

ORIGINAL ARTICLES

Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus

P Wiener¹, D Burton¹, P Ajmone-Marsan², S Dunner³, G Mommens⁴, IJ Nijman⁵, C Rodellar⁶, A Valentini⁷ and JL Williams¹

¹Roslin Institute, Edinburgh, Roslin, Midlothian EH25 9PS, UK; ²Institute of Zootechnics, Catholic University of S Cuore, via Emilia Parmense, 84, I-29100 Piacenza, Italy; ³Dpto. Produccion Animal, Facultad Veterinaria, Avda Puerta de Hierro, Madrid 28040, Spain; ⁴Dr. van Haeringen Polygen (VHP), Sint-Lenaartsebaan 29, B-2390 Malle, Belgium; ⁵Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL Utrecht, The Netherlands; ⁶Lab. Genetica Bioquimica Y Grupos Sanguineos, Fac. Veterinaria Miguel Servet 177, 50-03 Zaragoza, Spain; ⁷Istituto di Zootecnica, Universita' della Tuscia via C. de Lellis, 01100 Viterbo, Italy

This paper explores patterns of genetic diversity near a locus known to have been under selection. The myostatin gene (*GDF-8*) has been shown to be associated with double muscling, a phenotype selected for in a number of cattle breeds. We examined population genetic parameters for microsatellite loci at varying distances from *GDF-8* in double-muscled (DM) and non-double-muscled (non-DM) cattle breeds in order to assess patterns of diversity. A theoretical analysis was also performed to predict the patterns of

diversity expected under different scenarios. We found differences in the patterns of heterozygosity, allele diversity and linkage disequilibrium between DM and non-DM breeds. However, there were some exceptions to the predicted patterns. These are discussed in light of the histories of the breeds and the potential for using microsatellite diversity for mapping trait genes in livestock populations. *Heredity* (2003) **90**, 350–358. doi:10.1038/sj.hdy.6800257

Keywords: linkage disequilibrium; hitchhiking; myostatin; double muscling; Belgian blue cattle

Introduction

The exploitation of linkage disequilibrium (LD) between genes of large effect and neutral markers has been proposed for the identification of both production-related genes in livestock (Riquet *et al.*, 1999) and disease-associated genes in humans (Huttley *et al.*, 1999). Linkage disequilibrium is particularly likely to arise between closely linked loci, but it tends to break down fairly quickly with distance, therefore a very dense genome map would be required for whole-genome linkage disequilibrium mapping (Kruglyak, 1999; but see Ott, 2000 for an alternative view). However, LD may be promoted and maintained for longer periods under particular conditions, as has been seen in genomic regions of reduced recombination (eg, Begun and Aquadro, 1991) and in structured populations (eg, Peterson *et al.*, 1999, and references therein).

In addition, population genetics theory predicts that selection will promote LD via the hitchhiking effect (Maynard Smith and Haigh, 1974), which will alter the pattern of genetic variation near a selected locus (eg, Ohta and Kimura, 1975; Kaplan *et al.*, 1989; Stephan *et al.*, 1992; Slatkin, 1995; Barton, 1998; Slatkin and Wiehe, 1998; Barton, 2000; Kim and Stephan, 2000, 2002). The impact of hitchhiking will be particularly pronounced on highly

variable loci such as microsatellites (Wiehe, 1998) such that allelic diversity and heterozygosity distributions may be used to infer selection (Schlötterer *et al.*, 1997; Payseur *et al.*, 2002). While the theory predicts that, on a qualitative level, LD will increase and variability will decrease under directional selection, for the purposes of identifying genes of large effect, it becomes a quantitative issue as to whether the associations between these genes and neutral markers will be strong enough and exist over large enough genomic distances to allow the localization of major genes by this approach.

The goal of our study was to examine patterns of association between a gene under strong selection and linked, neutral markers. We focused on *GDF-8* (the 'myostatin' locus), a gene located at the centromeric end of bovine chromosome 2 (Charlier *et al.*, 1995) whose functional protein is a negative growth regulator of skeletal muscle. Mutations in the gene have been shown to result in the double-muscling phenotype in some breeds of cattle (Grobet *et al.*, 1997; McPherron and Lee, 1997). This phenotype is an extreme form of muscle development, characterized by a large increase in muscle mass. Double muscling was actively selected for in several cattle breeds in the last century (Arthur, 1995).

In our analysis, we use a number of population genetic measures to compare microsatellite loci on bovine chromosome 2 for nine breeds of cattle, three of which have a high frequency of double-muscled DM indivi-

duals while the other six are not known to have DM individuals. Using simulations, we also explore factors that may affect the likelihood of detecting hitchhiking, and thereby the potential for localizing a gene using selection mapping.

Materials and methods

Breed samples

The DM breeds included were Asturiana de los Valles (Spain), Belgian Blue (Belgium) and Piedmontese (Italy). Asturiana and Belgian Blue are known to carry the same variant at the *GDF-8* locus, which is different from that found in Piedmontese (Dunner *et al*, 1997; McPherron and Lee, 1997; Grobet *et al*, 1998). The majority of the Belgian Blue and Asturiana samples were homozygous for an 11-bp deletion (*mh*) previously reported in these breeds (Belgian Blue: 96% *mh/mh*, 4% *+/mh*; Asturiana: 88% *mh/mh*, 10% *+/mh*, 2% *+/+*) (Dunner *et al*, 1997; Grobet *et al*, 1997). We did not know the frequency of the double-muscling mutation in the Piedmontese sample used, but the estimated frequency of double muscling for that population is close to 100%. The non-double-muscling (non-DM) breeds studied were Aberdeen Angus, Ayrshire, British Friesian, Charolais and Hereford (all from the UK) and Toro de Lidia (Spain). The Aberdeen Angus, Charolais and Hereford are beef breeds and Ayrshire and Friesian are dairy breeds. Toro de Lidia is used for bull-fighting. Approximately 50

animals were sampled from the above breeds, excluding sibs and parent-offspring pairs where possible.

Microsatellite genotyping and map information

DNA was extracted from blood or semen and genotyped according to the protocols described in Wiener *et al* (2000). All of the 18 markers used are located on bovine chromosome 2 (Figure 1). Approximately half of the markers were chosen to cover the region near *GDF-8* as densely as possible, using all known microsatellite markers that worked reliably in our hands, and the other half distributed along the rest of chromosome 2 where the choice of markers was arbitrary, again using known microsatellite markers that were reliable in our hands. Where possible, marker positions were estimated from the USDA linkage map (<http://sol.marc.usda.gov>). The positions of several of the markers not positioned on the USDA map were estimated from the original source (Grobet *et al*, 1997). Many of the markers used are tightly linked and some have overlapping map positions, therefore, marker positions should be taken as estimates. All breeds were typed for all markers except that Asturiana de los Valles was not typed for BM3627 because of technical problems.

Data analysis

Heterozygosities, numbers of alleles and allele distributions were determined for each breed for each marker. The relation between these measures and physical

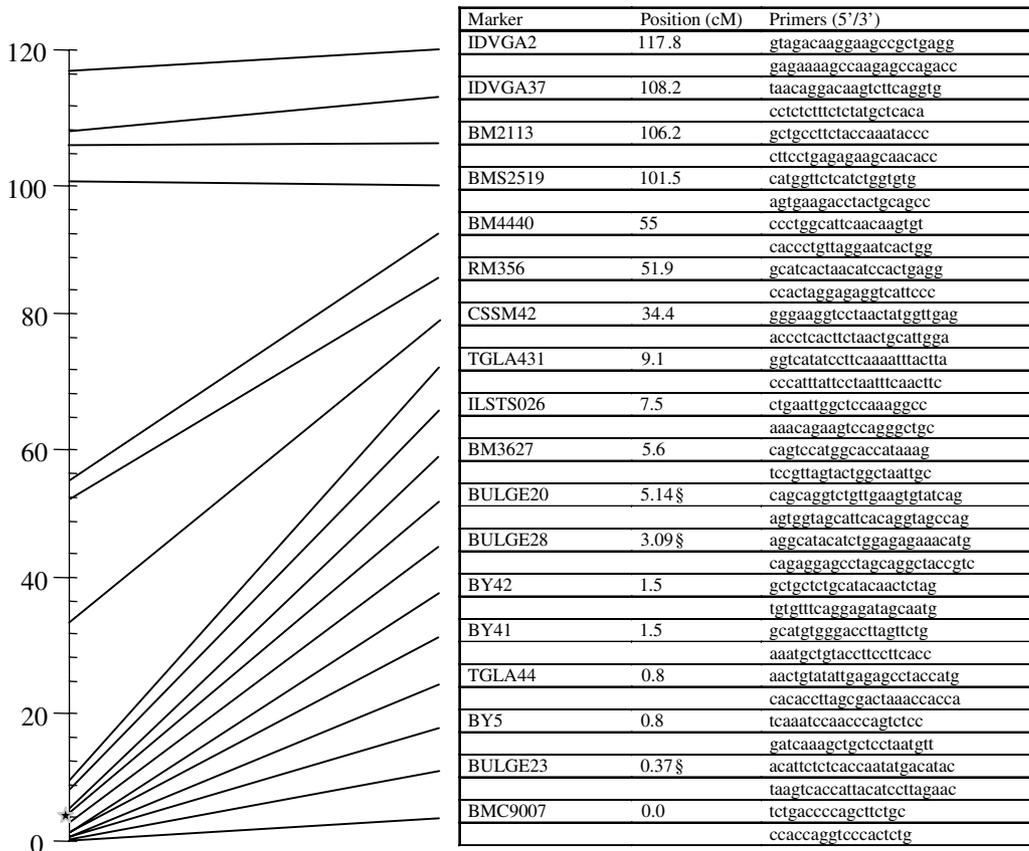


Figure 1 Bovine chromosome 2 markers used in study (* represents the position of *GDF-8*). Map positions based on USDA linkage map (<http://sol.marc.usda.gov>) except where indicated by § (map positions estimated from Grobet *et al*, 1997).

distance from *GDF-8* was assessed by regression. 'Adjusted heterozygosities' for the DM breeds were calculated as heterozygosity divided by the average heterozygosity of the non-double-muscling breeds for each marker position. This adjustment was applied to account for the variation in numbers of alleles between the different microsatellites. To compare the between-breed and within-breed variances, analysis of variance (ANOVA) was performed for each marker, with breed considered as a treatment. The relation between the resulting *F*-ratio (a measure of the contribution of breed to the overall variance) and physical distance from *GDF-8* was assessed by regression.

An exact, two-tailed Hardy–Weinberg test (Weir, 1996) was used to test for deviations from HWE using Genepop v.3.1d (<http://wbioimet.curtin.edu.au/genepop>). For markers with fewer than five alleles, the complete enumeration method of Louis and Dempster (1987) was used to calculate the exact *P*-value. For loci with five or more alleles, a Markov chain method was used to estimate the exact *P*-value (Guo and Thompson, 1992). The latter was implemented using the Genepop default settings. *P*-values were adjusted by the Bonferroni correction to account for multiple (18 or 17) tests. Significance was assessed at the (adjusted) $P < 0.05$ level.

Genotypic LD (defined by Weir, 1996) was calculated in Genepop. For each two-locus pair, a contingency table of genotype combinations was constructed, where rows represent genotypes at the first locus and columns represent genotypes at the second locus. An exact test was used to calculate the probability (*P*) of rejecting the null hypothesis of random association between loci for the observed contingency table although there was random association (Type I error). This probability would be computed exactly by summing the probabilities of all contingency tables with the same row and column sums and with the same or smaller probabilities. As this is computationally infeasible for multiallelic loci, the Markov chain method of Raymond and Rousset (1995) was used to estimate the exact *P*-value. This was implemented using Genepop default settings. *P*-values were adjusted by the Bonferroni correction to account for multiple ($153 = 18 \times 17/2$ or $136 = 17 \times 16/2$) tests. Two-locus LD was considered significant for adjusted *P*-values less than 0.05.

Simulations

The increase in frequency of a selected allele and its effect on the dynamics of neutral linked loci was examined in a simulation of a cattle population. The population of animals (N_{fem} , N_{mal}) was generated with n_{alleles} alleles at n_{loci} neutral loci linked to the double-muscling gene. All alleles were initiated at equal frequency in the population. The favoured allele at the selected locus was introduced into the population at a frequency of $1/2N$. The phenotype (eg, muscle score) was assumed to be normally distributed with the mean determined by a sex-specific population mean (μ_{fem} , μ_{mal}) and the (additive) effect of the selected locus (*a*). The variance term (σ^2) encompassed both environmental and background genetic factors. The trait value thus took on a value of $\mu_{\text{sex}} + a_{\text{genotype}} + \sigma$. Truncation selection (Figure 2), the standard approximation for selection on managed populations (see eg Falconer and Mackay, 1996), was imposed on the population using the

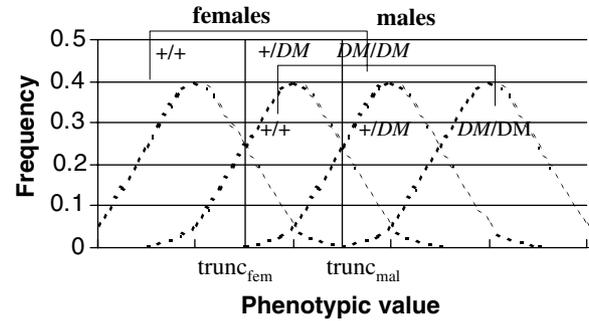


Figure 2 Distribution of phenotypes according to sex and genotype and truncation selection cutoffs used in simulation study.

parameters $\text{trunc}_{\text{fem}}$ and $\text{trunc}_{\text{mal}}$, points of truncation for females and males (to allow the effect of the gene to differ between sexes). The population was followed for *gen* generations of gamete production, random mating and truncation selection; the probability of recombination was calculated as a function of physical distance using Haldane's mapping function (Lynch and Walsh, 1998). The population size was held constant throughout the simulation. For all parameter sets, at least 100 simulations were run. The selected allele was frequently lost because of genetic drift (the actual numbers of simulations where the allele was maintained are given in Table 3).

Parameter values were chosen to mimic as far as possible the history of the Belgian Blue breed as this has been well documented. Computer memory limitations and computing times required that population sizes used for most analyses were approximately one-tenth of true breed population sizes ($N_{\text{fem}} = 45\,000$, $N_{\text{mal}} = 5\,000$), however, where possible, the dynamics of more realistic population sizes were examined ($N_{\text{fem}} = 450\,000$, $N_{\text{mal}} = 50\,000$). The number of alleles at the neutral loci (n_{alleles}) took the values 2 and 10, reflecting the range of alleles found at typical bovine microsatellite loci. Most analyses focused on the dynamics of a single locus 1 cM from the double-muscling gene as it was expected that LD would span a limited genetic distance. To determine the pattern of hitchhiking over a larger genetic distance, additional analyses looked at loci ranging between 1 and 50 cM from the selected locus.

Phenotypic values were based on a visual assessment of muscle conformation on beef cattle employed in the UK (1–15 scale) and our previous calculations of additive effect in a double-muscling cattle breed (Wiener *et al*, 2002). The population means were assumed to be $\mu_{\text{fem}} = 7$ and $\mu_{\text{mal}} = 9$ with standard deviation (σ) equal to 1. The additive effect of the double-muscling mutation was assumed to be equivalent to the average difference between males and females ($a_{+/+} = -2$, $a_{+/DM} = 0$ and $a_{DM/DM} = +2$), where DM refers to the selected allele. The selection model varied the points of truncation; $\text{trunc}_{\text{fem}}/\text{trunc}_{\text{mal}} = 6/8$ was the 'basic' model. Variants included 4/6 (weakest selection), 5/7 (weaker selection), 7/9 (stronger selection) and 8/10 (strongest selection). Selection was imposed on the offspring produced after each mating. That is, an animal was chosen for the next generation if its phenotype exceeded the truncation point, until the requisite number of males and females were produced.

The basic model assumed no mutation, no use of artificial insemination and nonoverlapping generations, however, as these assumptions are known not to reflect actual cattle populations, they were relaxed to test the effects of including these three factors but given computational constraints, only for the smaller population size. Two mutation rates were used, based on estimates for microsatellite loci from the literature (see Valdes *et al*, 1993): 0.003 and 0.02 per gamete per generation. Since these estimates are high compared to other published estimates of mutation rates (see Goldstein *et al*, 1996), the simulations should overestimate the breakdown of LD. In the two-allele model, one allele was changed to the other with these mutation probabilities. In the 10-allele model, stepwise mutation (Valdes *et al*, 1993; Goldstein *et al*, 1995) was assumed, with equal probability of gain or loss of one microsatellite repeat. For simplification, the order of the alleles formed a circle (so that the allele following allele 10 was allele 1). The effect of this assumption was tested by use of an alternative mutation model of reflecting boundaries (ie allele 1 could mutate only to allele 2 and allele 10 only to allele 9) with no differences seen between the two models (see Results).

When artificial insemination (AI) was included, an AI pool of 400 males was created, separate from the general male population, and AI from this pool accounted for 50% of the matings. Individuals were chosen to enter the AI pool based on an AI-specific truncation point ($trunc_{semen} = 9$). When the assumption of nonoverlapping generations was relaxed, an overlap of five generations for the males was allowed (ie males remained in the mating pool for five generations instead of just one). The effects of locus-specific mutation were also considered by modifying the mutation scheme such that the mutation rate for each locus was adjusted from the basic rate by 10^x ($-1 < x < 1$).

The relation between heterozygosity and distance from the selected locus was assessed by nonlinear regression (curve-fitting). Regression was performed on the results of each simulation run and on the average heterozygosity over all runs.

Results

Comparisons of DM and Non-DM breeds

Heterozygosity, adjusted heterozygosity and number of alleles were examined as functions of distance from *GDF-8*. We used the pattern of results seen in the simulations (see below) to derive our predicted pattern of results and thus fitted exponential curves to the data (asymptotic regression; $y = A + BR^x$, where x is the distance from *GDF-8* and y is the heterozygosity, adjusted heterozygosity or number of alleles), with the measured variable increasing with distance from *GDF-8* ('right sense'). The R parameter defines the rate of exponential increase (a higher R reflects a faster response of the variable to small changes in distance near *GDF-8*). The fit of the data to the relevant curve was measured by the significance of the fit and the percentage variation accounted for by the fitted curve (V).

Adjusted heterozygosity (Figure 3) significantly increased with distance from *GDF-8* for two of the double-muscling breeds, Asturiana and Piedmontese ($R = 0.802$,

$V = 57\%$, $P = 0.001$ and $R = 0.827$, $V = 25\%$, $P = 0.045$, respectively). The fit to an exponential curve was not significant for Belgian Blue.

The fit of the (nonadjusted) heterozygosity data (Table 1) to exponential curves were significant for the DM breeds ($R = 0.734$, $V = 53\%$, $P = 0.002$ for Asturiana; $R = 0.753$, $V = 40\%$, $P = 0.008$ for Belgian Blue; $R = 0.683$, $V = 51\%$, $P = 0.002$ for Piedmontese). However, in addition to the DM breeds, the heterozygosity values of Charolais, a beef breed, also fit an exponential curve ($R = 0.753$, $V = 34\%$, $P = 0.018$). While none of the other breeds had significant fits to exponential curves (Ayrshire had a 'left sense' rather than the predicted 'right sense' relation), heterozygosity tended to increase with distance from *GDF-8* in all breeds. Furthermore, the other two beef breeds (Aberdeen Angus and Hereford) had nearly significant asymptotic regression curves that explained a substantial proportion of the variation ($R = 0.753$, $V = 17\%$, $P = 0.093$ for Aberdeen Angus; $R = 0.753$, $V = 19\%$, $P = 0.083$ for Hereford).

Allele number (Table 1) also increased with distance from *GDF-8* in all breeds, but was not significant in the non-DM breeds. Asymptotic regression provided a good fit to the data for the double-muscling breeds with both Asturiana and Piedmontese significant ($R = 0.739$, $V = 50\%$, $P = 0.003$ and $R = 0.720$, $V = 31.7\%$, $P = 0.022$, respectively) and Belgian Blue nearly significant ($R = 0.626$, $V = 22\%$, $P = 0.058$).

Some loci near *GDF-8* had only a few alleles, for example, BY41 at 1.5 cM and BULGE28 at ~3.09 cM. However, the low allele number was not limited to double-muscling breeds; for BULGE28, Aberdeen Angus, Charolais and Hereford each had only a single allele, whereas Asturiana, Belgian Blue, Piedmontese and Friesian had two alleles each (Ayrshire had five alleles and Toro de Lidia had three). For BY41, Aberdeen Angus, Asturiana, Ayrshire, Belgian Blue, Piedmontese

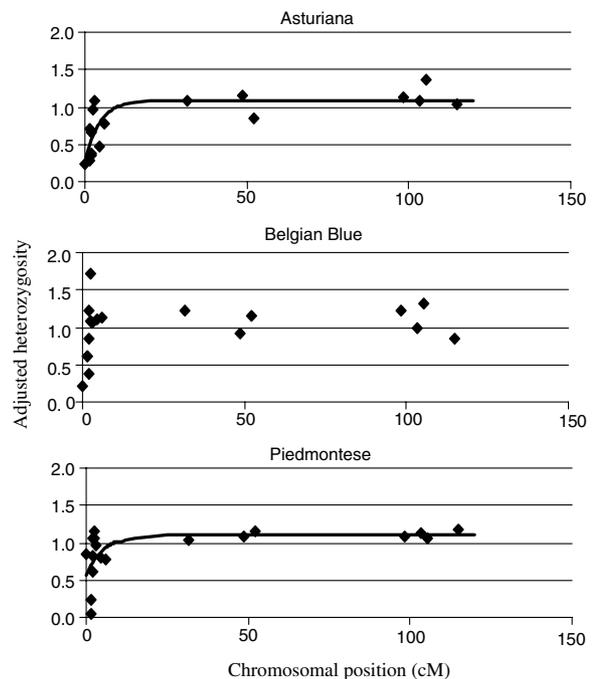


Figure 3 Adjusted heterozygosity as a function of distance from *GDF-8* for DM breeds.

Table 1 Heterozygosities and numbers of alleles for each breed and marker

Marker	Position	Aberdeen Angus		Asturiana de los Valles		Ayrshire		Belgian Blue		British Friesian		Charolais		Hereford		Piedmontese		Toro de Lidia	
		Het	Alleles	Het	Alleles	Het	Alleles	Het	Alleles	Het	Alleles	Het	Alleles	Het	Alleles	Het	Alleles	Het	Alleles
BMC9007	0.00	0.59	5	0.76	5	0.73	6	0.74	7	0.78	5	0.62	5	0.68	5	0.67	5	0.76	6
BULGE23	0.37	0.63	5	0.60	4	0.50	5	0.67	5	0.77	5	0.71	5	0.42	3	0.71	4	0.70	4
BY5	0.80	0.73	6	0.22	6	0.60	8	0.80	9	0.78	8	0.68	9	0.55	6	0.68	6	0.54	6
TGLA44	0.80	0.70	8	0.53	9	0.89	8	0.68	7	0.83	7	0.84	8	0.80	7	0.49	8	0.71	8
BY41	1.50	0.35	2	0.13	2	0.44	2	0.27	2	0.58	3	0.35	4	0.56	4	0.02	2	0.37	2
BY42	1.50	0.57	5	0.47	4	0.63	5	0.40	3	0.81	5	0.69	7	0.64	4	0.16	5	0.57	6
BULGE28	3.09	0.00	1	0.02	2	0.25	5	0.02	2	0.14	2	0.00	1	0.00	1	0.08	2	0.19	3
BULGE20	5.14	0.39	4	0.21	4	0.56	6	0.21	4	0.50	4	0.75	7	0.49	3	0.46	3	0.61	4
BM3627	5.60	0.13	3			0.07	4	0.33	2	0.07	2	0.22	4	0.40	5	0.20	2	0.24	3
ILSTS026	7.50	0.52	5	0.26	5	0.60	5	0.62	4	0.46	5	0.66	5	0.72	3	0.46	4	0.45	5
TGLA431	9.10	0.65	5	0.50	10	0.87	8	0.72	11	0.53	8	0.80	12	0.44	5	0.49	11	0.54	6
CSSM42	34.40	0.73	6	0.80	12	0.55	9	0.90	10	0.87	11	0.88	6	0.74	5	0.76	8	0.62	8
RM356	51.90	0.38	2	0.53	5	0.28	2	0.43	3	0.54	3	0.66	4	0.41	3	0.50	5	0.49	2
BM4440	55.00	0.81	8	0.57	10	0.54	9	0.78	11	0.72	10	0.69	10	0.68	6	0.78	11	0.60	7
BMS2519	101.50	0.58	5	0.86	10	0.89	7	0.91	7	0.87	7	0.88	7	0.57	7	0.81	7	0.74	9
BM2113	106.20	0.74	8	0.83	8	0.74	5	0.76	7	0.75	6	0.78	7	0.92	7	0.85	8	0.65	7
IDVGA37	108.20	0.39	3	0.74	8	0.50	5	0.72	5	0.38	3	0.65	5	0.62	3	0.59	5	0.76	9
IDVGA2	117.80	0.70	6	0.71	8	0.70	6	0.58	8	0.53	6	0.80	7	0.71	5	0.80	7	0.70	8

and Toro de Lidia had only two alleles. The Piedmontese sample was essentially fixed for a single allele at this locus (at 99% frequency). The other breeds had three or more alleles.

Significant deviations from Hardy–Weinberg equilibrium and LD for each breed are shown in Table 2. There was no evidence for deviation from Hardy–Weinberg equilibrium in the DM breeds although there were loci with an excess of homozygotes in five non-DM breeds (Aberdeen Angus, Ayrshire, Charolais, Friesian and Toro de Lidia, predominantly at BM3627). There was significant LD between chromosome 2 loci in the DM breeds and also in some non-DM breeds. Asturiana had 10 such pairs, Belgian Blue and Piedmontese each had six. The beef breeds had more marker pairs with significant LD than did the dairy breeds; Aberdeen Angus had a particularly large number of significant pairs (9). Toro de Lidia also had nine significant pairs. We also compared levels of LD between the four markers at the telomeric end of chromosome 2 (opposite end of the chromosome from *GDF-8*). Out of 6 possible significant pairs, Belgian Blue had one and Toro de Lidia had two.

The proportion of the variance in allele size because of breed is shown in Figure 4. The ratio of the mean-squared error between breeds to the residual (within-

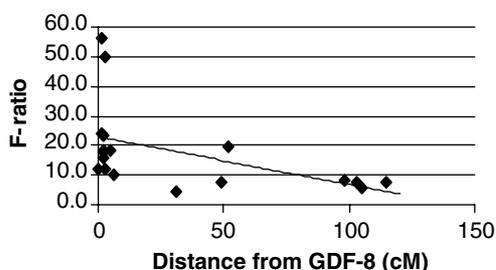


Figure 4 Ratio of variance in allele size because of breed to residual variance for each marker (all breeds included) as a function of distance from *GDF-8*.

breed) mean-squared error (ie the *F*-ratio) is plotted as a function of distance from *GDF-8*. Our expectation was that the ratio of these variances should be greater for markers near the gene because of selection on *GDF-8* in some breeds but not others and because the DM breeds carry different alleles at this locus. There was a significant (negative) linear trend for the *F*-ratio (slope = -0.161 , $V = 19\%$, $P = 0.038$).

Simulated population under selection

Figure 5a shows the average final heterozygosity per locus (over 24 runs) as a function of chromosomal

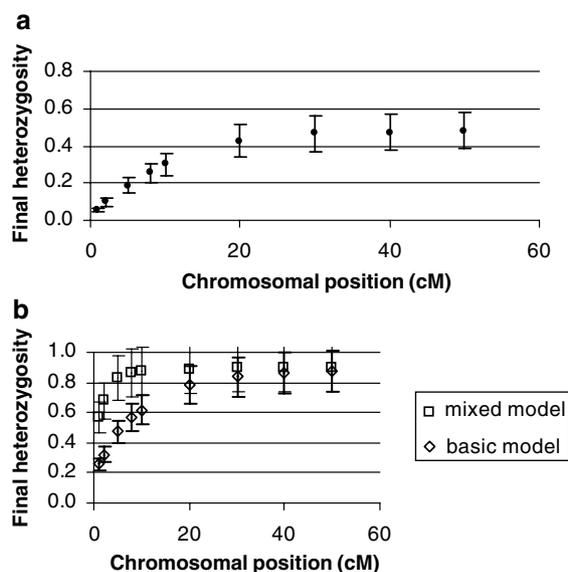


Figure 5 (a) Heterozygosity at marker loci after 33 generations of evolution under the basic simulation model with two alleles (initial heterozygosity = 0.5). (b) Heterozygosity at marker loci with 10 alleles after 33 generations of evolution (basic model) and after 50 generations of weak selection followed by 33 additional generations of stronger selection (mixed model). Standard error bars are shown.

Table 2 Results from tests of deviations from Hardy–Weinberg equilibrium and LD. HW deviations are deficiencies of heterozygote breeds

	Markers with significant deviations from HWE	Significant LD pairs (distance in cM)
Double-muscled breeds		
Asturiana de los Valles		BMC9007/BY5 (0.8) BMC9007/TGLA44 (0.8) BMC9007/BULGE20 (5.14) BULGE23/TGLA44 (0.43) BULGE23/BY42 (1.13) BULGE23/BULGE20 (4.77) BY5/TGLA44 (0) TGLA44/BY41 (0.7) TGLA44/BY42 (0.7) BY42/BULGE20 (3.64)
Belgian Blue		BMC9007/BULGE23 (0.37) BMC9007/BY41 (1.5) BULGE23/BY5 (0.43) BULGE23/TGLA44 (0.43) BY41/BY42 (0) BMS2519/BM2113 (4.7)
Piedmontese		BMC9007/BULGE23 (0.37) BMC9007/BY5 (0.8) BULGE23/BY5 (0.43) BULGE23/TGLA44 (0.43) BY5/TGLA44 (0) TGLA44/BY42 (0.7)
Non-double muscled beef breeds		
Aberdeen Angus	BM3627	BMC9007/BULGE23 (0.37) BMC9007/BY5 (0.8) BMC9007/TGLA44 (0.8) BULGE23/TGLA44 (0.43) BY5/TGLA44 (0) TGLA44/ILSTS026 (6.7) BY41/TGLA44 (0.7) BY42/TGLA44 (0.7) BY41/BY42 (0.7)
Charolais	BM3627 BM2113	BY5/TGLA44 (0) BY5/BY41 (0.7) BY5/BMS2519 (100.7) TGLA44/BULGE20 (4.34) BY41/BY42 (0) BULGE20/TGLA431 (3.96)
Hereford		BMC9007/TGLA44 (0.8) BULGE23/BY5 (0.43) BY5/TGLA44 (0) BY41/TGLA44 (0.7) RM356/BM4440 (3.1)
Other breeds		
Ayrshire	BULGE28 BM3627	BMC9007/TGLA44 (0.8) ILSTS026/IDVGA2 (109.7)
British Friesian	BM3627	BULGE23/TGLA44 (0.43) BY5/TGLA44 (0)
Toro de Lidia	BM3627 BM4440	BMC9007/BY42 (1.5) BULGE23/BY5 (0.43) BULGE23/TGLA44 (0.43) BULGE23/BY42 (1.13) BY5/TGLA44 (0) BY42/BULGE20 (3.64) BULGE28/BULGE20 (2.05) BMS2519/IDVGA37 (6.7) BMS2519/IDVGA2 (16.3)

distance from the selected locus after 33 generations of evolution under the basic model (initial heterozygosity at all loci was 0.50). Final heterozygosity increased with distance from the selected locus. There was a significant exponential relation ($R = 0.9112$, $V = 99.8\%$, $P < 0.001$) between mean final heterozygosity and chromosomal position (the individual R values varied between 0.77 and 0.95); the curve increased rapidly within 10 cM of the selected locus, reaching a plateau between 20 and 30 cM. To examine the hitchhiking effect under modifications of the basic model, the ratio of final heterozygosity (after 33 generations) to initial heterozygosity at a locus 1 cM from the selected locus was compared for the different scenarios presented in Table 3. There was no dramatic difference in this ratio between the basic model and models assuming either smaller or larger population size, artificial insemination, overlapping generations, multiple (10) alleles, stronger selection or smaller allelic effect. Mutation (both circular and alternative models) and weaker selection did significantly increase the heterozygosity ratio; locus-specific mutation rates further increased the impact of mutation on heterozygosity. With the higher mutation rate, heterozygosity was almost unchanged by selection. Under very weak selection, heterozygosity within 1 cM was reduced by 50%. A combination of these two factors resulted in a reduction of heterozygosity by 40%.

Additional simulations were performed to try to identify realistic conditions under which the hitchhiking effect would be lost. Thus, simulations were performed in which there was very little selection ($trunc_{fem}/trunc_{mal} = 3/3$; to prevent the selected allele being lost from the population) for 50 generations with 10 alleles at the microsatellite loci, followed by the imposition of selection (basic model, as described above) for another 33 generations. A low mutation rate was applied. The average heterozygosity at the end of these 33 further generations (at 1 cM from the selected gene) was 0.568 (final/max heterozygosity = 0.63), which is substantially greater than that under the basic model with the same mutation rate (final = 0.256; final/max = 0.28). Figure 5b shows the final heterozygosity as a function of chromosomal position for the basic mutation model (as in Figure 5a but with 10 alleles rather than two and mutation allowed) and the model described above (mixed model) where 50 generations of low selection preceded the 33 generations of basic selection. Again, asymptotic regression gave a statistically significant fit to the results (basic model: mean $R = 0.9066$, $V = 99.9\%$, $P < 0.001$, individual R values between 0.83 and 0.96; mixed model: mean $R = 0.6769$, $V = 99.8\%$, $P < 0.001$, individual R values between 0.27 and 0.88). When the period of low selection was included in the model, the heterozygosity profile became flat closer to the selected locus than for the basic model, as reflected in the lower value of R .

Discussion

A detailed analysis of bovine chromosome 2 demonstrates detectable effects of selection on *GDF-8*, the locus associated with double-muscling in several cattle breeds. We compared several population genetic measures between three DM and six non-DM breeds and found that they differed in their patterns of heterozygosity and

Table 3 Results from simulation of phenotypic selection at 1cM from selected locus

	<i>n</i>	Avg. final heterozygosity	SE	final heterozygosity max heterozygosity
Basic	33	0.046	0.003	0.09
AI=50%	41	0.040	0.006	0.08
Five-generation overlap	27	0.057	0.010	0.11
10 alleles	24	0.092*	0.008	0.10
10 × popn size	28	0.054	0.002	0.11
1/10 × popn size	20	0.048	0.007	0.10
Weaker seln	24	0.096	0.018	0.19
Stronger seln	25	0.037	0.006	0.07
Very strong seln	13	0.025	0.002	0.05
High mutation rate	29	0.470	0.001	0.94
Moderate mutation rate	25	0.212	0.012	0.42
Moderate mut (10 alleles)	32	0.245*	0.006	0.27
Alternative mutation model (10 alleles)	32	0.252	0.008	0.28
Locus-specific mutation (10 alleles)	35	0.325*	0.037	0.36
Very weak selection	13	0.245	0.015	0.49
Smaller allelic effect	32	0.056	0.006	0.11
Moderate mut + very weak seln (10 alleles)	17	0.532*	0.022	0.59

Maximum (=initial) heterozygosity is 0.50 for all cases with two alleles and 0.90 for all cases with 10 alleles (these are indicated with a *).

numbers of alleles along chromosome 2, although some non-DM breeds showed similarity to the DM breeds in their heterozygosity patterns. When heterozygosity was adjusted to account for intermarker diversity, the double-muscling breeds showed a significant correlation between heterozygosity and distance from *GDF-8*. LD between pairs of loci was greater within DM breeds than dairy breeds. However, there was also a large amount of LD within the beef breeds and a breed used for bull-fighting. LD results were also consistent with Dunner *et al* (1997), in that there was greater LD within Asturiana than Belgian Blue, suggesting a more recent introduction of the allele into Asturiana. Finally, the ratio of between- to within-breed variance in allele size increased with proximity to *GDF-8*, consistent with differences in selection history and allele frequencies between the breeds.

The results from our analyses of chromosome 2 were not as dramatic as seen in the basic model we simulated. Modifications of the basic model indicate several factors that could account for this difference. A very high mutation rate of the microsatellites near *GDF-8* could explain why the reduction in heterozygosities near *GDF-8* was not as large as might have been expected and why LD was not as strong. It seems unlikely that by chance, all (or most) of the microsatellites we chose had higher than average mutation rates, although between-locus variability in mutation rates could contribute to reducing the hitchhiking effect. A combination of weak selection and moderate mutation could possibly explain the pattern of results.

Alternatively, it may be that the history of double-muscling in some breeds is similar to the model we examined with a period of weak selection followed by stronger selection. If so, this could explain why the chromosome 2 results differed from the predictions of the basic model. Under this 'mixed' scenario, double-muscling mutations are actually old and have been maintained at low frequency in these breeds under very weak selection. Then, at some point in recent history, breeders began to select more strongly for the allele and it rose to high frequency or fixation. This would be

consistent with the wide diversity of myostatin haplotypes seen in European cattle breeds (Dunner *et al*, 2003). Under this scenario, the LD between *GDF-8* and neutral markers at the beginning of the selective sweep (the period where strong selection was imposed) would be lower than that for a model where strong selection coincided with the introduction of the mutation. As a result, the probability of detecting a hitchhiking effect would be reduced. This could certainly be the case for the Belgian Blue breed where the mutation is known to have existed prior to World War I (Compère *et al*, 1996), but until the 1950s, the breed was maintained as 'dual purpose,' providing both milk and beef (Hanset, 1982). Our own data from a study on South Devon cattle, another breed with double-muscling, indicate that even if farmers strongly favour the DM phenotype, effective selection when the trait is rare is likely to be weak. This is because of the difficulties in assessing the genotype on heterozygous individuals (Wiener *et al*, 2002).

Our results and previously published work on Asturiana de los Valles (Dunner *et al*, 1997) demonstrate that selection on *GDF-8* has left a stronger mark on this breed than the other two DM breeds. This indicates that the double-muscling mutation (the same mutation as found in Belgian Blue) has been in the Asturiana a shorter time than in the other two breeds. While there is historical documentation that the Belgian Blue and Piedmontese mutations date back prior to World War I (Masoero and Poujardieu, 1982; Compère *et al*, 1996), the earliest definitive cases of double-muscling in Asturiana date from the 1920s.

There could be confounding factors in the comparison of DM and non-DM breeds. If there is another gene near to *GDF-8* under selection in all breeds, then differences between DM and non-DM breeds might be overshadowed. Alternatively, *GDF-8* may itself have been under selection in the non-DM breeds, but perhaps for different alleles, for example, for those which increase the lean muscle mass of an animal (Lin *et al*, 2002). This could help to explain the fact that beef breeds showed greater LD than dairy breeds and that beef breeds had significant and nearly significant asymptotic regression

curves for heterozygosity as a function of distance from *GDF-8*. The presence of a QTL associated with several fat traits has been reported for Canadian beef cattle in the same region as *GDF-8* (Schimpf *et al*, 2000). There may have been selection on such a linked QTL or on *GDF-8* itself in the non-DM beef breeds. In a previous study, we analysed samples from these breeds for non-synonymous sequence variants within the coding regions of *GDF-8* and none were found (Smith *et al*, 2000 and unpublished data). However, it is possible that variants do exist in promoter regions of the gene. Alternatively, an excess of LD could be caused by population structure rather than selection. Thus, it is possible that peculiarities (eg, recent admixture of previously isolated populations) in the history of Aberdeen Angus and Toro de Lidia, which showed particularly high levels of LD, led to a great excess of LD throughout the genome. The elevated number of significant LD pairs at the opposite end of chromosome 2 suggests that this may be the case for Toro de Lidia. Calculating LD for markers on other chromosomes would help to resolve this issue.

One motivation for this project was to consider the utility of a 'selection mapping' (Kohn *et al*, 2000) approach in livestock where genetic diversity patterns could be used to help map genes affecting production traits. The results from our study indicate that such an approach would not have clearly identified the region of the myostatin gene at least at the density of markers used in this study. A similar approach was used more successfully in a study of natural rat populations under strong selection for resistance to anticoagulants. In comparing several population genetic parameters, Kohn *et al* (2000) found large differences between rat populations where resistance had evolved to high levels and those where resistance was low or nonexistent. They used patterns of LD for 26 microsatellite markers in five rat populations to improve positioning of the anticoagulant-resistance gene in rats, which had already been localized to a 6-cM region. These results indicate that there is a stronger hitchhiking effect than found in the DM cattle breeds. This suggests that the mutation for anticoagulant resistance is relatively recent and that selection pressure imposed by anticoagulant use is stronger than breeders' preferences for double-muscling, which is not surprising. Our study shows that before implementing such an approach for mapping trait genes in livestock, it will be important to have information on the age and history of selection of the trait in particular breeds so that breeds can be chosen to maximize the chances of success. Alternative methods that utilize haplotype data (eg, Sabeti *et al*, 2002) may turn out to be more powerful, but have the disadvantage of requiring additional data to infer phase.

Acknowledgements

We are grateful for donations of genetic material from the Aberdeen Angus Cattle Society, the Ayrshire Cattle Society of Great Britain and Ireland and Genus plc. Comments from CS Haley and JA Woolliams on this work are greatly appreciated. Thanks also to Hans Lenstra for his coordination of the wider European project. This work was funded by the European Commission, contract CT98-118.

References

- Arthur PF (1995). Double muscling in cattle: a review. *Aust J Agric Res* **46**: 1493–1515.
- Barton NH (1998). The effect of hitch-hiking on neutral genealogies. *Gen Res.* **72**: 123–133.
- Barton NH (2000). Genetic hitchhiking. *Philos Trans R Soc London B* **355**: 1553–1562.
- Begun DJ and Aquadro CF (1991). Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the *yellow-achaete* region. *Genetics* **129**: 1147–1158.
- Charlier C, Coppieters W, Farnir F, Grobet L, Leroy PL *et al* (1995). The *mh* causing double-muscling in cattle maps to bovine Chromosome 2. *Mamm Genome* **6**: 788–792.
- Compère G, Buren R, Hanset R (1996). *The Belgian Blue: The History of a Cattle Breed and its Conquest of the World*. Casterman, Tournai: Belgium.
- Dunner S, Charlier C, Farnir F, Brouwers B, Canon J *et al* (1997). Towards interbreed IBD fine mapping of the *mh* locus: double-muscling in the *Asturiana de los Valles* breed involves the same locus as in the *Belgian Blue* cattle breed. *Mamm Genome* **8**: 430–435.
- Dunner S, Miranda ME, Amigues Y, Cañon J, Georges M, Hanset R *et al* (2003). Haplotype diversity of the *myostatin* gene among beef cattle breeds. *Genet Sel Evol* **35**: in press.
- Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*. Addison Wesley Longman Limited, Harlow, UK.
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463–471.
- Goldstein DB, Zhivotovsky LA, Nayar K, RuizLinares A, Cavalli-Sforza LL, Feldman MW (1996). Statistical properties of the variation of linked microsatellite loci: Implications for the history of human Y chromosomes. *Mol Biol Evol* **13**: 1213–1218.
- Grobet L, Martin LJR, Poncelet D, Pirottin D, Brouwers B *et al* (1997). A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat Genet* **17**: 71–74.
- Grobet L, Poncelet D, Royo LJ, Brouwers B, Pirottin D *et al* (1998). Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mamm Genome* **9**: 210–213.
- Guo SW, Thompson EA (1992). Performing the exact test of Hardy–Weinberg proportions for multiple alleles. *Biometrics* **48**: 361–372.
- Hanset R (1982). Muscular hypertrophy as a racial characteristic: the case of the Belgian Blue breed. In: King JW, Ménissier F (eds) *Muscle Hypertrophy of Genetic Origin and its Use to Improve Beef Production*, Martinus Nijhoff Publishers, The Hague, pp 437–449.
- Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999). A scan for linkage disequilibrium across the human genome. *Genetics* **152**: 1711–1722.
- Kaplan NL, Hudson RR, Langley CH (1989). The 'hitchhiking effect' revisited. *Genetics* **123**: 887–899.
- Kim Y, Stephan W (2000). Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* **155**: 1415–1427.
- Kim Y, Stephan W (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.
- Kohn MH, Pelz H-J, Wayne RK (2000). Natural selection mapping of the warfarin-resistance gene. *Proc Natl Acad Sci USA* **97**: 7911–7915.
- Kruglyak L (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**: 139–144.
- Lin J, Arnold HB, Della-Fera MA, Azain MJ, Hartzell DL *et al* (2002). Myostatin knockout in mice increases myogenesis

- and decreases adipogenesis. *Biochem and Biophys Res Commun* **291**: 701–706.
- Louis EJ, Dempster ER (1987). An exact test for Hardy–Weinberg and multiple alleles. *Biometrics* **43**: 805–811.
- Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc.: Sunderland, MA.
- McPherron AC, Lee S-J (1997). Double muscling in cattle due to mutations in the myostatin gene. *Proc Natl Acad Sci USA* **94**: 12457–12461.
- Masoero G, Poujardieu B (1982). A note on the recent history of Piemontese cattle surveying the genetic determination of muscle hypertrophy. In: King JWB, Ménéssier F (eds.) *Muscle Hypertrophy of Genetic Origin and its use to Improve Beef Production*, Martinus Nijhoff Publishers: The Hague, pp 450–459.
- Maynard Smith J, Haigh J (1974). The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35.
- Ohta T, Kimura M (1975). The effect of a selected linked locus on heterozygosity of neutral alleles (the hitchhiking effect). *Genet Res* **25**: 313–326.
- Ott J (2000). Predicting the range of linkage disequilibrium. *Proc Natl Acad Sci* **97**: 2–3.
- Payseur BA, Cutter AD, Nachman MW (2002). Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. *Mol Biol Evol* **19**: 1143–1153.
- Peterson RJ, Goldman D, Long JC (1999). Effects of worldwide population subdivision on *ALDH2* linkage disequilibrium. *Genome Res* **9**: 844–852.
- Raymond M, Rousset F (1995). An exact test for population differentiation. *Evolution* **49**: 1280–1283.
- Riquet J, Coppieters W, Cambisano N, Arranz J-J, Berzi P *et al* (1999). Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. *Proc Natl Acad Sci USA* **96**: 9252–9257.
- Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ *et al* (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.
- Schimpf RJ, Winkelman-Sim DC, Buchanan FC, Aalhus JL, Plante Y *et al* (2000). QTL for marbling maps to cattle chromosome 2, In: *ISAG 2000 Conference Abstract Book*, Food Animal Biotechnology Center, University of Minnesota, p. 48.
- Schlötterer C, Vogl C, Tautz D (1997). Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* **146**: 309–320.
- Slatkin M (1995). Hitchhiking and associative overdominance at a microsatellite locus. *Mol Biol Evol* **12**: 473–480.
- Slatkin M, Wiehe T (1998). Genetic hitch-hiking in a subdivided population. *Gen Res* **71**: 155–160.
- Smith JA, Lewis AM, Wiener P, Williams JL (2000). Genetic variation in the bovine myostatin gene in UK beef cattle: Allele frequencies and haplotype analysis in the South Devon. *Anim Genet* **31**: 306–309.
- Stephan W, Wiehe THE, Lenz MW (1992). The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor Popul Biol* **41**: 237–254.
- Valdes AM, Slatkin M, Freimer NB (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**: 737–749.
- Weir BS (1996). *Genetic Data Analysis II*. Sinauer Associates, Inc.: Sunderland, MA.
- Wiehe T (1998). The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. *Theor Popul Biol* **53**: 272–283.
- Wiener P, Maclean I, Williams JL, Woolliams JA (2000). Testing for the presence of previously identified QTL for milk production traits in new populations. *Anim Genet* **31**: 385–395.
- Wiener P, Smith JA, Lewis AM, Woolliams JA, Williams JL (2002). Muscle-related traits in cattle: the role of the myostatin gene in the South Devon breed. *Gen Sel Evol* **34**: 221–232.