Spectrum of Nonrandom Associations Between Microsatellite Loci on Human Chromosome 11p15

Carlos Zapata,* Santiago Rodríguez,* Guillermo Visedo* and Felipe Sacristán[†]

*Departamento de Biología Fundamental, Facultad de Biología, Universidad de Santiago, 15782 Santiago de Compostela, Spain and [†]Hospital Juan Canalejo, 15006 La Coruña, Spain

ABSTRACT

Most evidence about nonrandom association of alleles at different loci, or gametic disequilibrium, across extensive anonymous regions of the human genome is based on the analysis of overall disequilibrium between pairs of microsatellites. However, analysis of interallelic associations is also necessary for a more complete description of disequilibrium. Here, we report a study characterizing the frequency and strength of both overall and interallelic disequilibrium between pairs of 12 microsatellite loci (CA repeats) spanning 19 cM (14 Mb) on human chromosome 11p15, in a large sample (810 haplotypes deduced from 405 individuals) drawn from a single population. Characterization of disequilibrium was carried out, taking into account the sign of the observed disequilibria. This strategy facilitates detection of associations and gives more accurate estimates of their intensities. Our results demonstrate that the incidence of disequilibrium over an extensive human chromosomal region is much greater than is commonly considered for populations that have expanded in size. In total, 44% of the pairs of microsatellite loci and 18% of the pairs of alleles showed significant nonrandom association. All the loci were involved in disequilibrium, although both the frequency and strength of interallelic disequilibrium were distributed nonuniformly along 11p15. These findings are especially relevant since significant associations were detected between loci separated by as much as 17–19 cM (7 cM on average). It was also found that the overall disequilibrium masks complicated patterns of association between pairs of alleles, dependent on their frequency and size. We suggest that the complex mutational dynamics at microsatellite loci could explain the allele-dependent disequilibrium patterns. These observations are also relevant to evaluation of the usefulness of microsatellite markers for fine-scale localization of disease genes.

M ULTILOCUS genetic theory represents an alternative to the classical theory of population genetics, which considers single genes as the units of selection (Franklin and Lewontin 1970; Lewontin 1974). The study of the phenomenon of nonrandom association of alleles at different loci (gametic disequilibrium) plays a critical role in distinguishing both theories, because if disequilibria were shown not to be relevant in populations, multilocus genetics would then merely be an extension of single gene systems. It follows that characterization of the frequency and intensity of gametic disequilibria across genomes is a necessary step for determining the conceptual framework of population genetics. Furthermore, disequilibrium is a good indicator of recent mutations, genetic drift, bottlenecks, stratification or admixture, and the demographic history of populations (Hill and Robertson 1968; Nei and Li 1973; Ohta 1982; Slatkin 1994; Kruglyak 1999). At a practical level, disequilibrium analysis is successful in achieving fine-scale localization of Mendelian disease genes, (e.g., Kerem et al. 1989; Hästbacka et al. 1994) since it incorporates the effects of historic recombination events between disease mutation and genetic markers. However, the usefulness of disequilibrium in localizing multifactorial disease genes remains an open question (Terwilliger and Weiss 1998; Jorde 2000). Importantly, it must be underlined that previous information on the distribution of disequilibria in a certain region (background disequilibrium) is relevant for designing and interpreting a mapping experiment (Freimer et al. 1997; Terwilliger and Weiss 1998).

Much empirical work over the past 30 years, based on protein loci, seemed to demonstrate that disequilibrium in outcrossing species was important only for very tightly linked protein loci (Hedrick et al. 1978; Lewontin 1985). As pointed out by Lewontin (1985), "the lack of abundant associations between genes in natural populations seemed to make the further development of multilocus studies less relevant." With the advent of disequilibrium studies between DNA polymorphisms, this traditional view did not change essentially. In large

outbred populations, disequilibrium seemed to exist only for markers within ∼0.1 cM (≈100 kb) of DNA (Bodmer 1986; Kruglyak 1999; Ott 1999).

This prevailing view began to change during the past decade. It was demonstrated that the lack of statistical power of traditionally used tests for disequilibrium was actually an important barrier, hindering detection of disequilibria between allozyme loci in populations (Zapata and Alvarez 1992). More powerful statistical techniques revealed that disequilibria of weak to moderate intensity between loosely linked allozyme loci are relatively frequent in populations, even between loci separated by recombination frequencies as large as 12 cM (Zapata and Alvarez 1992). More recently, investigations of the distribution of disequilibria between pairs of single nucleotide polymorphisms (SNPs) and microsatellites, over large (≥4 cM) anonymous regions of human chromosomes, have provided some revealing results (Peterson *et al.* 1995; Laan and Pääbo 1997; Huttley *et al.* 1999; Taillon-Miller *et al.* 2000; Wilson and Goldstein 2000). First, a substantial fraction of pairs of loci are nonrandom associated. Percentages of significant pairs seem to depend very much on the demographic history of populations, ranging from 5–14% (in populations that have undergone rapid growth) to 69% (in stable populations). Second, significant disequilibrium often occurs within ∼4 cM, although a few associations have been detected between polymorphisms separated by >18 cM. Third, the distribution of disequilibrium seems to be nonuniform across the human genome. These findings disclose that disequilibrium extends over wider regions of the human genome than formerly thought, which reopens questions about the importance of the multilocus genetic systems. In addition, the available evidence strongly suggests that the human genome is not only a structural and functional mosaic (Bernardi 1995) but also an evolutionary mosaic characterized by disequilibrium-rich and disequilibrium-poor regions. The implications of such a complex structure for the construction of meaningful multilocus genetic models present us with an extraordinary challenge.

Nevertheless, our current knowledge of the levels of disequilibrium along extensive human chromosomal regions is still limited, in some respects. So far, associations between pairs of microsatellite loci over extensive anonymous human regions have been characterized by testing the null hypothesis of overall gametic equilibrium and by using probabilities resulting from the significance tests as a measure of the strength of overall disequilibrium. This kind of analysis is useful, although clearly insufficient to discern the actual amount (frequency and strength) of disequilibrium existing along human chromosomes. Overall disequilibrium analysis has the advantage of summarizing disequilibria between all possible pairs of alleles at multiallelic loci in a single test of significance and as a single measure of disequilibrium

intensity, which can be particularly useful for loci with a high number of alleles, such as microsatellite markers. Nevertheless, an analysis based solely on overall disequilibrium necessarily forfeits important information underlying multiallelic systems. It must be remembered that the null hypothesis of no overall disequilibrium specifies that there is no disequilibrium for any pair of alleles at two loci. In this way, a given locus pair is judged to be in significant overall disequilibrium irrespective of whether one or many pairs of alleles deviate from random association. Undoubtedly, quantification of the levels of disequilibrium across the human genome requires an assessment of how many of the possible pairs of alleles across locus pairs are in significant disequilibrium. On the other hand, probabilities resulting from the significance tests are not the best tools for determining the strength (weak, moderate, or strong) of disequilibria, or for comparisons across loci, because statistical power is dependent on the sample size, the statistical tests, the number of alleles, their frequencies, and whether the association is positive or negative (Brown 1975; Weir and Cockerham 1978; Thompson *et al.* 1988; Zapata and Alvarez 1992, 1993, 1997a; Slatkin 1994; Ott and Rabinowitz 1997). In fact, decades of statistical studies have established that the best tools for measuring the strength of association are the coefficients or indices of association (Kendall and Stuart 1979; Agresti 1990; Everitt 1997). In addition, estimates of the strength of disequilibrium, using coefficients of association, are needed to establish points of contact between empirical and theoretical research, as the theory of multilocus genetic systems has been framed in terms of these coefficients.

Alleles at multiallelic loci cannot be reasonably considered as a homogeneous whole, and thus, in addition to overall disequilibrium, allele-dependent disequilibrium analyses are clearly needed. A multiallelic locus is a very complex genetic system, comprising alleles differentiated in frequency, age, and evolutionary history (Kruglyak 1999). This complexity is enhanced in the particular case of microsatellite loci. Microsatellites are tandem arrays of short (no more than six bases long) repeats, with complex mutational dynamics. The primary mechanism of mutation at microsatellites is thought to be polymerase slippage during replication, resulting in most of the observed cases in an increase or decrease in the array size of one repeat unit (Schlötterer and Tautz 1992; Macaubas *et al.* 1997). Several lines of evidence indicate that constraints on allele size must exist, including the absence of very long alleles (Garza *et al.* 1995; Goldstein and Pollock 1997). Different possible mechanisms for the regulation of repeat number at microsatellite loci have been proposed, such as bias in the mutation process, selective constraints, tendency of mutability to increase with array size, imperfections in repeated sequences generating a decrease in the mutation rate, and recombinational instability due

to large differences between pairs of alleles at a locus in a given individual (GARZA *et al.* 1995; JIN *et al.* 1996; GOLDSTEIN and POLLOCK 1997; SAMADI *et al.* 1998; FALUSH and IWASA 1999; ELLEGREN 2000). These considerations suggest the possible occurrence of variable patterns of disequilibrium, depending on the frequency and size of the alleles, at microsatellite loci. Surprisingly, these patterns have not yet been investigated for rather loosely linked microsatellite loci.

In this article, we describe the amount of overall and interallelic disequilibrium between all possible pairs of 12 dinucleotide (CA repeats) microsatellites, located on the telomeric region of the short arm of human chromosome 11 (11p15), from a large sample of the Galician population [northwest (NW) Spain]. Distribution of nonrandom associations along 11p15 and allele-dependent disequilibrium patterns were also investigated. The significance and intensity of overall and interallelic disequilibria were obtained by taking into account the sign of the deviations from random association. This new strategy for analyzing disequilibria between multiallelic loci provides both more statistical power for detecting nonrandom associations and more precise estimates of the strength of the disequilibrium. The results of our investigations provide new insights concerning two-locus disequilibrium across extensive regions of the human genome.

MATERIALS AND METHODS

Study subjects: Peripheral blood samples were obtained from 405 random unrelated individuals (169 males and 236 females) and five families (53 individuals). They were all Caucasian individuals of Galicia (NW Spain). Genomic DNA was isolated from blood samples using the QIAamp kit (QIAGEN, Chatsworth, CA). All samples were collected with the approval of the appropriate institutional review board.

Microsatellite markers and genotyping: All individuals were genotyped for 12 dinucleotide $(CA)_n$ repeat loci spanning chromosome region 11p15. Genotyping was performed using the polymerase chain reaction (PCR). The loci typed in this study ordered from telomere (D11S4177) to centromere (D11S926); their chromosomal assignments [Location Data Base (LDB), http://cedar.genetics.soton.ac.uk/public_html/ldb.html] and amplification primers (GÉNÉTHON, ftp://ftp.genethon.fr/pub/Gmap/Nature-1995) are shown in Table 1. PCR amplifications were performed in accordance with conditions described in GÉNÉTHON, with some modifications to optimize each marker. General amplification reactions were carried out in a final volume of 25 μl, containing 120 ng genomic DNA, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 0.1% Triton X-100, 1.5 mM MgCl₂, 50 pmol of each primer, 200 μM dNTP mix, and 1 unit *Taq* DNA polymerase. Glycerol 10% (v/v) was added to the PCR buffer to maximize efficient specific amplification (LU and NÈGRE 1993) and to avoid eventual preferential amplification (WEISSENSTEINER and LANCHBURY 1996). Cycling parameters for PCR were 35 cycles of denaturation (94° for 40 sec) and annealing (55° for 30 sec) according to GÉNÉTHON. An elongation step (72° for 2 min) completed the process after the last annealing step. All reactions were performed in GeneAmp PCR system 2400 thermocyclers (Perkin-Elmer, Norwalk, CT). The amplified products were

resolved on horizontal ultrathin-layer polyacrylamide gel electrophoresis, under nondenaturing conditions, using a discontinuous borate/formate buffer system according to HAAS *et al.* (1994), with some modifications. DNA was visualized with silver staining (BUDOWLE *et al.* 1991). The resolution of the electrophoretic system allowed precise determination of allele lengths, ranging in size from 72 to 231 bp (Table 1), and could even ultimately distinguish single-nucleotide differences between DNA bands. The size of the alleles was estimated using side-by-side comparisons with genotypes determined in one Centre d'Étude du Polymorphisme Humain (CEPH) individual (134702), and allelic ladders were constructed by mixing amplified samples of previously validated genotypes. A segregation analysis was performed for each marker in those families to ensure reliable identification of genotypes that might otherwise be compromised by the occurrence of shadow bands. In addition, reproduction of specific pattern bands of heterozygotes was also performed by amplification of pooled DNA samples for the corresponding homozygotes. The results obtained by means of this approach agreed with the results obtained by inheritance analysis and confirmed the specific band patterns of each genotype. Most of the samples were run on at least two different gels for each locus and were scored independently by two readers to check for concordance.

Estimation of haplotype frequencies and deviations from Hardy-Weinberg proportions: Maximum-likelihood estimates of two-locus haplotype frequencies were obtained from genotype data (405 individuals), using an expectation-maximization (EM) algorithm (DEMPSTER *et al.* 1977; SLATKIN and EXCOFFIER 1996). To ensure that the global maximum-likelihood estimate was found, the EM algorithm was run using 100 different random initial conditions. The EM algorithm is based on the assumption that genotype frequencies at each locus are in Hardy-Weinberg proportions (HWP). Deviations from HWP for each locus, in terms of the excess of homozygotes, were estimated with the *f*-statistic, with significance tested using the square of the ratio of the estimate to its standard error for $f = 0$, which follows a chi-square (χ^2) distribution with 1 d.f (CURIE-COHEN 1982; ROBERTSON and HILL 1984). Estimates of *f* and its significance were also computed for those different alleles at each locus.

Estimation of gametic disequilibrium: Two-locus disequilibrium was studied at two different levels. One analysis estimated the disequilibrium for each pair of alleles or haplotype. The second involved an overall disequilibrium analysis, condensing the information of disequilibrium between all alleles at the two loci. Two-locus haplotypes with singletons or alleles occurring once in the sample were excluded from the disequilibrium analysis.

Let us consider two loci A and B, having A_i ($i = 1, \ldots, k$) and B_j ($j = 1, \ldots, l$) alleles, respectively. Let p_i and q_j be the frequencies of alleles i and j, respectively, and X_{ij} be the relative frequency of the haplotype A_iB_j. Disequilibrium for each possible haplotype A_iB_j may be considered separately by collapsing the data into A_i *vs.* not A_i ($\bar{A}_i$) at the A locus and B_j *vs.* not B_j ($\bar{B}_j$) at the B locus. In this way, the full array of possible two-locus haplotypes is partitioned into $k \times l$ separate 2×2 contingency tables (WEIR and COCKERHAM 1978; KARLIN and PIAZZA 1981). It is easy to verify that the frequencies of the four resulting "haplotype" classes A_iB_j, $A_iB_{\bar{j}}$, $A_{\bar{i}}B_j$, and $A_{\bar{i}}B_{\bar{j}}$ become X_{ij}, $p_i - X_{ij}$, $q_j - X_{ij}$, and $1 - p_i - q_j + X_{ij}$, respectively. We consider A_iB_j and $A_{\bar{i}}B_{\bar{j}}$ to be the haplotype classes in coupling, $A_iB_{\bar{j}}$ and $A_{\bar{i}}B_j$ to be the repulsion classes. The strength of disequilibrium was analyzed in terms of D' coefficients because of their useful properties (ZAPATA and VISEDO 1995; ZAPATA and ALVAREZ 1997b; ZAPATA 2000). The strength of disequilibrium between pairs of alleles at the two loci was

Locus	Chromosomal location	PCR primers (5′ to 3′)	Range of alleles (bp)
D11S4177	11p15.5	GACGGTGAACCCAGTCATT TGTCAGCACAGAGGCAGAGT	177–217
D11S1318	11p15.5	CCCGTATGGCAACAGG TGTGCATGTNCATGAGTG	121–145
D11S4146	11p15.5	AACACGAGGTTAAGCAGAG GAATGAAGAATTTTCCAAACTAC	193–219
D11S4124	11p15.5	GAACACAATCCCATTTATATTAGC TCAGCCTCACAAAGTGC	163–181
D11S4181	11p15.5	GGGCACCTGTAATCCCA GAACTGAGACCAAGAACATTATTCC	193–231
D11S1323	11p15.5	TGCTGCTTAGAATGAGTAGATGTC CTCTATGAAGTTGGAGTCTAGGTTG	197–213
D11S1760	11p15.5	GATCTCAAGTGTTTCCCCAC AAACGATGTCTGTCCACTCA	72–108
D11S1331	11p15.4	GCTGCTTCCATGAGAGGATACTG GCAGAGCCCTTTGCAGTCTT	183–217
D11S4188	11p15.3	TAACACCATCAAATAGACCAGG TTTTATCCAGAAGGCAGCA	101–133
D11S909	11p15.3	GATATAACACCAAAAGCGCG GGTATTCTTACAGCACAAAAGTTCT	95–131
D11S4121	11p15.2	ACCTCCCACCCCAGAT TCAGTTGTTCTCAGTGTAGTTTATG	92–124
D11S926	11p15.2	TGAGTGTGATGTATATGCTCATT CAAGAGGTCAATTCTGGTGT	131–145

measured by $D'_{ij} = D_{ij}/D_{max}$, where $D_{ij} = X_{ij} - p_i q_j$ and $D_{max} = \min[p_i(1 - q_j), (1 - p_i)q_j]$ when $D_{ij} > 0$ or $D_{max} = \min[p_i q_j, (1 - p_i)(1 - q_j)]$ when $D_{ij} < 0$ (LEWONTIN 1964; WEIR and COCKERHAM 1978; HEDRICK 1987). The D'_{ij} coefficient potentially ranges from -1 to $+1$ (ZAPATA 2000). A measure of the strength of overall disequilibrium between all alleles at two loci is $D' = \Sigma_{i=1}^{k}\Sigma_{j=1}^{l} p_i q_j |D'_{ij}|$, which makes use of the absolute values of D'_{ij} weighted by the frequencies of the haplotypes expected at gametic equilibrium (HEDRICK 1987). It was recently shown that the range of the D' measure of overall disequilibrium is only slightly dependent on the polymorphisms at the loci, varying from 0 to a maximum value close or equal to 1 (ZAPATA 2000).

The null hypothesis of gametic equilibrium for each pair of alleles ($D_{ij} = 0$) was tested by $\chi^2_{ij} = nD^2_{ij}/p_i(1 - p_i)q_j(1 - q_j)$, which approximates a χ^2 distribution with 1 d.f., where n is the number of individuals sampled (WEIR and COCKERHAM 1978; WEIR 1979). We used the χ^2 test with Yates's correction to avoid spurious rejection of the null hypothesis when expectations are too small (EVERITT 1997). The significance of overall disequilibrium between pairs of loci was tested using likelihood-ratio statistics that, for large sample sizes, have χ^2 distributions, with $(k - 1)(l - 1)$ d.f. (SLATKIN and EXCOFFIER 1996). These tests compare the likelihood of the observed phenotypes based on the estimated haplotype frequencies with that based on the expected haplotype frequencies assuming no association. The significance of the observed likelihood ratio was found by computing the null distribution of this ratio under the hypothesis of gametic equilibrium, using a permutation procedure (SLATKIN and EXCOFFIER 1996). We used a sample of 16,000 permutations for each locus pair (GUO and THOMPSON 1992). These calculations were carried out with Arlequin software (SCHNEIDER *et al.* 2000).

Estimation of disequilibrium by sign: Overall disequilibrium

includes both positive and negative interallelic gametic associations. It was demonstrated that the statistical power of the chi-square test, for detecting disequilibrium between two loci with two alleles each, depends strongly on the sign of disequilibrium (BROWN 1975; THOMPSON *et al.* 1988; ZAPATA and ALVAREZ 1993). Thus, the power of the chi-square test is higher for detecting positive than negative disequilibrium when coupling haplotype classes contain the most or the least frequent alleles. On the other hand, it was recently found that for the two-allele, two-locus case, the sampling variance of D' estimates is smaller for positive than for negative disequilibrium of the same intensity, if haplotypes with the more extreme expected values are considered to be the coupling classes and allele frequencies at the two loci are not 0.5, under otherwise equivalent conditions (ZAPATA *et al.* 1997). Differences between positive and negative disequilibria in terms of statistical power and sampling variance become considerable when disequilibrium is not intense and allele frequencies are extreme.

The aforementioned observations can be extended to multiallelic systems because significance of each interallelic association by χ^2_{ij}-statistic and its intensity by D'_{ij} coefficient are obtained by collapsing the data to a 2×2 contingency table, which has the same framework as those systems of two loci with two alleles each. For multiallelic systems, it is easy to verify that if $p_i < 0.5$ and $q_j < 0.5$, haplotypes involving the most or the least frequent "alleles" turn out to be the coupling classes (A_iB_j and A_iB_j). If so, positive interallelic associations can be more easily detected by the chi-square tests. In addition, they provide a more accurate estimation of the strength of disequilibrium because positive D'_{ij} estimates will have smaller sampling variance than negative D'_{ij} ones of the same intensity under otherwise equivalent conditions. The reverse is true when repulsion classes (A_iB_j, A_iB_j) contain the most or the least frequent alleles. The significance of each positive (or

negative) interallelic association was determined by χ^2_{ij}-statistic (one sided) with Yates' correction.

We derived two different measures of the strength of global disequilibrium, depending on the sign of the interallelic associations, which are defined as

$$D'(+) = \sum_i \sum_j p_i q_j(+) D'_{ij}(+) / \sum_i \sum_j p_i q_j(+)$$

$$D'(-) = \sum_i \sum_j p_i q_j(-) D'_{ij}(-) / \sum_i \sum_j p_i q_j(-),$$

where $p_i q_j(+)$ and $p_i q_j(-)$ are the expected frequencies of the haplotypes with positive $[D'_{ij}(+)]$ and negative $[D'_{ij}(-)]$ association, respectively. The range of $D'(+)$ and $D'(-)$ also varies from 0 to a maximum value close or equal to 1 (C. ZAPATA, unpublished results).

Significance of overall positive gametic disequilibrium was tested by means of standard simultaneous-inference statistical procedures. Thus, the null hypothesis of overall gametic equilibrium specifies that there is no disequilibrium for any pair of alleles (*i.e.*, $D_{ij} = 0$) and was tested against the alternative $D_{ij} > 0$ from the individual χ^2_{ij} (one sided) with Yates' correction. Because of the large number of $D_{ij} > 0$ involved at each locus pair, it is likely that at least one χ^2_{ij} would be nominally significant, even if there were no real disequilibrium. The usual experimental error rate of 0.05 for multiple comparisons was controlled using the Bonferroni method (HOLM 1979; RICE 1989; ZAPATA AND ALVAREZ 1993). In a similar way, significance of negative overall gametic disequilibrium can be tested.

Relationship of disequilibrium with frequency of recombination and physical distance: The correlation between recombination frequency (or physical distance) and the strength of association between loci was investigated using the Pearson product-moment correlation coefficient (r) and Kendall's nonparametric coefficient of rank correlation (τ; SOKAL and ROHLF 1995). The significance of the relationship was assessed using Mantel's matrix-comparison test, which does not require that matrix elements be independent (MANTEL 1967; SMOUSE *et al.* 1986). Recombination frequencies (sex averaged) for all possible pairs of the 12 loci were obtained from GÉNÉTHON, using KOSAMBI's (1944) map function, which assumes crossover interference. Available recombination frequencies from GÉNÉTHON for other intermediate markers, between our pairs of loci, were used for calculations. Physical distances between all pairs of microsatellites were obtained from LDB.

RESULTS

Polymorphisms and Hardy-Weinberg proportions: A total of 405 individuals were scored for their genotype at 12 microsatellite loci (CA repeats) distributed along 11p15 (Figure 1). The allele sizes and their frequency distributions at the 12 loci are shown in Figure 2. The distributions of allele frequencies vary considerably from locus to locus and appear to be bimodal at some loci. In addition, alleles of extreme size at the loci generally occur at very low frequencies. Levels of polymorphism at the 12 loci are shown in Table 2. The number of alleles per locus ranged from 8 (D11S926) to 15 (D11S1331) and averaged 12. An analysis of individual loci shows that the number of alleles is always higher in our sample than in that reported by GÉNÉTHON from a smaller sample size (data not shown). Of the 142 alleles, 69 (49%) had a frequency ≤3%. This high
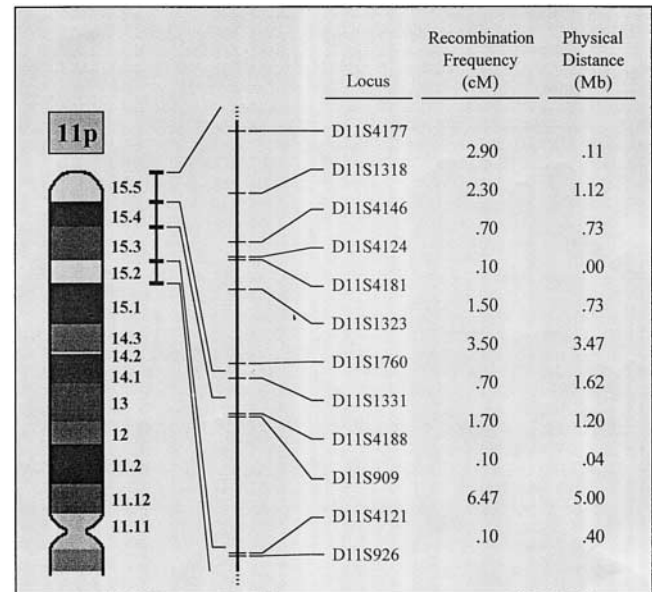

FIGURE 1.—Ideogram of human chromosome 11p with relative positions of microsatellite loci analyzed at 11p15. Sex-averaged recombination frequencies and physical distances between adjacent markers are shown.

proportion of alleles occurring at low frequency distributes quite uniformly over loci. All systems are highly polymorphic, and unbiased expected allelic diversity estimates range from 0.62 (D11S1323) to 0.85 (D11S-4121), with an average (±SE) of 0.75 ± 0.02. Overall, the distributions of allele frequencies and the levels of polymorphism found in our population are comparable to those reported in other studies of dinucleotide repeat markers (VALDES *et al.* 1993; DEKA *et al.* 1995; LAAN and PÄÄBO 1997; WEI *et al.* 1999).

From our genotypic data, it is not possible to distinguish the coupling and repulsion double heterozygotes, and estimates of two-locus haplotype frequencies were obtained using the EM algorithm. We examined whether there is evidence of deviations from HWP, as any departure from these conditions may lead to erroneous estimates of haplotype frequencies (FALLIN and SCHORK 2000). It is noteworthy that we used a large sample size, which increases the power of the tests and provides more opportunities for detecting deviations from HWP. Small expected frequencies of low-frequency alleles can lead to spurious rejection of the null hypothesis. Because of the many alleles of low frequency in our sample, we examined deviations from HWP in two data sets. One included all alleles, and the other included alleles with frequencies >3%. Table 2 shows that the agreement between the observed and the expected frequencies of heterozygotes is fairly good at any one locus. There are only two statistically significant differences (D11S1331 and D11S4121) among the 24 comparisons (both data sets). Analysis of deviations from HWP for each allele revealed that significant devia-
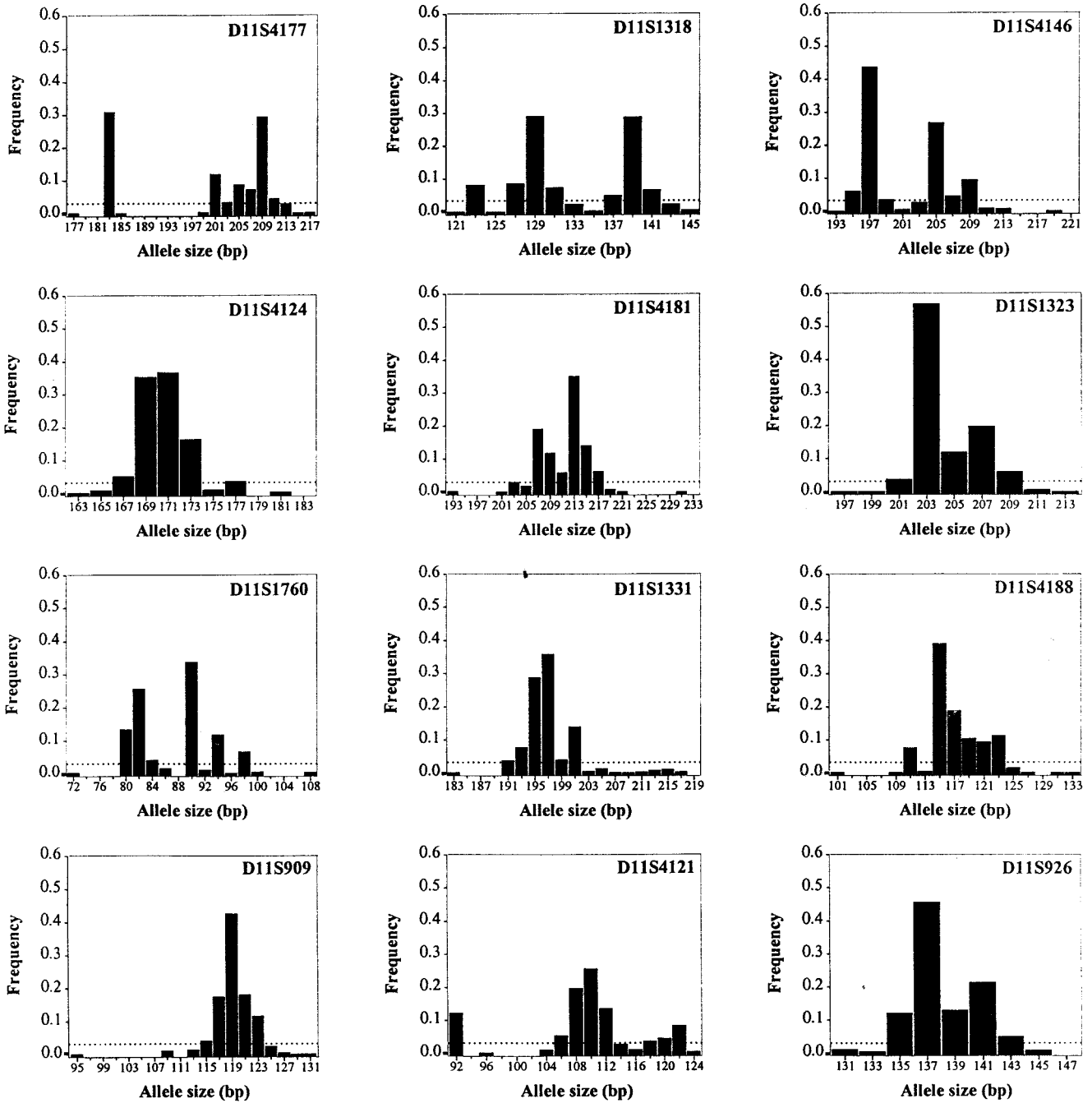
FIGURE 2.—Distribution of allele size frequencies at the 12 microsatellite loci.

tions arise only for alleles occurring at very low frequencies. Thus, of the 142 alleles, only 4, occurring at frequencies of 0.4% (D11S1331), 0.6% (D11S4121), 0.9% (D11S1331), and 2% (D11S909), were significant. No significant deviations from HWP were detected for any locus at common allelic frequencies (>3%). The number of significant cases is smaller than that expected from type I error, even without correction for multiple tests. In addition, only small and nonsystematic departures from HWP were detected by the f-statistic. It appears, therefore, that genotype data from the loci were generally consistent with their HWP within this population. This observation guarantees a reliable estimation of the haplotype frequencies estimated with the EM algorithm.

Interallelic disequilibrium: A total of 6797 two-locus haplotypes can be evaluated between the alleles of the 12 microsatellite loci studied when only alleles occurring more than once in our sample (122 alleles) are considered. Figure 3a shows the distribution of the magnitude of interallelic disequilibrium measured by D'_{ij} across all the pairwise comparisons. It can be seen

TABLE 2

Variability and deviations from HWP

Locus	n	No. of alleles	Observed heterozygosity	Expected heterozygosity	$f \pm$ SE	P value
D11S4177	405	13	0.795	0.788	0.005 ± 0.020	0.734
	375	7	0.781	0.772	0.001 ± 0.021	0.978
D11S1318	405	13	0.817	0.805	-0.001 ± 0.011	0.943
	359	7	0.797	0.779	-0.012 ± 0.020	0.588
D11S4146	405	12	0.736	0.720	-0.004 ± 0.002	0.793
	363	6	0.705	0.689	0.001 ± 0.023	0.982
D11S4124	405	9	0.723	0.710	0.001 ± 0.018	0.946
	383	5	0.708	0.694	0.008 ± 0.027	0.741
D11S4181	405	13	0.770	0.797	0.012 ± 0.024	0.414
	373	7	0.751	0.782	0.030 ± 0.024	0.153
D11S1323	405	9	0.625	0.617	0.001 ± 0.018	0.941
	396	5	0.616	0.611	0.004 ± 0.026	0.866
D11S1760	405	12	0.807	0.782	-0.015 ± 0.000	0.310
	370	6	0.789	0.763	-0.027 ± 0.020	0.251
D11S1331	405	15	0.775	0.758	0.061 ± 0.048	0.000
	367	6	0.757	0.728	-0.014 ± 0.021	0.542
D11S4188	405	13	0.765	0.773	0.009 ± 0.027	0.519
	382	6	0.751	0.758	0.027 ± 0.025	0.235
D11S909	405	12	0.723	0.738	0.024 ± 0.039	0.113
	361	5	0.695	0.709	0.028 ± 0.029	0.293
D11S4121	405	13	0.837	0.846	0.045 ± 0.028	0.002
	358	8	0.821	0.827	0.016 ± 0.022	0.412
D11S926	405	8	0.696	0.711	-0.002 ± 0.017	0.914
	385	5	0.681	0.693	0.001 ± 0.025	0.978
Mean $\pm$ SE	405 ± 0	12 ± 1	0.756 ± 0.017	0.754 ± 0.017	0.011 ± 0.006	
	373 ± 3	6 ± 0	0.738 ± 0.017	0.734 ± 0.017	0.005 ± 0.005	

Top lines, all the alleles; bottom lines, alleles at frequency >3%. Unbiased expected heterozygosities were calculated as $h = 2n \, (1 - \Sigma_{i=1}^{k} p_i^2)/2n - 1$, where p_i is the frequency of the allele i in a sample of n genotypes drawn from the population (NEI 1987).

that D'_{ij} values span the whole range of possible values (from -1 to $+1$), but the proportion of haplotypes with negative deviations from random association is considerably higher than that with positive deviations. Of the 6797 haplotypes, 4287 (63%) showed negative deviations, whereas 2510 (37%) gave positive ones. This discrepancy is due mainly to the high proportion of D'_{ij} values that are -1. Thus, a total of 2839 of the 6797 (42%) D'_{ij} are -1, indicating the absence of many of the possible haplotypes. In contrast, only 0.6% (42/6797) of D'_{ij} are found to be $+1$. On the other hand, only 364 of 6797 (5%) pairs of alleles are in significant disequilibrium ($P < 0.05$) by the χ^2 test (two sided).

The distribution of D'_{ij} values can be more easily understood by taking into account that all the alleles, except one (at the D11S1323 locus), had a frequency <0.5 in the sample (Figure 2). This means that most coupling haplotype classes in the 2×2 contingency tables used to obtain the D'_{ij} estimates involve the most or the least frequent alleles (see MATERIALS AND METHODS). This structure of the haplotype classes has two important consequences. First, sampling error would be sufficient to explain the high proportion of $D'_{ij} = -1$, because those haplotypes carrying two alleles at very low frequen-

cies are unlikely to be detected in the sample. Indeed, the proportion of $D'_{ij} = -1$ decreased dramatically (from 42 to 8%) when rarer alleles (frequency $\leq 3\%$) were removed from the analysis (Figure 3b). Second, the sampling variance of negative interallelic associations is expected to be greater than that of positive ones, as described in MATERIALS AND METHODS. Therefore, negative associations will tend to overestimate disequilibrium intensity. In addition, negative associations will be less likely to be detected by significance tests in comparison to positive ones. These predictions are well supported by the analysis of our data. Thus, mean values of $D'_{ij}(-)$ and $D'_{ij}(+)$ across loci were -0.756 ± 0.006 and 0.164 ± 0.004, respectively. These differences are not completely explained by the high occurrence of both $D'_{ij} = -1$ and alleles at a low frequency. If both $D'_{ij} = \pm 1$ and rarer alleles ($\leq 3\%$) are ignored, mean values of $D'_{ij}(-)$ and $D'_{ij}(+)$ were -0.258 ± 0.007 and 0.082 ± 0.002, respectively. On the other hand, percentages of significant $D'_{ij}(-)$ and $D'_{ij}(+)$ values calculated by the χ^2 test with Yates' correction (one sided) were 4% (160/4287) and 18% (456/2510), respectively.

Positive interallelic disequilibrium: The aforementioned observations make it advisable to confine our
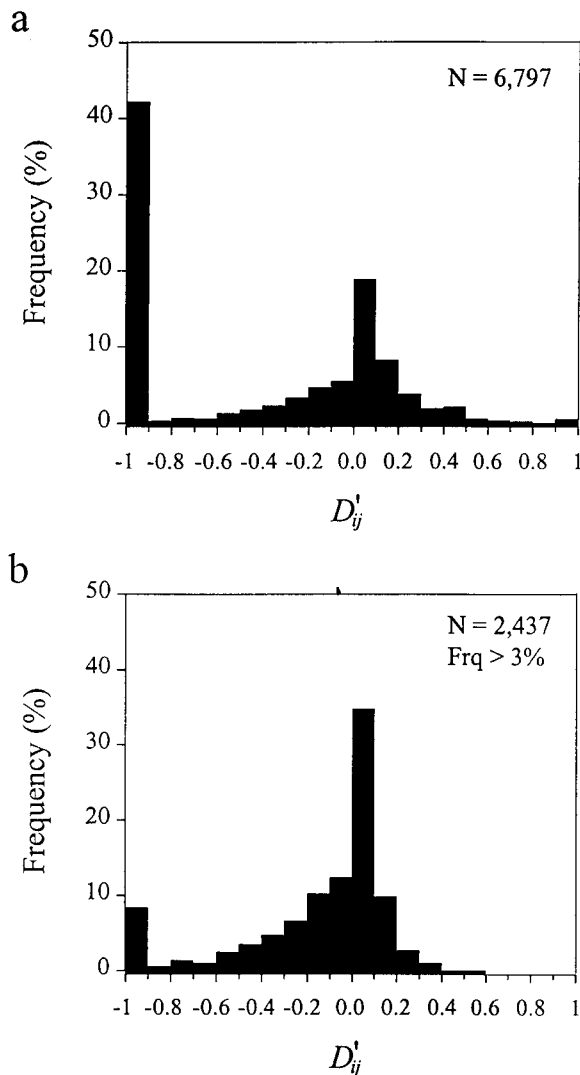
FIGURE 3.—Frequency distribution of D'_{ij} estimates across pairs of loci. (a) All alleles. (b) Alleles at frequencies >3%. *N* is the total number of two-locus haplotypes.

and size of the alleles allows us to exclude that the detected associations can be explained merely by type I error.

Overall disequilibrium: Pairs of loci in significant overall disequilibrium by likelihood-ratio χ^2 tests, as well as their intensities by D' disequilibrium coefficient, are given in Table 3. Only 4 of 66 pairwise comparisons (6%) were significantly different from 0 ($P < 0.05$). These results could be explained by type I error and seem to suggest that little global disequilibrium occurs between microsatellites located at 11p15. Nevertheless, likelihood-ratio χ^2 tests consider both positive as well as negative interallelic disequilibrium, which might diminish statistical power.

Positive overall disequilibrium: Overall disequilibrium, analyzed by considering only those positive interallelic associations, presents quite a different picture. Pairs of loci in significant overall positive disequilibrium are shown in Table 4. Also shown is the lowest probability of those Yates' correction χ^2_{ij} values for each locus pair, which allows us to reject the null hypothesis of no global disequilibrium (*i.e.*, all $D_{ij} = 0$) by using the Bonferroni criterion. It is noteworthy that a substantial number of pairs of loci are now in significant overall disequilibrium, despite the fact that the Bonferroni correction is believed to be very conservative (*i.e.*, maintains the null hypothesis too frequently; see ROTHMAN 1990; ZERBA *et al.* 1991). Thus, 29 of 66 pairwise comparisons (44%) are significant ($P < 0.05$). This proportion is much too large to be attributed to chance alone. All 12 loci that we studied were involved in the detected disequilibria, and no particular locus was preferentially involved in disequilibria. Every one of the 12 loci was involved in 4–6 significant pairwise comparisons, with the exception of D11S1323, which appeared in only 3 significant pairs of loci (Table 4). D' (+) estimates over significant locus pairs ranged from 0.050 to 0.138 and averaged 0.094 ± 0.004 (Table 4). Therefore, it seems that the global disequilibrium across pairs of loci fluctuates within a narrow range of values of weak intensity.

Disequilibrium depending on the allele frequency: The objective of this analysis was to ask whether the amount of positive disequilibrium is related to the allele frequency at loci. To investigate this, we stratified the sample of two-locus haplotypes (2510) into RR, RC, and CC, where R and C indicate rarer (frequency $\leq 3\%$) and more common (frequency $>3\%$) alleles, respectively. Then the amount of disequilibrium was analyzed separately for each haplotype data set. The results are shown in Table 5. It can be seen that disequilibrium is not randomly distributed with respect to these three haplotype data sets. The percentage of significant disequilibria is considerably higher in RR haplotypes (61%) than in RC (17%) and CC (14%) haplotypes. Differences of disequilibrium between RR haplotypes and any of the other haplotype classes are highly significant ($P \leqslant 0.001$) based on χ^2 tests for 2 × 2 contingency tables

attention to only interallelic disequilibrium with positive sign. Cases of significant positive interallelic disequilibrium (456/2510) are distributed across all 66 possible pairs of loci (Figure 4). However, percentages of significant disequilibrium between pairs of alleles, as well as their intensities, were not distributed uniformly across locus pairs. Percentages ranged from 6% (D11S4188 × D11S926) to 42% (D11S4124 × D11S1760) and averaged $18 \pm 0.9\%$. The means of $D'_{ij}(+)$ over significant comparisons ranged from 0.171 ± 0.066 (D11S4188 × D11S926) to 0.565 ± 0.251 (D11S1323 × D11S1331) and averaged 0.345 ± 0.011. It must be noted that a substantial proportion of alleles are in moderate or strong disequilibrium (Figure 5). By way of illustration, 144 of 456 (32%) significant positive nonrandom associations exhibited $D'_{ij}(+)$ values >0.40. As is shown in detail below, the occurrence of patterns of significant interallelic associations dependent on the frequency
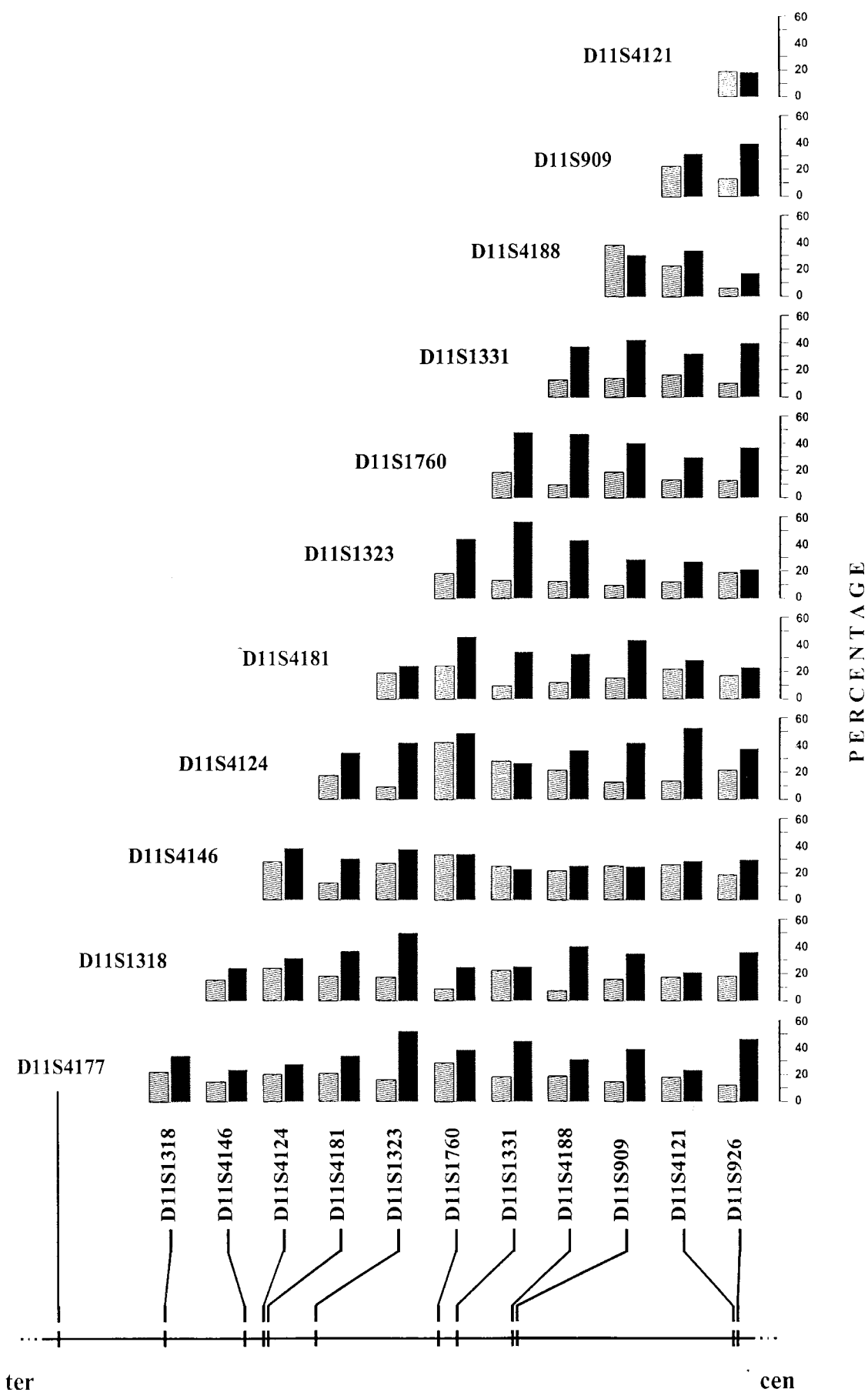
FIGURE 4.—Amount of positive interallelic associations within each locus pair. Percentages of pairs of alleles in significant disequilibrium (hatched bars), along with their mean intensities of association, as measured by the $D'_{ij}(+)$ coefficient (solid bars), are shown.
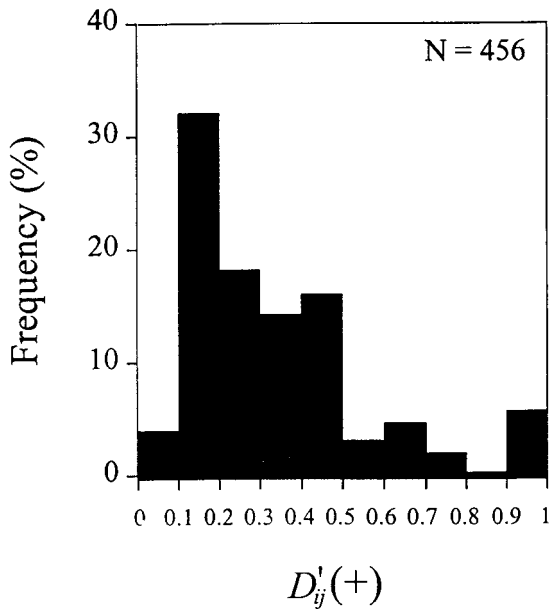
FIGURE 5.—Frequency distribution of the strength of positive association between pairs of alleles, as measured by the $D'_{ij}(+)$ coefficient, along 11p15. N is the total number of two-locus haplotypes in significant positive disequilibrium.

(1 d.f.). In addition, although CC haplotypes show somewhat less disequilibrium than RC haplotypes, differences are still significant ($P < 0.05$).

Overall, these results provide evidence that the frequency of significant disequilibrium depends strongly on the allele frequencies. This cannot be exclusively attributable to differences in statistical power, because the power of the χ^2 test is expected to decline as allele frequencies become extreme (BROWN 1975; ZAPATA and ALVAREZ 1992, 1993). Therefore, it is noteworthy that we found more disequilibrium in haplotypes bearing rarer alleles. However, it could also be argued that 2×2 contingency tables with low expected counts may lead to too liberal χ^2 tests (*i.e.*, that reject the null hypothesis too frequently). If so, some disequilibrium occurring between rarer alleles (frequency $\leq 3\%$) could be artifactual. However, the χ^2 test is robust and generally provides actual significance levels close to or smaller than the usual nominal level of 0.05 (UPTON 1982; D'AGOSTINO *et al.* 1988; ZAPATA and ALVAREZ 1997a). This conclusion applies generally for different sample sizes (30–500) and combinations of allele frequencies

TABLE 3

Significant overall nonrandom associations between loci

Locus pair	cM	Mb	D'	P value
D11S4177 × D11S1318	2.90	0.11	0.237	0.0326
D11S4146 × D11S1760	5.78	4.92	0.189	0.0174
D11S4124 × D11S1760	5.09	4.19	0.229	0.0000
D11S4188 × D11S909	0.10	0.04	0.282	0.0000

(from 0.5 to 0.9) at two loci (ZAPATA and ALVAREZ 1997a). Nevertheless, to minimize the effects of possible false positives, we used the χ^2 test with Yates' correction, which is commonly considered to be conservative (CAMILLI and HOPKINS 1978; UPTON 1982; D'AGOSTINO *et al.* 1988). In support of this, we obtained empirical type I errors (*i.e.*, rejecting a true null hypothesis) of the χ^2 with Yates' correction by Monte Carlo simulations under the conditions of our experiment. We constructed populations for a system of two loci with two alleles each, with genotype frequencies following HWP and haplotype frequencies at gametic equilibrium, for given allelic frequencies. Then, 10,000 replicate random genotype samples of size 405 were drawn from each of these populations with replacement. Haplotype frequencies were determined from each of the 10,000 random genotype samples according to HILL (1974). The χ^2 test with Yates' correction (one sided) was computed for each sample, and type I error probabilities were calculated as the proportion of significant positive disequilibria at the $\alpha = 0.05$ nominal significance level. Simulation results show that those empirical type I errors of the χ^2 test with Yates' correction are always less than the designed type I error of $\alpha = 0.05$, especially when allele frequencies are extreme (Table 6). Therefore, it can be concluded that those significant interallelic disequilibria detected at 11p15 are genuine and not an artifact resulting from the lack of rigor in the tests for significance. On the contrary, it is likely that disequilibrium between rarer alleles is being underestimated.

Disequilibrium depending on the allele size: We examined whether the amount of positive disequilibrium is related to allele size. Figure 2 shows that all alleles at the upper and lower extremes of the size distributions (except allele 92 at locus D11S4121) are found at low frequencies ($\leq 3\%$). Accordingly, those alleles in the left and right tails of the size distributions that were present at a frequency $\leq 3\%$ and that were separated by alleles more frequent than 3% were designated extreme-size (E) alleles, while the remaining alleles were designed intermediate-size (I) alleles. For example, only alleles 121, 143, and 145 at the D11S1318 locus are E alleles under this criterion (see Figure 2). Having established the categories of allele sizes, we stratified two-locus haplotypes as EE, EI, and II, and disequilibrium was analyzed for each haplotype class separately (Table 7). The results obtained show that disequilibrium is very heterogeneous across these haplotype data sets. Thus, EE haplotypes exhibit much greater relative amounts of disequilibrium than do the remaining haplotypes (Table 7). We found that disequilibrium occurs in 74% of EE haplotypes, but in only 21 and 15% of EI and II haplotypes, respectively. These differences in the amount of disequilibrium become highly significant ($P < 0.001$) when evaluated by χ^2 tests in 2×2 contingency tables. In addition, this pattern of disequilibrium dependent on allele size cannot be explained by varia-

TABLE 4

Significant positive overall nonrandom associations between loci

Locus pair	cM	Mb	$D'(+)$	α'	P value
D11S4177 × D11S1318	2.90	0.11	0.110	0.001220	0.001099
D11S4177 × D11S4181	5.97	1.74	0.050	0.001042	0.001003
D11S4177 × D11S4188	13.10	8.75	0.106	0.001563	0.000509
D11S4177 × D11S4121	19.01	13.78	0.085	0.001000	0.000257
D11S4177 × D11S926	19.09	13.38	0.065	0.001515	0.000038
D11S1318 × D11S4146	2.30	1.12	0.112	0.001064	0.000068
D11S1318 × D11S1323	4.59	2.58	0.127	0.001724	0.000247
D11S1318 × D11S1760	8.04	6.04	0.057	0.001087	0.000248
D11S1318 × D11S909	10.45	8.89	0.083	0.001316	0.000453
D11S1318 × D11S926	16.56	13.49	0.124	0.001515	0.001501
D11S4146 × D11S4124	0.70	0.73	0.103	0.001563	0.000010
D11S4146 × D11S4181	0.80	0.73	0.097	0.001042	0.000038
D11S4146 × D11S4188	8.13	7.74	0.119	0.001351	0.000018
D11S4146 × D11S4121	14.39	12.77	0.100	0.001087	0.001053
D11S4146 × D11S926	14.48	12.37	0.115	0.001563	0.000174
D11S4124 × D11S1760	5.09	4.19	0.120	0.001613	0.000068
D11S4124 × D11S1331	5.78	5.81	0.077	0.001282	0.001141
D11S4124 × D11S909	7.55	7.04	0.097	0.001563	0.001545
D11S4181 × D11S1323	1.50	0.73	0.094	0.001923	0.000945
D11S4181 × D11S1760	4.99	4.19	0.091	0.001111	0.000001
D11S4181 × D11S909	7.45	7.04	0.063	0.001282	0.000257
D11S4181 × D11S4121	13.65	12.04	0.074	0.000909	0.000851
D11S1323 × D11S1331	4.20	5.08	0.079	0.001667	0.001385
D11S1760 × D11S4188	2.40	2.81	0.101	0.001190	0.000508
D11S1760 × D11S909	2.50	2.85	0.119	0.001163	0.000886
D11S1331 × D11S4121	8.23	6.23	0.085	0.000909	0.000001
D11S1331 × D11S926	8.32	5.83	0.086	0.001250	0.001197
D11S4188 × D11S909	0.10	0.04	0.138	0.001724	0.000003
D11S909 × D11S926	6.56	4.59	0.063	0.001667	0.000166
Mean ± SE	7.55 ± 1.02	5.96 ± 0.82	0.094 ± 0.004		

Significant ($P < 0.05$) nonrandom associations by Bonferroni criterion. P value is the smallest probability of the χ^2_{ij} values with Yates' correction at each locus pair. α' is the level of significance (0.05) corrected by Bonferroni.

tions in recombination, since the mean of the recombination frequency within each haplotype class is very similar ($\sim$7 cM). To minimize differences in statistical power among comparisons that arise from variable polymorphisms associated with size classes, the data were reanalyzed excluding two-locus haplotypes with alleles at frequencies >3%. The results of this analysis, however, do not qualitatively change the conclusions, which retain the same aforementioned pattern of size-dependent disequilibrium (Table 7). Therefore, the occurrence of interallelic disequilibrium is not only dependent on the allele frequency but also on the size of the alleles.

Relationship of disequilibrium with frequency of recombination and physical distance: The 12 microsatellites chosen for this study span $\sim$19 cM, or 14 Mb, on chromosome 11p15. Recombination frequencies (in centimorgans) and physical distances (in megabases) for pairwise comparisons in significant overall disequilibrium are shown in Table 4. It can be seen clearly that disequilibrium is not confined to the closest pairs of loci. The largest recombination frequencies between the loci in disequilibrium were 17–19 cM (three pairs), whereas the largest physical distances were 12–14 Mb (six pairs). The recombination frequency and physical distance across significant pairs averaged 7.55 ± 1.02 cM and 5.96 ± 0.82 Mb, respectively. Note that disequilibrium was detected even between D11S4177 and D11S4121, the most distant pair in the set that we studied (19 cM and 14 Mb). We determined whether significant overall nonrandom associations decrease as recombination frequency or physical distance increases. Since loci located more distantly from one another undergo recombination more frequently, they are expected to exhibit less disequilibrium than more closely linked loci if evolutionary agents generating or randomizing nonrandom associations act uniformly across the chromosomal region studied. For the comparison, we classified the number of significant pairwise comparisons according to the same arbitrary criterion used in previous disequilibrium studies (PETERSON *et al.* 1995; LAAN and PÄÄBO 1997). The proportion of significant tests was 4

TABLE 5

Nonrandom associations depending on the allele frequency

Haplotype data set		No. of haplotypes	Interallelic nonrandom associations		
			Significant ($P \leq 0.05$)	Nonsignificant ($P > 0.05$)	% of significant
RR	Set 1	125	76	49	61
RC	Set 2	1202	210	992	17
CC	Set 3	1183	170	1013	14
	Total	2510	456	2054	

$$\chi_1^2 \text{ (set 1 } vs. \text{ set 2)} = 125.73; P \ll 0.001$$
$$\chi_1^2 \text{ (set 1 } vs. \text{ set 3)} = 159.60; P \ll 0.001$$
$$\chi_1^2 \text{ (set 2 } vs. \text{ set 3)} = 4.28; P = 0.04$$

R, rare allele (frequency $\leq 3\%$); C, common allele (frequency $>3\%$).

out of 10 for loci that were located 0–2 cM apart and 25 out of 56 for loci that were >2 cM distant. With respect to the physical distance, the number of significant tests was 7 out of 18 for loci located 0–2 Mb and 22 out of 48 for those located >2 Mb distant. The 2×2 contingency tables drawn for the numbers of significant and nonsignificant pairs were not significant for recombination frequency ($\chi^2 = 0.07$, $P = 0.79$) or physical distance ($\chi^2 = 0.26$, $P = 0.61$).

We looked for a negative relationship of the strength of overall and interallelic disequilibrium with the recom-

TABLE 6

Simulated type I error probabilities of the chi-square test (one sided) with Yates' correction for the null hypothesis $D_{ij} = 0$ and the alternative hypothesis $D_{ij} > 0$

p_i	q_j	Type I error probability
0.0025	0.0025	0.0118
	0.0198	0.0229
	0.0296	0.0188
	0.1	0.0217
	0.3	0.0080
	0.5	0.0002
0.0198	0.0198	0.0305
	0.0296	0.0344
	0.1	0.0359
	0.3	0.0395
	0.5	0.0409
0.0296	0.0296	0.0343
	0.1	0.0345
	0.3	0.0387
	0.5	0.0405
0.1	0.1	0.0430
	0.3	0.0459
	0.5	0.0456
0.3	0.3	0.0443
	0.5	0.0430
0.5	0.5	0.0446

Number of genotypes (n) = 405; nominal significance level of 0.05. Each type I error probability was obtained from 10,000 replicate random genotype samples.

bination frequency and physical distance for the 12 microsatellite loci spanning much of the 11p15 chromosome region. Only the pairs of alleles and loci in significant disequilibrium were included in this analysis. A weak but significant negative correlation was found by the Mantel test (one sided) between $D'(+)$ and recombination frequency ($r = -0.226$, $\tau = -0.177$, $P = 0.037$, $n = 29$). In contrast, a negative but not significant correlation was detected with physical distance ($r = -0.179$, $\tau = -0.151$, $P = 0.079$, $n = 29$). On the other hand, $D'_{ij}(+)$ values do not appear to be correlated with recombination frequency ($r = -0.068$, $\tau = -0.041$, $P = 0.098$, $n = 456$) or with physical distance ($r = -0.074$, $\tau = -0.042$, $P = 0.071$, $n = 456$). However, we found significant negative correlations between $D'_{ij}(+)$ and recombination frequency when only alleles present at higher frequencies are considered, and indeed the strength of the correlation showed an increasing trend with allele frequency (*e.g.*, $r = -0.192$, $\tau = -0.095$, $P = 0.019$, $n = 96$ for alleles at a frequency $>6\%$; and $r = -0.284$, $\tau = -0.128$, $P = 0.017$, $n = 50$ for alleles at a frequency $>9\%$). A similar trend was detected between the strength of interallelic disequilibrium and physical distance ($r = -0.198$, $\tau = -0.071$, $P = 0.015$, $n = 96$ for alleles at a frequency $>6\%$; and $r = -0.209$, $\tau = -0.079$, $P = 0.055$, $n = 50$ for alleles at a frequency $>9\%$).

DISCUSSION

This study is the most complete characterization to date of disequilibrium patterns between pairs of microsatellite markers over an extensive (19 cM and 14 Mb) anonymous region of the human genome. We provide, for the first time, estimates of the frequency and strength of both overall and interallelic disequilibria on the basis of 12 microsatellite loci (CA repeats) spanning human chromosome 11p15. We also derived a novel approach for analyzing disequilibrium between pairs of multiallelic loci on the basis of the sign of the interallelic

TABLE 7

Nonrandom associations depending on the allele size

Haplotype data set		No. of haplotypes	Interallelic nonrandom associations		
			Significant ($P \leq 0.05$)	Nonsignificant ($P > 0.05$)	% of significant
EE	Set 1	47 (47)	35 (35)	12 (12)	74 (74)
EI	Set 2	871 (65)	182 (36)	689 (29)	21 (55)
II	Set 3	1592 (13)	239 (5)	1353 (8)	15 (38)
	Total	2510 (125)	456 (76)	2054 (49)	

$$\chi^2_1 \text{ (set 1 } vs. \text{ set 2)} = 70.9 \text{ (4.28)}; P \ll 0.001 \text{ (} = 0.039)$$
$$\chi^2_1 \text{ (set 1 } vs. \text{ set 3)} = 115.9 \text{ (5.94)}; P \ll 0.001 \text{ (} = 0.015)$$
$$\chi^2_1 \text{ (set 2 } vs. \text{ set 3)} = 13.8 \text{ (1.24)}; P = 0.002 \text{ (} = 0.265)$$

E, extreme allele size; I, intermediate allele size. Values for two-locus haplotypes with both alleles at a frequency ≤3% are shown in parentheses.

associations. Our results demonstrate that the statistical power is much greater, for given allelic frequencies, in detecting positive rather than negative interallelic associations, which is in agreement with previous evidence on two-allele systems (Brown 1975; Thompson *et al.* 1988; Zapata and Alvarez 1993). At the same time, positive interallelic associations were found to provide more accurate estimates of the strength of disequilibria between multiallelic systems than negative associations, which greatly overestimate the strength of disequilibrium. Application of this strategy allowed us to discover new and relevant features concerning disequilibria between microsatellite markers across an extensive anonymous region of a human chromosome.

We found that all the loci studied were involved in significant disequilibrium, and disequilibrium was thus distributed over the entire 11p15 region. In total, 44% of pairs of loci were in significant overall disequilibrium, whereas 18% of pairs of alleles deviated significantly from random association. The strength of the significant overall disequilibria lay within a narrow interval of low values [$D'(+)$ from 0.050 to 0.138]. In contrast, the strength of the significant interallelic disequilibria was very heterogeneous, ranging from weak to strong, with a substantial proportion (32%) of pairs showing strong nonrandom association [$D'_{ij}(+) > 0.40$]. This does not contradict the finding that the strength of overall disequilibria across pairs of loci was weak: $D'(+)$ is a measure of overall disequilibrium, which will adopt low values if the proportion of haplotypes at gametic equilibrium outweighs the proportion of haplotypes in disequilibrium.

Overall, the amount of disequilibrium detected in 11p15 is considerable given the high recombination frequencies for loci in disequilibrium (7 cM on average). Disequilibrium was detected even between the most distant locus pair in the set studied (D11S4177 × D11S4121, 19 cM and 14 Mb). Table 8 shows that the percentage of pairs of microsatellites in overall disequilibrium in 11p15 (44%) is much higher (∼3–9 times higher) than previously reported for pairs of dinucleotide microsatellites along extensive anonymous regions of human chromosomes in fast-growing populations (5–14%; Peterson *et al.* 1995; Laan and Pääbo 1997). Unfortunately, there are too many differences in experimental design between our study and these previous studies to assess whether the conflicting findings imply real differences in the frequency of disequilibrium among chromosome regions and/or populations: both the analytical methods used and marker densities were different, the sample size was larger in our study, and haplotypes were inferred in our study from individuals from a population sample, whereas Peterson *et al.* used family-structured data and Laan and Pääbo obtained haplotypes of the X chromosome directly from a population sample of males. Nevertheless, among-study variations in marker density cannot satisfactorily explain the observed differences in frequency of disequilibrium, because the mean recombination distance for pairs of loci showing significant disequilibrium is consistently greater (generally 25–40 times greater) in 11p15 than in the other chromosomes. Our observations also suggest that the previously reported frequencies of overall disequilibrium over extensive human chromosome regions may be underestimates. The usual tests of significance for multiallelic loci, based on both positive and negative deviations from random association, are clearly less powerful. By way of illustration, the percentage of pairs of loci in significant disequilibrium decreased from 44 to 6% in our population when standard likelihood-ratio tests were used for testing the null hypothesis of overall equilibrium. Thus the statistical procedure used can have dramatic effects on outcome. Clearly, this suggests a need to reanalyze the data obtained in previous studies, considering the sign of the interallelic associations.

A variety of factors may cause disequilibrium in populations, such as founder effects, genetic drift, stratification or admixture of populations, selection, and recent

TABLE 8

Comparison of the amount of significant overall nonrandom associations between pairs
of microsatellite loci at different human chromosomes in expanded populations

Population	Chromosome region/cM	Sample size (individuals)	No. of loci	No. of locus pairs	Significant locus pairs	
					%	Mean of recombination distance (cM)
Galicians (present study)[a]	11p15/19.1	405	12	66	44	7.55 ± 1.02
Finns[b]	4/>80	50	32	496	5	5.21 ± 1.72
Finns[c]	Xq13/4.0	80	7	21	10	0.20 ± 0.10
Estonians[c]	Xq13/4.0	45	7	21	5	0.30
Swedes[c]	Xq13/4.0	41	7	21	14	0.20 ± 0.01

[a] For details on the demographic history see SALAS *et al.* (1998).

[b] Data from PETERSON *et al.* (1995).

[c] Data from LAAN and PÄÄBO (1997).

mutations (see Introduction for references). Founder effects appear to be an unlikely explanation for the observed disequilibria in 11p15, as the available genetic evidence on the demographic history of the Galician population is compatible with a population expansion model in Europe during the Upper Paleolithic, and there is no evidence for recent founder effects (SALAS *et al.* 1998). Genetic drift is strongly inhibited in rapidly expanding populations (SLATKIN 1994) and is a highly unlikely cause for the high levels of disequilibrium detected between distant microsatellites. For an effective population size of N_e and an amount of recombination c between two loci, the expected value for drift-generated disequilibrium can be calculated approximately as $E(r^2) \approx 1/(1 + 4N_e c)$ (HILL and ROBERTSON 1968). Considering that the approximate N_e of human populations is in the order of 10,000 (NEI and GRAUR 1984; TAKAHATA and SATTA 1997), and c is 7 cM (on average) for pairs of microsatellites in disequilibrium at 11p15, $E(r^2)$ is only 4×10^{-4}. In contrast, an admixture of genetically differentiated populations generates associations between pairs of loci with different allele frequencies over populations, regardless of whether the locus pairs are or are not linked (NEI and LI 1973; CHAKRABORTY and WEISS 1988; STEPHENS *et al.* 1994). Population subheterogeneity can be detected as a deviation from HWP since the combination of subgroups differing in allele frequencies produces a deficiency of heterozygotes. However, genotype data for the 12 loci considered in our study are in close agreement with HWP, so population subdivision seems to be an unlikely explanation for the observed disequilibria. Although a single generation of random mating generates HWP, historical admixture between populations could result in significantly elevated disequilibria over large genome regions for many generations (CHAKRABORTY and WEISS 1988; STEPHENS *et al.* 1994). Nevertheless, Galicia is a region situated in the northwest corner of the Iberian Peninsula that has remained relatively isolated from the rest of the Peninsula, with high emigration rates throughout the centuries and almost no immigration (SALAS *et al.* 1998). Alternatively, the loci monitored may be directly or indirectly under selection. Many of the uses of microsatellite loci in population biology depend on the assumption that they are neutral, although microsatellites may in fact be neutral markers located within blocks of genes subject to selection. Nevertheless, it seems unlikely that selection can explain the persistence of disequilibrium between distant microsatellites at 11p15.

A more likely explanation for the observed disequilibria may be related to the mutational dynamics of microsatellite loci. The analysis of patterns of disequilibrium dependent on allele frequency and size offers a new approach for interpreting associations between the microsatellites occurring in 11p15 on the basis of their mutational dynamics. An interesting finding of this analysis is that alleles present at low frequency (≤3%), as well as alleles of extreme size, tend to be much more in disequilibrium than the remaining alleles. This can be interpreted as a consequence of the high mutation rate (say 10^{-3}) and complex mutational dynamics of microsatellite loci (see Introduction for references). Many of the disequilibria detected between rarer alleles might be the result of relatively recent mutational events. Low-frequency alleles are more likely to have arisen recently in the population (by the introduction of new mutations) than alleles of moderate frequency, because time is necessary for new alleles to become common (KRUGLYAK 1999). Undoubtedly, the decay of disequilibrium between distant loci is mainly dependent on the fraction of recombination and on the number of generations passed since a single mutation occurred. Disequilibrium caused by recently introduced mutations is short-lived, and its dilution might still be in progress.

With regard to the higher frequency of disequilibrium in haplotypes bearing alleles of extreme size, previous studies already reported evidence for a relationship between disequilibrium and allele size for tightly linked microsatellite loci in human populations. A positive cor-

relation between the sizes of two alleles in each haplotype was reported for two pairs of microsatellites separated by 212 and 7 kb (SHERRINGTON *et al.* 1991; PENA *et al.* 1994). PENA *et al.* (1994) suggested that this relationship could be explained if the mutational mechanisms controlling the variability of repeat numbers in tightly linked microsatellite loci are cooperative. However, this interpretation, based as it is on some form of regional mutational mechanism, cannot explain our observations for distant loci. In addition, we found no evidence for positive allele size association; *i.e.*, the largest and smallest alleles at one locus were not preferentially associated with the largest and smallest alleles at the other locus, respectively. An alternative explanation for our data might again be based on the mutational dynamics of microsatellite loci: specifically, that selective or nonselective constraints on the number of repeats dictate higher rates of turnover for alleles of extreme size. This scenario would predict that new allele variants of extreme size are continuously arising in disequilibrium, which is consistent with the observation that these alleles occur at very low frequencies in our population as in other populations (VALDES *et al.* 1993; SCHLÖTTERER *et al.* 1998; WEI *et al.* 1999). In contrast, common alleles are likely to be older in the population. They are generally of intermediate size and are consequently subject to fewer constraints and less frequent turnover. On this view, recombination would be more effective in breaking up disequilibrium between common alleles, given their stability and the time that has passed since the occurrence or introduction of associations. Persistence of some disequilibria between common alleles could reflect the levels of background disequilibrium in 11p15.

Gametic disequilibrium analysis in humans has proven extremely useful for fine mapping of a large number of loci that have major effects in rare Mendelian diseases. In contrast, the usefulness of whole-genome association studies, recently proposed as a means of identifying the numerous genes of weak effect that underlie susceptibility to common diseases, remains controversial (RISCH and MERIKANGAS 1996; TERWILLIGER and WEISS 1998; JORDE 2000; RISCH 2000). It was pointed out that the occurrence of useful levels of gametic disequilibrium over large distances is a key prerequisite for strategies of this type because it reduces the number of markers required (COLLINS *et al.* 1999; KRUGLYAK 1999; BOEHNKE 2000). However, the finding of significant disequilibria between distant pairs of microsatellites on 11p15 does not necessarily mean that these disequilibria would be useful for association studies, since our sample was very large. It might be argued that the genetic distance over which significant disequilibrium can be detected increases with sample size. Thus, large sample sizes would be able to detect weak long-distance disequilibria that are not relevant to association studies. KRUGLYAK (1999) considered that levels of association be-

tween putative disease variants and marker loci of >0.1, as measured by the coefficient d^2, can be useful for whole-genome association studies. For comparison with Kruglyak's study, we calculated d^2, considering that either locus in each pair might represent a putative disease locus in the same way as in EAVES *et al.* (2000). The results of this analysis show that useful levels of disequilibrium (*sensu* KRUGLYAK 1999) extend over large genetic distances in 11p15. By way of illustration, 18% of the significant associations between very distant alleles (16–19 cM), show $d^2 \geq 0.20$ [$D'_{ij}(+) \geq 0.47$]; indeed d^2 was in some cases as high as 0.42 [$D'_{ij}(+) = 0.74$].

We detected no global relationship of the strength of interallelic disequilibrium with recombination frequency or with physical distance. A variety of factors may disturb the expected negative relationship between disequilibrium strength and recombination frequency (or physical distance) for distant markers (PETERSON *et al.* 1995; JORDE 2000). However, the above-mentioned variation in the ages of interallelic associations within each pair of microsatellites could be a major factor explaining why the strength of disequilibrium between pairs of alleles does not seem to be correlated with recombination frequency in 11p15. In support of this, we observed increasingly strong negative correlations between strength of interallelic disequilibrium and recombination frequency and between strength of interallelic disequilibrium and physical distance, with increasing allele frequencies at the two loci. The trend can be explained by assuming that more common, and probably older, alleles were exposed for a longer period to recombination than recent variants present at low frequency. This allele-dependent pattern is of significance for mapping disease genes, since the use of gametic disequilibrium in gene mapping assumes a predictable negative relationship between strength of disequilibrium and genetic distance between polymorphisms. Our analysis suggests that more frequent and older alleles at microsatellite loci may be more useful in disequilibrium mapping than rarer alleles. Extension of our analysis to other chromosome regions and populations is clearly needed to enable more general evaluations of the relationship between disequilibrium strength and genetic distance and assessment of the usefulness of microsatellite markers for localizing genes on the basis of disequilibrium.

The amount of disequilibrium may differ greatly across chromosomes, so there is a clear need for a whole-genome gametic disequilibrium map (HUTTLEY *et al.* 1999; KRUGLYAK 1999; BOEHNKE 2000; TAILLON-MILLER *et al.* 2000). Evaluation of disequilibrium at the whole-genome level is critical to successful localization of disease genes by disequilibrium, because the strength of disequilibrium between alleles at a marker locus (or loci) and a disease-predisposing allele must be greater than the background disequilibrium or disequilibrium among marker loci in anonymous regions of the ge-

nome (Freimer *et al.* 1997; Gordon *et al.* 2000). It was pointed out that "Quantifying the degree of such 'background' LD (LD, linkage or gametic disequilibrium) is a crucial undertaking in paving the way for whole genome association studies; to convince ourselves that LD between a disease phenotype and marker loci is meaningful, we must first be assured that we are not simply detecting background LD" (Freimer *et al.* 1997). We found that useful disequilibrum levels between pairs of microsatellite markers extend over the entire 11p15 region. However, our observations also suggest that microsatellites are not only markers of disequilibrium occurring in the chromosomal regions surrounding them (*i.e.*, markers of background disequilibrium), since a substantial fraction of the disequilibrium detected between microsatellites in 11p15 seems to be a consequence of their complex mutational dynamics. If so, using microsatellites as markers of background disequilibrium across the human genome requires us to determine what proportion of the disequilibrium detected is attributable to the marker itself. Further quantification of the levels of background disequilibrium for a given chromosome region, using genetic markers with very different mutational dynamics (such as microsatellites and SNPs), will be needed to separate background disequilibrium from marker-dependent disequilibrium.

LITERATURE CITED

Agresti, A., 1990 *Categorical Data Analysis.* John Wiley & Sons, New York.

Bernardi, G., 1995 The human genome: organization and evolutionary history. Annu. Rev. Genet. **29:** 445–476.

Bodmer, W. F., 1986 Human genetics: the molecular challenge. Cold Spring Harbor Symp. Quant. Biol. **51:** 1–13.

Boehnke, M., 2000 A look at linkage disequilibrium. Nat. Genet. **25:** 246–247.

Brown, A. H. D., 1975 Sample sizes required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. **8:** 184–201.

Budowle, B., R. Chakraborty, A. M. Giusti, A. J. Eisenberg and R. C. Allen, 1991 Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am. J. Hum. Genet. **48:** 137–144.

Camilli, G., and K. D. Hopkins, 1978 Applicability of chi-square to 2×2 contingency tables with small expected cell frequencies. Psychol. Bull. **85:** 163–167.

Chakraborty, R., and K. Weiss, 1988 Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc. Natl. Acad. Sci. USA **85:** 9119–9123.

Collins, A., C. Lonjou and N. E. Morton, 1999 Genetic epidemiology of single-nucleotide polymorphisms. Proc. Natl. Acad. Sci. USA **96:** 15173–15177.

Curie-Cohen, M., 1982 Estimates of inbreeding in a natural population: a comparison of sampling properties. Genetics **100:** 339–358.

D'Agostino, R. B., W. Chase and A. Belanger, 1988 The appropriateness of some common procedures for testing the equality of two independent binomial populations. Am. Stat. **42:** 198–202.

Deka, R., L. Jin, M. D. Shriver, L. M. Yu, S. Decroo *et al.*, 1995 Population genetics of dinucleotide $(dC-dA)_n \cdot (dG-dT)_n$ polymorphisms in world populations. Am. J. Hum. Genet. **56:** 461–474.

Dempster, A. P., N. M. Laird and D. B. Rubin, 1977 Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B **39:** 1–38.

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

Ellegren, H., 2000 Heterogeneous mutation processes in human microsatellite DNA sequences. Nat. Genet. **24:** 400–402.

Everitt, B. S., 1997 *The Analysis of Contingency Tables.* Chapman & Hall, London.

Fallin, D., and N. J. Schork, 2000 Accuracy of haplotype frequency estimation for biallelic loci, via the Expectation-Maximization algorithm for unphased diploid genotype data. Am. J. Hum. Genet. **67:** 947–959.

Falush, D., and Y. Iwasa, 1999 Size-dependent mutability and microsatellite constraints. Mol. Biol. Evol. **16:** 960–969.

Franklin, I., and R. C. Lewontin, 1970 Is the gene the unit of selection? Genetics **65:** 707–734.

Freimer, N. B., S. K. Service and M. Slatkin, 1997 Expanding on population studies. Nat. Genet. **17:** 371–373.

Garza, J. C., M. Slatkin and N. B. Freimer, 1995 Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol. Biol. Evol. **12:** 594–603.

Goldstein, D. B., and D. D. Pollock, 1997 Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J. Hered. **88:** 335–342.

Gordon, D., I. Simonic and J. Ott, 2000 Significant evidence for linkage disequilibrium over a 5 cM region among Afrikaners. Genomics **66:** 87–92.

Guo, S. W., and E. A. Thompson, 1992 Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics **48:** 361–372.

Haas, H., B. Budowle and G. Weiler, 1994 Horizontal polyacrylamide gel electrophoresis for the separation of DNA fragments. Electrophoresis **15:** 153–158.

Hästbacka, J., A. De La Chapelle, M. M. Mahtani, G. Clines, M. P. Reeve-Daly *et al.*, 1994 The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell **78:** 1073–1087.

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

Hedrick, P., S. Jain and L. Holden, 1978 Multilocus systems in evolution. Evol. Biol. **11:** 101–184.

Hill, W. G., 1974 Estimation of linkage disequilibrium in randomly mating populations. Heredity **33:** 229–239.

Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Holm, S., 1979 A simple sequentially rejective multiple test procedure. Scand. J. Stat. **6:** 65–70.

Huttley, G. A., M. W. Smith, M. Carrington and S. J. O'Brien, 1999 A scan for linkage disequilibrium across the human genome. Genetics **152:** 1711–1722.

Jin, L., C. Macaubas, J. Hallmayer, A. Kimura and E. Mignot, 1996 Mutation rates among alleles at a microsatellite locus: phylogenetic evidence. Proc. Natl. Acad. Sci. USA **93:** 15285–15288.

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

Karlin, S., and A. Piazza, 1981 Statistical methods for assessing linkage disequilibrium at the HLA-A, B, C loci. Ann. Hum. Genet. **45:** 79–94.

Kendall, F. B. A., and A. Stuart, 1979 *The Advanced Theory of Statistic. Vol 2. Inference and Relationship.* Charles Griffin & Company, London.

Kerem, B. S., J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox *et al.*, 1989 Identification of the cystic fibrosis gene: genetic analysis. Science **245:** 1073–1080.

Kosambi, D. D., 1944 The estimation of the map distance from recombination values. Ann. Eugen. **12:** 172–175.

Kruglyak, L., 1999 Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. **22:** 139–144.

LAAN, M., and S. PÄÄBO, 1997 Demographic history and linkage disequilibrium in human populations. Nat. Genet. **17:** 435–438.

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations: heterotic models. Genetics **49:** 49–67.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change.* Columbia University Press, New York.

LEWONTIN, R. C., 1985 Population genetics. Annu. Rev. Genet. **19:** 81–102.

LU, H. L., and S. NÈGRE, 1993 Use of glycerol for enhanced efficiency and specificity of PCR amplification. Trends Genet. **9:** 297.

MACAUBAS, C., L. JIN, J. HALLMAYER, A. KIMURA and E. MIGNOT, 1997 The complex mutation pattern of a microsatellite. Genome Res. **7:** 635–641.

MANTEL, N., 1967 The detection of disease clustering and a generalized regression approach. Cancer Res. **27:** 209–220.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NEI, M., and D. GRAUR, 1984 Extent of protein polymorphism and the neutral mutation theory. Evol. Biol. **17:** 73–118.

NEI, M., and W. H. LI, 1973 Linkage disequilibrium in subdivided populations. Genetics **75:** 213–219.

OHTA, T., 1982 Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc. Natl. Acad. Sci. USA **79:** 1940–1944.

OTT, J., 1999 *Analysis of Human Genetic Linkage.* The Johns Hopkins University Press, Baltimore.

OTT, J., and D. RABINOWITZ, 1997 The effect of marker heterozygosity on the power to detect linkage disequilibrium. Genetics **147:** 927–930.

PENA, S. D. J., K. T. SOUZA, M. ANDRADE and R. CHAKRABORTY, 1994 Allelic associations of two polymorphic microsatellites in intron 40 of the human von Willebrand factor gene. Proc. Natl. Acad. Sci. USA **81:** 723–727.

PETERSON, A. C., A. DI RIENZO, A. E. LEHESJOKI, A. DE LA CHAPELLE, M. SLATKIN et al., 1995 The distribution of linkage disequilibrium over anonymous genome regions. Hum. Mol. Genet. **4:** 887–894.

RICE, W. R., 1989 Analyzing tables of statistical tests. Evolution **43:** 223–225.

RISCH, N. J., 2000 Searching for genetic determinants in the new millennium. Nature **405:** 847–856.

RISCH, N., and K. MERIKANGAS, 1996 The future of genetic studies of complex human diseases. Science **273:** 1516–1517.

ROBERTSON, A., and W. G. HILL, 1984 Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. Genetics **107:** 703–718.

ROTHMAN, K. J., 1990 No adjustments are needed for multiple comparisons. Epidemiology **1:** 43–46.

SALAS, A., D. COMAS, M. V. LAREU, J. BERTRANPETIT and A. CARRACEDO, 1998 mtDNA analysis of the Galician population: a genetic edge of European variation. Eur. J. Hum. Genet. **6:** 365–375.

SAMADI, S., F. ERARD, A. ESTOUP and P. JARNE, 1998 The influence of mutation, selection and reproductive systems on microsatellite variability: a simulation approach. Genet. Res. **71:** 213–222.

SCHLÖTTERER, C., and D. TAUTZ, 1992 Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20:** 211–215.

SCHLÖTTERER, C., R. RITTER, B. HARR and G. BREM, 1998 High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. Mol. Biol. Evol. **15:** 1269–1274.

SCHNEIDER, S., D. ROESSLI and L. EXCOFFIER, 2000 Arlequin, ver. 2000. Genetics and Biometry Laboratory, University of Geneva, Switzerland. Available from http://anthropologie.unige.ch/arlequin/.

SHERRINGTON, R., G. MELMER, M. DIXON, D. CURTIS, B. MANKOO et al., 1991 Linkage disequilibrium between two highly polymorphic microsatellites. Am. J. Hum. Genet. **49:** 966–971.

SLATKIN, M., 1994 Linkage disequilibrium in growing and stable populations. Genetics **137:** 331–336.

SLATKIN, M., and L. EXCOFFIER, 1996 Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. Heredity **76:** 377–383.

SMOUSE, P. E., J. C. LONG and R. R. SOKAL, 1986 Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. **35:** 627–632.

SOKAL, R. R., and F. J. ROHLF, 1995 *Biometry.* W. H. Freeman, New York.

STEPHENS, J. C., D. BRISCOE and S. J. O'BRIEN, 1994 Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am. J. Hum. Genet. **55:** 809–824.

TAILLON-MILLER, P., I. BAUER-SARDIÑA, N. L. SACCONE, J. PUTZEL, T. LAITINEN et al., 2000 Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat. Genet. **25:** 324–328.

TAKAHATA, N., and Y. SATTA, 1997 Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc. Natl. Acad. Sci. USA **94:** 4811–4815.

TERWILLIGER, J. D., and K. M. WEISS, 1998 Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr. Opin. Biotechnol. **9:** 578–594.

THOMPSON, E. A., S. DEEB, D. WALKER and A. G. MOTULSKY, 1988 The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am. J. Hum. Genet. **42:** 113–124.

UPTON, G. J. C., 1982 A comparison of alternative tests for the 2×2 comparative trial. J. R. Stat. Soc. A **145:** 86–105.

VALDES, A. M., M. SLATKIN and N. B. FREIMER, 1993 Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics **133:** 737–749.

WEI, C. C., F. T. CHIANG, K. S. LIN and L. I. LIN, 1999 The spectrum of microsatellite loci on chromosomes 7 and 8 in Taiwan aboriginal populations: a comparative population genetic study. Hum. Genet. **104:** 333–340.

WEIR, B. S., 1979 Inferences about linkage disequilibrium. Biometrics **35:** 235–254.

WEIR, B. S., and C. C. COCKERHAM, 1978 Testing hypothesis about linkage disequilibrium with multiple alleles. Genetics **88:** 633–642.

WEISSENSTEINER, T., and J. S. LANCHBURY, 1996 Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. BioTechniques **21:** 1102–1108.

WILSON, J. F., and D. B. GOLDSTEIN, 2000 Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. Am. J. Hum. Genet. **67:** 926–935.

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

ZAPATA, C., and G. ALVAREZ, 1992 The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila.* Evolution **46:** 1900–1917.

ZAPATA, C., and G. ALVAREZ, 1993 On the detection of nonrandom associations between DNA polymorphisms in natural populations of *Drosophila.* Mol. Biol. Evol. **10:** 823–841.

ZAPATA, C., and G. ALVAREZ, 1997a On Fisher's exact test for detecting gametic disequilibrium between DNA polymorphisms. Ann. Hum. Genet. **61:** 71–77.

ZAPATA, C., and G. ALVAREZ, 1997b Testing for homogeneity of gametic disequilibrium among populations. Evolution **51:** 606–607.

ZAPATA, C., and G. VISEDO, 1995 Gametic disequilibrium and physical distance. Am. J. Hum. Genet. **57:** 190–191.

ZAPATA, C., G. ALVAREZ and C. CAROLLO, 1997 Approximate variance of the standardized measure of gametic disequilibrium D'. Am. J. Hum. Genet. **61:** 771–774.

ZERBA, K. E., A. M. KESSLING, J. DAVIGNON and C. F. SING, 1991 Genetic structure and the search for genotype-phenotype relationships: an example from disequilibrium in the Apo B gene region. Genetics **129:** 525–533.

Communicating editor: D. CHARLESWORTH