

The extent and distribution of linkage disequilibrium in a multi-hierarchic outbred canine pedigree

Xiang-Yang Lou,¹ Rory J. Todhunter,² Min Lin,¹ Qing Lu,¹ Tian Liu,¹ Zuoheng Wang,¹ Stuart P. Bliss,² George Casella,¹ Gregory M. Acland,³ George Lust,³ Rongling Wu¹

¹Department of Statistics, 533 McCarty Hall C, University of Florida, Gainesville, Florida 32611, USA

²Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA

³James A. Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA

Received: 4 February 2003 / Accepted: 9 April 2003

Abstract

A canine integrated linkage-radiation map has been recently constructed by using microsatellite markers. This map, with a good coverage of the canine genome, allows for a genome-wide search for the extent and distribution of linkage disequilibrium derived from linkage and evolutionary forces. In this study, we genotyped an outbred pedigree between Labrador retriever and Greyhound breeds with a set of microsatellite markers (240) from the canine linkage map. Linkage disequilibrium was measured between all syntenic and nonsyntenic marker pairs. Analysis of syntenic pairs revealed a significant correlation (-0.229 , $P < 0.001$) between linkage disequilibrium and genetic distance (log transformed). Significant linkage disequilibria were observed more frequently between syntenic pairs spaced <40 cM than those spaced >40 cM. There is a clear trend for linkage disequilibrium to decline with marker distance. From our results, a genome-wide screen with markers at low to moderate density (1–2 per 10 cM) should take full advantage of linkage disequilibrium for quantitative trait locus mapping in dogs. This study supports the appropriateness of linkage disequilibrium analysis to detect and map quantitative trait loci underlying complex traits in dogs.

Gametic linkage disequilibrium or linkage disequilibrium (LD) describes the nonrandom association of alleles among different loci in a population. Histor-

ically, LD analysis has interested geneticists because the extent and distribution of LD can be used to infer the evolutionary history of population size and structure (Lewontin 1964, 1988; Hedrick 1987). This interest has been intensified with the advent of advanced molecular technologies, especially single nucleotide polymorphism (SNP) (Taillon-Miller et al. 2000) and microsatellite markers (Farnir et al. 2000), which have the power to detect polymorphic genes across the entire genome. More recently, LD has also been widely used for fine-scale mapping of genes predisposing to human diseases (reviewed in Ardlie et al. 2002) and complex traits in animals (Haley 1999). LD-based mapping, making full use of historical recombination events, can position a quantitative trait locus (QTL) to small chromosomal segments in the order of <1 cM (Ardlie et al. 2002). For example, LD mapping has been successfully used to map Hirschsprung's disease in a large, inbred Mennonite kindred (Puffenberger et al. 1994).

The efficacy of LD analysis depends upon the level of LD in the population studied, its distribution and heterogeneity across the genome, its relationship with genetic or physical distances, and the informativeness of mapping markers used. Kruglyak (1999) estimated by simulation that useful levels of LD are unlikely to extend beyond 3 kb in the human. Although this theoretical prediction was supported by a molecular study (Dunning et al. 2000), many other studies found that strong LD could extend beyond 1 Mb (Taillon-Miller et al. 2000; Boehnke 2000). A number of molecular experiments have been performed in humans to investigate and interpret the differences in LD between different populations (Reich et al. 2001), different historical stages of the same population (Wilson and Goldstein 2000), or different chromosomal regions (Taillon-Miller et al. 2000).

Correspondence to: R. Wu, E-mail: rwu@stat.ufl.edu

Apart from its application for mapping human populations, the extent and distribution of linkage disequilibrium in livestock species have been examined for high-resolution mapping of QTL for agriculturally important traits (Haley 1999; Farnir et al. 2000; McRae et al. 2002). A limited number of LD studies in animal species suggested that LD was generally greater, and also extended for greater distances, in livestock than human owing to frequent occurrence of the evolutionary forces causing LD, such as genetic drift, population structure, admixture, selection, and small effective population sizes (Haley 1999). Despite its significant importance in designing a high-resolution QTL mapping experiment, exact information about the distance beyond which LD extends across the canine genome is lacking. In cattle, Farnir et al. (2000) found that, although microsatellite markers spaced <5 cM display strong LD, considerable LD can extend for at least 20 cM, and that a high proportion of significant disequilibrium between marker pairs occurs between unlinked markers on different chromosomes. Similar conclusions about the relationship between LD and map distances were also observed in sheep (McRae et al. 2002). The frequent occurrence of LD between distantly spaced markers or unlinked markers in livestock implies the importance of developing a strategy to avoid false positive results (spurious LD) when LD is used to fine-map genes for complex traits.

As part of our ongoing canine genome project, this study explores the possibility of using LD to map quantitative trait loci affecting canine hip dysplasia (CHD) in a multi-hierarchical pedigree constructed by mating dysplastic Labrador retrievers with unaffected Greyhounds (Todhunter et al. 1999). Comparative studies suggested that these two dog breeds display remarkable discrepancies in morphological features related to CHD (Todhunter et al. 1999). Seven founding Greyhounds and five founding Labrador retrievers were intercrossed, followed by backcrossing F_1 's to the Greyhounds and Labrador retrievers and intercrossing the F_1 's. A series of subsequent intercrosses among the progeny at different generation levels maximized phenotypic ranges in CHD-related quantitative traits and the chance to detecting substantial LD (Todhunter et al. 2003a, 2003b; Bliss et al. 2002). In total, 147 dogs from this outbred population were genotyped at 240 microsatellite markers scanning 39 pairs of chromosomes (Mellersh et al. 1997, 2000; Breen et al. 2001). Before LD-based mapping approaches can be useful for mapping CHD in this pedigree, two fundamental questions should be addressed: (1) How does intrachromosomal LD extend over genetic distance? (2) Does gametic phase disequilibrium exist

between nonsyntenic loci in dog? The first question is related to the determination of appropriate marker map density in target canine genomic regions, whereas the second question is related to the likelihood of detecting spurious LD in QTL mapping. In this article, these two questions are explored through a genome-wide LD analysis.

Materials and methods

Animal pedigree. Seven Greyhound founders and eight Labrador retriever founders are intercrossed to generate a number of F_1 families. These F_1 families were backcrossed to the founders and intercrossed to form F_2 families, and backcross dogs were bred. The offspring at different hierarchical generation levels were crossed, as shown in Fig. 1. In this study, we sampled 147 dogs from different generations for LD analysis.

Genotype determination. Breen et al. (2001) constructed an integrated linkage-radiation hybrid (RH) map of the canine genome. The RH map comprises 1078 microsatellites, 320 dog gene markers, and 102 chromosome-specific markers, and the linkage map contains 354 canine-specific microsatellite markers. These two maps were integrated through sets of common markers, 251 RH/genetic, 102 RH/cytogenetic, and 52 linkage/cytogenetic, distributed on both maps. From this integrated map, we genotyped 240 microsatellite markers, 142 of which are from the linkage map, and 98 of which are from the RH map (Table 1). These 240 markers are distributed in 39 different chromosomes as determined by Breen et al. (2001). Microsatellite genotyping was performed at the NHLBI Mammalian Genotyping Service (Marshfield, Wis.).

Measurements of linkage disequilibrium. As expected, a number of alleles ranging from 2 to 25 were observed at all these microsatellite markers genotyped in the canine pedigree. We used two different approaches to measure LD between a pair of markers. The first approach was to estimate the LD (D_{ij}) between all possible nonalleles from two given markers \mathcal{A} , with alleles A_i ($i = 1, \dots, u$), and \mathcal{B} , with alleles B_i ($i = 1, \dots, v$), and then to calculate an overall measurement of the LD. D_{ij} is expressed as

$$D_{ij} = p_{ij} - p_i q_j,$$

where p_{ij} is the estimated frequency of gamete $A_i B_j$, and p_i and q_j are the estimated frequencies of alleles A_i and B_j , respectively. We derived the EM algorithm to obtain the maximum likelihood estimates (MLEs) of p_{ij} , and then to obtain the MLEs of p_i , q_j and D_{ij} by

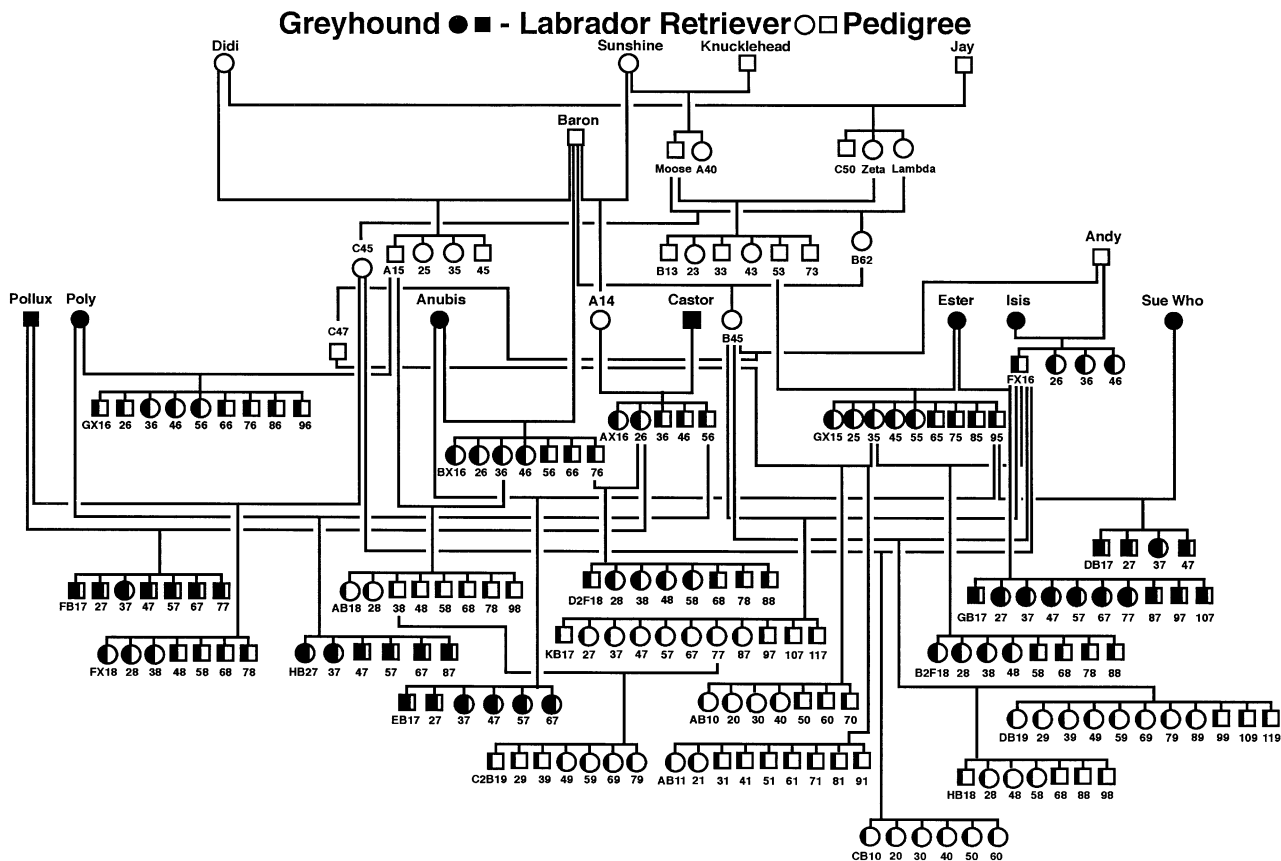


Fig. 1. Diagram of an outbred pedigree in dog. Squares and circles represent males and females, respectively. Filled and open portions of each symbol represent the proportion of Greyhound and Labrador retriever alleles, respectively, possessed by that dog.

solving a system of regular equations. The maximum value that D_{ij} can have depends strongly on allele frequency, expressed as

$$D_{\max} = \begin{cases} \min[p_i q_j, (1 - p_i)(1 - q_j)]; & D_{ij} < 0 \\ \min[p_i(1 - q_j), (1 - p_i)q_j]; & D_{ij} > 0 \end{cases}$$

A normalized measure of LD that avoids this dependence on allele frequency is $D'_{ij} = D_{ij}/D_{\max}$ (Lewontin 1964). The degree of D_{ij} is often assessed by its absolute value. The overall measurement of LD between two markers is calculated by

$$D' = \sum_{i=1}^u \sum_{j=1}^v p_i q_j |D'_{ij}|,$$

where u and v are the number of alleles at markers \mathcal{A} and \mathcal{B} , respectively (Hedrick 1987).

The second approach for measuring LD is to collapse multiple alleles to simulate di-allelic markers (Weir 1996). By treating the most common allele as A or B and the rest of the alleles as non- A (\bar{A}) or non- B (\bar{B}) at markers \mathcal{A} and \mathcal{B} , respectively, the LD and normalized LD were estimated by the pro-

cedure above for multiallelic markers. For di-allelic markers, another useful measure of LD, the square of the correlation coefficient between the two markers, $r^2 = D^2/p_{\bar{A}P}p_{\bar{B}P}$, was also calculated. Weiss and Clark (2002) discussed several desirable genetic and statistical properties of r^2 .

Testing for the significance of linkage disequilibrium. The significance of LD between markers A and B was tested by using a likelihood ratio approach. The likelihood value under the null hypothesis $D' = 0$ is calculated by assuming that $p_{ij} = p_i q_j$. Note that only when all D'_{ij} 's are 0, $D' = 0$ because D' is a sum of absolute values. The log-likelihood ratio test statistics were calculated by comparing the likelihoods under the null and alternative hypotheses.

The critical threshold can be estimated through Monte Carlo simulations. Using the same sample size as for the dog pedigree in this study, we simulated the observations of two-marker genotypes under the null hypothesis $D' = 0$. The simulated data were then subjected to LD analysis, and then the log-likelihood ratios (LRs) were calculated. The simulation was repeated 1000 times to determine the empirical dis-

Table 1. The calculated map distances for the 240 microsatellite markers chosen from the linkage map and RH map published in Breen et al. (2001) for 38 autosomes and 2 sexual chromosomes in canine

Marker name	cM	Marker name	cM	Marker name	cM	Marker name	cM
FH2294	0.00	FH2226	9.23	PEZ8*	0.00	CFA27	59.00
FH2598	10.89	CFA07	81.00	REN310J13*	15.18	REN51i12	0.00
C01.643	3.21	C08.618	6.80	REN50B03*	21.00	REN309N19*	8.10
REN47D17*	16.41	FH2138	8.60	FH2321	30.18	REN146G17*	8.10
FH2309	1.31	REN288F11*	9.46	COS 15	6.80	FH2585	17.20
D01505*	5.57	FH2144	8.64	CFA17	25.00	DTRCN14*	15.40
C01.424	8.30	C08.410	7.20	AHT130	0.00	C28.176	3.80
C01.246	18.60	PEZ11*	8.80	FH2429	13.00	CFA28	55.00
C01.673	14.55	CFA08	62.00	FH3010*	14.56	CPH9*	0.00
FH2313	3.77	REN177B24*	0.00	REN47J11*	6.99	REN45F03*	7.22
FH2016	12.40	REN278L10*	5.78	REN42L13	28.55	FH2328	13.45
CFA01	102.00	C09.474	17.71	REN183B03*	0.16	FH2609*	23.21
AHT111	39.90	FH2186	2.10	FH2356	4.64	CFA29	39.00
FH2132*	3.50	REN75M10*	14.04	CFA18	68.00	LEI-1F11	0.00
C02.894	3.50	FH2263	3.46	FH2380	0.00	AHTH134Ren*	12.44
REN150M24*	14.52	GALK1	12.40	FH2279	2.90	REN245M07*	3.86
FH2613*	11.22	CFA09	55.00	REN91114*	10.11	REN50N18*	8.10
FH2225	0.56	C10.865	0.00	AHT124	4.59	REN105i08*	26.30
FH2608*	2.20	C10.602	18.08	PEZ3	14.30	REN248F14*	4.65
AHT132	2.20	FH2422	11.11	CFA19	32.00	FH2050	3.44
C02.609	6.90	ZUBECA1*	10.70	FH2158	0.00	CFA30	59.00
FH2087U	13.10	C10.16	13.20	REN193A22*	20.80	REN239G04*	33.13
FH3006*	0.88	FH2293	6.24	CPH16	5.70	REN265M13*	11.26
CFA02	99.00	FH2537	15.48	FH2528	22.90	FH2582	1.91
FH2302	0.00	CFA10	76.00	REN55P21*	20.27	REN43H24*	8.17
FH2107	11.00	DGN13*	0.00	CFA20	58.00	FH2189*	22.79
PEZ12	10.70	C11.873	16.41	FH2312	0.00	CFA31	53.00
CPH19	1.50	REN147O02*	8.85	FH2603	8.90	AHT127*	0.00
REN47O24*	6.19	FH2019*	8.85	FH2441	7.20	FH2238	9.88
C03.895	12.81	FH2319	7.50	REN285A14*	2.94	REN41D20*	6.32
FH2541	7.70	FH2096	19.40	FH2233	18.06	CPH2	4.08
FH2131	12.60	AHT137	3.50	CFA21	37.00	CFA32	21.00
FH2137	10.80	CFA11	71.00	FH2538	0.00	FH2165	0.00
CFA03	73.00	C12.852	0.00	C22.279	15.40	REN291M20*	23.40
FH2457	0.00	C12.406	10.30	FH2109	7.00	REN98D17*	8.31
AHT103	14.40	REN260I06*	9.81	REN107H05*	4.90	FH2361	3.40
FH2097	4.90	PEZ5	12.73	REN262G14*	5.93	FH2507	1.20
REN160J02*	3.66	FH2401	16.74	REN42F10*	11.40	CFA33	36.00
FH2142	8.04	FH2054	1.39	CFA22	58.00	REN314H10*	0.00
FH2412	22.40	C05101*	1.40	FH2001	0.00	REN53L08*	13.50
FH2534	12.70	C07003*	26.51	C23.745	16.30	FH2377	18.74
REN298N18*	27.02	FH2202*	19.58	C23.123	10.10	REN160M18*	19.57
CFA04	74.00	CFA12	63.00	FH2626*	10.23	CFA34	51.00
CPH14	0.00	REN166I13*	24.98	REN46F18*	2.54	REN172L08*	0.00
C05.377	11.80	REN65L04*	2.67	FH2508	7.24	REN94K23*	3.85
C05.771	18.50	AHT121*	10.88	CFA23	63.00	REN214H22*	6.28
REN192M20*	11.01	REN307K04*	9.03	FH2079	0.00	REN01G01	11.91
CPH18	2.69	C13.391	9.03	AHT125	11.70	CFA35	27.00
ZUBECA6	18.40	CFA13	56.00	FH2261	4.00	REN85C13*	0.00
REN265H13*	5.96	PEZ10	2.20	FH2168	9.70	FH2611*	11.54
REN42N13*	2.58	FH2258	15.00	CFA24	33.00	CFA36	8.00
FH2140	5.86	REN289L09*	8.48	C25.213	8.60	AHT135*	0.00
FH2594	5.10	REN169D01*	10.09	FH2006	14.10	FH2532	18.36
CFA05	85.00	FH2547	11.33	FH2087L	17.90	H10101*	7.29
C06.636	0.00	C14.866	5.80	FH2324	2.60	AHT133	4.51
FH2370	8.80	FH2060	16.00	FH2318	1.30	REN67C18*	6.00
FH2119	5.50	CFA14	70.00	REN166C13*	14.32	CFA37	47.00
FH2164	8.60	FH2360	10.87	CFA25	51.00	AHTH91Ren	0.00
REN149M14*	3.10	FH2088*	3.49	C26.733	0.00	REN02C20	36.90
FH2561	18.60	REN230G12*	5.53	FH2130	9.40	REN86G15*	9.28
CFA06	49.00	REN303E22*	7.11	N41	8.50	REN164E17*	2.92

Table 1. continued

Marker name	cM	Marker name	cM	Marker name	cM	Marker name	cM
FH2301	0.00	FH2535	9.14	REN01O23*	12.60	CFA38	39.00
FH2581	13.40	RVC1*	4.27	AHTK211	15.40	FH2584	25.00
C09703*	12.72	REN06C11*	12.27	CFA26	53.00	REN130F03*	13.50
FH2201	2.98	CFA15	60.00	PEZ6	0.00	FH2985*	19.30
REN149P06*	10.15	REN130B10*	0.00	LEI002	23.80	FH2548	5.80
FH2174	10.15	REN292N24*	16.30	REN277O05*	5.49	FH3027*	4.35
1B10	5.10	C16.147	10.81	REN208N23*	11.79	CFAx	73.00
REN162C04*	11.93	FH2175	0.80	C27.442	12.01	AF192268	
VIASD10*	5.13	CFA16	54.00	FH2289	2.80	CFAy	

The chromosome names (CFA01 – CFA39, CFAx) and total lengths of linkage groups are given in bold. The markers with asterisks are from the RH map, whereas those without asterisks are from the linkage map.

tribution of the LRs and the critical value at the significance level $\alpha = 0.05$ from the distribution.

Results

Breen et al. (2001) reported an integrated linkage-radiation hybrid map of the canine genome comprising 1800 markers. These markers are grouped and ordered in 38 autosomes and two sexual chromosomes, from which 240 microsatellite markers displaying a good coverage of each chromosome were chosen for LD analysis in this study. A total of 147 dogs sampled from an outbred pedigree of seven Labrador retriever and eight Greyhound founders were genotyped for this battery of 240 microsatellite markers (Breen et al. 2001; Richman et al. 2001). The average number of alleles observed in these founders is 6.2 for all the markers, whereas the average heterozygosity measured in the overall population is 0.643. The markers with a high number of alleles tend to be more heterozygous, but the markers with a small number of alleles display various degrees of heterozygosity (Fig. 2).

The distances between adjacent markers were estimated in centiMorgans (cM) for the linkage map and in centiRays₅₀₀₀ (cR₅₀₀₀) for the RH map (Breen et al. 2001). To calculate the relationship between distances on the RH map and genetic distances, the cR/cM ratio was calculated for pairs of adjacent markers that mapped to the same RH and linkage groups. The calculated genetic distances for the 240 markers and chromosome numbers on which these markers are located are listed in Table 1.

The extent of LD was evaluated for each marker pair by analyzing all possible alleles (the multiallelic model) by using the formulae described in the Materials and methods. The extent of LD was first examined for pairs of syntenic markers. Of all 683 syntenic marker pairs considered, 645 (94.5%) were statistically significant ($\alpha = 0.01$), indicating that substantial intrachromosomal LD exists throughout the canine genome in this pedigree. Some of the

syntenic marker pairs were highly significantly associated; for example, the *P*-value for the first two markers on chromosome CFA01 was 1.87e-52. We next examined the relationship between the extent of LD and genetic distance for different marker pairs. But this relationship cannot be precisely examined with measures of LD from the multiallele model, because marker pairs with more alleles tend to display stronger LD than those with fewer alleles (see the simulation result presented at the end of this section). We evaluated the extent of LD for marker pairs, using the di-allelic model by collapsing multialleles into two alleles, the common allele vs. the rest. We performed a simulation study to investigate the effect of this allele grouping strategy on the power to detect disequilibrium.

The square of the correlation coefficient (r^2) between different markers estimated from the di-allelic model (see the Materials and methods) allows for the examination of the relationship between LD and marker distance (Fig. 3). The value of r^2 was significantly negatively correlated with log-transformed genetic distance, with a correlation coefficient of -0.229 ($P < 0.001$). The value of r^2 reduces to a half from marker pairs <5 cM apart to markers spaced 10–20 cM, and further reduces, but to a lesser extent, when markers are more distant. There is substantial variation among the r^2 values for closer marker pairs than for more distant ones (Fig. 3). This variation reduced with marker distance until markers were spaced 40 cM or more. Thus, it is suggested that high levels of LD extend to 20–40 cM throughout the canine genome.

The decrease of LD between marker pairs with their increased distance can also be observed in terms of the normalized LD (D') estimated from the di-allelic model, but with the trend less evident than that observed in terms of r^2 (results not shown). For example, two distant markers (>60 cM apart) may still have a large value of $D' = 1$. The reason for this is that D' is more sensitive than r^2 to allele frequencies; it will tend to be high if there is an allele in low frequency at either

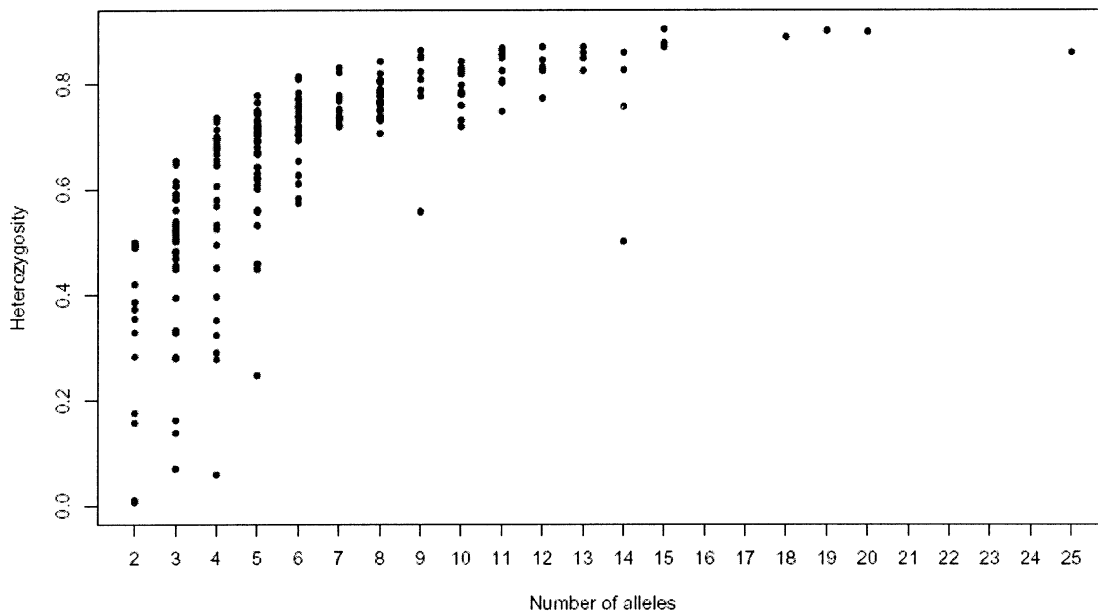


Fig. 2. Relationship between the number of alleles and the heterozygosity observed at 240 microsatellite markers in the dog population.

of the two markers. Thus, the high values of D' at marker distance >40 cM are likely to reflect the low allele frequency of these markers.

A one-way ANOVA comparing mean r^2 (log-transformed) across 39 pairs of chromosomes detected no evidence for interchromosomal heterogeneity in LD ($F_{38,644} = 0.91$, $P = 0.632$). As shown in Fig. 4, however, the distributions of r^2 varied from chromosome to chromosome. For example, Chrs 2, 5, 13, and 15 have r^2 values that decay significantly with genetic distance. The degree with which r^2 decays with genetic distance is low for Chrs 14, 17, 18, 30 and 31. For other chromosomes, the extent and distribution of LD was intermediate between these two extremes. Non-uniform distribution of LD across the canine genome suggests different impacts of evolutionary forces on targeted chromosomal regions.

We also investigated possible gametic phase disequilibrium between non-syntenic markers. Among 26,869 nonsyntenic marker pairs, significant LD was observed eightfold as often as expected under random segregation (11017/26,869 or 41%). The average r^2 value was 0.25 for nonsyntenic marker pairs, therefore quite similar to the r^2 value for distant (>60 cM) syntenic markers.

Simulations. Our LD analysis in this study is based on microsatellite markers characterized by multiple alleles. Direct analysis of multiallelic microsatellites relies upon the estimates of a large number of linkage disequilibria between different pairs of nonalleles at two given markers. This, on one

hand, provides a powerful means of accurately specifying the genetic structure of the population under consideration, but, on the other hand, is likely to result in biased estimates for the relationship between the extent of LD and genetic distance. It is possible that collapsing multiallelic microsatellites to simulate di-allelic loci provides an effective analysis of this relationship. Here, we performed simulation studies to compare the effect on the estimate of LD of the multiallelic compared with the di-allelic models.

Suppose two four-allelic markers, at each of which there is an identical frequency (0.25) among four alleles. These two markers produce a total of 16 disequilibria, 15 of which are independent because the 16 disequilibria are summed to 1. The observations were simulated under different frequencies of disequilibrium (0/15, 3/15, and 15/15), different degrees of disequilibrium (none, weak, and strong) and different sample sizes (100, 200, and 400). As expected, the power to detect disequilibrium from both multiallelic and di-allelic models increased with sample size, increased degree of disequilibrium, and higher frequency (Table 2). The two models generated almost identical results when not all disequilibria occurred between nonalleles at two different markers. When all possible disequilibria occurred, however, the multiallelic model displayed higher power to detect disequilibrium than the di-allelic model. For low disequilibrium, a large sample size was necessary to detect disequilibrium with the multiallelic model. We also estimated the mean square errors of the maximum likelihood estimates

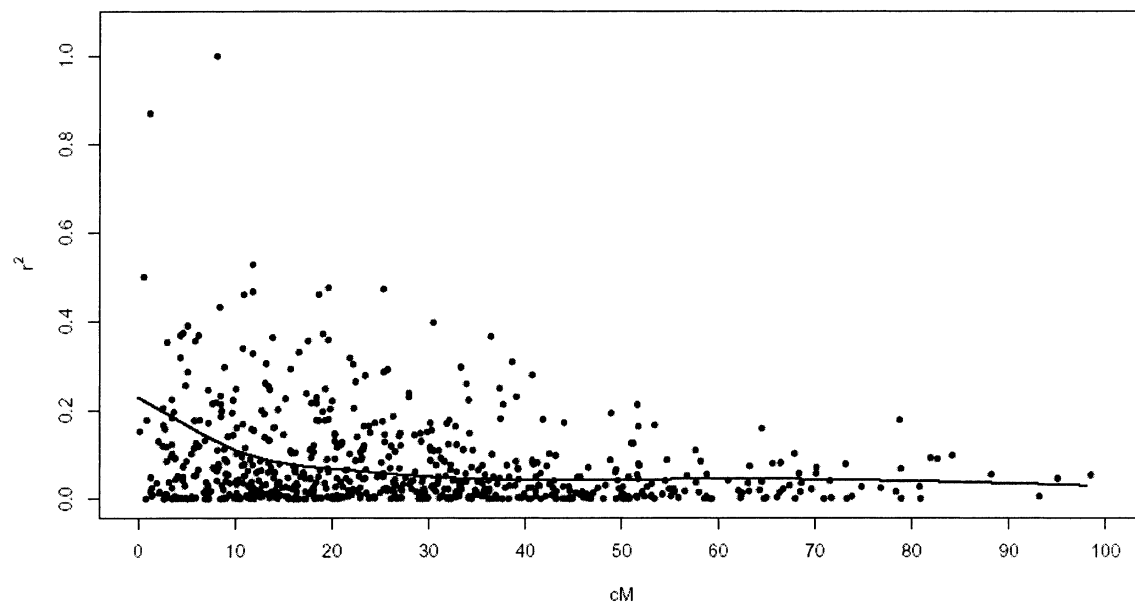


Fig. 3. Distribution of r^2 values observed between syntenic marker pairs as a function of genetic distance in centiMorgan (cM).

of disequilibria and found that they are not different between the analytical models (results not shown).

It is interesting to note that the normalized LD, D' , can be overestimated when the method of allele grouping is used (results not shown). This may be due to the fact that allele grouping based on the common allele vs. rest amplifies the difference in allele frequency. When the rare allele at one marker is associated with the common allele at the other marker, i.e., D is negative, this increased difference of allele frequency overestimated D' because it was obtained by dividing D by the product of the frequencies of two rare alleles, each from a marker. For this reason, if multiple alleles are grouped in terms of the common allele vs. the rest, r^2 (displaying a lower sensitivity to allele frequency) rather than D' is recommended to describe the pattern of LD distribution through the genome.

Discussion

The extent and distribution of LD can be used to estimate and predict the origin and evolutionary history of a population (Lewontin 1964). The occurrence of LD is more frequent in a population that has undergone recent admixture, a small effective population size, and intense selection. All of these processes were experienced in the domestic dog, which makes it a particularly attractive model for the study of LD. According to Ostrander and Kruglyak (2000), the purebred dogs of today are derived from a limited genetic pool and, thus, are expected to exhibit substantial LD over their genome.

By analyzing a battery of microsatellite markers genotyped from a published integrated linkage-radia-

tion hybrid map (Breen et al. 2001), we investigated genome-wide linkage disequilibrium (LD) in a canine pedigree constructed with multiple Greyhound founders and Labrador retriever founders. We found remarkable LD across all 39 chromosomes and its appreciable decline with genetic distance in dogs. Two striking findings of our study make it particularly similar to two earlier analyses of cattle (Farnir et al. 2000) and sheep (McRae et al. 2002). First, the levels of LD between marker pairs were significantly negatively correlated with genetic distance (log-transformed) and high levels of LD extended to 5–10 cM throughout the canine genome. It is likely that this considerable extension reflects a narrow bottleneck in the history of dog populations during their domestication (Ostrander and Kruglyak 2000). A recent study of the pattern of phylogeographic variation suggests that all domestic dogs, in spite of their remarkable discrepancies in morphology and anatomy, originate in a single common gene pool of wolves in East Asia (Savolainen et al. 2002). In practice, this extension can guide us to choose the minimum number of markers needed to scan the genome. Because LD declined, to some extent, with increasing distance, a low or intermediate density of markers (1 or 2 per 10 cM) would be sufficient for first-pass LD screening in our canine pedigree. This argument can be further tested through theoretical simulations as performed by Kruglyak (1999), who advocated a high marker density for genome-wide LD mapping with SNPs.

Second, substantial LD occurred between pairs of unlinked markers on different chromosomes. This may be due to a limited number of founders used in our pedigree. But it may also support a view of a

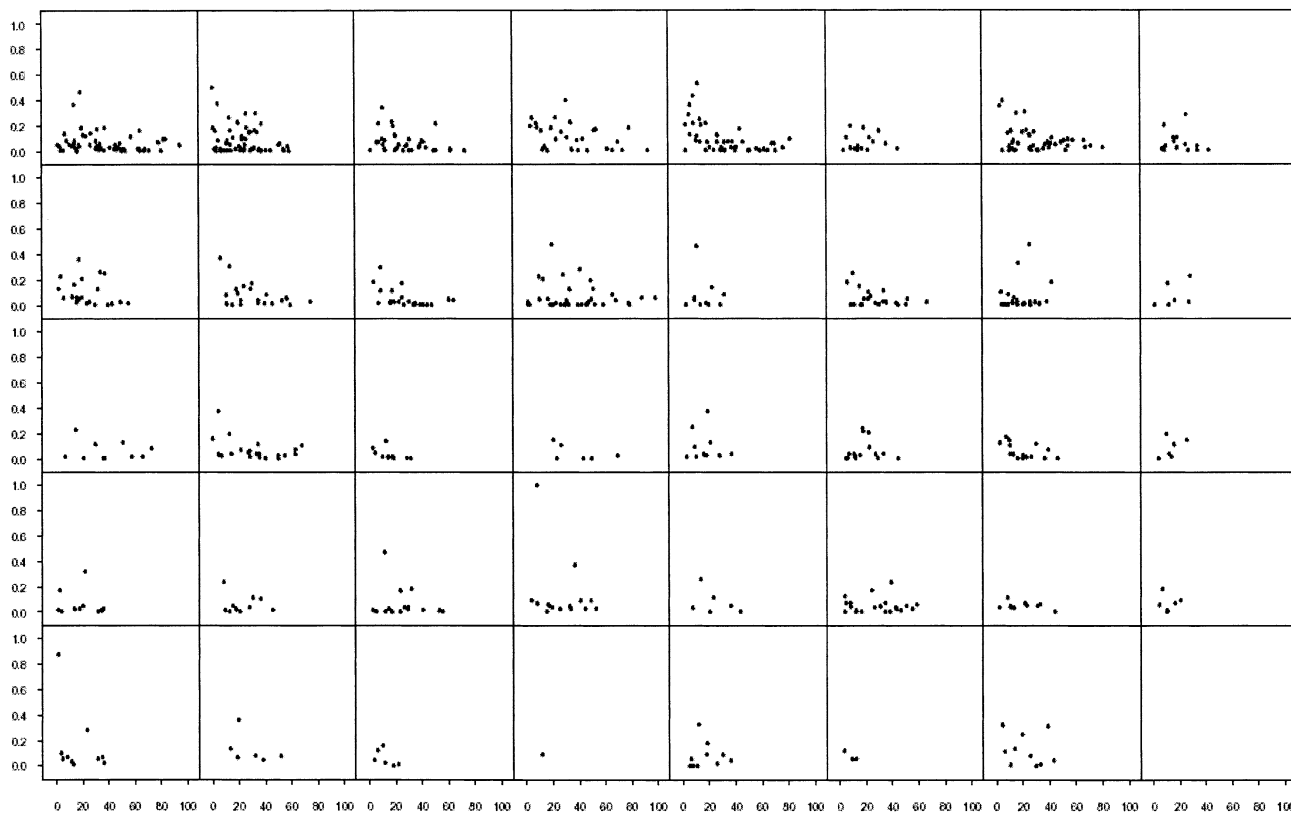


Fig. 4. Interchromosomal heterogeneity in the extent and distribution of LD among 39 chromosomes. The first, second, third, and fourth row include chromosomes CFA01–CFA10, CFA11–CFA20, CFA21–CFA30, and CFA31–CFA39 in order from the left to right, respectively.

narrow bottleneck in general in dogs (Savolainen et al. 2002). Population genetics theory predicts that disequilibrium due to evolutionary forces (rather than tight linkage) should have declined to negligible levels for nonsyntenic markers, provided that the population was randomly mating and reasonably large. The common occurrence of LD between nonsyntenic loci suggests that LD mapping as a means to locate genes underlying complex traits can be problematic in dog populations. But a newly developed strategy combining linkage and linkage disequilibrium analysis can be used to overcome this limitation of LD mapping alone. Wu and Zeng (2001) and Wu et al. (2002) discussed theoretical models for a joint analysis of linkage and linkage disequilibrium in natural populations. A similar idea was used by Farnir et al. (2002) and Meuwissen et al. (2002) to successfully map quantitative trait loci for economically important traits in a 1 ~3-cM marker interval.

We used microsatellite markers to understand the structure and organization of the canine genome as opposed to commonly used di-allelic SNPs. The reasons are two-fold: (1) a number of dense microsatellite markers have already been used in genetic studies in humans (Hall et al. 2002) and other or-

ganisms (Breen et al. 2001; Farnir et al. 2000; McRae et al. 2002), and (2) di-allelic markers may have less power to detect disequilibrium than multiallelic markers, especially in outcrossing species (Chapman and Wijsman 1998). A major problem related to the use of microsatellites is how to measure multiallelic LD. Hedrick's (1987) estimator of the normalized LD between all pairs of alleles at two multiallelic markers is relatively robust to variation in allele frequency, but it becomes imprecise and is computationally demanding when the number of alleles increases. This is especially the case in our study, in which the number of alleles is more than 6–8 for many markers, but only a limited dog population was available. Attempts have been made to collapse multiallelic markers into di-alleles based on (1) the common allele vs. the remaining alleles grouped together (Weir 1996), (2) the allele size that groups the alleles according to the bimodal distribution of allele frequency vs. size, and (3) the method that uses the observed allelic associations between marker pairs to determine the groupings (Cox et al. 1998). Theoretically, the third method can capture the most information about multiallelic LD, but it is not straightforward in practice. The common allele

Table 2. Comparisons in the power to detect disequilibrium from the multiallelic and di-allelic models

Size/frequency of LD	Sample size		
	100	200	400
		Multiallelic model	
No disequilibrium	5	9	4
D = 0.01 (3/15)	7	10	17
D = 0.04 (3/15)	65	98	100
D = 0.01 (15/15)	19	30	73
D = 0.04 (15/15)	100	100	100
		Di-allelic model	
No disequilibrium	6	5	4
D = 0.01 (3/15)	6	7	21
D = 0.04 (3/15)	59	87	100
D = 0.01 (15/15)	8	9	24
D = 0.04 (15/15)	61	86	100

grouping strategy was used in this study. We performed a simulation study to compare the power to detect disequilibrium from the multiallelic model (making use of all multiple alleles) and the di-allelic model (grouping the multiple alleles into two types). Unless disequilibria between all possible pairs of nonalleles occur, the di-allelic model generally has power similar to the multiallelic model. But our simulation also showed that D' may not be a good indicator of LD by using the grouping strategy where $D < 0$, because this grouping amplifies the differences in allele frequency. When rare alleles were present, skewed D' was also observed in a real-life example by McRae et al. (2002). Being less sensitive to allele frequencies, we recommend using r^2 for LD analysis based on collapsed multiallelic markers.

Most of current LD analyses are based on the assumption that the population studied is randomly mating. When this assumption is violated, as is often observed in practice, the introduction of the coefficient of Hardy-Weinberg disequilibrium is necessary. Recently, Yang (2000, 2002) proposed a multilocus statistic to examine zygotic associations in non-equilibrium populations. Different disequilibria due to a single locus or multiple loci can be summarized in such a statistic. It would be interesting to incorporate Yang's zygotic association theory to map QTL in an arbitrary natural population.

In this article, we present a genome-wide search for the extent and distribution of linkage disequilibrium by using markers randomly chosen from the published canine maps. From an evolutionary perspective, the findings reported can shed light on the impacts of various evolutionary forces on the canine genome during domestication (see also Farnir et al. 2000 in cattle and McRae et al. in sheep). But these findings are far from an understanding of the fine structure and function of the genome in

terms of positional cloning of target genes. In humans, detailed analyses of haplotype variation and structure have been studied for individual chromosomes with single nucleotide polymorphisms (Dunning et al. 2000; Taillon-Miller et al. 2000; Dawson et al. 2002). Our work provides a first step toward the construction of a fine-structured map of LD in canine and, ultimately, the identification of genetic variants associated with any complex phenotype of interest.

Acknowledgments

We thank two anonymous referees for constructive comments on this manuscript, which led to a better presentation. This work was supported by a grant from the Morris Animal Foundation, NIH grant AR36554, the Consolidated Research Grant Program, the Cornell Advanced Technology, Biotechnology Program, Nestle Purina Inc., NALBI Mammalian Genotyping Service, Marshfield Medical Research Foundation, Marshfield, WI, Cornell University College of Veterinary Medicine Unrestricted Alumni Funds, and an Outstanding Young Investigators Award (No. 30128017) of the National Natural Science Foundation of China to R. Wu. We thank Margaret Vernier-Singer, Alm Jo Williams, and Liz Grisanzio for expert technical assistance.

References

1. Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3, 299–309
2. Bliss S, Todhunter RJ, Quaas R, Casella G, Wu RL et al. (2002) Quantitative genetics of traits associated with hip dysplasia in a canine pedigree constructed by mating dysplastic Labrador Retrievers with unaffected Greyhounds. *Am J Vet Res* 63, 1029–1035

3. Boehnke M (2000) A look at linkage disequilibrium. *Nat Genet* 25, 246–247
4. Breen M, Jouquand S, Renier C, Mellersh CS et al. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Mamm Genome* 11, 1784–1795
5. Chapman NH, Wijsman EM (1998) Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. *Am J Hum Genet* 63, 1872–1885
6. Cox A, Camp NJ, Nicklin MJH, di Giovine FS, Duff GW (1998) An analysis of linkage disequilibrium in the interleukin-1 gene cluster, using a novel grouping method for multiallelic markers. *Am J Hum Genet* 62, 1180–1188
7. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S et al. (2002) A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 418, 544–548
8. Dunning AM, Durocher F, Healey CS, Teare MD, McBride SE et al. (2000) The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am J Hum Genet* 67, 1544–1554
9. Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N et al. (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Res* 10, 220–227
10. Farnir F, Grisart B, Coppieters W, Riquet J, Berzi P et al. (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
11. Haley C (1999) Advances in quantitative trait locus mapping. In: *From J.L. Lush to Genomics: Visions for Animal Breeding and Genetics*, JCM Dekkers, SJ Lamont, MF Rothschild (eds) (Ames, Iowa: Iowa State University Press), pp 47–59
12. Hall D, Wijsman EM, Roos JL, Gogos JA, Karayiorgou M (2002) Extended intermarker linkage disequilibrium in the afrikaners. *Genome Res* 12, 956–961
13. Hedrick P (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* 117, 331–341
14. Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22, 139–144
15. Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49, 49–67
16. Lewontin RC (1988) On measures of gametic disequilibrium. *Genetics* 120, 849–852
17. McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM et al. (2002) Linkage disequilibrium in domestic sheep. *Genetics* 160, 1131–1122
18. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K et al. (1997) A linkage map of the canine genome. *Genomics* 46, 326–336
19. Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C et al. (2002) An integrated linkage-radiation hybrid map of the canine genome. *Mamm Genome* 11, 120–130
20. Meuwissen THE, Karlsten A, Lien S, Olsaker I, Goddard ME (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161, 373–379
21. Ostrander EA, Kruglyak L (2000) Unleashing the canine genome. *Genome Res* 10, 1271–1274
22. Puffenberger EG, Hosoda K, Washington SS, Nakao K, deWit D, Yanagisawa M and Chakravart A (1994) A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* 79, 1257–1266
23. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC et al. (2001) Linkage disequilibrium in the human genome. *Nature* 411, 199–204
24. Richman M, Mellersh CS, Andre C, Gailbert F, Ostrander EA (2001) Characterization of a minimal screening set of 172 microsatellite markers for genome-wide screens of the canine genome. *J Biochem Biophys Methods* 47, 137–149
25. Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298, 1610–1613
26. Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T et al. (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat Genet* 25, 324–328
27. Todhunter RJ, Acland GM, Olivier M, Williams AJ, Vernier-Singer M et al. (1999) An outcrossed canine pedigree for linkage analysis of hip dysplasia. *J Hered* 90, 83–92
28. Todhunter RJ, Casella G, Bliss SP, Lust G, William AJ et al. (2003a) Power of a dysplastic Labrador Retriever-Greyhound pedigree for linkage analysis of hip dysplasia. *Am J Vet Res* 64, 418–424
29. Todhunter RJ, Bliss SR, Quaas SR, Lust G, Casella G et al. (2003b) Genetic structure of susceptibility traits for hip dysplasia and microsatellite informativeness of an outcrossed canine pedigree. *J Hered* 94, 39–48.
30. Weir BS (1996) *Genetic Data Analysis II*. (Sunderland Mass.: Sinauer Associates)
31. Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 18, 19–24
32. Wilson JF, Goldstein DB (2000) Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am J Hum Genet* 67, 926–935
33. Wu RL, Zeng Z-B (2001) Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* 157, 899–909
34. Wu RL, Ma C-X, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 160, 779–792
35. Yang R-C (2000) Zygotic associations and multilocus statistics in a nonequilibrium diploid population. *Genetics* 155, 1449–1458
36. Yang R-C (2002) Analysis of multilocus zygotic associations. *Genetics* 161, 435–445