Linkage Disequilibrium in the Domesticated Pig

Jérémie Nsengimana,* Philippe Baret,* Chris S. Haley† and Peter M. Visscher‡,1

*Université Catholique de Louvain, Faculté d'Ingénierie Biologique, Agronomique et Environnementale, Unité de Génétique, 1348 Louvain-la-Neuve, Belgium, †Roslin Institute (Edinburgh), Midlothian EH25 9PS, United Kingdom and ‡Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

Manuscript received July 23, 2003
Accepted for publication December 9, 2003

ABSTRACT

This study investigated the extent of linkage disequilibrium (LD) in two genomic regions (on chromosomes 4 and 7) in five populations of domesticated pigs. LD was measured with D' and tested for significance with the Fisher exact test. Effects of genetic (linkage) distance, chromosome, population, and their interactions on D' were tested both through a linear model analysis of covariance and by a theoretical nonlinear model. The overall result was that (1) the distance explained most of the variability of D', (2) the effect of chromosome was significant, and (3) the effect of population was significant. The significance of the chromosome effect may have resulted from selection and the significance of the population effect illustrates the effects of population structures and effective population sizes on LD. These results suggest that mapping methods based on LD may be valuable even with only moderately dense marker spacing in pigs.

LINKAGE disequilibrium (LD) is population-wide nonrandom association of alleles at different genetic loci. Measures of LD can provide information on population structure and dynamics, including effective population size, and can be used to map genes or quantitative trait loci (QTL). Genome-wide LD studies in livestock have shown that LD extends over large genetic map distances (>30 cM) in sheep (McRae *et al.* 2002) and dairy cattle (Farnir *et al.* 2000) populations. In both studies, LD was common and statistically significant. Long-range LD was also found for two genomic regions (chromosomes 4 and 6) in a sample from the dairy cattle populations in the United Kingdom (Tenesa *et al.* 2003). Hayes *et al.* (2003) inferred past effective population size in dairy cattle from haplotype frequencies and also detected LD spanning >10 cM. To our knowledge, these are the only studies that have explored the level of LD in livestock and the results obtained contrast sharply with the extent of LD in human populations, which ranges from 3–5 kb to hundreds of kilobases (*e.g.*, Pritchard and Przeworski 2001; Reich *et al.* 2001; Ardlie *et al.* 2002; Kaessmann *et al.* 2002). LD might exist at a larger distance in livestock than in human populations due to intensive artificial selection accompanied by a reduction in effective population size (Boichard *et al.* 1996; Haley 1999).

Although Holstein-Friesian dairy cattle are under intensive selection, linkage analyses have indicated the presence of QTL that are still segregating (*e.g.*, Georges *et al.* 1995; Coppieters *et al.* 1998). Farnir *et al.* (2000) tested whether the nonsyntenic LD observed in this population between chromosome regions harboring QTL—and potentially coselected—is greater than the LD between anonymous regions. However, no evidence of a selection effect was found and a simulation study indicated that random drift alone can explain the observed LD. Nevertheless, as outlined by these authors, a lack of evidence about the selection effect on LD does not imply the absence of this effect. For example, there may not have been sufficient statistical power to detect a significant selection effect. It is well established that selection can cause LD between unlinked loci that contribute to phenotypes undergoing selection (*e.g.*, Lewontin 1964; Bulmer 1971; Ardlie *et al.* 2002).

While Farnir *et al.* (2000) did not analyze the effect of selection on LD between linked loci (hitchhiking effect; see Guiyun *et al.* 1998; Ardlie *et al.* 2002), some observations provide evidence of this effect in UK dairy cattle (Tenesa *et al.* 2003). LD was most significant between markers lying in the region known to harbor QTL involved in milk yield and composition (Wiener *et al.* 2000; Tenesa *et al.* 2003). However, the LD analysis by Tenesa *et al.* (2003) was based on a small number of individuals and markers, so random sampling effects are likely to be large.

In domestic sheep, McRae *et al.* (2002) observed long-range LD in two data sets from two different breeds. The first breed was Coopworth, which is a young hybrid between the breeds Border Leicester and Romney (~10 generations old). Given the young age of this population and the intensive selection that reduced its effective size, the observed high level of LD was not surprising. The second breed was Romney, which is also under

[1]*Corresponding author:* University of Edinburgh, Institute of Cell, Animal and Population Biology, W. Mains Rd., Edinburgh EH9 3JT, United Kingdom. E-mail: peter.visscher@ed.ac.uk

TABLE 1

Characteristics of analyzed populations

Population	Breed	Mgc[a]	N_e[b]	Sires	Dams	Progeny	QTL[c]
A	Large White	8	200	10	156	431	7G, 7F
B	Duroc/Large White	5	85	10	166	421	4G, 4F, 7F
C	Yorkshire/Large White	5	60	10	94	385	7F
D	Large White	8	300	11	141	429	7G
E	Landrace	10	190	12	135	461	4G, 7F

[a] Minimum number of generations closed.
[b] Effective population size estimated by breeding companies.
[c] Chromosome (4 or 7) on which Nagamine *et al.* (2003) found evidence of segregating QTL for growth rate (G) or back fat (F). There is more evidence of both QTL on SSC7 than on SSC4.

intensive selection with a smaller population size than Coopworth. A lower LD was expected in this parental line, compared to its daughter line (Coopworth). However, the observed LD was of the same magnitude in both breeds, indicating a greater impact of the reduction in the population size compared to admixture. The direct effect of selection was not analyzed. The high level of LD observed in all aforementioned studies on LD in livestock may be utilized to perform fine-mapping studies of QTL. This was supported by the rapid decline of LD at low genetic map distance (5–10 cM), while it was constant at larger distances (Farnir *et al.* 2000; McRae *et al.* 2002).

Commercial pigs are under intense selection in populations that are typically of small effective size (<100). However, hybridization has occurred in the past and occasionally new synthetic lines are created through crossbreeding. This study aims to assess the level of LD in five populations of commercial pigs. Two chromosome regions were investigated, one on chromosome 4 (SSC4) and one on chromosome 7 (SSC7). As these regions have been reported to harbor QTL affecting growth rate and fat deposition in a number of pig breeds including the analyzed populations (Knott *et al.* 1998; Walling *et al.* 2000; Nagamine *et al.* 2003), it might be expected that the selection has influenced LD. In addition to estimating LD in these populations, the joint effects of genetic map distance, chromosome, and population on LD were quantified.

MATERIALS AND METHODS

Data: We used the same data as those used in a previous study on QTL variation for growth rate and obesity between and within lines of pigs (Nagamine *et al.* 2003). These data consisted of samples from five different populations provided by five different pig genetic companies. The five populations were either pure European breeds or established synthetic lines obtained by crossing European breeds at least 10 years before the study took place (Table 1). In the analyzed region of SSC4, there was evidence of the segregation of QTL influencing growth rate and back fat deposition in one population (population B) and a QTL for growth rate in one additional population (population E; see Table 1 and Nagamine *et al.* 2003). On SSC7, there was evidence for QTL in all five populations. There was strong evidence of QTL for both growth rate and back fat deposition in one population (population A), a QTL for back fat in three other populations (populations B, C, and E), and a QTL for only growth rate in one additional population (population D; see Table 1). The population structure for each of the five samples was composed of a number of full-sib and half-sib families.

Sires, dams, and their male progeny were genotyped for 15 microsatellite markers, chosen for their heterozygosity and technical tractability, which spanned 68 cM (29 cM on SSC4 and 39 cM on SSC7) as described by Nagamine *et al.* (2003). The goal was to have at least 5 informative markers per chromosome; thus the entire set of 15 markers was genotyped in a few individuals. Overall, missing genotypes amounted to 10–25%, depending on the population and on the chromosome. All sires and dams were genotyped while only progeny with extreme phenotypes were genotyped (≤25% of the lower and upper tail of the distribution).

In each population, a linkage map was estimated with the CRI-MAP package (Green *et al.* 1990) and compared to the published maps (http://www.thearkdb.org). A joint linkage analysis of all marker data across all five populations provided a consensus map that was used in subsequent analyses (see Table 2 and Nagamine *et al.* 2003).

Locus heterozygosity was estimated as the proportion of individuals with two different alleles at the locus among parents. An average heterozygosity across all markers of the same chromosome was computed for each population. At every marker locus and within each population, genotype proportions were tested for Hardy-Weinberg equilibrium (HWE), using an exact test (Guo and Thompson 1992).

Haplotype reconstruction: Multilocus haplotypes in each half-sib offspring were determined from its genotype and those of its parents. "Diplotypes" are defined as phased genotypes, *i.e.*, multilocus genotypes with reference to the haplotypes on which the alleles reside. In 95% of progeny/marker combinations, the paternal or maternal origin of each allele in the offspring was unambiguous. In the remaining 5%, the offspring and both parents were heterozygous for the same alleles and the parental origin of the alleles could not be resolved. These alleles were ignored in subsequent analyses.

By grouping progeny of each sire and of each dam, we obtained a set of gametes transmitted by each parent. Among these gametes, some will be exact copies of the parental haplotypes while others are recombinants. In estimating LD, only parental haplotypes were used as they represent a sample from the outbred population. Using only parental haplotypes in LD estimation makes the study independent of the selective genotyping of progeny, since all parents were genotyped irre-

TABLE 2

Genetic markers used in the study and their relative positions

Marker	Relative position (cM)	
	This study	USDA map[a]
	SSC4	
S0001	0	0
SW45	12	14
SW35	12	14
SW839	16	20
S0107	17	24
S0217	20	28
SW841	24	29
S0073	29	33
	SSC7	
SW1354	0	0
S0064	6	8
SWR1078	9	11
SWR1344	17	26
TNFβ	28	36
SW2019	30	38
S0102	39	48

[a] See http://www.genome.iastate.edu/maps/marcmap.html

spective of their own phenotypes. Haplotypes from each sire and each dam were identified using a simple algorithm, based on a comparison of their genotypes to those of their mates and their progeny.

Linkage disequilibrium analysis: Allele frequencies and pairwise haplotype frequencies were estimated from their counts in the parental generation for each population. For a pair of loci A and B, D' was estimated as

$$D' = D'_{AB} = \sum_{i}^{N_A}\sum_{j}^{N_B} p_i q_j |D'_{ij}|$$

with

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

and

$$D_{ij} = p_{ij} - p_i q_j$$
$$D_{max} = \min[p_i q_j, (1 - p_i)(1 - q_j)] \quad \text{if } D_{ij} < 0$$
$$D_{max} = \min[(1 - p_i) q_j, p_i (1 - q_j)] \quad \text{if } D_{ij} > 0,$$

where p_i and q_j are frequencies of alleles i and j on markers A and B, respectively, p_{ij} is the frequency of the pairwise haplotype ij, and N_A and N_B are the total numbers of alleles at markers A and B, respectively (Lewontin 1964; Hedrick 1987).

The statistical significance of allelic associations was estimated with the Monte Carlo extension of the Fisher exact test for contingency tables (Slatkin 1994) implemented in the ARLEQUIN software (Schneider *et al.* 2000). For this approach, the observed counts of pairwise haplotypes in a given population constitute a sample of a multinomial distribution and their probability can be obtained from this distribution. The statistical significance (P value) of the allelic association is estimated as the cumulative probability of observing the sample or any less likely sample with the same marginal and total haplotype counts (Weir 1996). However, in ARLEQUIN (Schneider *et al.* 2000), instead of enumerating all possible samples less likely than the analyzed sample, a Monte Carlo chain is used to explore efficiently the space of all possible contingency tables (Slatkin 1994). Following Farnir *et al.* (2000) and McRae *et al.* (2002), the P values of the test statistic were not corrected for multiple testing, because a too stringent type I error rate may result in loss of power to detect LD (Tenesa *et al.* 2003).

Effects on D' of marker distance, chromosome, and population as well as their interactions were tested using two different methods: (i) analysis of covariance with a general linear model and (ii) fitting a theoretical nonlinear model to the data. For the linear model, the population and the chromosome were analyzed as fixed factors while the log-transformed distance between markers was a covariate,

$$D'_{ijkl} = \mu + c_i + p_j + \log d_k + c_i p_j + c_i \log d_k + p_j \log d_k + c_i p_j \log d_k + \varepsilon_{ijkl},$$

where D'_{ijkl} is LD between two markers separated by distance k on chromosome i in population j, μ is the average LD across all pairs of syntenic loci along the two chromosomes in all populations, c_i is the mean effect of chromosome i, p_j is the mean effect of population j, d_k is the mean effect of genetic map distance k, and ε_{ijkl} is the residual. Each value of D'_{ijkl} is weighted by the number of haplotypes used in its estimation. This model assumes normality of residuals and homogeneity of variance. We fitted the effect of a log-transformed distance rather than the distance itself because a linear relationship is expected between D' and the log-transformed distance (see, *e.g.*, McRae *et al.* 2002).

For the nonlinear model, estimates of D' were fitted as an exponential function of genetic distance. We estimated the parameters of the model and tested the effects of the population and the chromosome on these parameters in a two-step procedure. First, the parameters from the nonlinear model were estimated for each population-chromosome combination, using a least squares approach, and second, the estimated parameters were treated as dependent variables in a linear model. In theoretical and simulation studies, it was shown that patterns of LD with respect to the genetic distance can be fitted with an exponential covariance function, commonly used to model spatial processes (Morton *et al.* 2001; Nsengimana and Baret 2002). The standardized form of such a covariance function was used in this study to fit the estimates of D',

$$D' = rs + (1 - rs)\exp\left(\frac{-3d}{R}\right),$$

where d is the genetic distance, rs is the *residual* D' corresponding to the spatially independent component, and R is the *range, i.e.*, the distance at which the spatially correlated part of D' is equal to 5% of its maximum value (Christakos 1992). This model was applied separately to the estimates of D' along SSC4 and SSC7 within each population and parameters rs and R were estimated through a least-squares approach, weighted by the number of haplotypes used to estimate each value of D'. Using a multivariate analysis of variance (MANOVA), we tested the hypothesis that the joint parameters $\{rs, R\}$ were different between chromosomes and between populations (both fixed factors),

$$R = \mu_R + c_i + p_j + \varepsilon_R$$
$$rs = \mu_{rs} + c_i + p_j + \varepsilon_{rs},$$

with μ_{rs} and μ_R the means of rs and R on the two chromosomes and across the five populations, c_i the mean effect of the chromosome i, p_j the mean effect of population j, and ε_{rs} and

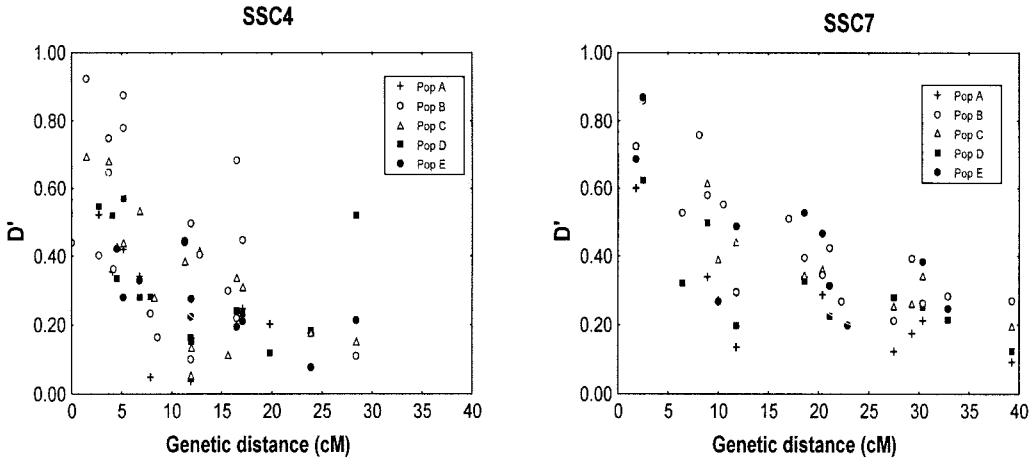

FIGURE 1.—The coefficient D' between syntenic markers on SSC4 and SSC7 within each of the five populations (A–E).

ε_R residuals of the models. The sampling correlation between these two parameters (R and rs) was estimated.

RESULTS

The mean numbers of alleles per marker on SSC4 and SSC7 were 5.5 and 10.7, respectively. The locus heterozygosity in the parents varied from 0.56 to 0.68 on SSC4 and from 0.65 to 0.80 on SSC7 according to population. The highest locus heterozygosity was observed in populations C (synthetic Yorkshire/Large White) and E (Landrace), while the lowest was observed in population A (Large White). In the five populations, 62 tests of HWE genotype proportions were performed and only 3 of them were significant at the 5% type I error rate: marker SW35 (on SSC4) with a P value of 0.008 in population D and marker SWR1078 (on SSC7) with P values of 0.02 in population B and 0.003 in population C. These tests showing the absence of HWE represent 4.8% of the total, practically the same as the frequency expected by chance.

The coefficient D' was estimated between ~25 pairs of syntenic markers and between ~30 nonsyntenic pairs within each population (124 syntenic and 164 nonsyntenic in all five populations). Along each of the two chromosomes, D' decreased as the distance between loci increased (Figure 1). The highest observed values of D' were similar on both chromosomes and they correspond to a distance close to zero. However, the decline of LD as a function of the marker distance was faster on SSC4 than on SSC7. This is also shown by the mean D' across all pairs of syntenic loci. The mean D' is higher on SSC7 than on SSC4 and this difference is highly significant in populations C and E (Table 3).

The lowest mean D' on both chromosomes and between nonsyntenic markers was observed in population A while population B had the highest (Table 3). Population E is unusual in that it had the second-lowest mean D' on SSC4 with the second-highest mean D' on SSC7 and between nonsyntenic markers.

Using a Monte Carlo extension of the Fisher exact test, we estimated the significance level (P value) of the observed marker association. Under the null hypothesis of random allelic association, the expected cumulative distribution of P values is on the diagonal of each graph in Figure 2. The distribution of the observed P values between nonsyntenic markers was close to this diagonal, while the distribution corresponding to syntenic markers on SSC4 and SSC7 departed from this diagonal, with the lowest P values being overrepre-

TABLE 3

Mean D' and standard deviation between linked and unlinked markers

Population	Mean $D' \pm \sigma$ on SSC4	Mean $D' \pm \sigma$ on SSC7	SSC4 *vs.* SSC7 (P value)[a]	$D'_{unlink} \pm \sigma$[b]
A	0.213 ± 0.097	0.300 ± 0.098	0.050*	0.114 ± 0.050
B	0.393 ± 0.181	0.492 ± 0.151	0.097	0.223 ± 0.098
C	0.304 ± 0.126	0.441 ± 0.122	0.008**	0.142 ± 0.061
D	0.270 ± 0.118	0.349 ± 0.106	0.077	0.133 ± 0.057
E	0.267 ± 0.124	0.461 ± 0.101	<0.001***	0.166 ± 0.056
Average	0.290 ± 0.131	0.409 ± 0.116	0.027*	0.156 ± 0.062

[a] The significance of the difference in D' between chromosomes at P values of *0.05, **0.01, and ***0.001.
[b] D'_{unlink} is the mean D' between unlinked markers. For linked markers, the means are adjusted for the genetic map distance.

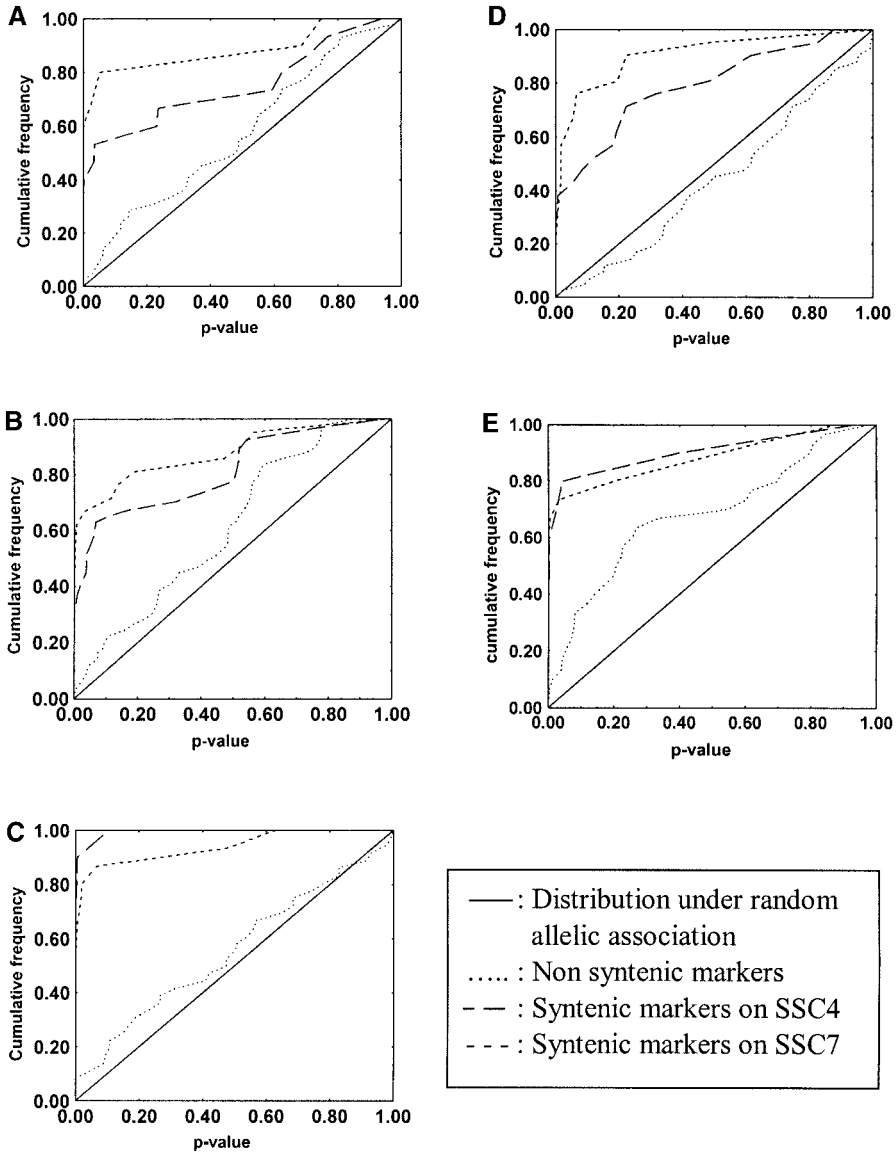

FIGURE 2.—Cumulative frequency of *P* values for LD significance on SSC4 and SSC7 and between nonsyntenic markers in the five populations (A–E).

sented (Figure 2). This indicates clearly significant LD between linked loci in all five populations and lower LD between unlinked loci.

A linear model was used to test the effects of distance, population, chromosome, and their interactions on D' (Table 4). Of the effects fitted, most of the differences in D' are explained by the genetic distance ($P < 0.0001$).

The chromosome effect on D' was also significant ($P = 0.027$) while the effects of the population and all the interactions were not significant (Table 4).

The significant difference between chromosomes may indicate a selection effect. In fact, effects of QTL underlying selected traits are significant on SSC7 in all five populations, while significant QTL on SSC4 are present in two populations only (see Table 1 and NAGAMINE *et al.* 2003). To further test the hypothesis of chromosome effect, we separated populations in two groups: group 1 includes the three populations where QTL were identified on SSC7 and no such QTL was on SSC4 (populations A, C, and D) and group 2 contains two populations for which QTL effects were significant on both SSC4 and SSC7 (populations B and E, see Table 1).

If there is a selection effect on LD, then we would expect a significant difference between chromosomes in group 1 and a nonsignificant difference in group 2 due to the absence (group 1) or presence (group 2) of QTL on SSC4. We obtained significance P values of 0.06 and 0.23 in group 1 and group 2, respectively (see Table 5). Although there is no significance at level 5% in both groups, this result indicates that an effect of selection cannot be discarded.

The exponential function was applied to the estimates of D' along each chromosome and within each population (Figures 3 and 4). This model fits the estimates of D' between all syntenic markers with a determination coefficient of 0.45–0.80.

According to this model, D' is 1 at the genetic map distance of zero and decreases with an increasing dis-

TABLE 4

Effects of the main factors and their interaction on *D′* between linked markers

Source of variation[a]	d.f.[b]	*F*	*P* value
Distance	1	144.37	<0.0001
Population	4	0.31	0.872
Chromosome	1	5.03	0.027
Distance × population	4	1.00	0.414
Distance × chromosome	1	0.21	0.645
Population × chromosome	4	0.07	0.991
Distance × population × chromosome	4	0.21	0.933

[a] The distance is logarithmically transformed.
[b] The residual degrees of freedom are 103.

tance to stabilize at a nonzero value (the residual LD) which varies between 0.150 and 0.215 on SSC4 and between 0.208 and 0.340 on SSC7 according to the population (Table 6).

The average of this component of D' between the two chromosomes and across all five populations is 0.222, greater than the mean D' between nonsyntenic markers (0.156 ± 0.062). This difference may be explained by the lack of information to infer *rs* accurately, as we covered 30 and 40 cM on SSC4 and SSC7, respectively, but *rs* corresponds theoretically to "very large map distances."

Under the exponential covariance model, the extent of the spatially correlated part of D' (*i.e.*, the range) varies from 9.6 to 21.8 cM on SSC4 and from 8.9 to 32.6 cM on SSC7, according to populations (Table 6).

Parameters *R* and *rs* are estimated simultaneously in a fitting procedure. In the five populations of pigs and for chromosomes SSC4 and SSC7, the relationship between *R* and *rs* is illustrated in Figure 5. On SSC4, there is a correlation of 0.87 between *R* and *rs* with a significance *P* value of 0.06, while this correlation is absent on SSC7 (corr = 0.00, $P = 1$). Overall, for both chromosomes, the correlation between *R* and *rs* is 0.38 and it is not significant ($P = 0.31$). To account for this relationship between *R* and *rs* when testing the effects of the population and the chromosome, we performed a MANOVA. Two different tests of this analysis were used: Wilks' lambda and Pillai's trace. Both tests are transformed into a Fisher test before the computation of a corresponding *P* value that indicates the significance level. Both effects of the chromosome and the population on {*R*, *rs*} are significant for each of the two tests, with *P* values of 0.05 for the population effect and 0.03 for the chromosome effect.

DISCUSSION

In this study we have quantified the extent of LD in two chromosomal regions in five commercial pig populations. To our knowledge, this is the first report of LD in pig populations. In all five populations and for both chromosomes, a high level of LD was observed between linked markers (Figure 1) and it was found to be significant, as the cumulative frequency of *P* values from the Fisher exact test departed from its expected distribution under the random allelic association (Figure 2). Between unlinked markers, LD was not significant since the cumulative frequency of *P* values was similar to its expectation under the hypothesis of the absence of LD (see Figure 2).

McRae *et al.* (2002) showed that D' can be upwardly biased when it is estimated with a small number of haplotypes. In this study, we used 184–302 haplotypes and, according to the model of McRae *et al.* (2002), these sample sizes may have introduced a bias of ~0.04–0.06 on D'. Since the sample size bias was the same for linked and unlinked markers, general conclusions of the study are expected to be robust with respect to sample size. As noted by McRae *et al.* (2002), many other studies have used the coefficient D' with a sample size as small

TABLE 5

Effects of the main factors and their interactions on *D′* between linked markers in two groups of populations

	Group 1[b]			Group 2[c]		
Source of variation[a]	d.f.[d]	*F*	*P* > *F*	d.f.[d]	*F*	*P* > *FS*
Distance	1	106.66	<0.0001	1	42.14	<0.0001
Population	2	0.10	0.90	1	0.67	0.42
Chromosome	1	3.67	0.06	1	1.48	0.23
Distance × population	2	1.58	0.21	1	0.10	0.75
Distance × chromosome	1	0.45	0.50	1	0.00	0.95
Population × chromosome	2	0.09	0.92	1	0.07	0.79
Distance × population × chromosome	2	0.03	0.97	1	0.70	0.41

[a] The distance is logarithmically transformed.
[b] Group 1: three populations with selected QTL on SSC7 only.
[c] Group 2: two populations with selected QTL on both SSC4 and SSC7.
[d] The residual degrees of freedom are 68 in group 1 and 35 in group 2.

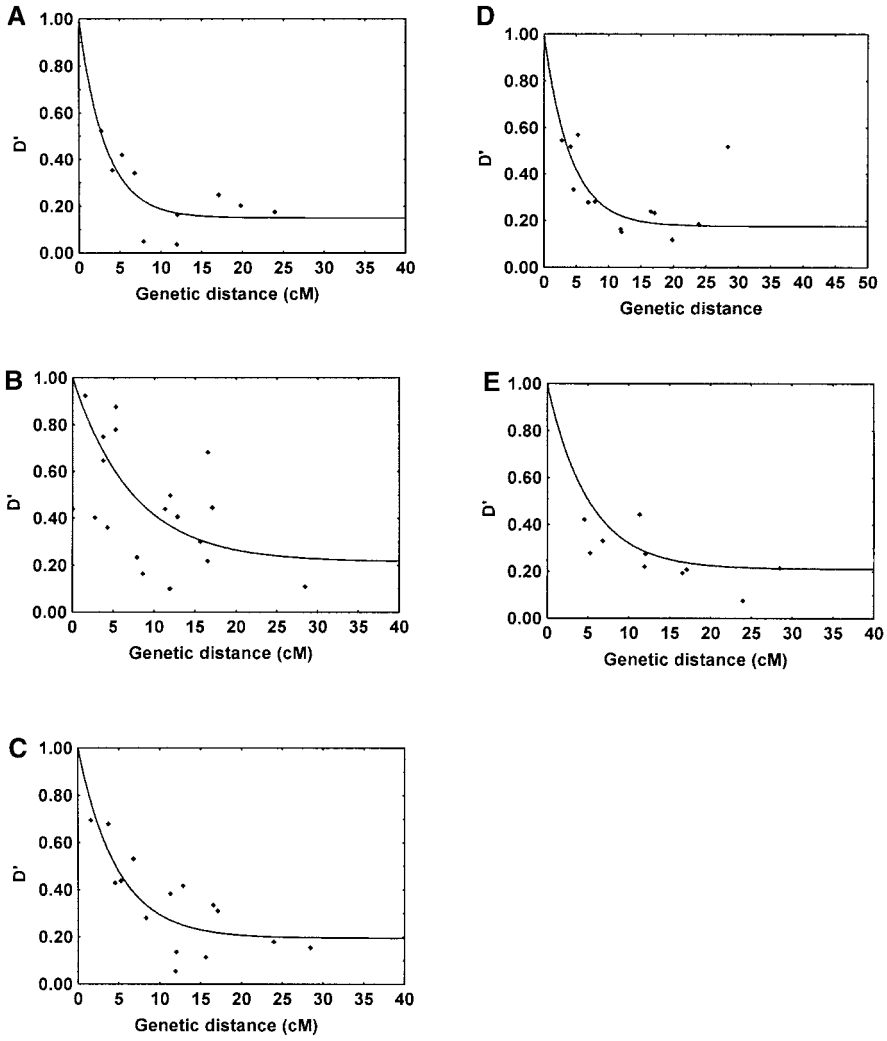

FIGURE 3.—The exponential function applied to D' between syntenic markers on SSC4 in the five populations (A–E).

as 50 haplotypes. When using the correlation measure r^2 for LD between two biallelic markers with a recombination fraction of θ, the relationship between LD, effective population size (N_e), and number of haplotypes (n) is, approximately,

$$E(r^2) = \mathrm{var}(r) = 1/(1 + 4N_e\theta) + 1/n$$

(following WEIR and HILL 1980). This expression clearly shows the effect of both finite population size and sample size on LD.

An ideal measure of LD would not depend on allele frequencies; however, no measures of LD are completely independent of allele frequencies. MCRAE *et al.* (2002) suggested simultaneously using a coefficient of LD and the statistical significance of the marker association to disentangle the relationship between the LD measure and allele frequency. In this study, we measured LD with the coefficient D' and the statistical significance of the marker association was computed with a Monte Carlo extension of the Fisher exact test (SLATKIN 1994). We chose to use D' because it is applicable to polymorphic markers, it is less dependent than other measures on allelic frequencies (see, *e.g.*, ZAPATA and VISEDO 1995), and it facilitates the comparison of our results with those observed in sheep and cattle (FARNIR *et al.* 2000; MCRAE *et al.* 2002; TENESA *et al.* 2003).

LD was analyzed at three levels: between populations, between chromosomes (within populations), and along individual chromosomes. At the population level, the global pattern of LD was similar in all five populations (Figure 1). Given the heterogeneity of demographic histories (see Table 1), a different level of LD might be expected in these populations. However, the test of ANCOVA indicated a nonsignificant population effect ($P = 0.872$). This nonsignificance of the population effect can be explained by the small sample sizes of our experiments (number of haplotypes). However, tests of MANOVA on the joint parameters {*R, rs*} obtained by adjusting the exponential function to D' indicate significance of both population and chromosome effects. This illustrates that fitting D' with a theoretical nonlinear model could be more powerful than an empirical linear model in the detection of significant effects.

Pairwise comparisons of populations revealed signifi-

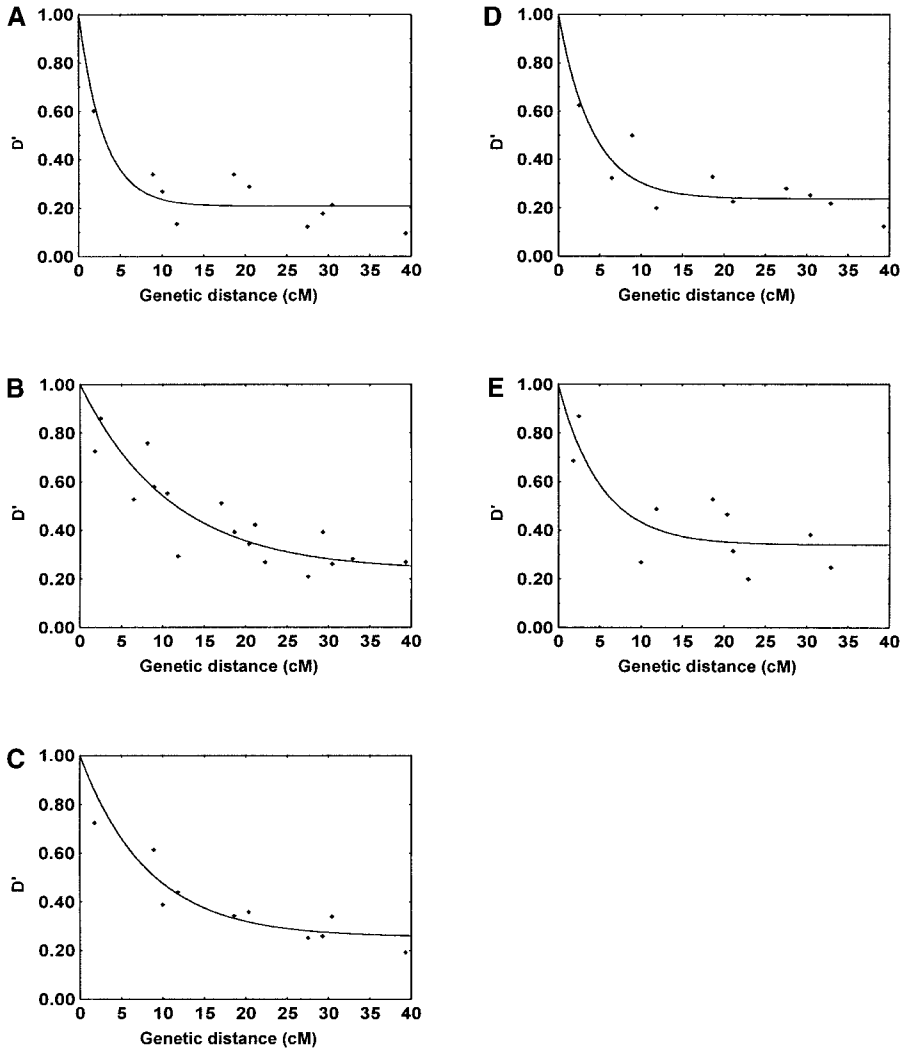

FIGURE 4.—The exponential function applied to D' between syntenic markers on SSC7 in the five populations (A–E).

cant difference between the two chromosomes in three out of five populations (A, C, and E, see Table 3), resulting in an overall significant chromosome effect ($P = 0.027$). The difference in levels of LD on SSC4 and SSC7 cannot be attributed to a sample size bias as the number of haplotypes was similar for both chromosomes in each population. It also cannot be attributed to the differences in the lengths of analyzed chromosome segments as the ANCOVA included the effect of genetic map distance. A putative explanation of the observed differences in the mean D' between the two analyzed chromosomes is the presence/absence of QTL underlying growth rate and fat deposition for which all five populations are selected. In four out of five populations, the highest D' was observed on the chromosome for which effects of QTL underlying one or two selected traits are the most significant (SSC7 in populations A, C, and D; SSC4 in population B; see Tables 1 and 3 and NAGAMINE *et al.* 2003). It is possible that a hitchhiking effect is prevailing in these populations, *i.e.*, that the observed differences in LD between chromosomes are caused by selection. This was confirmed when we considered two groups of populations (see Table 5): the chromosome effect is less important ($P = 0.23$) when both chromosomes show significant effects of selected QTL than in the case of significant QTL effects on one chromosome ($P = 0.06$). The selection effect on LD can be tested if the effects and frequencies of QTL alleles are known. However, this information was not available in the study of NAGAMINE *et al.* (2003).

Genetic map distance between markers was more significant than the other tested factors ($P < 0.0001$; see Table 4) in explaining variation in D'. This relationship between D' and genetic distance fits an exponential function (Figures 3 and 4). Parameters of this function have a simple biological interpretation: rs is the component of D' independent of distance and R is the distance at which D' drops to rs.

Unlike other LD studies in livestock, we analyzed five separate populations and presented results for individual chromosomes. This allowed us to test the effect of different factors (population, chromosome, and genetic distance) and their interactions. In addition, we fitted LD with a theoretical model, which provided us with

TABLE 6

Parameters of the exponential function of the genetic map distance applied to D' per chromosome and per population

	Chromosome 4		Chromosome 7	
Population	R (Morgans)	rs	R (Morgans)	rs
A	0.096	0.150	0.089	0.208
B	0.218	0.215	0.326	0.234
C	0.143	0.195	0.245	0.256
D	0.123	0.176	0.123	0.236
E	0.152	0.211	0.152	0.340

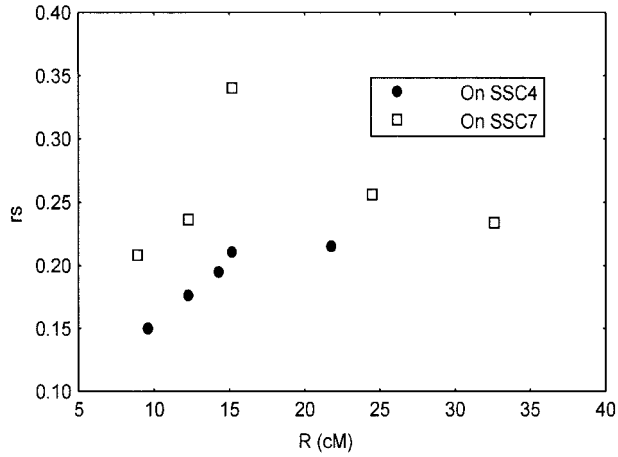

FIGURE 5.—Relationship between R and rs in five populations of pigs, for two chromosomes (SSC4 and SSC7).

interesting parameters in a comparative framework (*i.e.*, R and rs). As these parameters are not known for cattle and sheep populations, we can make comparisons only between the average levels of LD. Between linked loci, the level of LD in pigs, cattle, and sheep is comparable (global patterns of D' and significance levels). Between unlinked loci, LD was not significant in UK dairy cattle (TENESA *et al.* 2003), in New Zealand sheep (MCRAE *et al.* 2002), and in our five populations of pigs (distribution of P values close to its expectation under a random allelic association; see Figure 2), while it was highly significant in the Dutch cattle (FARNIR *et al.* 2000). A possible reason for this difference is the population effective size. Values of N_e that we have for the five populations of pigs vary between 60 and 300 (Table 1), while N_e appears to be <50 in the Holstein-Friesian population (BOICHARD *et al.* 1996).

Another possible explanation of differences in significance of LD in the three studies is the number of haplotypes used. We used 184–302 haplotypes in this study and MCRAE *et al.* (2002) used ~270 haplotypes, so both studies have less statistical power than FARNIR *et al.* (2000), which had 581–1254 haplotypes. TENESA *et al.*'s (2003) study had low statistical power because of a small sample size (50 individuals, *i.e.*, ≤100 haplotypes) and because they applied a Bonferroni correction. Among these three studies, the highest value of mean D' between unlinked loci was observed in Tenesa *et al.*'s study (0.39), while it was of the same magnitude in those of McRae *et al.* (0.20) and Farnir *et al.* (0.12–0.20) and in this study (0.11–0.22, see Table 3). According to the model of MCRAE *et al.* (2002), the bias on D' resulting from samples size is ~0.04–0.06 in our study, 0.00–0.02 in the Dutch cattle (FARNIR *et al.* 2000), 0.025–0.05 in sheep (MCRAE *et al.* 2002), and 0.13 in the UK cattle (TENESA *et al.* 2003).

Therefore, the most likely explanation of the differences in observed LD and its statistical significance between these studies is sample size. Our study was based upon 15 markers that covered ~70 cM in two chromosome regions, for five different populations. To our knowledge, there is no other comparable study in livestock populations. For each combination of population × chromosome, 184–302 haplotypes were available, which is higher than that of most LD studies in human populations (*e.g.*, <100 haplotypes were used in HUTTLEY *et al.* 1999 and REICH *et al.* 2001). While we can expect that more empirical data on LD will be available in the future with probably more specifically planned experiments, we think that our samples were large enough to support the conclusions from this study.

The observed level of LD in pigs indicates that QTL fine mapping may be effective with the presently available marker density. LD-based gene mapping methods are expected to be more powerful than classical methods of linkage analysis with smaller samples. At the distance of $R/3$, D' is equal to $rs + (1 - rs)/e$, according to the function used. Since the average value of rs is 0.22, this corresponds to $D' = 0.5$. If we consider values of $D' > 0.5$ as "useful" LD for mapping purposes, then the corresponding chromosome segments are ~3–10 cM in our populations. This suggests that powerful genome-wide association studies are feasible in commercial pig populations at marker densities of 5–10 cM, so that no QTL is >3–5 cM from the nearest marker with $D' > 0.5$, and many QTL will be in LD with markers with D' closer to 1.0. Thus, for a twofold increase in genotyping effort per animal relative to a linkage study, more power of detection is achieved for the same sample size or fewer animals are necessary to achieve the same power as a linkage study. Note also that these results imply that candidate gene studies in pigs that purport to find associations with phenotypic trait variation could reflect associations with causative loci some distance from the candidate gene itself.

We are grateful to Eric Le Boulengé, Fréderic Farnir, and Xavier Draye for their helpful suggestions and comments on an early version of the manuscript. We thank the referees for helpful comments and suggestions. We thank the commercial partners Cotswold, JSR Healthbred, PIC International, Rattlerow Ltd., and Newsham Ltd. for their generous support in supplying blood or tissue samples and phenotypic information. We acknowledge the support of the Belgian Fund for

Research in Agriculture and Industry, the Fond Spécial de Recherche program of the Catholic University of Louvain, the Ministère de la Communauté Française de Belgique, and the Roslin Institute. This study was partially funded by the United Kingdom Biotechnology and Biological Sciences Research Council under the Sustainable Livestock Production LINK program.

LITERATURE CITED

Ardlie, K. C., L. Kruglyak and M. Seielstad, 2002 Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet. **3:** 299–309.

Boichard, D., L. Maignel and E. Verrier, 1996 Analyse généalogique des races bovines laitières françaises. INRA Prod. Anim. **9:** 323–335.

Bulmer, M. G., 1971 The effect of selection on genetic variability. Am. Nat. **105:** 201–211.

Christakos, G., 1992 *Random Field Models in Earth Sciences.* Academic Press, San Diego.

Coppieters, W., A. Kvasz, F. Farnir, J.-J. Arranz, B. Grisart *et al.*, 1998 A QTL with major effect on milk yield and compositions maps to bovine chromosome 14. Mamm. Genome **9:** 540–544.

Farnir, F., W. Coppieters, J.-J. Arranz, P. Berzi, N. Cambisano *et al.*, 2000 Extensive genome-wide linkage disequilibrium in cattle. Genome Res. **10:** 220–227.

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto *et al.*, 1995 Mapping quantitative trait loci controlling milk production by exploiting progeny testing. Genetics **139:** 907–920.

Green, P., K. Falls and S. Crooks, 1990 *Documentation for CRI-MAP Version 2.4.* Washington University School of Medicine, St. Louis.

Guiyun, Y., D. D. Chandee and D. W. Severson, 1998 Evidence for genetic hitchhiking effect associated with insecticide resistance in *Aedes aegypti.* Genetics **148:** 793–800.

Guo, S., and E. Thompson, 1992 Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics **48:** 361–372.

Haley, C. S., 1999 Advances in QTL mapping, pp. 47–59 in *Proceedings of From Lush to Genomics: Visions for Animal Breeding and Genetics*, edited by J. C. M. Dekkers, S. J. Lamont and M. F. Rothschild. Iowa State University, Ames, IA.

Hayes, B. J., P. M. Visscher, H. C. McPartlan and M. E. Goddard, 2003 Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. **13:** 635–643.

Hedrick, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331–341.

Huttley, G. A., M. W. Smith, M. Carrington and S. J. O'Brien, 1999 A scan for linkage disequilibrium across the human genome. Genetics **152:** 1711–1722.

Kaessmann, H., S. Zöllner, C. Gustafsson, V. Wiebe, M. Laan *et al.*, 2002 Extensive linkage disequilibrium in small human populations in Eurasia. Am. J. Hum. Genet. **70:** 673–685.

Knott, S. A., L. Marklund, C. S. Haley, K. Andersson, W. Davis *et al.*, 1998 Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. Genetics **149:** 1069–1080.

Lewontin, R. C., 1964 The interaction between selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.

McRae, A. F., J. C. McEwan, K. G. Dodds, T. Wilson, A. M. Crawford *et al.*, 2002 Linkage disequilibrium in domestic sheep. Genetics **160:** 1113–1122.

Morton, N. E., W. Zhang, P. Taillon-Miller, S. Ennis, P.-Y. Kwok *et al.*, 2001 The optimal measure of allelic association. Proc. Natl. Acad. Sci. USA **98:** 5217–5221.

Nagamine, Y., C. S. Haley, A. Sewalem and P. M. Visscher, 2003 QTL variation for growth rate and obesity between and within lines of pigs (*Sus scrofa*). Genetics **164:** 629–635.

Nsengimana, J., and P. V. Baret, 2002 Linkage disequilibrium in inbred populations: a geostatistical approach. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, Vol. 32, pp. 681–684.

Pritchard, J. K., and M. Przeworski, 2001 Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. **69:** 1–14.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti *et al.*, 2001 Linkage disequilibrium in the human genome. Nature **411:** 199–204.

Schneider, S., D. Roessli and L. Excoffier, 2000 *Arlequin: A Software for Genetic Data Analysis. Version 2.000.* Genetics and Biometry Laboratory, University of Geneva, Geneva.

Slatkin, M., 1994 Linkage disequilibrium in growing and stable populations. Genetics **137:** 331–336.

Tenesa, A., S. A. Knott, D. Ward, D. Smith, J. L. Williams *et al.*, 2003 Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. J. Anim. Sci. **81:** 617–623.

Walling, G. A., P. M. Visscher, L. Andersson, M. F. Rothschild, L. Wang *et al.*, 2000 Combined analyses of data from quantitative trait loci mapping studies: chromosome 4 effects on porcine growth and fatness. Genetics **155:** 1369–1378.

Weir, B. S., 1996 *Genetic Data Analysis II.* Sinauer Associates, Sunderland, MA.

Weir, B. S., and W. G. Hill, 1980 Effect of mating structure on variation in linkage disequilibrium. Genetics **95:** 477–488.

Wiener, P., I. MacLean, J. L. Williams and J. A. Woolliams, 2000 Testing for the presence of previously identified QTL for milk production traits in new populations. Anim. Genet. **31:** 385–395.

Zapata, C., and G. Visedo, 1995 Gametic disequilibrium and physical distance. Am. J. Hum. Genet. **57:** 190–191.

Communicating editor: J. B. Walsh