Animal (2012), **6:8**, pp 1206–1215 © The Animal Consortium 2012
doi:10.1017/S1751731112000067

animal

Genome-wide associations for fertility traits in Holstein–Friesian dairy cows using data from experimental research herds in four European countries

D. P. Berry[1†], J. W. M. Bastiaansen[2], R. F. Veerkamp[3], S. Wijga[2], E. Wall[4], B. Berglund[5] and M. P. L. Calus[3]

[1]*Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Co. Cork, Ireland;* [2]*Animal Breeding and Genomics Centre, Wageningen University, 6708 WD Wageningen, The Netherlands;* [3]*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 8200 AB Lelystad, The Netherlands;* [4]*Sustainable Livestock Systems Group, Scottish Agricultural College, EH25 9RG Midlothian, United Kingdom;* [5]*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, PO Box 7023, S-75007 Uppsala, Sweden*

(Received 25 March 2011; Accepted 22 September 2011; First published online 30 January 2012)

Genome-wide association studies for difficult-to-measure traits are generally limited by the sample population 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 traditional measures of fertility, as well as a fertility phenotype derived from milk progesterone profiles. Traditional fertility measures investigated were days to first heat, days to first service, pregnancy rate to first service, number of services and calving interval (CI); post-partum *interval to the commencement of luteal activity (CLA) was derived using routine milk progesterone assays. Phenotypic and genotypic data on 37 590 single nucleotide polymorphisms (SNPs) were available for up to 1570 primiparous cows. Genetic parameters were estimated using linear animal models, and univariate and bivariate genome-wide association analyses were undertaken using Bayesian stochastic search variable selection performed using Gibbs sampling. Heritability estimates of the traditional fertility traits varied from 0.03 to 0.16; the heritability for CLA was 0.13. The posterior quantitative trait locus (QTL) probabilities, across the genome, for the traditional fertility measures were all $<$0.021. Posterior QTL probabilities of 0.060 and 0.045 were observed for CLA on SNPs each on chromosome 2 and chromosome 21, respectively, in the univariate analyses; these probabilities increased when CLA was included in the bivariate analyses with the traditional fertility traits. For example, in the bivariate analysis with CI, the posterior QTL probability of the two aforementioned SNPs were 0.662 and 0.123. Candidate genes in the vicinity of these SNPs are discussed. The results from this study suggest that the power of genome-wide association studies in cattle may be increased by sharing of data and also possibly by using physiological measures of the trait under investigation.*

Keywords: genome-wide association, fertility, cow, progesterone

Implications

Traditional genetic selection for fertility is made difficult by the low heritability of traditional fertility phenotypes. This limitation can be minimised by increasing the heritability of the trait(s) or by using genomic information that explains a large proportion of the genetic variation in fertility. In the present study, both approaches were investigated. Heritability estimates greater than some traditional fertility traits were observed for the *post-partum* commencement of luteal activity quantified using milk progesterone concentration. Specific regions of the genome associated with fertility were also observed, albeit they only explained a small proportion of the total genetic variance in the traits.

Introduction

The well-documented deterioration in reproductive performance in international populations of Holstein–Friesian dairy cattle (Royal *et al.*, 2000) has led to the broadening of international breeding goals to now include functional traits such as fertility (Miglior *et al.*, 2005). However, the low heritability of many traditional fertility traits (Pryce and Veerkamp, 2001; Berry *et al.*, 2003; Wall *et al.*, 2003), coupled with the long time period necessary to acquire some of the phenotypes (e.g. calving interval, CI), imply that selection

† E-mail: donagh.berry@teagasc.ie

for these traits using traditional methods is more difficult than selection, for example, for milk yield traits. Fertility phenotypes defined using information on hormonal profiles are, in general, more heritable (Veerkamp *et al*., 2000; Royal *et al*., 2002; Petersson *et al*., 2007), partly because of a possible reduction in data recording errors or minimising the impact of preferential treatment of some animals such as *post-partum* voluntary waiting periods to service. The generation of such hormonal profiles on sufficient animals for routine use in genetic evaluations is, nonetheless, currently not economically feasible. Use of genomic information in genetic evaluation of animals, however, has the potential to overcome some of the limitations of traditional methods of genetic evaluations.

The recent availability of fast-throughput genotyping platforms with dense genome-wide markers is reducing the cost of genotyping considerably, resulting in accurate phenotypes on large numbers of animals now being the limiting factor in association analyses. Most studies to date that attempted to identify quantitative trait loci (QTLs) in dairy cattle have used specific study designs such as (grand-) daughter designs (Weller *et al*., 1990; Holmberg and Andersson-Eklund, 2006; Höglund *et al*., 2009), and have therefore been generally limited to routinely available phenotypes. We are not aware of any genome-wide association study for traditional measures of fertility in dairy cattle that has been undertaken using actual cow phenotypic data. Furthermore, we are not aware of any study that has attempted to locate genomic regions associated with fertility phenotypes derived from progesterone profiles, most likely because of the expense associated with collecting such data on a sufficiently large sample population. In this study, we collate cow genotypic and phenotypic information from research herds in Ireland, the United Kingdom, the Netherlands and Sweden and relate genetic markers across the genome to the fertility phenotypes using a Bayesian stochastic search variable selection statistical approach.

Material and methods

The data were from Teagasc, Moorepark, Ireland; Scottish Agricultural College, United Kingdom; Wageningen UR Livestock Research, the Netherlands and the Jälla herd of the Swedish University of Agricultural Science. In the present study, these datasets will be referred to as 'Ireland', 'UK', 'the Netherlands' and 'Sweden', respectively. Phenotypic data were available on 2031 Irish, 1018 UK, 725 Dutch and 225 Swedish primiparous Holstein cows. A detailed description of the experimental treatments imposed on the animals in the different countries is provided elsewhere (Veerkamp *et al*., 2000; Horan *et al*., 2005; Petersson *et al*., 2006; Pollot and Coffey, 2008).

Sampling procedures for milk and quantification of the progesterone concentration in milk have previously been described in detail for the data originating from Ireland (Horan *et al*., 2005), the UK (Pollot and Coffey, 2008), the Netherlands (Veerkamp *et al*., 2000) and Sweden (Petersson *et al*., 2006). Milk sampling for progesterone was undertaken on Mondays, Wednesdays and Fridays in both Ireland and the UK, and data between the years 2001 and 2004 in Ireland and between 2003 and 2005 in the UK were included in this study. Between the years 1991 and 1998, inclusive milk sampling for progesterone in the Netherlands was undertaken on Mondays and Fridays and between 2003 and 2004, inclusive milk sampling was undertaken on Mondays and Thursdays. In Sweden, milk sampling for progesterone was undertaken twice weekly between 1989 and 2009. Quantification of milk progesterone concentration across countries was undertaken on raw milk.

Definition of fertility phenotypes

Two distinct types of fertility traits were defined: (1) traditional fertility traits and (2) a fertility trait derived from progesterone profiles. The traditional fertility traits in the present study were days from calving to first observed heat (CFH) or first service (CFS), calving interval (CI), number of services (NS) and pregnancy rate to first service (PRFS). CFH and CFS were restricted to be between 2 and 150 days and between 10 and 200 days, respectively. CI records had to be between 300 and 800 days. Cows were coded as pregnant to first service if they had no second service, were not diagnosed as 'not pregnant' by transrectal ultrasound and if the number of days from first service to next calving was between 265 and 300 days. Because of the experimental treatments imposed, no traditional fertility information other than CFH was available from the Netherlands.

Post-partum interval to the commencement of luteal activity (CLA) was defined as the number of days from calving to the first occurrence of two consecutive test-day records with a milk progesterone concentration of $\geqslant$3 ng/ml (Royal *et al*., 2000). Number of records per trait are summarised in Table 1.

Genotypic data

All animals had phenotypic information on a range of performance traits and were genotyped with the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) containing 54 001 single nucleotide polymorphisms (SNPs). A total of 65 cows with genotype call rates $<$95% were not included. SNPs were included in the association analyses if they fulfilled all of the following quality criteria: (1) GCscore $>$0.20 and GTscore $>$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. χ^2 test statistic $<$600). The GCscore and GTscore are quality measures on the genotype calls from the genotyping assay. Following these edits, 37 590 SNPs from 1956 cows remained. Checks for Mendelian inconsistencies between pedigree and SNP data were first performed for all genotyped parent–offspring pairs and among sibs (Calus *et al*., 2011b). Following the removal of animals that did not pass parentage verification, a total of 1887 cows from Ireland, the UK, the Netherlands and Sweden remained. Chromosome number and positions of the SNPs on the BovineSNP50 were obtained from the UMD3.0 bovine genome assembly from the University of Maryland.

Table 1 *Heritability estimated from the entire dataset as well as the number of records, mean and phenotypic standard deviation for each of the fertility traits within each country*

		Ireland			UK			The Netherlands[†]			Sweden[‡]		
Trait	Heritability	*n*	Mean	s.d.	*n*	Mean	s.d.	*n*	Mean	s.d.	*n*	Mean	s.d.
Traditional fertility													
CFH	0.16 (0.044)	443	53	18.1	1008	57	29	535	56	30			
CFS	0.11 (0.033)	2017	80	24	1013	76	23				193	111	39
CI	0.05 (0.029)	1629	377	31	912	396	59				118	418	54
NS	0.07 (0.027)	2017	1.9	1.1	1013	2.1	1.5				193	2.1	1.3
PRFS	0.03 (0.024)	1711	0.52	0.50	918	0.46	0.50				193	0.43	0.50
Progesterone derived													
CLA	0.13 (0.060)	195	36.1	18.4	70	34.9	20.9	680	36.0	18.7	212	34.8	23.1

CFH = calving to first heat; CFS = calving to first service; CI = calving interval; NS = number of services; PRFS = pregnancy rate to first service; CLA = commencement of luteal activity.
[†]No data available on calving to first service, number of services, pregnancy rate to first service or calving interval for the Netherlands.
[‡]No data available on calving to first heat for Sweden.

Variance component estimation

Genetic and residual (co)variances for the different fertility traits were estimated using animal linear mixed models in ASREML (Gilmour *et al*., 2009). Fixed effects included in the model for all traits were the contemporary groups of country-experimental treatment-year and country-year-season of calving. For the estimation of variance components for PRFS, days since calving at the time of insemination was also included as a class effect (<30 days, 31 to 60 days, 61 to 90 days, 91 to 120 days, >120 days) in the model to account for differing pregnancy rate by days post calving (Berry *et al*., 2011). Genetic and residual covariances were estimated using 15 bivariate analyses. Pedigree information on each genotyped animal was traced back to at least four generations where available; a total of 7619 animals were included in the pedigree file.

Genome-wide association analysis

The genome-wide association analysis was conducted using a model that estimates effects for all SNPs simultaneously. The general model used to estimate SNP effects was

$$y_{ij} = \mu_j + \textit{fixed effects}_{ij} + \textit{animal}_{ij} + \sum_{k=1}^{nloc}\sum_{l=1}^{2} \textit{SNP}_{ijkl} + e_{ij}$$

where y_{ij} is the phenotypic record of animal i, μ_j is the overall mean for trait j, $\textit{fixed effects}_{ij}$ are the fixed effects pertinent to animal i for trait j, $\textit{animal}_{ij}$ is the random polygenic effect of animal i for trait j, $\textit{SNP}_{ijkl}$ is a random effect for allele l on trait j at locus k of animal i and e_{ij} is a random residual for animal i. The fixed effects included in the models were the same as those used in the estimation of the variance components. This model was used both for univariate and bivariate analyses.

The polygenic effect was fitted with a numerator pedigree relationship matrix. SNP effects, denoted as $\textit{SNP}_{ijkl}$, were estimated as $q_{ijkl} \times v_{jk}$, as described by Meuwissen and Goddard (2004), where q_{ijkl} is the size of the effect of allele l at locus k and v_{jk} is a scaling factor in the direction vector for locus k that scales the effect at locus k for trait j. The variance of the direction vector $v_{.k}$ is denoted as **V..** The prior distribution for **V** in this case was

$$p(\mathbf{V}..) = \chi^{-2}(\mathbf{S}_0(..), 10)$$

where $\mathbf{S}_{0(no)}$ is chosen such that it reflects the total genetic (co)variance between traits n and o, divided by the total number of SNPs. **V..** was sampled from the following conditional m variate-inverted Wishart distribution with $(nloc + 10)$ degrees of freedom:

$$\mathbf{V}..|\mathbf{v}..,I. \sim IW_m(\mathbf{S}_0(..) + \mathbf{SZ}_{(..)}, nloc - m - 1 + 10)$$

where $\mathbf{SZ}_{(..)} = \sum_{k=1}^{nloc} v'_{.k} v_{.k}$, $nloc$ = number of evaluated marker loci and 10 is the number of degrees of freedom for the prior distribution.

In the model, a QTL indicator (I_k) was sampled for each locus k, which either had the value of 0 or 1. $\mathbf{S}_{0(..)}$ is chosen such that it reflected the total genetic (co)variances of traits n and o, divided by the total number of expected QTLs instead of the number of SNPs. **V..** was sampled from an inverted Wishart distribution as follows:

$$\mathbf{SZ}_{(..)} = \sum_{k=1}^{nloc} v'_{.k} v_{.k} (I_k + (1 - I_k) \times 100)$$

where the QTL indicator I_k was sampled from

$$I_k|v_{.k}, \mathbf{V} \sim \textit{Bernoulli}\left[\frac{\phi(v_{.k};\, \mathbf{0},\, \mathbf{V}) \times p_k}{\phi(v_{.k};\, \mathbf{0},\, \mathbf{V}) \times p_k + \phi\left(v_{.k};\, \mathbf{0},\, \frac{\mathbf{V}}{100}\right) \times (1 - p_k)}\right]$$

where p_k is the prior QTL probability, that is, the prior probability that I_k is equal to 1. Prior QTL probabilities used in the analyses reflected the prior assumption that 100 QTL underlie each of the traits. The model is referred to as Bayesian Stochastic Search Variable Selection (BSSVS; e.g. Verbyla *et al*., 2009).

The residual variance was assumed to be normally distributed $N(\mathbf{0}, \mathbf{R})$, where $\mathbf{R}$ is the $m \times m$ residual covariance matrix and m is the number of traits (1 or 2). The polygenic variance was assumed to be normally distributed $N(\mathbf{0}, \mathbf{A} \otimes \mathbf{G_A})$, where $\mathbf{A}$ is the numerator relationship matrix and $\mathbf{G_A}$ is the $m \times m$ polygenic covariance matrix. Matrices $\mathbf{R}$ and $\mathbf{G_A}$ were both sampled in the Gibbs sampler from an inverted Wishart distribution, with a uniform prior distribution. The BSSVS model was performed using Gibbs sampling. All univariate models were run for 50 000 cycles, discarding the initial 10 000 as burn-in. All bivariate models were run for 100 000 cycles, discarding the initial 20 000 as burn-in. To limit the required computational effort, only one Markov chain Monte Carlo chain was run for all models, except for those including CLA. For those, four chains were run for all univariate models and three chains were run for all bivariate models. For CLA, results were averaged across chains.

The BSSVS model yields posterior QTL probabilities for each SNP locus that are calculated as the average of the QTL indicator I_k across all cycles after the burn-in. Bayes factors (Kass and Raftery, 1995) were used to assess the posterior QTL probabilities for all SNP loci. The Bayes Factor, B_{12}, was calculated as follows. Let $\boldsymbol{y}$ be the data, H_1 is the hypothesis that the marker is linked to a QTL and H_2 is the hypothesis that the marker is not linked to a QTL. Then $Pr(H_1)$ is the prior probability of the first hypothesis H_1 and $Pr(H_2)$ is the prior probability of the alternative hypothesis, H_2. Similarly, $Pr(H_1|\boldsymbol{y})$ is the posterior probability of H_1 and $Pr(H_2|\boldsymbol{y})$ is the posterior probability of H_2. Using Bayes theorem, a Bayes factor comparing hypotheses H_1 and H_2, for a single SNP, is

$$B_{12} = \frac{Pr(H_1|y)}{Pr(H_2|y)} \Big/ \frac{Pr(H_1)}{Pr(H_2)}$$

As this is the ratio of the posterior odds to the prior odds, it can be expressed as

$$B_{12} = \frac{Pr(H_1|y)}{1 - Pr(H_1|y)} \Big/ \frac{Pr(H_1)}{1 - Pr(H_1)}$$

Bayes factors >3.1 indicate 'substantial evidence' in favour of the first hypothesis, that is, that the SNP is associated with a QTL. Furthermore, Bayes factors >10.1 were assumed to indicate 'strong evidence' for association with a QTL, and Bayes factors >30.1 were assumed to indicate 'very strong evidence' (Jeffreys, 1961).

Results

Summary statistics for each of the fertility traits within each country are in Table 1, together with heritability estimates across all data. Heritability estimates for the traditional fertility traits varied from 0.03 (PRFS) to 0.16 (interval from CFH). The heritability of CLA was 0.13.

Genetic correlations among traits were generally stronger than their respective phenotypic correlation and were, in general, of the same sign (Table 2). The interval traits (i.e. CFH, CFS, CI and CLA) were all strongly genetically correlated (0.37 to 0.99) with each other; the respective phenotypic correlations varied from 0.12 to 0.63. Genetically longer CIs were associated with poor PRFS and more services (Table 2).

Genome-wide associations

The posterior QTL probabilities at each SNP were <0.021, with the exception of CLA (Figures 1 and 2). The maximum posterior probability for CFH was 0.020 for SNP ARS-BFGL-NGS-104801 located on BTA 8; the next highest posterior probability for CFH was 0.014 for SNP Hapmap50203-BTA-107053 located on BTA 4. The maximum posterior probability for CFS was 0.010 for SNP BTB-01412031 located on BTA 20, whereas the next highest posterior probability for CFS was 0.009 for SNP ARS-BFGL-NGS-13749 located on BTA 6. The maximum posterior probability for NS was 0.006 for Hapmap39421-BTA-65907 located on BTA 4, whereas the maximum posterior probability for CI was 0.005 for ARS-BFGL-NGS-14867 located on BTA 19. The maximum posterior probability for PRFS was 0.005 for BTB-01390173 located on BTA 13. The proportion of the genetic variance explained by the top two SNPs in each of the traits varied from 0.002% to 0.012%.

Posterior QTL probabilities of >0.04 were observed for CLA (Table 3) on BTA 2 (BTA-49769-no-rs; posterior probability of 0.060) and BTA 21 (BTA-12468-no-rs; posterior probability of 0.045). The SNP on BTA 2 explained 0.51% of the genetic variance in CLA, whereas the SNP on BTA 21 explained 0.35% of the genetic variance in CLA. The Bayes

Table 2 *Genetic (above diagonal; standard errors in parenthesis) and phenotypic† (below diagonal) correlations among the fertility traits*

Trait	CFH	CFS	CI	NS	PRFS	CLA
CFH		0.99 (0.12)	0.95 (0.22)	0.05 (0.27)	−0.13 (0.38)	0.52 (0.29)
CFS	0.63		0.97 (0.18)	−0.14 (0.25)	−0.13 (0.38)	0.37 (0.34)
CI	0.25	0.42		0.66 (0.17)	−0.92 (0.20)	0.87 (0.33)
NS	0.02	−0.09	0.75		−0.90 (0.16)	0.10 (0.41)
PRFS	−0.12	0.03	−0.56	−0.72		−0.29 (0.52)
CLA	0.29	0.12	0.17	0.01	−0.03	

CFH = calving to first heat; CFS = calving to first service; CI = calving interval; NS = number of services; PRFS = pregnancy rate to first service; CLA = commencement of luteal activity.
†Standard errors of the phenotypic correlations were $\leqslant 0.052$.

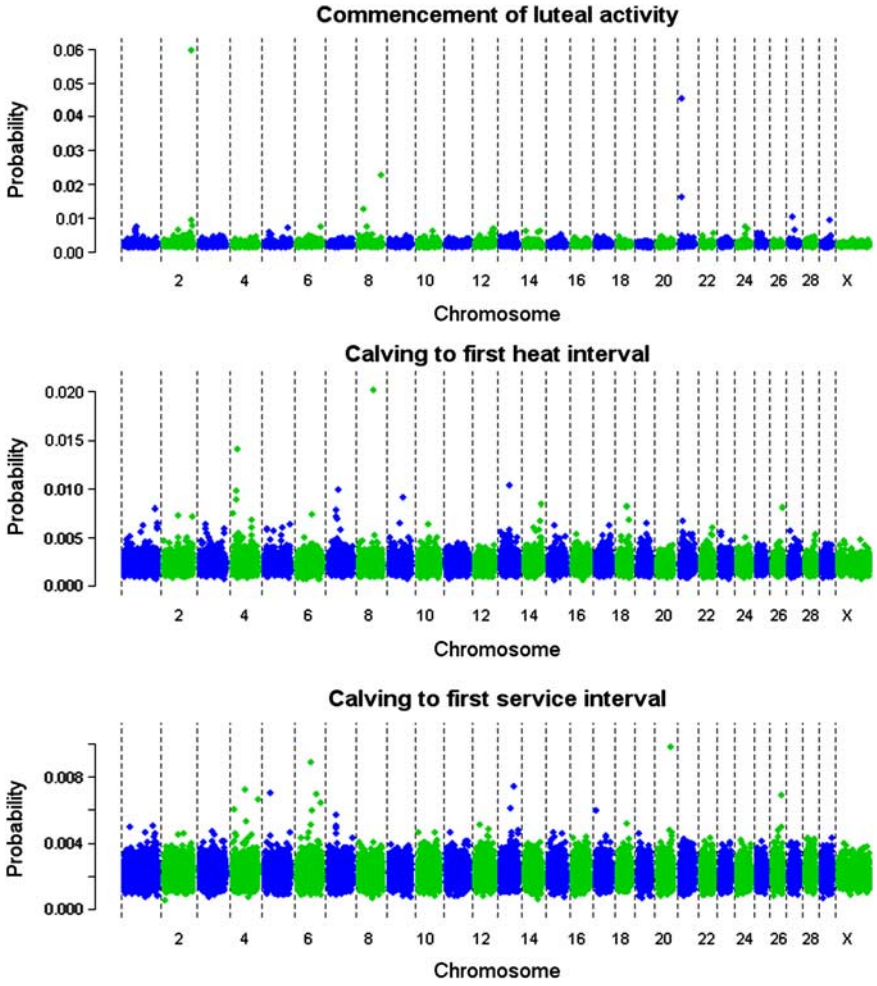

Figure 1 Posterior quantitative trait locus probabilities, from top to bottom, for *post-partum* commencement of luteal activity, calving to first heat interval and calving to first service interval for each of the autosomes from left (BTA 1) to right (BTA 29) and the X-chromosome at the far right.

factors of BTA-49769-no-rs and BTA-12468-no-rs were 24 and 18, respectively, suggesting strong evidence of a QTL.

Figure 3 illustrates the QTL probabilities at each SNP for the five bivariate analyses that included CLA. The posterior probabilities of the 10 most significant SNPs from the univariate analysis of CLA when included in each of the bivariate analyses are detailed in Table 3. The large posterior QTL probabilities for CLA on BTA 2 and BTA 21 remained on the same SNPs, but they were larger compared with the univariate analysis of CLA. The posterior probability of 0.060 for CLA at SNP BTA-49769-no-rs was increased to 0.094, 0.121, 0.162, 0.662 and 0.162 when CLA was included in a bivariate analysis with CFH, CFS, NS, CI and PRFS, respectively (Figure 3 and Table 3). The posterior probability of 0.045 for CLA at SNP BTA-12468-no-rs on BTA 21 was increased to 0.052, 0.152, 0.072, 0.123 and 0.135 when CLA was included in a bivariate analysis with CFH, CFS, NS, CI and PRFS, respectively (Figure 3 and Table 3). The Bayes factor for BTA-49769-no-rs and BTA-12468-no-rs when included in the bivariate analysis with CI was 880 and 52, respectively (Figure 4), indicating a very strong support of a QTL.

Table 4 summarises the correlations between SNP effects and genomic breeding values of animals from the pairwise bivariate analyses. The correlations between SNP effects were similar to the respective correlations between animal genomic breeding values. Correlations were also similar to the respective correlations estimated using traditional linear models using the numerator relationship matrix based on recorded pedigree (Table 2).

Discussion

The success of genome-wide association studies is a function of, amongst others, the heritability of the trait under investigation and the number of phenotypic records on that trait (Daetwyler *et al.*, 2008), as well as the genetic architecture of the trait (i.e. influenced by a large or small number of genes). Large population sizes for genome-wide association analysis are required for lower heritability traits such as traditional measures of fertility (Pryce and Veerkamp, 2001; Berry *et al.*, 2003; Wall *et al.*, 2003); this can be somewhat overcome by using fertility-estimated breeding values on genotyped sires as the phenotype under investigation; nonetheless,

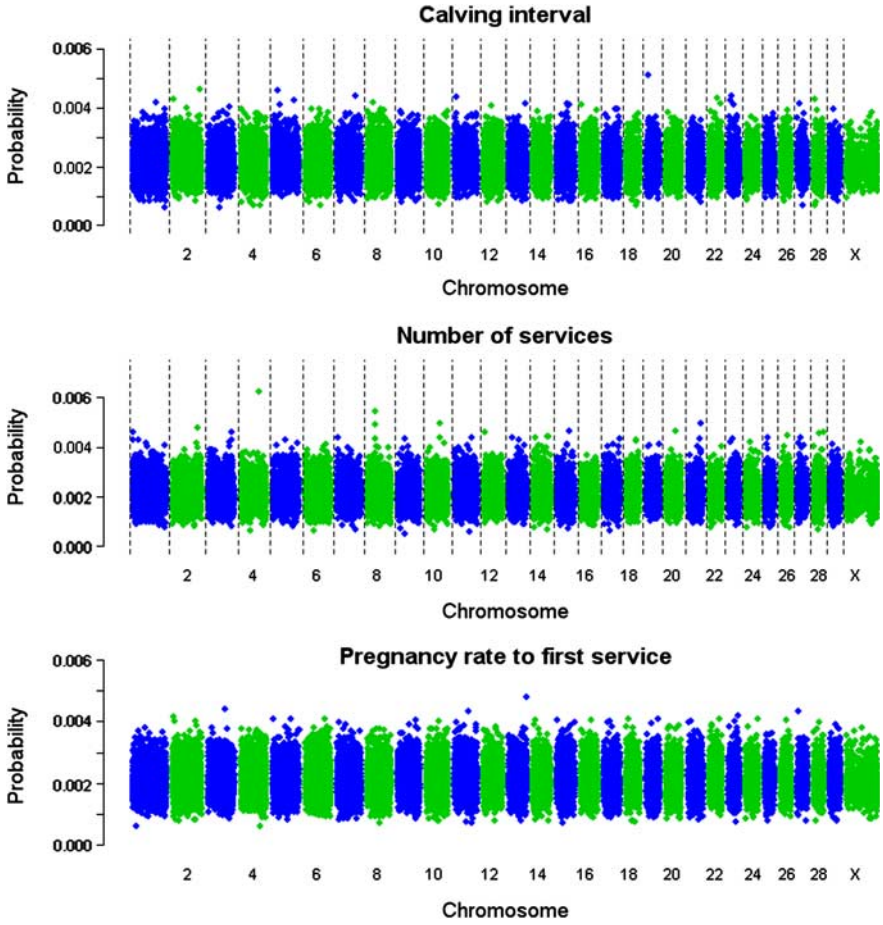

Figure 2 Posterior quantitative trait locus probabilities, from top to bottom, for calving interval, number of services and pregnancy rate to first service for each of the autosomes from left (BTA 1) to right (BTA 29) and the X-chromosome at the far right.

Table 3 *Ten SNPs with greatest posterior probabilities for associations with CLA from the univariate analysis, as well as their respective posterior probabilities when included in each of the five bivariate analyses. Also included is the chromosome (BTA) and position on the chromosomes for each SNP*

				Bivariate				
SNP name	BTA	Position	Univariate	CFH	CFS	CI	NS	PRFS
BTA-49769-no-rs	2	130141723	0.060	0.094	0.121	0.662	0.162	0.162
BTA-12468-no-rs	21	9375095	0.045	0.052	0.152	0.123	0.072	0.135
BFGL-NGS-118553	8	108035534	0.023	0.018	0.021	0.038	0.037	0.031
Hapmap42495-BTA-28432	21	9412024	0.016	0.030	0.034	0.012	0.026	0.031
ARS-BFGL-NGS-52642	8	25958375	0.013	0.037	0.010	0.067	0.019	0.013
ARS-BFGL-NGS-10959	27	12900677	0.010	0.009	0.013	0.018	0.024	0.014
ARS-BFGL-NGS-29244	29	36711948	0.010	0.024	0.010	0.016	0.016	0.013
ARS-BFGL-NGS-33709	2	130187033	0.010	0.005	0.010	0.008	0.009	0.012
BFGL-NGS-118362	2	134636106	0.008	0.006	0.015	0.007	0.009	0.004
BFGL-NGS-117704	6	113166606	0.008	0.008	0.004	0.008	0.009	0.011

SNPs = single nucleotide polymorphisms; CFH = calving to first heat; CFS = calving to first service; CI = calving interval; NS = number of services; PRFS = pregnancy rate to first service; CLA = commencement of luteal activity.

phenotypes on a large number of cows are still required to generate accurate sire-estimated breeding values. The low heritability of fertility traits can be due to a multitude of factors including: (1) the trait is truly under poor additive genetic control relative to non-additive and non-genetic factors; (2) the statistical model used to estimate the variance components is too simplistic relative to reality; (3) recording errors adding to random noise; and (4) preferential treatment that cannot be fully accounted for in the statistical model. The use of hormonal profiles, related to fertility performance, can aid in

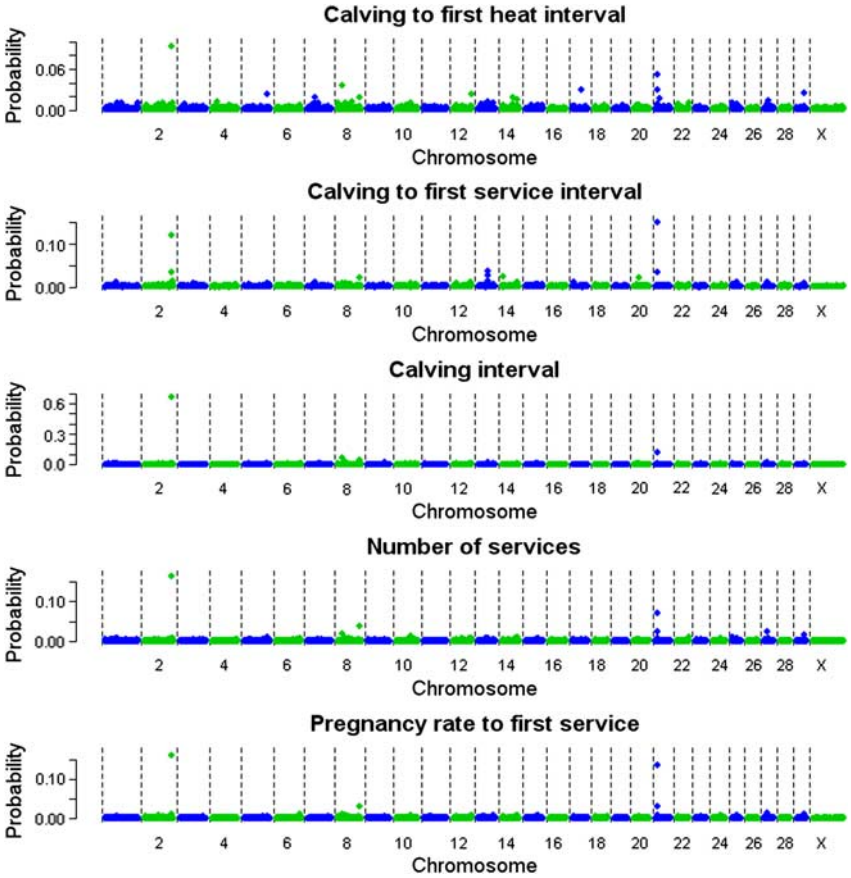

Figure 3 Posterior quantitative trait locus probabilities from the bivariate analyses with commencement of luteal activity, from top to bottom, for calving to first heat interval, calving to first service interval, calving interval, number of services and pregnancy rate to first service for each of the autosomes from left (BTA 1) to right (BTA 29) and the X-chromosome at the far right.

circumventing some of the limitations of traditional measures, and previous research has shown that heritability estimates for fertility traits derived from progesterone profiles tend to be greater than traditional measures of fertility (Veerkamp *et al.*, 2000; Royal *et al.*, 2002; Petersson *et al.*, 2007). Nonetheless, the quantity of data on progesterone profiles at an individual animal level is limited mainly because of the expense of collecting these data. The objective of this study, therefore, was to collate fertility information, including that derived from routine progesterone assays, on Holstein–Friesian cows from multiple countries and attempt to locate genomic regions associated with these traits.

Phenotypic performance

Mean phenotypic fertility performance in the present study was similar to previous reports from Ireland (Horan *et al.*, 2005), the UK (Pollot and Coffey, 2008), the Netherlands (Veerkamp *et al.*, 2000) and Sweden (Petersson *et al.*, 2006), all of which included some of the data used in the present study. However, the mean CLA was longer in the present study compared with the aforementioned studies, but this is more than likely attributable to only first parity animals being included in the present study unlike the other studies, with the exception of Veerkamp *et al.* (2000), which included pluriparae as well. Petersson *et al.* (2006) reported longer CLA in primiparae compared to pluriparae.

Genetic parameters

The low heritability of the traditional fertility traits in the present study corroborates the many previous studies on genetic parameters for traditional fertility measures (Pryce and Veerkamp, 2001; Berry *et al.*, 2003; Wall *et al.*, 2003). The heritability estimates of CFH and CFS, however, are slightly larger than that generally observed from field data, which may be attributable to an expected greater accuracy of recording in experimental herds. The heritability estimate for CLA in the present study (0.13) is slightly lower than estimates of 0.16 reported elsewhere (Veerkamp *et al.*, 2000; Royal *et al.*, 2002; Petersson *et al.*, 2007), although some degree of similarity is expected as the data used in the present study was also included in some previous estimates (Veerkamp *et al.*, 2000).

The phenotypic correlations among the different fertility traits are in general agreement with previous studies. Using data independent of the present study, Royal *et al.* (2002) reported identical correlations to those in the present study of 0.12 and −0.03 between CLA and both CFS interval and

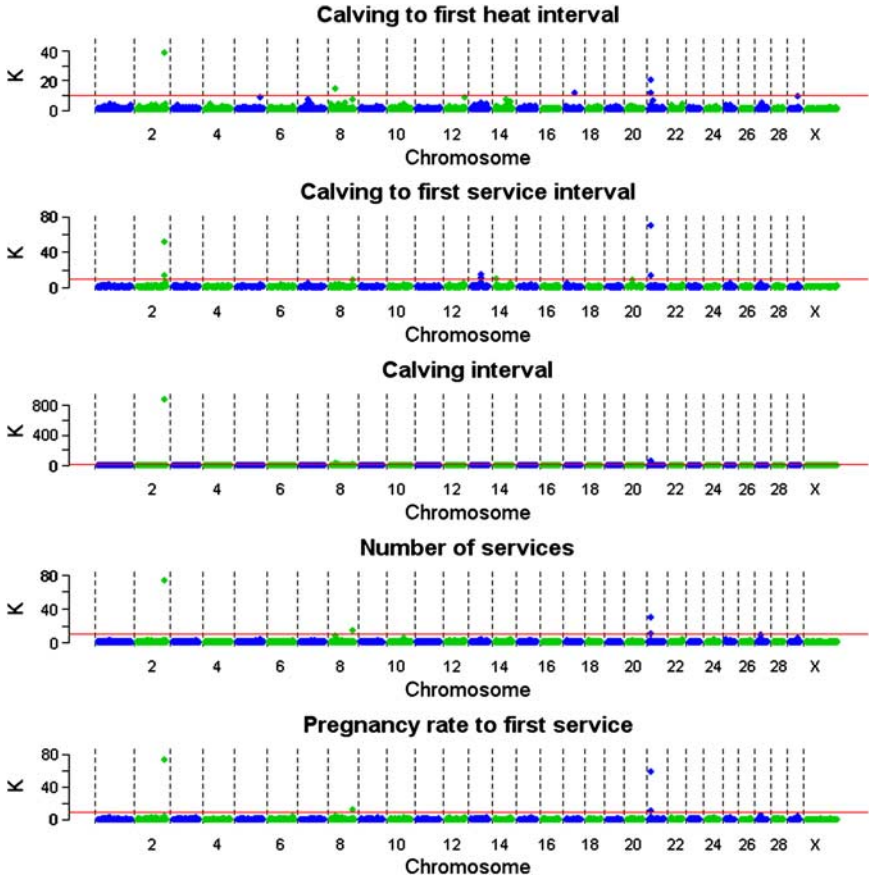

Figure 4 Bayes factor from the bivariate analyses with commencement of luteal activity, from top to bottom, for calving to first heat interval, calving to first service interval, calving interval, number of services and pregnancy rate to first service for each of the autosomes from left (BTA 1) to right (BTA 29) and the X-chromosome at the far right. Horizontal line represents a Bayes factor of 10.

Table 4 *Correlations between estimated SNP effects (above diagonal) and animal genomic breeding values (below diagonal) from the pairwise bivariate analyses*

Trait	CFH	CFS	CI	NS	PRFS	CLA
CFH		0.99	0.99	0.18	−0.45	0.85
CFS	0.99		0.99	0.12	−0.43	0.84
CI	0.99	0.99		0.88	−0.96	0.99
NS	0.11	0.12	0.98		−0.98	0.37
PRFS	−0.43	−0.45	−0.98	−0.99		−0.80
CLA	0.86	0.78	0.99	0.22	−0.74	

SNPs = single nucleotide polymorphisms; CFH = calving to first heat; CFS = calving to first service; CI = calving interval; NS = number of services; PRFS = pregnancy rate to first service; CLA = commencement of luteal activity.

PRFS, respectively. Genetic correlations among the traditional fertility variables were generally similar to those estimated on primiparae in larger datasets (Wall *et al.*, 2003). Nonetheless, the genetic correlations between the traditional fertility traits and CLA were different from those reported elsewhere, albeit the standard errors in the present study and in most of the other studies were large, attributable to a combination of low heritability of the fertility traits and the relatively small datasets. Royal *et al.* (2002) reported a correlation for CLA of −0.03 (s.e. = 0.27) and 0.49 (s.e. = 0.47) with CFS and PRFS, respectively, which are of opposite sign to the genetic correlations reported in the present study. Nonetheless, the genetic correlations in the present study were substantiated by the correlations between both the SNP effects and the genomic breeding values estimated in the bivariate genome-wide association analysis, where the covariances assumed between the traits in the analyses were estimated separately within the analyses themselves.

Considerable genetic variation was present for the different fertility traits evaluated in the present study, indicating that substantial genetic improvement can be achieved if the accuracy of identifying genetically superior animals is high. Achieving high accuracy of selection can be achieved through extensive phenotyping (of relatives and) of animals or through the use of genomics to augment the information from relatives and the animal itself.

Genome-wide associations

The model used to estimate the genome-wide associations in the present study is a model that is used to estimate genomic breeding values in genomic selection. Its performance for this purpose has been presented elsewhere for

univariate applications (Calus *et al*., 2008), and for univariate and bivariate applications compared with other models, both for genomic prediction (Calus and Veerkamp, 2011) and QTL mapping (Calus *et al*., 2011a). By estimating all SNP effects simultaneously, the rate of false positives is controlled and estimated SNP effects are less prone to overestimation compared with models where effects of only one SNP at the time are estimated. The dataset used in the present study could be considered relatively small for the detection of SNPs associated with low heritability traits. However, clear associations were identified with at least two SNPs. The association analysis for CLA was also undertaken with each country's data removed from the analysis, with the exception of the data from the Netherlands, which would have resulted in a large loss of data. The two SNPs BTA-12468-no-rs and BTA-49769-no-rs remained the most strongly associated SNPs with CLA, although the posterior probabilities were reduced to 0.007 (Swedish data removed) to 0.054 (Irish data removed) for BTA-49769-no-rs and to 0.018 (UK data removed) to 0.045 (Irish data removed) for BTA-12468-no-rs.

Although relatively large Bayes factors were obtained for CLA at two SNPs in the univariate analysis, these factors increased when CLA was included in a bivariate analysis. The degree of increase was greatest for CI and PRFS. Part of this increase may be attributable to additional information from the second trait, especially on genotyped animals that had no phenotype for CLA. Phenotypes for either of the traditional fertility traits were available on 1570 animals, whereas phenotypes for CLA were available on only 898 animals.

The BTA-49769-no-rs SNP that had a Bayes factor of 880 when estimated in the bivariate analysis with CI is a 'decisive' indication (Jeffreys, 1961) of a QTL at this location. The SNP is in an intergenic region on BTA 2 and a search for genes within 500 kb from this SNP revealed 16 genes at distances ranging from 9 to 462 kb; one gene was within 500 kb of BTA-12468-no-rs at a distance of 433 kb downstream. All genes were screened using Entrez Gene for functional annotation (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene) that would relate the genes to fertility traits in either cattle or humans.

The only gene near BTA-12468-no-rs was *Bos taurus* arrestin domain containing 4 (ARRDC4), which has a function in neurological processes. ARRDC4 was relatively distant from BTA-12468-no-rs compared with the average distance at which linkage disequilibrium (LD) decays in Holstein cattle (McKay *et al*., 2007).

Of the 16 genes near BTA-49769-no-rs, three have functions that relate to fertility. The first gene, 242 kb upstream from BTA-49769-no-rs, was '*Bos taurus* inhibitor of DNA binding 3, dominant negative helix-loop-helix protein' (ID3), which controls the growth of vascular cells through signalling that is induced by oestrogen (Felty and Porther, 2008). However, studies that have attempted to relate mutations in the ID3 gene to fertility in cattle were not found. The second gene, located 462 kb upstream from BTA-49769-no-rs, was *Bos taurus* fucosidase, alpha-L-1, tissue (FUCA1), which is broadly distributed over the membrane systems of human sperm and has a role in the intimate species signature interactions between the sperm and oocyte (Venditti *et al*., 2007). Even though FUCA1 was found to play a clear role in fertility, the evidence reported suggests a role in male rather than female fertility; although both are correlated (Berry *et al*., 2011). The third gene, located 199 kb upstream from BTA-49769-no-rs, was E2F transcription factor 2, which plays a crucial role in the control of cell cycle and action of tumour suppressor proteins in humans.

Several other studies have attempted to locate genomic regions associated with reproductive performance in cattle (Holmberg and Andersson-Eklund, 2006; Huang *et al*., 2010; Pryce *et al*., 2010; Sahana *et al*., 2010; Schulman *et al*., 2011). Although not all studies have documented significant associations with different measures of reproductive performance on either BTA 2 or BTA 21 (Holmberg and Andersson-Eklund, 2006; Pryce *et al*., 2010), significant associations with SNPs on BTA 2 and BTA 21 have been reported in some studies (Huang *et al*., 2010; Sahana *et al*., 2010; Schulman *et al*., 2011). Of particular note were the two SNPs, ARS-BFGL-NGS-36151 and BTB-00117780, reported by Huang *et al*. (2010) on BTA 2 to be associated with fertilisation rate in a Holstein population independent to that used in the present study. On the basis of the UMD3.0 bovine genome assembly from the University of Maryland and the SNPs retained in the present study for inclusion in the association analysis, both SNPs reported by Huang *et al*. (2010) flanked the BTA-49769-no-rs SNP shown to be associated with CLA in the present study. The ARS-BFGL-NGS-36151 SNP was the fifth SNP upstream (214 kb) from BTA-49769-no-rs, whereas BTB-00117780 was the 19th SNP downstream (1.25 Mb) from BTA-49769-no-rs. LD (r^2) between all three SNPs in the population of animals used in the present study was low (0.04 to 0.07). However, it is still possible that all three SNPs could be individually tagging the same causative mutation(s) in the region, and that the LD between associated SNP in both studies is simply low because of differences in local LD patterns between both populations.

Conclusion

The low heritability of the traditional fertility traits agree, albeit they were slightly greater in the present study, with the general consensus in the literature that traditional fertility measures are lowly heritable. Two regions of the genome with a relatively high probability of harbouring a mutation associated with CLA were identified, with the probability increasing when the genome-wide association was undertaken in a bivariate analysis; nonetheless, the proportion of the genetic variance explained by these regions was small.

Acknowledgements

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.

References

Berry DP, Buckley F, Dillon P, Evans RD, Rath M and Veerkamp RF 2003. Genetic relationships among body condition score, body weight, milk yield and fertility in dairy cows. Journal of Dairy Science 86, 2193–2204.

Berry DP, Evans RD and Mc Parland S 2011. Evaluation of bull fertility in dairy and beef cattle using cow field data. Theriogenology 75, 172–181.

Calus MPL, Meuwissen THE, de Roos APW and Veerkamp RF 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178, 553–561.

Calus MPL, Mulder HA and Veerkamp RF 2011a. Estimating genomic breeding values and detecting QTL using univariate and bivariate models. BMC Proceedings 5 (suppl. 3), S5.

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

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

Daetwyler H, Villanueva B and Woolliams JA 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 3, 1–8.

Felty Q and Porther N 2008. Estrogen-induced redox sensitive Id3 signaling controls the growth of vascular cells. Atherosclerosis 198, 12–21.

Gilmour AR, Cullis BR, Welham SJ and Thompson R 2009. ASREML reference manual. New South Wales Agriculture, Orange Agricultural Institute, Orange, NSW, Australia.

Höglund JK, Buitenhuis AJ, Guldbrandtsen B, Su G, Thomsen B and Lund MS 2009. Overlapping chromosomal regions for fertility traits and production traits in the Danish Holstein population. Journal of Dairy Science 92, 5712–5719.

Holmberg M and Andersson-Eklund L 2006. Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. Journal of Dairy Science 89, 3664–3671.

Horan B, Mee JF, O'Connor P, Rath M and Dillon P 2005. The effect of strain of Holstein–Friesian cow and feeding system on postpartum ovarian function, animal production and conception rate to first service. Theriogenology 63, 950–961.

Huang W, Kirkpatrick BW, Rosa GJM and Khatib K 2010. A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. Animal Genetics 41, 570–578.

Jeffreys H 1961. The theory of probability, 3rd edition. Oxford University Press, United Kindom.

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

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias E, Neto K, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF and Moore SS 2007. Whole genome linkage disequilibrium maps in cattle. BMC Genetics 8, 74.

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.

Miglior F, Muir BL and Van Doormaal BJ 2005. Selection indices in Holstein cattle of various countries. Journal of Dairy Science 88, 1255–1263.

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

Petersson K-J, Berglund B, Strandberg E, Gustafsson H, Flint APF, Wooliams JA and Royal MD 2007. Genetic analysis of postpartum measures of luteal activity in dairy cows. Journal of Dairy Science 90, 427–434.

Pollot GE and Coffey MP 2008. The effect of genetic merit and production system on dairy cow fertility, measured using progesterone profiles and on-farm recording. Journal of Dairy Science 91, 3649–3660.

Pryce JE and Veerkamp RF 2001. The incorporation of fertility indices in genetic improvement programmes. In Fertility in the high-producing dairy cow (ed. M.G. Diskin), 223–236. British Society of Animal Science Occasional Publication No. 26, Edinburgh, Scotland.

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME and Hayes BJ 2010. A validation genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. Journal of Dairy Science 93, 3331–3345.

Royal MD, Darwash AO, Flint APF, Webb R, Woolliams JA and Lamming GE 2000. Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility. Animal Science 70, 487–501.

Royal MD, Flint APF and Woolliams JA 2002. Genetic and phenotypic relationships among endocrine and traditional fertility traits and production traits in Holstein–Friesian dairy cows. Journal of Dairy Science 85, 958–967.

Sahana G, Guldbrandtsen B, Bendixen C and Lund MS 2010. Genomic-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. Animal Genetics 41, 579–588.

Schulman NF, Sahana G, Iso-Touro T, McKay SD, Schnabel RD, Lund MS, Taylor JF, Virta J and Vikki JH 2011. Mapping of fertility traits in finnish Ayrshire by genomie-wide association analysis. Animal Genetics, doi:10.1111/j.1365-2052.2010.02149.x.

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

Venditti JJ, Donigan KA and Bean BS 2007. Crypticity and functional distribution of the membrane associated alpha-L-fucosidase of human sperm. Molecular Reproduction and Development 74, 758–766.

Verbyla KL, Hayes BJ, Bowman PJ and Goddard ME 2009. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. Genetics Research 91, 307–311.

Wall E, Brotherstone S, Woolliams JA, Banos G and Coffey MP 2003. Genetic evaluation of fertility using direct and correlated traits. Journal of Dairy Science 86, 4093–4102.

Weller JI, Kashi Y and Soller M 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy-cattle. Journal of Dairy Science 73, 2525–2537.