r^2}$ between the QTL locus and 30 multiallelic marker loci as functions of the distance and QTL frequency in G200 in the LD (linkage disequilibrium) scenario. QTL mutation introduced in G100. Population size = 500 individuals.

is useful for mapping genes affecting simply-inherited Mendelian traits. The frequency of maximum LD was consistently higher for MST compared to SNP. This was in agreement with the results by others [10,30] who found that the statistical power to test disequilibrium increased as the number of marker alleles increased.

Results for the LE scenario are in Figure 2 and Table II. Because QTL frequency in G200 was generally higher in the LE scenario, some of the QTL classes were different from the LD scenario. Trends in the LD measures for the LE scenario were similar to the LD scenario: power increased with less extreme QTL allele frequencies. For the same frequency classes, the levels of D' in G200 were lower in the LE scenario compared to the LD scenario for both MST and SNP. In the LE scenario, unlike the LD scenario, the frequency of maximum disequilibrium measured by r^2 was higher than that measured by D'. This suggests that the optimum linkage disequilibrium

Table II. The frequency (%) that the maximum disequilibrium was with the marker within a specified distance in the linkage equilibrium (LE) simulation scenario. The base population (G0) was simulated assuming linkage equilibrium with a QTL frequency of 0.2. The results are from generation 200 (G200).

Distance (cM)	Marker type		QTL frequency (P)				
			$0 < P < 0.1$	$0.1 < P < 0.15$	$0.15 < P < 0.2$	$0.2 < P < 0.3$	$P > 0.3$
0.5	MST[1]	D'	63.1	80.7	80.2	83.1	87.0
		r^2	68.2	83.7	88.4	87.2	92.1
	SNP[2]	D'	40.8	56.2	58.0	62.5	64.1
		r^2	51.0	58.7	66.1	71.5	73.9
1.0	MST	D'	79.5	90.8	93.6	94.2	95.7
		r^2	84.4	95.0	97.7	97.6	98.6
	SNP	D'	58.0	74.8	73.9	79.8	79.3
		r^2	70.8	77.8	90.1	89.0	90.3
1.5	MST	D'	87.3	95.2	95.8	96.8	97.7
		r^2	91.8	97.8	98.8	98.8	99.8
	SNP	D'	70.9	83.6	83.2	86.9	86.9
		r^2	81.2	86.1	97.0	96.3	96.1

[1] multiallelic markers; [2] biallelic markers.

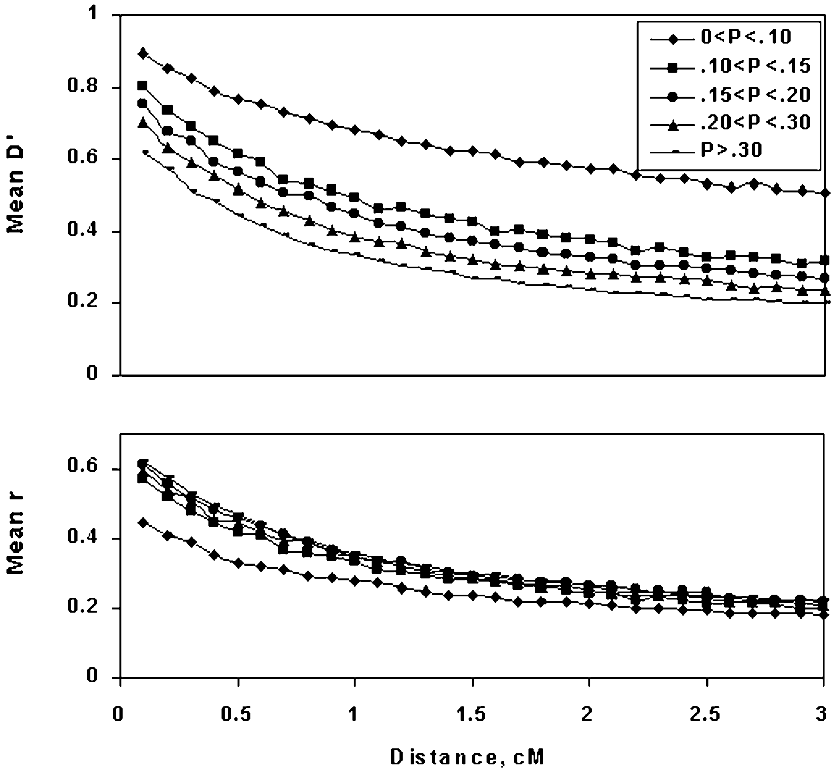

Figure 2. Mean values of D' and $r = \sqrt{r^2}$ between the QTL locus and 30 multiallelic marker loci as functions of the distance and QTL frequency in G200 in the LE (linkage equilibrium) scenario. The base population ($n = 500$) was simulated assuming linkage equilibrium.

measure for mapping single disease genes with complete penetrance may depend on the genetic heterogeneity of the trait: D' may be better than r^2 in the usual LD model, whereas r^2 may be preferred if there are several original mutations.

Measures of linkage disequilibrium usually have high variability [9, 12, 13, 39]. The variances of D' and r between the QTL and MST markers are plotted in Figure 3 as functions of distance and QTL frequency. The variance of D' was low for markers close to the QTL then increased up to 1 cM and slowly decreased afterwards except when QTL frequency was less than 0.05 where the variance continued to increase. The explanation for the behaviour of this class is not evident to us. The variance of r had a more stable behaviour, and it monotonically decreased as the distance from the QTL increased with no differences among classes of QTL frequency, *i.e.*, the variance of r was less influenced by the QTL frequency. Note that the variance of D' was larger than

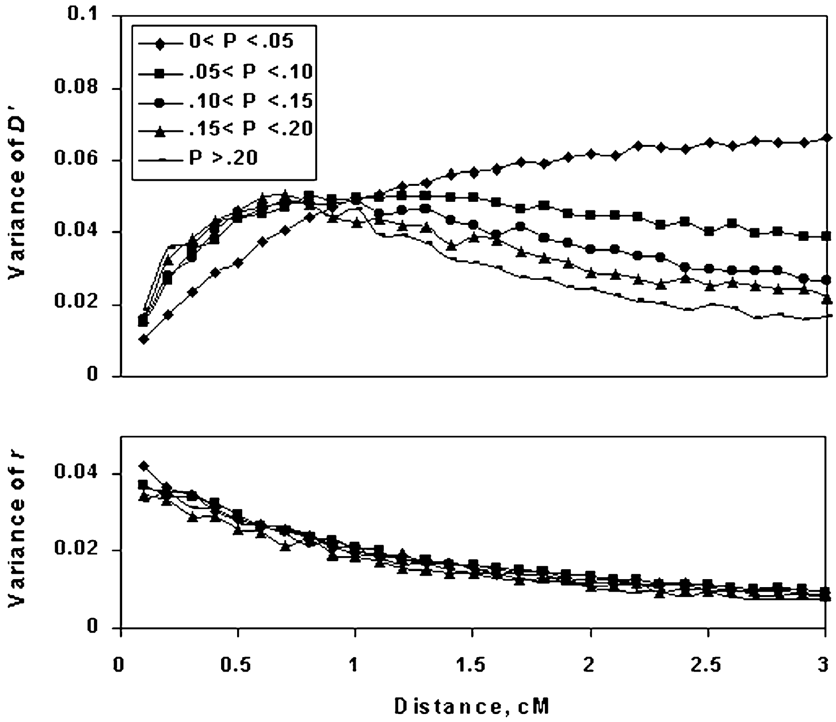

Figure 3. Variances of linkage disequilibrium measures as functions of marker distance and QTL frequency in G200 in the LD (linkage disequilibrium) scenario using multiallelic markers.

the variance of r except for markers very close to the QTL. These observations indicate that r is more stable and more consistent than D'. It is difficult to explain the variation in LD measures as these are potentially affected by several factors including sample size, allele number and allele frequency [39]. In any case, the overall decrease in variance as we move away from the QTL makes sense because it is expected that LD decreases with distance and there will be less uncertainty about this decrease when the marker is located farther away.

3.2. QTL Mapping

Figure 4 presents mean significance levels (P-values) for the F-statistic to test the association between the multiallelic marker loci and QTL for the LD scenario. On average, P-values decreased as the distance from the QTL decreased indicating a higher significant association. Mean P-values also decreased as the QTL frequency and QTL effect increased, showing that the power and accuracy of LD mapping will be higher when the QTL allele

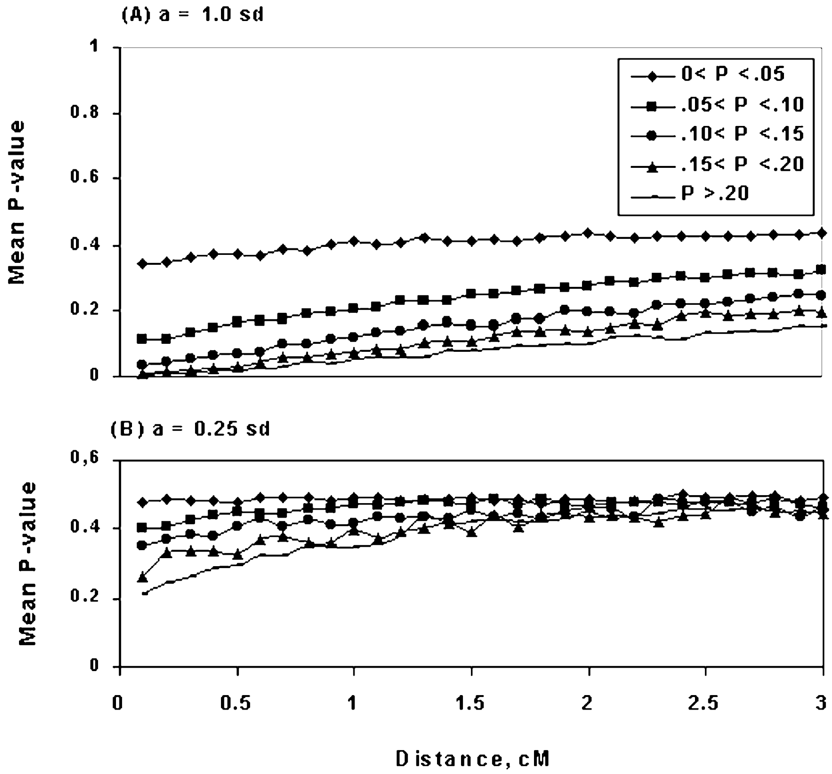

Figure 4. Mean *P*-values for the test of the association between QTL and multiallelic markers as a function of distance and QTL frequency in G200 in the LD (linkage disequilibrium) scenario for QTL effects of 1.0 sd (A), and 0.25 sd (B). Population size = 500.

frequencies are moderate than when they are extreme, which is in agreement with the results from the previous section. Multiallelic markers showed lower mean *P*-values (*i.e.* more significant) compared to biallelic markers (data not shown).

Table III shows the power to map the QTL within 0.5, 1.0, and 1.5 cM from its true position by the class of QTL frequency in G200 for the LD simulation scenario. The power increased as a function of the QTL effect and frequency. For a QTL with an effect of 1 sd and using multiallelic markers, in 95.1% of the replicates the most significant association was with a marker within 1 cM from the true position of the QTL when the QTL frequency exceeded 0.20 (heritability > 0.25). The percentage was equal to 44.1 when QTL frequency was less than 0.05 (heritability < 0.10). For a QTL of 0.25 sd, the percentage decreased to 55.7% for QTL frequency > 0.20 (heritability > 0.075) and 21.6% for QTL frequency < 0.05 (heritability < 0.025).

Table III. The power (%) to map a QTL within a specified distance using multiallelic (MST) and biallelic (SNP) markers in the linkage disequilibrium (LD) scenario. The base population (G0) was simulated assuming linkage equilibrium. The QTL mutation was introduced in G100. The results are from generation 100 after introduction of the QTL mutation (G200).

Distance (cM)	Marker type	QTL effect	QTL frequency (P)				
			$0 < P < 0.05$	$0.05 < P < 0.1$	$0.1 < P < 0.15$	$0.15 < P < 0.2$	$P > 0.2$
0.5	MST	1.0	29.1	54.3	68.1	76.4	81.8
		0.5	17.2	37.2	55.4	63.1	76.9
		0.25	11.4	18.0	27.0	31.1	38.2
	SNP	1.0	20.4	40.9	54.7	63.4	73.3
		0.5	16.0	30.7	41.2	49.8	63.7
		0.25	10.3	13.8	21.6	30.8	36.8
1.0	MST	1.0	44.1	73.6	84.9	90.7	95.1
		0.5	28.8	54.0	70.4	79.3	91.1
		0.25	21.6	29.7	37.7	45.5	55.7
	SNP	1.0	36.1	60.1	75.5	81.4	93.0
		0.5	28.6	47.6	61.4	72.2	80.4
		0.25	23.1	25.8	39.9	49.3	49.2
1.5	MST	1.0	54.7	84.4	90.9	95.0	98.0
		0.5	38.5	64.7	80.2	85.9	94.4
		0.25	31.0	37.8	49.0	56.1	66.1
	SNP	1.0	49.0	75.3	85.9	90.8	95.9
		0.5	40.1	59.5	72.5	81.8	88.2
		0.25	33.8	38.3	54.8	64.5	64.2

For the LE scenario, the trends in mean *P*-values (not shown) and power (Tab. IV) as a function of the QTL effect and frequency were similar to those of the LD scenario. For the same QTL frequency, the power was higher in the LE scenario than in the LD scenario when using MST but was similar for SNP. This is interesting because it is usually assumed that LD mapping will be more useful when there is a single founder haplotype. This occurred because in the LE scenario there was probably more than one allele in association with the trait locus. Note that we did not test the association of a particular allele to the trait but, rather, the global association between all marker alleles and the trait. Thus, in the LD scenario we expect that a single allele is associated with the trait, making the other allele effects "blur" the global marker association. In contrast, in the LE scenario it may well happen that more than one allele become correlated with the trait. This is true for multiallelic markers. For biallelic markers, the test reduces to that of a single allele which explains why there was no difference between the two scenarios in the case of SNP.

We have seen that power increases with less extreme QTL frequencies; this is logical because, other things being equal, the proportion of the phenotypic variance explained by the QTL (*i.e.* heritability) increases at intermediate frequencies. However, the increase in power may not be only due to the increase in heritability. The proportion of maximum disequilibrium also increased with an increased QTL frequency (Tabs. I and II) which may have resulted in an increased power. Abecasis *et al.* [1] found that the power to detect LD increased as the trait allele frequency increased but the power was maximum when the trait locus and marker allele frequencies were similar. Luo *et al.* [21] compared regression, ANOVA, and maximum-likelihood analyses to detect LD between a marker locus and QTL in samples from a random mating population. They found that, given the genetic variance explained by a trait locus, the power of regression and ANOVA tests is relatively independent from the allele frequency. They also suggested that the regression model is the preferred test as far as the power is concerned.

Mapping power using MST markers was higher than when using SNP. Power is expected to increase as the number of marker alleles increases because larger degrees of freedom are expected under the multiple marker allele model [21]. The great advantage, of course, of single nucleotide polymorphisms is that they are much more abundant than highly polymorphic markers like microsatellites. This result was based on a single marker analysis. The problem of the low information content of SNP markers will be reduced when haplotypes are analysed rather than each marker individually.

The empirical variances of significance levels (*P*-values) for association between MST markers and the QTL are given in Figure 5A. The variance increased as the distance from the QTL increased, except for very small frequencies where it was consistently high. Variation in *P*-values among five

Table IV. The power (%) to map a QTL within a specified distance using multiallelic (MST) and biallelic (SNP) markers in the linkage equilibrium (LE) simulation scenario. The base population (G0) was simulated assuming the linkage equilibrium with a QTL frequency of 0.2. The results are from generation 200 (G200).

Distance (cM)	Marker type	QTL effect	QTL frequency (P)				
			$0 < P < 0.1$	$0.1 < P < 0.15$	$0.15 < P < 0.2$	$0.2 < P < 0.3$	$P > 0.3$
0.5	MST	1.0	56.1	81.5	84.4	86.2	89.7
		0.5	34.2	64.4	68.6	76.3	83.3
		0.25	16.0	30.9	32.6	41.5	50.1
	SNP	1.0	36.6	54.4	61.2	67.6	70.5
		0.5	23.8	41.2	51.9	59.4	65.0
		0.25	16.6	20.8	28.9	32.9	40.0
1.0	MST	1.0	72.5	92.4	95.1	97.4	97.4
		0.5	49.0	79.3	83.0	89.9	93.0
		0.25	28.8	43.3	46.1	58.3	64.2
	SNP	1.0	54.1	76.3	83.2	85.0	88.4
		0.5	37.9	59.3	72.3	78.1	81.5
		0.25	28.6	36.7	45.8	47.3	58.7
1.5	MST	1.0	81.8	96.2	97.4	98.8	99.0
		0.5	59.5	85.3	89.4	94.5	96.7
		0.25	37.1	51.4	56.8	67.6	73.0
	SNP	1.0	65.1	85.5	90.6	92.2	93.8
		0.5	48.9	70.9	81.1	86.2	88.0
		0.25	39.0	49.4	58.4	59.4	69.4

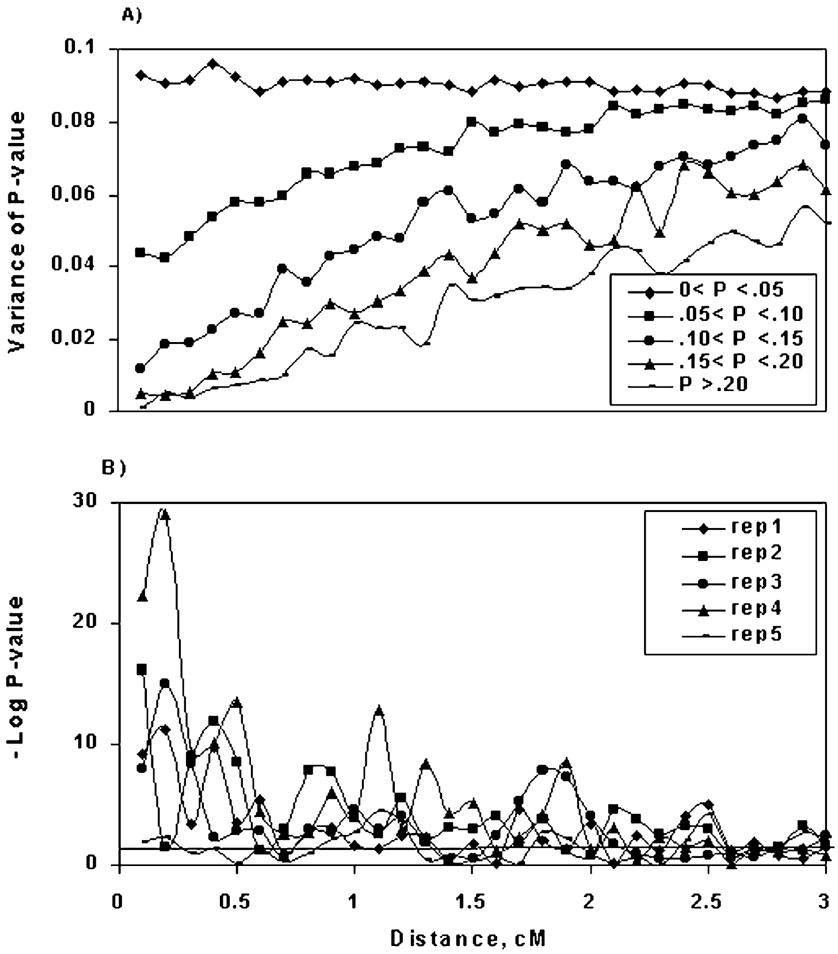

Figure 5. (A) Variances of P-values as functions of marker distance and QTL frequency in G200 in the LD (linkage disequilibrium) scenario using multiallelic markers. (B) Variation in P-values (on $-\log$ scale) among five replicates with a QTL frequency > 0.2. QTL effect = 1 sd. The horizontal line in B corresponds to a P-value $= 0.05$. Note that because of the $-\log$ scale, the values above the line are significant at a 5% level.

independent random replicates is also shown in Figure 5B as a function of genetic distance for QTL frequency > 0.20. It is interesting to realise that not only the expected power of LD mapping increases with intermediate QTL frequencies, it will also be less risky (less variable). In turn, the variability increases with distance, making the pattern of P-values hard to interpret. The risk of finding a spurious significant value increases as we move away from the QTL.

So far we presented results when the length of the chromosomal region was 3 cM. For the sake of completeness, we also present a simulated 18 cM region with a QTL and 40 multiallelic marker loci (30 markers as before and an additional 10 markers with 1.5 cM spacing). Simulations were carried out under the LD scenario with a QTL effect of 1 sd. Levels of LD further decreased beyond 3 cM (results not shown). More importantly, the percentage of replicates at which maximum disequilibrium and significant *P*-values occurred for markers within 0.5, 1.0, and 1.5 cM, were very similar to the case where only the 3 cM region was used. In fact, a maximum disequilibrium occurred very rarely beyond 3 cM (less than 5% for a QTL frequency < 0.05 and less than 0.5% for a QTL frequency > 0.05). In more than 96% of the replicates the most significant association was with a marker within 3 cM from the QTL. This reinforces the popular idea that LD may not be useful in genome scans for QTL but can be useful in fine-scale mapping using a dense marker map. In practice, linkage analysis can be used in genome-wide scans as a first step to narrow the position of the QTL to a few centimorgans and then linkage disequilibrium can be used to fine map the QTL [24]. In a simulation study, Kruglyak [18] concluded that a useful level of linkage disequilibrium is unlikely to extend beyond an average distance of 3 kb in human populations. In a more recent study, Hall *et al.* [7] found significant evidence for LD in the Afrikaners extending over a 6-cM range but LD decayed significantly beyond 3 cM distances in the other populations they examined. In domestic animal populations, Farnir *et al.* [4] and McRae [23] reported that high levels of LD extend over tens of centimorgans. However, in both studies, LD was frequently observed between unlinked markers. Thus, it is not clear from their results whether LD can be used for fine mapping in these populations but, certainly, there is a great need of assessing the extent of LD in livestock selected populations.

3.3. Effect of population size

In the previous simulations we considered a population size of 500 individuals. We also simulated a population of 200 individuals as in the LD scenario with a QTL effect of 1 sd using MST markers. Mean values of LD were higher for a population of 200 than a size of 500 individuals. This was as expected because genetic drift is higher in the smaller population. However, the percentage of replicates at which maximum disequilibrium between QTL and markers occurred within 0.5, 1.0, and 1.5 cM from the QTL, was lower for the population of 200. This was likely due to the higher variation in LD in the smaller population. The difference was larger with D' than with r^2. For example, for QTL frequency > 0.20, the percentage of replicates that maximum LD occurred with markers within 1 cM was 81.1 for D' and 87.8

for r^2 for N_e of 200 individuals compared to 97.7 for both D' and r^2 for the N_e of 500 individuals. The power to detect association with markers within a specified distance decreased by 5 to 20% for N_e of 200 compared to N_e of 500 individuals. This is not surprising because the power of regression tests depends on the sample size.

3.4. Effect of selection and admixture

So far, we have considered a random mating natural population where LD is influenced by recombination, mutation, and drift. Other historical events that have important effects on LD include selection and admixture. Selection is particularly important in domestic animal and plant populations. Selection for an improved phenotype of the quantitative trait is expected to increase the frequency of the QTL but, at the same time, decreases the effective population size. Thus, in the light of the previous results, selection is expected to increase the power of association mapping particularly for low frequent trait alleles, provided that the effective population size is not too small.

Admixture or migration of individuals between populations of differing allele frequencies creates a linkage disequilibrium that may extend over large distances [38]. Admixture is often considered a liability in LD mapping [15]. This is because population stratification caused by admixture can lead to spurious associations between markers and unlinked loci [26,29]. To account for population stratification, researchers often use family-based tests such as the transmission-disequilibrium test (TDT) by Spielman *et al.*, [32], at the expense of decreasing power.

To assess the effect of admixture we simulated a population of 500 individuals as in the LD scenario using 40 MST markers in a chromosome region of 18 cM. We then allowed 1% of the individuals to migrate each generation into this population from another population in the LE scenario with a QTL frequency of 0.20 in the base population. Thus both populations differed in the QTL frequency and in their history. The results from these simulations provided no evidence for spurious associations beyond 3 cM. For low frequency classes, the frequency of maximum disequilibrium was even higher for markers close to the putative QTL than in the case with no admixture. For example the frequency of maximum D' within 1 cM was 72% with admixture and 65.4% without admixture for a QTL frequency < 0.05. The results were very similar in both cases for QTL frequency > 0.10. The same trends were found for the power of the association test. Note that this migration model is different from the usual model in human genetics; here we considered a continuous migration between populations such as that occurring when, *e.g.*, a breeder regularly imports foreign animals. Human population studies rather tend to consider populations where admixture occurred only once.

4. CONCLUSIONS

The simulation results presented in this study showed that on average, LD decreased as a function of recombination distance. However, LD measures had high variability and were influenced by gene frequency. Due to this variability it is not unlikely to find high LD with the more distanced markers from the QTL but rarely beyond 3 cM from the gene locus of interest for N_e of 500.

The power to detect a significant association between QTL and nearby markers depends on the heritability of the QTL as well as on the amount of LD. The power of LD mapping using regression analysis is satisfactory when the QTL frequency is intermediate and its effect is large. More powerful, multilocus approaches may be required to map QTL with small heritability. Our results suggest that, in single marker analysis, multiallelic markers may be more useful than biallelic markers in LD mapping of quantitative trait loci.

One important observation from the results presented herein is that although maximum disequilibrium occurred with the closest marker more frequently than with any other marker, the presence of high LD between the QTL and nearby markers does not necessarily mean significant P-values. In real life, analysis is usually based on one replicate of data and it is not surprising to find significant association with the more distant markers. This is because of the high variability in LD which for much of it, may not reflect recombination at all [25].

ACKNOWLEDGEMENTS

Funding for this work was provided by project No. 20 (2001–2002) of the “Bureau des ressources génétiques”, and a project within “Action en bioinformatique” of the “ministère de la Recherche” (France).

REFERENCES

[1] Abecasis G.R., Cookson W.O.C., Cardon L.R., The power to detect linkage disequilibrium with quantitative traits in selected samples, Am. J. Hum. Genet. 68 (2001) 1463–1474.

[2] Devlin B., Risch N., A comparison of linkage disequilibrium measures for fine-scale mapping, Genomics 29 (1995) 311–322.

[3] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, 4th edn., Longman, Essex, 1996.

[4] Farnir F., Coppieters W., Arranz J.J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mni M., Nezer C., Simon P., Vanmanshoven P., Wagenaar D., Georges M., Extensive genome-wide linkage disequilibrium in cattle, Genome Res. 10 (2000) 220–227.

[5] Griffiths A.M., Tavaré S., Simulating probability distributions in the coalescent, Theor. Popul. Biol. 46 (1994) 131–159.

[6] Guo S.-W., Linkage disequilibrium measures for fine-scale mapping: a comparison, Hum. Hered. 47 (1997) 301–314.

[7] Hall D., Wijsman E.M., Roos J.L., Gogos J.A., Karayiorgou M., Extended intermarker linkage disequilibrium in the Afrikaners, Genome Res. 12 (2002) 956–951.

[8] Hästbacka J., de la Chappelle A., Kaitila I., Sistonen P., Weaver A., Lander E., Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland, Nat. Genet. 2 (1992) 204–211.

[9] Hedrick P.W., Gametic disequilibrium measures: proceed with caution, Genetics 117 (1987) 331–341.

[10] Hedrick P.W., Thompson G., A two-locus neutrality test: applications to humans, *E. coli*, and lodgepole pine, Genetics 112 (1986) 135–156.

[11] Hill W.G., Robertson A., Linkage disequilibrium in finite populations, Theor. Appl. Genet. 38 (1968) 226–231.

[12] Hill W.G., Weir B.S., Maximum-likelihood estimation of gene location by linkage disequilibrium, Am. J. Hum. Genet. 54 (1994) 705–714.

[13] Hudson R.R., The sampling distribution of linkage disequilibrium under an infinite alleles model without selection, Genetics 109 (1985) 611–631.

[14] Hudson R.R., The how and why of generating gene genealogies, in: Takahata N., Clark A.G. (Eds.), Mechanics of Molecular Evolution, Sinauer, Sunderland, 1993, pp. 23–36.

[15] Jorde L.B., Linkage disequilibrium and the search for complex disease genes, Genome Res. 10 (2000) 1435–1444.

[16] Kaplan N., Hill W.G., Weir B.S., Likelihood methods for locating disease genes in nonequilibrium populations, Am. J. Hum. Genet. 56 (1995) 18–32.

[17] Kerem B.S., Romens J.M., Buchanan J.A., Markiewics D., Cox T.K., Lehesjoki A., Koskiniemi J., Norio R., Tirrito S., Sistonen P., Lander E.S., de la Chapelle A., Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping, Hum. Mol. Genet. 2 (1993) 1229–1234.

[18] Kruglyak L., Prospects for whole-genome linkage disequilibrium mapping of common disease genes, Nat. Genet. 22 (1999) 139–144.

[19] Lewontin R.C., The interaction of selection and linkage. I. General considerations: heterotic models, Genetics 49 (1964) 49–67.

[20] Lewontin R.C., On measures of gametic disequilibrium, Genetics 120 (1988) 849–852.

[21] Luo Z.W., Tao S.H., Zeng Z.B., Inferring linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations, Genetics 156 (2000) 457–467.

[22] MacCluer J.W., VandeBerg J.L., Read B., Ryder O.A., Pedigree analysis by computer simulation, Zoo. Biol. 5 (1986) 147–160.

[23] McRae A.F., McEwan J.C., Dodds K.G., Wilson T., Crawford A.M., Slate J., Linkage disequilibrium in domestic sheep, Genetics 160 (2002) 1113–1122.

[24] Meuwissen T.H.E., Goddard M.E., Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci, Genetics 155 (2000) 421–430.

[25] Nordborg M., Tavaré S., Linkage disequilibrium: what history has to tell us, Trends Genet. 18 (2002) 83–90.
[26] Pritchard J.K., Przeworski M., Linkage disequilibrium in humans: models and data, Am. J. Hum. Genet. 69 (2001) 1–14.
[27] Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S., Structure of linkage disequilibrium and phenotypic associations in the maize genome, Proc. Natl. Acad. Sci. 98 (2001) 11479–11484.
[28] Risch N., Merikangas K., The future of genetic studies of complex human diseases, Science 273 (1996) 1516–1517.
[29] Schulze T.G., McMahon F.J., Genetic association mapping at the crossroads: Which test and why? Overview and practical guidelines, Am. J. Med. Genet. 114 (2002) 1–11.
[30] Slatkin M., Linkage disequilibrium in growing and stable populations, Genetics 137 (1994) 331–336.
[31] Slatkin M., Disequilibrium mapping of a quantitative-trait locus in an expanding population, Am. J. Hum. Genet. 64 (1999) 1765–1773.
[32] Spielman R.S., McGinnis R.E., Ewens W.J., Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM), Am. J. Hum. Genet. 52 (1993) 506–516.
[33] Terwilliger J.D., Weiss K.M., Linkage disequilibrium mapping of complex disease: fantasy or reality?, Curr. Opin. Biotechnol. 9 (1998) 578–594.
[34] Weir B.S., Genetic data analysis II, 2nd edn., Sinauer, Sunderland, 1996.
[35] Weisberg S., Applied linear regression, 2nd edn., John Wiley, New York, 1985.
[36] Weiss K.M., Clark A.G., Linkage disequilibrium and the mapping of complex human traits, Trends Genet. 18 (2002) 19–24.
[37] Wilson I.J., Balding D.J., Genealogical inference from microsatellite data, Genetics 150 (1998) 499–510.
[38] Wilson J.F., Goldstein D.B., Consistent long range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population, Am. J. Hum. Genet. 67 (2000) 926–935.
[39] Zapata C., Carollo C., Rodriguez S., Sampling variance and distribution of the D'measure of overall gametic disequilibrium between multiallelic loci, Ann. Hum. Genet. 65 (2001) 395–406.
[40] Zöllner S., von Haeseler A., A coalescent approach to study linkage disequilibrium between single-nucleotide polymorphisms, Am. J. Hum. Genet. 66 (2000) 615–628.