Genet. Res., Camb. (2007), **89**, *pp.* 1–6. © 2007 Cambridge University Press
doi:10.1017/S0016672307008634 Printed in the United Kingdom

Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms

H. ZHAO[1], D. NETTLETON[2] AND J. C. M. DEKKERS[1]*
[1] *Department of Animal Science and Center for Integrated Animal Genomics, 239 Kildee Hall, Iowa State University, Ames, IA 50011, USA*
[2] *Department of Statistics, 111A Snedecor Hall, Iowa State University, Ames, IA 50011, USA*

(*Received 6 October 2006 and in revised form 29 December 2006 and 7 February 2007*)

Summary

Effectiveness of marker-assisted selection (MAS) and quantitative trait locus (QTL) mapping using population-wide linkage disequilibrium (LD) between markers and QTLs depends on the extent of LD and how it declines with distance between markers and QTLs in a population. Marker–QTL LD can be predicted from LD between markers. Our previous work evaluated LD measures between multi-allelic markers as predictors of usable LD of multi-allelic markers with QTLs. Since single nucleotide polymorphisms (SNPs) are the current marker of choice for high-density genotyping and LD-mapping of QTLs, the objective of this study was to use LD between multi-allelic markers to predict LD among biallelic SNPs or between SNPs and QTLs. Observable LD between multi-allelic markers was evaluated using nine measures. These included two pooled and standardized measures of LD between pairs of alleles at two markers based on Lewontin's LD measure, two pooled measures of squared correlations between alleles, one standardized measure using Hardy–Weinberg heterozygosities, and four measures based on the chi-square statistic for testing for association between alleles at two loci. The standardized chi-square measure that best predicted usable LD between multi-allelic markers and QTLs, based on our previous work, overestimated usable SNP–SNP or SNP–QTL LD. Instead, three other measures were found to be good predictors of usable SNP–SNP or SNP–QTL LD when LD is generated by drift. Therefore, the LD measure between multi-allelic markers that is best for predicting usable LD in a population depends on the type of markers (i.e. multi-allelic or biallelic) that will eventually be used for QTL mapping or MAS.

1. Introduction

Effectiveness of marker-assisted selection (MAS) and quantitative trait locus (QTL) mapping using population-wide linkage disequilibrium (LD) between markers and QTLs depends on the extent of LD and how it declines with distance in a population. Although marker–QTL LD cannot be observed directly, it can be predicted from LD between markers. Zhao *et al.* (2005) evaluated nine LD measures between multi-allelic markers as predictors of usable LD between the same group of markers and biallelic QTLs. When LD is generated by drift, a standardized chi-square statistic ($\chi^{2\prime}$) was recommended to quantify the amount and extent of usable LD in a population for QTL mapping and MAS based on multi-allelic markers (Zhao *et al.*, 2005).

While highly polymorphic microsatellite (MS) markers are still often used in genome-wide linkage analysis to track inheritance of chromosome regions, recently biallelic single nucleotide polymorphism (SNP) markers have been receiving more attention in genetics research. In addition to the abundance of SNPs in the genome, recent advances in technology have made large-scale SNP genotyping rapid, accurate and inexpensive (Kwok, 2001). High-density SNP maps are now available for both human and several livestock species. For example, the public SNP database contains 9·2 million candidate human SNPs (International HapMap Consortium, 2005),

* Corresponding author. Telephone: +1 (515) 2947509. Fax: +1 (515) 2949150. e-mail: jdekkers@iastate.edu

and a genetic variation map for the chicken genome containing 2·8 million SNPs has been constructed (International Chicken Polymorphism Map Consortium, 2004).

These exciting developments of dense SNP maps present tremendous opportunities for high-resolution LD mapping of QTLs. Within a closed breeding population in livestock, LD is limited to closely linked loci due to many generations of recombination. Therefore, high-density SNP genotyping enables detection and fine-mapping of QTLs in outbred populations using historical LD, and resulting QTLs can immediately be implemented for MAS (Dekkers & Hospital, 2002; Grapes *et al*., 2004; Meuwissen & Goddard, 2000). A crucial issue in using high-density SNP maps is the extent of LD among SNPs or between SNPs and QTLs, which affects the power of LD mapping and effectiveness of MAS and is needed to determine the SNP density that is required to obtain a given power to detect QTLs. However, since MS markers are still frequently used or MS genotypes may be available from previous studies in a population, it is of interest to predict the extent of LD that exists in a population among SNPs or between SNPs and QTLs based on LD between available MS markers, which is the objective of this study. This research has practical implications because, before collecting data on SNPs, it is important to know how many SNPs and what sort of density will be needed. The genotype data that are available on MSs in many populations can help us address this prior to designing SNP panels and collecting SNP data.

2. Materials and methods

The methods in this paper are the same as in Zhao *et al*. (2005). Briefly, observable LD between multi-allelic marker pairs was evaluated using nine alternate measures. These included two pooled and standardized measures of LD between pairs of alleles at two markers based on Lewontin's LD measure (denoted D' and D_{hap}), two pooled measures of squared correlations between alleles (r^2 and r^2_{hap}), one standardized measure using Hardy–Weinberg heterozygosities (D^*), and four measures based on the chi-square statistic for testing for association between alleles at two loci (χ^2, χ^2_{df}, $\chi^{2\prime}$ and χ^2_{tr}). Definitions of these measures are given in Appendix. For LD between biallelic markers, $D' = D_{hap}$ and

$$r^2 = r^2_{hap} = D^* = \chi^2_{df} = \chi^{2\prime}.$$

These nine measures of LD between multi-allelic markers were evaluated for their ability to quantify LD among biallelic SNPs on a 100 cM chromosome in simulated populations. In generation 0, MSs with 2, 4, 6, 8 or 10 equi-frequent alleles were simulated at 0, 2, …, 100 cM and SNPs with two equi-frequent alleles at 1, 3, …, 99 cM. All loci were in Hardy–Weinberg and linkage equilibrium in generation 0. LD was generated by drift by 100 generations of random mating of N parents ($N = 50$, 100, 150 or 200), which was found to be sufficient to reach a steady-state situation with regard to the level of LD (Zhao *et al*., 2005). Data on segregating loci in generation 100 were used for analysis.

Estimates of SNP–SNP LD were obtained from LD between a pair of SNPs in our simulation and measured by r^2 and D'. LD measured by r^2 is equivalent to usable LD between biallelic markers and QTLs (also assumed biallelic) (Zhao *et al*., 2005). Because many studies have used D' to evaluate multi-allelic marker LD (Farnir *et al*., 2000; McRae *et al*., 2002; Nsengimana *et al*., 2004; Tenesa *et al*., 2003), we evaluated the ability of multi-allelic D' to predict biallelic D'.

To assess and compare the decline in LD with distance ($\leqslant 20$ cM) for SNP–SNP LD and MS–MS LD, the function $LD_d = 1/(1 + 4\beta d)$ (Sved, 1971) was fitted to the LD data that were generated for each replicate, where LD_d is LD at distance d morgans, as measured by SNP–SNP r^2 or D' or by a MS–MS LD measure, and β is a parameter that is related to effective population size (N_e = actual population size for the idealized populations that were simulated (Falconer & Mackay, 1996)). A weighted least squares regression was used to estimate β for each simulated data set, as described in Zhao *et al*. (2005).

Following the same criteria as described in Zhao *et al*. (2005), LD curves predicted from different measures of MS–MS LD were compared with SNP–SNP LD measured by: (1) r^2 to find which multi-allelic marker measure best predicts usable SNP–SNP LD, and (2) D' to find which multi-allelic marker measure best predicts SNP–SNP LD based on D'.

3. Results

(i) *Decline of LD with distance*

The observed relationships of SNP–SNP LD and MS–MS LD with distance for a representative replicate with a population size of 100 and 4 alleles per MS marker are illustrated in Fig. 1. Usable SNP–SNP LD measured by r^2 was relatively high at short distances and declined rapidly with distance (Fig. 1*a*). Similar declines were observed when r^2, r^2_{hap}, D^*, χ^2, χ^2_{df} (Fig. 1*b*), $\chi^{2\prime}$ and χ^2_{tr} were used to measure MS–MS LD. The SNP–SNP LD measured by D' was strongly inflated relative to SNP–SNP r^2, and high LD values were obtained even for markers that approached equilibrium (results not shown). The same

https://doi.org/10.1017/S0016672307008634 Published online by Cambridge University Press

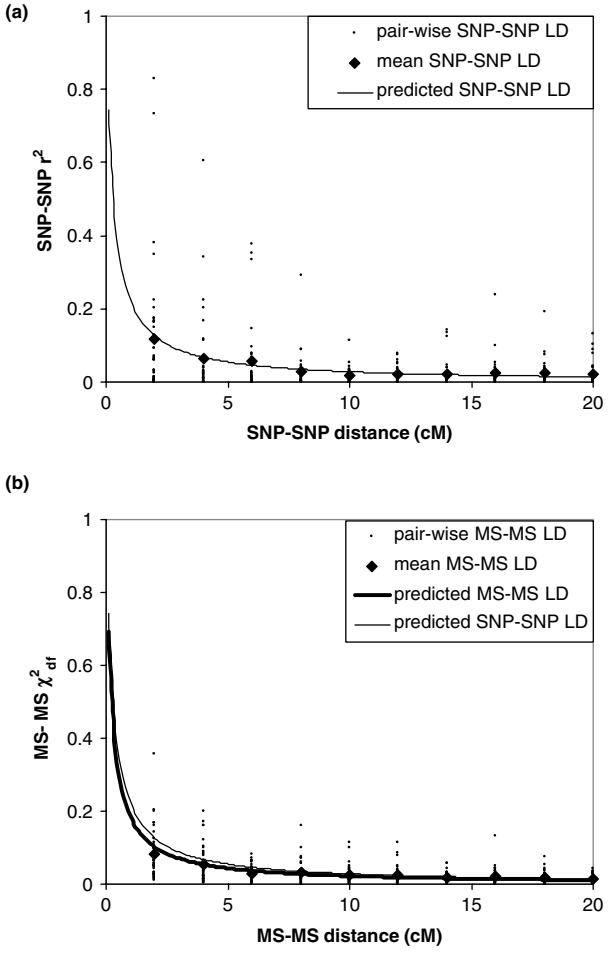

Fig. 1. Observed relationships of SNP–SNP LD measured by r^2 (*a*) and microsatellite–microsatellite (MS–MS) LD measured by χ^2_{df} (*b*) against map distance for a representative replicate with a population size of 100 and four alleles per MS marker in generation 0. LD at a distance d morgans was predicted from $\hat{LD}_d = 1/(1+\hat{\beta}d)$ where $\hat{\beta}$ was obtained from the simulated data for each LD measure.

was true for MS–MS LD measured by D' and D_{hap} (results not shown).

To assess the decline of LD with distance, the equation $LD_d = 1/(1+4\beta d)$ was fitted to the sample data for the replicate pictured in Fig. 1. Estimates were $\hat{\beta} = 86{\cdot}8$ for SNP–SNP r^2 and 3·3 for SNP–SNP D', and 5·4, 5·4, 92·0, 89·8, 93·5, 110·4, 42·6 and 24·1 for MS–MS LD measured by D', D_{hap}, r^2, r^2_{hap}, D^*, χ^2_{df}, $\chi^{2\prime}$ and χ^2_{tr}, respectively. Measure χ^2 was not used to estimate β because of its non-standardized scale. The LD curve predicted from SNP–SNP r^2 was very close to LD curves predicted from MS–MS LD measured by r^2, r^2_{hap}, D^* and χ^2_{df} (Fig. 1*b*). The LD curve predicted from SNP–SNP D' was close to LD curves predicted from MS–MS LD measured by D' and D_{hap} (results not shown). Based on mean LD at a given distance (Fig. 1), the estimated curves appeared to provide a good fit to the data for all LD measures except for D' and D_{hap} due to their inflated values at larger distances.

(ii) *Comparison of LD curves predicted from SNP–SNP and MS–MS LD*

Results in this section are based on analysing 100 replicates for each of the 20 combinations of population size and number of marker alleles. All LD measures were evaluated except χ^2.

The mean $\hat{\beta}$ across 100 replicates obtained from SNP–SNP and MS–MS LD for each simulated scenario is shown in Table 1. The mean $\hat{\beta}$ for MS–MS LD measured by r^2, D^* and χ^2_{df} was very close to the mean $\hat{\beta}$ for SNP–SNP r^2, and they all provided good estimates of N_e (Table 1). With more than two alleles per MS marker in generation 0, the mean estimates of $\hat{\beta}$ obtained from MS–MS $\chi^{2\prime}$ were much lower than the mean $\hat{\beta}$ for SNP–SNP r^2 (Table 1). Because of the relationship between LD at a given distance and β based on the equation $LD_d = 1/(1+4\beta d)$, this implies that measure $\chi^{2\prime}$ overestimated usable SNP–SNP LD. The mean $\hat{\beta}$ for MS–MS LD measured by D' and D_{hap} was very close to the mean $\hat{\beta}$ for SNP–SNP D' (Table 1).

The relationship between MS–MS LD and SNP–SNP LD for a given population was further analysed using estimates $\hat{\beta}$ obtained from each replicate. Table 2*a* illustrates the relationship of $\hat{\beta}$ for SNP–SNP LD measured by r^2 with $\hat{\beta}$ for MS–MS LD across the 20 simulated cases with varying population sizes and numbers of MS marker alleles. A good linear relationship was observed for MS–MS LD measured by r^2, D^* and χ^2_{df}, with a correlation of 0·93, 0·93, 0·94 and slope of 1·0, 1·0, 0·8, respectively (Table 2*a*). Corresponding relationships were poorer for $\chi^{2\prime}$ and for the other MS–MS LD measures (Table 2*a*). The mean of the squared difference (MSE) averaged over 100 replicates between LD predicted based on SNP–SNP r^2 and MS–MS LD measured by r^2, D^* and χ^2_{df} was low for all 20 simulated cases (Table 3). Therefore, usable SNP–SNP LD is best predicted by MS–MS LD measured by r^2, D^* and χ^2_{df}, but not by $\chi^{2\prime}$.

Corresponding relationships of $\hat{\beta}$ for SNP–SNP LD measured by D' with $\hat{\beta}$ for MS–MS LD are in Table 2*b*. The relationship appeared to be linear for MS–MS LD measured by D' and D_{hap}. Correlations were 0·79 and 0·83 and slopes were 0·90 and 0·83 for D' and D_{hap}, respectively (Table 2*b*), while slopes ranged from 0·02 to 0·16 for the other MS–MS LD measures (Table 2*b*). The MSE between LD predicted based on SNP–SNP D' and MS–MS LD measured by D' and D_{hap} was much lower than the other MS–MS LD measures for all 20 simulated cases (results not shown). This implies that D' and D_{hap} between multi-allelic markers, although not

https://doi.org/10.1017/S0016672307008634 Published online by Cambridge University Press

Table 1. *Mean estimates of the decline of LD with distance ($\hat{\beta}$) over 100 replicates based on SNP–SNP LD and microsatellite–microsatellite (MS–MS) LD for simulated data based on different combinations of number of MS marker alleles in generation 0 and population size*

No. of MS alleles in generation		Population size	SNP–SNP LD		MS–MS LD				
0	100		D'	r^2	D'	r^2	D^*	χ^2_{df}	$\chi^{2\prime}$
2	2	50	2·6	55·0	2·5	55·7	55·7	55·7	55·7
	2	100	4·1	101·9	4·2	101·5	101·5	101·5	101·5
	2	150	6·9	146·4	6·6	148·6	148·6	148·6	148·6
	2	200	10·1	190·0	9·9	193·0	193·0	193·0	193·0
4	2·2	50	2·6	53·7	2·7	49·1	49·3	51·1	35·5
	2·8	100	4·2	102·2	5·5	92·6	93·8	104·6	46·2
	3·3	150	6·3	146·5	8·7	130·3	133·0	156·1	56·0
	3·6	200	9·9	190·6	11·3	176·4	178·9	207·0	69·0
8	2·4	50	2·4	56·5	2·7	46·9	47·2	49·8	29·8
	3·5	100	4·2	103·5	5·7	87·3	89·6	107·5	33·1
	4·4	150	6·3	146·6	7·4	128·3	131·7	162·6	37·3
	5·2	200	10·0	190·6	8·3	170·8	175·5	219·0	41·5

The number of MS alleles in generation 100 is the average of the mean number of alleles across MS markers still segregating in generation 100 over 100 replicates. Results for MS–MS LD measured by D_{hap}, r^2_{hap} and χ^2_{tr} (not shown) can be found in Zhao *et al*. (2005).

Table 2. *Correlation and slope of the regression of the decline of LD with distance ($\hat{\beta}$) estimated from each replicate for (a) SNP–SNP r^2 and (b) SNP–SNP D' on $\hat{\beta}$ estimated for different measures of microsatellite–microsatellite (MS–MS) LD*

	MS–MS LD							
	D'	D_{hap}	r^2	r^2_{hap}	D^*	χ^2_{df}	$\chi^{2\prime}$	χ^2_{tr}
(*a*) SNP–SNP r^2								
Correlation	0·86	0·89	0·93	0·88	0·93	0·94	0·35	0·76
Slope	16·82	15·25	1·01	1·08	1·00	0·82	0·43	2·99
(*b*) SNP–SNP D'								
Correlation	0·79	0·83	0·87	0·82	0·87	0·87	0·34	0·72
Slope	0·90	0·83	0·05	0·06	0·05	0·04	0·02	0·16

Data are based on 100 replicates simulated for each of the 20 combinations of population size (50, 100, 150 or 200) and number of MS marker alleles (2, 4, 6, 8 or 10) in generation 0.

Table 3. *The mean of the squared difference (MSE) between LD predicted based on SNP–SNP r^2 and different measures of microsatellite–microsatellite (MS–MS) LD at 1, 2, …, 20 cM for simulated data generated from different combinations of population size and number of MS marker alleles in generation 0*

No. of MS alleles in generation 0	Population size	MS–MS LD			
		D'	r^2	$\chi^{2\prime}$	χ^2_{tr}
2	50	244·2	0·6	0·6	31·4
	100	168·2	0·1	0·1	15·8
	150	114·4	0·0	0·0	7·6
	200	74·7	0·0	0·0	3·5
4	50	216·5	0·4	1·8	29·3
	100	127·0	0·1	3·0	13·2
	150	80·5	0·0	2·9	8·1
	200	61·8	0·0	2·3	5·5
8	50	216·0	0·5	3·8	31·7
	100	121·7	0·1	7·4	16·8
	150	96·6	0·0	7·9	14·0
	200	89·9	0·0	7·7	12·5

Values are the average MSE over 100 replicates multiplied by 1000 for each combination. Results for MS–MS LD measured by D_{hap} (not shown) were similar to those for D', and results for r^2_{hap}, D^* and χ^2_{df} were similar to those for r^2.

recommended for measuring LD, can predict SNP–SNP LD based on D'.

4. Discussion and conclusions

Before collecting data on SNPs for high-density genotyping and LD-mapping of QTLs, it is important to know how many SNPs and what sort of SNP density are needed to obtain a given power to detect QTLs, which depends on the extent of LD that exists in a population among SNPs. Our study shows that LD between available MS markers measured by r^2, D^* and χ^2_{df} are good predictors of usable SNP–SNP LD when LD is generated by drift. Although the focus

https://doi.org/10.1017/S0016672307008634 Published online by Cambridge University Press

was on predicting SNP–SNP LD using LD between MSs, our conclusions also apply to relating LD between other multi-allelic markers to LD among SNPs or between SNPs and QTLs.

Assuming a drift model, the decline of LD with distance estimated from SNP–SNP r^2 provides good estimates of N_e. Hence, in order to predict usable LD between SNPs, the observed LD between MSs should be measured by a statistic that approximately estimates N_e, regardless of the number of alleles left at a MS marker after drift. Such a measure was found by our simulation studies to be r^2, D^* and χ^2_{df}, which were invariant to the number of MS marker alleles remaining in generation 100; all provided good estimates of N_e (Table 1).

Zhao *et al*. (2005) showed that $\chi^{2\prime}$, a measure of LD between multi-allelic markers, is the best predictor of usable LD of multi-allelic markers with QTLs for the purpose of QTL detection and MAS. However, as demonstrated here, $\chi^{2\prime}$ overestimates usable LD among SNPs or between SNPs and QTLs, because it reflects not only N_e but also the number of marker alleles that remain in the generation under consideration (Zhao *et al*., 2005). Therefore, the LD measure between multi-allelic markers that is best for predicting usable LD in a population depends on the type of markers (i.e. multi-allelic or biallelic) that will eventually be used for QTL mapping or MAS.

The definitions of χ^2_{df} and $\chi^{2\prime}$ suggest that they are proportional to each other, with a ratio of $[\max(k, m)-1]$, where k and m are the numbers of alternate alleles at two MS markers (see Appendix). For a large N_e and $k=m$, this implies a constant ratio of $(k-1)$, and the estimate of the decline of LD with distance (β) should reflect this ratio. Because the number of alleles remaining in generation 100 varied across MS markers, this is approximately what is observed in Table 1. For $N_e=200$, the ratio of $\hat{\beta}$ for χ^2_{df} versus $\chi^{2\prime}$ was 3 for $k=m=4$ in generation 0, and the ratio was 5 for $k=m=8$, because on average a number of alleles are lost due to drift (Table 1).

Because many previous studies have used D' to evaluate multi-allelic marker LD (Farnir *et al*., 2000; McRae *et al*., 2002; Nsengimana *et al*., 2004; Tenesa *et al*., 2003), we also evaluated its ability to predict D' for biallelic loci. We found that SNP–SNP LD based on D' can be predicted from LD between multi-allelic markers measured by D' and D_{hap}. However, they are not recommended to quantify LD due to their inflated LD estimates (Zhao *et al*., 2005).

In our simulation, the smallest distance between SNPs is 2 cM, but we see no reason that we can not extrapolate our results to SNPs at shorter distances. Although our study is based on simulated populations where LD was generated by drift alone, the conclusions are expected to hold for populations that are under selection or subject to mutation, as reasoned by Zhao *et al*. (2005).

Appendix. Definitions of nine LD measures between multi-allelic markers

The first two measures are based on Lewontin's normalized LD measure (Lewontin, 1964) weighted by the product of allele frequencies:

$$D' = \sum_{i=1}^{k}\sum_{j=1}^{m} p(A_i)\,p(B_j)\left|\frac{D_{ij}}{D_{ij}^{\max}}\right|$$

(Hedrick, 1987), or weighted by haplotype frequencies:

$$D_{hap} = \sum_{i=1}^{k}\sum_{j=1}^{m} p(A_iB_j)\left|\frac{D_{ij}}{D_{ij}^{\max}}\right|$$

(Karlin & Piazza, 1981), where k and m are the numbers of alternate alleles at locus A and B, respectively, $p(A_i)$ is the frequency of allele A_i at locus A, $p(B_j)$ the frequency of allele B_j at locus B, $p(A_i\,B_j)$ the frequency of haplotype A_iB_j, and

$$D_{ij} = p(A_iB_j) - p(A_i)p(B_j),$$

$$D_{ij}^{\max} = \min[\,p(A_i)\,p(B_j),\ (1-p(A_i))\,(1-\,p(B_j))\,] \quad \text{when } D_{ij}<0,$$

and

$$D_{ij}^{\max} = \min\,[\,p(A_i)\,(1-p(B_j)),\ (1-p(A_i))\,p(B_j)\,] \quad \text{when } D_{ij}\geqslant 0.$$

The next two measures are based on pooling the square of the correlation between A_i and B_j, denoted by r^2_{ij}, based on allele frequencies:

$$r^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} p(A_i)\,p(B_j)\,r^2_{ij},$$

or based on haplotype frequencies:

$$r^2_{hap} = \sum_{i=1}^{k}\sum_{j=1}^{m} p(A_iB_j)\,r^2_{ij},$$

where $r^2_{ij} = \dfrac{D^2_{ij}}{p(A_i)\,(1-p(A_i))\,p(B_j)\,(1-p(B_j))}$ (Hill & Robertson, 1968).

Using Hardy–Weinberg heterozygosities at two loci, the fifth measure is

$$D^* = \frac{D^2}{H_A H_B}$$

https://doi.org/10.1017/S0016672307008634 Published online by Cambridge University Press

(Maruyama, 1982; Hedrick & Thomson, 1986; Hedrick, 1987), where

$$D^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} D_{ij}^2, H_A = 1 - \sum_{i=1}^{k} p^2(A_i)$$

and

$$H_B = 1 - \sum_{j=1}^{m} p^2(B_j).$$

The final four measures are related to the chi-square statistic to test for independence between alleles at two loci. The chi-square statistic is

$$\chi^2 = 2N \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{D_{ij}^2}{p(A_i)\,p(B_j)},$$

where N is the sample size and $2N$ is the number of haplotypes that occurs in the sample. Three standardized measures of χ^2 with values between 0 and 1 are:

$$\chi_{df}^2 = \frac{\chi^2}{2N(k-1)(m-1)}$$

(Hedrick & Thomson, 1986; Hedrick, 1987), where $(k-1)(m-1)$ is equal to the degrees of freedom of χ^2;

$$\chi^{2\prime} = \frac{\chi^2}{2N\,(l-1)}$$

(Yamazaki, 1977), where $l = \min(k, m)$; and

$$\chi_{tr}^2 = \frac{\chi^2}{\chi_{\max}^2}$$

(Zhao *et al.*, 2005), where $\chi_{\max}^2$ is a sharp upper bound for the maximum of χ^2. Note that χ_{df}^2 and $\chi^{2\prime}$ are proportional to each other, with a ratio of $[\max(k, m) - 1]$.

We are very grateful to Rohan Fernando for providing us with programs used in our simulations. We also thank Morris Soller, Eli Heifetz and Janet Fulton for their help and valuable discussion. The editor and reviewers are acknowledged for suggestions that resulted in substantial improvements of the manuscript. This research was supported by State of Iowa Hatch and Multi-state Research Funds.

References

Dekkers, J. C. M. & Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews: Genetics* **3**, 22–32.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th edn. Harlow, UK: Addison-Wesley Longman.

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. (2000). Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**, 220–227.

Grapes, L., Dekkers, J. C. M., Rothschild, M. F. & Fernando, R. L. (2004). Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. *Genetics* **166**, 1561–1570.

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

Hedrick, P. W. & Thomson, G. (1986). A two-locus neutrality test: application to humans, *E. coli* and Lodgepole pine. *Genetics* **112**, 135–156.

Hill, W. G. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226–231.

International Chicken Polymorphism Map Consortium (2004). A genetic variation map for chicken with 2·8 million single-nucleotide polymorphisms. *Nature* **432**, 717–722.

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* **437**, 1299–1320.

Karlin, S. & Piazza, A. (1981). Statistical methods for assessing linkage disequilibrium at the HLA-A, B, C loci. *Annals of Human Genetics* **45**, 79–94.

Kwok, P. Y. (2001). Methods for genotyping single nucleotide polymorphisms. *Annual Review of Genomics and Human Genetics* **2**, 235–258.

Lewontin, R. C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67.

Maruyama, T. (1982). Stochastic integrals and their application to population genetics. In *Molecular Evolution, Protein Polymorphism and the Neutral Theory*, pp. 151–166. Tokyo: Japan Scientific Societies Press.

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

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

Nsengimana, J., Baret, P., Haley, C. S. & Visscher, P. M. (2004). Linkage disequilibrium in the domesticated pig. *Genetics* **166**, 1395–1404.

Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology* **2**, 125–141.

Tenesa, A., Knott, S. A., Ward, D., Smith, D., Williams, J. L. & Visscher, P. M. (2003). Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *Journal of Animal Science* **81**, 617–623.

Yamazaki, T. (1977). The effects of overdominance on linkage in a multilocus system. *Genetics* **86**, 227–236.

Zhao, H., Nettleton, D., Soller, M. & Dekkers, J. C. M. (2005). Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genetical Research* **86**, 77–87.