Human Molecular Genetics, 2000, Vol. 9, No. 7 **1011–1019**

ARTICLE

Single nucleotide polymorphism and linkage disequilibrium within the TCR α/δ locus

M. F. Moffatt[+], J. A. Traherne[+], G. R. Abecasis and W. O. C. M. Cookson[§]

University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK

Received 19 November 1999; Revised and Accepted 11 February 2000

Much attention is being given to the identification of common disease genes through whole-genome linkage disequilibrium (LD) screens with single nucleotide polymorphisms (SNPs). Simulation studies have suggested that useful LD is unlikely to extend beyond 3 kb, and that > 500 000 SNPs may be needed for comprehensive coverage of the genome. The TCR α/δ locus on chromosome 14q contains many V, J and D segments that combine with constant domains to produce either an α or a δ chain of the T cell receptor. Multiple SNPs have been recognized within the V segments, and it has been suggested that variation within the locus may modify the course of autoimmune and allergic diseases. We have examined LD within an 850 kb section of the TCR α/δ locus on chromosome 14q by typing 24 V gene segment SNPs and two microsatellites. One hundred and fifty-nine nuclear and extended families were genotyped in order to derive haplotypes, and the pair-wise LD between SNPs was investigated in 600 haplotypes from unrelated individuals (the parents). The mean extent of useful LD was much greater than suggested by simulations: significant LD was relatively common at 250 kb and was detectable beyond 500 kb. The mean extent of LD was twice as far between alleles of low frequency than between common alleles. The distribution of LD was highly irregular and concentrated in three distinct islands. The results differ from those obtained by simulation, and if they are typical of other genomic regions, suggest that the minimum number of markers necessary for comprehensive LD mapping may be reduced by at least an order of magnitude.

INTRODUCTION

The positional cloning of genes underlying common diseases relies either on the detection of genetic linkage or on the identification of linkage disequilibrium (LD) through allelic association between markers and disease. LD mapping may detect weaker effects than genetic linkage studies (1), so that systematic genome-wide association mapping may be used to identify the genes underlying common complex diseases (1–4). Most attention in genome-wide association studies has been directed at single nucleotide polymorphisms (SNPs) because they are abundant and amenable to large-scale detection and genotyping.

The ability to detect association between marker alleles and disease depends critically on the nature of LD between disease alleles and surrounding markers. The possible extent of LD has been recently evaluated in simulated data, with the conclusion that useful LD is unlikely to extend beyond 3 kb (5). This implies that 500 000 to 3 000 000 SNPs may be necessary for genome-wide LD mapping (5).

The TCR α/δ locus on chromosome 14q contains many V, J and D segments that combine with constant domains to produce either an α or a δ chain of the T cell receptor. Multiple SNPs have been recognized within the V segments (6–9), and it has been suggested that variation within the locus may modify the course of autoimmune (10,11) and allergic (12) diseases.

We have therefore assessed the extent and characteristics of LD between 24 SNPs and two microsatellites within the V segments of the TCR α/δ locus (Table 1). The region of study covered 850 kb and has been sequenced in its entirety (13), so that exact physical distances could be correlated with LD measurements. Multiple pair-wise comparisons of LD between markers were possible, covering a wide range of physical distances between markers from 0 to 850 kb apart.

Many genetic studies of complex disorders are based on family panels. In the present investigation families were genotyped. This meant that haplotypes could be deduced directly from family data rather than inferred from unrelated individuals, and that LD relationships could subsequently be defined with increased precision from the haplotypes.

[+]These authors contributed equally to the work
[§]To whom correspondence should be addressed. Tel: +44 1865 287607; Fax: +44 1865 287578; Email: william.cookson@ndm.ox.ac.uk

Table 1. Polymorphisms within the TCR α/δ locus

Gene	SNP	Old gene name	Frequency	Position (kb)	Heterozygosity
AV1S1	AV1S1A	AV7S1	0.44	0	0.49
″	AV1S1B	″	0.41	0	0.48
AV6S1	AV6S1	AV5S1	0.08	146	0.15
AV8S3	AV8S3	AV1S4	0.44	229	0.49
AV13S1	AV13S1	AV8S1	0.41	245	0.48
–	D14S50[a]	–	–	254	0.78
AV12S2	AV12S2	AV2S1	0.06	263	0.11
AV8S4	AV8S4	AV1S2	0.43	268	0.49
hADV14S1	hADV14S1A	AV6S1-hADV104S1	0.39	298	0.48
″	hADV14S1B	″	0.22	299	0.34
″	hADV14S1C	″	0.39	299	0.48
AV8S6	AV8S6	AV1S3	0.26	356	0.38
AV17S1	AV17S1	AV3S1	0.20	375	0.33
AV21S1	AV21S1	AV23S1	0.50	429	0.50
hADV23S1	hADV23S1	AV17S-hDV106S1	0.04	462	0.08
AV26S1	AV26S1A	AV4S2	0.30	498	0.42
″	AV26S1B	″	0.32	498	0.44
AV27S1	AV27S1A	AV10S1	0.25	523	0.38
″	AV27S1B	″	0.10	523	0.17
hADV29S1	hADV29S1	AV21S1-hDV105S1	0.26	537	0.38
AV30S1	AV30S1	AV29S1	0.03	542	0.06
AV26S2	AV26S2	AV4S1	0.26	575	0.38
hADV38S2	hADV38S2	AV14S1-hDV8S1	0.22	653	0.34
AV39S1	AV39S1	AV27S1	0.24	675	0.36
hDV102S1	hDV102S1	hDV102S1	0.23	794	0.36
–	TCR1 (FCA.TA1)[a]	–	–	842	0.77

Twenty-four SNPs and two microsatellites were typed within the locus. New and old gene names are given. The frequency of the rare SNP allele is given in each case.
[a]Microsatellites.

RESULTS

Measurement of linkage disequilibrium

We studied 810 subjects from 158 British (UK) and Australian (AUS) Caucasian families. The Australians were predominately of British ancestry, with 10% of families owing their origins to southern Europe. Allele frequencies at the TCR α/δ locus and multiple other loci did not differ between populations ($P > 0.1$), and so the data were pooled.

The families contained 339 unrelated individuals (the founders) and 471 descendants. The average family size was 5.13. Fifty percent of families contained four individuals, 39% contained five individuals, and 11% had six or more individuals. Thirty-four percent of the subjects were atopic (showing signs of IgE-mediated allergy) and 18% were asthmatic, compared with 42% atopic and 14% asthmatic for the unselected Busselton population. Ninety-three percent of subjects had given DNA and were able to be genotyped. Haplotypes were estimated by the SimWalk2 program. Thirty-one obligatory double recombinants resulting from a single marker were likely to represent genotyping errors and were therefore removed from the analysis. The subsequent analysis was based on the remaining 19 518 genotypes.

The extent of LD between each marker was measured on ~300 unrelated individuals (the parents) (mean = 295.7 ± 11.6 SD), corresponding to ~600 haplotypes. LD was expressed as D′ (14). This measure performs well in several situations (15,16) and, unlike the measurement d^2 used in recent simulations (5), is symmetric (so that marker A compared with marker B gives the same result as marker B compared with marker A). D′ was measured by two methods: directly from haplotypes; and, ignoring family information, by inference to maximize founder haplotype likelihoods with the expectation maximization (EM) algorithm. The statistical significance of LD between markers was calculated from haplotypes by χ^2 analysis of allele by allele contingency tables.

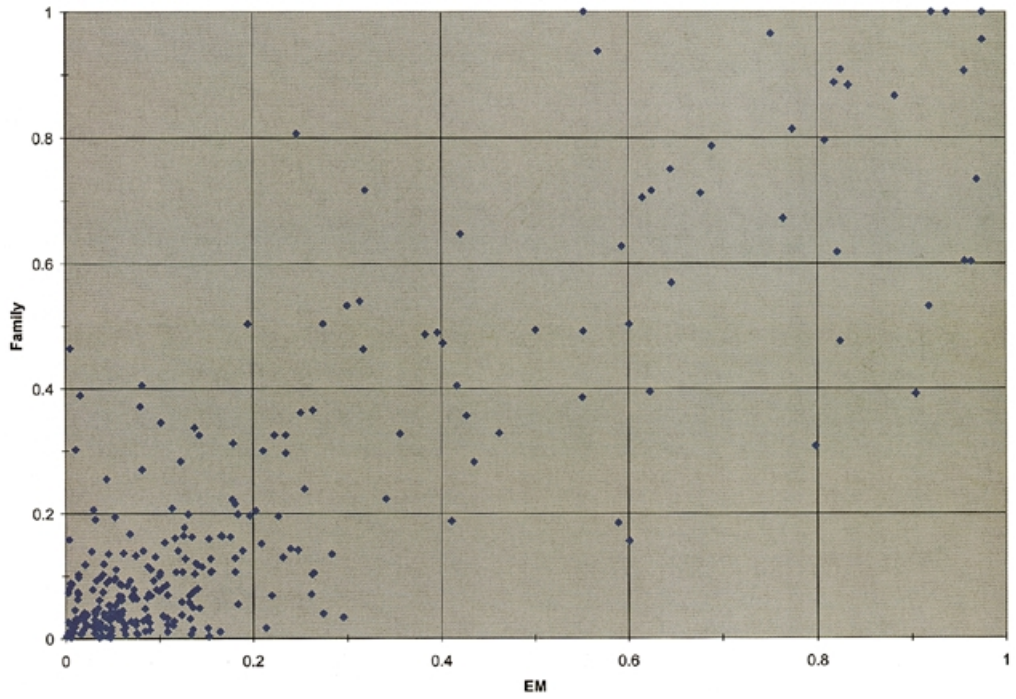

Figure 1. Plot of D′ in unrelated subjects, measured directly from family-derived haplotypes and by inference using the EM algorithm. Haplotypes were created from family data using the SimWalk2 program.

Direct and inferred measurements of disequilibrium were correlated (Spearman's Rank Correlation Coefficient, r_s = 0.67, df = 274, $P << 10^{-10}$), but were in some instances quite different (Fig. 1). For some families Simwalk found alternative sets of founder haplotypes to be equally likely. However, estimates of disequilibrium only varied marginally between equally likely haplotypes (Spearman's Rank Correlation Coefficient = 0.93 ± 0.01, over 100 distinct haplotype sets), and a single set was selected at random for further analyses. Although the mean D′ between all 276 SNP pairs was 0.20 for both methods, only 33 marker pairs showed a significant D′ ($P < 0.05$) by inference, while 61 marker pairs were in significant disequilibrium according to haplotypes inferred from family data. Assuming families contained more information than isolated individuals, these results indicated that inference of D′ using the EM method on unrelated individuals was less precise than using family data, and so the rest of the analyses were based on measurements of D′ from the haplotypes derived by SimWalk2.

Linkage disequilibrium and physical distance

D′ was seen to be strongly correlated with distance (Fig. 2). High levels of LD were not always significant, due to the occurrence of rare alleles in the marker pair. Significant LD was relatively common at 250 kb, and was still detected at distances >500 kb.

Assuming the marker and disease alleles have similar frequencies, the useful power to detect association depends on a value of D′ which approximates to values of >0.33 (5). Beyond this value the requirements of sample size become prohibitive (5). It can be seen that values of D′ >0.33 often extended beyond 300 kb (Fig. 2). These distances are considerably greater than the 3 kb suggested by simulation studies (5).

Inspection of the family data showed that there were 21 recombination events found in 1004 meioses. Recombinations were evenly spread across the locus, with no obvious clusters. The observed recombination fraction (θ) of 0.02 between the beginning and end of the region was higher than the 0.008 that might be anticipated from the physical distance between markers, so that the extent of LD cannot be attributed to a reduced recombination frequency within the locus.

It has been suggested that alleles of moderate frequency are likely to be older than rarer alleles, because time is necessary for new alleles to become common (5). The extent of LD declines with the age of a polymorphism, and it is therefore possible that the degree of LD between markers may be a function of their allele frequencies.

The marker pairs in the region were accordingly divided into three categories of haplotype: likely ancient haplotypes (both SNPs with all alleles having a frequency of >30%), haplotypes of likely intermediate age (both SNPs with all alleles having a frequency of >20%) and likely recent haplotypes (at least one SNP allele with a frequency of <20%). The results show a significant effect of allele frequency on the extent of LD, with combinations of rare alleles showing more LD than common alleles (Fig. 3).

Models for the decay of LD with distance were then fitted to the three categories of marker pairs (Fig. 3, inset). The exponential decay parameter (β) was larger for common SNPs (3.9×10^3), intermediate for moderately frequent SNPs (2.8×10^3) and smallest for rare SNPs (1.8×10^3). A mean D′ of 0.33 was seen to extend to ~30 kb for likely ancient haplotypes, and beyond 60 kb for likely recent haplotypes.

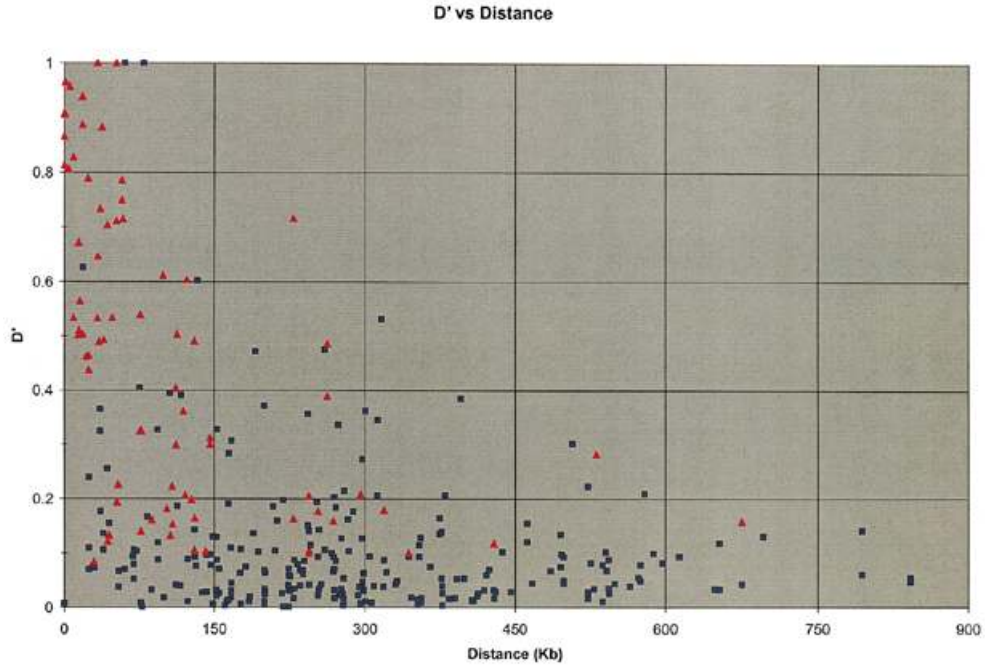

Figure 2. Disequilibrium versus physical distance. Red triangles represent significant marker association ($P < 0.05$).

It is recognized that common diseases may be due to common variants in or around disease causing genes (1–3). Common in this context may still imply a frequency of <20%, so that the extended LD around alleles of this frequency may still be valuable in association studies.

Haplotypes in regions of linkage disequilibrium

Strategies for the generation of SNP maps assume, at least tacitly, that LD is distributed evenly throughout the genome (2–4), or that the distribution is sufficiently predictable for regional marker spacing to be chosen to reflect the pattern of LD (5). Data from some studies have suggested that this may not be the case (17,18). In our data, a wide range of LD was seen at any particular physical distance (for example between 0 and 150 kb, Fig. 2). Therefore, we examined the regional distribution of LD (Fig. 4). The distribution of LD was found to be highly irregular, with significant LD concentrated into three regions. These regions were signified as TCRα/δ-A, TCRα/δ-B and TCRα/δ-C.

TCRα/δ-A contained the *AV6S1*, *AV8S3*, *AV13S1*, *AV12S2* and *AV8S4* SNPs and covered ~120 kb. TCRα/δ-B contained three SNPs in the *hADV14S1* gene, together with SNPs in *AV8S6* and *AV17S1* and covered ~80 kb. TCRα/δ-C contained two SNPs in the *AV27S1* gene, together with SNPs in *hADV29S1*, *AV30S1* and *AV26S2* and covered ~50 kb.

Each of the clusters was examined for common haplotypes, and the frequencies compared with random by simulation. Haplotypes were observed in each cluster that appeared at frequencies higher than expected at random (Table 2). A significant reduction in haplotype diversity was also observed, as in each case there were far fewer haplotypes than were possible (Table 2).

Table 2. Haplotype diversity in the TCRα/δ-A, TCRα/δ-B and TCRα/δ-C disequilibrium clusters

	Observed	Expected ± SEM
TCRα/δ-A		
Most frequent haplotype	0.269	0.165 ± 0.011
Next frequent haplotype	0.244	0.132 ± 0.008
Distinct haplotypes	21	26.19 ± 1.266
w/ frequency > 0.01	11	12.10 ± 1.180
w/ frequency > 0.05	5	7.97 ± 0.171
TCRα/δ-B		
Most frequent haplotype	0.285	0.171 ± 0.012
Next frequent haplotype	0.203	0.116 ± 0.008
Distinct haplotypes	19	31.11 ± 0.833
w/ frequency > 0.01	9	21.70 ± 1.485
w/ frequency > 0.05	5	5.61 ± 0.780
TCRα/δ-C		
Most frequent haplotype	0.379	0.362 ± 0.012
Next frequent haplotype	0.235	0.131 ± 0.007
Distinct haplotypes	15	22.63 ± 1.378
w/ frequency > 0.01	9	11.54 ± 0.992
w/ frequency > 0.05	5	4.41 ± 0.591

The frequencies of common haplotypes are compared with the expected values derived from100 simulations (assuming no disequilibrium).

The TCR repertoire is not random, and is influenced by germ-line (genomic) polymorphism (19–21). It is not clear

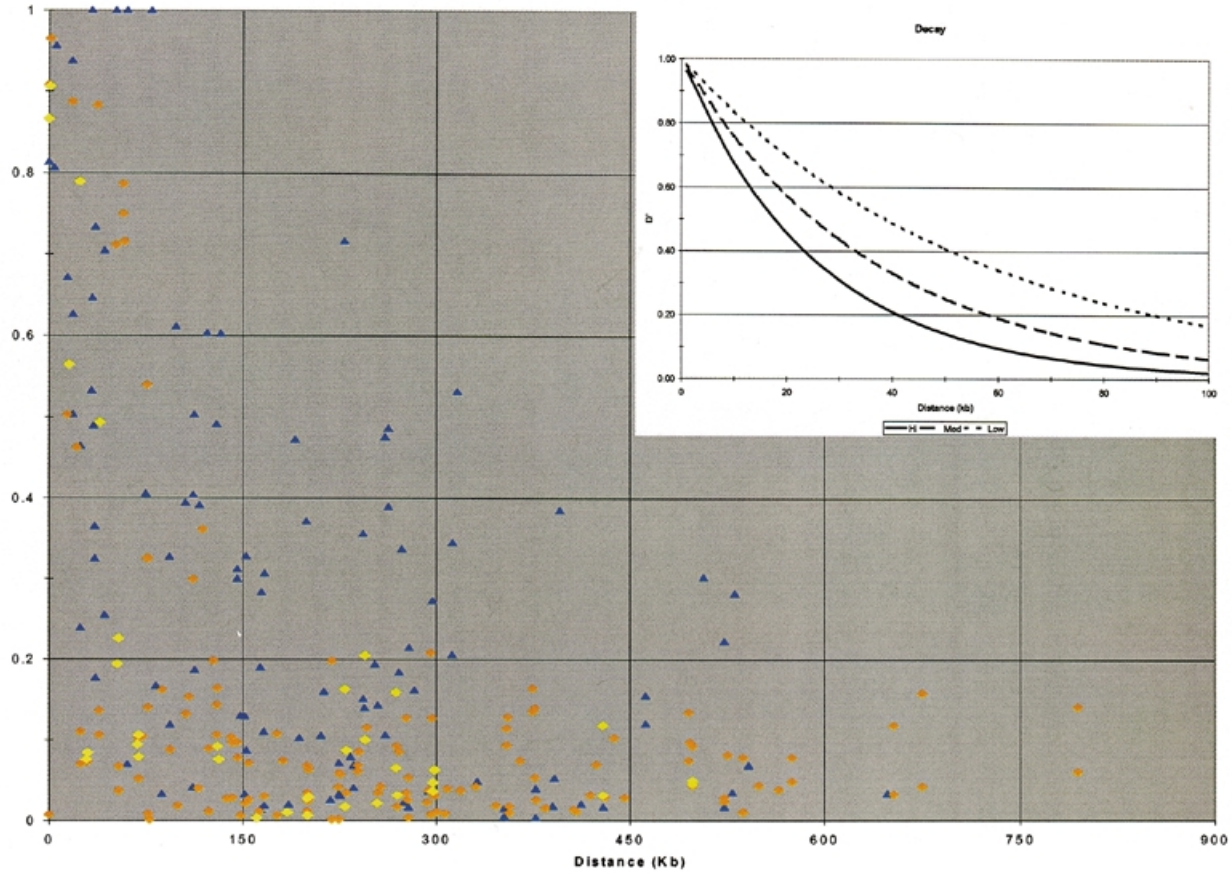

Figure 3. Decay of disequilibrium for different SNP types. Pairs in blue include one SNP with a rare allele (<20%). Pairs in orange include SNPs with moderately common alleles (<30%). Yellow pairs include SNPs with two very common alleles (>30%). Microsatellite markers were excluded from this analysis. The inset shows the results of a model for the decay of LD fitted to the data for each of the three sets of markers.

whether the variation in repertoire arises from polymorphism within the V segments themselves, or with regulatory or other sequences. The recognition of the common TCR α/δ haplotypes (Table 3) may therefore assist in the exploration of disease-associated variation within the locus.

Linkage disequilibrium with microsatellite alleles

Microsatellite repeats are common throughout the genome, and a high density microsatellite map is well established. The utility of tests of association with microsatellite alleles has not been examined in any detail. It is of interest therefore that the microsatellite D14S50 is present within the TCRα/δ-A cluster. Examination of the repeats within each of the three common TCRα/δ-A haplotypes showed three distinct patterns of distribution (Fig. 5). Each distribution contained a central dominant allele, with less common alleles spread on either side. These findings are consistent with mutation from a founder haplotype containing a single microsatellite allele, and suggest that microsatellites may have a higher mutation rate than that of SNPs. The overlap in distributions means that a single microsatellite allele may represent several haplotypes (such as allele 4 in Fig. 5), potentially (but not invariably) confounding the detection of association between a microsatellite allele and any disease-causing variant with which it may be in LD. The high mutation rate and the overlap in distribution both suggest that the detection of association with microsatellite alleles is likely to be more problematic than with SNPs.

CONCLUSIONS

Overall, the results show that LD can extend considerable distances, and suggest that the simulation studies of Kruglyak (5) may have given results that are unduly pessimistic. The discrepancy may be due to several factors. It is possible that positive selection within the population at large may have taken place for particular combinations of TCR α/δ V segments. However, LD that extends to distances similar to the present findings has been observed previously in a more limited study (22). It is also possible that the model of population growth used for simulation studies is unrealistic, particularly as it has been suggested that the comparatively recent history of humanity has been that of relatively isolated tribes characterized by high levels of endogamy and inbreeding (23). The families were selected for the presence of allergic disease, but the unrelated individuals studied for LD contained a high proportion of normal individuals, and haplotype frequencies did not differ between normal and allergic subjects. Whatever the reason for the discrepancy with simulated models, similar factors affecting the distribution of LD are likely to act around many genes predisposing to common diseases, and estimations

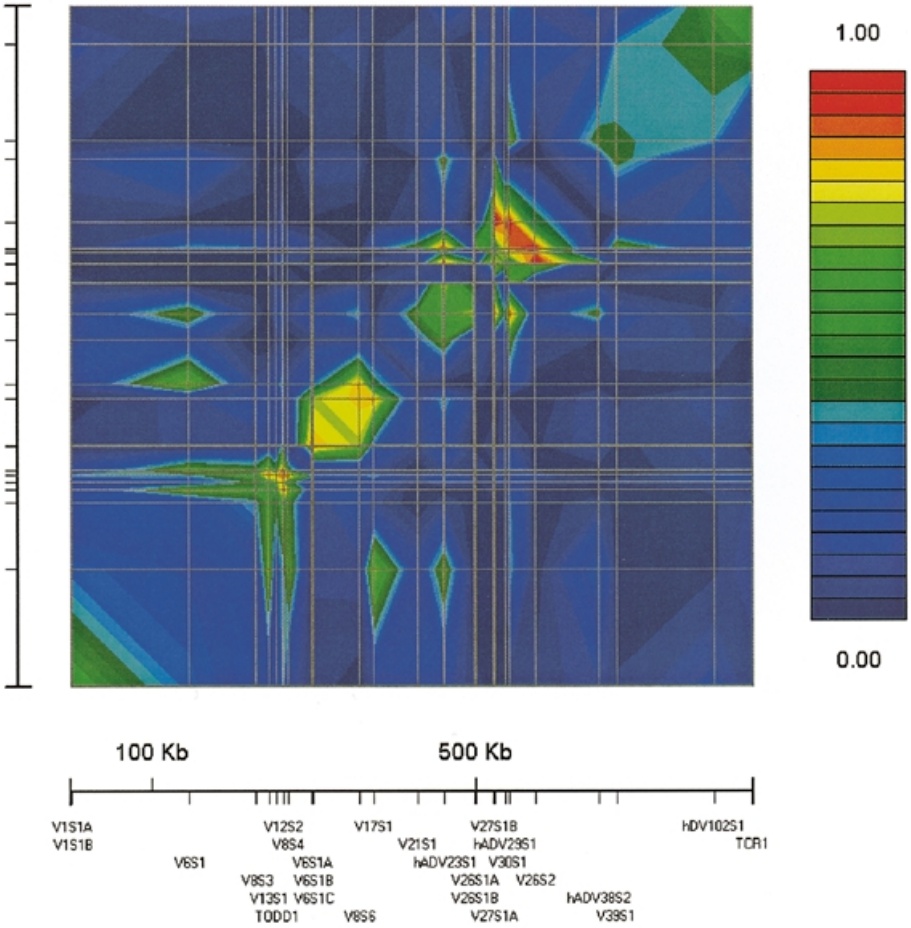

Figure 4. Distribution of pair-wise linkage disequilibrium (D′) across the TCR locus. The axis of the region runs along the diagonal from the bottom left to the top right of the figure. Multiple pair-wise comparisons between markers are shown, with pair-wise disequilibrium statistics colour coded (bright red and dark blue are opposite ends of the scale) and plotted at the marker locations. The plot was completed by interpolation.

based on LD of the desirable density of the SNP map may be reduced by at least a factor of 10.

Although the extent of LD observed in this study may be encouraging for systematic LD mapping of disease genes, the highly irregular fine structure of LD observed in the TCR α/δ locus suggests a different set of problems. The success of genome-wide screens for genetic linkage to many disorders has relied on the comparatively even distribution of recombination events which have occurred within very few generations. The distribution of LD on the other hand is a function of recombination and stochastic mutation over thousands of generations. If, as is likely, LD in the TCR α/δ locus is typical of that found elsewhere in the genome, effective LD mapping strategies may require a systematic knowledge of the evolution of LD relationships between individual markers.

MATERIALS AND METHODS

Two panels of subjects were studied. Panel A consisted of 410 Caucasian subjects within 88 nuclear families sub-selected for the presence of atopic disease from an Australian random population sample of 230 families (24). Panel B contained 410 Caucasian British individuals from 66 nuclear and five extended pedigrees ascertained through members with asthma or rhinitis (24).

SNPs in AV/DV gene segments were genotyped either by restriction fragment length polymorphism (RFLP) analysis or sequence specific oligonucleotide (SSO) probing (http://www.well.ox.ac.uk/asthma/TCR). In general, the choice of method was initially dependent on sequence availability. The latter also influenced the PCR strategy with some AV/DV gene segments being amplified with a semi-nested PCR approach to obtain gene specificity (8). Detailed PCR conditions are given on the web page. In general reactions were performed with 100 ng of DNA, 200 µM each dNTP, 0.5 µM each primer and 0.45–1 U of *Taq* DNA polymerase. Magnesium chloride concentration varied and is as stated on the web page. When semi-nested PCR was necessary, a 1/100 dilution of the first round product was made and 1 µl used in the second round PCR. For all PCRs, standard cycling conditions were implemented with variation in annealing temperature dependent on the PCR primers. On occasion a hot-start was found to be necessary. Typically 27–40 rounds of amplification were adopted for first round PCRs with 15 cycles for any second round amplifications.

Table 3. Most frequent haplotypes in each disequilibrium cluster

TCRα/δ-A								
	Frequency							
Rank	UK	AUS	Total	AV6S1	AV8S3	AV13S1	AV12S2	AV8S4
1	0.26	0.28	0.27	1	1	1	1	2
2	0.24	0.25	0.24	1	2	2	1	1
3	0.13	0.15	0.14	1	1	1	1	1
4	0.06	0.07	0.06	1	1	2	1	1
5	0.08	0.04	0.06	1	2	1	1	2
TCRα/δ-B								
	Frequency			hADV14S1	hADV14S1	hADV14S1		
Rank	UK	AUS	Total	A	B	C	AV8S6	AV17S1
1	0.27	0.32	0.29	1	1	1	1	1
2	0.24	0.17	0.20	2	1	2	2	1
3	0.19	0.17	0.18	1	2	1	1	1
4	0.10	0.12	0.11	2	1	2	1	2
5	0.06	0.05	0.06	1	1	1	1	2
TCRα/δ-C								
	Frequency			AV27S1	AV27S1			
Rank	UK	AUS	Total	A	B	hADV29S1	AV30S1	AV26S2
1	0.39	0.37	0.38	1	1	1	1	1
2	0.25	0.22	0.24	1	1	1	1	2
3	0.15	0.16	0.15	2	1	2	1	1
4	0.08	0.08	0.08	2	1	1	1	1
5	0.07	0.07	0.07	1	2	2	1	1

Haplotype frequencies are given for the British (UK) and Australian (AUS) subjects, as well as for the pooled data. The rarer allele is signified as 2 in each instance.

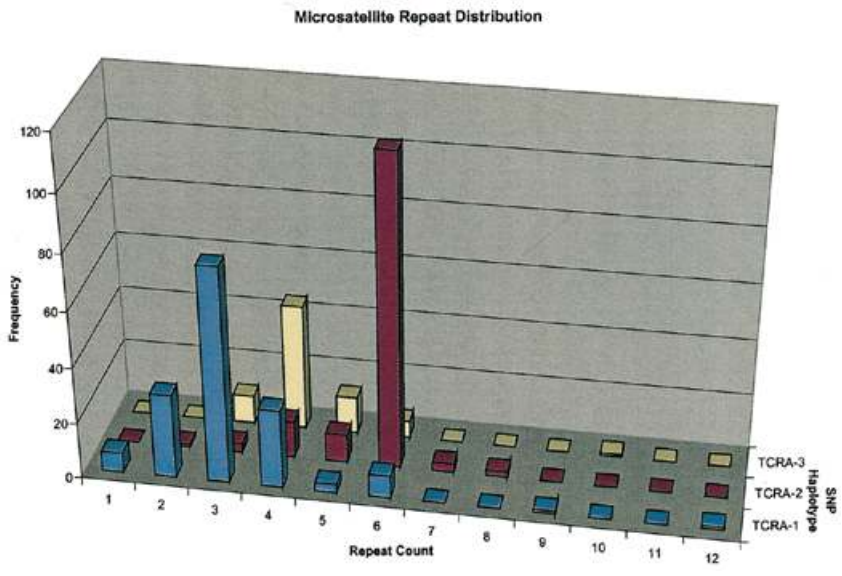

Figure 5. Distribution of D14S50 microsatellite alleles within common TCRα/δ-A cluster SNP haplotypes. The distribution of the frequency of individual microsatellite alleles is shown for each of the three most common TCRα/δ-A cluster haplotypes.

For SSO analysis, 25 µl PCRs were performed. An aliquot (5 µl) of each reaction was run on an agarose gel to verify successful amplification. The remaining 20 µl were used to create dot blot filters for SSO hybridizations. Non-radioactive SSO probing was performed as described previously (25). Sequences of allele specific probes and stringent wash temperatures used are given on the web page. Sequenced controls of all possible genotypes were included on each filter.

For RFLP analysis, 15 µl PCRs were performed. Successful amplification was assessed by agarose gel electrophoresis of 5 µl of product. This was done for five or 10 individuals from each 96-well plate of amplifications. Digests were then set up using 3–5 µl of PCR product in a final digest volume as specified on the web page. Details of the restriction enzyme used, the number of units per digest and the incubation temperature are also given on the web page. Time of digestion varied between 1 and 2 h. Restriction digestion fragments were resolved either on 2–3%(w/v) agarose gels or 4%(w/v) agarose gels (3:1 NuSieve GTG:LMP agarose) depending on the size of products. Sequenced controls were included in each set of digests. Microsatellites were typed as described previously (24).

Genotypes for all SNPs were checked independently by two individuals.

The nomenclature of the Vα/δ gene segments has recently changed (13), so that both the old and new gene names are given in Table 1. The new nomenclature, in which V segments are numbered consecutively according to their position, is used thereafter in the tables and text.

The most likely haplotypes for each pedigree were determined by the SimWalk2 program [(26) E. Sobel, 1998; ftp://watson.hgen.pitt.edu/pub/simwalk2], under the assumption of no linkage disequilibrium. For haplotyping, recombination fractions were estimated from physical distances with the approximation 1 centimorgan ≈ 1 megabase. As the number of untyped individuals was small, multiple runs of the SimWalk2 program with different random seeds gave similar results, as expected. The full set of estimated haplotypes used in this analysis is available at http://www.well.ox.ac.uk/asthma/TCR .

Using the haplotype set identified by SimWalk2, the standardized disequilibrium coefficient D′ was calculated as defined by Hedrick (16) and the strength of association was measured by a standard contingency table χ^2 test. Rare alleles (<7%) were pooled when analysing microsatellite markers. Estimates of D′ and significance were compared with those derived from maximum likelihood estimates of founder haplotype frequencies obtained by the EM algorithm of the Arlequin program (27).

The decay of disequilibrium was modelled assuming the simple model $D' = (1 - \theta)^b$, where decay of disequilibrium from its maximum value of 1 is expected to be a function of recombination fraction θ and a parameter β reflecting age of mutation and other factors, including drift and population size. To estimate β, we fitted the model $D' = (1 - \theta)^{\beta}(1 - \alpha) + \alpha$ by least squares, where α allows for the positive bias in measurement of D′.

Disequilibrium across the locus was plotted by the GOLD program (Abecasis, 1999; http://www.well.ox.ac.uk/asthma/GOLD). The horizontal and vertical axes were scaled according to physical distances between markers. For each marker pair, the pair-wise disequilibrium statistics were colour coded and plotted at their Cartesian co-ordinates. The plots were completed by interpolation.

The significance of reductions in haplotype diversity were evaluated by simulation. Random haplotypes were generated under the assumption of linkage equilibrium by randomly shuffling alleles across all haplotypes. These random haplotypes were then used to determine the number of haplotypes, and the frequency of the most common haplotype that might be observed, conditional on the sampled allele frequencies and in the absence of disequilibrium. The *P* values from 100 simulations were reported.

ACKNOWLEDGEMENTS

The study was funded by the Wellcome Trust and the National Asthma Campaign. We are grateful to Deborah Nickerson for advice on the sequence of the TCR α/δ locus and to Lon Cardon for many helpful discussions on the analyses of SNP data.

REFERENCES

1. Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.
2. Lander, E.S. (1996) The new genomics: global views of biology. *Science*, **274**, 536–539.
3. Collins, F.S., Guyer, M.S. and Chakravati, A. (1997) Variations on a theme: cataloging human DNA sequence variation. *Science*, **278**, 1580–1581.
4. Lai, E., Riley, J., Purvis, I. and Roses, A. (1998) A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. *Genomics*, **54**, 31–38.
5. Kruglyak, L. (1999) Prospects for whole-genome mapping of common disease genes. *Nature Genet.*, **22**, 139–144
6. Boysen, C., Carlson, C., Hood, E., Hood, L. and Nickerson, D.A. (1996) Identifying DNA polymorphisms in human TCRA/D variable genes by direct sequencing of PCR products. *Immunogenetics*, **44**, 121–127.
7. Wei, S., Charmley, P. and Concannon, P. (1997) Organization, polymorphism, and expression of the human T-cell receptor AV1 subfamily. *Immunogenetics*, **45**, 405–412.
8. Cornélis, F., Pile, K., Loveridge, J., Moss, P., Harding, R., Julier, C. and Bell, J. (1993) Systematic study of human αβ T cell receptor V segments shows allelic variations resulting in a large number of distinct T cell receptor haplotypes. *Eur. J. Immmunol.*, **23**, 1277–1283.
9. Reyburn, H., Cornélis, F., Russell, V., Harding, R., Moss, P. and Bell, J. (1993) Allelic polymorphism of human T-cell receptor Vα gene segments. *Immunogenetics*, **38**, 287–291.
10. Hauser, S.L. (1995) T-cell receptor genes. Germline polymorphisms and genetic susceptibility to demyelinating diseases. *Ann. N. Y. Acad. Sci.*, **756**, 233–240.
11. Ibberson, M., Peclat, V., Guerne, P.A., Tiercy, J.M., Wordsworth, P., Lanchbury, J., Camilleri, J. and So, A.K. (1998) Analysis of T cell receptor Vα polymorphisms in rheumatoid arthritis. *Ann. Rheum. Dis.*, **57**, 49–51.
12. Moffatt, M.F., Schou, C., Faux, J.A. and Cookson, W.O.C.M. (1997) Germline *TCR-A* restriction of immunoglobulin E responses to allergen. *Immunogenetics*, **46**, 226–230.
13. GenBank Accession numbers AE000658 AE000659 AE000660 AE000661 AE000662.
14. Lewontin, R.C. (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, **49**, 49–67.
15. Devlin, B. and Risch, N. (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, **29**, 311–322.
16. Hedrick, P.W. (1997) Gametic disequilibrium measures: proceed with caution. *Genetics*, **117**, 331–341.
17. Clark, A.G., Weiss, K.M., Nickerson, D.A., Taylor, S.L., Buchanan, A., Stengard, J., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E. and Sing, C.F. (1998) Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.*, **63**, 595–612.

18. Keavney, B., McKenzie, C.A., Connell, J.M., Julier, C., Ratcliffe, P.J., Sobel, E., Lathrop, M. and Farrall, M. (1998) Measured haplotype analysis of the angiotensin-I converting enzyme gene. *Hum. Mol. Genet.*, **7**, 1745–1751.
19. Gulwani-Akolkar, B., Posnett, D.N., Janson, C.H., Grunewald, J., Wigzell, H., Akolkar, P., Gregersen, P.K. and Silver, J. (1991) T cell receptor V-segment frequencies in peripheral blood T cells correlate with human leukocyte antigen type. *J. Exp. Med.*, **174**, 1139–1146.
20. Loveridge, J.A., Rosenberg, W.M.C., Kirkwood, T.B.L. and Bell, J.I. (1991) The genetic contribution to human T-cell receptor repertoire. *Immunology*, **74**, 246–250.
21. Moss, P.A.H., Rosenberg, W.M.C., Zintzaras, E. and Bell, J.I. (1993) Characterization of the human T cell receptor α-chain repertoire and demonstration of a genetic influence on Vα usage. *Eur. J. Immunol.*, **23**, 1153–1159.
22. Jorde, L.B., Watkins, W.S., Carlson, M., Groden, J., Albertsen, H., Thliveris, A. and Leppert, M. (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *Am. J. Hum. Genet.*, **54**, 884–898.
23. Thompson, E.A. and Neel, J.V. (1997) Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. *Am. J. Hum. Genet.*, **60**, 197–204.
24. Moffatt, M.F., Hill, M.R., Cornélis, F., Schou, C., Faux, J.A., Young, R.P., James, A.L., Ryan, G., le Souef, P., Musk, A.W. *et al.* (1994) Genetic linkage of T-cell receptor α/δ complex to specific IgE responses. *Lancet*, **343**, 1597–1600.
25. Moffatt, M.F. and Cookson, W.O.C.M. (1997) Tumour necrosis factor haplotypes and asthma. *Hum. Mol. Genet.*, **6**, 551–554.
26. Sobel, E. and Lange, K. (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am. J. Hum. Genet.*, **58**, 1323–1337.
27. Excoffier, L. and Slatkin, M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, **12**, 921–927.