doi:10.1111/j.1365-2052.2005.01400.x

Genome-wide linkage disequilibrium in two Japanese beef cattle breeds

M. Odani*, A. Narita*, T. Watanabe†, K. Yokouchi†, Y. Sugimoto†, T. Fujita‡, T. Oguni§, M. Matsumoto§ and Y. Sasaki*

*Laboratory of Animal Breeding and Genetics, Graduate School of Agriculture, Kyoto University, Sakyo, Kyoto, Japan. †Shirakawa Institute of Animal Genetics, Nishigo, Fukushima, Japan. ‡Oita Prefectural Institute of Animal Industry, Kuju, Oita, Japan. §Kumamoto Agricultural Research Center, Goushi, Kumamoto, Japan

Summary

There is little knowledge about the degree of linkage disequilibrium (LD) in beef cattle. This study aims to perform a genome-wide search for LD in Japanese Black and Japanese Brown beef cattle and to compare the level of LD between these two breeds. Parameter D' (the LD coefficient) was used as a measure of LD, and LD was tested for significance of allelic associations between syntenic and between non-syntenic marker pairs. Effects of breed, chromosome, genetic map distance and their interactions with D' were tested based on least squares analyses. Both breeds showed high levels of LD, which ranged over several tens of cM and declined as the marker distance increased for syntenic marker pairs. A rapid decline of the D' value was observed between markers that were spaced 5 and 20 cM apart. LD was significant in most cases for marker pairs <40 cM apart but was not significant between non-syntenic loci. The pattern of LD found in these two breeds was similar to that previously published for dairy cattle. The D' value between breeds was not significantly different ($P > 0.05$), but the interaction between breed and chromosome was highly significant ($P < 0.001$). Genetic selection seems to have caused the heterogeneity of the D' values among chromosomes within breed. These results indicate that LD mapping is a useful tool for fine-mapping quantitative trait loci of economically important traits in Japanese beef cattle.

Keywords beef cattle, fine-mapping, linkage disequilibrium, selection.

Introduction

In recent years, the number of genetic markers linked to quantitative trait loci (QTL) of economically important traits has dramatically increased in livestock (e.g. Andersson 2001). After detection of the QTL-linked markers, two important issues need to be resolved: (i) the identification of causal genes controlling the traits; and (ii) the identification of QTL-linked markers for marker-assisted selection (MAS) to improve the accuracy of genetic evaluation (Fernando & Grossman 1989). To date, at least two excellent experiments have produced convincing evidence of causal mutations for QTL or quantitative trait nucleotides in livestock (Van Laere *et al.* 2003; Grisart *et al.* 2004).

Address for correspondence

Yoshiyuki Sasaki, Laboratory of Animal Breeding and Genetics, Graduate School of Agriculture, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan.
E-mail: sasaki@kais.kyoto-u.ac.jp

Accepted for publication 24 October 2005

Linkage disequilibrium (LD) is a non-random association of alleles at different genetic loci in a population. LD has received much attention as it provides a potential way of fine-mapping a QTL region underlying a trait (Terwilliger & Weiss 1998). By making full use of historical recombination events, LD mapping has the potential to position QTL to a small chromosomal segment, perhaps on the order of <1 cM (Ardlie *et al.* 2002), and it has been used to attain mapping resolution down to the sub-cM level (Van Laere *et al.* 2003; Grisart *et al.* 2004). The usefulness of LD in fine-mapping and MAS depends on the degree of LD, the distribution and heterogeneity of LD across the genome and its relationship with genetic map or physical distances in the population.

Previous studies on genome-wide LD in livestock species have shown that considerable LD spans large genetic distances (>20 cM) in Dutch dairy cattle (Farnir *et al.* 2000), New Zealand sheep (McRae *et al.* 2002) and canine (Lou *et al.* 2003) populations. These studies have shown significant associations not only between syntenic loci but also between pairs of unlinked markers on different chromo-

somes. Thus, LD analysis in livestock may require a simultaneous test for linkage and association to avoid false-positive results because of the associations between non-syntenic loci (Farnir *et al.* 2000). Long-range LD was also observed for two bovine genomic regions (chromosomes 4 and 6) in a sample of the dairy cattle population in the UK (Tenesa *et al.* 2003) and for two porcine genomic regions (chromosomes 4 and 7) in five populations of domesticated pigs (Nsengimana *et al.* 2004). Hayes *et al.* (2003) detected LD spanning >10 cM in dairy cattle, and there was extensive LD in the US Holstein population reported by Vallejo *et al.* (2003). These results in livestock species contrast starkly with the extent of LD in human populations, which ranges from 3–5 kb to hundreds of kb (e.g. Pritchard & Przeworski 2001; Ardlie *et al.* 2002; Kaessmann *et al.* 2002). The longer-range LD in livestock than in humans might be caused by evolutionary forces such as genetic drift, admixture, selection and small effective population size, which are common in livestock (Haley 1999).

Although the high levels of LD observed in some livestock species can be exploited for fine-mapping QTL and MAS, little is known about the degree of LD in beef cattle. In Japan, there are two main beef breeds: the Japanese Black and the Japanese Brown. The Japanese Black is famous for high quality meat with prominent intramuscular fat deposition or marbling, whereas the Japanese Brown has a larger mature size and a faster growth rate than the Japanese Black. The objective of this study was to assess and compare genome-wide LD between the Japanese Black and the Japanese Brown cattle breeds.

Materials and methods

Data

The Japanese Black pedigree consisted of one sire and his 162 half-sib progeny with 31 maternal grand-sires sampled from 30 fattening farms in Oita Prefecture. The progeny were genotyped, but the dams of the progeny were not. A battery of 246 autosomal microsatellite loci was used to measure genome-wide LD. These 246 markers were distributed across 29 chromosomes. The Japanese Brown pedigree consisted of one sire and his 406 half-sib progeny with 71 maternal grand-sires with known genotypes sampled from 14 fattening farms in Kumamoto Prefecture. The dams of the progeny were not genotyped. A battery of 156 autosomal microsatellite loci dispersed over the genome was used. Microsatellite genotyping was performed as described (Ihara *et al.* 2004). Marker order and interval were determined according to the Shirakawa–USDA linkage map (Ihara *et al.* 2004).

Haplotype reconstruction

Initially, the marker linkage phase of each sire was reconstructed by identifying the frequently cosegregating alleles at linked loci in his half-sib progeny. The paternal allele at each marker was identified in all progeny. In the event that the inherited paternal allele was ambiguous because the sire and his offspring had the same heterozygous genotype, we selected an inherited paternal allele one-by-one conditionally on the transmission probability based on information of the ascertained paternal allele at adjacent marker(s) and the recombination probability between the markers. From this, the paternal haplotype for each offspring was inferred. The maternal allele was inferred by removing the paternal allele from the genotype of offspring. LD was assessed with haplotypes of the dams. Repeating this process 100 times (multiple imputations) resulted in 100 sample sets of maternal haplotypes for every offspring. We assessed LD for the 100 sample sets and obtained estimates by taking the average of 100 values of LD.

LD analysis

Frequencies of alleles and pair-wise haplotypes were estimated from their counts in the maternal gametes within each breed. Both LD coefficients and statistical significance of allelic associations between markers were computed as recommended by McRae *et al.* (2002). The LD coefficient as measured by D' allows simple comparison with the results of Farnir *et al.* (2000).

Following Hedrick (1987), LD between two multiallelic loci A and B was measured as:

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

where u and v are the respective number of alleles at the two marker loci, p_i and q_j are the frequencies of marker i allele at locus A and marker allele j at locus B respectively, and $|D'_{ij}|$ is the absolute value of Lewontin's (1964) normalized LD measure calculated as:

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

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

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

with x_{ij} being the frequency of pair-wise haplotype A_iB_j.

The significance level (α) of allelic associations was estimated using the Monte-Carlo approximation of Fisher's exact test for contingency tables (Slatkin 1994). This approach treats the observed counts of pair-wise haplotypes in a population as a sample of a multinomial distribution and their probability can be obtained from the distribution. The value of α for a given marker pair can be estimated as the cumulative probability of finding a table with the same marginal and total allele counts that has a probability equal to or lower than that of the observed table (Weir 1996). The estimation of α in this study was based on the simulations of 17 000 contingency tables under the null hypothesis of random allelic association.

The computing algorithm was based on the conventional Monte-Carlo method (Guo & Thompson 1992) with a minor modification for the test of LD. The observed cumulative frequency distribution of α-values was compared with that expected under the hypothesis of random allelic association as done by Farnir *et al.* (2000), McRae *et al.* (2002) and Nsengimana *et al.* (2004).

Least-squares analysis was used to test the D' values among syntenic marker pairs, which was used to test for interpopulational and interchromosomal variation in LD. Breed and chromosome were treated as fixed factors and the log-transformed distance between markers was used as a covariate in the general linear model:

$$D'_{ijk} = \mu + b_i + c_j + m(\log m_k - \log \bar{m}) + b_i c_j + b_i m(\log m_k - \log \bar{m}) + c_j m(\log m_k - \log \bar{m}) + \varepsilon_{ijk},$$

where D'_{ijk} is the D' value between two markers separated by distance k on chromosome j in breed i, μ is the average D' value across all the pairs of syntenic loci along the 29 chromosomes in the two breeds, b_i is the effect of breed i, c_j is the effect of chromosome j, m is the partial regression coefficient on marker distance, m_k is the genetic map distance k, $\bar{m}$ is the average distance between the markers, and ε_{ijk} is the residual. The value of D'_{ijk} was adjusted by the number of haplotypes according to the model of McRae *et al.* (2002). The $\log_{10}$-transformed distance was used instead of the original distance because of a linear relationship between D' and the log-transformed distance (e.g. McRae *et al.* 2002). Least-squares analysis of variance was performed using the GLM procedure of SAS (SAS Institute, Inc., Cary, NC, USA).

Results

The average number of alleles observed in the half-sibs was 6.3 and 6.7 for the Japanese Black and the Japanese Brown respectively, whereas the average heterozygosity of the half-sibs was 0.646 and 0.705 respectively. Total map length and the average interval between markers were 2820 cM (Haldane map) and 12.4 cM for the Japanese Black and 2795 and 20.3 cM for the Japanese Brown respectively.

Linkage disequilibrium was estimated for 1000 syntenic and 29 135 non-syntenic marker pairs in Japanese Black and 372 syntenic and 11 718 non-syntenic marker pairs in Japanese Brown. Of the 1000 syntenic marker pairs in Japanese Black, 96 pairs were separated by <10 cM and 157 pairs by 10–20 cM. Of 372 syntenic marker pairs in Japanese Brown, 11 pairs were separated by <10 cM and 58 pairs by 10–20 cM.

The extent of LD was measured for syntenic marker pairs within breed. The distribution of the D' value as a function of genetic map distance in the two breeds is shown in Fig. 1a. High levels of LD were found, which ranged over several tens of cM in both breeds. The D' value declined rapidly between 5- and 20-cM spaced markers, gradually between 20- and 40-cM spaced markers, and finally reached a more or less constant value for greater spaced markers. The difference in the D' value between the two breeds was about 0.07 after reaching a constant value. The mean values (SD) of the D' value for syntenic marker pairs were 0.251 (0.109) and 0.163 (0.075) for the Japanese Black and the Japanese Brown, respectively, with a mean difference of

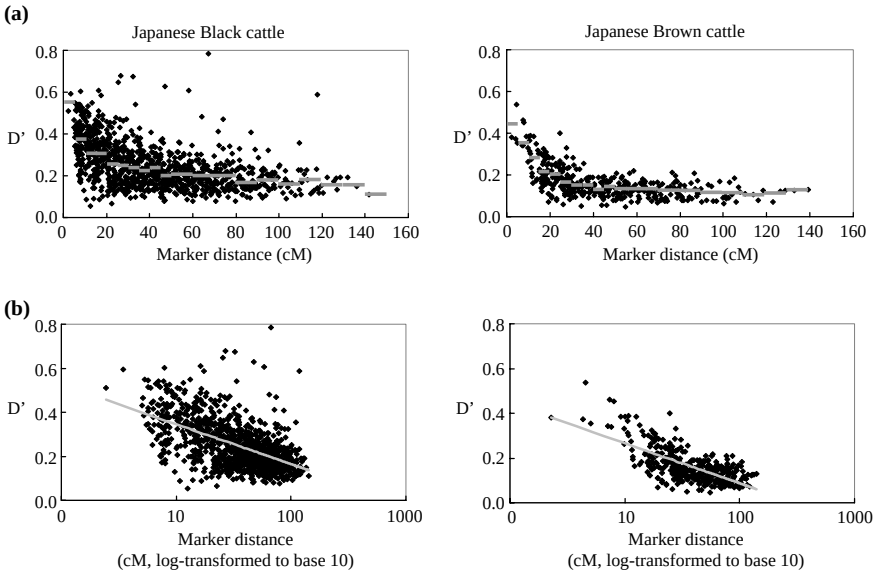

Figure 1 (a) Distributions of observed D' between syntenic marker pairs as a function of genetic map distance (cM) in Japanese Black and Japanese Brown cattle. The grey bars correspond to the average D' values for marker pairs every 5 cM (0–50 cM) or every 10 cM (<50 cM). (b) Distributions of the D' value between syntenic marker pairs as a function of log-transformed marker distance. The grey line is the regression of the D' value on log-transformed marker distance.

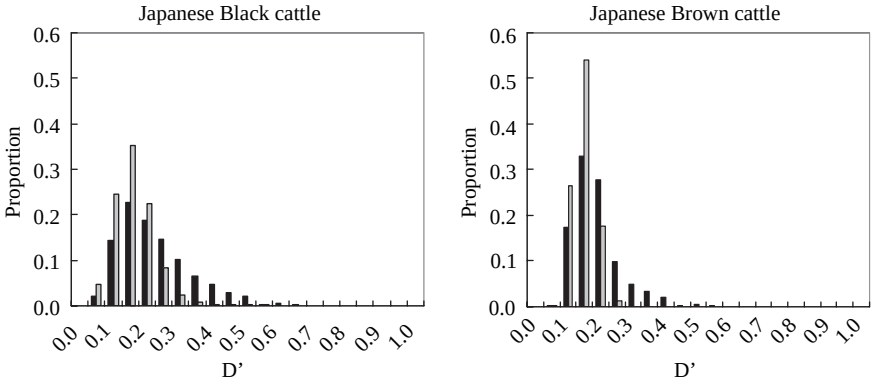

Figure 2 Frequency distributions of the *D′* value in Japanese Black and Japanese Brown cattle. Syntenic marker pairs are indicated by black bars, while non-syntenic marker pairs are indicated by grey bars.

0.088. The relationship between log-transformed marker distance and the *D′* value within breed is shown in Fig. 1b. The regression of *D′* on log-transformed marker distance was $y = -0.1812\ (0.0085)x + 0.527\ (0.0133)$ for the Japanese Black and $y = -0.1785\ (0.0096)\ x + 0.4444\ (0.0154)$ for the Japanese Brown. (The standard error is shown in the parenthesis.) Both regression coefficients were highly significant ($P < 0.0001$), suggesting that *D′* declined significantly as the marker distance increased. There was no significant breed difference in the regression coefficients.

Gametic disequilibrium was evaluated for non-syntenic marker pairs. The mean values (SD) of the *D′* value for non-syntenic marker pairs were 0.189 (0.080) and 0.122 (0.033) for Japanese Black and Japanese Brown, respectively, with a difference of 0.067 between the two breeds. Although the frequency distributions of the *D′* value for syntenic and non-syntenic marker pairs overlapped substantially as shown in Fig. 2, the average *D′* value of syntenic loci was significantly higher ($P < 0.0001$) than that of non-syntenic loci for both breeds according to *t*-test with Welch's correction (Ichihara 2001).

Linkage disequilibrium was tested based on the statistical significance of allelic associations between markers. The cumulative distribution of *P*-values from the significance tests is shown in Fig. 3. For syntenic marker pairs, *P*-values were grouped by marker distance. In both breeds, significant LD was observed frequently for marker pairs <40 cM apart, as the cumulative frequency of *P*-values departed largely from the distribution expected under the random allelic association. Overall, significant LD was observed more frequently in Japanese Brown than in Japanese Black. About 5.5% (about the same frequency as expected by chance) of the studied non-syntenic marker pairs showed significant LD at a 5% significance level in the Japanese Black when compared with 10.8% in the Japanese Brown.

Most of the variation in *D′* was explained by marker distance ($P < 0.001$; Table 1). The effect of chromosome, the interaction between breed and chromosome, and the interaction between chromosome and marker distance were also significant ($P < 0.001$), while the effect of breed was not significant ($P > 0.05$) (Table 1). Three-way interaction of the three main factors was non-significant as reported by Nsengimana *et al.* (2004) and was excluded from our model.

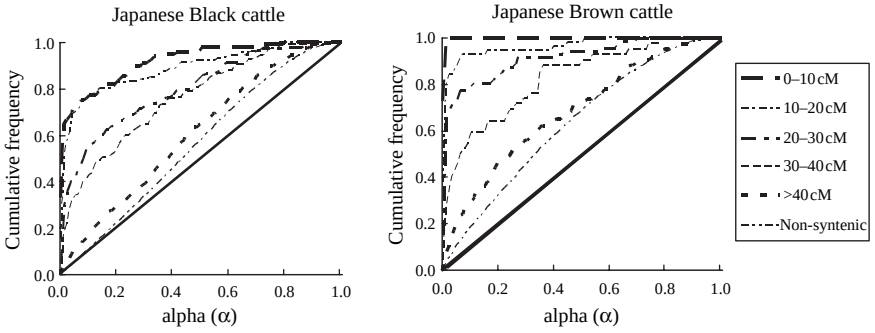

Figure 3 Cumulative frequency distributions of α-values from the significance test of allelic associations between syntenic and between non-syntenic marker pairs in Japanese Black and Japanese Brown cattle. Syntenic marker pairs were grouped by genetic map distance. The diagonal line represents the expected distribution under random allelic association.

Table 1 The least squares ANOVA table for the D' value among syntenic marker pairs.

Source of variation	d.f.	Mean square
Breed	1	0.000335
Chromosome	28	0.0210*
Marker distance	1	2.122*
Breed × chromosome	28	0.0132*
Breed × marker distance	1	0.00458
Chromosome × marker distance	28	0.0177*
Residual	1284	0.00606

D', linkage disequilibrium coefficient; d.f., degrees of freedom.
*$P < 0.001$.

Discussion

A genome-wide search for LD was performed on Japanese Black and Japanese Brown cattle. The transmission of maternal haplotypes to the offspring, which were sampled from many paternal half-sib progeny was used to measure LD, where the dams were considered to be a random sample of the female population. Multiple imputations were employed for the haplotype reconstruction of offspring in this study. Comparison of the results between single and multiple imputations indicates that the general pattern of the distribution of the D' value between the two methods was similar for each of the two breeds (results not shown). However, the D' value between pairs of markers, of which haplotype was ambiguous in more animals, largely fluctuated from imputation to imputation, suggesting that a single imputation must produce biased estimate of the D'. In contrast, multiple imputations can account for the uncertainty by means of complementing multiple times to one contingent value (Rubin 1996).

High levels of LD were found among syntenic loci in both breeds. The general pattern of the distribution of D' was similar between the two breeds. However, breed difference exists in the mean values of the D' value for both syntenic and non-syntenic marker pairs and in the D' value after reaching a constant value. McRae *et al.* (2002) showed that D' can be biased upwards when measured with a small number of haplotypes. These differences correspond approximately with the predicted difference of 0.06 from the model of McRae *et al.* (2002) considering the number of haplotypes in this study. This explained most of the actual difference between two breeds, which may be due to sample size difference. Furthermore, breed difference in average interval between markers could also be a contributing factor in the case of syntenic marker pairs. When the marker distance was lower, the D' value was higher (Fig. 1b).

Table 2 Effect of group of the chromosome with or without region(s) in which QTL for marbling were detected on the D' value among sytenic marker pairs.

Source of variation	d.f.	Mean square
Group	1	0.0441*
Marker distance	1	3.386**
Group × marker distance	1	0.0298
Residual	996	0.00817

D', linkage disequilibrium coefficient; QTL, quantitative trait loci; d.f., degrees of freedom.
*$P < 0.05$; **$P < 0.001$.

Linkage disequilibrium between non-syntenic loci was not significant in Japanese Black, but it was significant in Japanese Brown (about twice as often as expected under random allelic association). Ardlie *et al.* (2002) noted that P-values obtained from the test of significant departure from linkage equilibrium between loci depend largely on sample size and even a weak LD could become statistically significant due to a sufficiently large sample. Therefore, the discrepancy in significance tests between the two breeds is also attributable to the difference in sample size rather than in background genotypes.

A genome-wide search by Farnir *et al.* (2000) in dairy cattle showed similar pattern of LD to this study. The observable difference is that LD between non-syntenic loci was highly significant in dairy cattle but not in Japanese beef cattle. The study of Farnir *et al.* (2000) had a greater power of test than this study because they used 581 and 1254 haplotypes when compared with 162 and 406 haplotypes in this study. Nsengimana *et al.* (2004) reported that the difference in statistical significance between their study and Farnir *et al.* (2000) seems to be explained by the number of haplotypes. The results from our study were in agreement with the findings of population-wide LD in dairy cattle (Farnir *et al.* 2000; Hayes *et al.* 2003; Tenesa *et al.* 2003; Vallejo *et al.* 2003) and in other livestock populations (McRae *et al.* 2002; Lou *et al.* 2003; Nsengimana *et al.* 2004) in terms of the level of LD extended over great genetic map distances (about 20–40 cM).

The significant interaction between breed and chromosome indicates that there exists interchromosomal heterogeneity in the LD and that the heterogeneity differs between the two breeds. As described before, the two breeds studied have distinctive, different characteristics in terms of meat productivity and selection history, although they originated from the same native cattle in Japan. The interchromosomal heterogeneity in the LD between the two cattle breeds may result from differential selection, which was in agreement with the report of Nsengimana *et al.* (2004) in pigs.

Historically, the Japanese Black has been subjected to intensive selection for marbling. Using the same half-sib family of the Japanese Black, QTL associated with marbling were detected on seven chromosomes at a 5% chromosome-wise significance level (K. Yokouchi, T. Watanabe, T. Fujita, K. Shiga & Y. Sugimoto, personal communication, 2004). In order to elucidate the relationship between selection and the heterogeneity of the D' values, the 29 chromosomes

were divided into two groups with or without a previously detected QTL region. As shown in Table 2, there was significant difference ($P < 0.05$) between the two groups in the D' value among syntenic marker pairs, indicating that selection for marbling is relating with heterogeneity of the LD between chromosomes in the Japanese Black.

Acknowledgements

We are very grateful to Dr C.Y. Lin for the helpful advice and comments on the manuscript. The work was partly supported by the Ministry of Agriculture, Forestry and Fishery, Japan, and by the Japan Racing and Livestock Promotion Foundation.

References

Andersson L. (2001) Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* **2,** 130–8.

Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3,** 299–309.

Farnir F., Coppieters W., Arranz J.J. *et al.* (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10,** 220–7.

Fernando R.L. & Grossman M. (1989) Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution* **21,** 467–77.

Grisart B., Farnir F., Karim L. *et al.* (2004) Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 2398–403.

Guo S.W. & Thompson E.A. (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48,** 361–72.

Haley C. (1999) Advances in quantitative trait locus mapping. In: *From J. L. Lush to Genomics: Visions for Animal Breeding and Genetics* (Ed. by J.C.M. Dekkers, S.J. Lamont & M.F. Rothschild), pp. 47–59. Iowa State University, Ames, IA.

Hayes B.J., Visscher P.M., McPartlan H.C. & Goddard M.E. (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* **13,** 635–43.

Hedrick P.W. (1987) Gametic disequilibrium measures. Proceed with caution. *Genetics* **117,** 331–41.

Ichihara K. (2001) *Statistics for Bioscience*, pp. 71–114. Nankodo, Tokyo.

Ihara N., Takasuga A., Mizoshita K. *et al.* (2004) A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research* **14,** 1987–98.

Kaessmann H., Zollner S., Gustafsson A.C., Wiebe V., Laan M., Lundeberg J., Uhlen M. & Paabo S. (2002) Extensive linkage disequilibrium in small human populations in Eurasia. *American Journal of Human Genetics* **70,** 673–85.

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. *et al.* (2003) The extent and distribution of linkage disequilibrium in a multi-hierarchic outbred canine pedigree. *Mammalian Genome* **14,** 555–64.

McRae A.F., McEwan J.C., Dodds K.G., Wilson T., Crawford A.M. & Slate J. (2002) Linkage disequilibrium in domestic sheep. *Genetics* **160,** 1113–22.

Nsengimana J., Baret P., Haley C.S. & Visscher P.M. (2004) Linkage disequilibrium in the domesticated pig. *Genetics* **166,** 1395–404.

Pritchard J.K. & Przeworski M. (2001) Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* **69,** 1–14.

Rubin D.B. (1996) Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91,** 473–89.

Slatkin M. (1994) Linkage disequilibrium in growing and stable populations. *Genetics* **137,** 331–6.

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. *Journal of Animal Science* **81,** 617–23.

Terwilliger J.D. & Weiss K.M. (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? *Current Opinion in Biotechnology* **9,** 578–94.

Vallejo R.L., Li Y.L., Rogers G.W. & Ashwell M.S. (2003) Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population. *Journal of Dairy Science* **86,** 4137–47.

Van Laere A.S., Nguyen M., Braunschweig M. *et al.* (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425,** 832–6.

Weir B.S. (1996) *Genetic Data Analysis II. Methods for Discrete Population Genetic Data*, p. 117. Sinauer Associates Inc. Publishers, Sunderland, MA.