*Animal*, page 1 of 12 © The Animal Consortium 2012
doi:10.1017/S1751731112001152

animal

# Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries

R. F. Veerkamp[1†], M. P. Coffey[2], D. P. Berry[3], Y. de Haas[1], E. Strandberg[4], H. Bovenhuis[5], M. P. L. Calus[1] and E. Wall[2]

[1]*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 65, 8200 AB, Lelystad, The Netherlands;* [2]*Sustainable Livestock Systems Group, Scottish Agricultural College, Roslin Institute Building, Easter Bush Campus, Midlothian EH25 9RG, Scotland, UK;* [3]*Animal and Bioscience Research Department, Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, County Cork, Ireland;* [4]*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, PO Box 7023, S-75007 Uppsala, Sweden;* [5]*Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH, Wageningen, The Netherlands*

(Received 27 September 2011; Accepted 10 February 2012)

*Genome-wide association studies for difficult-to-measure traits are generally limited by the sample size with accurate phenotypic data. The objective of this study was to utilise data on primiparous Holstein–Friesian cows from experimental farms in Ireland, the United Kingdom, the Netherlands and Sweden to identify genomic regions associated with the feed utilisation complex: fat and protein corrected milk yield (FPCM), dry matter intake (DMI), body condition score (BCS) and live-weight (LW). Phenotypic data and 37 590 single nucleotide polymorphisms (SNPs) were available on up to 1629 animals. Genetic parameters of the traits were estimated using a linear animal model with pedigree information, and univariate genome-wide association analyses were undertaken using Bayesian stochastic search variable selection performed using Gibbs sampling. The variation in the phenotypes explained by the SNPs on each chromosome was related to the size of the chromosome and was relatively consistent for each trait with the possible exceptions of BTA4 for BCS, BTA7, BTA13, BTA14, BTA18 for LW and BTA27 for DMI. For LW, BCS, DMI and FPCM, 266, 178, 206 and 254 SNPs had a Bayes factor >3, respectively. Olfactory genes and genes involved in the sensory smell process were overrepresented in a 500 kbp window around the significant SNPs. Potential candidate genes were involved with functions linked to insulin, epidermal growth factor and tryptophan.*

**Keywords:** genome-wide association, feed intake, cow, live-weight

## Implications

Several single nucleotide polymorphisms (SNPs) were associated with live-weight, dry matter intake, body condition score and milk yield in a dataset combining data from research herds in four countries. In regions with significant SNP, potential candidate genes were identified for biological processes linked to taste and smell, a gene linked to the essential amino acid tryptophan that is involved with feeding motivation, as well as genes linked to regulation of insulin or epidermal growth factor.

## Introduction

There has been long running interest in how feed intake and feed efficiency should be taken into account in breeding decisions (see Veerkamp, 1998). Quantitative genetic studies have shown that genetic variation exists among animals in the feed utilisation complex, and in its component traits (e.g. feed intake, milk production and body weight; Veerkamp *et al.*, 2000; Berry *et al.*, 2007). Feed intake, as a trait, was initially targeted in breeding programs to try to reduce the quantity of feed required per unit of production, that is, improving feed efficiency. However, in the past two decades, interest has shifted towards the role of feed intake and its relationship with energy balance (EB), health and fertility. This shift in interest arose from the fact that current genetic selection for yield increases feed intake, but also results in a more negative energy balance (NEB) and more body tissue mobilisation during lactation (see Veerkamp *et al.*, 1993; Dillon *et al.*, 2006). As demonstrated by the size and direction of the genetic correlation between milk yield and feed intake and by the negative genetic correlation

† E-mail: Roel.Veerkamp@wur.nl

between milk yield and measures of EB, live-weight (LW; change) and body condition score (BCS; Veerkamp *et al.*, 2003), the expected correlated response in feed intake from selection on yield alone cannot cover the extra requirements needed for the increased yield. More recently, genetic selection for improvement in the feed utilisation complex has received increased interest as a means of reducing methane emission of dairy cows (Wall *et al.*, 2010).

Despite the reoccurring need to include the feed utilisation complex in dairy cattle breeding goals (i.e. improving feed efficiency, maintaining EB and/or reducing methane emission), there is, as yet, no direct selection practiced. This is primarily because of the large resource demand of measuring individual feed intake in dairy cows. This makes routine selection in breeding programmes too difficult, but also means that studies that estimated genetic parameters for the feed utilisation complex in dairy cows have generally been confined to a small number of animals from research herds (Svendsen *et al.*, 1994; Veerkamp *et al.*, 2000; Berry *et al.*, 2007). As a result, there has been considerable interest in predictors of the feed utilisation complex, like, for example, BCS (Koenen *et al.*, 2001; Banos *et al.*, 2004; Dal Zotto *et al.*, 2007), or other predictors of body energy state (Coffey *et al.*, 2003). However, studies with actual feed intake measures tend to have few animals recorded, which makes it difficult to establish the driving sources of genetic variation in feed efficiency (see Veerkamp *et al.*, 1993; Veerkamp, 2002) and optimal selection strategies due to lack of reliable genetic parameters.

Recent developments in genotyping technology have provided an alternative method of selection, in which DNA information is used. This would allow relatively cheap selection, compared with progeny testing for feed intake. Few studies have looked at the individual gene level for explaining genetic variation in feed intake. Examples are leptin, growth hormone or DGAT1 (Liefers *et al.*, 2002; Liefers *et al.*, 2005; Banos *et al.*, 2008; Oikonomou *et al.*, 2009), but the recent availability of fast-throughput genotyping platforms with dense genome-wide markers has reduced the cost of genotyping considerably, making whole-genome association studies a viable alternative. Potentially, functional processes and (quantitative trait loci) QTL that underlie the genetic variation in the feed utilisation complex might be identified. Nevertheless, accurate phenotypes on large numbers of animals and available in one dataset is the limiting factor in such association studies. To overcome this limitation, in this study we collated cow genotypic and phenotypic information from research herds in Ireland, the United Kingdom, the Netherlands and Sweden and related genetic markers across the genome to the phenotypes for the feed utilisation complex using a Bayesian Stochastic Search Variable Selection (BSSVS) statistical approach. The objective of this study was to identify putative genes and functional processes at gene level that might be involved in genetic variation in the feed utilisation complex. Fat–protein corrected milk yield is an important component of the feed utilisation complex, but has been given relatively little attention in this study, as much larger datasets are available in national programs.

## Material and methods

The data used in the present study originated from Teagasc, Moorepark, Ireland; the Langhill herd from Scottish Agricultural College, United Kingdom; two herds of Wageningen UR Livestock Research (NBZ and GEN), the Netherlands and the Jälla herd of the Swedish University of Agricultural Science. A detailed description of the experimental treatments imposed on the animals in the different countries is provided elsewhere, for Scotland (Veerkamp *et al.*, 1995; Pryce *et al.*, 1999; Coffey *et al.*, 2004), Ireland (Horan *et al.*, 2005), Sweden (Petersson *et al.*, 2006), the Netherlands (Veerkamp *et al.*, 2000; Beerda *et al.*, 2007), and a more detailed description of the merging of the data sources and variance components across the different herds is given by Banos *et al.* (2012). The analyses by Banos *et al.* (2012) illustrated the benefits of combining data across herds, and therefore the Swedish data were added in this study to increase the number of records for milk yield.

### *Phenotypic data*

Originally, phenotypic data were available on 2031 Irish, 1018 United Kingdom, 725 Dutch and 225 Swedish Holstein–Friesian cows. However, for this study only first lactation records were selected (Table 1), resulting in 1804 first lactation Holstein–Friesian cows with 66 116 test-day records collected up to 45 weeks (315 days) in lactation. Recorded traits were milk, fat and protein yield, LW, BCS, dry matter intake (DMI) and fat and protein corrected milk yield (FPCM). Extreme values for these traits (comprising $<1\%$) were excluded from further analysis, criteria used were for milk yield between 1 and 70 kg/day, with fat and protein yield $>0.025$ kg/day, LW had a minimum value of 100 kg and DMI of 1 kg/day. A description of data recording systems and frequency of measurements on each farm is described in detail by Banos *et al.* (2012). Data originated from different herds with different management and different frequencies of data recording and phenotypic data were pre-adjusted for management effects. For this purpose, a test-day model was fitted with a random permanent environmental animal effect as a sixth order polynomial (no genetic relationships included), a fixed effect for the mean lactation curve (fourth order polynomial) and a deviation of this mean curve for 10 management groups generated as an interaction between farm, nutritional treatment and milking frequency ($2\times$ or $3\times$ daily) combinations. Specific time-dependent random effects were fitted for year-month of milk test by management group (353 levels) and a specific treatment effect was fitted for experimental treatments during lactation for the cows in Ireland (81 levels). The model was fitted in ASReml (Gilmour *et al.*, 2009) and used to predict a full lactation curve for each cow. The average LW, BCS, DMI and FPCM of the predicted values for weeks 3 to 15 were used in the further analysis, for all animals with 10 or more observations in this period in the dataset. This period was chosen to maximise the number of cows with sufficient feed intake records, as a substantial proportion of the Dutch feed intake data used here was recorded till 100 days in milk only.

**Table 1** *The total number of cows with records, years of recording, management groups (HC, MC and LC, respectively; twice or three times a day milking (2 or 3, respectively), total number of weekly records for FPCM, LW, BCS and DMI separately and average number of weekly records for each cow per trait used in the pre-adjustment procedure*

| Country | Farm | No. of cows | Years of recording | Management groups | Total number of weekly records | | | | Average number of weekly records per cow | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | FPCM | LW | BCS | DMI | FCPM | LW | BCS | DMI |
| UK | Langhill | 558 | 1992 to 2010 | HC2, LC2, MC2 | 17 526 | 19 709 | 18 531 | 14 892 | 31.4 | 35.3 | 33.2 | 26.7 |
| NL | NBZ | 89 | 2003 to 2004 | HC2, HC3, LC2, LC3 | 1748 | 1825 | 623 | 1323 | 19.6 | 20.5 | 7.0 | 14.9 |
| SE | SLU | 209 | 1989 to 2009 | HC2 | 7172 | 0 | 0 | 0 | 34.3 | 0.0 | 0.0 | 0.0 |
| IE | Moorepark | 415 | 1998 to 2009 | LC2 | 16 814 | 13 487 | 6697 | 867 | 40.5 | 32.5 | 16.1 | 2.1 |
| NL | 't Gen | 545 | 1991 to 1998 | HC2 | 18 180 | 15 700 | 0 | 15 005 | 33.4 | 28.8 | 0.0 | 27.5 |

HC = high; MC = medium concentrate; LC = low concentrate; FPCM = fat and protein corrected milk; LW = live-weight; BCS = body condition score; DMI = dry matter intake; UK = United Kingdom; NL = the Netherlands; SE = Sweden; IE = Ireland.

*Genotypic data*

All animals with phenotypic information were genotyped with the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) containing 54 001 single nucleotide polymorphisms (SNPs). SNPs that fulfilled the following criteria were included in the association study: (1) GenCall (GC)score $>0.20$ and GenTrain (GT)score $>0.55$; (2) call rate $>0.95\%$; (3) minor allele frequency $>0.01$ in each country; and (4) no extreme deviation from Hardy–Weinberg Equilibrium (i.e. $\chi^2 < 600$). The GCscore and GTscore are quality measures on the genotype calls from the genotyping assay. After the quality control edits 37 590 SNPs remained. Checks for Mendelian inconsistencies between pedigree and SNP data were performed for all genotyped parent–offspring pairs and among sibs and animals with suspected erroneous pedigrees were removed (Calus *et al*., 2011). After these edits, 1629 animals with phenotypes and genotypes remained in the data. Missing SNPs were imputed using Beagle (Browning and Browning, 2007). When chromosome number and location of a SNP is presented in this study, these are on the BovineSNP50 from UMD3.0 bovine genome assembly from the University of Maryland as provided by Illumina (July 2010; Illumina Inc.). To identify functional genes in proximity of the SNP, the R package BioMart (Haider *et al*., 2009) was used based on the provided position on BTAU_4.0 (http://www.ensembl.org, Btau_4.0; April 2010). The package DAVID was used to cluster genes on functional process or gene function and to test whether certain processes and function were over-represented around the significant genes (Huang *et al*., 2009b).

*Linkage disequilibrium (LD)*

The persistency of the LD across the genome for different subpopulations is important for the power of the genome-wide association study. Following de Roos *et al*. (2008), persistency of LD ($r^2$ and $r$) were investigated for nine subpopulations in the complete dataset. The UK animals were split into the Langhill selection and control line animals (UKS and UKC), in the Netherlands there were three groups being NBZ herd, and the selected and not selected animals at the GEN herd (GENS and GENC), the Swedish (SLU) and Irish animals (TEA) formed one group each, and also the genotypes of 344 sires were used as an additional group (BULLS). It is important to note that all Irish animals were of North American ancestry. For each group, $r$ and $r^2$ were calculated between all neighbouring SNP pairs following Hill and Robertson (1986). The correlation between populations for $r$ was calculated across all marker distances. The decay in $r^2$ with marker distance was investigated by sorting on the distance between the neighbouring SNPs and taking the average $r^2$ and marker distance for bins of 50 SNPs. The average $r^2$ and distance were plotted and smoothed with a Loess Curve (smoothing parameter = 2/3) for each group. Pairs of SNPs closer together than 20 kbp were ignored, as many of these pairs had extremely low $r^2$. Finally, to quantify the decay in $r^2$ with distance between the pairs, it was assumed that $r^2 \approx 1/(4N_e c + 1)$, where $N_e$ is the effective population size and $c$ is the distance in Morgan (Sved, 1971). The combined parameter $N_e c$ was estimated for each group using the nonlinear least-squares ('nls') in the R package 'stats', which enabled to compare $N_e$ between the groups at a given genetic distance, albeit it is a very approximate estimate of the real $N_e$.

*Variance component estimation*

Genetic and residual (co)variances for the four traits defined above (LW, BCS, DMI and FPCM) were estimated using animal linear mixed models in ASReml (Gilmour *et al*., 2009). No fixed effects were included in the model because pre-adjustments of the data were made before analysis. Pedigree information on each genotyped animal was traced back as far as possible; a total of 9368 identities over 19 generations were included in the pedigree file.

*Genome-wide association analysis*

The BSSVS model (described by Verbyla *et al*., 2010) was used to sample whether SNPs were linked to a QTL or not. The model used was

$$y_i = \mu + a_i + \sum_{j=1}^{n} \sum_{k=1}^{2} SNP_{ijk} + e_i$$

where $y_i$ is the phenotypic record of animal $i$, $\mu$ is the overall mean, $a_i$ is the random polygenic effect of animal $i$, $n$ is

number of SNP markers, $SNP_{ijk}$ is a random effect for allele $k$ at locus $j$ of animal $i$ and $e_i$ is a random residual for animal $i$. SNP effects ($SNP_{ijk}$) were estimated as $q_{ijk} \times v_j$ (Meuwissen and Goddard, 2004), where $q_{ijk}$ is the size of the effect of allele $k$ at locus $j$ and $v_j$ is a scaling factor for locus $j$. The vectors $\boldsymbol{a}$ and $\boldsymbol{e}$ where assumed to be normally distributed, $\boldsymbol{a} \sim \mathbf{N}(\mathbf{0}, \mathbf{A}\sigma_u^2)$ and $\boldsymbol{e} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where $\mathbf{A}$ is the numerator relationship matrix and $\mathbf{I}$ is an identity matrix.

An indicator variable $I_j$ was introduced into the hierarchical model to allow for scenarios where a limited number of SNPs with large effects exist and for other scenarios where a large number of SNPs all had an infinitely small effect (the infinitesimal model). This indicator variable takes a value of 1 or 0, depending on whether the SNP was included with a large effect in the model or not. The QTL-indicator $I_j$ was sampled from

$$I_j | v_j, \sigma_v^2 \sim Bernoulli \leftarrow \left[ \frac{\phi\left(v_j;\, 0,\, \sigma_v^2\right)(1-\pi)}{\phi\left(v_j;\, 0,\, \sigma_v^2\right)(1-\pi) + \phi\left(v_j;\, 0, \frac{\sigma_v^2}{100}\right)\pi} \right]$$

where $\pi$ is the probability that $I_j$ is unequal to 1, which follows a Bernoulli distribution. The probability that $v_j$ is sampled from the distribution with large effects is $\phi\left(v_j;\, 0,\, \sigma_v^2\right)$ with variance $\sigma_v^2$ and mean 0. The prior QTL probability $(1 - \pi)$ used in the analyses had a value of 0.00027, reflecting the prior assumption that 10 QTL come from the distribution with large effects. This is equivalent to fixing $\pi$ at a value of 0.99973, which is the term commonly used for BayesB type models to denote the proportion of SNPs that are *not* included in the model and therefore have zero effect, whereas for BSSVS it defines the proportion of loci with effects sampled from the distribution with small effects. The posterior distribution of $I_j$ showed the proportion that the effects for locus $j$ were sampled from the distribution with large effects, across all cycles after the burn-in. SNPs with the highest posterior probability were expected to be most likely linked to a QTL. A more detailed description of the model is given by Calus and Veerkamp (2011).

The estimated variance components from the quantitative genetic analysis were used to infer prior variances for the BSSVS model to estimate SNP effects. The BSSVS model was performed using Gibbs sampling, each time the model was run for 50 000 cycles with 10 000 cycles discarded for burn-in. For the prior QTL variance it was assumed that 80% of the genetic variance was explained by the SNPs and the remaining 20% by the polygenic component in the model. This division was based upon 80% of the total additive variance to be explained by the Bovine50 beadchip used in this study (Daetwyler, 2009). The 80% of the genetic variance explained by the SNP was split again across 10 SNPs sampled from the distribution with a large effect and the remaining 37 580 SNPs were sampled from the distribution with small effect (i.e. with a SNP variance that is one hundred times smaller than the variances for a SNP with a large effect). Each trait was analysed with five different Gibbs chains of 50 000 cycles and average results of the five chains were used in the further study.

A SNP was assumed to be significant if the Bayes factor (BF; Kass and Raftery, 1995) was $\geqslant$3.1, which is termed a 'substantial' effect. The BF was calculated for each SNP by

$$BF = \frac{Pr(H_1|y)}{1 - Pr(H_1|y)} \div \frac{Pr(H_1)}{1 - Pr(H_1)}$$

where $H_1$ is the hypothesis that the marker is linked to a QTL, $Pr(H_1|y)$ is the posterior probability of a QTL being present and $Pr(H_1)$ is the prior probability of a QTL being present, $(1 - Pr(H_1|y))$ and $(1 - Pr(H_1))$ represent the posterior and prior probability for the alternative hypothesis, respectively. As many SNPs are closely linked and different sets of SNPs might give an equivalent good fit to the data in different chains, the sum of the posterior and prior QTL probabilities of a window of 10 SNPs (a SNP plus its closest neighbours) was also used. Using these summed probabilities might reveal regions with QTL that would be missed otherwise.

The genetic variance of trait $i$ was calculated for each SNP $k$ separately:

$$\sigma'^{2}_{ik} = 2p_k q_k \times a_{ik}^2$$

where $a_{ik}$ is the allele substitution effect of SNP $k$ on trait $I$, $p_k$ is the frequency of the minor allele of SNP $k$ and $q_k$ is the frequency of the major allele of SNP $k$, both obtained from the genotypes. The sum of the variances of all SNPs for trait $i$ was taken as the total genetic variance explained by the SNPs for trait $i$.

## Results

In the entire dataset, 1629 lactations remained (Table 2) after the selection criteria of at least 10 records up to week 15. For BCS mainly records from the UK Langhill herd complied with this criteria, as well as DMI records of the Netherlands and the United Kingdom. Heritability estimates of the four traits were between 0.37 and 0.58. Genetic correlations were of the same sign as their respective phenotypic correlation, but generally stronger (Table 2).

The correlations between $r$ across the genome were 0.95 and 0.97 between seven of the eight cow populations, and between 0.97 and 0.99 with the group of sires. The Langhill control line was somewhat different and $r$ had a correlation of 0.92 with the other cow populations and of 0.95 with the group of sires. The decay in $r^2$ was similar between the lines (Figure 1), and the combined $N_e c$ parameter was estimated to be 0.107, 0.100, 0.100, 0.107, 0.107, 0.105, 0.100 and 0.105 for Langhill selection and control line (UKC and UKS), NBZ herd, SLU herd, Teagasc herd, the selected and control GEN herd and the group sires, respectively. These $N_e c$ values gave equivalent decay curves for each group, and indicated that $N_e$ of the different subpopulations ranged from 12 462 to 13 344 at average marker distance of 0.08 cMorgans (623 generations ago).

**Table 2** *Number of cows (*n*) and estimated heritability, with its standard error (s.e.) per trait, and the genetic (below diagonal; standard errors in parenthesis) and phenotypic (above diagonal; standard errors in parenthesis) correlations among the traits*

| | | | | Correlations | | | |
|---|---|---|---|---|---|---|---|
| Trait | *n* | Heritability | s.e. | FPCM | DMI | BCS | LW |
| FPCM | 1629 | 0.37 | 0.05 | | 0.51 (0.03) | −0.10 (0.04) | 0.23 (0.03) |
| DMI | 970 | 0.53 | 0.08 | 0.41 (0.10) | | 0.17 (0.05) | 0.46 (0.02) |
| BCS | 564 | 0.58 | 0.09 | −0.09 (0.14) | 0.39 (0.13) | | 0.50 (0.03) |
| LW | 1416 | 0.47 | 0.07 | 0.39 (0.11) | 0.68 (0.08) | 0.52 (0.10) | |

FPCM = fat and protein corrected milk; LW = live-weight; BCS = body condition score; DMI = dry matter intake.

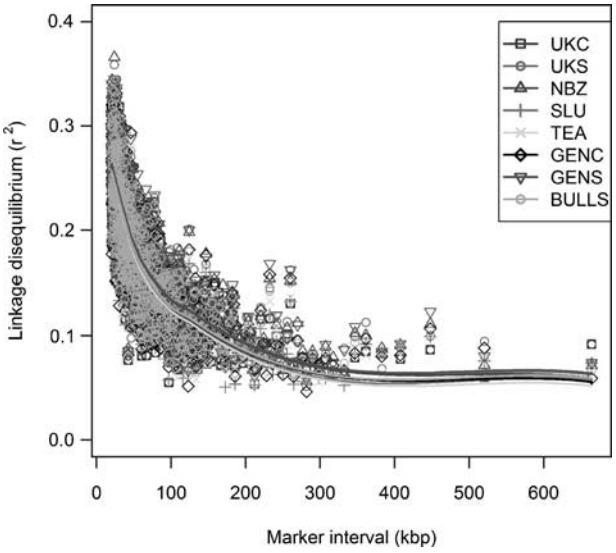


**Figure 1** Linkage disequilibrium (LD; $r^2$) as a function of interval between neighbouring markers for nine different groups: Langhill control and selection line (UKC and UKS), NBZ herd in the Netherlands, SLU herd in Sweden, Teagasc herd in Ireland (TEA), the selected and control herd in the Netherlands (GENS and GENC) and a group of 334 sires of these animals (BULLS).

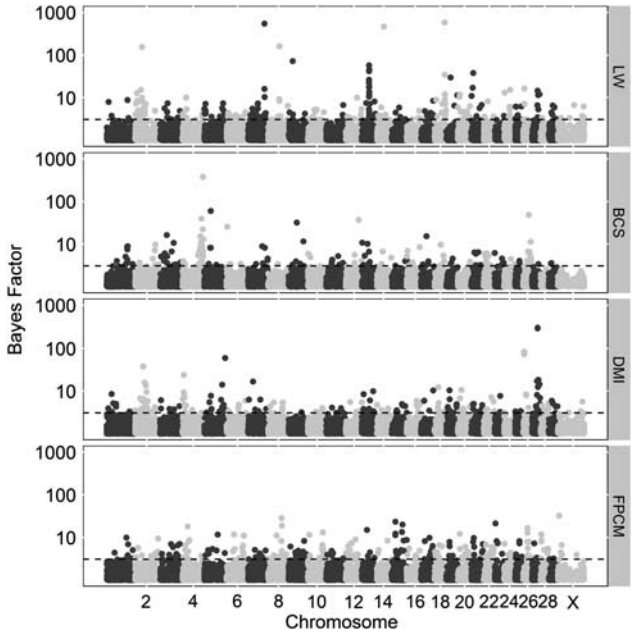


**Figure 2** Bayes factor (BF) corresponding to the posterior quantitative trait loci probabilities for each individual single nucleotide polymorphism, from top to bottom, for live-weight (LW), body condition score (BCS), dry matter intake (DMI) and fat and protein corrected milk yield (FPCM) for each of the autosomes from left (BTA1) to right (BTA29) and the X-chromosome at the far right. Dashed lines indicate BF 3.1.

Across all four traits investigated, 872 SNPs of 37 590 SNPs had a BF >3.1 for at least one trait (Figure 2), whereas 266, 178, 206 and 254 SNPs had a BF >3.1 for LW, BCS, DMI and FPCM, respectively. The number of SNPs with a BF >10 was 36, 18, 19, 22 for LW (Table 3), BCS (Table 4), DMI (Table 5) and FPCM (Table 6), respectively. Using the probabilities across a window of 10 SNPs, in total 300 regions remained significant and 113, 81, 57 and 49 SNPs had a BF above 3.1 for LW, BCS, DMI and FPCM, respectively (Figure 3). The variation explained by the SNPs on each chromosome was related to the size of the chromosome and was relatively consistent for each trait (Table 7), with the possible exceptions of BTA4 for BCS, BTA7, BTA13, BTA14, BTA18 for LW and BTA27 for DMI.

For LW the largest individual BF was on BTA18, BTA7 and BTA14 (Table 3), explaining 0.63%, 0.53% and 0.50% of the total genetic variance explained by SNPs, respectively. All 36 SNPs with a BF >10 explained 2.5% of the genetic variation. Furthermore, four SNPs at BTA13 each explained 0.02% or more of the genetic variance in LW and several other SNPs in close proximity to BTA13 were also associated with LW (Figure 4).

A SNP on BTA4 explained 0.68% of the genetic variance in BCS and all 18 SNPs with a BF >10 explained, in total, 1.2% of the genetic variation in BCS (Table 4). For DMI, one SNP on BTA27 explained 0.4% of the genetic variance (Table 5) with several SNPs on predominantly BTA2, BTA26 and BTA27 also explaining genetic variation in DMI. The 19 SNPs with a BF >10 explained 0.8% of the genetic variation.

Of the 872 SNPs associated (i.e. BF >3.1) with at least one of the phenotypes investigated, 31 were associated with two traits and 1 was associated with three traits. This SNP was ARS-BFGL-NGS-71055 on chromosome 27 (34904207 bp) affecting LW, DMI and BCS (Figure 6).

On BTA2, one SNP explained 0.17% of the genetic variance in LW (BF of 153.6) and 0.01% of the genetic variance in DMI (BF of 6.0). Two SNPs in the middle of the putative

**Table 3** *SNPs that explained most genetic variation for LW with a BF >10 with position on the genome, the posterior probability that the SNP effect is sampled from the distribution with a large effect, the significance (BF) and the variance explained by the single SNP*

| Chromosome | SNP name | rs | UMD3 position | Posterior probability | BF | SNP variance (BCS units$^2$) | Total variance (%) |
|---|---|---|---|---|---|---|---|
| 2 | BTB-00862061 | 42025600 | 8933349 | 0.004 | 12.7 | 0.06 | 0.01 |
| 2 | ARS-BFGL-NGS-38895 | 110915377 | 31793643 | 0.005 | 15.4 | 0.07 | 0.01 |
| 2 | ARS-BFGL-NGS-114956 | 110791598 | 39199347 | 0.044 | 153.6 | 0.96 | 0.17 |
| 7 | Hapmap46473-BTA-21133 | 41625106 | 92293717 | 0.141 | 546.8 | 2.96 | 0.53 |
| 7 | ARS-BFGL-NGS-5139 | 110804446 | 92474466 | 0.005 | 16.1 | 0.04 | 0.01 |
| 7 | ARS-BFGL-NGS-18900 | 110059753 | 93218452 | 0.003 | 10.3 | 0.04 | 0.01 |
| 8 | BTA-81701-no-rs | 41589869 | 64266000 | 0.046 | 161.6 | 0.97 | 0.17 |
| 9 | ARS-BFGL-NGS-104704 | 110042295 | 22843876 | 0.021 | 71.9 | 0.34 | 0.06 |
| 13 | ARS-BFGL-NGS-95823 | 109596793 | 40722564 | 0.003 | 10.5 | 0.04 | 0.01 |
| 13 | ARS-BFGL-NGS-41919 | 109564214 | 40777908 | 0.017 | 57.4 | 0.24 | 0.04 |
| 13 | ARS-BFGL-NGS-116093 | 109695286 | 40974987 | 0.012 | 42 | 0.18 | 0.03 |
| 13 | ARS-BFGL-NGS-113658 | 108948219 | 41015075 | 0.007 | 22.5 | 0.09 | 0.02 |
| 13 | BTA-32481-no-rs | 41629039 | 41111195 | 0.008 | 26.3 | 0.11 | 0.02 |
| 13 | ARS-BFGL-NGS-60607 | 109554041 | 41164722 | 0.005 | 17.7 | 0.07 | 0.01 |
| 13 | ARS-BFGL-NGS-101382 | 109542102 | 41366870 | 0.013 | 45.6 | 0.22 | 0.04 |
| 13 | ARS-BFGL-NGS-16279 | 110689635 | 41414256 | 0.005 | 18.3 | 0.08 | 0.01 |
| 13 | BTA-26412-no-rs | 109799865 | 41506674 | 0.003 | 10.7 | 0.04 | 0.01 |
| 13 | ARS-BFGL-NGS-102025 | 110274438 | 41529941 | 0.004 | 11.7 | 0.05 | 0.01 |
| 13 | Hapmap61089-rs29014590 | 29014590 | 42070556 | 0.004 | 13.7 | 0.05 | 0.01 |
| 14 | ARS-BFGL-BAC-21948 | 109179401 | 40512939 | 0.123 | 465.8 | 2.79 | 0.50 |
| 18 | ARS-BFGL-NGS-98028 | 41636749 | 57174711 | 0.149 | 581.4 | 3.54 | 0.63 |
| 18 | ARS-BFGL-NGS-109285 | 109478645 | 57589121 | 0.01 | 34.6 | 0.14 | 0.03 |
| 18 | Hapmap40906-BTA-121147 | 41664920 | 58551307 | 0.004 | 12.1 | 0.06 | 0.01 |
| 19 | UA-IFASA-6091 | 41584914 | 24521970 | 0.009 | 29.7 | 0.14 | 0.02 |
| 19 | UA-IFASA-8578 | 41580816 | 62760617 | 0.003 | 11.4 | 0.05 | 0.01 |
| 20 | ARS-BFGL-NGS-100777 | 41935177 | 9965585 | 0.003 | 10.6 | 0.07 | 0.01 |
| 20 | ARS-BFGL-NGS-71611 | 110815453 | 11035444 | 0.004 | 11.9 | 0.06 | 0.01 |
| 21 | ARS-BFGL-NGS-31507 | 110426938 | 2592694 | 0.003 | 11 | 0.05 | 0.01 |
| 21 | ARS-BFGL-NGS-100241 | 110440906 | 8955497 | 0.005 | 17.1 | 0.1 | 0.02 |
| 21 | Hapmap55337-rs29009998 | 29009998 | 11981034 | 0.011 | 37.8 | 0.18 | 0.03 |
| 24 | ARS-BFGL-NGS-63149 | 42039872 | 3721378 | 0.004 | 13.1 | 0.05 | 0.01 |
| 24 | Hapmap57214-rs29016076 | 29016076 | 34928812 | 0.005 | 15.4 | 0.09 | 0.02 |
| 26 | ARS-USMARC-Parent-DQ990834 | 29013727 | 8221270 | 0.005 | 16.4 | 0.07 | 0.01 |
| 27 | Hapmap40631-BTA-103396 | 41615000 | 31461111 | 0.004 | 14.5 | 0.06 | 0.01 |
| 27 | ARS-BFGL-NGS-29650 | 42634874 | 34326743 | 0.003 | 11.5 | 0.05 | 0.01 |
| 27 | Hapmap50424-BTA-63130 | 110518457 | 41726707 | 0.004 | 12.1 | 0.06 | 0.01 |

SNP = single nucleotide polymorphism; LW = live-weight; BF = Bayes factor; BCS = body condition score.

QTL region on BTA13 for LW (Figure 4) had a BF >3.1 for BCS and DMI as well. Similarly, SNPs associated with LW were in close proximity to the putative QTL region for BCS on BTA26 (Figure 5), and on BTA27 several SNPs were associated (i.e. BF >3.1) with FPCM, DMI, LW and BCS (Figure 6).

## Discussion

*Genetic variation*

The heritability estimates of 0.37 to 0.58 are similar to the estimates previously reported for these traits (Koenen and Veerkamp, 1998; Berry *et al.*, 2003). The increased DMI for heavier animals and animals that yield more milk is expected and agrees with previous studies as does the phenotypic and genetic correlations of 0.50 to 0.52 between BCS and LW (Berry *et al.*, 2002). The general pattern that the genetic variance explained by SNPs on each chromosome decreases with the decreasing size of the chromosome is consistent with a polygenic model. Exceptions to this general rule that genetic variance explained by SNPs is related to the size of the chromosome are BTA7, BTA13, BTA14 and BTA18 for LW, BTA4 for BCS and BTA27 for DMI. These chromosomes explain more variance than expected in the case that variance were uniformly distributed across the genome.

The chromosomes that explain more genetic variance than expected relative to their size are also the chromosomes with most significant SNPs for these traits. Nevertheless, SNPs explained only a small proportion of the overall variance. With a maximum of 0.7% of the variance explained by a single SNP, it is expected that thousands of QTL are required to explain the full variance and the effect is smaller than is expected given the power of the experiment.

**Table 4** *SNPs that explained most genetic variation for BCS with a BF >10 with position on the genome, the posterior probability that the SNP effect is sampled from the distribution with a large effect, the significance (BF) and the variance explained by the single SNP*

| Chromosome | SNP name | rs | UMD3 position | Posterior probability | BF | SNP variance (BCS units$^2$) | Total variance (%) |
|---|---|---|---|---|---|---|---|
| 3 | BTA-91560-no-rs | 41661577 | 38507659 | 0.005 | 16.3 | 0.0000064 | 0.02 |
| 3 | BTA-68406-no-rs | 41608825 | 77666001 | 0.003 | 10.7 | 0.0000045 | 0.02 |
| 4 | ARS-BFGL-NGS-31561 | 109395087 | 102433510 | 0.005 | 15.4 | 0.0000044 | 0.02 |
| 4 | ARS-BFGL-NGS-72134 | 110800831 | 106253833 | 0.003 | 11.3 | 0.0000036 | 0.01 |
| 4 | BTA-90241-no-rs | 41612738 | 107985246 | 0.012 | 39.4 | 0.0000119 | 0.04 |
| 4 | Hapmap28815-BTA-155553 | 81169737 | 116390964 | 0.007 | 22.2 | 0.0000065 | 0.02 |
| 4 | Hapmap60162-rs29015028 | 29015028 | 116485311 | 0.102 | 377.2 | 0.0001796 | 0.68 |
| 4 | ARS-BFGL-NGS-33132 | 41669146 | 119617605 | 0.004 | 13.1 | 0.0000037 | 0.01 |
| 5 | BTB-00225371 | 43435407 | 36578127 | 0.018 | 59.9 | 0.0000228 | 0.09 |
| 6 | BTB-01280976 | 42404150 | 4193024 | 0.008 | 25.7 | 0.0000075 | 0.03 |
| 9 | BTA-83597-no-rs | 41658682 | 46632366 | 0.009 | 31.8 | 0.000009 | 0.03 |
| 12 | ARS-BFGL-NGS-28261 | 110940900 | 72311316 | 0.011 | 36.6 | 0.0000121 | 0.05 |
| 13 | ARS-BFGL-BAC-11281 | 109707704 | 2217706 | 0.003 | 10.8 | 0.0000028 | 0.01 |
| 13 | ARS-BFGL-NGS-69389 | 109983109 | 32831454 | 0.003 | 10.1 | 0.0000034 | 0.01 |
| 17 | BTB-00676214 | 41840165 | 30540080 | 0.005 | 15.4 | 0.0000049 | 0.02 |
| 26 | BTB-02036417 | 43145430 | 32716808 | 0.014 | 48.2 | 0.0000185 | 0.07 |
| 26 | Hapmap52062-rs29027270 | 29027270 | 43479080 | 0.003 | 11.3 | 0.0000034 | 0.01 |

SNP = single nucleotide polymorphism; BF = Bayes factor; BCS = body condition score.

**Table 5** *SNPs that explained most genetic variation for DMI with a BF >10 with position on the genome, the posterior probability that the SNP effect is sampled from the distribution with a large effect, the significance (BF) and the variance explained by the single SNP*

| Chromosome | SNP name | rs | UMD3 position | Posterior probability | BF | SNP variance (BCS units$^2$) | Total variance (%) |
|---|---|---|---|---|---|---|---|
| 2 | Hapmap59844-rs29025329 | 29025329 | 44947955 | 0.011 | 37.9 | 0.00035 | 0.05 |
| 2 | BTB-01559051 | 42674303 | 53431931 | 0.005 | 15.9 | 0.00013 | 0.02 |
| 2 | ARS-BFGL-NGS-4659 | 109704643 | 60143778 | 0.004 | 13.6 | 0.00016 | 0.02 |
| 3 | BTA-68406-no-rs | 41608825 | 77666001 | 0.003 | 10.7 | 0.0000045 | 0.02 |
| 4 | ARS-BFGL-NGS-114611 | 110661405 | 8421365 | 0.007 | 23.9 | 0.00015 | 0.02 |
| 5 | BTA-116856-no-rs | 41618337 | 100826813 | 0.004 | 14.2 | 0.00009 | 0.01 |
| 5 | ARS-BFGL-NGS-14632 | 110117542 | 118501191 | 0.018 | 59.9 | 0.00048 | 0.06 |
| 7 | BTB-01074080 | 42232105 | 27973088 | 0.005 | 16.7 | 0.0001 | 0.01 |
| 13 | ARS-BFGL-BAC-11281 | 109707704 | 2217706 | 0.003 | 10.8 | 0.0000028 | 0.01 |
| 13 | ARS-BFGL-NGS-69389 | 109983109 | 32831454 | 0.003 | 10.1 | 0.0000034 | 0.01 |
| 18 | BTA-42769-no-rs | 41579997 | 20765346 | 0.004 | 12.2 | 0.00009 | 0.01 |
| 26 | BTB-01078268 | 42233405 | 7846224 | 0.024 | 83.3 | 0.00068 | 0.09 |
| 26 | BTB-01078331 | 42234268 | 7900988 | 0.021 | 72.5 | 0.00063 | 0.08 |
| 26 | Hapmap52062-rs29027270 | 29027270 | 43479080 | 0.003 | 11.3 | 0.0000034 | 0.01 |
| 27 | Hapmap43421-BTA-97718 | 41666656 | 28148660 | 0.005 | 17.2 | 0.00014 | 0.02 |
| 27 | Hapmap40631-BTA-103396 | 41615000 | 31461111 | 0.081 | 295.4 | 0.00303 | 0.40 |
| 27 | ARS-BFGL-NGS-112047 | 43727261 | 33442942 | 0.004 | 12.6 | 0.00012 | 0.02 |
| 27 | ARS-BFGL-NGS-108861 | 110320533 | 34758932 | 0.006 | 18.4 | 0.00014 | 0.02 |
| 27 | Hapmap50424-BTA-63130 | 110518457 | 41726707 | 0.004 | 14.4 | 0.00012 | 0.02 |

SNP = single nucleotide polymorphism; DMI = dry matter intake; BF = Bayes factor; BCS = body condition score.

An explanation for the small proportion of variance explained by each SNP might be that all SNPs were fitted simultaneously in the model and in comparison with single SNP analysis, this might have reduced the total variance explained by each SNP. There might not be enough discriminant power in the data to separate the SNPs close together, as several SNPs are expected to be in LD with each other, and the effects of the same QTL will be distributed across several SNPs. In addition, the average SNP effects across five chains were taken to avoid a single SNP being highly significant in one chain, but the effect being attributed to a neighbouring SNP in subsequent chain. The effect of several SNPs in close proximity of each other explaining each a little of the genetic variance is, for example, illustrated by

**Table 6** *SNPs that explained most genetic variation for FPCM with a BF >10 with position on the genome, the posterior probability that the SNP effect is sampled from the distribution with a large effect, the significance (BF) and the variance explained by the single SNP*

| Chromosome | SNP name | rs | UMD3 position | Posterior probability | BF | SNP variance (BCS units$^2$) | Total variance (%) |
|---|---|---|---|---|---|---|---|
| 0 | BTA-69349-no-rs | 43741236 | 0 | 0.004 | 13.2 | 0 | 0.01 |
| 4 | Hapmap38265-BTA-96973 | 41592066 | 30526269 | 0.005 | 18 | 0.001 | 0.02 |
| 5 | Hapmap46673-BTA-74066 | 41566931 | 75205122 | 0.003 | 11.7 | 0 | 0.01 |
| 6 | BTA-75902-no-rs | 43455987 | 42155077 | 0.004 | 11.7 | 0 | 0.01 |
| 6 | Hapmap36567-SCAFFOLD30438_8760 | 29016199 | 97937238 | 0.003 | 11 | 0 | 0.01 |
| 8 | ARS-BFGL-NGS-118200 | 110589243 | 49203188 | 0.003 | 11.1 | 0 | 0.01 |
| 8 | BTA-81945-no-rs | 41658302 | 77252882 | 0.009 | 28.7 | 0.001 | 0.02 |
| 8 | BTB-01139515 | 42296470 | 78701950 | 0.006 | 18.9 | 0.001 | 0.01 |
| 10 | ARS-BFGL-NGS-5922 | 110672262 | 9789845 | 0.003 | 11.5 | 0 | 0.01 |
| 10 | BTB-00441734 | 43651067 | 85944517 | 0.004 | 13.2 | 0 | 0.01 |
| 13 | ARS-BFGL-NGS-113236 | 109655563 | 28989009 | 0.004 | 15 | 0.001 | 0.01 |
| 15 | BTB-00582624 | 41745388 | 21138866 | 0.003 | 11.3 | 0 | 0.01 |
| 15 | BTB-02053380 | 43160689 | 21162454 | 0.004 | 12.3 | 0 | 0.01 |
| 15 | BTB-00582817 | 41746981 | 21401836 | 0.007 | 23.5 | 0.001 | 0.02 |
| 15 | ARS-BFGL-BAC-6489 | 110748087 | 58411725 | 0.004 | 13.4 | 0.001 | 0.01 |
| 15 | BTB-01059544 | 42218318 | 61098317 | 0.006 | 20 | 0.001 | 0.02 |
| 20 | Hapmap59100-rs29024884 | 29024884 | 11770435 | 0.004 | 12.1 | 0 | 0.01 |
| 23 | ARS-BFGL-NGS-106470 | 110809507 | 5047747 | 0.006 | 21.3 | 0.001 | 0.02 |
| 26 | Hapmap53060-rs29020888 | 29020888 | 25032529 | 0.005 | 16.7 | 0 | 0.01 |
| 26 | ARS-BFGL-NGS-25560 | 42097667 | 25407197 | 0.004 | 13 | 0 | 0.01 |
| 28 | BTA-100905-no-rs | 41612720 | 4866330 | 0.004 | 11.7 | 0 | 0.01 |
| 30 | Hapmap41447-BTA-118863 | 41573451 | 7268125 | 0.01 | 32.3 | 0 | 0.01 |

SNP = single nucleotide polymorphism; FPCM = fat and protein corrected milk; BF = Bayes factor; BCS = body condition score.

**Table 7** *Percentage of total SNP variance[†] of LW, BCS, DMI and FPCM explained by SNPs on each chromosome*

| Chromosome | LW (%) | BCS (%) | DMI (%) | FPCM (%) | Chromosome | LW (%) | BCS (%) | DMI (%) | FPCM (%) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.8 | 6.0 | 6.0 | 6.1 | 16 | 3.0 | 3.0 | 3.1 | 3.0 |
| 2 | 5.3 | 5.1 | 5.3 | 5.2 | 17 | 2.9 | 2.9 | 2.9 | 2.9 |
| 3 | 4.6 | 4.8 | 4.7 | 4.7 | 18 | **3.1** | 2.5 | 2.5 | 2.5 |
| 4 | 4.6 | **5.6** | 4.7 | 4.7 | 19 | 2.6 | 2.6 | 2.6 | 2.6 |
| 5 | 4.0 | 4.1 | 4.1 | 4.0 | 20 | 2.8 | 2.8 | 2.8 | 2.8 |
| 6 | 4.6 | 4.7 | 4.7 | 4.8 | 21 | 2.7 | 2.7 | 2.7 | 2.7 |
| 7 | **4.7** | 4.1 | 4.1 | 4.2 | 22 | 2.4 | 2.5 | 2.5 | 2.5 |
| 8 | 4.3 | 4.1 | 4.2 | 4.3 | 23 | 2.0 | 2.1 | 2.1 | 2.1 |
| 9 | 3.8 | 3.8 | 3.8 | 3.9 | 24 | 2.3 | 2.3 | 2.3 | 2.3 |
| 10 | 4.0 | 4.1 | 4.0 | 4.1 | 25 | 2.0 | 2.0 | 2.0 | 2.0 |
| 11 | 4.1 | 4.2 | 4.2 | 4.2 | 26 | 1.9 | 2.1 | 2.1 | 2.0 |
| 12 | 3.1 | 3.2 | 3.1 | 3.2 | 27 | 1.8 | 1.8 | **2.3** | 1.8 |
| 13 | **3.7** | 3.5 | 3.5 | 3.5 | 28 | 1.7 | 1.8 | 1.8 | 1.8 |
| 14 | **3.8** | 3.4 | 3.4 | 3.4 | 29 | 2.1 | 2.1 | 2.1 | 2.1 |
| 15 | 3.1 | 3.1 | 3.1 | 3.3 | X | 1.9 | 1.9 | 1.9 | 1.9 |

SNP = single nucleotide polymorphism; LW = live-weight; BCS = body condition score; DMI = dry matter intake; FPCM = fat and protein corrected milk.
[†]1.4% of the genetic variance was explained by SNPs not mapped to chromosomes yet.

the SNPs associated with LW on BTA13 (Figures 2 and 4). At the same time on BTA14 one SNP associated with LW was found, whereas no other neighbouring SNPs were found. However, this SNP explained only 0.5% of the variance, suggesting that the most likely scenario is that many mutations, all with a relatively small effect, explain the total genetic variance for LW, DMI and BCS.

*Functional processes relating to SNPs associated with DMI, LW and BCS*

In total, 728 genes and 500 unique genes were found within 500 kbp flanking the SNPs associated with LW, DMI or BCS using BioMart (Haider *et al*., 2009). Fourteen of these genes were present in proximity of the SNPs associated with both LW and DMI, and 27 in proximity of the SNPs associated with

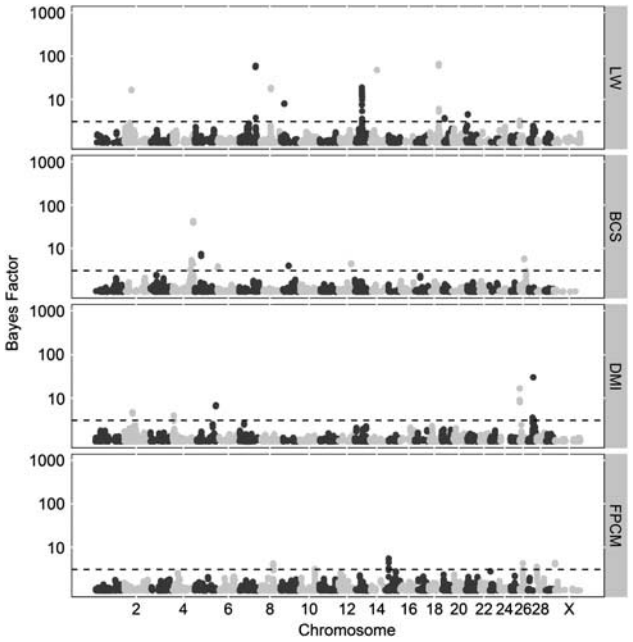


**Figure 3** Posterior Bayes factor (BF) corresponding to the posterior quantitative trait loci probabilities for a window of 10 neighbouring single nucleotide polymorphisms, from top to bottom, for live-weight (LW), body condition score (BCS), dry matter intake (DMI) and fat and protein corrected milk yield (FPCM) for each of the autosomes from left (BTA1) to right (BTA29) and the X-chromosome at the far right. Dashed lines indicate BF 3.1.

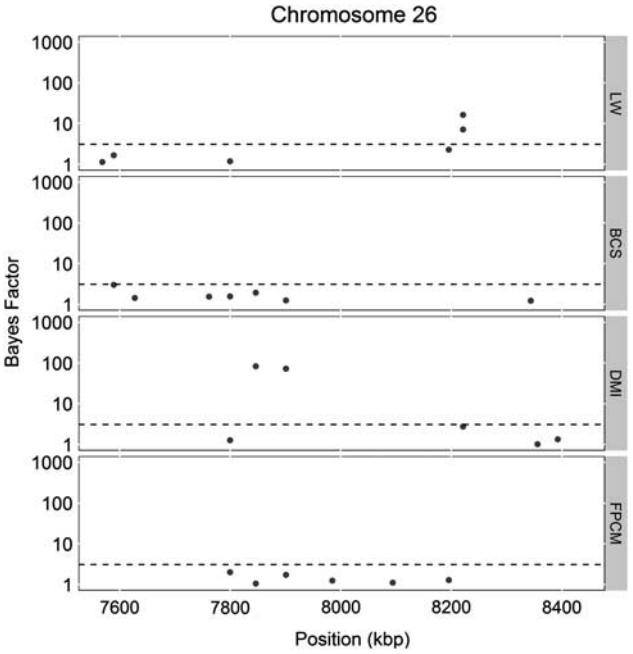


**Figure 5** Bayes factor (BF) corresponding to the posterior quantitative trait loci probabilities for single nucleotide polymorphism with BF $>1$ on a region of chromosome 26 (BTA 26) for live-weight (LW), body condition score (BCS), dry matter intake (DMI) and fat and protein corrected milk yield (FPCM). Dashed lines indicate BF 3.1.

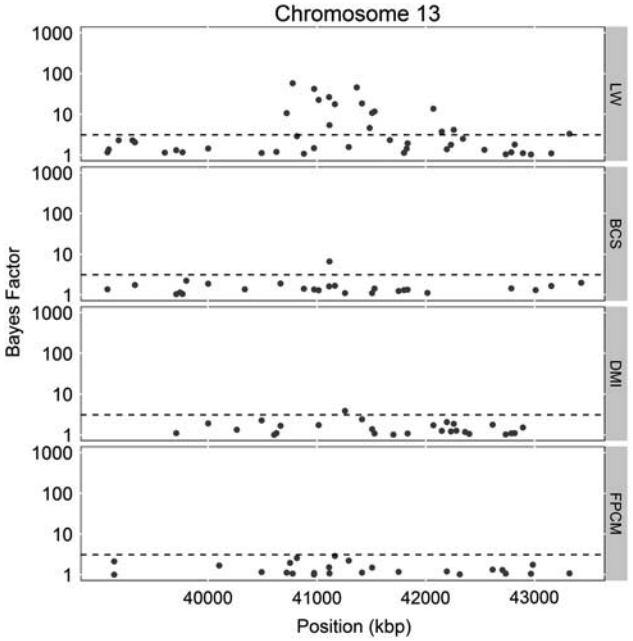


**Figure 4** Bayes factor (BF) corresponding to the posterior quantitative trait loci probabilities for single nucleotide polymorphism with BF $>1$ on a region of chromosome 13 (BTA13) for live-weight (LW), body condition score (BCS), dry matter intake (DMI) and fat and protein corrected milk yield (FPCM). Dashed lines indicate BF 3.1.

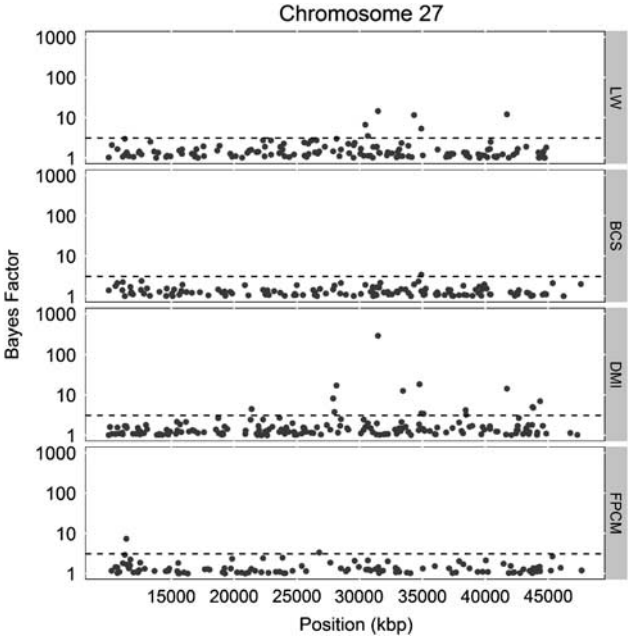


**Figure 6** Bayes factor (BF) corresponding to the posterior quantitative trait loci probabilities for single nucleotide polymorphism with BF $>1$ on a region of chromosome 27 (BTA 27) for live-weight (LW), body condition score (BCS), dry matter intake (DMI), and fat and protein corrected milk yield (FPCM). Dashed lines indicate BF 3.1.

both BCS and DMI. Clustering the 500 genes on gene function (Huang *et al.*, 2009a) showed that three gene groups were overrepresented. The first gene group (enrichment score 5.66) contained 16 genes involved with kallikrein, peptidase and trypsin. The second gene group contained five genes involved with cystan (enrichment score 1.86) and, interestingly, the third group contained 32 olfactory, taste receptor and pheromone receptor genes. The latter suggests

that smell and taste genes play an important role in the feed intake complex. When clustered on functional process, peptidase activity, epidermal growth factor (EGF)-like sensory smell, olfactory and catabolic processes were enriched.

Clustering was also performed for each of the genes associated to each trait separately. When the 227 genes close to the SNPs for BCS were clustered, the group of 16 olfactory genes remained enriched (score 2.40), but one cluster also contained the trypsin-related genes. Clustering the 111 genes in proximity of the SNPs associated with DMI gave one enriched group with zinc finger and ADAM metallopeptidase genes. Functionally, clustering gave a group with EGF-like processes, and a cluster with leptin-type processes, zinc- and iron-binding processes and peptidase. For the trait LW, 11 genes out of 257 were clustered for the kallikrein-related peptidase (enrichment 4.66) and cystatin (enrichment 1.65). Nevertheless, most of these processes have very general functions, and it is difficult to anticipate any direct target for application apart from perhaps the olfactory genes and sensory smell process that might indicate the importance of genetic variation in the genes affecting taste and smell for the feed intake complex of dairy cows.

*Candidate genes affecting DMI, LW and BCS*

Several candidate genes have been investigated in cattle. For example, the leptin gene on BTA4 (Van der Lende *et al.*, 2005; Banos *et al.*, 2008), LEPR on BTA3 (Liefers *et al.*, 2005; Banos *et al.*, 2008), SCD on BTA26 (Macciotta *et al.*, 2008), PGF on BTA10 (Seidenspinner *et al.*, 2011), SPP1 on BTA6 (Sheehy *et al.*, 2009), GH on BTA19 (Mullen *et al.*, 2010) and GHR on BTA20 (Banos *et al.*, 2008; Waters *et al.*, 2011). For all these candidate genes we looked at the groups of SNPs positioned surrounding these genes. However, no strong associations (BF $>$10) were found for the SNPs in the vicinity of these genes. This does not exclude that these genes play a role in the genetic variation for these traits, but the Bayesian variable selection method used here did not select any of these surrounding segregating SNPs. This might be because the effect is explained by many SNPs in LD with the important SNPs within these genes, or the SNPs used here are not in strong LD with functional SNPs. Therefore, for each of the traits new candidate genes were investigated by focussing first on the 22 SNPs that were identified with BF $>$3.1 and that were located in known genes (http://www.ensembl.org, Btau_4.0; April 2010) and second by identifying the genes around the SNPs with an extreme effect (BF $>$50).

For LW, seven SNPs had a BF $>$50, of which three SNPs had a very strong association (BF $>$460) and explained close to 0.5% of the genetic variation. The first SNP on BTA18 (ARS-BFGL-NGS-98028) is located in the syt3 gene. This gene is from a family of genes that are known to affect the exocytosis of insulin that in turn influences plasma insulin levels (Gauthier and Wollheim, 2008). In this gene, a putative QTL for LW was identified before in a beef F2 population (MacNeil and Grosz, 2002). Interestingly, ARS-BFGL-NGS-38895 on BTA2 was also associated with LW, and within 5 kbp of this SNP the gene GRB14 was found. This gene encodes a growth factor receptor-binding protein that interacts with insulin receptors and IGF receptors and may play a role in signalling pathways that regulate growth and metabolism. The second SNP with a strong association with LW was located on BTA7 (Hapmap46473-BTA-21133) and is located in the gene CETN3. The third SNP (ARS-BFGL-BAC-21948) with a very strong association (BF $>$460) was not located in a gene. Of the four other SNPs associated to LW with BF $>$50, only BTA-81701-no-rs was located in anks6 on BTA8.

For BCS, three SNPs with a large effect were located in known genes: BTB-00225371 on BTA5 (BF $>$59) in TMEM117 and, the two SNPs on BTA4 (ARS-BFGL-NGS-33132 and ARS-BFGL-NGS-31561) were mapped to genes Ptprn2 and CREB3L2.

For DMI, three SNPs were mapped on genes: BTB-01074080 was mapped on MEGF10, ARS-BFGL-NGS-108861 was mapped on IDO2 and Hapmap59844-rs29025329 was mapped on N-myc (and STAT) interactor. IDO2 might be an interesting candidate gene. The gene IDO2 plays an important role in tryptophan metabolism. Tryptophan is an essential amino acid that cannot be produced in the body. For that reason, tryptophan is well investigated and linked to feed intake in humans (Wolfe *et al.*, 1997), cattle (Choung and Chamberlain, 1992), mice (Coskun *et al.*, 2006) and pigs (Montgomery *et al.*, 1980). As discussed by Koopmans *et al.* (2006), tryptophan serves as the immediate precursor to serotonergic activity in the brain, and has been implicated in the regulation of many behavioural and physiological processes such as mood, aggression, susceptibility to stress, sleep patterns, but also the regulation of feed intake. Hence, IDO2 on BTA27 might be an ideal candidate gene to affect DMI.

Another set of candidate genes for DMI might be those involved in EGFs and transforming growth factors. The MEGF10 gene, in which a SNP that was associated with DMI was identified, clustered on function with three others genes that were in proximity to the significant SNPs (enrichment score 2.04). The three genes were ADAM2 on BTA27 (within 500 kbp of ARS-BFGL-NGS-108861 that was associated with DMI and within 5 kbp of ARS-BFGL-NGS-29650 that was associated with LW), ADAM3A on BTA27 (within 50 kbp of ARS-BFGL-NGS-29650 that was associated with LW) and NRG1 on BTA27 (close to Hapmap43421-BTA-97718 that was associated with DMI). EGFs are involved in cell recognition and division (Campbell *et al.*, 1990), and functional analysis suggest that these set of genes played a role in lean and obese mice (Fuller *et al.*, 2007).

Overall a few new candidate genes have been identified for the feed intake complex. These candidate genes are linked to insulin, EGF and tryptophan.

## Conclusion

Genome-wide association study of data collected across five experimental herds identified 872 SNPs associated with DMI, BCS, LW and FPCM. The genetic variance in each trait explained by the SNPs on a chromosome was generally proportional to the size of the chromosome. Next to genes involved in some general processes, the olfactory genes and

genes involved with sensory smell processes appeared to be present in the region of 500 kbp of the significant SNP(s). Potential candidate genes, based on function of the gene and location of a significant SNP in a gene, were linked to insulin, EGF and tryptophan.

## Acknowledgements

Ina Hulsegge, Dirkjan Schokker and Mari Smits are acknowledged for their invaluable help with the bioinformatics analysis. This study is part of the RobustMilk project (http://www.robustmilk.eu), which is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708. The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the Commission or its services. Scottish government is recognised for continued funding of the Langhill lines of dairy cattle.

## References

Banos G, Brotherstone S and Coffey MP 2004. Evaluation of body condition score measured throughout lactation as an indicator of fertility in dairy cattle. Journal of Dairy Science 87, 2669–2676.

Banos G, Woolliams JA, Woodward BW, Forbes AB and Coffey MP 2008. Impact of single nucleotide polymorphisms in leptin, leptin receptor, growth hormone receptor, and diacylglycerol acyltransferase (DGAT1) gene loci on milk production, feed, and body energy traits of UK dairy cows. Journal of Dairy Science 91, 3190–3200.

Banos G, Coffey MP, Veerkamp RF, Berry DP and Wall E 2012. Merging and characterising phenotypic data on conventional and rare traits from dairy cattle experimental resources in three countries. Animal 6, 1040–1048.

Beerda B, Ouweltjes W, Sebek LBJ, Windig JJ and Veerkamp RF 2007. Effects of genotype by environment interactions on milk yield, energy balance, and protein balance. Journal of Dairy Science 90, 219–228.

Berry DP, Buckley F, Dillon P, Evans RD, Rath M and Veerkamp RF 2002. Genetic parameters for level and change of body condition score and body weight in dairy cows. Journal of Dairy Science 85, 2030–2039.

Berry DP, Buckley F, Dillon P, Evans RD, Rath M and Veerkamp RF 2003. Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models. Journal of Dairy Science 86, 3704–3717.

Berry DP, Horan B, O'Donovan M, Buckley F, Kennedy E, McEvoy M and Dillon P 2007. Genetics of grass dry matter intake, energy balance, and digestibility in grazing Irish dairy cows. Journal of Dairy Science 90, 4835–4845.

Browning SR and Browning BL 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81, 1084–1097.

Calus MPL and Veerkamp RF 2011. Accuracy of multi-trait genomic selection using different methods. Genetic Selection Evolution 43, 26–40.

Calus MPL, Mulder HA and Bastiaansen JWM 2011. Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genetics Selection Evolution 43, 34–46.

Campbell ID, Baron M, Cooke RM, Dudgeon TJ, Fallon A, Harvey TS and Tappin MJ 1990. Structure–function-relationships in epidermal growth-factor (EGF) and transforming growth factor-alpha (TGF-ALPHA). Biochemical Pharmacology 40, 35–40.

Choung JJ and Chamberlain DG 1992. Protein nutrition of dairy-cows receiving grass-silage diets – effects on silage intake and milk-production of postruminal supplements of casein or soya-protein isolate and the effects of intravenous infusions of a mixture of methionine, phenylalanine and tryptophan. Journal of the Science of Food and Agriculture 58, 307–314.

Coffey MP, Simm G, Hill WG and Brotherstone S 2003. Genetic evaluations of dairy bulls for daughter energy balance profiles using linear type scores and body condition score analyzed using random regression. Journal of Dairy Science 86, 2205–2212.

Coffey MP, Simm G, Oldham JD, Hill WG and Brotherstone S 2004. Genotype and diet effects on energy balance in the first three lactations of dairy cows. Journal of Dairy Science 87, 4318–4326.

Coskun S, Ozer C, Gonul B, Take G and Erdogan D 2006. The effect of repeated tryptophan administration on body weight, food intake, brain lipid peroxidation and serotonin immunoreactivity in mice. Molecular and Cellular Biochemistry 286, 133–138.

Daetwyler HD 2009. Genome-wide evaluation of populations. Wageningen University, Wageningen, The Netherlands.

Dal Zotto R, De Marchi M, Dalvit C, Cassandro M, Gallo L, Carnier P and Bittante G 2007. Heritabilities and genetic correlations of body condition score and calving interval with yield, somatic cell score, and linear type traits in Brown Swiss cattle. Journal of Dairy Science 90, 5737–5743.

de Roos APW, Hayes BJ, Spelman RJ and Goddard ME 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics 179, 1503–1512.

Dillon P, Berry DP, Evans RD, Buckley F and Horan B 2006. Consequences of genetic selection for increased milk production in European seasonal pasture based systems of milk production. Livestock Science 99, 141–158.

Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ and Horvath S 2007. Weighted gene coexpression network analysis strategies applied to mouse weight. Mammalian Genome 18, 463–472.

Gauthier BR and Wollheim CB 2008. Synaptotagmins bind calcium to release insulin. American Journal of Physiology – Endocrinology and Metabolism 295, E1279–E1286.

Gilmour AR, Cullis BR, Welham SJ and Thompson R 2009. ASREML. Program user manual. NSW Agriculture, Orange Agricultural Institute, Orange, NSW, Australia.

Haider S, Ballester B, Smedley D, Zhang JJ, Rice P and Kasprzyk A 2009. BioMart Central Portal – unified access to biological data. Nucleic Acids Research 37, W23–W27.

Hill WG and Robertson A 1986. Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38, 226–231.

Horan B, Dillon P, Berry DP, O'Connor P and Rath M 2005. The effect of strain of Holstein–Friesian, feeding system and parity on lactation curves characteristics of spring-calving dairy cows. Livestock Production Science 95, 231–241.

Huang DW, Sherman BT and Lempicki RA 2009a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4, 44–57.

Huang DW, Sherman BT and Lempicki RA 2009b. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research 37, 1–13.

Kass RE and Raftery AE 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.

Koenen EPC and Veerkamp RF 1998. Genetic covariance functions for live weight, condition score, and dry-matter intake measured at different lactation stages of Holstein Friesian heifers. Livestock Production Science 57, 67–77.

Koenen EPC, Veerkamp RF, Dobbelaar P and De Jong G 2001. Genetic analysis of body condition score of lactating Dutch Holstein and Red-and-White heifers. Journal of Dairy Science 84, 1265–1270.

Koopmans SJ, Guzik AC, van der Meulen J, Dekker R, Kogut J, Kerr BJ and Southern LL 2006. Effects of supplemental L-tryptophan on serotonin, cortisol, intestinal integrity, and behavior in weanling piglets. Journal of Animal Science 84, 963–971.

Liefers SC, te Pas MFW, Veerkamp RF and van der Lende T 2002. Associations between leptin gene polymorphisms and production, live weight, energy balance, feed intake, and fertility in Holstein heifers. Journal of Dairy Science 85, 1633–1638.

Liefers SC, Veerkamp RF, Pas M, Delavaud C, Chilliard Y, Platje M and van der Lende T 2005. Leptin promoter mutations affect leptin levels and performance traits in dairy cows. Animal Genetics 36, 111–118.

Macciotta NPP, Mele M, Conte G, Serra A, Cassandro M, Dal Zotto R, Borlino AC, Pagnacco G and Secchiari P 2008. Association between a polymorphism at the stearoyl CoA desaturase locus and milk production traits in Italian Holsteins. Journal of Dairy Science 91, 3184–3189.

MacNeil MD and Grosz MD 2002. Genome-wide scans for QTL affecting carcass traits in Hereford × composite double backcross populations. Journal of Animal Science 80, 2316–2324.

Meuwissen THE and Goddard ME 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. Genetics Selection Evolution 36, 261–279.

Montgomery GW, Flux DS and Greenway RM 1980. Tryptophan deficiency in pigs – changes in food-intake and plasma-levels of glucose, amino-acids, insulin and growth-hormone. Hormone and Metabolic Research 12, 304–309.

Mullen MP, Berry DP, Howard DJ, Diskin MG, Lynch CO, Berkowicz EW, Magee DA, MacHugh DE and Waters SM 2010. Associations between novel single nucleotide polymorphisms in the *Bos taurus* growth hormone gene and performance traits in Holstein–Friesian dairy cattle. Journal of Dairy Science 93, 5959–5969.

Oikonomou G, Angelopoulou K, Arsenos G, Zygoyiannis D and Banos G 2009. The effects of polymorphisms in the DGAT1, leptin and growth hormone receptor gene loci on body energy, blood metabolic and reproductive traits of Holstein cows. Animal Genetics 40, 10–17.

Petersson KJ, Strandberg E, Gustafsson H and Berglund B 2006. Environmental effects on progesterone profile measures of dairy cow fertility. Animal Reproduction Science 91, 201–214.

Pryce JE, Nielsen BL, Veerkamp RF and Simm G 1999. Genotype and feeding system effects and interactions for health and fertility traits in dairy cattle. Livestock Production Science 57, 193–201.

Seidenspinner T, Tetens J, Habier D, Bennewitz J and Thaller G 2011. The placental growth factor (PGF) – a positional and functional candidate gene influencing calving ease and stillbirth in German dairy cattle. Animal Genetics 42, 22–27.

Sheehy PA, Riley LG, Raadsma HW, Williamson P and Wynn PC 2009. A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. Animal Genetics 40, 492–498.

Sved JA 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theoretical Population Biology 2, 125–141.

Svendsen M, Skipenes P and Mao IL 1994. Genetic correlations in the feed conversion complex of primiparous cows at a recommended and a reduced plane of nutrition. Journal of Animal Science 72, 1441–1449.

Van der Lende T, Te Pas MFW, Veerkamp RF and Liefers SC 2005. Leptin gene polymorphisms and their phenotypic associations. In Vitamins and hormones – advances in research and applications (ed. Gerald Litwack), vol. 71, pp. 373–404. Elsevier Academic press, San Diego, California, USA

Veerkamp RF 1998. Selection for economic efficiency of dairy cattle using information on live weight and feed intake: a review. Journal of Dairy Science 81, 1109–1119.

Veerkamp RF 2002. Feed intake and energy balance in lactating animals. In Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, CD-ROM Session 10-01.

Veerkamp RF, Simm G and Oldham JD 1995. Genotype by environment interactions: experience from Langhill. BSAP Occasional Publication 19, 59–77.

Veerkamp RF, Beerda B and van der Lende T 2003. Effects of genetic selection for milk yield on energy balance, levels of hormones, and metabolites in lactating cattle, and possible links to reduced fertility. Livestock Production Science 83, 257–275.

Veerkamp RF, Emmans GC, Oldham JD and Simm G 1993. Energy and protein utilization of cows of high and low genetic merit for milk solids production on a high and low input diet. Biological basis of sustainable animal production. Proceedings of the 4th Zodiac Symposium, Wageningen, The Netherlands, April.

Veerkamp RF, Oldenbroek JK, van der Gaast HJ and van der Werf JHJ 2000. Genetic correlation between days until start of luteal activity and milk yield, energy balance and live weights. Journal of Dairy Science 83, 577–583.

Verbyla KL, Calus MPL, Mulder HA, de Haas Y and Veerkamp RF 2010. Predicting energy balance for dairy cows using high density SNP information. Journal of Dairy Science 93, 2757–2764.

Wall E, Simm G and Moran D 2010. Developing breeding schemes to assist mitigation of greenhouse gas emissions. Animal 4, 366–376.

Waters SM, McCabe MS, Howard DJ, Giblin L, Magee DA, MacHugh DE and Berry DP 2011. Associations between newly discovered polymorphisms in the *Bos taurus* growth hormone receptor gene and performance traits in Holstein–Friesian dairy cattle. Animal Genetics 42, 39–49.

Wolfe BE, Metzger ED and Stollar C 1997. The effects of dieting on plasma tryptophan concentration and food intake in healthy women. Physiology & Behavior 61, 537–541.