Incorporating Mouse Genome
Mammalian Genome
Genes and Phenotypes

Prospects for whole genome linkage disequilibrium mapping in domestic dog breeds

Changbaig Hyun,[1,5] Lucio J. Filippich,[1] Rod A. Lea,[3,4] Graeme Shepherd,[3] Ian P. Hughes,[2] Lyn R. Griffiths[3]

[1]Companion Animal Sciences, School of Veterinary Science, University of Queensland, St Lucia, QLD 4072, Australia
[2]Immunogenetics Unit, School of Veterinary Science, University of Queensland, St Lucia, QLD 4072, Australia
[3]Genomics Research Centre, School of Health Science, Griffith University, Gold Coast, QLD 4111, Australia
[4]Institute for Molecular Systematics, School of Biological Sciences, Victoria University of Wellington, New Zealand
[5]Victor Chang Cardiac Research Institute, Level 6, 384 Victoria Street, Darlinghurst, NSW, 2010, Australia

Received: 22 January 2003 / Accepted: 22 May 2003

Abstract

Linkage disequilibrium (LD) mapping is commonly used as a fine mapping tool in human genome mapping and has been used with some success for initial disease gene isolation in certain isolated inbred human populations. An understanding of the population history of domestic dog breeds suggests that LD mapping could be routinely utilized in this species for initial genome-wide scans. Such an approach offers significant advantages over traditional linkage analysis. Here, we demonstrate, using canine copper toxicosis in the Bedlington terrier as the model, that LD mapping could be reasonably expected to be a useful strategy in low-resolution, genome-wide scans in pure-bred dogs. Significant LD was demonstrated over distances up to 33.3 cM. It is very unlikely, for a number of reasons discussed, that this result could be extrapolated to the rest of the genome. It is, however, consistent with the expectation given the population structure of canine breeds and, in this breed at least, with the hypothesis that it may be possible to utilize LD in a genome-wide scan. In this study, LD mapping confirmed the location of the copper toxicosis in Bedlington terrier gene (*CT-BT*) and was able to do so in a population that was refractory to traditional linkage analysis.

A number of problems are associated with the identification and isolation of disease-causing genes in canine populations. The first is the fact that in comparison with human, mouse, and some other domestic species, the genome map of the dog is relatively sparse. It is only since 1997, when Mellersh et al. (1997) published a 14-cM resolution linkage map, that the canine map has become a viable resource in gene identification. The canine map continues to improve, however, with the publication of a 10-cM resolution map (Neff et al. 1999) and a map containing a total of 1800 markers covering more than 90% of the dog genome (Breen et al. 2001). Notwithstanding this, the failure to detect linkage in studies by Dukes-McEwan and Jackson (2002) of certain multifactorial genetic diseases prompted this group to conclude that more detailed linkage maps were required. The second problem relates to the difficulty in obtaining sufficient numbers of informative matings from the canine population under investigation to perform a traditional linkage analysis. Matings are informative only if at least one parent is heterozygous at both disease and marker loci. To calculate Lod (Log of odds) scores for linkage, it is also a requirement that the relationships between all members of the pedigree are known. In addition, it is preferable if three generations of dogs are involved in the study to allow for the linkage phase of alleles to be determined. Thus, an alternative approach that does not rely on pedigree information would be ideal. Linkage disequilibrium (LD) mapping is one such technique; in canine populations, this should prove to be a powerful genetic tool for the identification of disease genes.

In linkage disequilibrium, we observe a particular marker allele to be associated with the disease allele more (or less) frequently than one would expect from

Subscribers may view supplemental material to this article on the *Mammalian Genome* website: http://www.springerlink.com

Correspondence to: I. Hughes; E-mail: i.hughes@uq.edu.au

 DOI: 10.1007/s00335-003-3006-0 • Volume 14, 640–649 (2003) • © Springer-Verlag New York, Inc. 2003

their individual frequencies. LD mapping is, thus, the identification of disease loci based on frequency differences in marker alleles between affected and unaffected groups within the population (Sheffield et al. 1998; Ardlie et al. 2002). The assumption (for the purposes of mapping) is that this association is due to individuals sharing haplotypes of disease and marker alleles that are identical because they were inherited from a common ancestor; that is, the haplotypes are identical by descent (IBD). The probability of haplotypes being IBD (rather than identical in state, IBS) is dependent upon a large number of factors related to the nature of the markers and disease locus, and population parameters such as size, growth dynamics, and history, as well as stochastic processes such as mutation and recombination. The effects of these factors will be discussed in more detail below, but for a thorough analysis of LD from the point of view of haplotype sharing, see Nolte and te Meerman (2002) and Nordborg and Tavaré (2002).

LD may arise, although rarely, because of a selective advantage conferred by the combination of two particular alleles at different loci. Most linkage, disequilibria arise, either after a mutational event, (e.g., normal allele to disease allele) or due to founder effect in which a single rare (disease) allele is included within a small founding population. In both cases, the disease allele will be, initially, associated with only one particular set of marker alleles (Terwilliger and Weiss 1998; Ardlie et al. 2002). Over the subsequent generations, in the absence of other influences, this LD will be broken down, largely owing to recombination events, and as such will disappear proportionally with the linkage distance between the disease locus and marker loci. Over many generations, linkage disequilibrium will remain only between the disease locus and very closely linked marker loci (Kruglyak 1997; Liu 1998). Detection of such a region is the basic strategy in gene localization by linkage disequilibrium mapping (Durham and Feingold 1997; Peltonen and Uusitalo 1997).

For LD to be of most use in a low-resolution, genome-wide scan (with markers spaced at 5–20 cM) for the detection of disease genes, the population in question should have certain structural and historical characteristics. These include the following: It is genetically isolated and originated from a relatively small number of founders (Wang et al. 1997; de la Chapelle and Wright 1998; Kruglyak 1999; Wright et al. 1999). Population expansion took place by growth rather than by immigration (de la Chapelle and Wright 1998; Kruglyak 1999). The mutation is recent, or the disease allele included in a founding population is rare in the parent population (Peltonen and Uusitalo 1997). Finally, the population under study is young, probably less than 100 generations (Wang et al. 1997; de la Chapelle and Wright 1998; Sheffield et al. 1998; Wright et al. 1999; Service et al. 2001).

These criteria can be understood if one views useful LD in terms of the genealogy of haplotypes and the probabilities of IBD. A small number of founders and small populations in general favor a faster coalescence to an ancestral haplotype because there are fewer potential parents in each generation. As there are fewer chances for mutation or recombination to occur, the probability of IBD is increased, and the extent of the shared haplotype is increased. Faster coalescence is also true for young populations. Small populations also generate LD due to genetic drift, but are particularly useful for mapping, there being less allelic heterogeneity, particularly among disease alleles. Rare haplotypes (which might include a rare disease allele) are more likely to be IBD. As there are fewer of them, they coalesce more quickly, leaving less time for mutation or recombination to create an IBS haplotype. While rare alleles and small population sizes increase the extent of an IBD region and the probability of IBD, it should be pointed out that they are also more influenced by genetic drift, which thus increases the variance of these measures (Nolte and te Meerman 2002; Nordborg and Tavaré 2002). Finally, it has been shown that a rapidly expanding population should be ideal for LD mapping and that, as haplotypes will coalesce faster (fewer parents in each previous generation), the probability of IBD will be higher than in stable populations (Service et al. 2001; Nolte and te Meerman 2002).

Certain human populations have been shown to be particularly suited to linkage disequilibrium mapping. The Finnish population has been the focus of many linkage disequilibrium mapping projects (reviewed in de la Chappelle and Wright 1998), as have the Bedouin-Arabs of the Negev (Sheffield et al. 1998). Very young founder populations such as those in Costa Rica, Quebec, and Newfoundland have been cited as amenable to LD mapping with a relatively sparse set of markers (Service et al. 1999, 2001). For a discussion on the effects of population demographics on LD, see Wright et al. (1999), Chapman and Thompson (2001), Service et al. (2001), Nordborg and Tavaré (2002), and Ardlie et al. (2002).

Pure-bred dog populations (breeds) fulfill the above criteria because they are genetically isolated sub-populations derived from a relatively small number of founding ancestors and are of a relatively recent origin (Vila et al. 1999). Most breeds were founded 30–120 years ago (Wilcox and Walkowicz

Table 1. Microsatellite markers used in this study

Marker/Repeat	*Sequence*	T_a (°C)[a]
FH2537/tetra	5′ TETAA AAA G TG TAG AGC TTT CTT CAA A 3′ 3′ ATT GAG ACC CAA GAG TGT TAG TG 5′	60
C10.781/(AG)12-14	5′ TET AC CTC CAA GAT GGC TCT TGA 3′ 3′ ACG TCG AGC TCC TGG CAT 5′	60
FH2293/(GAAA)43	5′ TET GA ATG CCC TTC ACC TTG AAA 3′ 3′ AGG AAA AGG AGA GAT GAT GCC 5′	60
C10.16/(AC)13	5′ HEX TT CTT GCT TCT TGA AGT AAG CC 3′ 3′ GAG TTC CAG ATC GAG TCC CA 5′	60
FH2422/tetra	5′ HEX TT GCC CGT CCT ATA CTC CTG 3′ 3′ CCA CAT GAT TTC ACT TGT ATA TGG 5′	60
FH2523/tetra	5′ HEX GG TTT AGT GCC AGC CTT CAG C 3′ 3′ TTC TAG CAG CCG GGA GTT TAT G 5′	60
C04107[b]/(GT)6GA(GT)11	5′ HEX TCA GCA ACT ATA CAT TTA AGA GGA 3′ 3′ CTG TCC CAT CTA AAG GAT AGG 5′	60
C10.602/(TC)15(CA)19	5′ FAM ACG AAT GAA AAC GGA GCA GCA 3′ 3′ CGA TAT TTT TCT CTC CCA CCC 5′	55–60
FH2339/tetra	5′ FAM TC CTT ATG ATA GTT TCC CTG TCT C 3′ 3′ CAA CTA ACA CAC CCA TCA CTT C 5′	60
C10.865/di-repeat	5′ HEX AG TGT ATG TAA GCC TGG AG 3′ 3′ TAA CTG ATG TTA TCA CTC TCT GC 5′	60

[a]Annealling temperature.
[b]For marker *C04107*, radioisotope labeling with α-^{32}P dCTP or γ-^{33}P ATP was also used.

1989), which corresponds to only 15–60 generations. The occurrence of severe genetic bottle-necks in many breeds within the last 100 years has also produced a founder effect. Another pertinent observation is that many genetic diseases of dogs are found exclusively, or almost so, in one particular breed (Hughes 1998; Ostrander et al. 2000), thus indicating the rarity of these mutations in the ancestral canine population. The potential for LD mapping in pure-bred dogs has been recognized by Ostrander and Kruglyak (2000), and the existence of LD over several tens of cM has been demonstrated in cattle (Riquet et al. 1999; Farnir et al. 2000) and sheep (McRae et al. 2002), which have breed structures similar to some breeds of dogs. Other, more popular breeds may have larger effective population sizes and, in these breeds, the usefulness of LD mapping for genome-wide scans cannot be easily inferred.

This study is the first attempt to assess practically the utility of genome-wide LD mapping in pure-bred dogs, and we have used copper toxicosis in Bedlington terriers as the disease model. Copper toxicosis in Bedlington terriers (*CT-BT*) is an autosomal recessive disease in which copper accumulates in the liver owing to impaired biliary copper excretion. This disease has many similarities to human copper storage diseases, especially Wilson's disease. The causative gene for *CT-BT* has remained elusive, although recent studies have initially assigned *CT-BT* to a 4.6-cM region on canine Chromosome (Chr) 10p26 (van de Sluis et al. 2000) and subsequently identified *MURR1* to be, in all likelihood, the disease-causing gene (van de Sluis et al. 2002). One particular microsatellite marker on Chr 10, *C04107*, has been used diagnostically, as it has been shown to be in complete LD with *CT-BT* (Yuzbasiyan-Gurkan et al. 1997). In fact, it has been demonstrated recently that *C04107* is located within the first intron of *MURR1* (van de Sluis et al. 2002).

The initial aim of this study is to correctly locate the *CT-BT* locus by the use of Chr 10 microsatellite markers and LD mapping, and secondly to determine over what distance LD can be detected. The study was also performed to compare LD mapping and linkage analysis as methods of gene localization in this population. Two methods of LD mapping were used. The first was to simply compare marker allele frequencies between normal and diseased groups by use of 2 × 2-table χ^2 test of association. The second was to calculate an estimate of the LD called D′ with an associated probability. D′ varies from 0 (no LD) to 1 (complete LD). For details, see Materials and methods section and Tables 1 and 2.

Materials and methods

Dogs. One hundred and thirty-one pure-bred Bedlington terriers (54 males and 79 females), aged between 4 weeks and 13 years, from eight different Bedlington terrier pedigrees were used in this study. Six dogs were eliminated because of non-Mendelian inheritance of marker alleles. Of the remaining 125 dogs, 24 were diagnosed with copper toxicosis based on the basis of a combination of results from blood

and hepatic copper tests, liver enzymes tests, and liver histopathology from biopsies (Hyun and Filipich, unpublished). It is believed that the vast majority of the 200 to 300 Bedlington Terrier dogs in Australia today are descendants of six dogs imported from the United Kingdom between 1963 and 1967 (Robertson et al. 1983).

Microsatellite markers. Ten microsatellite markers on Chr 10 were selected. (Fig. 1). These markers were numbered 1 to 10 from the centromere. Forward primers of each marker were labeled with FAM, TET, or HEX fluorescent dyes (PE Applied Biosystems, USA) (Table 1).

PCR reaction. DNA was isolated from canine whole blood by the method of Lahiri and Nurnberger (1991) or with the QIAamp DNA Mini kit (Qiagen, Germany) according to the manufacturer's instructions. Following extraction, DNA concentration was estimated by absorbance at 260 nm, and the solution was assayed for purity using the A_{260nm}/A_{280nm} absorbtion ratio. Following quantification, samples were diluted to a working concentration of 20 ng/μL. PCRs were then performed in a PC-960 air-cooled thermal cycler (Corbett Research, Australia). A 12.5-μL total volume was used, containing 20 ng genomic DNA, 25 μM of each dNTP, 200 nM of each forward and reverse primer, 1× PCR buffer (100 mM Tris-Cl, pH 8.4, 500 mM KCl, 0.01% gelatin), l0 μg acetylated BSA, 0.5 unit *Taq* polymerase, and 1.5 mM $MgCl_2$. DNA was initially denatured at 94°C for 4 min and then subjected to 35 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min, followed by a final extension at 72°C for 7 min.

Marker genotyping. PCR product DNA concentration was standardized by dilution (up to 10-fold) after electrophoresis of l μL of product on a 1% agarose gel. One μL of standardized PCR product was then mixed with 12.5 μL of standard TAMRA-labeled dye/formamide solution containing a 100-bp to 1000-bp ladder (Perkin Elmer, USA), denatured at 94°C for 4 min and then snap frozen for 5 min. Finally, the PCR products were separated by laser-induced fluorescence capillary electrophoresis by using an ABI Prism 310 Genetic Analyzer (Perkin Elmer, USA). Genotypes were then determined by reference to known size standards.

Linkage and LD analysis. Pedigree information and Chr 10 haplotypes for the 10 markers for 131 pure-bred Bedlington terriers were recorded by using Cyrillic® (Ver 2.1.3, Pedigree make-up program, Cherwell Scientific Publishing Ltd, UK). For para-

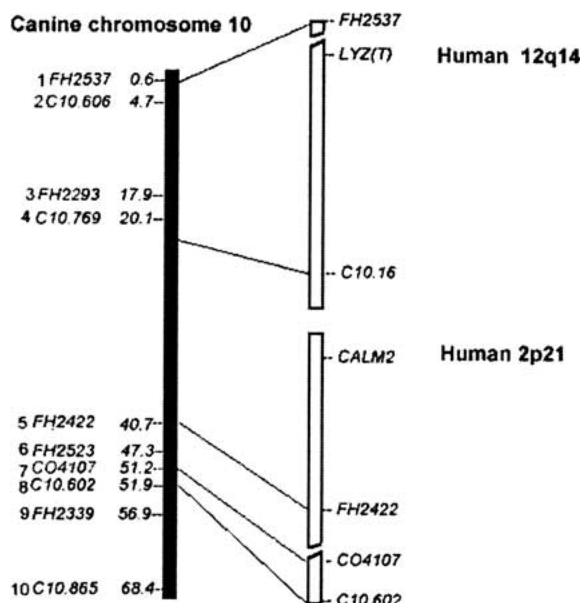

Fig. 1. Canine Chr 10. Based on Neff et al. (1999).

metric linkage analysis, two-point linkage analysis was performed with the linkage analysis software package, FASTLINK (Lathrop et al. 1984; Cottingham et al. 1993; Schäffer et al. 1994). The analysis was run under a recessive mode of inheritance and assumed isofrequent alleles. The GENEHUNTER program (Kruglyak et al. 1996; Kruglyak and Lander 1998; http://linkage.rockefeller.edu/soft/gh/) was also used to run a multi-point linkage analysis.

LD analysis was performed in two ways. First, the probability of an association between a particular allele and the presence of copper toxicosis was calculated. A 2 × 2-table χ^2 test of association was used to compare the frequency of the allele between affected and clinically normal groups. As most markers were multiallelic, the most associated allele was selected, and other alleles collapsed into a single, alternative class. This was done to ensure that expected frequencies in each cell were large enough to perform a valid χ^2 test. Following convention in linkage analysis (see Chotai 1984), a P value <0.0001 was considered significant evidence for linkage.

The second approach was to calculate the LD between each marker locus and the diagnostic marker locus, *C04107*. LD scores were calculated as follows. For two polymorphic loci A and B, let A_i be an allele at the A locus and B_j be an allele at the B locus. If the relative frequencies of A_i and B_j are p_i and p_j and the relative frequency of gamete A_iB_j, is p_{ij}, then the linkage disequilibrium between the alleles, D_{ij}, can be defined as

$$D_{ij} = p_{ij} - p_i p_j$$

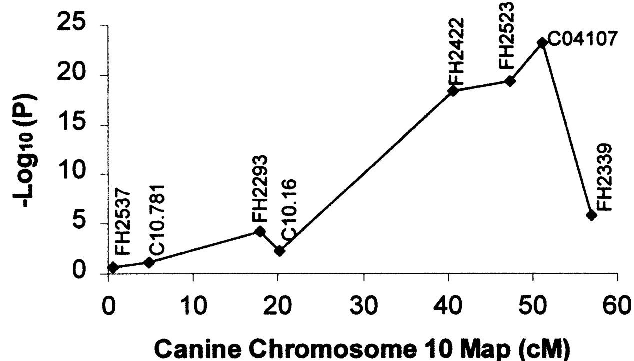

Fig. 2. Association χ^2 probabilities from Table 3 plotted (as $-\log_{10}P$) against map positions of markers indicating the most likely location of *CT-BT*.

being the difference between the observed frequency of A_iB_j and that expected under the assumption of independence. A more useful measure of LD between a particular pair of alleles is the standardized disequilibrium measure, which is defined as

$$D'_{ij} = \frac{D_{ij}}{D_{\max}}$$

where

$$D_{\max} = \min[p_ip_j, (1-p_i)(1-p_j)] \text{ when } D_{ij} < 0 \text{ or}$$

$$D_{\max} = \min[p_i(1-p_j), (1-p_i)p_j] \text{ when } D_{ij} > 0$$

(Lewontin 1964; Hedrick 1987). An overall measure of the disequilibrium between two loci, taking into consideration all alleles, is D', which is defined as

$$D' = \sum_{i=1}^{m}\sum_{j=1}^{n} p_ip_j|D'_{ij}|$$

where m is the number of alleles at the A locus and n is the number of alleles at the B locus. D' has a range from 0 (linkage equilibrium) to 1 (complete linkage disequilibrium) and was the measure of LD used in this study. It was calculated by using computer programs PM (Xie and Ott 1993; Zhao et al. 2000) and 2LD available from Jin Hua Zhao (http://www.iop.kcl.ac.uk/IoP/Departments/PsychMed/GEpiBSt/software.stm). The 2LD program also gave a standard deviation of the D' estimate and a probability of observing the calculated D' based on a χ^2 analysis of the observed haplotype frequencies compared with those expected under the hypothesis of linkage equilibrium (i.e., $D' = 0$).

Results

Markers. The number of alleles and their sizes for each Chr 10 marker is shown in Table 2. Of the ten

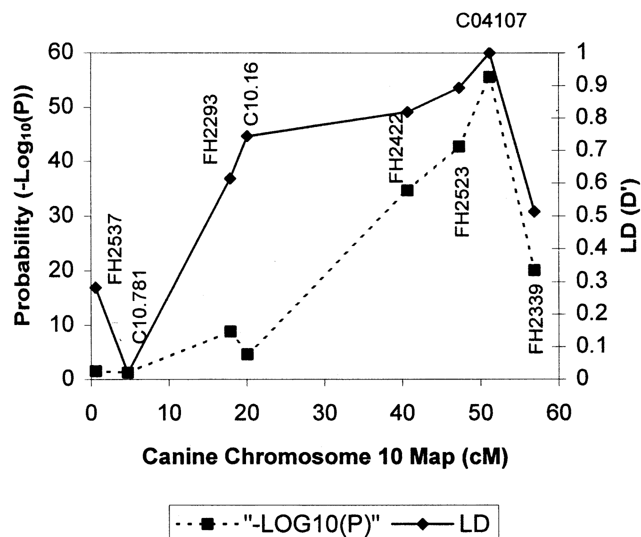

Fig. 3. D′ estimates for markers on canine Chr 10 and associated probabilities indicating the most likely location of *CT-BT*.

markers tested, *C10.602* and *C10.865* were monomorphic in this Bedlington terrier population.

Linkage and LD analysis. Traditional (parametric) linkage analysis was unable to detect linkage between the marker loci and disease locus in our Bedlington terrier population. This, initially surprising, result arose because, of all the matings in the population, none was informative. That is, either the mating type could not detect segregation at the *CT-BT* locus or the haplotype and disease status information was missing from a parent.

With respect to LD and χ^2 association analyses, six dogs were excluded because they showed non-Mendelian inheritance patterns, possibly owing to genotyping errors or incorrect pedigree information. Association χ^2 probabilities for each marker are shown in Fig. 2. Significant (P<0.0001) allelic association was detected for markers *FH2293*, *FH2422*, *FH2523*, the diagnostic marker *C04107*, and *FH2339*. In general, it can be seen from Fig. 2 that, with the exception of marker *C10.16*, the closer the markers were to *CT-BT*, the more significant the allelic association became. It is also of particular note that marker *FH2293*, located some 33.3 cM from *CT-BT*, was seen to be significantly ($P = 6.37 \times 10^{-5}$) associated with disease status.

D′ values calculated between each marker and *CT-BT* (*C04107*) ranged from 0.02 (*C10.781*) to 1.00 for the diagnostic marker, *C04107*. Six markers showed D′ values significantly ($P < 0.0001$) greater than 0. Again, marker *FH2293*, 33.3 cM from *CT-BT*, showed significant LD, with the closer markers showing increased significance (Fig. 3).

Table 2. Alleles and chromosome 10 map locations of microsatellite markers

Microsatellite marker	*Map location (cM)*	*Number of alleles*	*Sizes of alleles (bp)*
FH2537	0.6	3	164, 168, 175
C10.781	4.7	2	182, 189
FH2293	17.9	2	232, 278
C10.16	20.1	3	177, 180, 182
FH2422	40.7	4	191, 195, 199, 203
FH2523	47.3	3	384, 388, 448
C04107	51.2	2	162, 166
C10.602	51.9	1	175
FH2339	56.9	4	668, 676, 680, 684
C10.865	68.4	1	135

Discussion

Owing to the population structure of domestic dog breeds, we surmized that linkage disequilibrium should exist over extensive regions of the canine genome and that, in the dog, genome-wide mapping methods based on linkage disequilibrium may be an effective and indeed a preferred strategy in many instances. This potential has also been recognized by Ostrander and Kruglyak (2000). To test this theory, we chose as our model copper toxicosis in the Australian Bedlington Terrier population.

Until recently, it was generally thought that whole-genome LD mapping was not feasible for most human populations. For example, Kruglyak (1999) estimated that approximately 500,000 single nucleotide polymorphism (SNP) markers would be required in a normal human population, because it was estimated that useful LD existed only over a distance of 3 kb. Even in isolated populations, simulations performed by Kruglyak (1999) suggested that the LD approach would not be useful unless the founding population consisted of 100 individuals or fewer, or the frequency of the variant was less than 5%. Given this, however, there are now numerous reports of human populations which, due to their structure, have been used or could potentially be used in genome-wide mapping strategies based on LD (e.g., Wang et al. 1997; de la Chapelle and Wright 1998; Sheffield et al. 1998; Gordon et al. 2000; Chapman and Thompson 2001; Mohlke et al. 2001; Reich et al. 2001). Of more significance, however, from the point of view of canine population structure, have been the observations that LD extends for tens of centimorgans in the Dutch black-and-white dairy cattle population (Farnir et al. 2000), in the Holstein-Friesian population (Riquet at al. 1999), and in Coopworth and Romney sheep (McRae et al. 2002). What these results tell us is that the structure and history of populations is very important with respect to LD mapping and that, although we may be able to identify certain attributes of a population that suggest that the population will be amenable to this approach (e.g., Ardlie et al. 2002), empirical sampling is generally required to confirm this. That is, the level of background LD and the extent of LD across the genome will be different in each breed. As a final cautionary note, however, are two assertions by Terwilliger; first, that the variance of LD is so large that one should expect numerous examples of spuriously high LD (Terwilliger 2001). And second, that the use of isolated populations is apt to produce false-positive associations as, by definition, affected individuals will be more related to each other than to other members of the population (Terwilliger and Weiss 1998).

An additional factor suggesting that genome-wide LD mapping should be effective in canine populations is that, because a less dense set of markers is required, microsatellite markers rather than SNPs can be used. Microsatellites are multi-allelic and have been shown, under all circumstances, to be more powerful than SNPs in detecting LD (Chapman and Wijsman 1998; Sham et al. 2000). Microsatellites were the preferred markers used in this study and in the LD studies on domestic animals noted above (Riquet et al. 1999; Farnir et al. 2000; McRae et al. 2002).

It was our intention to compare two mapping strategies based on LD; namely, association mapping and D′ calculation, with a traditional linkage analysis approach. Linkage analysis by traditional methods requires mating of affected to carrier dogs or carrier to carrier dogs to identify the segregation of the disease gene. However, no such matings occurred in the pedigrees sampled, although our sample population constituted approximately half of the Bedlington terrier population in Australia, collected over 3 years. While this precluded any analysis of the relative powers of the tests, it did demonstrate, very pointedly, that pure-bred dog populations segregating for a particular disease allele often do not conform to the requirements for a traditional linkage analysis.

Both the χ^2 association method and direct D′ calculation identified significant association or LD up to 33.3 cM from the *CT-BT* locus. LD decreased with increasing distance from *CT-BT* with the closest marker, *FH2523*, displaying a D′ of 0.89 and the most distant markers (*FH2537* and *C10.781*) with D′ values not significantly different from 0. The significance of these D′ values also, in general, increased with proximity to *CT-BT*, as did the significance of the probabilities of the association tests. These relationships are shown in Figs. 2 and 3.

An exception to this trend was that the *P* values associated with both χ^2 and D′ for marker *FH2293* located 33.3 cM from *CT-BT* were more significant than the *P* values for marker *C10.16* located closer (31.1 cM) to *CT-BT*. This result is a reflection of the nearly optimal allele frequencies of *FH2293* (0.4 and 0.6) compared with those of *C10.16* (0.184 and 0.816, after the three alleles were collapsed into two). Thus, while D′ increased as expected, the significance of this result (and the χ^2 *P* value) decreased owing to the decreased sample space available for marker *C10.16*.

Probabilities associated with D′ values were consistently more significant than equivalent probabilities for the χ^2 association test. This is probably a reflection of the fact that D′ calculations include information from all alleles at the same time and is a direct estimate of LD. Conversely, the χ^2 association test requires the collapse of a multiallelic system into a biallelic system and measures a consequence of the existence of LD.

The question of what measure of LD to use and, more importantly, how to calculate the variance of that measure and hence gain insight into its significance is an area of much thought and debate. Improvements on the χ^2 test of association have been suggested, for example, by Terwilliger (1995) and Choulakian and Mahdi (2000). For multiallelic markers, Ardlie et al. (2002) claim that there is no satisfactory measure of LD, with the *P* value of Fisher's exact test (or the Chi Square test) being used as a proxy in most studies. D′ values are difficult to compute meaningfully for multiallelic markers, tend to be inflated by small sample sizes and extreme allele frequencies, and could be biased in populations of related individuals if haplotype frequencies are estimated with algorithms rather than directly counted (Zapata 2000; Ayres and Balding 2001; Zapata et al. 2001; Ardlie et al. 2002; Teare et al. 2002). An alternative to D′, called r^2 or Δ^2, is less affected by these factors but, again, is not easily applied to LD estimation from multiallelic markers (Ardlie et al. 2002; Teare et al. 2002). The present focus of much research is to develop new statistical procedures that are better able to represent and define the significance of associations resulting from linkage disequilibrium. Teare et al. (2002) showed that the bias in D′ values, or any other summary measure of LD, due to sample size and allele frequencies can be corrected to some extent by the use of bootstrapping. Another resampling procedure, the Markov chain Monte Carlo method, was used by Ayres and Balding (2001) to approximate not just the variance but the entire distribution of D′ values. Terwilliger (2001) argues that genome scanning experiments should be viewed as estimate problems rather than hypothesis tests; that is, defining the accuracy of an estimate of a gene's location, rather than finding the probability of an association under the null hypothesis of there being no gene. Terwilliger (2001) discussed how a 3-lod-unit support interval may be estimated around a putative gene location. Cordell and Elston (1999) also took the approach of obtaining a confidence interval for the location of a disease gene. These workers transformed LD measures as a function of location to fit a quadratic curve. The peak of this curve estimated the disease gene location. In a different approach entirely, artificial neural networks have been proposed to use pattern recognition to identify differences in genotypes between cases and controls (Curtis et al. 2001; Sherriff and Ott 2001). Haplotype pattern mining is another method that uses pattern recognition but is instead applied to haplotypes and uses a data mining algorithm to identify disease-associated haplotypes (Toivonen et al. 2000).

Of the two methods used in this study, the χ^2 association test was the simpler but seemingly less powerful. D′ estimation is more complex and its interpretation less straightforward. Undoubtedly, the behavior of D′ values and their variances, or other measures of LD, will be further investigated, while the power of association studies will be improved with more sophisticated statistics (e.g., Choulakian and Mahdi 2000).

In this study, LD was detected over a considerable distance (33.3 cM). This is an encouraging result from the point of view of the potential utility of LD in a genome-wide scan, but should be viewed realistically for a number of reasons. It is known that LD is not uniform across the genome (Zavattari et al. 2000; Goldstein 2001; Service et al. 2001), and as such we cannot infer a general level of LD across the genome from this result. Indeed, the observation in our own data of the *FH2339* marker located 5.7 cM on the telomeric side of *CT-BT* having a smaller D′ than loci at further distances on the centromeric side would highlight this. It is also possible that we are witnessing a "hitch-hiking" effect due to selection occurring against the disease allele at the *CT-BT* locus or that the *CT-BT* mutation is very recent; that is, ignore recent than the genetic isolation of the Bedlington Terrier breed. Having said this, however, this result along with the population history of the breed and the results seen in other domestic animals would suggest that LD should exist throughout the genome, and the number of markers required for a genome-wide screen would be in the hundreds rather than the hundreds of thousands. The extent and nature of LD will also be different among breeds, reflecting both differences in population history and

the stochastic nature of LD. Our knowledge of the very narrow founder base of the Bedlington Terrier breed in Australia suggests that we might expect more extensive LD in this breed than in others. We are currently conducting a microsatellite-based, genome-wide LD analysis of a number of breeds of dog, including the Bedlington Terrier, that should better demonstrate the extent of LD in domestic dog breeds. It will also allow investigation of genome-wide background LD. This leads to the potential problem of false-positive associations, as noted by Farnir et al. (2000), that arise owing to population structure and history. Genetic drift, admixture, and stratification in particular are known to produce spurious associations. Essentially, a high background LD will confuse the interpretation of LD measures made in genome mapping studies, as most methods of LD analysis assume linkage equilibrium between markers on control chromosomes. Thus, for LD mapping of disease genes to be successful, the LD arising from association with a disease allele must be able to be differentiated from background LD (Service et al. 2001).

For smaller sample sizes, the distance over which LD or association could be detected will be decreased, and hence the number of evenly spaced markers required would need to be increased. Simulations conducted with random sub-samples of our data (not shown) indicated that, even with very small sample sizes, LD could be detected up to 10 cM from the disease locus.

In conclusion, this study has shown that in the Australian Bedlington Terrier population it is possible, by using a linkage disequilibrium approach, to correctly localize a disease locus, namely, the *CT-BT* locus. In this population, LD behaved exactly as would be predicted, decreasing with chromosomal distance from the disease locus. In fact, the nature of this particular population precluded the use of a traditional linkage approach, and as such LD was superior to linkage analysis for gene isolation in this case. These results suggest that it may be possible to use LD in such a dog population as an initial genome screen. It is recognized that the statistics and interpretation of LD analyses are still not well defined; however, there is every reason to believe that these problems will be resolved adequately in the near future.

Acknowledgments

This study was supported by research grants from the Canine Control Council and John and Mary Kibble Trust. We are very appreciative of the cooperation and support of the members of the Bedlington Terrier Club of Queensland and of the veterinary practitioners providing case materials.

References

1. Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3, 299–309
2. Ayres KL, Balding DJ (2001) Measuring gametic disequilibrium from multilocus data. Genetics 157, 413–423
3. Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C et al. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. Genome Res 11, 1784–1795
4. Chapman NH, Thompson EA (2001) Linkage disequilibrium mapping; the role of population history, size, and structure. Adv Genet 42, 413–436
5. Chapman NH, Wijsman EW (1998) Genome screens using linkage disequilibrium tests, optimal marker characteristics and feasibility. Am J Hum Genet 63, 1872–1885
6. Chotai J (1984) On the lod score method in linkage analysis. Ann Hum Genet 48, 359–378
7. Choulakian V, Mahdi S (2000) A new statistic for the analysis of association between trait and polymorphic marker. Math Biosci 164, 139–145
8. Cordell HJ, Elston RC (1999) Fieller's theorem and linkage disequilibrium mapping. Genet Epidemiol 17, 237–252
9. Cottingham Jr RW, Idury RM, Schäffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53, 252–263
10. Curtis D, North BV, Sham PC (2001) Use of a neural network to detect association between a disease and multiple marker genotypes. Ann Hum Genet 65, 95–107
11. de la Chapelle A, Wright FA (1998) Linkage disequilibrium mapping in isolated populations. The example of Finland revisited. Proc Natl Acad Sci USA 95, 12416–12423
12. Dukes-McEwan J, Jackson IJ (2002) The promises and problems of linkage analysis by using the current canine genome map. Mamm Genome 13, 667–672
13. Durham LK, Feingold E (1997) Genome scanning for segments shared identical by descent among distant relatives in identical populations. Am J Hum Genet 61, 830–842
14. Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N et al. (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10, 220–227
15. Goldstein DB (2001) Islands of linkage disequilibrium. Nat Genet 29, 109–111
16. Gordon D, Simonic I, Ott J (2000) Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. Genomics 66, 87–92
17. Hedrick PW (1987) Gametic disequilibrium measures, proceed with caution. Genetics 117, 331–341

18. Hughes I (1998) Diagnosis and control of genetic disease in small animals, Part I. Basic genetic principles and the diagnosis of genetic disease. Aust Vet Pract 28, 124–135
19. Kruglyak L, (1997) What is significant in whole-genome linkage disequilibrium studies. Am J Hum Genet 61, 810–812
20. Kruglyak L (1999) Prospects of whole genome linkage disequilibrium mapping of common disease genes. Nat Genet 22, 139–144
21. Kruglyak L, Lander ES (1998) Faster multipoint linkage analysis using Fourier transforms. J Comput Biol 5, 1–7
22. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and non-parametric linkage analysis. A unified multipoint approach. Am J Hum Genet 58, 1347–1363
23. Lahiri DK, Nurnberger JI (1991) A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleic Acids Res 19, 5444
24. Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81, 3443–3446
25. Lewontin RC (1964) The interaction of selection and linkage II. Optimal model. Genetics 50, 757–782
26. Liu B-H (1998) Two locus models: using linkage disequilibrium. In: *Statistical Genomics*, Chapter 8. (Boca Raton FL; CRC Press LLC), pp 241–272
27. McRae AF, McEwen JC, Dodds KG, Wilson T, Crawford AM et al. (2002) Linkage disequilibrium in sheep. Genetics 160, 1113–1122
28. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K et al. (1997) A linkage map of the canine genome. Genomics 46, 326–336
29. Mohlke KL, Lange EM, Valle TT, Ghosh S, Magnuson VL et al. (2001) Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns. Genome Res 11, 1221–1226
30. Neff MW, Bromann KW, Mellersh CS, Ray K, Acland GM et al. (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. Genetics 151, 803–820
31. Nolte IM, te Meerman GJ (2002) The probability that similar haplotpes are identical by descent. Ann Hum Genet 66, 195–209
32. Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history tells us. Trends Genet 18, 83–90
33. Ostrander EA, Kruglyak L (2000) Unleashing the canine genome. Genome Res 10, 1271–1274
34. Ostrander EA, Galibert F, Patterson DF (2000) Canine genetics comes of age. Trends Genet 16, 117–124
35. Peltonen L, Uusitalo A (1997) Rare disease genes—lessons and challenges. Genome Res 7, 765–767
36. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC et al. (2001) Linkage disequilibrium in the human genome. Nature 411, 199–204
37. Riquet J, Coppieters W, Cambisano N, Arranz J-J, Berzi P et al. (1999) Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. Proc Natl Acad Sci USA 96, 9252–9257
38. Robertson HM, Studdert VP, Reuter RE (1983) Inherited copper toxicosis in Bedlington Terriers. Aust Vet J 60, 235–238
39. Schäffer AA, Gupta SK, Shriram K, Cottingham Jr RW (1994) Avoiding recomputation in linkage analysis. Hum Hered 44, 225–237
40. Service SK, Lang DWT, Freimer NB, Sandkuijl LA (1999) Linkage-disequilibrium mapping of disease genes by reconstruction of ancestral haplotypes in founder populations. Am J Hum Genet 64, 1728–1738
41. Service SK, Ophoff RA, Freimer NB (2001) The genome-wide distribution of background linkage disequilibrium in a population isolate. Hum Mol Genet 10, 545–551
42. Sham PC, Zhao JH, Curtis D (2000) The effect of marker characteristics on the power to detect linkage disequilibrium due to single or multiple ancestral mutations. Ann Hum Genet 64, 161–169
43. Sheffield VC, Stone EM, Carmi R (1998) Use of isolated inbred human populations for identification of disease. Trends Genet 14, 391–396
44. Sherriff A, Ott J (2001) Applications of neural networks for gene finding. Adv Genet 42, 287–297
45. Teare MD, Dunning AM, Durocher F, Rennart G, Easton DF (2002) Sampling distribution of summary disequilibrium measures. Ann Hum Genet 66, 223–233
46. Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56, 777–787
47. Terwilliger JD (2001) On the resolution and feasibility of genome scanning approaches. Adv Genet 42, 351–391
48. Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping, fantasy or reality? Curr Opin Biotechnol 9, 578–594
49. Toivonen HTT, Onkamo P, Vasko K, Ollikainen V, Sevon P et al. (2000) Data mining applied to linkage disequilibrium mapping. Am J Hum Genet 67, 133–145
50. van de Sluis B, Kole S, van Wolferen M, Holmes NG, Pearson PL et al. (2000) Refined genetic and comparative physical mapping of the canine copper toxicosis locus. Mamm Genome 11, 703–705
51. van de Sluis B, Rothuizen J, Pearson PL, van Oost BA, Wijmenga C (2002) Identification of a new copper metabolism gene by positional cloning in a purebred dog population. Hum Mol Genet 11, 165–173
52. Vila C, Maldonado JE, Wayne RK (1999) Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. J Hered 90, 71–77
53. Wang C-Y, Hawkins-Lee B, Ochoa B, Walker RD, She J-X (1997) Homozygosity and linkage disequilibrium mapping of the urofacial (Ochoa) syndrome gene to a 1-cM interval on chromosome 10q23–q24. Am J Hum Genet 60, 1461–1467

54. Wilcox B, Walkowicz C (1989) Atlas of dog breeds of the world. (Neptune City, NJ: TFH, Publications)
55. Wright AF, Carothers AD, Piratstu M (1999) Population choice in mapping genes for complex diseases. Nat Genet 23, 397–404
56. Xie X, Ott J (1993) Testing linkage disequilibrium between a disease gene and marker loci. Am J Hum Genet 53, 1107
57. Yuzbasian-Gurkan V, Blanton SH, Cao Y, Ferguson P, Li J et al. (1997) Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington Terriers. J Am Vet Res 58, 23–27
58. Zapata C (2000) The D′ measure of overall gametic disequilibrium between pairs of multiallelic loci. Evolution Int J Org Evol 54, 1809–1812
59. Zapata C, Rodriguez S, Visedo G, Sacristan F (2001) Spectrum of nonrandom associations between microsatellite loci on human chromosome11p15. Genetics 158, 1235–1251
60. Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A et al. (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations, demography, chromosome recombination frequency and selection. Hum Mol Genet 9, 2947–2957
61. Zhao JH, Curtis D, Sham PC (2000) Model-free analysis and permutation test for allelic associations. Hum Hered 50, 133–139

Web Site References

http,//linkage, rockefeller.edu/soft/gh/, GENEHUNTER documentation.

http,//www.iop.kcl.ac.uk/IoP/Departments/PsychMed/GEpiBSt/software.stm, King's College, London, Institute of Psychiatry, Division of Psychological Medicine, Section of Epidemiology and Biostatistics.