

Admixture and patterns of linkage disequilibrium in a free-living vertebrate population

J. SLATE* & J. M. PEMBERTON†

*Department of Animal & Plant Sciences, University of Sheffield, Sheffield, UK

†Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

Keywords:

admixture;
association mapping;
 D' ;
gametic disequilibrium;
gene mapping;
natural population.

Abstract

Linkage disequilibrium (LD), a measure of nonrandom association of alleles at different loci, is of great interest to evolutionary geneticists as it can be used to help identify loci that explain phenotypic variation. Surveys of the extent of LD across genomes have been carried out in a number of systems, most notably humans and model organisms. However, studies of natural populations of vertebrates have rarely been performed. Here, we describe an investigation of LD in a free-living island population of red deer *Cervus elaphus*. Relatively high levels of LD extended several tens of centimorgans, and significant LD was frequently detected between unlinked markers. The magnitude of LD varied depending on how the population was sampled. It also varied across different chromosomes, and was shown to be a function of sample size, intermarker distance and marker heterozygosity. A recent admixture event in the population led to an ephemeral increase in LD. Association mapping may be possible in this population, although a high 'baseline' level of LD could lead to false positive associations between marker loci and a trait of interest.

Introduction

When the alleles at one genetic locus are not independent of the alleles at another locus, the two loci are described as being in linkage disequilibrium (LD) (sometimes termed gametic disequilibrium). For a simple two-locus bi-allelic situation, the extent of LD can be estimated as:

$$D = x_{11} - p_1q_1$$

where x_{11} is the frequency of gametes with allele 1 at locus A and allele 1 at locus B, p_1 is the frequency of allele 1 at locus A and q_1 is the frequency of allele 1 at locus B. In other words, LD is the observed minus the expected (if the allelic state at locus A is independent of allelic state at locus B) frequency of gamete A_1B_1 .

Linkage disequilibrium is a central concept in many areas of population genetics because it is sensitive to recombination, genetic drift, selection, migration and inbreeding. Therefore, understanding the extent of LD is relevant to many evolutionary genetic topics including

the evolution and rate of genetic recombination (McVean, 2002; Otto & Lenormand, 2002; Awadalla, 2003) (and therefore the evolution of sex), the mechanistic basis of genetic correlations (Via & Hawthorne, 2005), explanations of why heterozygosity and fitness are often positively correlated (Hansson & Westerberg, 2002), and as a tool to measure effective population size (Hill, 1981) particularly of bottlenecked populations (Wang, 2005).

In recent years, interest in LD has risen further because the increasing availability of genome sequences and high marker density genetic maps has opened up possibilities to detect genes and mutations underlying quantitative genetic variation by association mapping (Terwilliger & Weiss, 1998; Kruglyak, 1999; Jorde, 2000; Pritchard *et al.*, 2000; Cardon & Bell, 2001; Ewens & Spielman, 2001; Weiss & Clark, 2002). Association (or LD) mapping can detect quantitative trait loci (QTL) because a marker genotype will be statistically associated with trait variation when the marker and a QTL are in LD. The power to detect such associations is a function of sample size, the magnitude of the QTL effect, marker density and the extent of LD. Because LD between a functionally important locus and a marker locus is expected to decline as a function of the distance between them, only those markers most closely linked to the QTL should be

Correspondence: Jon Slate, Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.
Tel.: +44 114 2220048; fax: +44 114 2220002;
e-mail: j.slate@sheffield.ac.uk

associated with the trait. The advantages of association mapping over family-based linkage mapping methods are that unrelated individuals can be screened, and that loci can be mapped with greater precision than conventional linkage mapping.

Measuring the extent and magnitude of LD is therefore important, to determine the optimal number of markers required for association mapping. The magnitude and extent of LD has now been described in various human populations (Pritchard & Przeworski, 2001; Reich *et al.*, 2001; Stephens *et al.*, 2001); in domesticated animals including dairy cattle (Farnir *et al.*, 2000; Tenesa *et al.*, 2003), sheep (McRae *et al.*, 2002), chickens (Heifetz *et al.*, 2005), pigs (Nsengimana *et al.*, 2004), thoroughbred horses (Tozaki *et al.*, 2005) and dogs (Lou *et al.*, 2003; Sutter *et al.*, 2004); in important crop plants including maize (Remington *et al.*, 2001; Stich *et al.*, 2005), wheat (Maccaferri *et al.*, 2005) and barley (Morrell *et al.*, 2005; Malysheva-Otto *et al.*, 2006); and in model organisms including *Drosophila melanogaster* (Langley *et al.*, 1974) and *Arabidopsis thaliana* (Nordborg *et al.*, 2002).

Across studies of LD some general patterns are beginning to emerge. Consistent with theory, virtually every empirical study shows that LD declines as a function of physical/recombination distance between markers. Typically the proportion of variance in LD explained by distance varies from a few per cent (Stich *et al.*, 2005) to around 45% (Abecasis *et al.*, 2001). However, the rate of decline with distance is variable across species, across populations and across different parts of the genome. In livestock, where effective population sizes are often small, significant LD can extend for tens of centimorgans (Farnir *et al.*, 2000; McRae *et al.*, 2002). In humans, the extent of LD is generally much lower (Pritchard & Przeworski, 2001), although it varies between populations (Wilson & Goldstein, 2000; Reich *et al.*, 2001; Varilo *et al.*, 2003). In selfing plants, such as *A. thaliana* and barley *Hordeum vulgare*, LD extends over surprisingly short distances (Nordborg *et al.*, 2002; Morrell *et al.*, 2005), at least when a global sampling regime is used.

In humans, the relationship between distance and LD has been examined at a finer scale than elsewhere. It is now clear that over short distances LD declines in a step-like manner rather than as a linear function of distance. The reason for this pattern is that local recombination rates vary as much as 10-fold, such that within a local region there are recombination hot spots and cold spots (McVean *et al.*, 2004). The consequence of this is that LD is arranged into 'blocks' of loci in absolute LD (Daly *et al.*, 2001; Goldstein, 2001) separated by recombination hot spots.

In addition to physical distance between markers, a number of other factors (biological and statistical) are known to shape patterns of LD. Directional and stabilizing selection can generate negative LD between loci, whereas disruptive selection and epistasis can lead to increased positive LD, even when the loci under selection are unlinked (Bulmer, 1971; Walsh & Lynch, 2007). LD is

also sensitive to genetic drift. In populations with low effective population size, high levels of LD between distantly linked or even unlinked loci can arise due to chance sampling effects, usually because rare allele combinations can be lost (Hill & Robertson, 1968). Finally, admixture can result in high levels of LD (Pritchard & Rosenberg, 1999; Pritchard *et al.*, 1999; Wilson & Goldstein, 2000) that subsequently decline due to recombination.

Statistical factors that influence LD include the choice of metric used to measure LD, as well as the sample size and the variability of the genetic markers. There are a large number of LD metrics available (reviewed in Devlin & Risch, 1995; see also Pritchard & Przeworski, 2001; Zhao *et al.*, 2005), and the choice of metric depends on some degree to the question being addressed. Hedrick's (1987) extension of Lewontin's D' is one of the most widely used, because it can be measure LD between multi-allelic markers. The main disadvantage of D' is that it tends to be inflated with small sample sizes and it has also been shown to be sensitive to marker variability (Abecasis *et al.*, 2001; McRae *et al.*, 2002; Zhao *et al.*, 2005). Zhao *et al.* (2005) recently compared LD metrics for measuring LD between a marker and a QTL and concluded that a standardized chi-squared statistic (χ^2) was the preferred measure for quantifying the amount of LD in a population.

Although there have been many studies of LD in recent years, there have been few attempts to quantify the magnitude and extent of LD in a natural vertebrate population. One notable exception is a recent study of collared flycatchers (Backström *et al.*, 2006), but that focused on one chromosome only. The aim of this study is to investigate patterns of LD in a wild population of red deer (*Cervus elaphus*) on the Isle of Rum, UK. This population has been the focus of a long-term individual-based ecological study for over three decades and has been used to study additive genetic correlations between fitness-related traits (Kruuk *et al.*, 2002), the genetic basis of heterozygosity–fitness correlations (Slate & Pemberton, 2002) and to map QTL for fitness related traits (Slate *et al.*, 2002b). An investigation of LD in this population will help to establish whether association mapping is feasible in the population, and enhance understanding of the causes of genetic correlations and heterozygosity–fitness correlations (HFCs). Here, we describe patterns of LD in the red deer genome, with a particular emphasis on the impact of a recent admixture event on genome-wide LD.

Materials and methods

The Rum study population

Red deer (*C. elaphus*) have been studied on the island of Rum (57°0'N, 6°20'W) since the late 1960s. Detailed descriptions of the study population and field methods are available elsewhere (Clutton-Brock *et al.*, 1982;

Kruuk *et al.*, 1999; Slate *et al.*, 2002b; Nussey *et al.*, 2005). The history of the study population is of direct relevance to this paper, however. Red deer were common on Rum until the early 18th century. Thereafter, the growth of the human population, the destruction of forest for sheep farming, and hunting led to a decline in the population, which went extinct around 1787. From the late 1840s onwards, the island was restocked with red deer from a variety of sources including Windsor Great Park, Berkshire in 1845 and 1887, Knowsley Great Park, Lancashire in 1850–1852, Perthshire, Scotland from 1887 onwards, and Warnham Park, Sussex in the late 1920s. The last introduction occurred in 1972 and is pertinent to this paper. In 1970, a hummel (an antlerless stag) from Braemar, Scotland, was crossed to several enclosed Rum hinds to investigate the inheritance of hummelism. All male progeny developed normal antlers, were vasectomized and then released onto the island. One of the offspring, Maxi, achieved considerable reproductive success (the vasectomy operation having been unsuccessful), siring at least 30 offspring and many subsequent descendants. The extended 'Maxi pedigree' was used to map QTL for birth weight in the population (Slate *et al.*, 2002b) because the pedigree contained many related large half-sibships and because it was hypothesized that novel additive genetic variation caused by the introduction would be segregating in Maxi's descendants.

Genetic markers

The Maxi pedigree comprises 364 genotyped individuals, of which 221 are Maxi descendants and the remainder are, for the purposes of this paper, founders. An additional 137 unknown animals were assigned dummy IDs because one or both of their haplotypes could be inferred from the genotypes of their offspring (see Haplotype inference). The animals used in the mapping pedigree were born between 1966 and 1996, and therefore include individuals born before and after the most recent introduction of deer into the study population. The pedigree has been typed for 90 microsatellite loci and three allozyme markers, described in Slate *et al.* (2002b). A genetic map has been constructed by following the co-segregation of markers within the Maxi pedigree and determining marker order and distances with the CriMap software (Green *et al.*, 1990); details in Slate *et al.*, 2002b). Marker order is consistent with a deer linkage map of > 700 markers developed from Père David's deer (*Elaphurus davidianus*)/red deer F1 hybrids backcrossed to red deer (Slate *et al.*, 2002a). Here, we analyse all linkage groups that contain at least three markers, giving a total of 76 markers located on 18 different linkage groups. Investigated linkage groups were 1, 4, 5, 7, 8, 11–15, 17–21, 23, 24 and 26. Markers and map positions are provided in Slate *et al.* (2002b). The number of markers per linkage group ranged from three to six, and linkage group lengths ranged from

18.9 cM (LG 26) to 94.8 cM (LG 19). Previous comparative genomic investigations (Slate *et al.*, 2002a) suggest that each linkage group is an independent chromosome.

Haplotype inference

In diploid organisms, LD is most readily estimated from haplotype rather than genotype marker data. Therefore, an initial step of the analysis was to infer the phase of the two haplotypes inherited by each individual. Haplotype phase was inferred using the Markov Chain Monte Carlo (MCMC) genetic descent algorithm implemented in SimWalk v2.89 (Sobel & Lange, 1996). Before the program could be run, the limits for the maximum number of individuals (MXPEO), the maximum number of founders per pedigree (MXFNDR) and the maximum number of mates per individual (MXMATE) were increased to 1024, 512 and 64, respectively, and the program re-compiled. Simwalk haplotype inference was performed on the University of Sheffield Unix computer cluster, Titania. Unknown parents were assigned dummy IDs and the entire pedigree was analysed simultaneously rather than by splitting it into constituent sibships. Default settings were used throughout, and each chromosome was analysed independently. Haplotype inference appeared to be robust as the number of inferred recombination events did not exceed expectations for any chromosome and the optimal pedigree was obtained before the simulated annealing process had reached the step threshold setting (default 85%) that would have been indicative of unsatisfactory haplotype inference. Furthermore, less than 0.05% of inferred haplotypes were triple recombinants and they only occurred on linkage groups > 60 cM long.

Measurement of LD

We used three different measures of LD. First, we used the Hedrick (1987) multi-allelic extension of the Lewontin (1964) normalized D' . This measure is widely reported in the LD literature, and so can be used to make a comparison with previous studies. It can range from zero to one regardless of allele frequencies, although it is sensitive to both allele frequency and sample size. Rare alleles and small data sets both tend to result in an inflation in D' . We also used the standardized chi-squared statistic χ^2 , that was introduced by Zhao *et al.* (2005). Those authors suggested that marker–marker χ^2 is a good indicator of marker–QTL χ^2 , but marker–marker D' is not necessarily a good indicator of marker–QTL D' .

D' between two multi-allelic markers was calculated as:

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

where u and v are the number of alleles at each marker, p_i is the frequency of allele i at the first marker and q_j is the frequency of allele j at the second marker. $|D'_{ij}|$ is the

absolute value of Lewontin’s normalized LD measure (Lewontin, 1964) calculated as:

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

where

$$D_{ij} = x_{ij} - p_i q_j$$

and

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

where x_{ij} is the frequency of gametes with alleles i at the first marker and j at the second marker, and p_i and q_j are the frequencies of allele i at the first marker and allele j at the second marker.

χ^2 was calculated as:

$$\chi^2 = \frac{\chi^2}{N(w - 1)}$$

where $w = \min(u, v)$, N is the number of haplotypes in the sample and

$$\chi^2 = N \sum_{i=1}^u \sum_{j=1}^v \frac{D_{ij}^2}{p_i q_j}$$

D' and χ^2 were both estimated using the Haploxt program, part of the GOLD package (Abecasis & Cookson, 2000). Note that Haploxt actually calculates the statistic Cramer’s V , which is the square root of χ^2 .

The third measure of LD was the statistical significance (P value) of the observed association between pairs of alleles, tested against a null hypothesis of linkage equilibrium. P values were estimated using an MCMC approximation of Fisher’s exact test, implemented in the population genetics software Arlequin 2.0 (Schneider *et al.*, 2000). A total of 100 000 alternative tables were explored by the Markov chain and estimated P values typically had low standard errors (< 0.001). No corrections for multiple testing were made, because we were primarily interested in the proportion of pairwise comparisons that were significant at $P < 0.05$.

Linkage disequilibrium was estimated between all pairs of markers, regardless of whether they were linked.

Data sets

Linkage disequilibrium was estimated from four different data sets, sample sizes of which are reported in Table 1. Data set 1 (All haplotypes) comprised all haplotypes inferred within the Maxi pedigree. Haplotypes in this data set are nonindependent in the sense that reproductively successful individuals will pass multiple copies of nonrecombined and recombined haplotypes onto their descendants. However, this data set may best reflect the sampling distribution of haplotypes within the overall population, because red deer have a polygynous mating system whereby harem-holding males monopolize mating opportunities and achieve high reproductive

Table 1 LD summary statistics for each data set and LD metric

Data set	N	D'										χ^2									
		0–10 cM	10–20 cM	20–30 cM	30–60 cM	Nonsynthetic	0–10 cM	10–20 cM	20–30 cM	30–60 cM	Nonsynthetic	0–10 cM	10–20 cM	20–30 cM	30–60 cM	Nonsynthetic					
1. All haplotypes	992 (632)	0.447 (0.139)	0.385 (0.118)	0.279 (0.082)	0.226 (0.058)	0.177 (0.060)	0.138 (0.078)	0.118 (0.069)	0.053 (0.019)	0.044 (0.023)	0.029 (0.017)	0.138 (0.078)	0.118 (0.069)	0.053 (0.019)	0.044 (0.023)	0.029 (0.017)					
2. Founders	554 (221)	0.378 (0.103)	0.330 (0.115)	0.250 (0.077)	0.201 (0.064)	0.193 (0.066)	0.108 (0.052)	0.100 (0.061)	0.051 (0.020)	0.040 (0.020)	0.038 (0.021)	0.108 (0.052)	0.100 (0.061)	0.051 (0.020)	0.040 (0.020)	0.038 (0.021)					
3. Maxi descendants	440 (352)	0.502 (0.162)	0.457 (0.154)	0.350 (0.102)	0.319 (0.145)	0.236 (0.112)	0.171 (0.106)	0.159 (0.080)	0.080 (0.052)	0.079 (0.061)	0.057 (0.067)	0.171 (0.106)	0.159 (0.080)	0.080 (0.052)	0.079 (0.061)	0.057 (0.067)					
4. Maxi Nondescendants	554 (281)	0.408 (0.093)	0.378 (0.104)	0.308 (0.094)	0.245 (0.090)	0.224 (0.090)	0.113 (0.049)	0.115 (0.071)	0.077 (0.037)	0.054 (0.027)	0.046 (0.027)	0.113 (0.049)	0.115 (0.071)	0.077 (0.037)	0.054 (0.027)	0.046 (0.027)					

N is the maximum number of haplotypes included in each data set; the number in parentheses indicates the mean number of resolved haplotypes across all marker pairs. Mean (SD) D' and χ^2 are reported for syntenic markers separated by 0–10, 10–20, 20–30 and 30–60 cM, as well as for nonsynthetic marker pairs. LD, linkage disequilibrium.

success within a breeding season (the rut). Data set 2 (Founders) comprised all 'founder' haplotypes. That is, progeny of haplotyped individuals were discarded, such that haplotypes were not (knowingly) double counted. This ensures that particular haplotypes are not over-represented due to high reproductive success of certain individuals within the Maxi pedigree. Data set 3 (Maxi descendants) comprised haplotypes from individuals who were known to be descended from Maxi. Data set 4 (Maxi nondescendants) comprised haplotypes from individuals that were not descendants of Maxi (to the best of our knowledge). A comparison between data sets 1 and 2 was made to evaluate the extent to which LD was inflated due to 'multiple counting' of haplotypes derived from successful individuals. Comparisons between data set 3 and data set 4 enabled us to test whether the most recent admixture event on the island, the introduction of Maxi, had led to an increase in LD across the genome.

Factors affecting LD

Syntenic markers

Previous studies have shown that D' is sensitive to sample size and marker variability, with smaller data sets and more variable markers typically yielding larger values of D' (McRae *et al.*, 2002). For syntenic (linked) marker pairs we fitted a linear model where sample size (\log_{10} transformed), intermarker distance (measured in Kosambi cM, \log_{10} transformed), mean heterozygosity and chromosome were fitted as main effects. Sample size, mean heterozygosity and distance were continuous variables, whereas chromosome was a categorical variable with 17 degrees of freedom. First-order interaction terms were also fitted in the full model. Model terms were tested by sequential removal from the full model, and were retained if the full model gave a significantly better fit (at $P < 0.05$) to the data than the reduced model. For the purposes of these analyses, linked markers were regarded as those that were separated by a distance of less than 60 cM.

Differences in LD between data sets

We examined whether there were differences in LD between data sets 1 and 2 and also between data sets 3 and 4. Models using syntenic marker data were constructed as described above, except data from both data sets (e.g. 1 and 2) were pooled and 'data set' was fitted as a two level categorical term.

Linkage disequilibrium between nonsyntenic markers was also examined to establish whether 'baseline' levels of LD differed between data sets. Models to identify factors that explained variation in LD between nonsyntenic markers were similar to those described above, except distance (which is effectively 50 cM between all unlinked marker pairs) and chromosome were excluded from the model.

Testing for evidence of selection at unlinked loci

It has been suggested that various forms of selection can lead to nonrandom associations between alleles at unlinked loci (Bulmer, 1971). Farnir *et al.* (2000) introduced a test for nonuniform nonsyntenic LD across the genome based on comparison of two linear models. Model 1 fits LD between all nonsyntenic marker pairs and the effects of the two chromosomes on which nonsyntenic loci map are tested. In model 2, an interaction term between the two chromosomes is included; this term is sensitive to unlinked pairs of loci that may be in high LD due to selection. The significance of this interaction term is estimated from:

$$\frac{(SSE_1 - SSE_2)(N - r(X_2))}{SSE_2(r(X_2) - r(X_1))} = F_{(r(X_2) - r(X_1), N - r(X_2))}$$

where SSE_1 and SSE_2 are the residual sum of squares for models 1 and 2, N is the number of nonsyntenic marker pairs (= 2434 here), and $r(X_1)$ and $r(X_2)$ are the rank incidence matrices for models 1 and 2 (in this case 18 and 153 respectively).

Temporal trends

We also examined temporal trends in LD by estimating LD from individuals born within each year of the long-term study. Because the number of individuals born in any given year was low (min = 1; max = 32), we pooled individuals born over three consecutive years, treating the middle year as the 'focal year'. This analysis was performed as a 'sliding window', such that consecutive data points were constructed from data sets that overlapped by 2 years. D' from nonsyntenic markers was measured, and corrected for sample size (each data point was taken as the residual from a regression of D' on \log_{10} sample size). Sample-size corrected mean D' was then estimated for each year to examine temporal patterns in LD.

Comparison with other studies

We were unaware of any formal comparison of levels of LD between different study populations or species, and in particular whether meaningful comparisons can be made without accounting for sample size or the distance between linked markers. Therefore, we compiled data where D' had been estimated between nonsyntenic markers from the literature, and examined the effect of sample size (number of haplotypes screened) on LD magnitude.

Results

General patterns

Linkage disequilibrium between syntenic markers declined as a function of distance regardless of data set or LD metric (Figs 1 and 2, Tables 1–2). Intermarker

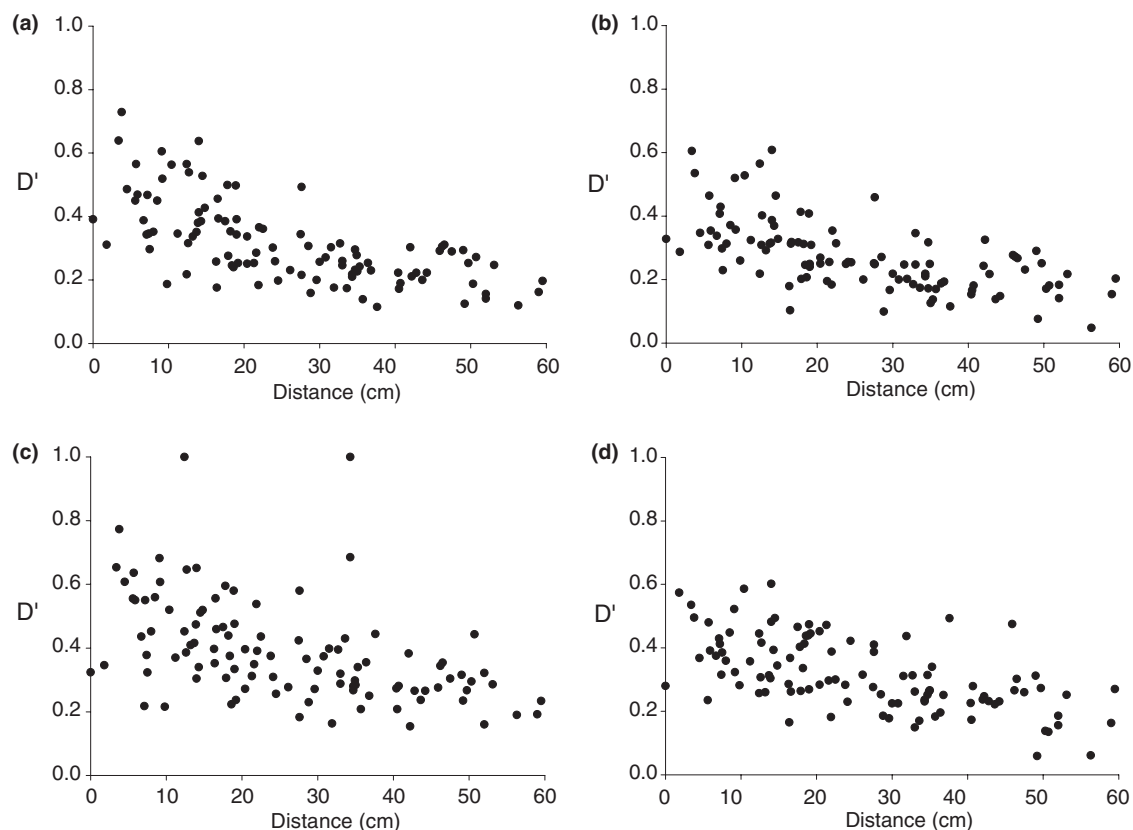


Fig. 1 D' as a function of distance between syntenic markers. Measured in (a) data set 1: All haplotypes, (b) data set 2: Founders only, (c) data set 3: Maxi descendants and (d) data set 4: Maxi nondescendants.

distance explained between 19% (data set 3, LD measured by D') and 32% (data set 1, LD measured by χ^2) of the variation in LD between syntenic markers (Table 3). Within data set 3 (Maxi descendants), LD declined as a function of increasing sample size (Table 3), i.e. marker pairs with relatively few haplotypes resulted in greater levels of LD. There was also evidence of interchromosomal variation in the extent of LD, even after correcting for the effects of distance and sample size (Table 3). Mean heterozygosity explained significant variation in χ^2 in one data set, but not in D' for any data set.

We examined the proportion of syntenic and nonsyntenic marker pairs that were in significant (as inferred by Fisher's exact test) LD (Figs 3 and 4). At least 86% of marker pairs separated by 20 cm or less were in significant (at $P < 0.05$) LD, regardless of data set (Table 2). Strikingly, for data sets 1 and 3 marker pairs separated by > 30 cm and even nonsyntenic marker pairs were often in significant LD (Table 2).

Differences between data sets

Linkage disequilibrium between syntenic markers was greater in data set 1 (All haplotypes) than in data set 2 (Founders only) whether measured by D' ($F_{1,189} = 12.84$,

$P < 0.001$) or by χ^2 ($F_{1,188} = 4.86$, $P = 0.029$). These differences are after accounting for the independent effects of intermarker distance, interchromosomal variation, variation in mean heterozygosity and sample size (number of haplotypes screened). Data set 1 also exhibited higher levels of LD than data set 2 among nonsyntenic marker pairs, regardless of LD metric (Fig. 3). In a model of D' between nonsyntenic marker pairs, 'data set' was significant as a first-order term ($F_{1,4863} = 4.22$, $P = 0.04$), and as an interaction with sample size ($F_{1,4861} = 7.44$, $P < 0.01$) and mean heterozygosity ($F_{1,4861} = 12.60$, $P < 0.001$). For χ^2 , data set was not significant as a main effect ($F_{1,4863} = 2.81$, $P = 0.094$), but was significant as an interaction term with sample size ($F_{1,4863} = 20.89$, $P < 0.001$).

Linkage disequilibrium was greater in data set 3 (Maxi descendants) than in data set 4 (Maxi nondescendants; Fig. 4, Tables 1–2). When LD was measured by D' between syntenic marker pairs, 'data set' was a significant explanatory term, both as a main effect ($F_{1,186} = 23.42$, $P < 0.001$) and as interaction terms with 'chromosome' ($F_{17,169} = 2.36$, $P < 0.01$) and 'sample size' ($F_{1,169} = 5.85$, $P < 0.05$). When LD was measured with χ^2 , 'data set' was significant as a first-order term ($F_{1,189} = 17.03$, $P < 0.001$) and as an interaction with 'distance' ($F_{1,188} = 4.58$, $P = 0.03$).

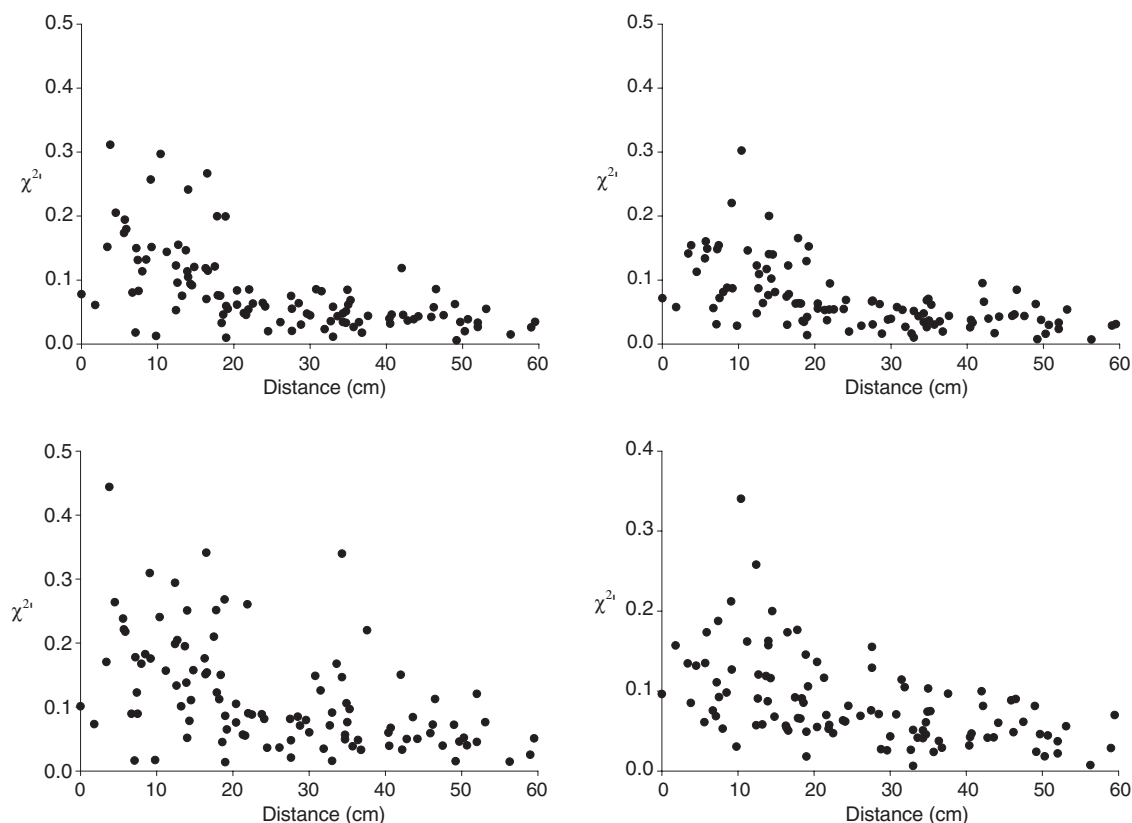


Fig. 2 χ^2 as a function of distance between syntenic markers. Measured in (a) data set 1: All haplotypes and (b) data set 2: Founders only. (c) Data set 3: Maxi descendants and (d) data set 4: Maxi nondescendants.

When LD was measured between nonsyntenic marker pairs, it was greater in data set 3 than in data set 4 regardless of LD metric (D' : $F_{1,4864} = 49.93$, $P < 0.001$; χ^2 : $F_{1,4862} = 5532.8$, $P \ll 0.001$). The interaction term between 'data set' and 'sample size' was also significant whether LD was measured by D' ($F_{1,4864} = 137.16$, $P < 0.001$) or χ^2 ($F_{1,4862} = 183.82$, $P < 0.001$), as was the interaction between 'data set' and mean heterozygosity when LD was measured by χ^2 ($F_{1,4862} = 80.87$, $P < 0.001$).

Testing for evidence of selection affecting nonsyntenic LD

Using estimates of D' between nonsyntenic markers in data set 1, we tested whether there was evidence of selection causing heterogeneity in levels on LD across the genome. The interaction term between the two chromosomes was not significant ($F_{135,2281} = 0.64$, $P \sim 0.99$), indicating that selection was not a major contributor to patterns of nonsyntenic LD.

Temporal analysis

D' between nonsyntenic markers was greater when estimated from animals born during the period that Maxi

was breeding (mean 0.37, SD 0.02) than when estimated from animals born outside of this period (mean 0.34, SD 0.01). Nonindependence between sampling years, make formal statistical comparisons between 'Maxi years' and 'Non-Maxi years' problematic. However, any possible effect of Maxi on LD magnitude is unlikely to be confounded with population density, because mean D' and population density are not associated ($r = 0.024$, $P = 0.91$).

Discussion

General patterns

Here we have estimated LD across the genome of a wild red deer population. Unsurprisingly, for a population of a species with a polygynous mating system and a history of recent admixture, significant levels of LD extended for tens of centimorgans along chromosomes. However, LD did decline with distance, even within a sample derived entirely from descendants of the most recently introduced immigrant to the study population. Depending on the data set and the LD metric, intermarker distance explained between 3% and 32% of the variance in LD among syntenic markers. Thus, the effects of recombination on

Table 2 The proportion of markers in significant ($P < 0.05$) LD in all data sets

Data set	Proportion				
	0–10 cM	10–20 cM	20–30 cM	30–60 cM	Nonsyntenic
1. All haplotypes	16/17 (0.94)	27/28 (0.96)	17/18 (0.94)	34/41 (0.83)	1867/2434 (0.77)
2. Founders	16/17 (0.94)	25/28 (0.89)	11/18 (0.61)	15/41 (0.37)	365/2434 (0.15)
3. Maxi descendants	16/17 (0.94)	24/28 (0.86)	17/18 (0.94)	35/41 (0.85)	1668/2434 (0.69)
4. Maxi nondescendants	16/17 (0.94)	27/28 (0.96)	16/18 (0.89)	26/41 (0.63)	1078/2434 (0.44)

The number and proportion of marker pairs in significant LD is reported for markers separated by 0–10, 10–20, 20–30 and 30–60 cM, as well as for nonsyntenic marker pairs. Statistical significance was determined by Fisher's exact test. Note the high frequency of nonsyntenic marker pairs in significant LD in data sets 1 and 3. LD, linkage disequilibrium.

Table 3 Factors that explain variance in LD between syntenic markers

Data set	d.f.	1. All individuals		2. Founders		3. Maxi descendants		4. Non-Maxi descendants	
		F	%	F	%	F	%	F	%
<i>D'</i>									
Distance	1	67.78***	31.8	43.01***	24.2	38.27***	19.2	43.03***	21.8
N haplotypes	1	1.24	0.6	2.94	1.7	31.10***	15.6	0.99	0.5
Chromosome	17	1.99*	15.9	1.49	14.3	2.17*	18.6	3.13***	26.9
χ^2									
Distance	1	54.98***	32.0	32.28***	29.1	45.53***	20.1	17.53***	22.6
N haplotypes	1	2.87	0.9	1.65	2.4	20.52***	15.9	1.26	0.1
Chromosome	17	3.16***	29.2	2.31**	23.4	2.41**	23.5	3.11***	28.6
Heterozygosity	1	3.29	1.4	0.52	0.3	8.35**	3.6	2.19	1.2

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Linear models were constructed that examined the proportion of variance in LD explained by intermarker distance (measured in centimorgans and \log_{10} transformed), the sample size (\log_{10} transformed number of haplotypes contributing to each estimate of marker–marker LD), interchromosomal variation in LD and marker variability (mean heterozygosity of the two markers contributing to each estimate of marker–marker LD). For each term the numerator degrees of freedom, *F*-ratio statistic and percentage of overall variance explained by that term are provided. Note that to compare the influence of different factors between data sets the minimal model is not always shown. Mean heterozygosity did not explain significant variation in *D'* in any data set and is not shown. Residual degrees of freedom are 104. LD, linkage disequilibrium.

patterns of LD are detectable, even within one to two generations of the most admixture event.

One striking observation was the high proportion of nonsyntenic marker pairs in significant LD in some data sets. Among data set 1 (All haplotypes) and data set 3 (Maxi descendants) 77% and 69% of nonsyntenic marker pairs were in significant LD (Figs 3 and 4). Among data sets 2 (Founders) and 4 (Maxi nondescendants) equivalent figures were 15% and 40% respectively. It appears likely that the introduction of Maxi caused the majority of nonsyntenic marker pairs to be in significant LD, although these associations should rapidly be broken down by independent assortment. Similar studies in dairy cattle (Farnir *et al.*, 2000) and domestic sheep (McRae *et al.*, 2002) have found that 10–30% of nonsyntenic markers were in LD, similar to the values we observe in our data sets that are least affected by recent admixture, although significantly more than expected in an equilibrium population. The best explanation for the patterns we observe are that population structure and

mating system caused some departures from equilibrium among nonsyntenic loci, but LD has been increased by the most recent admixture event.

We detected interchromosomal heterogeneity in LD among syntenic marker pairs, regardless of data set or LD metric employed. Previous studies have failed to detect interchromosomal variation in LD (McRae *et al.*, 2002; Lou *et al.*, 2003) with the exception of one study of domestic pigs that only analysed two chromosomes (Nsengimana *et al.*, 2004). In many of our models, the term 'chromosome' explained as much of the variation in LD as did intermarker distance (see Table 3). It is known that local recombination rates can vary over fine scales in many genomes (McVean *et al.*, 2004), causing heterogeneity in the magnitude of haplotype blocks (Stumpf & Goldstein, 2003), but that explanation is unlikely here as most of our markers are separated by tens of centimorgans, and recombination fraction was already fitted in the model. A more likely explanation is that the interchromosomal variation we detect is an artefact

Fig. 3 Proportion of syntenic and nonsyntenic marker pairs in significant ($P < 0.05$) linkage disequilibrium (LD). Data sets 1 and 2 are shown. Statistical significance was determined by Fisher's exact test. The vertical line is drawn at $P < 0.05$. A greater proportion of syntenic marker pairs are in significant LD than nonsyntenic marker pairs. More marker pairs were in significant LD in data set 1 than in data set 2.

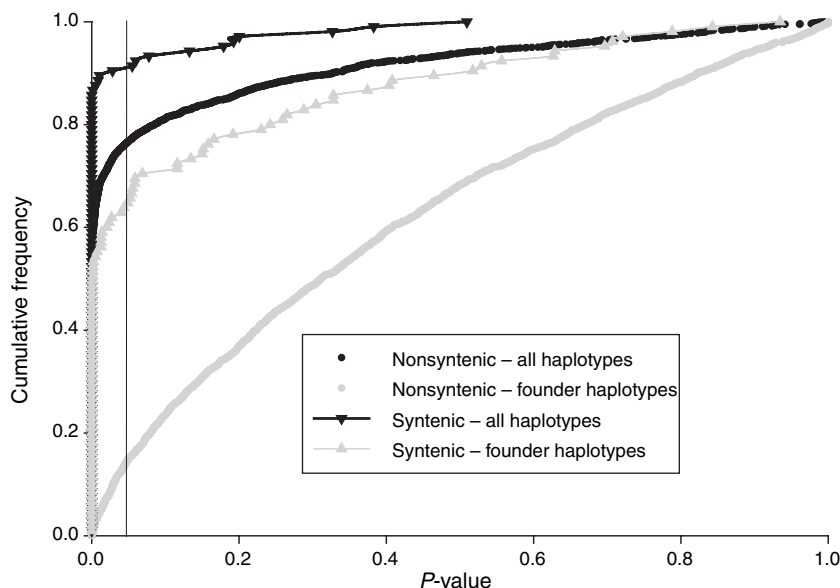
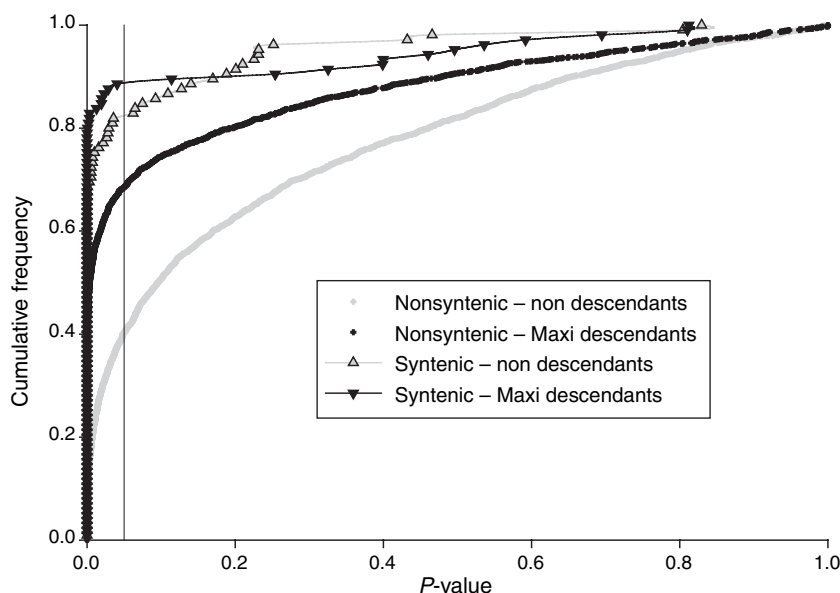


Fig. 4 Significant linkage disequilibrium (LD) is more often observed when calculated from Maxi descendants (data set 3) than nondescendants (data set 4). In data set 3 around 70% of nonsyntenic marker pairs are in significant LD.



caused by the large variance in LD metrics, combined with the relatively small number of marker pairs per chromosome (mean 5.8).

We also examined whether chromosome-by-chromosome interaction was a significant term in a general linear model that examined LD between nonsyntenic marker pairs. When fitted as a main effect chromosome was significant ($F_{17,2416} = 14.00$, $P < 0.001$), but the interaction term was not significant ($F_{135,2281} = 0.635$, $P = 0.99$). A significant interaction term would have been consistent with the 'Bulmer effect'; the idea that LD is the greatest between regions of the genome that

contain QTL for characters under selection (Bulmer, 1971; Walsh & Lynch, 2007). These data indicate that selection has not contributed to patterns of nonsyntenic LD in this population. A similar observation was made in an analysis of Dutch dairy cattle (Farnir *et al.*, 2000).

Comparisons between data sets

The general pattern of long-range LD and significant LD between nonsyntenic marker pairs was observed in all four data sets. However, there were some distinctions

between the data sets. After correcting for the effect of sample size, LD was significantly greater in data set 1 (All haplotypes) than data set 2 (Founders only). This observation is unsurprising, as some haplotypes will be represented many times in data set 1 among the descendants of reproductively successful individuals. This is likely to be particularly true for nonrecombinant haplotypes, although recombinant haplotypes may also be represented multiple times. It could be argued that data set 2 is unbiased due to the independence of sampled haplotypes, although data set 1 perhaps better represents the effect of a polygynous mating system on patterns of LD, because most of the nonindependent haplotypes are descended from a small number of highly successful males, who themselves may be related.

Linkage disequilibrium in data set 3 (Maxi descendants) was significantly greater than in data set 4. This pattern was apparent both for syntenic and nonsyntenic marker pairs. There are two likely causes of this observation. First, because Maxi and his son RED77 were the most successful individuals within the entire pedigree, nonindependence of haplotypes is likely to be greater among data set 3 than data set 4. However, an additional factor, recent admixture, is also likely to be important. Maxi was a cross between a hind from Rum and a stag from Braemar, north-east Scotland, born in 1970; he first bred in 1975. The last admixture event occurred ~50 years earlier, and did not involve deer from NE Scotland. Therefore, it is likely that novel haplotypes were present in Maxi and his descendants, which would have caused an increase in LD. A temporal analysis of patterns in LD among nonsyntenic markers provides firm support for an increase in LD directly attributable to Maxi mating in the population (Fig. 5). LD remained at a fairly constant level before Maxi bred, and then showed a sharp increase the year he first bred, which was main-

tained until he ceased breeding. Shortly after Maxi died, levels of LD returned to pre-1975 levels. The population size of the study area also changed (increased) during this period, which in theory could have caused LD to decline. However, temporal patterns of LD during this period do not appear to show any obvious relationship with changes in population density. In contrast, the years in which LD was the greatest coincide almost perfectly with the reproductive lifetime of Maxi. Certainly, these data are consistent with an ephemeral increase in LD caused by an admixture event.

Comparison with other studies

Most of the previously collected data on LD were obtained from humans or model organisms, with relatively little data collected in the wild. Direct comparisons between different studies are complicated because: (i) LD is often measured using different metrics; (ii) maps are often measured in different units, e.g. physical maps are typically measured in nucleotides and linkage maps in centimorgans; and (iii) most estimates of LD are sensitive to the number of individuals typed. However, it is possible to make comparisons between studies that measure D' between nonsyntenic markers. We compiled data from studies where this information was available (see Fig. 6). The effect of sample size on D' magnitude is particularly pronounced; it actually accounts for 58% of the variation between published data sets. After accounting for sample size it is apparent that the magnitude of background LD is reasonably consistent across studies, regardless of whether they were carried out on vertebrates or plants, or come from artificial or natural populations. The magnitude of LD observed here is not atypical relative to other populations.

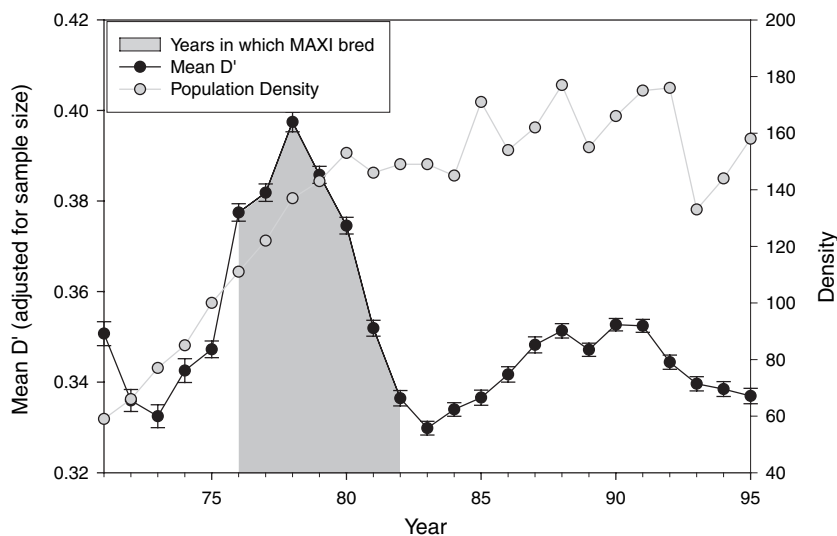


Fig. 5 Temporal peaks in LD coincide with Maxi's reproductive lifetime. To ensure adequate sample size, LD was measured as mean D' between nonsyntenic marker pairs calculated from animals born in a sliding 3-year window (e.g. 1969–1971, 1970–1972, 1971–1973, etc.) and corrected for sample size. The mean number of haplotypes per year was 72.6. Population density was estimated as the number of adult hinds resident in the study area. Note that LD was greatest in years where Maxi's offspring were born, and then declined to pre-Maxi levels. There is no apparent relationship between LD and population density.

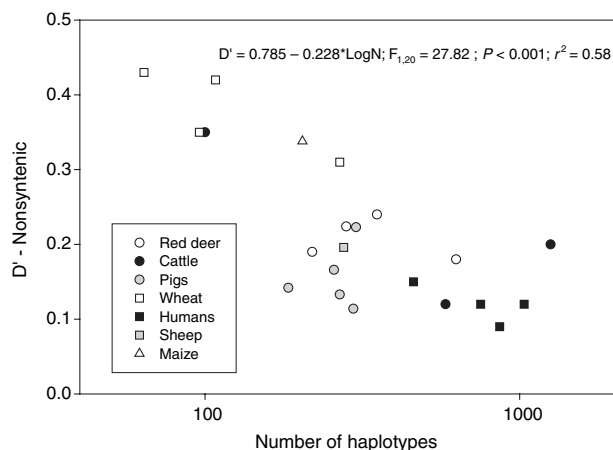


Fig. 6 A comparison of linkage disequilibrium (LD) in different studies. D' between nonsynthetic markers has been measured in a variety of systems and declines as a function of sample size. The magnitude of LD observed in red deer is not atypical compared with other systems. Data taken from red deer (this study), domestic cattle (Farnir *et al.*, 2000; Tenesa *et al.*, 2003), domestic pigs (Nsengimana *et al.*, 2004), domestic sheep (McRae *et al.*, 2002), cultivated maize (Remington *et al.*, 2001), cultivated wheat (Maccaferri *et al.*, 2005) and humans (Dunning *et al.*, 2000).

Different LD metrics

It has recently been suggested (Zhao *et al.*, 2005) that χ^2 is a useful LD metric because marker–marker χ^2 better reflects marker–QTL LD than is the case for other metrics. Although we do not specifically evaluate the ability of D' and χ^2 to estimate marker–QTL LD, a number of interesting observations are apparent. First, intra-chromosomal LD does appear to decline relatively more rapidly with distance when measured with χ^2 compared with D' (Figs 1–2). For example, χ^2 between markers spaced 0–10 cM apart was typically two times greater than for marker pairs separated by 20–30 cM, whereas D' between markers 0–10 cM apart was always less than twice as great as for markers 20–30 cM apart. In all data sets, intermarker distance explained marginally more of the variance in LD between syntenic markers when LD was measured with χ^2 than when it was measured with D' . χ^2 and D' appear to be equally sensitive to the number of sampled haplotypes. In all four data sets interchromosomal variation was greater when LD was measured with χ^2 , than when it was measured with D' . Given that it is useful to compare LD statistics across studies we recommend that future studies should report both LD metrics, regardless of which gives the best indication of marker–QTL LD.

Prospects for LD mapping

A decline in LD with distance would imply that LD mapping might be possible in this population. However,

high levels of LD were observed between distantly linked markers, and a high proportion of nonsynthetic markers were in significant LD (at the nominal significance $P < 0.05$ level). Therefore, association mapping studies may be possible with relatively low marker density genome scans, but the risk of type 1 error is high, and the prospects for fine-scale mapping once regions harbouring QTL are identified might be limited. The marker density of the red deer linkage map is currently too low to evaluate patterns of LD between more closely linked (<1 cM) loci.

The temporal patterns in LD we observed may also aid fine-scale mapping. It is possible that low marker density association mapping scans will be successful if applied to animals that are recent descendants (one to three generations) of Maxi, whereas finer scale mapping of QTL will be more successful in data sets where baseline LD is lower, e.g. a sample of later generation descendants.

Summary

In conclusion, we have described the most comprehensive survey of LD in a nonhuman wild vertebrate population to date. The general pattern is one of high levels of LD that decline as a function of distance, although significant LD is detected between nonsynthetic markers. A recent admixture event resulted in a pronounced ephemeral rise in LD. Association mapping studies may be possible in this population, although care will have to be taken to minimize the risk of type 1 error.

Acknowledgements

We thank Scottish National Heritage for permission to work on Rum; Tim Clutton-Brock, Fiona Guinness and Steve Albon for their long-term contributions to the project; and Angela Alexander, Ailsa Curnow, Sean Morris, Ali Donald and numerous volunteers for field data collection. Kati Csillery, Allan McRae and Dan Nussey provided valuable comments on a previous draft of the manuscript. The genotype data used in this work were generated while JS was a Biotechnology and Biological Sciences Research Council PhD student in JP's laboratory. The red deer research project on Rum is supported by the Natural Environment Research Council.

References

- Abecasis, G.R. & Cookson, W.O.C. 2000. GOLD – graphical overview of linkage disequilibrium. *Bioinformatics* **16**: 182–183.
- Abecasis, G.R., Noguchi, E., Heinzmann, A., Traherne, J.A., Bhattaharyya, S., Leaves, N.I., Anderson, G.G., Zhang, Y., Lench, N.J., Carey, A., Cardon, L.R., Moffatt, M.F. & Cookson, W.O.C. 2001. Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68**: 191–197.

- Awadalla, P. 2003. The evolutionary genomics of pathogen recombination. *Nat. Rev. Genet.* **4**: 50–60.
- Backström, N., Qvarnström, A., Gustafsson, L. & Ellegren, H. 2006. Levels of linkage disequilibrium in a wild bird population. *Biol. Lett.*
- Bulmer, M.G. 1971. Effect of selection on genetic variability. *Am. Nat.* **105**: 201–6.
- Cardon, L.R. & Bell, J.I. 2001. Association study designs for complex diseases. *Nat. Rev. Genet.* **2**: 91–99.
- Clutton-Brock, T.H., Guinness, F.E. & Albon, S.D. 1982. *Red Deer – Behavior and Ecology of Two Sexes*. Edinburgh University Press, Edinburgh.
- Daly, M.J., Rioux, J.D., Schaffner, S.E., Hudson, T.J. & Lander, E.S. 2001. High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**: 229–232.
- Devlin, B. & Risch, N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**: 311–322.
- Dunning, A.M., Durocher, F., Healey, C.S., Teare, M.D., McBride, S.E., Carlomagno, F., Xu, C.-F., Dawson, E., Rhodes, S., Ueda, S., Lai, E., Luben, R.N., Van Rensburg, E.J., Mannermaa, A., Kataja, V., Rennart, G., Dunham, I., Purvis, I., Easton, D. & Ponder, B.A.J. 2000. The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am. J. Hum. Genet.* **67**: 1544–1554.
- Ewens, W.J. & Spielman, R.S. 2001. Locating genes by linkage and association. *Theor. Popul. Biol.* **60**: 135–139.
- Farnir, F.W.C., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. & Georges, M. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* **10**: 220–227.
- Goldstein, D.B. 2001. Islands of linkage disequilibrium. *Nat. Genet.* **29**: 109–111.
- Green, P., Falls, K. & Crooks, S. 1990. *Documentation for CRI-MAP*. Washington University, St Louis, MO.
- Hansson, B. & Westerberg, L. 2002. On the correlation between heterozygosity and fitness in natural populations. *Mol. Ecol.* **11**: 2467–2474.
- Hedrick, P. 1987. Gametic disequilibrium measures: proceed with caution. *Genetics* **117**: 331–341.
- Heifetz, E.M., Fulton, J.E., O'Sullivan, N., Zhao, H., Dekkers, J.C.M. & Soller, M. 2005. Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* **171**: 1173–1181.
- Hill, W.G. 1981. Estimation of effective population-size from data on linkage disequilibrium. *Genet. Res.* **38**: 209–216.
- Hill, W.G. & Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**: 226–231.
- Jorde, L.B. 2000. Linkage disequilibrium and the search for complex disease genes. *Genome Res.* **10**: 1435–1444.
- Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22**: 139–144.
- Kruuk, L.E.B., Clutton-Brock, T.H., Rose, K.E. & Guinness, F.E. 1999. Early determinants of lifetime reproductive success differ between the sexes in red deer. *Proc. R. Soc. Lond. Ser. B* **266**: 1655–1661.
- Kruuk, L.E.B., Slate, J., Pemberton, J.M., Brotherstone, S., Guinness, F. & Clutton-Brock, T. 2002. Antler size in red deer: heritability and selection but no evolution. *Evolution* **56**: 1683–1695.
- Langley, C.H., Tobar, Y.N. & Kojima, K.-I. 1974. Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**: 921–936.
- Lewontin, R.C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49–67.
- Lou, X.Y., Todhunter, R.J., Lin, M., Lu, Q., Liu, T., Wang, Z.H., Bliss, S.P., Casella, G., Acland, G.M., Lust, G. & Wu, R.L. 2003. The extent and distribution of linkage disequilibrium in a multi-hierarchical outbred canine pedigree. *Mamm. Genome* **14**: 555–564.
- Maccaferri, M., Sanguineti, M.C., Noli, E. & Tuberosa, R. 2005. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Mol. Breed.* **15**: 271–289.
- Malysheva-Otto, L.V., Ganal, M.W. & Roder, M.S. 2006. Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genet.* **7**: 6.
- McRae, A.F., McEwan, J.C., Dodds, K.G., Wilson, T., Crawford, A.M. & Slate, J. 2002. Linkage disequilibrium in domestic sheep. *Genetics* **160**: 1113–1122.
- McVean, G.A.T. 2002. A genealogical interpretation of linkage disequilibrium. *Genetics* **162**: 987–991.
- McVean, G.A.T., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R. & Donnelly, P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* **304**: 581–584.
- Morrell, P.L., Toleno, D.M., Lundy, K.E. & Clegg, M.T. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc. Natl Acad. Sci. U.S.A.* **102**: 2442–2447.
- Nordborg, M., Borevitz, J.O., Bergelson, J., Berry, C.C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J.N., Noyes, T., Oefner, P.J., Stahl, E.A. & Weigel, D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* **30**: 190–193.
- Nsengimana, J., Baret, P., Haley, C.S. & Visscher, P.M. 2004. Linkage disequilibrium in the domesticated pig. *Genetics* **166**: 1395–1404.
- Nussey, D.H., Coltman, D.W., Coulson, T., Kruuk, L.E.B., Donald, A., Morris, S.J., Clutton-Brocks, T.H. & Pemberton, J. 2005. Rapidly declining fine-scale spatial genetic structure in female red deer. *Mol. Ecol.* **14**: 3395–3405.
- Otto, S.P. & Lenormand, T. 2002. Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**: 252–261.
- Pritchard, J.K. & Przeworski, M. 2001. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**: 1–14.
- Pritchard, J.K. & Rosenberg, N.A. 1999. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* **65**: 220–228.
- Pritchard, J.K., Stephens, M. & Donnelly, P.J. 1999. Correcting for population stratification in linkage disequilibrium mapping studies. *Am. J. Hum. Genet.* **65**: 528.
- Pritchard, J.K., Stephens, M., Rosenberg, N.A. & Donnelly, P. 2000. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**: 170–181.
- Reich, D.E., Cargill, M., Bolik, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R. & Lander, E.S. 2001. Linkage disequilibrium in the human genome. *Nature* **411**: 199–204.
- Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doeblay, J., Kresovich, S., Goodman, M.M. & Buckler, E.S. 2001. Structure of linkage disequilibrium and

- phenotypic associations in the maize genome. *Proc. Natl Acad. Sci. U.S.A.* **98**: 11479–11484.
- Schneider, S., Roessli, D. & Excoffier, L. 2000. *Arlequin Ver. 2.000: A Software for Genetic Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva.
- Slate, J. & Pemberton, J.M. 2002. Comparing molecular measures for detecting inbreeding depression. *J. Evol. Biol.* **15**: 20–31.
- Slate, J., van Stijn, T.C., Anderson, R.M., McEwan, K.M., Maqbool, N.J., Mathias, H.C., Bixley, M.J., Stevens, D.R., Molenaar, A.J., Beever, J.E., Galloway, S.M. & Tate, M.L. 2002a. A deer (subfamily *Cervinae*) genetic linkage map and the evolution of ruminant genomes. *Genetics* **160**: 1587–1597.
- Slate, J., Visscher, P.M., MacGregor, S., Stevens, D., Tate, M.L. & Pemberton, J.M. 2002b. A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). *Genetics* **162**: 1863–1873.
- 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.
- Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharya, T., Stanley, S.E., Jiang, R.H., Messer, C.J., Chew, A., Han, J.H., Duan, J.C., Carr, J.L., Lee, M.S., Koshy, B., Kumar, A.M., Zhang, G., Newell, W.R., Windemuth, A., Xu, C.B., Kalbfleisch, T.S., Shaner, S.L., Arnold, K., Schulz, V., Drysdale, C.M., Nandabalan, K., Judson, R.S., Ruano, G. & Vovis, G.F. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**: 489–493.
- Stich, B., Melchinger, A.E., Frisch, M., Maurer, H.P., Heckenberger, M. & Reif, J.C. 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor. Appl. Genet.* **111**: 723–730.
- Stumpf, M.P.H. & Goldstein, D.B. 2003. Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Curr. Biol.* **13**: 1–8.
- Sutter, N.B., Eberle, M.A., Parker, H.G., Pullar, B.J., Kirkness, E.F., Kruglyak, L. & Ostrander, E.A. 2004. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* **14**: 2388–2396.
- Tenesa, A., Knott, S.A., Ward, D., Smith, D., Williams, J.L. & Visscher, P.M. 2003. Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *J. Anim. Sci.* **81**: 617–623.
- Terwilliger, J.D. & Weiss, K.M. 1998. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr. Opin. Biotechnol.* **9**: 578–594.
- Tozaki, T., Hirota, K., Hasegawa, T., Tomita, M. & Kurosawa, M. 2005. Prospects for whole genome linkage disequilibrium mapping in thoroughbreds. *Gene* **346**: 127–132.
- Varilo, T., Paunio, T., Parker, A., Perola, M., Meyer, J., Terwilliger, J.D. & Peltonen, L. 2003. The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum. Mol. Genet.* **12**: 51–59.
- Via, S. & Hawthorne, D.J. 2005. Back to the future: genetic correlations, adaptation and speciation. *Genetica* **123**: 147–156.
- Walsh, J.B. & Lynch, M. 2007. *Evolution and Selection of Quantitative Traits*. Volume 2. http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html
- Wang, J. 2005. Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. Lond. Ser. B* **360**: 1395–1409.
- Weiss, K.M. & Clark, A.G. 2002. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet.* **18**: 19–24.
- Wilson, J.F. & Goldstein, D.B. 2000. Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am. J. Hum. Genet.* **67**: 926–935.
- Zhao, H., Nettleton, D., Soller, M. & Dekkers, J.C.M. 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet. Res.* **86**: 77–87.

Received 17 October 2006; revised 12 February 2007; accepted 15 February 2007