Am. J. Hum. Genet. 68:191–197, 2001

Extent and Distribution of Linkage Disequilibrium in Three Genomic Regions

Gonçalo R. Abecasis,[1] Emiko Noguchi,[1] Andrea Heinzmann,[1] James A. Traherne,[1] Sumit Bhattacharyya,[1] Nicholas I. Leaves,[1] Gavin G. Anderson,[1] Youming Zhang,[1] Nicholas J. Lench,[2] Alisoun Carey,[2] Lon R. Cardon,[1] Miriam F. Moffatt,[1] and William O. C. Cookson[1]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, and [2]Oxagen Ltd., Abingdon, United Kingdom

The positional cloning of genes underlying common complex diseases relies on the identification of linkage disequilibrium (LD) between genetic markers and disease. We have examined 127 polymorphisms in three genomic regions in a sample of 575 chromosomes from unrelated individuals of British ancestry. To establish phase, 800 individuals were genotyped in 160 families. The fine structure of LD was found to be highly irregular. Forty-five percent of the variation in disequilibrium measures could be explained by physical distance. Additional factors, such as allele frequency, type of polymorphism, and genomic location, explained <5% of the variation. Nevertheless, disequilibrium was occasionally detectable at 500 kb and was present for over one-half of marker pairs separated by <50 kb. Although these findings are encouraging for the prospects of a genomewide LD map, they suggest caution in interpreting localization due to allelic association.

Introduction

Linkage disequilibrium (LD) mapping can precisely locate genes of small effect and could be used to identify common disease genes in genomewide scans or to reduce the number of candidate genes in a region in which linkage has been established (Lander 1996; Risch and Merikangas 1996; Collins et al. 1997; Lai et al. 1998). In the presence of common disease alleles, or when the frequency of rare alleles is increased through selection, the sample sizes required for LD studies are much smaller than those for equivalently powered linkage studies (Risch and Merikangas 1996). Unlike traditional linkage studies, the power of LD mapping depends strongly on disease-allele frequencies and on the extent of disequilibrium between marker and disease alleles (Muller-Myhsok and Abel 1997; Tu and Whittemore 1999).

Although large-scale detection of single-nucleotide polymorphism (SNP) markers is already under way (Collins et al. 1998), data on genomewide LD patterns is limited. Simulations have suggested that LD may extend for <5 kb, even in relatively isolated populations, so that >1,000,000 equally spaced markers may be required for genomewide LD scans (Kruglyak 1999). However, reviews of published data provide examples of LD at distances >100 kb (Collins et al. 1999; Huttley et al. 1999), and there is evidence that LD patterns vary between populations (Goddard et al. 2000; Kidd et al. 2000). For example, LD is detectable at 500 kb in the APC gene region on chromosome 5 (Jorde et al. 1994) and significant LD between microsatellite loci has been shown to extend to 4.0 cM in some chromosomal regions (Huttley et al. 1999). Other studies have shown that the distribution of LD is irregular in a number of chromosomal regions (Clark et al. 1998; Rieder et al. 1999; Moffatt et al. 2000; Templeton et al. 2000). Here we examine the patterns of LD for 127 polymorphisms in three genomic regions, in a sample of 575 chromosomes from two white populations of British ancestry.

Received September 27, 2000; accepted for publication October 24, 2000; electronically published November 13, 2000.

Address for correspondence and reprints: Dr. William Cookson, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, England OX3 7BN. E-mail: william.cookson@ndm.ox.ac.uk

0002-9297/2001/6801-0018$02.00

Methods

Subjects

Two panels of subjects were studied. The first panel consisted of 410 white subjects within 88 nuclear families selected for the presence of atopic disease from an Australian random-population sample of 230 families (Moffatt et al. 1994). The second panel consisted of 410 white British individuals from 66 nuclear and 5 extended pedigrees, ascertained through members with asthma or rhinitis (Moffatt et al. 1994). The prevalence of asthma was 28% in British subjects and 18% in Australians, compared with a population prevalence of 14% in each locale. The studies were approved by regional ethics

committees, and all subjects gave written informed consent.

Markers and Genotyping

We examined 127 markers in three genomic regions: the IL1 cytokine cluster on chromosome 2q13 (containing 50 SNPs, 1 VNTR, and 12 microsatellites within 2.5 Mb) (Bhattacharyya 1998); a locus on chromosome 13q13 (36 microsatellites and 2 diallelic loci within 4.2 Mb) (Bhattacharyya 1998); and the TCR α/δ locus on chromosome 14q11 (24 SNPs and 2 microsatellites within 850 kb) (Moffatt et al. 2000). We are currently investigating these loci in an attempt to identify asthma-susceptibility genes. The majority of SNPs were identified through a pooled sequencing strategy, and the sample is biased toward markers with common alleles (table 1). The markers were not selected with reference to potential coding sequences.

SNPs and microsatellites were genotyped as described elsewhere (Daniels et al. 1996; Bhattacharyya 1998; Moffatt et al. 2000). Briefly, SNPs were genotyped either by PCR amplification followed by RFLP analysis or by sequence-specific oligonucleotide probing, as appropriate. Microsatellite markers were genotyped using fluorescently labeled oligonucleotide probes. All SNP genotype calls were checked independently by two individuals. To evaluate error rates, complete assays were repeated for 1,878 alleles, and 15 discrepancies consistent with Mendelian inheritance between assays were observed. Seven of these were identified in a check for double recombinants.

Haplotyping

Allele frequencies did not differ between the two populations (χ^2 test for independence, $P > .05$ for 99% of all markers), so the data were pooled. For each pedigree, haplotype sets were identified by the SIMWALK2 program (available by anonymous FTP [Sobel and Lange 1996]). In a preliminary analysis, genotypes that implied a double recombination event were identified and excluded from further analyses. To improve the performance of the SIMWALK2-simulated annealing algorithm, we increased the number temperature changes and the number of pedigree configurations sampled at each temperature and reduced the magnitude of each temperature change. The full set of estimated haplotypes used in this analysis is available at the Asthma Genetics Web site. For comparison, we also estimated founder haplotype frequencies for each pair of linked markers, using the expectation-maximization (EM) algorithm (Excoffier and Slatkin 1995). Rare alleles with a frequency <7% were pooled when microsatellite markers were analyzed.

Table 1

Distribution of SNP Allele Frequencies for 76 SNPs

Allele Frequency Range (%)	Count
<10	7
10–20	10
20–30	13
30–40	31
40–50	15

Pairwise Disequilibrium

Using either the founder haplotype sets identified by SIMWALK2 or the EM haplotype frequencies, the standardized multiallelic disequilibrium coefficient D' was calculated (Hedrick 1987). D' varies between 0 (no disequilibrium) and 1 (maximum disequilibrium). Maximum disequilibrium occurs only when some haplotypes have frequency 0; intermediate levels of disequilibrium imply recombination between markers or recurrent mutation. For each marker pair, the strength of association was measured by a standard contingency-table χ^2 test. Information, K, was defined as the ratio between a χ_1^2 and squared D' values (for multiallelic markers χ_n^2 values were converted to χ_1^2 values) (Collins et al. 1999).

Disequilibrium across each locus was plotted by the GOLD program (Abecasis and Cookson 2000). For each marker pair, GOLD plots the color-coded pairwise disequilibrium statistics at the Cartesian coordinates corresponding to marker location, and the plots are completed by interpolation.

Observed D' values were sorted according to distance between the corresponding marker pairs. For each set of 30 consecutive observations, a running average D' value and the proportion of values >.33 were estimated and plotted.

Decay of D'

A simple model for the decay of pairwise linkage disequilibrium over distance is $E(D'_t) = (1 - \theta)^t D'_0$, where $E(D'_t)$ is the expected value of D', $D'_0 = 1$ is the initial value of D', θ is the recombination fraction between the two loci, and t is the number of generations since $D' = D'_0$ (Hartl and Clark 1997). When D' is measured on a finite sample of chromosomes, the average of $|D'|$ is >0, even if $D' = 0$. We therefore modified the model through two additional parameters, to allow for a positive bias in estimates of D' as well as incomplete disequilibrium between tightly linked markers. These parameters were defined as D'_{low}, the minimum expected D' between markers, and D'_{high}, the maximum D' between tightly linked markers. The model $E(D'_t) = D'_{\text{low}} + (D'_{\text{high}} - D'_{\text{low}})(1 - \theta)^t D'_0$ allows D' to decay exponentially between these boundaries.

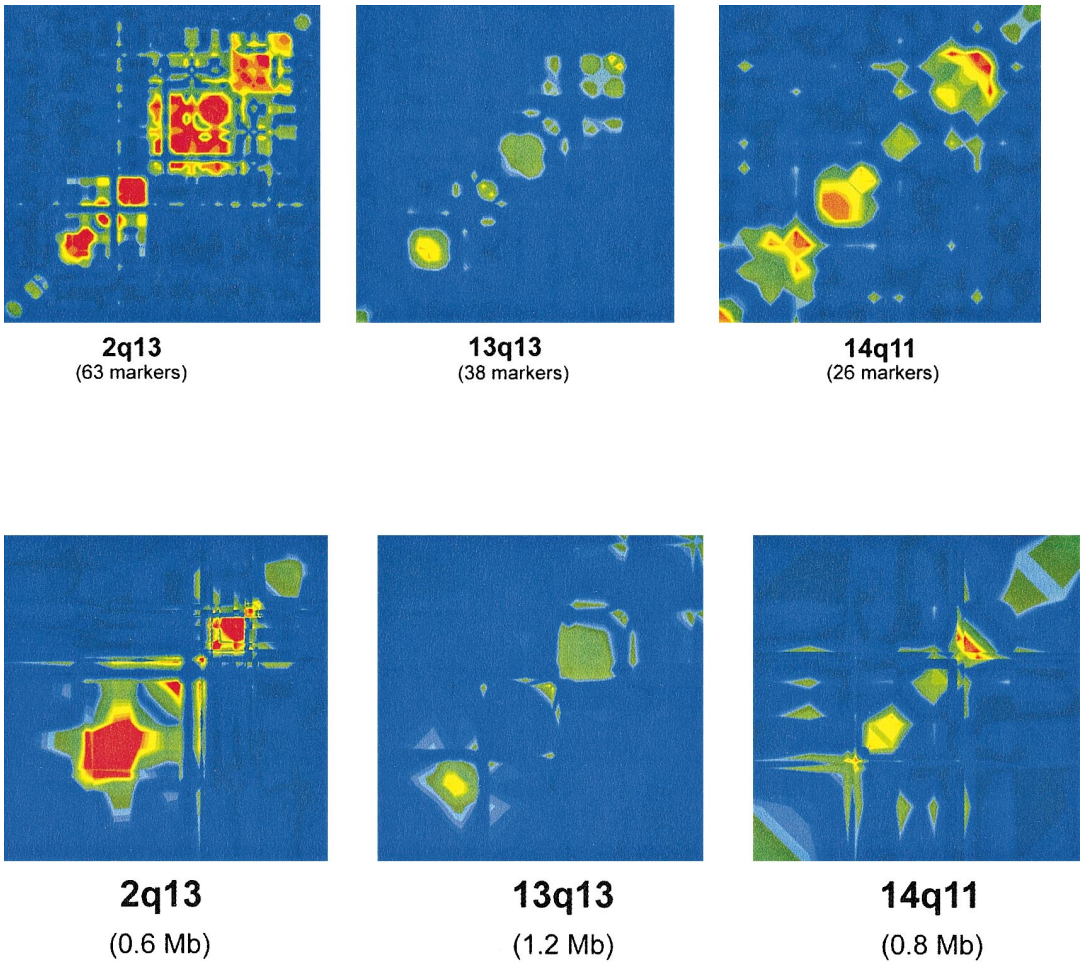

Figure 1 Distribution of LD in three genomic regions. *A,* Raw disequilibrium matrix. *B,* Adjusted for physical distance. Red, green, and blue represent strong disequilibrium ($D' \geqslant .8$), moderate ($D' \sim .5$), and weak ($D' \sim 0$) disequilibrium, respectively (Abecasis and Cookson 2000).

For model fitting, D'_0, was fixed at 1, and θ was estimated from physical distance using the approximation 1 cM = 1 Mb. The parameters D'_{high}, D'_{low}, and t were estimated to minimize the residual variability in D'. This residual variability was measured as the sum of squares $(\text{SS}) = \Sigma K_{ij}[D'_{ij} - E(D'_{ij})]^2$, where D'_{ij} is the observed disequilibrium coefficient between i and j, K_{ij} is the information on D'_{ij}, and $E(D'_{ij})$ is as defined above. (This measure is simply the standard root mean squared error in unstandardized form).

Results

Haplotyping

The pedigrees in the study included a total of ~900 meioses. The haplotype sets identified by SIMWALK2 implied a minimum of 58 recombinants (~23 expected if 1 Mb = 1 cM) for chromosome 2, 43 (~37 expected) for chromosome 13, and 14 (~8 expected) for chromosome 14. Pairwise D' coefficients were averaged over 10 alternative haplotype sets identified by SIMWALK2 and estimated from an average of 575 founder chromosomes for each marker pair.

The Spearman rank-order correlation between D' values estimated by haplotyping with SIMWALK2 and those estimated from founder genotypes with the EM algorithm was $r_s = .81$ (2,908 marker pairs, $P \ll 10^{-10}$). The correlation was greater for SNP-SNP pairs ($r_s = .86$, 1,502 pairs) than for other types of marker pairs ($r_s = .68$, 1,406 pairs). This result conforms to theoretical predictions, since the performance of the EM algorithm should degrade as the proportion of heterozygous (phase unknown) genotypes increases. Estimates of D' from haplotypes derived by SIMWALK2 were used in subsequent analyses.

The great variability of D' in these data is summarized in figure 1. As expected, LD tended to be most significant for nearby markers (those along the diagonal). Several

clusters of moderate to high disequilibrium could also be observed, which did not correspond directly to more densely typed regions. The irregularity was seen whether the maps were derived from SNPs (chromosome 14) or from microsatellites (chromosome 13) or their combination (chromosome 2).

LD versus Distance

The data included 2,108 marker pairs separated by <1 Mb (fig. 2*A*). At this scale, the negative correlation between D' and physical distance was striking and highly significant ($P \ll 10^{-10}$). Spearman's rank correlation coefficient (r_s) between D' and distance was −.49 for the combined 2,108 pairs and −.59 (1,439 pairs), −.61 (344 pairs), and −.46 (325 pairs) for chromosomes 2, 13, and 14, respectively.

Plotting the moving average of D' (fig. 2*B*) further demonstrated that the decay of disequilibrium with distance was subject to great stochastic variation. Even when markers were only separated by <1 kb of DNA, D' values were, on average, <1.

The power of association tests is proportional to D'^2, and a value of $D' = .33$ (corresponding to a 10-fold increase in the required sample size) is commonly taken as the minimum usable amount of LD. At distances <50 kb, more than one-half of all marker pairs exhibited a $D' > .33$ (fig. 2*C*). Although this proportion rapidly decreased with distance, some markers separated by up to 400 kb still exhibited moderate amounts of LD in all three regions.

Decay of D′

An exponential model for the decay of D' based on physical distance (using the approximation 1 cM = 1 Mb) explained 45% of the variation in the 2,908 D' coefficients estimated between pairs of linked markers (table 2). If complete disequilibrium between tightly linked markers was assumed (by fixing the parameter $D'_{\text{high}} = 1.0$), then the proportion of the variation explained dropped to 39%. Assuming that disequilibrium between unlinked markers was zero (by fixing the parameter $D'_{\text{low}} = 0.0$) reduced the proportion of variance explained to 41%. The best-fit model parameters give the baseline level of D' in this study as $D'_{\text{low}} = .08$, the expected disequilibrium between tightly linked markers as $D'_{\text{high}} = .69$, and the exponential decay time as $t = 1{,}927$. Discounting the inflation in observed LD implied by the D'_{low} parameter, this model suggested that low levels of disequilibrium ($D' < .33$) are expected for distances greater than ~35 kb.

Although the fitted parameters were similar in each genomic region (table 2), D'_{high} and the decay rate t appeared to be somewhat lower in the 13q13 region. Allowing for a different model for each of the three chromosomal regions (9 additional parameters) increased the overall proportion of variance explained from 45% to 46%. If the decay of LD was modeled separately for SNP-SNP pairs and other marker pairs (three additional parameters), the proportion of variation explained increased from 45% to 50%. Subdividing SNP-SNP pairs into three groups on the basis of allele frequencies (one allele frequency <20%, one allele frequency <30%, and both allele frequencies >30%, giving six additional parameters) further increased this by <1%.

Discussion

The measurement, extent, and variability of marker-marker disequilibrium are important factors in deciding the number of markers required for association scans and for interpreting marker-disease association results. The study used D' as a measure of allelic association because it has a simple interpretation, its scale is independent of allele frequency, and it is applicable to both SNP and microsatellite data (Hedrick 1987; Devlin and Risch 1995). The precision and variance of individual D' estimates suffer at low allele frequencies, and we allow for that by using a measure of information that increases with sample size and with marker heterozygosity. In our study, the sample size was large, and 90% of the SNPs had minor allele frequencies that were >10%, so that the results are unlikely to be biased by the use of D' to measure LD. Our results also indicate that haplotypes derived by the EM algorithm are less accurate than those derived from family data.

This study represents the largest examination of disequilibrium patterns carried out to date, and the findings illustrate the difficulties to be faced in LD mapping of complex traits. Our results show that significant disequilibrium in these regions can be detected at distances of 30–300 kb, which is an order of magnitude greater than those predicted by simulation (Kruglyak 1999). However, the distribution of LD is highly irregular. These findings are consistent with other studies of the extent and variability of disequilibrium in real data (Clark et al. 1998; Collins et al. 1999; Huttley et al. 1999; Rieder et al. 1999; Eaves et al. 2000; Templeton et al. 2000). If the extent of LD observed in our data were to be representative of the human genome, relatively small numbers of markers might cover the genome for a basic LD map for disease association studies.

Approximately 45% of the variation in LD could be explained by physical distance, and none of the other factors we considered (marker type, chromosomal location, and SNP allele frequency) could explain >5% of the remaining variation.

The biological reasons for the residual variation in distribution of LD are not yet known. They may include selective sweeps and genetic drift. However, the varia-

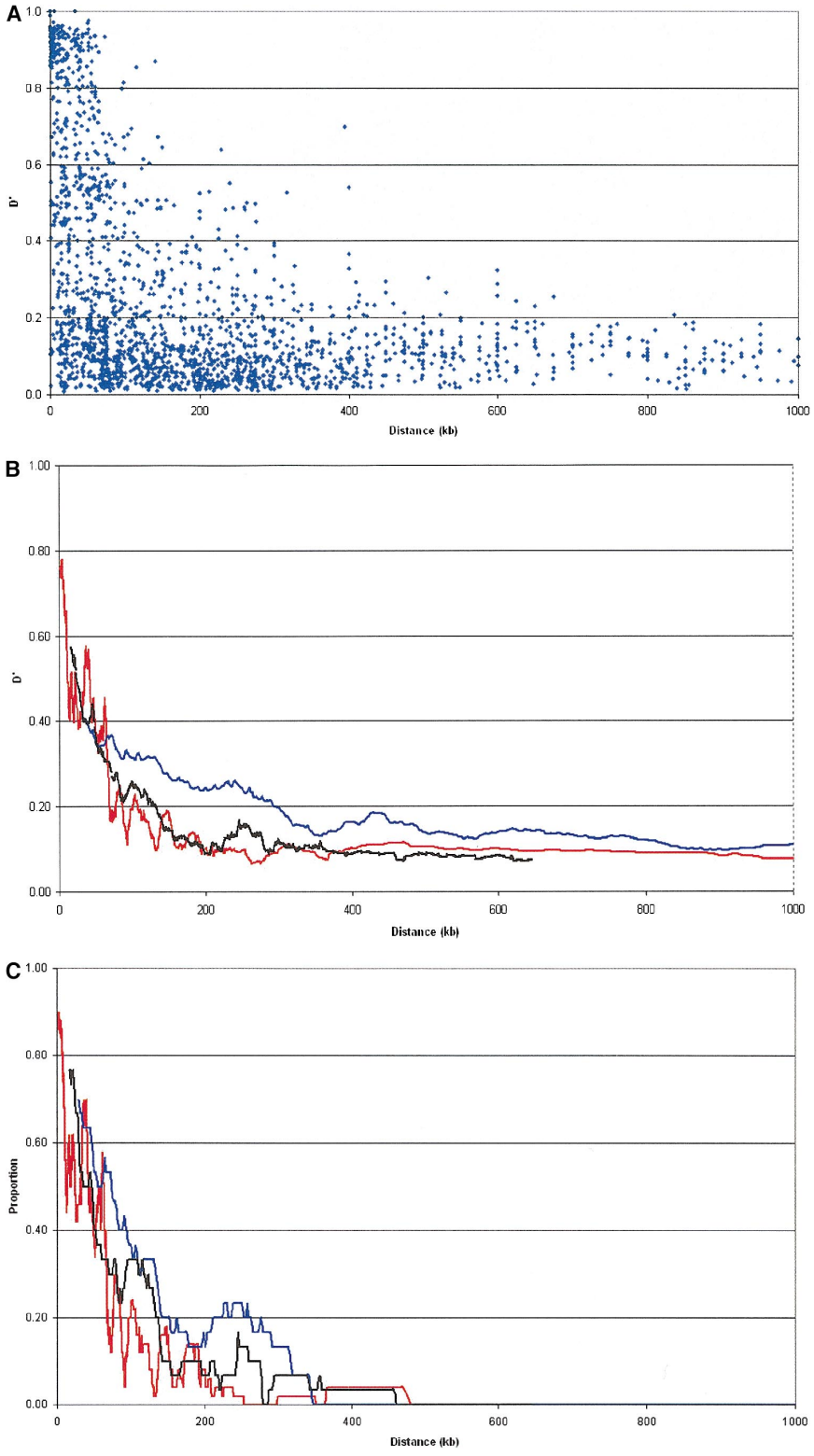

Figure 2 *A*, LD (*D′*) and physical distance for markers derived from three genomic regions. *D′* for marker pairs separated by <1 Mb is shown. *B*, Moving averages of *D′* and physical distance in three genomic regions. Data from chromosome 2 are shown in red, from chromosome 13 in blue, and from chromosome 14 in black. *C*, Frequency of useful *D′* (<.33) and physical distance in three genomic regions. Data from chromosome 2 are shown in red, from chromosome 13 in blue, and from chromosome 14 in black.

Table 2

Modeling the Decay of Disequilibrium with Distance Parameters for the Model $E(D'_t) = D'_{low} + (D'_{high} - D'_{low})(1 - \theta)^t$, Estimated for Each Genomic Region

	Model Parameters			Variance Explained	Predicted D' (kb)		
Region	D'_{high}	D'_{low}	t	(%)	10	20	30
2q13	.70	.07	1.8×10^3	45.3	.59	.48	.43
13q13	.49	.10	0.6×10^3	45.5	.46	.44	.42
14q11	.59	.06	2.1×10^3	45.5	.49	.41	.34
Overall	.69	.08	1.9×10^3	45.2	.58	.50	.42

tion may also be due to physical characteristics of DNA that induce regional differences in mutation or recombination rates. Physical causes of regional variation are likely to result in reproducible patterns of LD between populations, and the recognition of such physical factors may be important in effective localization of disease-causing genes by LD mapping (Majewski and Ott 2000).

The difference in extent of LD between the three chromosomal regions was small, with the decay rate t and expected levels of disequilibrium between tightly linked markers D'_{high} a little lower in the 13q13 region. This region was studied primarily with microsatellite markers, whereas the other two loci included mostly SNP markers. Recurrent mutation is more likely for microsatellite alleles (Brinkmann et al. 1998), and polyphyletic origins would explain lower predictions of disequilibrium for tightly linked 13q13 markers. Additionally, a more recent origin of microsatellite alleles could explain the slower observed decay of disequilibrium in this region.

Microsatellites reflect LD from multiple alleles simultaneously, and if LD also extends for longer distances around microsatellites, they may be more informative than SNPs for low-density scans or when searching for mutations of recent origin. Dense panels of microsatellite markers are already in use for genetic-linkage studies (Weissenbach et al. 1992), and there are 12,000 microsatellites in the public domain, so that these markers also merit consideration in LD mapping.

The extent of LD in our study also depends on the age of the population and the number of founders from which our subjects were derived. Kimura and Ota (1973) have derived expectations for the age (t) of a neutral polymorphism based on effective population size (Ne) and allele frequency. An SNP with minor allele frequency >.05 (such as the SNPs in this study) is expected to be older than 1,500 generations if Ne = 2,000 and older than 3,900 generations if Ne = 5,000. Our simple model for the decay of disequilibrium estimates the age of the haplotypes in our study to be $t \sim 2{,}000$, suggesting a relatively small effective population size for this British white population. This small effective population size might be explained by genetic drift (due to population bottlenecks) and geographically clustered mating (Thompson and Neel 1997), which would account for some of the differences with theoretical predictions based on large random-mating populations (Kruglyak 1999).

We have observed that, in some cases, LD can extend as far as 500 kb, and that useful LD may extend to 50 kb for 50% of markers. This suggests that a genome-wide LD map is feasible. However, the mean D' is <1 even for closely linked markers, and SNPs a few base pairs apart may show no LD. The results therefore indicate that care is required when interpreting allelic association as evidence of precise localization. In general, successful LD mapping will require a systematic understanding of local patterns of LD and haplotype evolution, as exemplified by the mapping of polymorphisms in the *ACE* gene, which controls circulating ACE levels (Farrall et al. 1999). Haplotype-based tests may be more powerful in the presence of low levels of LD or multiple disease alleles, but the relationship between haplotype variation and the power of haplotype-based tests is not well described. The accumulation of more-dense data in these and other regions will allow systematic investigation of the patterns and causes of haplotype variability.

Acknowledgment

The study was funded by the Wellcome Trust.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Anonymous FTP, ftp://watson.hgen.pitt.edu/pub/simwalk2 (for SIMWALK2)
Asthma Genetics, http://www.well.ox.ac.uk/asthma/ (for a full set of estimated haplotypes)
GOLD, http://www.well.ox.ac.uk/asthma/GOLD (for GOLD software)

References

Abecasis GR, Cookson WO (2000) GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183
Bhattacharyya S (1998) Dissecting the genetics of asthma. University of Oxford Press, Oxford
Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415
Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF (1998) Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am J Hum Genet 63:595–612
Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology

of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 96:15173–15177

Collins FS, Guyer MS, Charkravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1581

Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L (1998) New goals for the U.S. Human Genome Project: 1998–2003. Science 282:682–689

Daniels SE, Bhattacharrya S, James A, Leaves NI, Young A, Hill MR, Faux JA, Ryan GF, le Souef PN, Lathrop GM, Musk AW, Cookson WO (1996) A genome-wide search for quantitative trait loci underlying asthma. Nature 383: 247–250

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

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (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

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

Farrall M, Keavney B, McKenzie C, Delepine M, Matsuda F, Lathrop GM (1999) Fine-mapping of an ancestral recombination breakpoint in DCP1. Nat Genet 23:270–271

Goddard KA, Hopkins PJ, Hall JM, Witte JS (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. Am J Hum Genet 66:216–234

Hartl DL, Clark AG (1997) Principles of Population Genetics. Sinauer Associates, Sunderland, Mass

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331–341

Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999) A scan for linkage disequilibrium across the human genome. Genetics 152:1711–1722

Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region [see comments]. Am J Hum Genet 54:884–898

Kidd JR, Pakstis AJ, Zhao H, Lu RB, Okonofua FE, Odunsi A, Grigorenko E, Tamir BB, Friedlaender J, Schulz LO, Parnas J, Kidd KK (2000) haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. Am J Hum Genet 66: 1882–1899

Kimura M, Ota T (1973) The age of a neutral mutant persisting in a finite population. Genetics 75:199–212

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

Lai E, Riley J, Purvis I, Roses A (1998) A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. Genomics 54:31–38

Lander ES (1996) The new genomics: global views of biology. Science 274:536–539

Majewski J, Ott J (2000) GT repeats are associated with recombination on human chromosome 22. Genome Res 10: 1108–1114

Moffatt MF, Hill MR, Cornelis F, et al. (1994) Genetic linkage of T cell receptor α/δ complex to specific IgE responses. Lancet 343:1597–1600

Moffatt MF, Traherne JA, Abecasis GR, Cookson WO (2000) Single nucleotide polymorphism and linkage disequilibrium within the TCR alpha/delta locus. Hum Mol Genet 9:1011–1019

Muller-Myhsok B, Abel L (1997) Genetic analysis of complex diseases. Science 275:1328–1330

Rieder MJ, Taylor SL, Clark AG, Nickerson DA (1999) Sequence variation in the human angiotensin converting enzyme. Nat Genet 22:59–62

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping location scores, and marker-sharing statistics. Am J Hum Genet 58:1323–1337

Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF (2000) Recombinational and mutational hotspots within the human lipoprotein lipase gene. Am J Hum Genet 66:69–83

Thompson EA, Neel JV (1997) Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. Am J Hum Genet 60:197–204

Tu IP, Whittemore AS (1999) Power of association and linkage tests when the disease alleles are unobserved. Am J Hum Genet 64:641–649

Weissenbach J, Gypay G, Dib C, et al. (1992) A second generation linkage map of the human genome. Nature 359: 794–801