Application of BLUP in prediction of breeding values and estimation of SNP effects in dairy cattle

Sausan Aziz Ahmad

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

School of Biosciences
The University of Nottingham
Division of Animal Sciences
Sutton Bonington Campus
Loughborough
Leicestershire
LE12 5RD

December 2010

Abstract

Genetic selection for higher milk production has unavoidably resulted in a decline in fertility of dairy cattle in the UK due to the antagonistic correlation existing between fertility and milk yield. This trend in fertility has necessitated broadening the breeding programme to include fertility traits. However, the heritability of fertility traits currently used in the UK are of low heritability (h^2<0.05). Therefore, one approach is to use molecular markers in order to identify animals with the highest genetic merit for fertility.

The first aim of this project was therefore to test single nucleotide polymorphisms (SNPs) in candidate genes for their possible associations with fertility PTAs. A group of 408 dairy cows and bulls were genotyped at different loci (*GnRH receptor, FSH receptor, LH receptor, activin receptor*, and *neuropeptide Y receptor Y2 (NPYRY2))*. Seven SNPs were identified in the *GnRH receptor* gene. These were in the promoter and coding regions at positions -331, -108, 206, 260, 341, 383 and 410 relative to the translation start site. Two groups of SNPs were found to be in linkage disequilibrium, the SNPs at positions 206 and 383 being inherited together, as were those at -108, 260, 341 and 410. An association study between these SNPs and fertility PTAs revealed that the -108 SNP group was associated with favourable effects on fertility, reducing PTA for DFS by about 0.4 day. However, when *GnRHR* gene function was evaluated by measuring circulating LH level after administration of GnRH in pre-pubertal animals, no significant associations of the SNPs with LH levels were detected. Furthermore, five SNPs were identified in the *FSHR* at residues 502, 669, 596, 658, and 685, four of which (at positions 502, 669, 658, and 685), had deleterious effects on fertility PTAs, increasing CI and DFS, while decreasing NR56.

Inaccurate heat detection has been identified as a major contributor to low fertility in dairy cattle, and therefore the second aim of this project was to investigate oestrus expression traits for which BLUP breeding values were estimated. Electronic tags (Fullwood and Lily) were used to measure behavioural activities at oestrus in two separate groups of animals. The first group comprised 103 cows in their first parity. Oestrus behaviour traits were: the number of steps at oestrus (STEPS) and the

percentage increase in the number of steps at oestrus (STEPS%) relative to the number of steps over the 10 days preceding oestrus (BASE). Cows in this group were genotyped at 10 genes (*GnRHR, LHR, FSHR, oestrogen receptors α and β (ESR), activin receptor, leptin, ghrelin, neuropeptide Y (NPY)* and *NPYRY2* for SNP identification. Of these SNPs, only those located in the *GnRHR*, *oestrogen receptor α* and *NPY* were found to have statistically significant effects on oestrus traits. The second group consisted of 189 cows with 995 oestrus behaviour records. Oestrus traits were: the number of activity units (AU) at oestrus (ACTIVITY) and the percentage increase in the number of AU at oestrus (ACTIVITY%) over baseline activity (BASE) determined over a 4 day rolling average. In addition to the above mentioned 10 genes, these cows were genotyped at *signal transducer and activator of transcription 1* and 5A *(STAT1, STAT5A), growth hormone receptor (GHR), prolactin, and prolactin receptor* genes. In this study SNPs in the *GnRHR, LHR, FSHR, ESRα, GHR, activin receptor, STAT1, and STAT5* were found to have significant effects on oestrus expression. The genetic basis for oestrus behaviour was also investigated applying various animal models in this group. ACTIVITY, ACTIVITY% and BASE had moderate heritabilities of 0.19 to 0.24, 0.16, and 0.25 respectively. The breeding values of the cows were shown to be normally distributed, and ranged between -19 to 27, -18 to 20, and -20 to 28 for ACTIVITY, ACTIVITY%, and BASE respectively. These values are significant in terms of the ability of the stockman to reliably detect cows in heat.

There have in the past been few genetic analyses of oestrus behaviour characteristics, and there are few studies of SNP associations with fertility traits of the kind described here. Therefore these studies make a significant contribution to our understanding of the genetics of fertility in dairy cattle, and can be used by breeders as a means to improving fertility in the national dairy herd.

Table of Contents

Declaration

I hereby declare that this thesis is a presentation of my own work and effort and that it has not been submitted anywhere for any award. Wherever contributions of others are involved they have been acknowledged. I also acknowledge all assistance given to me during my studies.

Part of this work has been published as Derecka et al. (2009) *Animal Genetics* **41**, 329-331.

Acknowledgments

I would like to thank all people who have made this thesis possible.

It is difficult to overstate my deep and sincere gratitude to my supervisor, Professor Anthony Flint. His wide knowledge and his logical way of thinking have been of great value for me. His understanding, encouragement, and his great efforts to explain things clearly have provided a good basis for the present thesis. Throughout my thesis-writing period, he provided advice, support, and lots of good ideas. I would have been lost without him.

I would like to express my great appreciation to my second supervisor, Professor Phil Garnsworthy for his detailed and constructive comments, and for his important support particularly with oestrous work.

I am also grateful to Professor Keith Campbell for being always accessible and willing to help his students with their research.

I was delighted to cooperate with Dr Raphael Mrode and to attend his classes at the Scottish Agricultural College, Edinburgh. I warmly thank him, for his great advice and friendly help. His extensive discussions about BLUP and interesting comments have been very helpful during this study.

I owe my deepest gratitude to Professor Arthur Gilmour for his invaluable help in ASREML. After attending ASREML course, he has continued to answer my endless questions about oestrous data analyses (chapter 6 and 7); his ideas have had a remarkable influence on my work.

I also thank Dr Kamila Dercka for doing the genotyping. My sincere thanks are due to Professor Charlie Hodgman for his great help particularly in carrying out nucleosome residence analysis. I would like to thank Professor John Woolliams for his help with the data analysis.

I would like to extend my warmest thanks to all those who have helped me with my work in the Division of Animal Science. I would like to express sincere thanks to Monica Bagshaw for her kindness and continuous support.

I am indebted to my best friend in high school (Nahed), my best friend as an undergraduate (Ghada), and my best friends as a postgraduate student (Ghofran and Pooh), for helping me get through the difficult times, and for the emotional support they have provided.

Dr Jane Morris has been constant source of help, support, and inspiration. Despite her hectic schedule as a general practitioner at the Church Lane Health Centre, she has always been there for me whenever help was required.

I would like to show great appreciation to the Syrian government for giving me this opportunity to study in the UK and for all the financial support I have received throughout my studies.

My deepest gratitude goes to my family for their love and support throughout my life particularly, my brother Saad and my sister Aleen, who have always been so close to my heart despite being far away from me.

For the love and care they have given me for all my life, words cannot convey my deep thanks and respect to my mother, Jamilah, and my late father, Aziz. I will never forget his last words to me while I was leaving Syria to study in the UK “I am so proud of you, God bless you”.

Most importantly, I would like to thank a very special person, my husband Rafe, for being my confidant and my friend. Thanks for pushing me to be the best, and giving me a shoulder to lean on when times were hard.

Last but not least, my warm and loving thanks go to my two little angels, Rama and Maya, who have come as a gift into my life, and have brought the happiest end to my PhD studies. To them and to my wonderful husband I dedicated this work.

Abbreviations

A: Numerator relationship matrix.

$\hat{a}$: Estimated breeding value.

a: True breeding value.

AA: Amino acid.

AU: Number of activity units during oestrus.

AU%: Percentage increase in the activity units during oestrus.

AI: Artificial insemination.

ACTR: Activin receptor.

BASE: Average number of activity units over the 10 days preceding oestrus.

BCS: Body condition score.

BLP: Best linear prediction.

BLUP: Best linear unbiased prediction.

BMPR1B: Bone morphogenetic protein receptor type 1B.

bp: Base pair.

BV: Breeding value.

Cart-1: Cartilage homeoprotein 1.

CI: Calving interval.

CINS: Number of inseminations per conception.

CL: Corpus luteum.

CLA: Commencement of luteal activity.

CREBP: cAMP response element-binding protein.

DFE: Days to first detectable oestrus.

DFS: Days in milk to first service.

EB: Energy balance.

ESR: Oestrogen receptor gene.

ESRKO: Oestrogen receptor knockout mice.

FecB: The Booroola gene.

FecX^{I}: The Inverdale gene.

FI: Fertility index.

FPLA: First postpartum luteal activity.

FSHR: Follicle stimulating hormone receptor.

FSHβ: Follicle stimulating hormone β.

FT: Functional teats.

GATA-1: Transcription factor.

GEBV: Genomic estimated breeding value.

GDF-5: Growth differentiation factor-5.

GDF9B: Growth deferential factor 9B.

GH: Growth hormone.

GHS-R: Growth hormone secretagogue receptor (Ghrelin receptor).

GnRHR: Gonadotrophin releasing hormone receptor.

GPCR: G- protein coupled receptor.

GS: Genomic selection.

GSE: Gonadotroph-specific element.

H: Selection objective.

h^2: Heritability.

HIF: Hepatic leukemia factor.

HNF-4: Hepatocyte nuclear factor 4.

i: Intensity of selection.

I: Selection index.

LL: Log likelihood.

LHR: Luteinizing hormone receptor.

Lmo2: Transcription factor.

MAS: Marker assisted selection.

MY: Milk yield.

MME: Mixed model equation.

n: Number of records.

NBA: Number born alive.

NEB: Negative energy balance.

NFY: Nuclear transcription factor Y.

NKS: Homeodomain transcription factor.

NPY: Neuropeptide Y.

NR56: Non-return rate after insemination.

nt: Nucleotide.

Ob-R: Leptin receptor.

Oct-1: Octamer transcription factor.

OVX: Ovariectomized.

P: Variance covariance matrix.

PCR: Polymerase chain reaction.

pe: Permanent environmental effects

PG $F_{2\alpha}$: Prostaglandin $F_{2\alpha}$.

PIN: Production index.

PLI: Profit lifespan index.

PR: Progesterone receptor.

PRLR: Prolactin receptor.

PTA: Predicted transmitting ability.

QTL: Quantitative trait loci.

$r_{a,y}$: The accuracy of the of estimation of the breeding value.

RBP4: Retinol-binding protein 4.

re: Repeatability.

RFLP: Restriction fragment length polymorphism.

s.d.: Standard deviation.

s.e.: Standard error.

SD: Selection differential.

SNP: Single nucleotide polymorphisms.

STAT1: Signal transducer and activator of transcription 1.

STAT5A: Signal transducer and activator of transcription 5A.

SF-1: Steroidogenic factor-1.

STEPS: Average number of steps at oestrus.

STEPS%: Percentage increase in the number of steps at oestrus.

t: Intra-class correlation.

te: Temporarily environmental effects.

TGFβ: Transforming growth factor β superfamily.

TNB: Total number of piglets born.

TSS: Transcription start site.

WTNB: Total weight of animals born.

WNBA: Total weight of animals born alive.

X^2: Chi-sequare.

Y1-Y6: Neuropeptide receptor.

YOB: Year of birth.

To Rafe
Rama and Maya

Chapter 1: The genetics of fertility

1.1 Introduction

There is substantial evidence to suggest that fertility of modern dairy cow in the UK has fallen in the recent years, and infertility has become a major problem affecting the profitability of cattle breeders and dairy farmers. In parallel with the reported decline of fertility, there has been constant drive towards breeding for higher milk yield in dairy cattle (Royal et al., 2000a; Wathes et al., 2008; Dobson et al., 2008). This has been achieved by breed substitution from the British Friesian to the North American Holstein, for which the proportion in UK bloodlines has increased from 0% to 80% (Royal, 1999). This suggests that the change in the genetic makeup of modern dairy cows has contributed to the downward trend in fertility traits in UK dairy cattle, as an antagonistic genetic relationship has been established between production traits and fertility in dairy cattle (Royal et al., 2002a; Pryce et al., 2004). Further evidence for this relationship in the dairy industry has been accumulated from several countries in which cows with the highest level of production have expressed the highest incidence of infertility (Lucy, 2001; Roxström et al., 2001; Evans et al., 2006). UK figures indicated that the genetic selection for higher milk yield has resulted in an increase in milk production by 1000 litres per lactation per decade, while the conception rate of dairy cattle has declined by 1% annually in the period between 1960s to 2000 and the calving rate had become as low as 40% in 2000 (Figure 1.1; Royal et al., 2000a). Furthermore, in a survey carried out by the World Holstein Friesian Federation in order to summarize the status of fertility in Holstein population around the world, it has been concluded that fertility is a serious problem and actions need to be taken internationally and within each country (Sørensen et al., 2007).

In addition to the genetic basis of infertility, management changes such as the increase in average herd size in the UK (from 72 to 113 cows between 1996 and 2009; Defra, 2009; DairyCo, 2009) may also have contributed to the decline in the fertility of dairy herds, as heat observation and hygiene may have been affected through less time being spent per cow (Dobson et al., 2008; Dobson et al., 2009).

Moreover, high yielding cows have high energy and nutritional requirements in order to maximize milk production. The onset of lactation is associated with a prolonged period of negative energy balance (NEB) during which metabolic demand for high milk yield exceeds energy intake. So many cows may still be in NEB at the start of the breeding season and this consequently results in a longer interval to the onset of ovarian activity postpartum through its effect on LH pulse frequency (Butler, 2003, Garnsworthy et al., 2008). There is a strong positive relationship between conception rate and early commencement of ovulatory cycles postpartum (Butler & Smith, 1989; Royal et al., 2000a, Royal et al, 2002a; Darwash et al., 1997; Petersson et al., 2007). Recent studies have demonstrated the deleterious impact of NEB on oocyte and embryo developmental competence in the high producing dairy cow. NEB has resulted in changes in the microenvironment of the growing and maturing female gametes leading to the ovulation of a developmentally incompetent oocyte. Suboptimal microenvironment in the uterus caused by the NEB has also been associated with early embryonic loss (Leroy et al., 2008).

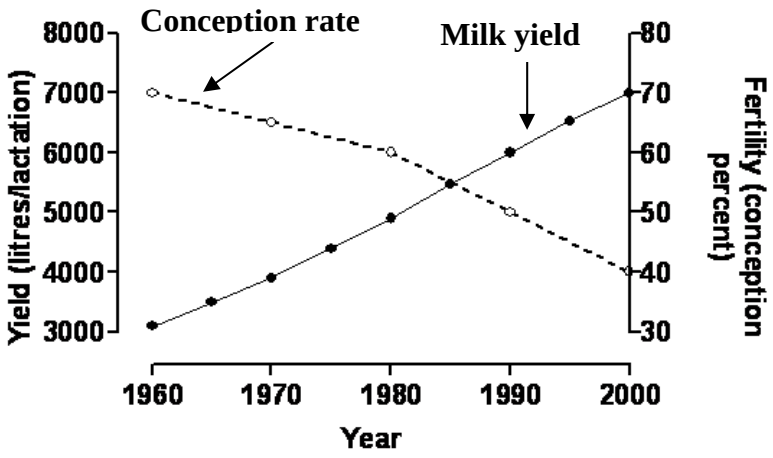


Figure 1.1 Trends in milk production and fertility with in the UK between 1960 and 2000 (Royal, 1999)

Poor fertility in dairy cattle is reflected in unfavourable genetic trends in all measures of fertility in the UK, such as calving interval (CI), days in milk to first service (DFS), non-return rate after insemination (NR56), and number of inseminations per

conception (CINS), with CI, DFS, and CINS increasing while NR56 decreasing (Figure 1.2; Wall et al., 2003). The increase in the number of inseminations per calving is associated with difficulties in detecting heat, and this leads to many cows being inseminated at an inappropriate time (Flint, 2006; Dobson et al., 2009).

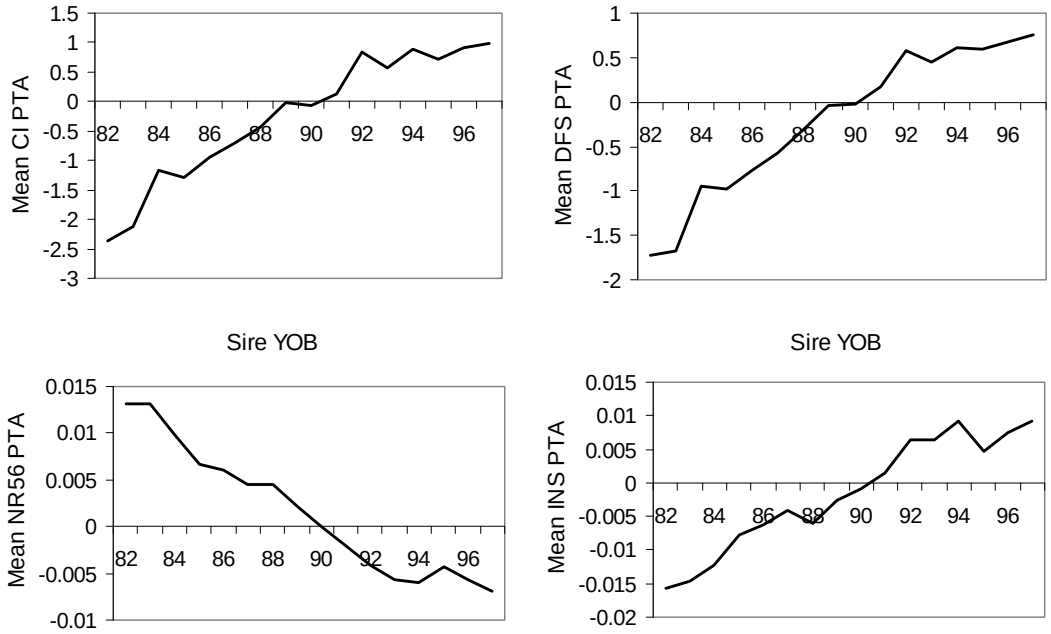


Figure 1.2 Genetic trends in calving interval (CI, days), days to first service (DFS, days), non-return rate (NR56, percent), and the number of inseminations per conception (CINS, count). X axis refers to YOB = year of birth (Wall, et al., 2003).

A recent survey has demonstrated that the average UK dairy cow survives only three lactations, and that infertility is the major cause for culling, reducing the availability of replacement heifers (Wathes et al., 2008). Therefore, many potential replacements never have the chance to enter the milking herd as 7.9% of calves are born dead and 3.4% die within 1 month. In addition, during the rearing phase, 6.7% of animals were reported to die before reaching first service at 15 months due to disease or accident and another 2.3% failed to conceive (Wathes et al., 2008). This decline in fertility has been supported by measurement of physiological indicators of fertility. Royal et al. (2000a) showed that lower fertility is accompanied by physiological changes in reproductive function, such as an increase in the incidence of atypical ovarian hormonal patterns from 32% to 44%. Furthermore, persistent corpora lutea in the first cycle postpartum and in subsequent cycles are of particular concern, as their

occurrence has increased from 7.3% to 18.2% and from 6.4% to 16.8% respectively. Furthermore, high incidence of prenatal losses associated with early embryonic mortality (about 40%), later embryo loss (up to 20% in high-yielding herds) or abortion (about 5%) have been reported in dairy cattle in the UK (Wathes et al., 2008).

In 2000b, Royal et al suggested three strategies to improve dairy cow fertility; these were: endocrine management, changes in herd nutrition, and genetic selection towards better fertility. Fertility traits, including traditional and endocrine fertility traits, are heritable (Wall et al., 2003; Royal et al., 2002a) and are genetically correlated with production traits including yield (Royal et al., 2002a) and linear type traits (Royal., 2002b, Wall et al., 2003). In addition, genetic selection for higher milk yield has been accompanied by poorer fertility; therefore control of fertility through genetic selection will in the long run offer a sustainable solution (Wall et al., 2003). In order to improve or prevent a further decline in fertility a fertility index for dairy cattle has been introduced, to provide the means by which milk producers can select for bulls with highly fertile daughters (Wall et al., 2003, Flint et al., 2002; 2003; 2004). Since its implementation in 2005, use of the fertility index in the UK has reversed the decline in fertility traits. At the genotypic level, CI has dropped by around three days during the period between 2001 and 2005 and now its level is equivalent to what it was in 1993, while at the phenotypic level, there has been a decline in DFS of about 5 days in the period between 2004 and 2008, DFS dropping to its level in 2000 of 85 days (Figure 1.3, Data of Eileen Wall, personal communication with Raphael Mrode). In line with this, many other countries have in recent years implemented genetic evaluation for fertility in dairy cattle which might have contributed to the improvement in fertility. This was the conclusion reached by the International Conference on Fertility in Dairy Cattle held in Liverpool in 2007 (Berglund et al., 2008).

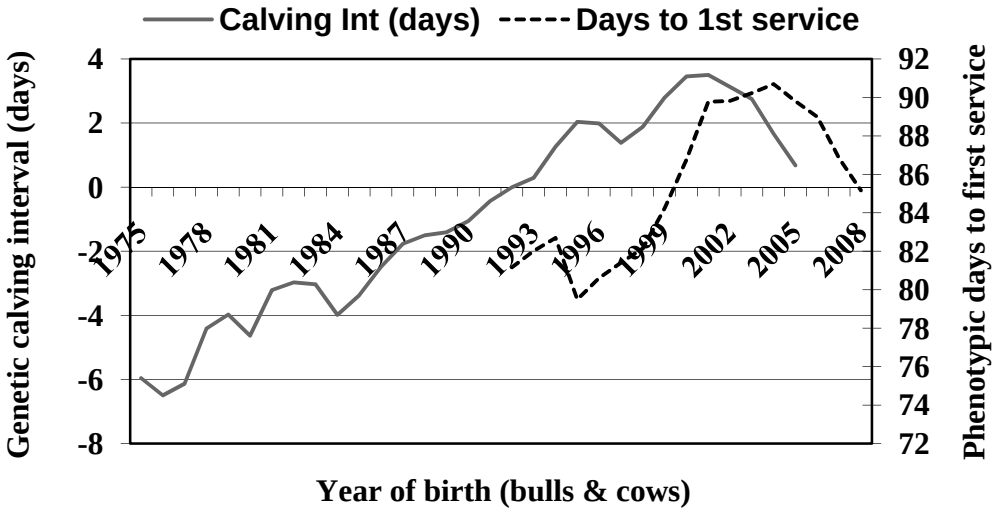


Figure 1.3 Genetic and phenotypic trends for CI and DFS respectively before and after the implementation of fertility index in 2005 (Data of Eileen Wall).

This chapter will review the genetic basis for fertility and discuss the development of the UK fertility index (FI). It will also show how the FI can be improved by using molecular markers, and review the evidence leading to particular genes being considered as candidate genes in SNP association studies.

Due to the high prolificacy rate observed in some breeds in sheep, ovulation rate in sheep considered an attractive topic in term of identifying genes that are involved in this process. This will be covered in the next section.

1.2 Litter size in sheep

1.2.1 Introduction

In sheep a large range in litter size has been observed both between different breeds and within the same breed. In most of the domestic breeds the average number of lambs at each lambing is one or two (Fabre et al., 2006). However, there is a small number of breeds which normally have a litter size of three or more, such as the Booroola Merino, Inverdale, Cambridge, Belclare, Coopworth etc (Montgomery et al., 1992; Davis, 2004). As a result, sheep have been an excellent model for identifying genes involved in the mechanisms underlying follicular growth and

ovulation rate (Fabre et al., 2006; Montgomery et al., 2001). Genetic studies have been carried out since 1980s to investigate the inheritance mode of genes affecting sheep prolificacy. It has been reported that the difference in ovulation rate and litter size might have resulted from the action of a single gene or closely linked group of genes (Davis, 2004). Booroola Merino and Inverdale breeds are among many breeds that have been investigated for possible genetic effects on litter size. It has been reported that the high fecundity expressed in these two breeds results from the action of genetic mutations with major effects on ovulation rate. Interestingly, both of these fecundity genes belong to the receptor of members of transforming growth factor β (TGFβ) and to the to the transforming growth factor (TGFβ) superfamily, respectively (McNatty et al., 2005). A commercial DNA testing service enables some of the mutations allocated to these genes to be used in genetic improvement programmes through marker assisted selection (Davis, 2005). The discovery of major genes that affect litter size will have a great impact on sheep breeding as an alternative to traditional means of selection, which have achieved slow progress. These genes can be introgressed into different breeds in different countries. These major genes may also have great impact on other species such as humans or cattle, where there is a good deal of evidence for genetic factors in multiple births (Montgomery et al., 1992; Davis, 2005).

1.2.2 The Booroola gene (*FecB*)

1.2.2.1 Phenotypic characteristics

The Booroola Merino is a highly fecund genotype developed through flock selection procedures from the medium Non-Peppin strain of the Australian Merino (Piper and Bindon, 1982). In the early 1980's, a study on litter size (Pier & Bindon, 1982) and ovulation rate (Davis et al., 1982) carried out on Booroola Merinos and their crosses revealed that the high prolificacy of the Booroola Merino may result from the action of a single autosomal major gene (or closely linked group of genes) with an additive effect on ovulation rate and a partially dominant effect on litter size. Homozygous and heterozygous carriers expressed higher ovulation rates of 3 and 1.5 respectively, and higher litter sizes of 1.5 and 1 respectively, compared to non-carriers (Davis,

2005). Piper & Bindon (1982) demonstrated that litter size for Booroola Merino ewes was higher than control Merino ewes with ranges of between 1 and 6, compared to 1 - 2 respectively. However, survival of Booroola lambs was considerably lower due to low survival rates of triple and higher order birth lambs (Table 1.1). Nonetheless, heterozygous Booroola ewes displayed higher fertility (5%), higher ovulation rates (80%) and higher litter size (50%), in comparison with the control ewes, despite lower overall lamb survival (64% vs 85%). Half-Booroola lambs also had a slower pre-weaning growth rate than lambs from control Merinos but this was not significant when litter size was taken into account.

The Booroola phenotype has been reported to follow a Mendelian pattern of segregation indicating that hyperprolificacy in the Booroola is influenced by a major gene, named *FecB*, that influences their ovulation rate and follicle size (Montgomery et al., 1992). On the basis of ovulation rate, genotypes can be segregated into homozygous carriers, heterozygous carriers, and non-carriers of the *FecB* gene recording five or more follicles, three or four follicles, and one or two respectively (Fabre et al., 2006).

Table 1.1 Least squares means ± s.e. for reproduction rate and its components in mixed age Booroola (2-7 years) and control (2-6 years) Merino ewes (Piper & Bindon, 1982).

Flock	Litter size (No)	Fertility (%)	Survival (%)	Reproduction rate (%)
Booroola	2.30±0.30	0.88±0.1	0.62±0.2	1.25±0.3
Control	1.30±0.03	0.92±0.1	0.84±0.2	0.98±0.3

In 1993 Montgomery et al developed a DNA marker test for the *FecB* gene and they proved that *FecB* maps to ovine chromosome 6, which corresponds to human chromosome 4. Two decades later Mulsant et al (2001) revealed that the *FecB* gene corresponds to a mutation in the coding region of the bone morphogenetic protein receptor type 1B (*BMPR1B* or activin-like kinase) gene. This finding was confirmed by two other studies (Wilson et al., 2001; Souza et al., 2001). *BMPR1B* ligands are BMPs which belong to the transforming growth factor β superfamily (TGFβ). There are 20 members of BMPs which share similar structures, having seven conserved cysteine residues. Six out of seven of these residues on each BMP unit form a three

dimensional structure called a cysteine knot, and the remaining residue forms a disulfide bridge between the two subunits to create a dimer. BMPs are synthesized as large inactive precursor proteins which need to undergo post-translational processing in order to become biologically active. This post-translational processing includes dimerisation and cleavage of the precursor protein into the mature active protein. BMPs are involved in the regulation of growth, differentiation, and apoptosis in a variety of tissues in addition to bone. They also play crucial roles in the reproductive system, particularly in the gonads (Shimasaki et al., 2004). The activin receptor (ActR-II) was the first to be defined as a BMP receptor, and this will be covered later in section 1.4.2.6. Another 4 BMP receptors have been characterised; these are ActR-IIB, AMHR-II (anti-Müllerian inhibiting substance receptor II), BMPR-II, and the TGF-β type-II receptor.

The BMP system is expressed in the ovaries by granulosa cells and oocytes from primary to the late antral follicle stage, and to a lesser extent, by the theca layer of antral follicles (Souza et al., 2001; Wilson et al., 2001). It is also expressed in the pituitary, hypothalamus, uterus, placenta, mammary gland, testis, and prostate (Shimasaki et al., 2004). The Booroola mutation is an A to G transition at nucleotide position 830 in a highly conserved area of the receptor resulting in a glutamine residue in non-carriers being replaced by arginine. This mutation was not detected in non-*FecB* gene carriers and was found to be strongly associated with the hyperprolificacy phenotype of Booroola ewes (Wilson et al., 2001, Souza et al., 2001). In 2002, Davis et al showed that Garole (Bengal) ewes possess the same mutation as the Booroola, suggesting that the Bengal sheep that arrived in Australia in 1972 are possibly the original source of the *FecB* mutation in the Booroola.

1.2.2.2 Physiological effects of the Booroola gene

The most obvious physiological effects of the *FecB* locus are on the number of ovulatory follicles in the ovary, ovulation rate and litter size. The increase in ovulation rate in *FecB* carriers results from a change in the criteria by which the follicles are selected, associated with the precocious development of a large number of small antral follicles which ovulate at a smaller size than non-carrier follicles,

leading to an increase in ovulation rate and multiple births (Fabre et al., 2006). The smaller ovulatory follicles of *FecBFecB* ewes contain a smaller number of granulosa cells and produce less oestradiol than those of control ewes. However, the increased number of ovulatory follicles in *FecBFecB* carriers offsets these differences resulting in the same amount of oestradiol being produced from both genotypes (Montgomery et al., 1992). Therefore, oestradiol-induced GnRH secretion is triggered at the same oestradiol level in both carriers and non-carriers of the *FecB* gene, leading to the ovulation and luteinisation of a large number of LH-responsive follicles in *FecB* gene carriers (McNatty et al., 2001).

The product of the *FecB* gene has no obvious effects on hypothalamic function. This suggests that its main effects are likely to be at the level of the pituitary gland or ovary, presumably by interfering with pituitary hormone composition, storage, and secretion, in addition to affecting the morphology and function of developing ovarian follicles. The pituitary gonadotrophin content is lower during fetal life in *FecBFecB* carriers in comparison with non-carriers but it tends to be higher in adult life. The FSH level in peripheral plasma was significantly higher in both neonatal and adult *FecB* carriers in comparison with non-carriers. There is no effect on LH level (Montgomery et al, 1992).

Moreover, in a study on granulosa cells in sheep it has been reported that the growth differentiation factor-5 (GDF-5) and bone morphogenetic protein-4 (BMP-4), two natural ligands of *BMPRIB*, have strong inhibitory effects on both basal and FSH-induced progesterone secretion, with no clear effect on cell proliferation. However, the granulosa cells in *FecBFecB* carriers are less susceptible in terms of progesterone secretion to the inhibitory effects of *BMPRIB* ligands than ovarian cells from non-carriers. Therefore, the mutation in *BMPRIB* modifies its activity, leading to partial loss of function. This results in premature differentiation of granulosa cells and an advanced maturation of ovulatory follicles leading to the acquisition of fewer granulosa cells by the follicles in gene carriers (Mulsant et al., 2001). Furthermore, because *FecB* gene carriers have higher levels of FSH than non-carriers (Montgomery et al., 1992), the *BMPRIB* mutation has been suggested to impair the inhibitory effects of BMPs on FSH receptor expression in granulosa cells.

Consequently, this increases the follicles' sensitivity to FSH (Fabre et al, 2006) and augments the expression of FSH-dependent markers for differentiation such as inhibin/activin subunits and their receptors and the LH receptor in granulosa cells of antral follicles. These markers have resulted in the premature differentiation of granulosa cells in follicles with smaller size in gene carriers compared with non-carriers (McNatty et al., 2001). However, despite the important differences in folliculogenesis between *FecB* gene carriers and non-carriers, oocytes from mature follicles produced a viable offspring with no obvious differences in fertility or embryo survival among all genotypes (Montgomery et al., 2001).

Few studies have been carried out on expression of the *FecB* gene in rams. It has been found that the *FecB* mutation may reduce fetal body weight in both males and females (Montgomery et al, 1992). In general, there is a significant genotypic difference in body weight of Booroola rams with *FecBFecB* rams being smaller both in stature and body weight compared with non-carriers. In Booroola rams no genotypic differences were found in testis size, Sertoli cell number, or in spermatogenic function (Montgomery et al, 1992).

The effects of the *FecB* gene on the phenotype of homozygous and heterozygous carriers in comparison with non-carriers are summarized in Table 1.2.

Table 1.2 A summary of the physiological effects of *FecB* gene carriers in both homozygous and heterozygous.

Homozygous (*FecBFecB*)	**Heterozygous (*FecBFec+*)**
Multiple ovulations (3)	Multiple ovulations (1.5)
Increased FSH concentration	Increased FSH concentration
Follicles mature at smaller size: • More responsiveness to FSH • Have fewer granulosa cells • Granulosa cells acquire LHR earlier	Follicles mature at smaller size: • More responsiveness to FSH • Have fewer granulosa cells • Granulosa cells acquire LHR earlier

1.2.3 The Inverdale gene (*FecX*^{I})

1.2.3.1 Phenotypic characteristics

In the early 1990s intensive studies were undertaken to investigate the high prolificacy of Romney ewes. These were mostly concerned with the offspring of a single foundation ewe with a history of 33 lambs born in 11 lambings, which produced a number of female offspring with high ovulation rates (Davis et al., 1991). Four progeny tests were consequently carried out to determine the inheritance pattern of ovulation rate, which showed that the high prolificacy resulted from the action of a new major gene affecting ovulation rate in this line. It was clear from the pattern of inheritance and the phenotype being displayed that this gene is carried on the X-chromosome, as ram carriers of the Inverdale gene passed on the gene to all of their daughters but not to any of their sons. This gene is known as the Inverdale gene and in view of its locus, it has been named *FecX*^{I} (Davis et al., 1991).

In 2000, Galloway et al proved that the effect of the *FecX*^{I} gene was due to a mutation in a homologous chromosomal locus syntenic to the human X chromosome where the bone morphogenetic protein 15 (*BMP15* also known *GDF9B*) is expressed. BMP15 is a member of the transforming growth factor β (TGFβ) superfamily and is particularly expressed in oocytes from the primary follicle stage onwards (Galloway et al., 2000; Montgomery et al., 2001). This mutation was reported to be a T to A transition at nucleotide position 92 of the mature peptide of *BMP15* causing the replacement of a valine residue with aspartic acid. This substitution has been suggested to impair the ability of the BMP15 to dimerize, resulting in a biologically inactive protein (Galloway et al., 2000). In heterozygous carrier ewes (*FecX*^{I} *FecX*$^{+}$) the *FecX*^{I} gene was found to increase ovulation rate and litter size. However, ewes which are homozygous carriers (*FecX*^{I} *FecX*^{I}) are infertile (Davis et al., 1992). Therefore, the heterozygous daughters (*FecX*^{I} *FecX*$^{+}$) are obtained commercially by crossing carrier rams with non-carrier ewes (Davis, 2005).

1.2.3.2 Physiological effects of the Inverdale gene

In heterozygous female carriers the *FecX*^{I} gene increases ovulation rate by about 1 and litter size by about 0.6 (Davis et al., 1991). However, mating rams with one copy of the gene with heterozygous daughters demonstrated that homozygous carriers have abnormal streak ovaries. Streak ovaries were found in nearly 50% of female progeny characterized by their small size, large numbers of primordial follicles, and few (and frequently abnormal) primary follicles. Subsequently, the ovaries were empty of follicles with more than one layer of granulosa cells. These findings indicate that normal development beyond the primary stage is impaired, suggesting that two copies of the *FecX*^{I} gene affect the earliest phase of follicular development (i.e. transformation of primordial follicles) and may in some instance lead to the development of abnormal structures. In addition, the streak ovaries in *FecX*^{I}*FecX*^{I} ewes were found to be about one eighth the size of normal ovaries (Davis et al., 1992). The germ cell number in the ovaries of homozygous *FecX*^{I}*FecX*^{I} carriers and non-carriers were the same during early fetal development. However, by day 105 of gestation the ovaries of homozygous *FecX*^{I}*FecX*^{I} fetuses had started to reveal abnormal structures such as oocytes with no follicular cells, follicles with degenerating oocytes or oocyte-free follicles (Montgomery et al., 2001).

Galloway et al (2000) reported that BMP15 mRNA is first expressed in primary follicles and is exclusively found in the oocytes of both carriers and non-carriers of the *FecX*^{I} gene. Nevertheless, given that homozygous *FecX*^{I}*FecX*^{I} ewes are sterile it appears likely that the mutation impairs the production of biologically active *BMP15* causing a complete loss of functional protein (Galloway et al., 2000). However, inactivation of only one copy of BMP15 in *FecX*^{I} heterozygotes increases ovulation rate. This is because BMP15 reduces the sensitivity of granulosa cells to FSH by suppressing expression of the FSH receptor. As a result, in heterozygous carriers the decline in bioactive *BMP15* reduces granulosa cell proliferation, leading to lower secretion of steroids and inhibin. This in turn increases FSH-induced granulosa cell responsiveness of the small follicles and consequently enhances the action of FSH, thereby accelerating follicular development and causing precocious ovulation (Montgomery et al., 2001, McNatty et al., 2005). Furthermore, it has been suggested

that the reduced level of active BMP15 allows other oocyte-derived growth factors to express their effects on granulosa cell proliferation and differentiation. This includes the growth differentiation factor GDF9, another member of the TGFβ superfamily, which is an essential factor in follicular growth and is expressed in all primordial and all growing follicles (Galloway et al., 2000).

$FecX^{I}FecX^{+}$ ewes have significantly more antral follicles than non-carriers, despite having similar numbers of primordial follicles. The granulosa cells in $FecX^{I}FecX^{+}$ ewes develop higher LH responsiveness and acquire LH receptors at smaller diameters than in non-carriers. Moreover, $FecX^{I}FecX^{+}$ ewes have fewer granulosa cells and smaller corpora lutea (CL) than non-carriers. However neither total weight of luteal tissue nor total ovarian secretion rate of progesterone was different between the different genotypes, and no gene-specific differences were noted for ovarian hormones or for FSH or LH. The presence of the LH receptor in smaller follicles and the smaller size of CL are consistent with the findings that follicles ovulate at smaller diameter in heterozygous carriers in comparison with non-carriers (Shackell et al., 1993). Therefore, it has been concluded that the $FecX^{I}$ gene in heterozygotes acts as a paracrine regulator on the ovaries without changing significantly ovarian or pituitary hormone secretion (McNatty et al., 2005). The effect of $FecX^{I}$ gene on the phenotype of homozygous and heterozygous carriers is illustrated in Table 1.3.

Table 1.3 Asummary of the physiological effect of $FecX^{I}$ in homozygous and heterozygous carriers.

Heterozygous carriers ($FecX^{I}FecX^{+}$)	**Homozygous carriers ($FecX^{I}FecX^{I}$)**
Multiple ovulation • More antral follicles • Follicle mature at smaller size	Streak ovaries • Small size • Large number of primordial follicles
Have fewer granulosa cells • Granulosa cells acquire LH receptor earlier	Follicles fail to grow normally beyond primary stage

1.2.4 Conclusion

Major genes with different effects on litter size and ovulation rate offer a new option by which to increase reproductive performance in sheep flocks throughout the world. These major genes have the advantage that they can be introgressed into any new breed without changing other breed characteristics. The introduction of the Booroola gene from the dwarf Garole breed in India to the fine-wool Merino in Australia and later to the long-wool Romney in New Zealand is a good example of the action of a major gene (Davis, 2005).

The discoveries of mutations in *BMPRIB* and *BMP15* and their physiological effects have generated a new understanding of the roles of BMPs as intra-ovarian regulators of folliculogenesis and ovulation rate (Fabre et al., 2006). In addition, the high expression of these two mutations in oocytes also indicates to their significant impact on determining the number of follicles that continue to the ovulation stage (McNatty et al., 2001).

These studies highlight the potential importance of genes involved in ovarian function as candidate genes for fertility control. As such, these genes represent important opportunities for marker assisted selection. Among the non-ruminant domestic species, the pig is possibly the best example of the application of genetic selection using molecular markers.

1.3 Litter size in pigs

1.3.1 Introduction

Litter size (fecundity) is one of the most economically important traits in the pig industry, and achieving a higher number of piglets weaned per sow is an important aim for pig breeders (Rothschild et al., 1996). Litter size in pigs varies from approximately 2 to 20 piglets per litter, with means from 9 to 11, depending on the breed. Phenotypic standard deviations are between 2.63 and 3 piglets, and heritability is 17% (Johnson et al., 1999). Conventional selection for higher litter size in pigs has

had limited success due to its moderate heritability and the sex limited nature of reproductive traits (Rothschild et al., 1996). Primarily, many quantitative trait loci (QTL) studies have been undertaken to identify regions that may have an impact on traits of interest in pigs. However, this method has had limited success due to small population size, moderate to large effects of the chromosomal regions identified, and the difficulty in identifying the gene or genes associated with known QTL. Therefore, in order to enhance the genetic improvement of litter size, a candidate gene approach has been employed. In this method genes examined for polymorphisms are selected based on the biological and physiological information about them (Rothschild et al., 2007). This approach has been successful in identifying genetic variants at a single locus. A number of polymorphic genes with significant impact on litter size in the pig have been identified such as the oestrogen receptor gene (ESR; Rothschild et al., 1996), FSHβ gene (Zhao et al., 1998), retinol-binding protein 4 gene (RBP4; Rothschild et al., 2000), and the prolactin receptor (PRLR; Vincent et al., 1998). The size of the effects of these genes ranges from an extra 0.25 to over 1 piglet per allele per gene copy with variations depending on the genetic background. These effects will be considered in detail below, as they represent candidate genes that may be used in other species, including dairy cattle.

1.3.2 The effect of polymorphic genes on litter size in pigs

1.3.2.1 The effect of oestrogen receptor gene (*ESR*) on litter size

The *ESR* has been chosen as a candidate gene for litter size due to its integral role in several reproductive pathways, which will be covered in detail in section 1.4.2.5. The Chinese Meishan is one of the best known prolific pig breeds, producing four more viable piglets per litter than American and European breeds. To determine a genetic cause of the increased litter size observed in Meishan pigs, Rothschild et al. (1996) investigated the effect of *ESR* gene polymorphisms on litter size in a divergent breed cross involving the Meishan. The results demonstrate that a specific allele B of the *ESR* locus is significantly associated with higher litter size in comparison with the other undesirable A allele. The effect of the *ESR* B allele seems to be additive on the litter of first parity. The differences between BB and AA female pigs carriers from a

synthetic line with 50% Meishan background is 2.3 more pigs in first parities and 1.5 more pigs on average over all successive parities. Moreover, pigs with Large White breed ancestry were also found to carry the beneficial *ESR* B allele with BB carriers tending to produce 1 more piglet born alive in comparison with AA carriers. Due to the differences, it has been speculated that the effects of *ESR* may depend on the genetic background. The presence of the favourable allele in the Large White might be through interbreeding of Chinese pigs with pigs in England before 1800, which eventually became the Large White breed (Rothschild et al., 1996).

The effect of the *ESR* B allele on litter size was investigated in a large population of pigs from four commercial pig lines. Three lines were of Large White origin and the fourth was a 1/4 Large White synthetic line. The *ESR* locus was tested for its possible effects on the total number of piglets born (TNB) and the number born alive (NBA). The results revealed the additive effects of *ESR* B allele on litter size to be about 0.42 and 0.39 piglets/litter for TNB and NBA in the first parity and 0.3 piglets/litter for TNB and NBA in later parities. No dominance effect was detected in the first parity, but in later parities a dominance effect of half of the value of the additive effect was found (Short et al., 1997). In a population of Yorkshire, Large White, and crossbred animals, *ESR* alleles were significantly associated with the total weight of animals born (WTNB) and the total weight of animals born alive (WNBA). BB dams expressed considerably lighter WTNB and WNBA in comparison with AA dams. The small birth weight observed in the Meishan is in agreement with the hypothesis that the Meishan breed is the origin of the B allele. There has been some indication of the possible effect of the ESR B allele on the total number of piglets alive at weaning but this effect was not significant. However, no significant effects of the ESR B allele were reported on TNB and NBA (Isler et al., 2002).

The discovery of the *ESR* polymorphisms is of a great commercial impact as it has no negative effects on other performance traits such as growth rate. It has been reported that BB females have slightly higher growth rates and are fatter, with higher numbers of functional teats (FT) (Rothschild et al., 1996). However, Short et al. (1997) demonstrated that the ESR B allele resulted in a small favourable effect on backfat thickness associated with a small decrease in average daily feed intake, while

no effect was found on growth rate. There was also a small but significant negative effect on teat number.

In 1996, Rotthschild et al reported that the discovery of the significant effect of the *ESR* locus on prolificacy has substantial economic value. The value of one extra pig per litter has been estimated in to be between $15 and $30 depending on market condition. This outcome demonstrates the benefits of the candidate gene approach to discovery QTL with significant effects on economic traits in pigs. Therefore, the commercial pig breeders have effectively incorporated these QTL in marker assisted-selection along with the traditional performance information in order achieve further improvement in the traits of economic importance. This eventually, will improve the efficiency of pig production by reducing the costs of production and increasing feed efficiency (Rothschild et al, 2007).

1.3.2.2 The prolactin receptor (*PRLR*)

The prolactin receptor (*PRLR*) was investigated as a potential gene influencing reproductive traits in pigs due to its involvement in reproduction. In pigs the *PRLR* gene has been mapped to chromosome 16 (Vincent et al., 1998). Polymorphisms have been detected and two alleles have been identified as A and B. Vincent et al. (1998) revealed that the A allele of the *PRLR* locus was linked to a significant additive effect on litter size in three commercial lines, involving the Meishan, Large White and Landrace. In Large whitle breed the AA animals have 0.66 pig per litter advantage in NBA over the AB and BB. While in Meishan, BB animals produce 0.33 pig/litter more than the AA animals. In Meishan the *PRLR* gene imposed significant dominance effects of 0.55 and 0.63 on TNB and NBA respectively. The highest effect was noticed in the Landrace, where the difference between the two homozygous genotypes was greater than one piglet per litter for both TNB and NBA. The positive effect of allele A on litter size was confirmed in another study carried out on Polish Large White and Landrace sow crosses. In the first parity, AA carriers have one more piglet for both the TNB and NBA in comparison with AB and BB carriers. However, in later parities this effect was not significant (Terman, 2005). Van Rens and Van der Lende, (2002) reported comparable results in a crossbreed of

Large White and Meishan piglets with AA gilts tending to have higher NBA of 2.45 than BB. However, allele A has been found to be negatively correlated with age at first oestrus and the number of FT. BB gilts had their first oestrus 41 days earlier than AA gilts and had significantly higher FT number. However, the magnitude and the direction of the gene effects on litter size vary depending on the genetic background of the line. In the Duroc line in Germany, the favourable allele of the *PRLR* was found to be the B allele, with BB animals having 0.71 more piglets than AA across all parties (Drogemuller et al., 2001).

Tomás et al. (2006) in a study on Iberian × Meishan crossbred sows have proposed that the effects of the *PRLR* gene on ovulation rate might be explained by the combined effects of multiple mutations rather than a single mutation in *PRLR*. *PRLR* DNA has been sequenced and 6 non-conservative SNPs have been identified within exon 10 which encodes the intracellular domain of the gene. These SNPs were at nucleotide positions: C1217T, C1283A, G1439A, T1528A, G1600A and G1789A relative to the translation start site. Eight haplotypes were found with different frequency in the porcine lines. The three major haplotypes were named: $PRLR_A$, $PRLR_B$, and $PRLR_C$. The other five were less frequent and showed a breed-specific distribution i.e. $PRLR_D$ and $PRLR_E$ were only found in Meishans. The SNPs at nucleotide positions C1217T and G1439A have significant additive effects on the number of corpora lutea (CL). An increase of 0.85 CL was recorded in sows expressing C1217T and G1439A in comparison with non-carriers. The SNP at G1789A which corresponds to the mutation (B>A) described by Vincent et al. (1998) has no effect on the number of CL. However, when haplotypes rather than single SNPs were analysed, a significant correlation was found. The $PRLR_B$ haplotype increased CL number by 2 while $PRLR_E$ displayed unfavourable effects on the number of CL. The AA genotype at position 1789 in Meishans has been found only in animals homozygous for $PRLR_E$, which has negative effects on ovulation rate. These results may clarify the lower litter size reported by Vincent et al. (1998) in the AA animals of the Meishan breed. Moreover, *PRLR* has been shown to have multiple pleiotropic effects. In addition to its effects on ovulation rate, there has been some indication for its possible role on the metabolism, suckling behaviour and the

viability of the newborn pigs. However, in this study no effects of *PRLR* were found on litter size (Tomás et al., 2006).

1.3.2.3 Follicle-stimulating hormone *β (FSHβ)*

Follicle stimulating hormone (FSH) influences follicular growth and development through the regulation of granulosa cell proliferation, differentiation, and steroidogenic function. In addition, FSH is considered to be a major regulator of ovulation rate in pigs, stimulating follicular recruitment while it blocks granulosa cell apoptosis. The expression of *FSH* receptor mRNA changes remarkably during late follicular development in pigs; *FSHR* tends to be high during the early follicular phase but decreases significantly as follicles grow and approach ovulation (Cárdenas & Pope, 2002). A positive genetic relationship has been demonstrated between plasma concentration of FSH and ovulation rate in pigs (Cassady et al., 2000). Plasma level of FSH was found to be moderately heritable in gilts at day 58 after birth (h^2 = 0.41; Cassady et al., 2000). Lines selected for higher ovulation rate and embryonic survival expressed higher breeding values than randomly selected animals for FSH concentration at 58 days in gilts and 90 days for both gilts and boars. Since FSH level can be measured in both sexes, higher selection intensities can be achieved by selecting directly for higher plasma concentration of FSH rather than through selecting for higher ovulation rate. Selection for higher FSH level was estimated to be 93% as effective in improving ovulation rate as direct selection for greater ovulation rate (Cassady et al., 2000). Zhao et al. (1998) found that sows homozygous for the beneficial allele (B) of the *FSHβ* gene subunit produced on average 2.53 and 2.12 more piglets for TNB and NBA respectively in comparison with homozygous carriers for the non-beneficial allele A. Despite a tendency for the effect of the B allele to decline in later parities, BB sows exhibited 1.5 piglets born per litter more than AA sows, with the B allele displaying a dominant mode of gene action. However, no genetic effects of *FSHβ* were detected on body weight at birth and body weight at 20 days, which may imply that there is no significant pleiotropic effect of the *FSHβ* locus on other performance traits. Linville et al. (2001) in a study on the effects of *FSHβ* on reproduction traits in advanced selection lines in the pig have reported a significant increase of the favourable allele of the *FSHβ* in lines

selected for higher ovulation rate and embryonic survival. However, when estimating the additive and dominance effects of the *FSH* gene it did not differ from zero.

1.3.2.4 Retinol-binding protein 4 (RBP4)

Harney et al. (1990) indicated that *RBP* mRNA is present in the peri-implantation conceptus, endometrium, and glandular epithelium in pigs. The presence of *RPB* mRNA in the conceptus indicates a possible role as a transporter of retinoids to the developing conceptus. Pig conceptus secretion of RBP starts at day 10 of gestation and continues throughout the peri-implantation period (day 10-16, Harney et al., 1990). Rothschild et al. (2000) in a study carried out on 1300 *RBP4* genotyped sows of six commercial lines with nearly 2800 litters, showed a significant additive effect associated with the *RBP4* gene. Homozygous carriers for the favourable allele had 0.50 pigs per litter for TNB and 0.26 for NBA higher than homozygous controls. This increase in litter size was not associated with a lower birth weight. RBP4 has been mapped to chromosome 14 where a small number of genes and no QTL affecting litter size have been identified. Therefore, it is not clear from this study whether *RBP4* is a major gene or whether it is correlated with a gene or genes affecting litter size. Drogemuller et al. (2001) have demonstrated that despite differences in the frequencies of the favourable allele A in three German lines of pigs, no significant effect was found among the different genotypes of *RBP4*. In a recent study by Munzo et al (2010) five intronic and one exonic polymorphisms were detected in the *RBP4* in a Chinese-European porcine line. A joint significant effect of one of the RBP4 intronic polymorphisms and ESR polymorphism of 0.61 ± 0.29 was detected on NBA.

1.3.3 Conclusion

In the pig industry, incorporating genetic markers with traditional selection methods has resulted in improving litter size in pigs. The increase in litter size associated with allelic substitution in the *ESR*, *PRLR*, *FSHβ*, and *RBP4* genes will result in substantial economic returns for commercial pig breeders. However, the variation in

gene effects associated with the genetic background of the lines imposes the necessity for different selection methods to be designed for each line and consideration has to be given to the possible pleiotropic effects of the genes.

As fertility of dairy cattle is the key topic of this thesis, the physiology and the major candidate genes for fertility in dairy cattle will be discussed in the following section.

1.4 Candidate genes for fertility in dairy cattle

To find candidate genes for fertility, the hormones involved in reproduction need to be identified. To extend the discussion of gene effects described in sheep and pigs in section 1.2 and 1.3, the physiology of reproduction in the cow is reviewed in this section.

1.4.1 Bovine oestrous cycle

The oestrous cycle may be divided into four phases which are hormonally controlled. These phases are: proestrus, oestrus, metoestrus, and diestrus. The proestrous and oestrous periods are influenced by the circulating level of oestrogen and are associated with the growth of the ovarian follicles. Metoestrus and diestrus are associated with the growth of the corpus luteum and are affected by the level of progesterone.

Proestrus is the period of preparation for mating and is characterized by follicular growth and oestrogen production, with a consequent increase in blood flow to the reproductive tract. The secretory activity of cervical and vaginal glands is stimulated leading to the production of a thin vaginal discharge. The proestrous period lasts for about 2 to 3 days and at the end of this phase cows exhibit oestrus.

The oestrous period is characterized by the behavioural demonstrations of heat which will be covered in section 1.5.4. During this period cows become markedly restless showing an increase in physical activities. Cows tend to accept bulls or stand to be

mounted by other cows. These behavioural changes are the result of the action of oestrogen on the central nervous system. After about 18 to 19 hours these clinical signs of oestrus start to decrease, and ovulation then occurs 10 to 11 hours after the end of oestrus. However, these signs of oestrous activity in cows can be easily missed, unless observed frequently (Hafez, 1993). Techniques for measuring oestrous behaviour will be covered in Section 1.5.5.

The period immediately following oestrus is metoestrus, which is the period in which ovulation occurs. As the egg is released from the follicle, the follicular cavity develops to form the corpus luteum (CL). Metoestrus lasts for only 2 to 3 days. It is not uncommon to observe a blood stained mucus discharge during this period, which results from the blood supply to the reproductive tract tissues during oestrus. During this period the egg is transported into the infundibulum of the oviduct and begins its course toward the uterus (which takes between 5 and 7 days).

The last and the longest period in the oestrous cycle is diestrus, during which the CL becomes fully functional. The essential function of the CL is to produce progesterone, a hormone that regulates several physiological functions. Progesterone prepares the uterus for pregnancy by increasing the thickness of the endometrium and developing the uterine glands in anticipation of embryo implantation. If fertilization occurs, progesterone secreted by the CL will maintain pregnancy through to the formation of the placenta. Whether pregnancy has resulted or not the CL will develop into a fully functional organ producing large amounts of progesterone and inhibiting other follicles from ovulating. If a fertilized egg reaches the uterus, the CL will be maintained throughout pregnancy. If on the other hand, the egg which reaches the uterus is not fertilized, the CL will remain functional for up to 19 days, but starts to degenerate at around day 16. The regression of the corpus luteum and proestrus occur simultaneously, thereby removing the inhibitory effect of progesterone on gonadotrophin secretion, and permitting a new oestrous cycle to start (Salisbury et al., 1987).

After calving cows exhibit anoestrus, the period between calving and first oestrus (postpartum commencement of luteal activity, CLA). The duration of this period is

controlled by several environmental, physiological and metabolic factors. Postpartum CLA is affected by season (being longest in the spring and shortest in the autum; Drwash et al., 1997), milk production (longer in high producing cows; Royal et al., 2002), frequency of milking (cows milked twice a day start cycling earlier than cows milked four times; Hafez, 1993), and by the genetic potential for milk production (Royal et al., 2002a). CLA is also influenced by the rate of uterine involution, the rate of follicular growth, the peripheral and central level of gonadotrophins, peripheral concentration of oestrogen and progesterone, and by the change in body weight and energy intake (Butler, 2000). During pregnancy the cow's placenta produces large quantities of progesterone which has negative effects on the activities of the hypothalamus and pituitary. So returning to cyclicity after calving is associated with the recovery of the hypothalamus and pituitary from the suppression caused by the preceding pregnancy, allowing oestrus and ovulation under the influence of follicle stimulating hormone (FSH) and luteinizing hormone (LH) secretion (Butler & Smith, 1989). During this period, cows may ovulate without showing oestrus (the phenomenon known as 'silent heat'). This occurs because oestrous behaviour requires progesterone withdrawal to precede oestrogen secretion, and before the first ovulation there can be no corpus luteum, and therefore no progesterone withdrawal (Dobson et al., 2008, Lopez et al., 2004).

1.4.2 Hormones regulating oestrus

The bovine oestrous cycle is regulated by endocrine and neuroendocrine mechanisms. The hormones involved are secreted by the hypothalamo-pituitary-gonads axis. Changes in the levels of these hormones regulate follicular waves, the timing of ovulation, and the length of the oestrous cycle.

During proestrous the circulating concentration of FSH increases markedly resulting in the recruitment of a group of follicles. FSH reaches a maximum level immediately before ovulation, and thereafter decreases. During the presence of the dominant follicle the concentration of FSH continues to be low, preventing the commencement of a new follicle wave. At the time of FSH secretion and follicular growth, the amount of oestradiol produced by the follicle increases and reaches a peak at the day

of oestrus (Salisbury et al., 1987). The high level of oestradiol causes standing oestrus and behavioural changes associated with oestrus. A high concentration of oestradiol in the absence of progesterone has positive effects on secretion of gonadotrophin releasing hormone (GnRH) by the hypothalamus. This increase in the level of GnRH secretion leads to an increase in LH secretion, which reaches a maximum level (the LH surge) shortly before ovulation. The LH surge results in ovulation and the formation of the corpus luteum (Hafez, 1993). LH plays a key role in regulating the growth and development of the dominant follicle. In addition, LH is necessary for maintenance of the corpus luteum (CL) and for stimulating the production of progesterone by the CL. Progesterone, produced by the CL prepares the uterus to receive the fertilized egg and allows the uterus to maintain pregnancy. If pregnancy does not occur, the uterus releases prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$) which determines the lifespan of the CL and the concentration of FSH increases, inducing a new follicle wave (Hafez, 1993).

The activity of the hypothalamus is also mediated through higher centres in the brain which allow for the integration of neural signals from other non-reproductive systems such as those controlling EB and the response to photoperiod. In dairy cows, NEB is directly related to the postpartum initiation of ovarian activities. This effect of NEB has been speculated to be through the effect of a neuroendocrine opioid inhibiting LH pulses (Butler & Smith, 1989).

GnRH, LH, FSH, oestradiol and progesterone all act through binding to specific receptors. Therefore, these receptors play fundamental roles in ensuring the cow conceives, and they are candidate genes in terms of identification of DNA markers for fertility. In recent years it has become evident that mutations and sequence polymorphisms in the receptors for gonadotrophins and GnRH are involved in some forms of human hypogonadism and pathological pubertal maturation (Huhtaniemi, 2002). Moreover, single nucleotide polymorphisms (SNPs) associated with fertility have already been identified in the bovine hypothalamic-pituitary-gonadal axis (Hastings et al., 2006). The above mentioned hormones, together with some other hormones which are known to have an impact on energy metabolism such as leptin and ghrelin, will be discussed in more detail in the following sections.

1.4.2.1 Gonadotrophin releasing hormone receptor *(GnRHR)*

GnRH acts through binding and activating the GnRH receptor (GnRHR), a member of the seven-transmembrane domain G protein-coupled receptor (GPCR) family (Fan et al., 1995). However, GnRHR differs from other G-protein coupled receptors in the loss of the cytoplasmic carboxy terminal domain (Millar et al., 2004). The human *GnRHR* gene, which is composed of three exons and two introns (Fan et al., 1995), codes for a protein of 328 amino acids, as does the bovine gene (Kakar et al., 1993). The GnRH receptor has been detected primarily in pituitary gonadotroph cells, in addition to the gonads, placenta, adrenal glands, the central nervous system and some neoplastic tissues (Stojikovic et al., 1994). However, the bovine GnRHR was localized in the pituitary but not in many extrapituitary tissues such as hypothalamus, testis, corpus luteum and ovary (Kakar et al., 1993). Binding of the GnRHR to its ligand evokes the release of the gonadotrophic hormones, FSH and LH, which results in stimulation of the steroidogenic and gametogenic functions of gonads in both sexes (Millar, 2004).

Given the importance of GnRH and its receptor in the control of reproduction, DNA sequencing was carried out to determine the structure of the gene encoding the bovine GnRH receptor, and genetic correlations were determined for associations between single nucleotide polymorphisms identified in the gene and PTAs for fertility traits. The findings of this study are presented in Chapter 4.

1.4.2.2 Luteinizing hormone receptor *(LHR)*

The luteinizing hormone receptor (LHR) is a 7-transmembrane domain G-protein coupled receptor expressed in the ovary, testis and uterus. The LHR plays a key role in the gonads: in the ovary, the LH receptor was detected in theca cells, differentiated granulosa cells, and luteal cells; in the testis, it is found exclusively in Leydig cells (Ascoli et al., 2002). The expression of LHR in granulosa cells is critical to the establishment and maintenance of the dominant follicles (Xu et al., 1995) and may also allow the follicles to ovulate and luteinize (Inskeep et al., 1988). The LHR is also expressed in extragonadal tissues such as the uterus, sperm, prostate, adrenals,

and rat brain (Ascoli et al., 2002), although in most cases the significance of extragonadal expression is unclear. The activation of the *LHR* through binding to its ligand, LH, stimulates the production of androgen, oestrogen and progesterone. Thereby, LH triggers ovulation and maintains progesterone production by the corpus luteum in females and increases testosterone production in males.

DNA sequencing revealed that the *LHR* contains 10 introns and 11 exons, exons 1 to 10 coding for the N-terminal extra-cellular domain and exon 11 for the rest of the molecule (Themmen & Huhtaniemi, 2000). Inactivating and activating mutations in the *LHR* have been identified as the cause of several endocrine diseases in human, for example the first activating point mutation described in males caused early onset of gonadotrophin-independent precocity. In addition, around 20 inactivating mutations have been discovered in the *LHR* ranging from partial to complete inactivation causing in males a wide range of phenotypes from mild undervirilisation to complete failure of masculinisation (Huhtaniemi, 2002). Bovine *LHR* gene sequencing revealed that there are three SNPs in the coding region of the gene. These are: a mis-sense mutation (TGG→TGT) identified at codon 467 causing an amino acid substitution from tryptophan to cysteine; a silent mutation (CTC→CTT) at position 490 leaving the leucine residue unchanged; and a mis-sense mutation (CAG→CAT) at position 527 changing the residue from glutamine to histidine. These SNPs were present in 4 haplotypes which were found to be related to variation in the fertility traits, CI, DFS, as well as in the production index (PIN). In particular, the effect of the haplotype TCT was associated with a decrease in CI, DFS and PIN, but had no significant effects on NR56, CINS and BCS (Hastings et al., 2006).

1.4.2.3 Follicle stimulating hormone receptor *(FSHR)*

As described above, FSH plays a central role in mammalian reproductive function. In the ovaries, FSH stimulates follicular maturation and oestrogen biosynthesis in granulosa cells, and together with LH regulates ovulation. In the testis, it stimulates a range of functions in Sertoli cells which consequently provide physical and biochemical support for spermatogenesis (UIIoa-Aguirre & Timossi, 1998). Like its partner gonadotrophin LH, FSH is released by the pituitary gland and acts by binding

to specific receptors located mainly on the surface of Sertoli cells in the testis and granulosa cells in the ovary. Together with the LH receptor, the FSH receptor belongs to the family of G-protein-coupled receptors, displaying a large extracellular domain specialized in the binding of the hormone (Simoni et al., 1997). The *FSHR* gene consists of 10 exons and nine introns. Exons 1-9 encode the extracellular domain, while exon 10 encodes the transmembrane and the intracellular domains (Nordhoff et al., 1999). Only one activating mutation in the *FSHR*, asp567gly, has been described in a hypophysectomised male who had normal spermatogenesis in the face of undetectable levels of gonadotrophins (Huhtaniemi & Themmen, 2001). On the other hand a small number of inactivating mutations have been identified in the *FSHR* gene, and these have important effects. Inactivating mutations in *FSHR* in women caused symptoms ranging from FSH-responsiveness amenorrhoea to total suspension of follicular development including suspended pubertal maturation, depending on the completeness of the inactivation. Moreover, inactivating mutations in men cause a decline in testicular size and a reduction in the quality and quantity of spermatogenesis without causing azoospermia (Huhtaniemi, 2002).

1.4.2.4 Progesterone receptors *(PR)*

The steroid hormone progesterone plays a central role in the reproductive events associated with ovulation, luteinisation, pregnancy establishment and maintenance, in addition to its role in controlling the proliferation, differentiation, and development of mammary and uterine tissues (Lydon et al., 1995). The physiological effects of progesterone hormone are mediated through binding to specific intracellular progesterone receptors (*PRs*) that are expressed in two isoforms, *PR-A* and *PR-B* (Mulac-Jericevic & Conneely, 2004). Both isoforms, which are members of the nuclear receptor superfamily of transcription factors, differ only in their N-terminal ends, *PR-B* containing 164 amino acids more than *PR-A*. Both isoforms are encoded by the same gene using two different promoters and transcription factors (Vegeto et al., 1993). Although PR-A and PR-B share several structural domains, their expression varies in reproductive tissues due to the distinct regulation of their own progesterone-dependent target genes (Mulac-Jericevic & Conneely, 2004).

Genetic effects of the PRs have been studied using gene knockout techniques. Transgenic adult female mice carrying null mutations of the *PRs* were infertile. This reflected numerous abnormalities of the ovary, uterus, and mammary glands as well as impaired sexual behaviour, suggesting that the PR plays a role as a transcription factor with pleiotropic effects that are required for normal structure and function of the reproduction system (Lydon et al., 1995). *PR-A* knockout mice displayed a normal mammary response to progesterone but suffered from severe uterine hyperplasia and ovarian abnormalities. In contrast, *PR-B* knockout mice showed normal biological responses of the ovary or uterus to progesterone but displayed reduced pregnancy-associated mammary gland morphogenesis (Mulac-Jericevic et al., 2003). Many SNPs have been identified in the human progesterone receptors; two were identified in the coding region at positions S344T and G393G, and two in the promoter region +44 C/T, +331 G/A. The latter was reported to be associated with risk of human endometrial cancer by increasing expression of *PR-B* (De Vivo et al., 2002).

1.4.2.5 Oestrogen receptors *(ESR)*

Oestrogen receptors, in common with other nuclear receptors, are transcription factors, which after binding to their ligand (oestrogen) are capable of regulating gene expression (Kuiper *et al.*1996). In mammals, oestrogen regulates many vital processes including the development and functioning of the reproduction system in both males and females, including the mammary gland, ovary, vagina, oviduct and uterus, testis, epididymis and prostate. In addition, oestrogen plays an important role in many other parts including liver, bone, pituitary and cardiovascular cells (Gustafsson, 1999).

Two isoforms of the oestrogen receptor, α and β are known; each is encoded by a separate gene and localised on different chromosomes 6 and 14 respectively (Gustafsson, 1999). ESRα was the only identified oestrogen receptor until the discovery of specific oestrogen binding sites that can still be observed in an ESRα knockout mouse (ESRKO). Apart from the abnormalities noticed in the reproductive organs and some fertility problems in both male and female homozygous mutants,

there were no obvious problems in prenatal sexual development. These results were unexpected as oestrogen is well known to play a key role in breast and uterus development in addition to its effects in cardiovascular disease and in preventing bone loss after the menopause (Lubahn et al., 1993). This suggested the existence of a second receptor for oestrogen (*ESRβ*) which was initially cloned and isolated from rat prostate and ovary DNA. *ESRβ* was found to be highly homologous to *ESRα* particularly in the DNA binding domain (95% amino acid identity). There was less homology in the ligand binding domain (only 55% amino acid identity), leading to ligand selectivity between the receptors (Kuiper et al., 1996). Moreover, ESRβ was located in the central nervous system, the cardiovascular system, urogenital tract, kidney and lungs. ESRα was also localized in the ovary but at a lower level than ESRβ and with no characteristic cellular localization. However, ESRα is the dominant receptor in the mammary gland and uterus (Gustafsson, 1999).

Many single mutations within the human *ESRα* sequence have been reported, many being silent mutations. However, some of the mutations identified in *ESRα* change the protein sequence and have been found in a variety of tissues and diseases, including breast cancer, endometrial cancer, and cardiovascular disease. Few point mutations have been identified in the *ESRβ*, none of which was associated with the phenotypes analyzed (Herynk & Fuqua, 2004).

Due to the numerous functions of oestrogens, the oestrogen receptors are considered candidate markers for production and functional traits in farm animals. However, few studies have been carried out on *ESR* polymorphisms in farm animals. In 1996 Rothchild et al proposed the *ESR* gene as a candidate gene for prolificacy in pigs (see section 1.3.2.1). A SNP was identified in the *ESR* which was found to be significantly associated with the number of piglets borne alive.

The above mentioned steroid hormons (estrogen and progestron) exert their effects through non-transcriptional pathway which is highlighted in the next section.

- **Non genomic action of steroid hormones**

In addition to the traditional action of steroid hormones through the regulation of transcriptional process involving nuclear receptors, a great body of evidence has

been accumulated about the non-transcriptional mechanism of signal transduction through steroid hormone receptors (Simoncini and Genazzani, 2003). These effects on cellular function were explained by the presence of signal-generating steroid receptors on the cell membrane and have been called non-genomic steroid effects (Falkenstein et al., 2000). Non-genomic steroid effects have been characterized as being very rapid (acting in seconds or minutes) and not to be compatible with the involvement of RNA or protein synthesis, as they can be observed in highly specialized cells which do not carry out mRNA and protein synthesis or even in cells lacking steroid nuclear receptors. Non-genomic actions can also be observed with steroids bound to high molecular weight molecules which are unable to pass through cell membranes, and therefore do not enter target cells, These effects cannot be blocked by inhibitors of mRNA and protein synthesis or by antagonists of the classic genomic steroid receptors (Revelli et al., 1998). The integrated actions of steroid hormones through genomic and non-genomic mechanisms have been demonstrated to be implicated in a variety of physiological and pathophysiological processes (Simoncini and Genazzani, 2003).

1.4.2.6 Activin receptors *(ACTR)*

Activins are dimeric proteins composed of disulphide-linked activin polypeptide β subunits (activin-A {βA βA}, activin-AB {βA βB}, activin-B {βB βB}) that belong to the transforming growth factor β (TGF-β) superfamily (Chen et al., 2006). Activins were initially extracted from gonadal fluids on the basis of their role in the ovarian-pituitary axis, particularly in stimulating FSH release from pituitary. Activins are expressed also in the granulosa cells of developing follicles, ovarian theca cells, granulosa–lutein cells of the corpus luteum, placental trophoblast, testicular Sertoli and Leydig cells, in addition to many non-reproductive organs such as the adrenal cortex, bone marrow, anterior pituitary and brain (Knight, 1996). Therefore, activins have been described as multifunctional proteins due to their roles in different biological activities such as cell proliferation, differentiation, apoptosis, metabolism, homeostasis, immune response and endocrine function (Chen et al., 2006).

Activins express their effects through binding to two types of specific receptors, type I (ACTRI) and type II (ACTRII). Both activin receptors are transmembrane proteins with ligand binding activity in the extracellular domain and serine/threonine kinase activity in the intracellular domain. The activin type II receptors which comprise two subtypes called ACTRIIA and ACTRIIB are the primary ligand-binding proteins, which bind activin with high affinity and subsequently phosphorylate and activate ACTRI. Subsequently, the activated ACTRI phosphorylates and activates Smad(s) proteins, which accumulate in the nucleus, binding to the promoter region of the target genes, and regulate their expression (Chen et al, 2006). In addition to activin receptors, the actions of activin are also modulated by follistatin which binds activins with high affinity, blocks the interaction between activins and their receptors, and consequently inhibits the signalling activity of activins (Welt et al., 2002).

Activins are implicated in a variety of reproductive processes such as embryonic development and folliculogenesis. In rat and sheep, exogenous activin A administration results in an increase in FSH release. Activin A is produced at a high level around the time of luteo-follicular transition associated with a high level of FSH, and this is important for the onset of the follicular development for the next cycle. In addition, activin plays important roles in granulosa cell differentiation and steroidogenesis (Knight, 1996). Moreover, activins have significant effects on the growth of small follicles and on the stimulation of granulosa cell proliferation alone and in combination with FSH. Activin A knockout mice exhibit reduced fertility while mice lacking both activin A and activin B are sterile and both have shown increased numbers of functional number of CLs (Pangas et al., 2007). Moreover, activin A plays a functional role in the developmental competence of bovine embryos and regulates the development of the preimplantation embryo. Adding activin A to culture medium increases the development rate of bovine embryos produced *in vitro* to the blastocyst stage (Yoshioka et al., 1998).

1.4.2.7 Neuropeptide Y *(NPY)*

Neuropeptide Y (*NPY*) is a 36-amino acid neurotransmitter peptide which plays a fundamental role in the neural regulation of feed intake, EB (Miner, 1992), and reproductive hormone secretion (Bauer-Dantoin et al., 1992; Crowley & Kalra, 1988). Many studies have implied the involvement of *NPY* neurons in the generation of the preovulatory LH surge. This action occurs at two levels, the hypothalamus and pituitary, in the reproductive axis. NPY significantly stimulates the release of hypothalamic GnRH *in vitro* in ovariectomized (OVX) and oestrogen-pretreated rats. It also potentiates GnRH-induced LH secretion directly from the pituitary in OVX and oestrogen-pretreated rats but not in OVX and oestrogen untreated rats. The dual actions of *NPY* at both sites has suggested its possible role in mediating ovarian feedback signals for generating preovulatory GnRH and LH surges (Crowley & Kalra, 1988). This action has been supported by findings of a significant increase in NPY gene expression in the hypothalamus in association with preovulatory LH surge during proestrous in rats; this increase was not observed during metestrous. A high level of NPY is accompanied by a high level of oestrogen, suggesting the possible involvement of oestrogen in the preparation of the pituitary gland for the action of GnRH partially through increasing its sensitization to the effects of NYP (Bauer-Dantoin et al., 1992). However, the stimulatory effect of NPY on GnRH/LH release observed in the rat was not consistent with its effects in ewes. NPY administrated centrally in OVX and oestrogen-pretreated ewes resulted in either a delay or a complete blockage of the preovulatory LH surge, indicating the effects of NPY on the LH surge are predominant negative. These results were confirmed since the highest level of NPY mRNA expression during oestrous have been detected during the luteal phase when gonadotrophin secretion is at its lowest level. This effect on the LH surge has been proposed to be through blockade of the positive-feedback action of oestrogen (Estrada et al., 2003).

Furthermore, OVX- cows showed a decline in LH level associated with an increase in growth hormone (GH) secretion when injected with NPY. These results were in agreement with the reported elevated *NPY* expression accompanied by low LH and high GH levels during feed restriction and poor body condition in other species

(Thomas et al., 1999). In dairy cattle, NEB results in a delay in the time of the first ovulation through inhibition of LH pulse frequency (Butler, 2000). These results suggest the possible involvement of NPY in the mechanism by which undernutrition impairs fertility and might also explain the negative association between reproduction performance and body condition in cattle (Thomas et al., 1999). One of the possible mechanisms by which NPY acts on hypothalamic neurons might be by transferring signals on nutritional status to the growth and reproductive axes (Thomas et al., 1999). In rats and sheep, feed intake significantly increases within 30 minutes after central NPY administration. However, peripherally injected NPY does not affect feed intake, suggesting its central modulation of feed intake particularly within the hypothalamic nucleus (Miner, 1992).

The action of NPY is mediated through its Y-receptors which belong to the G protein-coupled receptor superfamily. Six subtypes of NPY receptors were identified (Y1-Y6) based on their affinities for agonists and later by cloning and characterization of five of the receptor subtypes (Blomqvist & Herzog, 1997). The effect of NPY in rat hypothalamic tissues is mediated throughY1 receptors (Crowley & Kalra, 1988). However, injection of Y1 receptor antagonist in ewes does not change the onset of the oestrogen-induced LH surge, excluding the Y1 receptor from involvement in LH regulation in the ewe. The inhibitory effect of NPY on the reproductive axis in ewes was reported to be through the Y2 receptor, while the Y1 receptor was reported to affect food intake (Clarke et al., 2005). Meanwhile, the effects of NPY on increasing feed intake and inhibiting oestrous behaviour in Syrian hamsters has been reported to be through different receptors subtypes, namely Y5 and Y2 respectively (Corp et al., 2001).

Three SNPs have been identified in the bovine NPY gene: 2 A/G SNPs, and 1 T/C SNP. These SNPs have significant effects on growth and body weight with AAC haplotype carriers acquiring better growth and heavier body weight. These SNP carriers also showed a tendency for better feed efficiency. None of these SNPs affects daily dry matter intake, which is not consistent with NPY as an appetite stimulator (Sherman et al., 2008).

1.4.2.8 Leptin

Leptin is a 176-amino acid protein mainly produced and secreted by adipocytes. Many studies have demonstrated that leptin is expressed in some other tissues such as placenta, ovaries, mammary gland, brain, pituitary and stomach (Friedman & Halaas, 1998). Leptin expression is mediated by many hormones such as insulin, growth hormone and glucocorticoids. As a result, leptin is involved in a variety of physiological processes such as the control of food intake and energy expenditure, regulation of the onset of puberty, fertility, and during pregnancy (Margetic et al., 2002). The leptin receptor is a glycoprotein with a single transmembrane spanning region of which six isoforms were found in different tissues. These isoforms originate from a single gene and share identical extracellular and transmembrane domains, but with variable lengths of intracellular domains (Liefers et al., 2005a). Leptin receptors are distributed in many organs including liver, kidneys, lungs, small intestine, testes, ovaries, placenta, pancreas and adipose tissue (Margetic et al., 2002).

Leptin plays an important role in reducing food intake. Leptin administration in mice lacking leptin (homozygous for mutation in the *ob/ob* gene) results in a decline in food intake and body weight (Friedman & Halaas, 1998; Margetic et al., 2002). Liefers et al. (2003) found that during pregnancy in dairy cows leptin levels were high and declined to a nadir at parturition. A high level of leptin during pregnancy was not associated with a decrease in food intake or metabolic efficiency (Margetic et al., 2002). It has been speculated that the increase in the level of adiposity and in the expression of leptin in adipose tissue might have contributed to this increase in leptin levels during pregnancy. This increase in leptin level seems to be contradictory to what is known about leptin as a satiety factor for which many mechanisms have been proposed. During pregnancy the increase in concentration of the soluble form of leptin receptor may have resulted in an increase in the bound inactive form of leptin causing leptin resistance, assuming that leptin-binding protein inhibits leptin signalling. Moreover, a decrease in the long-active form of the leptin receptor at hypothalamus level during pregnancy also leads to fewer signals being sent to reduce food intake (Margetic et al., 2002; Liefers et al., 2005a; Liefers et al., 2003). The

decline in leptin level towards parturition might be due to a change in energy metabolism before lactation (Liefers et al., 2005a).

Furthermore, during early lactation, cows experience conditions of NEB due to the increased demands on fat stores for lactation, body maintenance and growth, while reproduction receives the lowest priority. Cows in NEB express lower plasma leptin concentration, less food consumption, and lower live weight in comparison with cows in positive EB. Therefore, leptin level can be considered as a sensor of EB in cows during lactation. Moreover, during the postpartum period the circulating leptin level is negatively correlated with the time to first observed oestrus (FE), with no relationship with first postpartum luteal activity (FPLA). A difference of 40 days was found between mean FPLA and FE with a correlation of 0.26 implying a possible relationship between leptin level and oestrous expression (Liefers et al., 2003). NEB suppresses LH pulse frequency, resulting in a delay in the first ovulation (Butler, 2000). Short-term fasted heifers showed a considerable reduction in the frequency of LH pulses and this effect was antagonised by leptin administration, which was reported to act at the hypothalamic level to maintain GnRH secretion (Maciel et al., 2004).

Furthermore, leptin was reported to have positive effects on fertility; for example, *ob/ob* mice are sterile. Administration of leptin in leptin deficient (*ob/ob*) mice results in activation of the reproductive axis and in the restoration of fertility. Leptin-treated females have a significant elevation in the serum LH level and an increase in ovarian and uterine size. Males also experience a significant increase in FSH level associated with an increase in testicular size and sperm counts. Furthermore, leptin administration positively affects the secretion of GnRH, LH, and FSH (Margetic et al., 2002). Leptin has also been demonstrated to be a signal that informs the brain whether fat stores are sufficient to support the high energy requirements of reproduction and may also accelerate the onset of puberty in normal female rodents (Margetic et al., 2002). Leptin then stimulates GnRH release from the hypothalamus leading to the onset of puberty (Maciel et al., 2003).

Many SNPs have been identified in the bovine leptin gene. In the promoter region, 3 of the 20 identified SNPs (-147, -197, and -282) explain most of the variation in prepartuim leptin concentration. A combination of these SNPs can be used to detect leptin roles during pregnancy. Moreover, SNPs at positions -963, -1457 and -578 were found to be associated with EB, FPLA and protein percentage respectively. These SNPs can be used as genetic markers in the genetic selection for better fertility and EB without having a significant impact on milk yield (Liefers et al., 2005b). Three SNPs in the promoter region at positions -2470, -1239 and -963 in addition to one SNP in the coding region (tyr7phe) were significantly associated with milk production traits. However, the SNP at position -1457 had shown no association with calving interval or survival in dairy cattle (Giblin et al., 2010). Another missense mutation C>T was identified on exon 20 in the bovine leptin receptor, and carriers of this SNP have significantly lower leptin levels during late pregnancy (Liefers et al., 2004).

1.4.2.9 Ghrelin receptor (*GHS-R)*

Ghrelin is a 28-amino acid peptide, recently identified as the endogenous ligand for the growth hormone secretagogue receptor (GHS-R). Ghrelin derived from both hypothalamus and stomach was found to regulate pituitary growth hormone (GH) secretion in a distinct dose-dependent manner (Kojima & Kanjawa., 2005). Ghrelin was also found in the pituitary gland where it acts as an autocrine or endocrine regulator of GH release (Garcia et al., 2007). GHS-R is a G protein-coupled receptor which is expressed at high levels in the hypothalamus and pituitary. Moreover, GHS-R expression has also been found in testis, ovary, pituitary, heart, lung, liver, kidney, pancreas, stomach, adipose tissue, and immune cells. Therefore, ghrelin, through operating at different levels, was reported to be involved in many biological functions including energy homeostasis and reproduction (Kojima & Kanjawa., 2005).

Ghrelin expression in the cyclic and pregnant rat ovary changes throughout pregnancy and during oestrus, with the lowest level during proestrus and maximum values in dioestrus. This cyclic change of ghrelin expression with the maximum level

detected during the luteal stages in parallel with progesterone profile suggests a possible effect in the development and function of the corpus luteum in rat (CL; Caminos et al., 2003). During the course of pregnancy ghrelin was detected at its highest level in an early stage and it decreased gradually during the latter half of pregnancy. The function of the CL required for maintaining pregnancy during the first week of gestation is regulated by LH and prolactin (PRL), while the second half of pregnancy is regulated both by the action of lactogen and androgen produced by the placenta, and by oestrogen and progesterone produced by CL. Therefore, it has been suggested that ovarian ghrelin expression during early pregnancy is modulated by the action of LH and prolactin while placental lactogen is primarily involved in placental ghrelin expression during the latter half of pregnancy (Caminos et al., 2003). Moreover, ghrelin expression in testis is highly selective for Leydig cells and under the hormonal control of pituitary LH. LH administration raises levels of testicular ghrelin RNA. Conversely, ghrelin inhibits pulsatile LH in female rat, and testosterone secretion in male rat. Therefore, elevated ghrelin levels may result in suppression of the reproductive axis (Garcia et al., 2007).

Ghrelin expression in the hypothalamic nucleus, which has important effects on food intake, is highly suggestive of a possible role of ghrelin in food intake. Peripherally or centrally injected ghrelin in freely-fed rats results in an increase in food intake and GH secretion (Wren, et al., 2000). Tschöp et al. (2000) showed that continuous central administration of ghrelin results in a dose-dependent increase in food intake and body weight. In addition, there is a fluctuation in plasma ghrelin associated with feeding status, increasing during fasting and decreasing after feeding. In beef cattle, Wertz-Lutz et al. (2006) suggested that ghrelin plays an important role in feeding behaviour and energy homeostasis. An increase in plasma ghrelin concentrations was observed prior to feeding in cattle, followed by a postprandial decrease. Roche et al. (2006) reported a positive relationship between genetic selection for increased milk production and plasma ghrelin concentration, dry matter intake, and GH level. Therefore, NEB in lactating cows was proposed to stimulate ghrelin and GH secretion (Bardford & Allen, 2008). Taking into account the impact of nutritional status on reproduction prompted the idea that ghrelin may operate at central and peripheral levels as a key signal for energy status to the reproductive axis.

Collectively, ghrelin was recognized as a potent orexigenic agent by acting at central endocrine levels, providing a link between the stomach, hypothalamus and pituitary and playing an important role in EB through stimulating food intake and adiposity (Garcia et al., 2007).

1.4.3 Conclusion

The decline in fertility in modern dairy cattle is complex and many factors have been suggested to have an effect on this trait (Figure 1.4). Most of these factors can be categorized as pathological (mastitis, lameness), physiological (high milk yield, reproduction hormones), and managerial (nutrition, oestrous detection). So, different genes that play a role in these pathways may be expected to have an impact on fertility, resulting in hundreds of genes potentially being involved in reproduction in dairy cattle (which is clearly a polygenic trait). In addition, there are complicated interactions among these factors and many of these genes have multiple effects. For example, the incidence of lameness is associated with lower oestrus expression and lower fertility (Dobson et al., 2008). On the other hand, high producing cows were found to be more susceptible to lameness and this was associated with the low BCS in these cows (Espejo et al., 2006). Furthermore, nutritional status of the cows, which is affected by management and genetics, has major effects on fertility in dairy cattle. For instance, cows overconditioned at calving or in a state of undernutrition early lactation will develop more sever NEB and subsequently lower fertility through reducing LH and FSH pulse frequency (Butler and Smith, 1989; Butler, 2003). Many genes that have direct effects on feeding behaviour also have significant effects on fertility (leptin, ghrelin etc.). The effects of some of the genes that affect fertility are investigated in more detail in the following chapters.

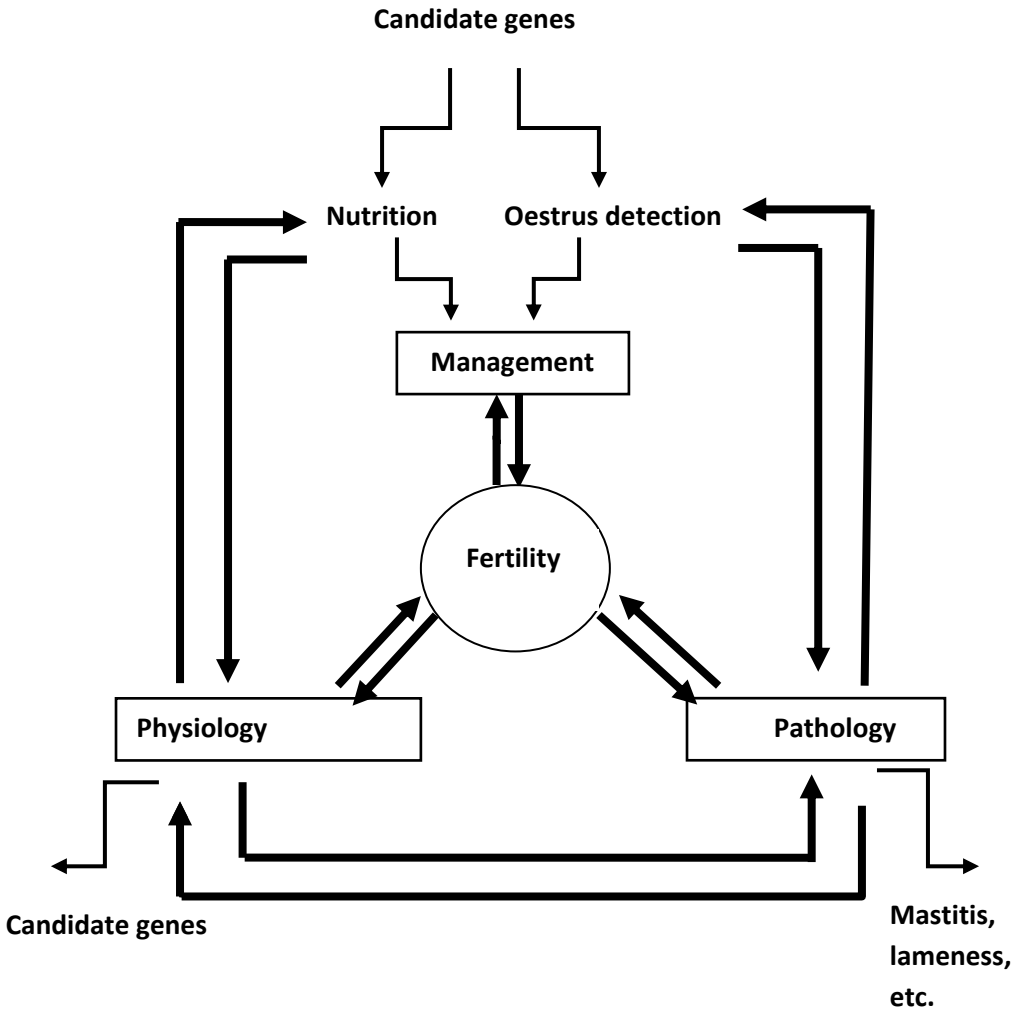


Figure 1.4 Different factors affecting fertility in dairy cattle.

Selection indices are used to describe the magnitudes of genetic effects of the kind described above (which will be covered exensively in the following chapter). This approach has been used successfully in the dairy industry, as described in the following section.

1.5 The UK fertility index of dairy cattle

To establish the UK fertility index for dairy bulls (which consists of the predicted breeding values of the bulls that are estimated depending on their daughters performance for the various traits of interest), databases of different fertility traits including management traits (calving interval, traits derived from insemination

records) and traits genetically correlated with fertility such as production traits and body condition score were used to estimate breeding values (Flint et al., 2003).

As reproductive performance is an important determinant of dairy production efficiency, many studies have been carried out to investigate trends in fertility traits in high producing cows and to derive genetic correlations among the various fertility and milk yield traits. This is a necessary pre-requisite for the establishment of the fertility index.

1.5.1 Data available in the UK for fertility index calculation

In the UK dairy industry recording of fertility information is relatively poor as there is a requirement only to record calving dates, and participation in a recording scheme is voluntary (Kadarmideen & Coffey, 2001; Flint et al., 2003). The uncertainty surrounding fertility data can be overcome by using information available on other traits which have higher heritability and are genetically correlated with fertility traits, such as milk yield (Wall et al., 2003; Royal et al., 2002a), condition score and linear type traits (Pryce et al., 2000; Royal et al., 2002b). To generate fertility indexes, a database was compiled from data held by National Milk Records plc, Cattle Information Services Ltd and Holstein UK and other sources (Kadarmideen & Coffey, 2001). After the evaluation of these databases and the establishment of editing rules where possible, four fertility traits (calving interval (CI), days in milk to first service (DFS), number of inseminations (CINS), and non-return rate after 56 days of conception (NR56), also called management traits) and two other correlated traits, milk yield at 110 days of lactation (MY) and body condition score (BCS) were used in the analysis (Wall et al., 2003). These six traits were used for establishing the fertility index but each has advantages and disadvantages:

1.5.1.1 Calving interval

Calving interval (CI) is defined as the number of days between two consecutive calvings. CI is an economically important trait which is recorded reliably in the UK, as recording CI is a legal requirement. On the other hand, for CI to be recorded cows must have undergone two consecutive calvings. Therefore, depending on CI as a

selection tool is limited as CI data will not be available when making decisions by breeding companies for breeding young bulls. Furthermore, CI records are only available for cows selected to have a second calving as many cows are culled for subfertility, and therefore the data are not randomly censored. In addition CI data are open to management bias in cases where the breeders delay insemination for high yielding cows (Flint et al., 2002, 2003). However records on relatives (mothers and sisters) can be used to estimate a bull's proof (a bull has a "proof" when his daughters have been assessed for fertility, so from the process of proving a bull, fertility PTAs will be obtained which then can be listed in an index) before his daughters have a second calf.

1.5.1.2 Insemination data

In an analysis of insemination data in the UK it was revealed that only 10% of herds that take part in recording schemes have all service dates; 7% had only 40-50% of the information, while 15% failed to record any service dates (Kadarmideen & Coffey, 2001). Many insemination events were not reported to milk recorders by producers due to the long period between recorders' visits which will allow time for more than one insemination, of which only the last is usually reported as being important in the milk recording scheme. Moreover, there is a clear tendency for farmers to report only the insemination which has resulted in a significant output such as conception while ignoring those that do not. In part this results from the NMR form for recording insemination dates, which has sufficient space only for one date to be entered; as a cow may have 2 inseminations between a recorder's monthly visits, the earlier date is likely to be lost. As a result of the special nature of insemination data, careful editing is required in order to be able to include it in the fertility index, and methods for carrying this out were developed by Kadarmideen & Coffey (2001). Different insemination traits were derived from the insemination records. The most important traits are:

DFS: DFS has been shown to have the highest heritability among insemination data (h^2 = 0.035; Wall et al., 2003) with a strong and positive correlation with CI. DFS is available earlier than CI and therefore it can be used to obtain information related to

CI before CI can be estimated. DFS data are compiled from a single insemination record, reducing the potential errors in recording the trait while all the other insemination data need more than one record. On the other hand, the insemination required to be recorded is the first after calving, and where data is lost, as suggested above, it is likely to be the earlier insemination date. This trait is also open to management bias as in the case of CI, such as delaying insemination in high yielding cows. CI and DFS data provide information on success of conception to first service which gives a reference point for all fertility traits (Flint et al., 2002).

NR56: NR56 identifies the pregnancy status of a cow 56 days after insemination and reflects the ability of a cow to hold a pregnancy over a period of early gestation when pregnancy loss is more likely to occur. NR56 reflects a physiological process as the success rate of conception following a timely insemination is high, and most pregnancy loss occurs during the first two months. NR56 has a low correlation with yield, and therefore adds information to the index. However, the heritability of this trait is low (h^2 = 0.018; Wall et al., 2003) so data on second calving are needed for better interpretation of insemination data when the quality of records is poor. On the other hand, adding this trait to the fertility index allows UK index data to be compared across countries, where NR56 is internationally recorded. NR56 is coded 1 if a cow did not return to service 56 days after first insemination and 0 otherwise (Flint et al., 2003).

1.5.1.3 Milk yield

Milk yield has moderate heritability (h^2 = 0.33; Wall et al., 2003) and a high genetic correlation with physiological traits of fertility such as commencement of luteal activity after calving (CLA) and the occurrence of persistent corpora lutea as well as the traditional measures of fertility (Royal et al., 2000a, Royal et al., 2002a). There are a number of traits that can be derived from milk yield such as total yield per conception (305 day yield), or yield at a definite time during lactation such as milk yield at day 50 (close to the time of the first insemination after calving), or yield at day 110 (close to the time of the first conception), milk yield at the third recorded test (peak yield), or at the time of condition score measurement (Brotherstone at al.,

2002). An early inclusion of yield data in the index required measuring yield early in lactation rather than total lactation yield, which cannot be readily used until the end of lactation.

1.5.1.4 Body condition score (BCS)

BCS was revealed to be a useful indicator of energy balance (EB), as the decline in energy efficiency obtained from food intake during lactation affects live weight (Butler & Smith, 1989). BCS has a moderate heritability (h^2 = 0.237; Wall et al, 2003) with a high and unfavourable genetic correlation with milk yield indicating sires with higher breeding value for milk yield are more likely to have daughters with lower BCS (Pryce et al., 2000; Wall et al., 2003). High yielding cows are vulnerable to NEB especially in early lactation which confirms the negative relationship between yield and EB, as cows metabolise body tissue to maintain milk production (Butler & Smith, 1989; Pryce et al., 2000). BCS is a valuable trait for fertility prediction as it can be measured early in lactation, during which it is more highly correlated with CI than at any other stages of lactation; thus cows that are thinner are expected to have longer CI (Pryce et al., 2000). In addition, BCS has a high genetic correlation with physiological measures of fertility derived from milk progesterone such as CLA, with cows in strong NEB experiencing poorer fertility and longer interval to CLA (Royal et al., 2002b). BCS observed during the first lactation is recorded in the field on a scale of 1 to 9, where 1 = thin and 9 = fat for animals used in the classification scheme operated by Holstein UK (Royal et al., 2002b).

1.5.1.5 Linear type traits

Linear type traits such as chest width, angularity etc., are applied in many breeding schemes in the UK as they provide a good indicator of body condition and live weight. Linear type traits are genetically correlated with CI and are a good indicator of CLA (Pryce et al., 2000; Royal et al., 2002b). More angular, thin-chested cows tend to have longer CI (Pryce et al., 2000) which was also supported by the unfavourable relationship between CLA and PTAs for linear type traits (Royal et al., 2002b).

1.5.2.1 Calculation of genetic parameters and predicted transmitting ability

As there are many factors contributing to the change in fertility traits (genetics, management and environment), the challenge in calculating fertility parameters is to separate the genetic effects from those due to management. This could be achieved through utilizing information from pedigrees which identifies the proportion of the differences observed among families due to parentage (Flint et al., 2002). The correlations between the different fertility traits and milk production have been demonstrated to be unfavourable. Furthermore, the heritability of some fertility traits (management traits) is relatively low in comparison with other physiological traits such as CLA as they are affected by management decisions (Royal et al., 2002a, Wall et al., 2003). The heritabilities of management traits range between 0.018 and 0.035 (Wall et al., 2003) while a physiological traits such as CLA has a heritability ranging between 0.16 and 0.23 (Royal et al., 2002b). Nonetheless, despite a low heritability, there is substantial genetic variation among bulls for their daughters' fertility traits, indicating that there is room for genetic improvement (Wall et al., 2003). The selection index makes use of these variations in helping producers to choose bulls with daughters having both good fertility and high milk yield, while avoiding bulls with poor fertility PTAs (Flint, 2006). Currently, only management traits are included in the fertility index in the UK, and this may affect the reliability of fertility proofs. Therefore, due to the low quality of data recording, large numbers of animals are required to be included in the evaluations, and this can be done as large numbers of records are available through the milk recording companies.

PTAs for the following 6 traits were calculated for all UK bulls being used since 1985. This calculation involved estimation the heritabilities of the traits in addition to the genetic relationships between them (Wall et al., 2003; Table 1.4).

Table 1.4 Genetic correlations between fertility and production traits

	CI	BCS	MILK	DFS	NR56	CINS
CI	**0.033**	-0.14	0.27	0.67	-0.45	0.61
BCS		**0.237**	-0.44	-.063	-0.30	0.29
MILK			**0.329**	0.49	-0.25	0.06
DFS				**0.037**	0.24	-0.12
NR56					**0.018**	-0.94
INS						**0.020**

- Heritabilities are on the diagonal.

The genetic parameters obtained from this evaluation were then used to estimate the PTAs for each bull and for each trait applying BLUP methodology, in which a sire-maternal grandsire model was used (Wall et al., 2003). In this model, the A matrix incorporated in the MME includes only the sires of the sire and maternal grandsires, therefore only the breeding values for the sires and maternal grandsires are estimated (Mrode, 2005). Herd-year-season, month of calving, and age at calving were fitted as fixed effects while the effect of sire was fitted as a random effect. During the estimation of milk yield and BCS, the stages of lactation at test and classification have to be considered respectively. Different units were used to estimate the fertility traits, with NR56 given as a percentage, CINS as a number and CI, DFS in days. In addition, as fertility traits are changing with time, fertility proofs relate to a specific time. The estimated ranges of the PTAs for CI, DFS, NR56, and CINS in 95% of the bulls were 10 days (-5 to +5d), 7 days (-3 to 4d), 0.14% (-7% to +7), and 0.3 for CINS (-0.15 to +0.15 insemination) respectively (Figure 2.2; Wall et al., 2003).

Moreover, a conversion formula was used for appropriate inclusion of foreign bull proofs in the UK fertility index. The bulls can then be ranked depending on their PTAs for each fertility trait. These PTAs can then be combined into an index which provides an accurate indicator of the best bulls for both fertility and milk yield (Flint et al., 2004).

1.5.2.2 Calculation of economic values

It is well known that producers tend to be more concerned about their profits when choosing bulls for breeding, as considering fertility could add an additional cost in terms of reducing milk yield. For this reason, and also in order for the fertility index to be practical to use, it should be included in a more comprehensive profit index. This requires identification of the economic costs of the traits involved. As these traits are correlated with each other, it is necessarily to avoid overlapping cost estimations (i.e. double counting). The profit lifespan index used in the UK (£PLI) reflecting yield and lifespan already accounts for some costs associated with subfertility, as for instance the effect of culling on lifespan includes culling for subfertility. After identifying the economic value of fertility traits in pounds sterling, the fertility index can be readily incorporated into £PLI to provide £PLIF which includes the cost of fertility. In order for the selection index to be simple and to be valid internationally, the economic cost values are published in terms of CI and conception rate (NR56). The economic values for these traits were agreed on introduction of the index to be -£0.31 per each additional day of calving interval and £1.56 per one percent increase in conception rate. At its introduction the final formula by which the fertility index was calculated was as follows:

FI= (PTACI X -£0.31) + (PTA NR56 X £1.56) (Flint et al, 2004).

1.5.3 Future improvement of the index

1.5.3.1 Milk progesterone

As discussed earlier, the traits used to construct the FI are open to management bias which explains their relatively low heritability (h^2 = 0.018-0.037; Wall et al., 2003). Therefore, in the long term it will not be a sustainable solution to improve fertility by traditional phenotypic selection approaches as they do not reflect the actual physiological characteristics of the cows. So in order for the selection index to be more effective in improving fertility and identifying the most fertile animals, management and environmental effects must be reduced (Royal et al., 2000a). One approach is through measuring physiological parameters, such as the time of the

commencement of luteal activity postpartum (CLA), which are less affected by management bias (Veerkamp et al., 1997, Royal et al., 2000a). Milk progesterone level is measured thrice weekly from shortly after calving and continued until a minimum of 100 days of lactation. These measurements can be used then to determine CLA from the first elevation of milk progesterone level of ≥3ng/ml (Darwash et al., 1997; Royal et al., 2000a). This information will help to identify on average when the daughters of each bull have returned to cyclicity after calving. A significant and unfavourable genetic correlation exists between milk yield traits and endocrine fertility traits. Thus, CLA tends to be longer in cows with higher genetic merit for milk yield. The magnitude of this relationship is estimated to be an increase of about 1.4 days in CLA for every 1kg daily increase in milk yield over the first 100 days of lactation (Veerkamp et al., 1997). Furthermore, CLA was found to be positively correlated with calving interval (CI) indicating that cows with a genetically longer CI are more likely to have longer CLA. CLA increases by 1.1% (approximately 0.3d) for every 1 day increase in CI. A negative genetic correlation was established between CLA and body condition score, implying that cows that tend to be thinner are more likely to have longer intervals to CLA postpartum. Every unit increase in BCS (on a scale of 1 to 9) is associated with a decline in CLA by 22.4% (approximately 6 days; Royal et al., 2002b; Royal et al., 2002a). The heritability of CLA was found to be within the range of 0.16 and 0.23 (Royal et al., 2002b). Therefore, it has been concluded that endocrine measurements for fertility such as CLA might provide useful tools to improve the reliability of fertility breeding values when used with the traditional fertility traits of the selection index (Royal et al., 2002b). In the future, it is likely that milk progesterone recorded in automated parlours using biosensors will be incorporated into the selection index (Wall et al., 2003).

1.5.3.2 Juvenile predictor

Moreover, as bulls cannot have fertility proofs until they have daughters in milk, including them in the fertility index takes at least 4 years. So it would beneficial if the bulls with fertile daughters could be identified early in life before they are progeny tested. This approach is feasible as genes controlling fertility are expressed

early in life. As mentioned before there are many hormones regulating ovarian function such as gonadotrophin releasing hormone (GnRH) which is secreted from the hypothalamus and then travels to the pituitary gland where it will trigger the secretion of gonadotrophin (LH or FSH). These hormones stimulate the development of eggs and sperm in the ovaries and testes. So an experiment has been conducted by Royal and her colleagues to measure circulating LH level in response to administration of GnRH. This response is a measure of the ability of the pituitary to react to GnRH, which was found to be heritable in heifers at 5 months of age and to be genetically correlated with fertility in adulthood (Royal, 1999). This approach provides the means for pre-pubertal prediction of fertility and since the same genes control fertility in both males and females, so the pre-pubertal response in bull calves would also be expected to be correlated to their daughters' fertility. Recently, measures of somatotrophic hormones at six month of age have also been suggested to provide an indication to cow's performance over the first three lactations; due to the involvement of somatotrophic axis in the regulation of both fertility and milk production. It has been found that higher growth hormone pulse amplitude and lower insulin like growth factor-I (IGF-I) were associated with delayed ovulation in the first lactation, thus offering a new juvenile predictor for fertility in dairy cattle (Wathes et al., 2008)

1.5.3.3 Molecular markers

Molecular markers for fertility traits are considered to be another method for improving fertility index as they are highly heritable ($h^2=1$) and predictable (Hastings et al., 2006). These molecular markers are parts of the animal's DNA which are close to (linked to) the genes controlling fertility and in some cases they may be in the genes themselves. As reviewed above, there are many single nucleotide polymorphisms causing important effects on prolificacy and ovulation rate. Polymorphisms have also been found in several human genes to be associated with reproductive pathologies, such as mutations in the luteinising hormone receptor (Huhtaniemi, 2002). To identify mutations affecting fertility, the fertility trait PTAs of the sires in the fertility index can be tested for association with genotypes of candidate genes in their daughters. This method will provide an opportunity to

improve fertility by identifying sires carrying genes of particular importance for fertility before the sires are bred. This approach is the basis for a large part of the work carried out in this thesis (see Chapters 4 and 5). The impact of molecular markers on breeding program will be dicussed in in more details in the secetion 1.6.

1.5.4 The impact of oestrous detection on fertility of dairy cattle

One of the major causes of poor fertility in a dairy herd is the failure to detect oestrus due to the poor expression of oestrous behaviour, which consequently results in longer calving intervals (Lyimo et al., 1999). Standing heat is one of the most obvious symptoms by which to detect oestrus, and is a good predictor of the time of ovulation (as it occurs on average 26.4 ± 5.2h before ovulation; Roelofs et al., 2005a). However, in many studies only a limited proportion of cows are reported to display standing heat (50%; Van Eerdenburg et al., 2002), particularly when only one cow is in oestrus at a time (Roelofs et al., 2005a). Furthermore, the duration of oestrus in terms of the standing heat period was found to be as short as 8-9h, which might contribute to the low rate of oestrous detection (Van Eerdenburg et al., 2002). Other behavioural oestrous expression signs such as restlessness, mounting, sniffing and chin resting have also been found to be markedly elevated during oestrus. Therefore, a scoring system has been developed for these behavioural signs for use in oestrous detection (Lyimo et al., 1999; Roelofs et al., 2005a). These signs are highly correlated with the time of ovulation (Van Eerdenburg et al., 2002) and with the length of the oestrous period. By depending on visual observation of these behavioural signs, the oestrous period was reported to last approximately 20h (Lyimo et al., 1999). Furthermore, a high correlation of 0.7 was established between oestradiol level and visual signs of oestrous behaviour and its duration, suggesting that use of visual signs of oestrus improves the reliability and efficiency of identification of the correct time for insemination (Lyimo et al., 1999). However, variations in the expression of oestrous signs have been associated with many factors. Oestrous signs were more intense and lasted for longer periods in primiparous cows compared with multiparous, and also when more animals were in oestrus at the same time (Roelofs et al., 2005a). High milk yield and the resultant NEB have also been reported to be one of the reasons for the low expression of

oestrous behaviour in modern dairy cattle (Harrison et al., 1999; Yaniz et al., 2006). Furthermore, mastitis and lameness are associated with poorer expression of oestrus. Lame non-ovulating cows expressed 50% less intense oestrus, 50% lower luteal phase milk progesterone level and a lower circulating oestradiol level during the follicular phase than ovulating cows. It has been demonstrated that the stress of lameness results in a reduced LH pulsatility, which is required to stimulate oestradiol secretion by the dominant follicle and consequently causes failure to initiate the LH surge. As a result there is no oestrous behaviour and no ovulation (Dobson et al., 2008).

Because of the poor expression of obvious heat in cows, and since oestrous detection is time consuming and farmers cannot spend the time required for accurate detection of oestrus, an interest has arisen in the electronic detection of oestrus. As cows walking activity increases during oestrus, electronic activity tags or pedometers can be used to quantify these activities and to facilitate the automated prediction of the oestrous period (Limo et al., 1999; Van Eerdenburg et al., 2002; Roelofs et al., 2005b; Løvendahl and Chagunda, 2010). Pedometers attached to the cows' legs provide constant and diurnal monitoring of the cows behaviour, offering an accurate and efficient aid in oestrous detection when accompanied by visual oestrous detection (Van Eerdenburg et al., 2002). Furthermore, these devices can be used as reliable tools for accurate prediction of ovulation time, thereby improving fertilization rate. It has been reported that ovulation occurred 29.3± 3.9 h after the onset of increased activity (Roelofs et al., 2005b), which is consistent with the established positive relationship between fertility (pregnancy rate) and the increase in walking activities during oestrus (Lopez-Gatius et al., 2005).

Many studies have been performed to test the agreement between visual and walking activity oestrous cues. The duration of behavioural oestrous signs was on average 2h longer compared to pedometer oestrus. The intensity of both behavioural signs and walking activities increased when more animals were in oestrus at the same time, resulting in a higher percentage (95%) of oestruses detected. Moreover, the percentage of oestruses detected by pedometer was higher when cows expressed longer and more intense behavioural oestrus (Roelofs et al., 2005b).

Different factors have been associated with the variation observed in the proportion of overt oestruses detected by pedometer. Days from calving to first episode of high activity is negatively correlated with BCS in early lactation, suggesting that activity monitoring may provide information about other traits (Løvendahl and Chagunda, 2010). This is in agreement with the previously established negative association between BCS and fertility traits (Royal et al., 2002). Furthermore, increases in milk yield, parity, and insemination at the warm time were associated with lower walking activity in dairy cattle (López-Gatius et al., 2005).

Genetic variation in oestrous traits based on activity measurements have been reported (Løvendahl and Chagunda, 2009). The first detected high-activity episode after calving was shown to be a measure of days to first detectable oestrus (DFE). DFE was reported to occur at 44 days in milk in Holstein cows with a heritability of 0.18 ± 0.07 and a repeatability of 0.18 confirming the genetic component of the trait. This therefore shows that oestrus detected by activity monitoring is amenable to improvement through genetic selection. The slightly higher heritability of DFE in comparison with traditional fertility traits (h^2= 0.03; Wall et al., 2003) has led to the suggestion that this trait may be used in genetic selection for improved fertility.

In order for electronic oestrous detection records to be implemented in breeding schemes, it is necessary to predict breeding values, and this, together with the identification of molecular markers for higher activity during oestrus, are the subject of Chapters 6 and 7.

1.6 The contribution of molecular genetics to selective breeding

1.6.1 Introduction

In animal breeding, most of the traits of economic importance such as litter size, milk yield, disease resistance etc are multifactorial traits that are influenced by both genetic and environmental factors. Frequently they involve more than one gene. These traits are called quantitative traits (QT) and the genes that control them are called quantitative trait loci (QTL) (Simm, 2000). Molecular markers, which are

polymorphisms or variations in genes at the DNA level (Beuzen et al., 2000), will help in identifying parts of the chromosomes that are associated with particular QT. Therefore, combining information on molecular markers that are linked to QTL along with phenotypic information in a breeding programme (so called marker assisted selection; MAS) will result in a more accurate evaluation and therefore selection of animals with high genetic merit particularly for traits that are sex-limited or of low heritability (Mrode, 2005).

1.6.2 Types of molecular markers

At the DNA level three main types of variations have been described including: restriction fragment length polymorphisms (RFLP), microsatellites, and single nucleotide polymorphisms (SNP) (Beuzen et al., 2000) and these will be discussed below.

RFLP: RFLP are variations between individuals in the number and positions of the recognition sites at which restriction enzymes cut DNA. Therefore, some mutations will cause a DNA sequence to gain or lose the ability to be cleaved by a particular restriction enzyme. So the principle of the RFLP method for detecting polymorphisms depends on finding differences or similarities between DNA fragments of different animals after digestion of DNA samples with specific restriction enzymes. Thereafter, the DNA fragments of different length can be separated by gel electrophoresis. RFLP often occur in functional genes, and therefore they have great value as potential markers (Beuzen et al., 2000; Simm, 2000). However, the fact that RFLP is a gel-based technique makes it inappropriate for high-throughput screening. In addition, taking into account that not all mutations will result in the creation or deletion of restriction sites, the application of RFLP is limited (Beuzen et al., 2000).

Microsatellites: Microsatellites are repeated sequences of two to six base pairs throughout the genome. The level of mutation in the length of microsatellites is high, which makes them valuable markers for genes controlling economic traits in animal breeding. As a result, RFLP has largely been replaced by microsatellites for building

genetic maps for human and animal species. The reason for this replacement is the high number of alleles at each microsatellite which results in high heterozygosity values. This consequently reduces the number families needed to construct genomic maps. Moreover, microsatellites can be detected by the polymerase chain reaction (PCR) followed by allele sizing using a gel-based approach, which makes them easy to analyse (Vignal et al., 2002). Nonetheless, the requirement for gel-based DNA sizing represents a weakness in accurately defining microsatellite length (Beuzen et al., 2000).

SNPs: SNPs which have recently come to prominence in MAS, are the most common form of DNA variation. They are caused by the alteration of a single base, or the addition or deletion of one or more nucleotides in a DNA sequence (Beuzen et al., 2000). For such substitution in the DNA sequence to be considered a SNP, the least frequent allele must occur with a frequency of 1% or more (Vignal et al., 2002). SNPs may fall within the coding or non-coding regions of a gene, with SNPs located within the coding region being of particular importance as they are more likely to affect the biological function of the protein through a change in amino acid (AA) sequence. These are termed non-synonymous mutations. On the other hand, SNPs might be described as silent (or synonymous) mutations when the SNP results in no change in the AA sequence. These SNPs, although not changing the AA sequence, were found to be functional as well. For instance they might affect the function of exonic splicing enhancers (ESEs). ESEs are short sequences within exons which encourage exon recognition by the cell's splicing mechanisms. Therefore, SNPs in the ESEs can affect mRNA splicing which is an important step in creating the mRNA that is involved in protein synthesis via the process of translation (Mooney, 2005). In addition, SNPs may also affect the expression or translation of gene products by interrupting the regulatory region, a short sequence of DNA where regulatory proteins such as the transcription factors can bind, thus controlling gene expression and protein synthesis (Ngan et al., 1999).

Furthermore, SNPs that are located in the non-coding region of genes were also found to affect gene transcription by introduction or deletion of transcription factor binding sites (Pottier et al., 2007; Cheng et al., 2002). The effects of SNPs on transcription factor binding sites were tested as part of this project (Chapter 4). These

SNPs also may create novel splice donor/acceptor sites (Chillon et al., 1995) or interfere with messenger RNA regulation. Rates of mRNA translation may be affected by changes in the mRNA secondary structure (Duan et al., 2003) or by changes in codon adaptation index (causing usage bias of codons that code for the same AA, thus some codons are used more often than others; Carlini, et al., 2001), which in turn may affect protein folding during translation. Furthermore, SNPs that affecting transcription factor binding sites are more likely to exert these effects when they are located in an area of the chromosome that is more accessable to tranasciption factors, and this can be measured by nucleasome formation potential (nucleosome packaging). The nucleasomes are the major component of the chromatin, consisting of 147bp of DNA wrapped around an octamer comprising pairs of four histone molecules. It has been demonstrated that nucleasome packaging is one of the key factors underlying the specific function of genomic DNA, as decreasing or increasing nucleosome density will facilitate or inhibit binding of transcription factors respectively (Levitsky et al., 2001). This approach was tested in Chapter 4.

The widespread application of SNPs in selective breeding reflects their abundance throughout the genome (for example; in the human genome which comprises a total of 2.9 giga bp there is a SNP every 1000bp on average; Venter et al, 2001). This makes them more convenient markers than other polymorphisms when they are located near or in any locus of interest. As indicated above, SNPs located in coding sequences may change protein function, therefore accounting for significant variation among individuals for traits of interest. Furthermore, SNPs are more suitable for long term selection than other types of polymorphism due to the stability in their inheritance mode. Lastly, SNPs have been proved to be more suitable for high throughput genetic analysis than other polymorphisms (Beuzen et al., 2000).

1.6.3 SNP genotyping

Various methods have been used to detect SNPs. Some of these require prior information on the sequence of the polymorphic site. These methods are mainly characterized by two main features: the production of allele-specific molecular

reaction products and then the separation and detection of the products of DNA synthesis. Restriction enzyme cutting, single strand DNA conformation and heteroduplexes, and primer extension are examples of different techniques that have been employed in the detection of SNPs. Primer extension, which is a reliable method, has been used in identifying many SNPs in the present work. Basically primer extension is a two-step process that first involves the use of an oligonucleotide to prime DNA synthesis by a polymerase in the area immediately upstream of the polymorphism. DNA polymerase extends the strand by adding a single base that is complementary to the SNP nucleotide. Thereafter, this extended base can be detected in order to determine the SNP allele through fluorescent labelling (Vignal et al., 2002),.

1.6.4 Molecular markers in assisted selection

Different approaches have been employed to identify molecular markers for different traits including mainly association analysis and genome-wide selection analysis (Beuzen et al., 2000). Genomic selection depends on the discovery of a large number of polymorphic markers spread over an area of the genome. Thereafter, quantitative trait loci (QTL) associated with the phenotypic traits of interest can be identified using specific experimental designs combined with suitable statistical methods. Subsequently, the discovered QTL can be used in selection programmes (Haley, 2009; covered later in section 2.6). On the other hand, a candidate gene approach depends on the selection of genes that have been identified to be involved in the physiological regulation of a trait. Therefore, it is expected that variation in these genes will have an effect on the trait of interest. Subsequently, polymorphisms between animals in these genes are explored and linked to the phenotypic data (Veerkamp and Beerda, 2007).

If the genes that influence traits of interest are known mapped genes, these genes can be considered as candidate genes for an association study to investigate whether they account for any variation in the traits of interest. This approach, which has been employed in many species, has been used extensively during the current work. On the other hand if the traits are not well studied, genome-wide selection may have

advantages over the candidate gene approach, as it might help in identifying new genes of potential impact as markers for economically important traits (Simm, 2000; Beuzen et al., 2000).

1.6.5 Conclusion

Using association analysis and genome-wide analysis, several QTL have been identified and are now used in genetic selection by breeding companies (Beerda et al., 2008). However, finding significant effects of candidate genes on quantitative traits may not purely reflect the actual effects of the gene under consideration due to the presence of other closely-linked genes (linkage disequilibrium). Therefore, the effects of a candidate gene on a trait might be due to linkage disequilibrium with an actual QTL (Beuzen et al., 2000).

The incorporation of data from SNP studies into statistical models for derivation of breeding values, the so called genome-wide selection, is an area of quantitative genetics which is currently under development (this will be covered in section 2.6). In the mean time, information from association studies will continue to enhance our knowledge about the different genes that might have an impact on fertility in dairy cattle. In order to deduce the connexion between molecular markers and phenotypic measurements the concept of calculation the effects of these markers must be introduced; this being covered in the next section.

1.7 Calculation of the possible effect of allelic substitution

The effect of allelic substitutions will be discussed in the context when they are present singly or in combination (haplotype) on a gene or chromosome as follows:

1.7.1 Effect of allelic substitution at one locus

The phenotypic value of an individual *P* for a particular trait can be expressed as:

$$P = G + E = (A + D + I) + E$$

where: G is the genotypic value of an individual for a particular trait and this is equal to the sum of the additive value *A*, the dominance effect *D* and the epistatic effect *I* ; E represents the environmental effects.

The additive value is the sum of the average additive effects of each individual gene an individual carries. The dominance effect accounts for the effect of putting polymorphic genes together at a single locus (the within locus interaction). The epistatic effect represents the interaction between genes at different loci. In the absence of dominance and epistatic effects, the genotypic value is equal to the additive value at all the loci affecting a trait, and it is said that the genes "act additively".

By considering the variation at only one locus (SNP), the genotypic value consists of additive and dominance effects and these can estimated as follow:

Let A and B represent two alleles at one locus, let the genotypic value of one homozygote be *a*, that for the other homozygote be *–a*, and that for the heterozygote be *d* as illustrated in Figure 1.5:

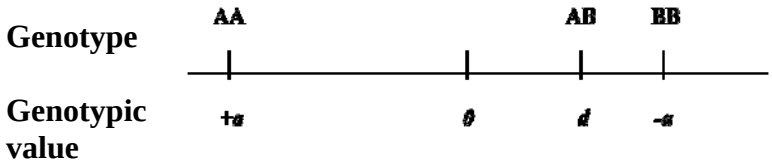


Figure 1.5 Arbitrary assigned genotypic value for the three possible genotypes (redrawn after Falconer, 1960).

In this example the value of the heterozygote *d* depends on the degree of dominance. If there is no dominance then $d = 0$; if A is dominant over B then *d* is positive; if B is dominant over A then *d* is negative. In the case of complete dominance *d* is equal to $+a$ or $-a$ and if there is overdominance, d is greater than $+a$ or less than $-a$. The

estimate of the additive effect in this case represents half the difference between the two homozygotes, whereas the dominance coefficient is the deviation of the heterozygote from this estimate (Falconer, 1960). The slope of a regression line drawn between the three genotypes represents the additive effect of the gene (Falconer, 1960).

1.7.2 Effect of allelic substitution at multiple loci (haplotype analysis)

Methods based on individual single nucleotide polymorphisms (SNPs) may lead to significant findings; however they cannot estimate the effects of the co-presence of several polymorphisms on the same gene or chromosome (a haplotype) on the phenotype (Tregouet et al., 2002). Therefore, haplotype analysis has become an essential step when investigating an association between several polymorphisms within a gene and a phenotype (Tregouet and Garelle, 2007). A haplotype can be defined as a combination of marker alleles on a single gene or chromosome (Zhao et al., 2003). The advances in molecular technologies now make it possible for hundreds of thousands of genetic polymorphisms to be studied in population samples of reasonable sizes. One major aspect of haplotype analysis is to identify linkage disequilibrium patterns (markers display statistical dependence and tend to be inherited together) in different regions and different populations (Zhao et al., 203) and this offers a promising tool in deducing population histories and identifying genetic variants underlying complex traits, and may also provide information about factors affecting inter-dependency among genetic markers (Liu et al., 2008). Furthermore, haplotype-based analysis may help to distinguish the true effect of a polymorphism from what is due to its linkage disequilibrium with other variants (Tregouet and Garelle, 2007). Importantly, haplotypes were found to provide better markers for unknown functional variants than single nucleotide polymorphisms (SNPs). In addition, haplotype analysis may define functional haplotypes whose effects cannot be predicted from the individual effects of each of the variants comprising these haplotypes (Tregouet and Garelle, 2007). Therefore, it has been suggested that full haplotypic information should be exploited in order to better characterize the role of a candidate gene (Tregouet et al., 2002). Haplotype markers were also employed in genomic selection as an alternative to single marker

genotypes (discussed in 2.6) as this will result in increasing the linkage disequilibrium between the haplotype markers and the QTL in comparison with a single marker genotype. Subsequently, this will increase the proportion of variance that is explained by the haplotype markers in comparison with single marker genotypes, and this will increase the accuracy of the estimate of the QTL effect, particularly when large number of phenotypic and genotypic records is available (Goddard and Hayes, 2007).

As discussed in section 1.6, SNPs might be located in different regions of the gene, and upon their location they might be assigned different names. The principles of abbreviating SNPs name in association with their location are discussed in the next section.

1.8 Nomenclature

The Human Genome Organization Gene Nomenclature Committee (HGNC) has approved a name and symbol for each known gene (short-form abbreviation). All approved gene symbols are stored in the HGNC Database (Wain et al., 2002). Each symbol is unique and the HGNC ensure that each gene is only given one approved gene symbol. In general, gene names should be brief and specific and should express the character or function of the gene, but should not go in much detail in describing everything known about it. The first letter of the symbol is the same as that of the name (such as luteinizing hormone receptor gene, LHR) in order to facilitate alphabetical listing and grouping. Gene names are written using American spelling. Moreover, a nomenclature system has been suggested for the description of changes (mutations and polymorphisms) in DNA and protein sequences as discussed below (Wain et al., 2002).

Description at the DNA Level

• Nucleotide changes start with the nucleotide number and the change follows this number; {nucleotide position}{sequence changed nucleotide}{type of change}{sequence new nucleotide} e.g., 206 G>A (see tables 3.4; 3.5).

• Substitutions are designated by ">" such as for example 206 G>A denotes that at nucleotide (nt) position 206 of the reference sequence a G is changed to a A.
• Deletions are designated by "del" after the deleted interval (followed by the deleted nucleotide). For example, a deletion SNP at nucleotide position 1588 in the LHβ was identified in this project (see Table 3.4) and this can be expressed as 1588del (alternatively 1588-delC) which denotes a C deletion from nucleotide 1588. Insertions are designated by the same way as deletion SNP with the substitution of "del" with "ins,"
• Variation in the length of the short sequence repeats, e.g., in ACTGTGTGCC (A is nt 100), can be expressed as 102(TG)3-15 with nucleotide at position 102 containing the first TG-dinucleotide which is found repeated from 3 to 15 times in the population.
• Intron mutations are expressed relative to the intron number (preceded by "IVS") or cDNA position; positive numbers starting from the G of the GT splice donor site, negative numbers starting from the G of the AG splice acceptor site. For instance, five intronic SNPs were identified on the *activin receptor2B* gene at nucleotides positions 45, 46, 86, 95 and 503 as part of this project (see table 3.4) and these can be denoted as IVS3 45C>T , IVS3 46T>G, AVS9 86G>A, AVS1 95G>A and AVS4 C>T respectively; with IVS3 45C>T for instance denotes the C to T substitution in intron 3 at the cDNA level sequence 45 (den Dunnen and Antonarakis, 2000))

Description at the protein level

The expressions of sequence variations at the protein level describe the deduced consequence and not the nature of the mutation, and follow the following principles:

• The codon for the initiator Methionine is codon 1.
• Stop codons are designated by X. For instance, R97X denotes a change of Arginine at amin acid position 96 to a termination codon.
• The single letter amino acid code is recommended, but the three letter code is acceptable (Table 1.5).
• Amino acid changes are described in the format {code first amino acid changed}{amino acid position}{code new amino acid or type of change}. For

example, FSHR-T658S denotes a Threonine at residue 658 is substituted by a Serine in the *FSHR* gene (see Table 3.4).

• Deletions are designated by "del" after the amino acid position such as the polymorphisms expressed as T97-C102del which denotes a change in the amino acid from Threonine at residue 97 to Cysteine at residue 102 are deleted. Insertions are designated by the same way as deletion but with "del" substituted by "ins" (den Dunnen and Antonarakis, 2000).

Table 1.5 Three letter and single letter abbreviations assigned for the amino acids.

Amino acid	Three-letter symbol	Single-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asn +Asp	Asx	B
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Gln + Glu	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Selenocysteine	Sec	U
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Chapter 2: Breeding values and their prediction

2.1 Introduction

The breeding value of an animal is that part of its genotype that is transmitted to its offspring. Thus, to improve the next generation, parents with the highest breeding values should be selected. Currently, it is not possible to directly estimate the genetic values of animals, so the true breeding value of an animal can never be known. However, one can come close to it using phenotypic observations of an animal to estimate or predict its breeding value, and consequently animals are selected based on their predicted breeding values. An animal's phenotypic characteristics are the result of genetic and environmental factors; therefore, they reflect partially the genetic merit of an animal. Thus the critical point, the estimated breeding value ($\hat{a}$) should be an unbiased estimator of the true breeding value (a), and the regression coefficient of the true breeding value on the estimated breeding value should be unity. Furthermore, the estimated breeding value should be as accurate as possible, reflecting the need for the correlation between $\hat{a}$ and a to be as high as possible. True breeding values and phenotypes are assumed to be normally distributed so the relationship between them is linear. Consequently, linear regression of true breeding values on phenotypes can be used to estimate breeding values.

The method employed for the prediction of breeding values depends on the type and amount of information available on animals that are candidates for selection. In the simplest case the performance of the animal itself for the trait of interest can be used to predict the breeding value. Repeated measurements of animal performance can also be used for more accurate estimation of the breeding value. Furthermore, phenotypic measurements of animals' relatives can also be used, because relatives have some of their genes in common such as for example when the animal itself has no record, its breeding value can be predicted from the breeding values of it parent (Falconer, 1960).

Hazel in 1943 developed a method allowing the combination of all information available on the animal and its relatives for several traits, in order to get a better

estimate of the animal's breeding value. This method is called a selection index or best linear prediction (BLP). BLP requires all fixed effects to be known. However, these are seldom known, therefore Henderson (1949) introduced a more accurate method to estimate the fixed effects and to predict the breeding values simultaneously. This procedure is based on the so-called "Mixed Model Equations" (MME). MME combine the properties of a linear model (least square) and a selection index (based on regression of the breeding values on the phenotypic records obtained from different sources) and can simply be reduced to a selection index when no adjustments for environmental effects are required (Henderson, 1984).

This method, which is classified as best linear unbiased prediction (BLUP), requires the inverse of the additive genetic relationship matrix A among animals to be known. The genetic relationship between two animals is the probability that the genotypes of the two animals, for a gene taken at random, are identical by descent (Mrode, 2005). The A matrix is usually computed from the pedigree information (methods for constructing pedigree information for BLUP analysis is highlighted in section 3.3.1) to describe the genetic variances and covariances of the complete population (Henderson, 1973). However, the resulting matrix is generally large and sometimes it cannot be inverted. Therefore, Henderson (1976) described a method to write the inverse of (A) directly from the pedigree information without the need to compute (A) itself. Different BLUP models have been applied depending on the information available on animals to be evaluated. The BLUP technique has become one of the most attractive and widely used tools for animal evaluation in breeding programmes. In the subsequent sections the prediction of breeding values using different sources of information is discussed.

2.2 Information on the animal

2.2.1 Single measurement per animal

When each animal in the population has only one phenotypic record, animals are selected purely on their own phenotypic values (Falconer, 1960). The estimated breeding value $\hat{a}$ for each animal can be calculated by regression of the animal's

breeding value a, on its phenotypic value y (Figure, 2.1).The true breeding values and phenotypes are assumed to be normally distributed resulting in a linear relationship between them (Cameron, 1997). The regression equation can be written as:

$$\hat{a}_i = b(y_i - \mu)$$

where: $\hat{a}_i$ is the predicted breeding value for animal i, b is the regression coefficient of the true breeding value a_i on the phenotypic performance y_i and μ is the mean performance of animals in the same management group (which is assumed to be known).

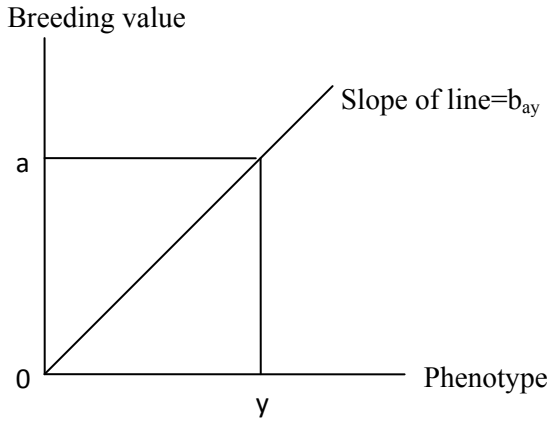


Figure 2.1 Regression of the true breeding value on phenotype (redrawn after Cameron, 1997)

With one measurement on animal performance, the regression coefficient of additive genetic merit on phenotype is equal to the heritability:

$$b = \frac{\sigma_{ay}}{\sigma_y^2} = \frac{\sigma(a, a+e)}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_y^2} = h^2.$$

where: σ_{ay} is the covariance between the true breeding value *a* and phenotypic performance y , σ_y^2 is the phenotypic variance, σ_a^2 is the additive gentic variance, e is the sum of the random enviromnetal effects and non additive genetic effects (dominance and epistatic).

So the best estimate of an individual's breeding value is the product of the heritability and its phenotypic value (Falconer, 1960).

The accuracy of estimation of the breeding value is defined as the correlation between the selection criterion, in this case the phenotypic measurement, and the true breeding value. It is a measure of the precision by which the breeding value has been calculated and it helps in the evaluation of different selection criteria. The higher the accuracy, the better the estimated breeding value as a predictor of the true breeding value is going to be. In some cases the accuracy of estimation is described in terms of reliability (r^2) which is the squared correlation between the selection criterion (the phenotype) and the true breeding values:

$$r_{ay} = \frac{cov\,(a,y)}{\sigma_a \sigma_y} = \frac{\sigma_a^2}{\sigma_a \sigma_y} = h.$$

where: r_{ay} is the accuracy of the estimation of the breeding value, cov(*a, y*) is the covariance between the true breeding value and the phenotype, σ_a and σ_y are the additive genetic and phenotypic standard deviations respectively.

So with single measurements per animal the reliability equals h^2 (Mrode, 2005).

If animals are selected based on their breeding value, and the animals with the highest genetic merit are selected to be parents, then the response to selection can be defined as the differences between the mean phenotypic values of the progeny of the selected parents compared to the parental population before selection. The selection differential (SD), which is a measure of the magnitude of selection applied, is the difference between the mean phenotypes of the selected parents and parental generations. The predicted response to selection depends on both the proportion of animals selected (the intensity of selection *i*) and the regression coefficient *b* of additive genetic merit on phenotype. Considering that the trait is normally distributed the expected response to selection (*R*) on the basis of a single record per animal (Falconer, 1960) is:

$$R = b.SD = ih^2\sigma_y = ir_{a,y}^2\sigma_y \qquad \text{since} \quad r_{a,y} = h$$

where: SD is equal to the intensity of selection *i* multiplied by the phenotypic standard deviation σ_y.

The variance of the predicted breeding value $var(\hat{a}_i)$, which is also a measure of the precision with which the breeding value is estimated, can be written as follows (Cameron, 1997):

$$var(\hat{a}_i) = var(by) = var(h^2 y)$$

$$= h^4\sigma_y^2 = r_{a,y}^2\, h^2\sigma_y^2 = r_{a,y}^2\sigma_a^2.$$

2.2.2 Repeated measurements per animal

When more than one measurement of the same trait such as milk yield in successive lactations or litter size in successive pregnancies can be made on an individual, its breeding value may be predicted from the mean of these measurements (Mrode, 2005). With repeated records, the phenotypic variance can be partitioned into variance within individual, (measuring the differences between the performances of the same individual), and variance between individuals, (measuring the permanent differences between individuals). The within-individual variance is completely environmental, caused by temporary or localised changes in environment between successive performances. The between-individual variance is partly environmental and partly genetic, the environmental part being caused by factors that affect the individuals permanently. The permanent or non-localised environmental effects result in an increase in the covariance between records of an individual. Therefore, the variance of measurements *var(y)* could be partitioned as:

$$var(y) = var(g) + var(pe) + var(te)$$

where *var(g)* is the genetic variance including both additive and non-additive effects, *var(pe)* is the variance contributed by the permanent environmental effect, and *var(te)* is the variance resulting from temporary environmental effects. The ratio of the between-individual variance to the total phenotypic variance measures the intra-class correlation *t* between the repeated measurements of the same individual, which is also known as the repeatability (Falconer, 1960; Mrode, 2005):

$$t = \frac{var(g) + var(pe)}{var(y)}. \qquad (1)$$

From (1):

$$\frac{var(te)}{var(y)} = 1 - t. \qquad (2)$$

The essential assumption enclosed in the principle of repeatability is that the multiple measurements are measurements of what is genetically identical. So, the genetic correlation between all pairs of the records is equal to one (Falconer, 1960). Moreover, it is assumed that all records have equal variance and that the environmental correlations between all pairs of records are equal. Assuming that $\tilde{y}_i$ is the mean of n records on animal *i* for particular trait, the predicted breeding value $\hat{a}$ can be written as (Mrode, 2005):

$$\hat{a} = b(\tilde{y}_i - \mu)$$

where *b* is the regression coefficient of the animal true breeding valye *a* on the mean of of n records $\tilde{y}_i$:

$$b = \frac{cov(\hat{a}, \tilde{y}_i)}{var\ (\tilde{y}_i)}.$$

Therefore, derivation of the regression coefficient requires both the covariance between the genotype and the mean of n measurements, and the variance of the mean of n measurements (Mrode, 2005):

$$cov(\hat{a}, \tilde{y}_i) = cov\left(\hat{a}, g + pe + \frac{\sum te}{n}\right) = \sigma_a^2.$$

assuming that the covariance between additive genetic effect with non-additive genetic effects and environmental effects, are zeros.

The phenotypic variance ($var(\tilde{y}_i)$) of the estimated mean ($\tilde{y}$) for n repeated measurements per animal (Falconer, 1960) is:

$$var(\tilde{y}_i) = var(g) + var(pe) + \frac{var(te)}{n}. \qquad (3)$$

Substituting the items in (3) with the corresponding values from (1) and (2) gives:

$$var(\tilde{y}_i) = \left[t + \frac{(1-t)}{n} \right] \sigma_y^2.$$

where: t is the repeatability, σ_y^2 is the phenotypic variance of the trait, n is the number of records.

The phenotypic variance of the mean of n measurements is less than the phenotypic variance of the trait. As the number of repeated measurements increases, the influence of the temporary environmental effect on the mean measurement will be reduced (Falconer, 1960; Mrode, 2005). Therefore, the heritability of the mean measurement is greater than the heritability when one measurement is taken per animal, as the genetic variance contributes a substantial proportion of the phenotypic variance (Cameron, 1997).

Substituting the previously derived covariance and variance back into the regression formula gives the value of the regression coefficient of the additive genetic merit on the mean of n measurements in terms of the number of records n, the repeatability t and the heritability h^2 of the trait:

$$b = \frac{\sigma_a^2}{\left[t + \frac{(1-t)}{n} \right] \sigma_y^2} = \frac{nh^2}{1 + (n-1)t}.$$

So the regression coefficient increases as the number of measurements and the heritability increases, and more weight is given to the phenotypic mean for prediction of breeding value (Mrode, 2005).

The accuracy of the estimated breeding value $r_{a\tilde{y}}$ is:

$$r_{a\tilde{y}} = cov(a, \tilde{y})(\sigma_a, \sigma_{\tilde{y}})$$

where: $cov(a, \tilde{y})$ is the covariance between the true breeing value and the mean of n repeated measurements, $\sigma_a, \sigma_{\tilde{y}}$ are the additive genetic and phenotypic standard deveiation respectively.

$$r_{a\tilde{y}} = \frac{\sigma_a^2}{\sigma_a \sqrt{\frac{t + (1-t)}{n} \sigma_y^2}}$$

$$= h\sqrt{\frac{n}{1+(n-1)t}}$$

$$= \sqrt{\frac{nh^2}{1+(n-1)t}} = \sqrt{b}.$$

From the above equation, it can be concluded that there is an increase in the accuracy with repeated measurements in comparison with single records, which depends on the number of records and the value of the repeatability. Thus, increasing the number of measurements results in a reduction in the variance due to within individual variance that contributes to the phenotypic variance, and this reduction in the phenotypic variance constitutes the gain in accuracy. When *t* is high, and there is little temporary environmental variance, the gain in accuracy will be small with repeated measurements. When *t* is low, the increase in accuracy is worthwhile as the number of records increases. However, the gain in accuracy falls rapidly as the number of records increases, and in most situations it is not cost effective to take more than two measurements (Falconer, 1960). The gain in accuracy from repeated measurements compared with single measurements can be derived as the ratio of accuracy from repeated records (r_n) to that from single records (r) (Mrode, 2005):

$$\frac{r_n}{r} = \frac{\sqrt{\frac{h^2}{t+\frac{(1-t)}{n}}}}{h} = \sqrt{\frac{1}{t+\frac{(1-t)}{n}}}.$$

where *t* is the repeatability.

As mentioned earlier, repeated measurements are assumed to be genetically identical and therefore the variance between them is due to temporary environmental variance. However, this assumption is not valid when successive records, such as milk yield in successive lactations, are not entirely controlled by the same genes, or in other words the developmental and physiological process are not the same in all lactations. In this case the variance within individual is not purely environmental, and this results in an increase in the variance between the means of individuals which arises from what is known as interaction between genotype and environment. This additional variance may be enough to halt the gain achieved from repeated measurements. Therefore,

records must be corrected for these factors before using the means for the prediction of breeding values (Falconer, 1960).

The expected response to selection R_n on the basis of the mean of repeated measurements (Cameron, 1997) is:

$$R_n = b.SD = \left[\frac{nh^2}{1+(n-1)t}\right] i\sigma_{\tilde{y}}$$

$$R_n = ih^2\sigma_y \sqrt{\frac{n}{1+(n-1)t}}.$$

where: *b* is the regression coefficient of the true breeding value on the mean of n measurements, *SD* is the selection deferential.

Given that the regression coefficient is the square of the accuracy of the predicted breeding value, so the response can be expressed in terms of the accuracy as:

$$R_n = ih\sigma_y\, r_{ay}$$

$$= ir_{ay}\sigma_a\,.$$

So the response to selection can be predicted if the accuracy of the predicted breeding value and the selection intensity are known (Cameron, 1997).

2.3 Information from relatives

Measurements on an individual's relatives can be employed to estimate its breeding value once proper genetic relationships between the individual and its relatives are applied.

2.3.1 Information from progeny

Measurements on progeny can also be used to predict an individual's breeding value, and this is used mainly when the trait can only be recorded in one sex. Thus progeny

testing is commonly applied in the dairy industry to estimate the breeding value of a bull for milk quality and quantity on the basis of his daughters' records. Similarly, progeny testing can be used to predict breeding values for reproductive performance and carcass composition (Simm, 2000; Cameron, 1977). The prediction of an individual's breeding value depending on its progeny's records is comparable to using repeated measurements on the individual, with consideration of the genetic relationship between an individual and its progeny (Cameron, 1997). Assuming that $\tilde{y}$ is the mean of single measurements of *n* half-sib progeny of sire *i* (parental half-sib), the breeding value of the sire (Mrode, 2005) is:

$$\hat{a}_i = b(\tilde{y} - \mu)$$

where *b* is the regression coefficient of the animal's breeding value (BV) on its progeny's mean records:

$$b = \frac{cov(a, \tilde{y})}{var(\tilde{y})}.$$

The covariance between the individual's BV and the mean of its progeny's measurements, which is equivalent to the covariance between the individual's BV and one measurement of its progeny (Falconer, 1960; Cameron, 1977), is:

$$cov(a, \tilde{y}) = cov(a, \frac{1}{2}a_s + \frac{1}{2}a_d + \frac{\sum e}{n})$$

where a_s and a_d are the sire and dam breeding values respectively. So for half-sibs:

$$cov(a, \tilde{y}) = \frac{1}{2}cov(a, a_s) = \frac{1}{2}\sigma_a^2.$$

Using the same principle of the mean of n repeated measurements, the variance of the progeny mean, $\bar{y}$, is:

$$\sigma_{\tilde{y}}^2 = \left[t + \frac{1-t}{n}\right]\sigma_y^2$$

where: *t* is the intra-class correlation between progeny. For half-sib progeny:

$$t = \frac{\frac{1}{4}\sigma_a^2}{\sigma_y^2} = \frac{1}{4}h^2$$

assuming that there is no environmental covariance between the half-sib records (Falconer, 1960). The regression coefficient of predicted breeding value on the mean of n progeny $\bar{y}$, is:

$$b = \frac{\frac{1}{2}\sigma_a^2}{\left[t + \frac{(1-t)}{n}\right]\sigma_y^2} = \frac{\frac{1}{2}h^2\sigma_y^2}{\left[\frac{1}{4}h^2 + \frac{(1-\frac{1}{4}h^2)}{n}\right]\sigma_y^2}$$

$$= \frac{2nh^2}{nh^2 + (4-h^2)} = \frac{2n}{n + \frac{(4-h^2)}{h^2}} = \frac{2n}{n+k}$$

with: $k = \frac{(4-h^2)}{h^2}$ and is constant for any given heritability.

The regression coefficient depends on the heritability and the number of progeny and approaches two as the number of progeny increases (Cameron, 1977).

The accuracy of the predicted breeding value depending on progeny information can be derived (Mrode, 2005) as:

$$r_{a,\tilde{y}} = \frac{cov(a,\tilde{y})}{\sqrt{var(a)var(\tilde{y})}}$$

$$= \frac{\frac{1}{2}h^2\sigma_y^2}{\sqrt{h^2\sigma_y^2\left(\frac{1}{4}h^2 + \frac{\left(1-\frac{1}{4}h^2\right)}{n}\right)\sigma_y^2}}$$

$$= \sqrt{\frac{nh^2}{nh^2 + (4-h^2)}} = \sqrt{\frac{n}{n+k}}$$

which approaches unity as the number of daughters increases. Therefore, the reliability of the predicted breeding value equals $\frac{n}{n+k}$.

The expected response to selection when selection is based on the mean of half-sibs is of the same form as that for the mean of repeated records, with *t* referring to the intra-class correlation between half-sibs, and consideration being taken of the genetic relationships between the individual and its progeny by using pedigree file (Cameron, 1997):

$$R_n = \frac{1}{2} i h^2 \sigma_y \sqrt{\frac{n}{1 + (n-1)t}}$$

However, progeny selection suffers from the serious drawback of a long generation interval in some specieces (such as the case in dairy cattle), as the selection of parents cannot be carried out until the phenotypes of their offspring have been measured (Falconer, 1960; Hazel, 1943).

2.3.2 Information from sibs

Measurements on sibs are mainly used when the measurements of the traits of interest cannot be measured directly on the candidates for breeding such as information on reproductive traits, carcass composition or meat quality (Falconer, 1960). So the genetic merit of an individual is predicted on the basis of the mean of its sibs' measurements $\tilde{y}$, but the individual is not measured. Measurements on sibs are similar to having repeated measurements on the individual but one has to take into account the genetic relationship r_g between the individual and its sibs. The genetic relationship between two animals is the probability that the genotypes of the two animals, for a gene taken at random, are identical by descent. With selection depending on sib measurements, the regression coefficient is calculated by dividing the covariance between the animal's breeding value and the sibs' mean by the variance of the sib mean. With selection on sib information, the predicted breeding value of the individual is:

$$\hat{a} = b(\tilde{y} - \mu).$$

As with repeated measurements, the covariance between an individual's breeding value and its sibs' mean measurements is the covariance between an individual's breeding value and one sib's measurements (Cameron, 1977). The covariance is

$r_g\sigma_a^2$, where r_g is the genetic relationship between the individual and its relatives, which is one half for full-sibs and a quarter for half-sibs (Falconer, 1960). The variance of the sibs' mean is:

$$\sigma_{\bar{y}}^2 = \left[\frac{1+(n-1)t}{n}\right]\sigma_y^2$$

which is similar to the variance of the repeated measurements of an individual derived before, with *t* being the intra-class correlation between sibs measurments. If the individuals are half-sibs, then $t = \frac{1}{4}h^2$ while if they are full sibs then $t = \frac{1}{2}h^2 + c^2$ where c^2 is the maternal effect, which also includes the common environmental effects and non-additive genetic effects (Cameron, 1997). Therefore, the regression coefficient of the individual's breeding value on the mean of its n sibs (Falconer, 1960) is:

$$b = \frac{r_g\sigma_A^2}{\left[\frac{1+(n-1)t}{n}\right]\sigma_y^2} = \frac{nr_gh^2}{1+(n-1)t}$$

When measurements are taken on sibs, then the accuracy of the predicted breeding value is obtained from:

$$r_{a\tilde{y}} = \frac{cov(a,\hat{y})}{\sqrt{var(a)Var(\tilde{y})}} = \frac{r_g\sigma_a^2}{\sigma_a\sqrt{\left[\frac{1+(n-1)t}{n}\right]\sigma_y^2}}$$

Therefore, the reliability (square of the accuracy) is:

$$r_{a\tilde{y}}^2 = \frac{nr_g^2h^2}{1+(n-1)t}$$

where: r_g is the genetic relationship between the individual and its relatives, *t* is the the correlation between sibs measurements.

When measurements are taken only on half sibs, then $r_g = \frac{1}{4}$ and $t = \frac{1}{4}h^2$ and the square of the accuracy of the predicted breeding value can be written as (Cameron, 1997):

$$r_{A\hat{A}}^2 = \frac{1}{4}\frac{n}{n+k} \qquad \text{where } k = \frac{4-h^2}{h^2}.$$

2.3.3 Information from parents

In the case where no record is available on an individual, its breeding value can be predicted from the predicted breeding values of its sire (*s*) and its dam (*d*). As each parent transmits half of its genes to its progeny, so the predicted breeding value of their progeny (Cameron, 1997) is:

$$\hat{a} = \frac{1}{2}(\hat{a}_s + \hat{a}_d)$$

where $\hat{a}_s$ and $\hat{a}_d$ are the sire's and dam's predicted breeding values respectively.

The regression coefficient of predicted breeding value on average parental predicted breeding value is:

$$\mathrm{b} = \frac{\mathrm{cov}(\mathrm{a}, \hat{\mathrm{a}})}{\mathrm{var}(\hat{\mathrm{a}})}.$$

The covariance between the breeding value and the average parental predicted breeding value is:

$$\mathrm{cov}(\mathrm{a}, \hat{a}) = \mathrm{cov}\left[\frac{1}{2}(\mathrm{a_s} + \mathrm{a_d}), \frac{1}{2}(\hat{\mathrm{a}}_\mathrm{s} + \hat{\mathrm{a}}_\mathrm{d})\right]$$

$$= \frac{1}{4}[\mathrm{cov}(\mathrm{a_s}, \hat{\mathrm{a}}_\mathrm{s}) + \mathrm{cav}(\mathrm{a_d}, \hat{\mathrm{a}}_\mathrm{d})].$$

Assuming the sire and dam breeding values are not correlated, the covariance is:

$$= \frac{1}{4}[r_s^2 + r_d^2]\sigma_a^2 \quad \text{since } r_{a\hat{a}}^2 = \frac{var(\hat{a})}{var(a)}.$$

The variance of the average parental predicted breeding values is:

$$\mathrm{var}(\hat{a}) = \mathrm{var}\left[\frac{1}{2}(\hat{a}_s + \hat{a}_\mathrm{d})\right] = \frac{1}{4}[r_s^2 + r_d^2]\sigma_a^2.$$

So the regression coefficient of predicted breeding value on the average parental predicted breeding value is unity.

The response to selection is:

$$R = b.SD$$

where *SD* is the selection differential for average parental predicted breeding values. As *b=1*, the response to selection is simply the selection differential for average parental predicted breeding values:

$$R = SD = \mathrm{i}\sigma_{\hat{\mathrm{a}}} = \frac{1}{2}\sqrt{(\mathrm{r}_s^2 + \mathrm{r}_\mathrm{d}^2)}\sigma_\mathrm{a} = \frac{1}{2}\sqrt{(\mathrm{r}_s^2 + \mathrm{r}_\mathrm{d}^2)}\mathrm{h}\sigma_\mathrm{y}$$

and as the accuracy of the parental predicted breeding values increases, then the response to selection on average parental predicted breeding value increases.

The accuracy of the animal's breeding value depending on predicted breeding values of its parents is:

$$r = \frac{1}{2}\sqrt{(\mathrm{r}_\mathrm{S}^2 + \mathrm{r}_\mathrm{D}^2)}$$

With the accuracy of each parent being one, the accuracy of the animal's predicted breeding value would be $\frac{1}{2}\sqrt{2} = \sqrt{\frac{1}{2}}$, which is equivalent to the accuracy when predicting breeding values depending on large number of records of half-sibs. The maximum value of the accuracy is limited to $\sqrt{\frac{1}{2}}$ as a result of Mendelian sampling, as the animal's two genes at a specific locus are not similar to both genes of either parent (Cameron, 1997).

The formulas for the regression coefficient of the breeding value on the mean phenotypic measurements (repeated measurements per animal, sib measurements and half-sib progeny measurements) are summarized in Table 2.1.

Table 2.1 Formula for the regression coefficient of breeding value on mean phenotypic measurements.

	Repeated measurements	**Sib measurements**	**measurements on half-sib Progeny**
Regression coefficient	$\frac{nh^2}{1+(n-1)t}$	$\frac{nr_g h^2}{1+(n-1)t}$	$\frac{\frac{1}{2}nh^2}{1+(n-1)\frac{1}{2}h^2}$

- Where *t* is the intra-class correlation, r_g is the genetic correlation between sibs.

2.4 Selection index methodology

So far, only a single information source has been used to estimate an animal's breeding value. However, in practice many traits might affect an animal's practical value, although they may do so to different degrees. Genetic improvements might be desired for several traits which differ in variability, heritability, economic importance, and in the genetic and phenotypic correlations between them. Furthermore, there are various sources of information regarding different traits, some coming from the animal's own performance and some from the animal's relatives, for traits that are expressed once or more in lifetime (Hazel, 1943). All these records may provide a better estimate of the individual's breeding value, as all relatives have different proportions of their genes in common. It is important to reflect this in genetic improvement programmes, so animals are selected on combinations of traits and records. With multiple information sources, multiple regressions of breeding values on phenotypes should be applied. This is called selection index methodology, and was introduced by Hazel in 1943. The resulting regression expression is called an index *I* which was proved to be the best linear prediction (BLP) of an individual breeding value and it is the basis on which individuals are ranked for selection (Henderson, 1973). Selection on the basis of an index, where proper weight is given to each trait, was proved to be more efficient than selection for one trait at a time or for several traits with independent culling levels. Hazel stated that the purpose of a selection index should be to attain the maximum genetic gain towards a desirable economic goal. This economic return to the livestock industry should be of primary importance and was defined as the expected return from a unit change in each trait. The genotypic and phenotypic parameters should be used to estimate the index weightings that maximize the correlation of an index with the genotypic value in terms of net profit (Hazel, 1943; Hazel et al., 1994). The principles of constructing a selection index which permits maximum genetic improvement are given below.

2.4.1 Selection objective

The selection objective (*H*), the starting point of animal breeding, defines the direction in which it is desirable to improve the population. It represents the trait(s)

that are to be improved (Simm, 2000). The total genetic improvement which can be achieved by selection among groups of animals is defined as the sum of the genetic gain attained for several traits of economic importance. So the average genetic superiority of a selected group over the group from which it was selected is:

$$\bar{H} = w_1\bar{a}_1 + w_2\bar{a}_2 + \cdots w_n\bar{a}_n$$

where: $\bar{a}_i$ is the genetic gain for each trait and w_i is the relative economic weight. Moreover, animals vary in their phenotypes as well as in their genotypes for each of the several traits. So the aggregate value of an animal is the sum of its genotype for each trait, weighted by its relative economic value and this can be expressed as:

$$H = w_1a_1 + w_2a_2 + \cdots w_na_n.$$

So to improve the population, *H* for each individual has to be estimated and subsequently individuals with the highest estimated value for *H* are selected. As the phenotypic performance of an animal is not affected by its genotype alone, but also by environmental factors, dominance and epistasis, the accurate selection of animals with the highest values for *H* is not a straightforward procedure. Therefore, genetic improvement through selection should be applied indirectly by direct selection on the phenotypic measurements of each animal for the several traits, and these are represented by the selection criterion or selection index *I* (Hazel, 1943). This is analogous to breeding value estimation for a single trait, an index *I* being used to estimate *H* for selection candidates. For example, the selection objective of fertility in dairy cattle is to improve PTAs for CI and NR56 and these have been assigned an economic values of -0.31 and 1.56£ respectively, giving a selection objective of a form:

FI = PTACI X -0.31 + PTANR56 X 1.56

2.4.2 Selection criteria

Selection criteria (*I*) are the set of traits on which measurements are available for candidates for selection or their relatives and from which breeding values are estimated (Cameron, 1997). After defining the overall breeding value or the aggregate breeding value, Hazel developed a method to derive index weightings (b_i)

for the index phenotypic traits which will maximize the correlation between the selection objective or the aggregate genotype and the selection criterion R_{IH} (Hazel et al., 1994). A general selection index can be written as:

$$I = b_1X_1 + b_2X_2 + \cdots b_nX_n$$

where: Xi represents the phenotypic performance of an animal or its relatives for several traits, and b_n are the multiple regression coefficients by which the phenotypic measurements information are weighted.

The problem dealt with is to find weights in such a way that the information from different sources is optimally used to rank animals on their expected genetic merit. This can be achieved by minimizing the average square prediction error, that is minimizing $(H - I)^2$. This is equivalent to maximizing the correlation (R_{IH}) between the true breeding value and the index, which is also called the accuracy of prediction. These procedures result in a set of equations which are solved simultaneously in order to obtain the values of b (Hazel, 1943; Mrode, 2005). The set of equations to be solved for b is:

$$b_1p_{11} + b_2p_{12} + \ldots\ldots\ldots b_mp_{1m} = w_1 g_{11} + w_2 g_{12} + \ldots\ldots w_m g_{1m}$$

$$b_1p_{21} + b_2p_{22+} \ldots\ldots\ldots b_mp_{2m} = w_1g_{12} + w_2g_{22} + \ldots\ldots\ldots w_mg_{2m}$$

$$\vdots \qquad\qquad \vdots$$

$$b_mp_{m1} + b_2p_{m2+} \ldots\ldots\ldots b_mp_{mm} = w_1g_{m1} + w_2g_{12} + \ldots\ldots w_m\, g_{mm}$$

where p_{mm} and g_{mm} are the phenotypic and genetic variances, respectively, for individual or trait m; and p_{mn} and g_{mn} are the phenotypic and genetic covariances, respectively, between individuals or traits m and n, w_i represents the economic values for the traits in the selection objectives. These equations can be expressed in matrix notation as:

$$Pb = Gw$$

Therefore:

$$b = P^{-1}Gw$$

Where *P* is the variance and covariance matrix for observations, *G* is the covariance matrix between observations and breeding value to be predicted, *w* is the vector of the economic weights.

Therefore the selection index equation is:

$$I = (\hat{a}_i) = P^{-1}Gw(y-\mu)$$

$$= b'(y-\mu),$$

where μ represents the estimates of environmental influences on the observations, which are assumed to be known without error (Mrode, 2005).

It should be noted that some traits in the aggregate breeding value may not be in the index as they are difficult to measure or may not be available until late in life. These traits may be replaced in the index with other traits that are highly correlated and are easy to measure or occur early in life (Cameron, 1997).

2.4.3 Accuracy of selection criteria

As before, the accuracy (R_{IH}) is the correlation between *I* and *H*; the higher the correlation the better the index is going to be as a predictor of the true breeding value. Knowledge of the accuracy helps in evaluating different indices based on different traits and in testing whether a particular trait is worth including in an index or not (Mrode, 2005). From the above definition the accuracy can be written as (Hazel, 1943; Mrode, 2005; Cameron, 1997):

$$R_{IH} = \frac{cov(I,H)}{\sigma_I \sigma_H}$$

To obtain this, first the variance of the index needs to be calculated:

$$\sigma_I^2 = var\,(b_1X_1) + var\,(b_2\,X_2) + \cdots + 2b_1b_2cov(X_1,X_2) + \cdots$$

$$= b_1^2 var\,(X_1) + b_2^2 var\,(X_1) + \cdots + 2b_1b_2\mathrm{cov}(X_1,X_2) + \cdots$$

$$\sigma_I^2 = b_1^2\mathrm{p}_{11} + b_1^2\mathrm{p}_{22} + \cdots + 2\mathrm{b}_1\mathrm{b}_2\mathrm{p}_{12} + \cdots$$

Or in matrix notation:

$$\sigma_I^2 = \acute{b}Pb.$$

As $b=P^{-1}Gw$ the variance of the index can be written as:

$$\sigma_I^2 = \acute{G}wP^{-1}Gw.$$

The covariance between the true breeding value for trait or individual *i* and the index is:

$$cov(H_i, I) = cov(H_i, b_1X_1) + cov(H_i, b_2X_2) + \cdots + cov(H_i, b_jX_j)$$

which can be written as:

$$cov(H_i, I) = \sum_{j=1}^{m} b_j g_{ij} w$$

where g_{ij} is the genetic covariance between traits or individuals i and j and m is the number of traits or individuals on the index.

In matrix notation the covariance can be written as:

$$cov(H_i, I) = \acute{b}Gw.$$

Expressing b as $P^{-1}G\,w$ gives:

$$cov(H_i, I) = \acute{G}wP^{-1}Gw = \sigma_I^2.$$

Therefore:

$$R_{IH} = \frac{\sigma_I^2}{\sigma_H \sigma_I} = \frac{\sigma_I}{\sigma_H} = \sqrt{\frac{\sum_{j=1}^{m} b_j g_{ij} w}{\sigma_H^2}}$$

2.4.4 Response to selection

In selection for several traits in a selection objective, the response to selection will be the sum of the individual responses in each trait. The correlated response (*CR)* in trait Y_j, to selection on the selection criterion can be derived as:

$$CR_j = b_{jI} SD_I = i_I \frac{COV(Y_j, I)}{\sqrt{var(I)}} = i_I \frac{\acute{b} G_j}{\sqrt{\acute{b} P b}}$$

Where: b_{jI} is the regression coefficient of the trait Y_j on the selection criterion I, G_j is the j[th] column of the G matrix.

The economic value of the response in the selection objective (CR_w) is the sum of the correlated responses for each trait in the selection objective multiplied by its relative economic values (Cameron, 1997):

$$CR_w = i \frac{\acute{b} G w}{\sqrt{\acute{b} P b}} = i \frac{\acute{G} w P^{-1} G w}{\sqrt{\acute{b} P b}} = i \frac{\sigma_I^2}{\sigma_I} = i\sigma_I = i R_{IH} \sigma_H, \qquad as\ b = P^{-1} G w$$

Therefore, the genetic gain which can be obtained by selecting for several traits simultaneously with a group of animals is determined by a standardized selection differential, the multiple correlations between both the aggregate genotype and the selection index (the accuracy of selection R_{IH}), and the genetic variability (Hazel, 1943).

In summary, to predict the breeding value for animals that have different information sources, different sets of index weights have to be derived, and therefore animals can be ranked correctly for selection decisions depending on their index values. However, a selection index (best linear prediction) is not free of bias, and the reason bias exists is that animals were assumed to belong to the same environmental group, and phenotypic differences were due to genetic differences rather than to a combination of genetic and environmental effects. One approach would be to estimate the environmental effects and then to pre-adjust each animal's record for fixed or environmental effects and these are assumed to be known. However, in practice these are in fact seldom known, particularly when no prior data exist for new subclasses of fixed effect or new environmental effects (Henderson, 1976a; Hazel, 1943). Furthermore, the index equations to be solved need the inverse of the covariance matrix for observations, and this may not be computationally feasible particularly if the number of relatives is large (Henderson, 1997b; Mrode, 2005). Selection index methodology requires the genetic covariances between individuals to be known in order to construct the genetic covariance matrix which is used in the prediction of breeding values. However, selection indices use records on certain

close relatives of an animal in order for its breeding value to be predicted, ignoring the information which could be provided by other relatives (Hazel, 1943, Simm, 2000). For these reasons the use of selection indices for genetic evaluation has been superseded by mixed model procedures which were introduced by Henderson (1949). Mixed model equations combine the properties of a linear model (least square) and a selection index (combination of information from different sources) and can simply be reduced to a selection index when no adjustments for environmental effects are required (Mrode, 2005).

2.5 Best linear unbiased prediction (BLUP)

2.5.1 Introduction

Over the last few decades, breeding values estimated by best linear unbiased prediction (BLUP) incorporated into overall economic indices have become widely applied in animal breeding. In BLP all fixed effects were assumed to be known. However, in practice, all fixed effects are seldom known, therefore the selection index is not appropriate. In this case best linear unbiased prediction (BLUP) which was developed by Henderson (1949) is optimum, and can be used to predict the breeding values and the environmental effects simultaneously (Henderson, 1976). BLUP can correct for all the possible environmental factors that affect the records, which are estimated simultaneously with the breeding values.

The properties of the BLUP procedure are as follows:

- Best: maximization of the correlation between the true breeding value and the predicted breeding value.
- Linear: predicted breeding values are a linear function of observations.
- Unbiased: estimation of the values of the random variables (such as animal breeding values) and of estimable functions of fixed effects are unbiased, the unknown true breeding values being distributed around the predicted breeding values.

- Prediction: involves prediction of the true breeding value.

Henderson (1949) developed a set of equations that simultaneously generate the best linear unbiased estimator (BLUE) of the fixed effects and the best linear unbiased predictor (BLUP) for the random effects, and these equations were called mixed model equations (MME) (Henderson, 1988). Moreover, Henderson (1973; 1976) showed that the genetic relationships among all animals that are to be evaluated can be incorporated in BLUP evaluations of their breeding values. This method requires the inverse of the numerator relationship matrix (A) of these animals. However, computing (A) then inverting it might become impossible with large numbers of animals. Therefore, Henderson (1976) introduced a fast and simple method for computing the inverse of (A) directly from pedigree information without the need to construct the (A) matrix itself.

BLUP is widely used in genetic evaluation of domestic animals because of its desirable statistical properties, which have been enhanced by the steady development of computing power. BLUP was originally used to predict sire breeding values depending on progeny measurements, but it has since been extended to deal with various prediction problems, and to predict the breeding values of all the animals in a pedigree (Henderson, 1973). Currently, there are many computer packages for BLUP evaluations such as AsReml, which has been used for analysis of oestrous behaviour characteristics in the present work. In the following sections the theory underlying BLUP is presented, together with its major applications.

2.5.2 The linear mixed model

Linear models are the most common type of statistical models used in animal breeding to predict the random effects of animals based on their phenotypic performance. Linear models form the basis of BLUP as it provides the ability to correct breeding values for fixed effects and allows them to be estimated simultaneously (Henderson, 1973; 1975). The general mixed model for such a prediction is:

$$y = Xb + Za + e. \qquad (4)$$

The observations are a function of the fixed effects, the genetic merits of the individuals and an error term where:

y = $n \times 1$ is an observations vector for n number of records,

$b = p \times 1$ is vector of fixed effects; p is the number of levels for fixed effects,

$a = q \times 1$ is vector of random animal effects; q is the number of levels for random effect,

$e = n \times 1$ is vector of non-observable random residual effects,

X is the design matrix of order $n \times p$, which relates records to fixed effects,

and Z is a design matrix of order $n \times q$, which relates records to random animal effects.

Both X and Z are termed incidence matrices in which each element consists of either a zero or one, depending on the level of the fixed or random effect at which each animal is classified.

It is assumed that:

- The expectation (E) to the variables is: *E(Y)=Xb, E(a)= E(e)=0*
- Residual effects, including random environmental and non-additive genetic effects, are independently distributed with variance σ_e^2 therefore, $var\ (e) = I\sigma_e^2 = R, var\ (a) = A\sigma_a^2 = G$ and $cov\ (a, e) = cov\ (e, a) = 0$ where: G is a known matrix and A is the numerator relationship matrix.

It follows then that:

$$var(y) = V = var(Za + e) = ZG\dot{Z} + R.$$

The solution to equation (4) is based on the following principles:

- The linear functions of *b* and *a* indicate that *K′b+ a* is to be predicted (the predictand) using the linear function of the records *y* that is *L′y* (the predictor) given that *K′b* is estimable. The predictor *L′y* is chosen such that:

$$E(\hat{a}) = E(a).$$

- The average square error of prediction, that is $E(\hat{a} - a)^2$ is minimal. This minimization leads to the BLUP of *a*:

$$\hat{a} = BLUP(a) = G\grave{Z}V^{-1}\left(y - X\grave{b}\right).$$

$$\grave{L}y = \grave{K}\hat{b} + G\grave{Z}V^{-1}\left(y - X\grave{b}\right),$$

where: $\hat{b} = (\grave{X}V^{-1}X)\grave{X}V^{-1}y$ is the generalized least-square solution (GLS) for *b* and $\grave{K}\hat{b}$ is the BLUE of $\grave{K}b$, the best linear unbiased predictor is BLP with the GLS estimator of the fixed effects substituted with the corresponding parameter (Henderson, 1973; 1988). However, the solutions for *a* and *b* require the inverse of *V* which is not always computationally feasible, particularly for large numbers of records. Therefore, Henderson (1949) developed a set of equations that simultaneously generate the best linear unbiased estimator (BLUE) $\acute{k}\hat{b}$ of any set of estimable function $\acute{k}b$, and the best linear unbiased predictor (BLUP) $\hat{a}$ of *a*, without requiring the inverse of *V*, these equations being called mixed model equations (MME):

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}.$$

Since R is an identity matrix it can be factorized from both sides of the equation:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \; where\; \alpha = \sigma_e^2 \; / \sigma_a^2. \qquad (5)$$

The solutions of *a* and *b* from the MME were proved to be the GLS of *b* and BLUP of *a* which have the following properties:

- The solutions are unbiased in the sense that the predictor has the same expectation as the variable to be predicted (the predictand).
- Under the class of linear unbiased prediction, they minimize the error of prediction; that is they maximize the correlation between the predictor and predictand and increase the possibility of correctly ranking animals.

So the MME method provides a powerful computational tool that can estimate environmental and genetic effects simultaneously. It is also considered a general model in terms of the flexibility of the choice of what terms can be included in the model, and which are random and which are fixed (Henderson, 1988; Mrode, 2005). It also helps in eliminating the possible bias in predicting the breeding values due to selection and culling where no prior estimate of the population parameters is available. Selection can lead to changes in genetic variance resulting from inbreeding and linkage disequilibrium. So, by assuming that *y*, *a*, and *e* are multivariate and normally distributed (implying that traits are controlled by a large number of loci), and both G and R for the base population are known or at least known to proportionality, MME can account for selection, and yield the maximum likelihood and BLUE of the predictand (Kennedy et al, 1988).

2.5.3 Sire models

Most early applications of BLUP for the prediction of breeding values were based on sire evaluations, particularly in dairy cattle where progeny testing of sires is practiced. In this model, only the effects of sires (i.e. half of their breeding values) are predicted from observations on their progeny (Henderson, 1973). Therefore, reducing the number of equation required in comparison with the animal model (discussed in the next section), for example we may have 100 sires in a dataset of 100.000 recorded animals therefore making for computational ease (Mrode, 2005).

The equation of the sire effects is:

$$y = Xb + Zs + e.$$

The terms are as defined in equation (4) where s is the vector of the sire's random effect, Z relates records to sires and:

$$var(s) = A\sigma_s^2.$$

$$var(y) = ZA\grave{Z}\sigma_s^2 + R$$

where A is the numerator relationship matrix for sires, $\sigma_s^2 = 0.25\sigma_a^2$ and $R = I\sigma_e^2$. The MME are exactly the same as in (5) except that$\alpha = \sigma_e^2 \ /\sigma_s^2 = \left(\frac{4-h^2}{h^2}\right)$.

However, the estimated breeding values obtained from the sire model may lack accuracy, for instance when sires have few progeny (Henderson, 1973), or because there is no correction for differences between dams. This model assumes that all progeny of a sire are from different dams and all dams are expected to be form similar populations all with the same expected mean. In practice, dams might belong to different breeds, and in addition, dams may be selected over time making younger dams better than older dams (Mrode, 2005).

2.5.4 The animal model

BLUP under the animal model has become the international reference for animal evaluation. It is a set of many different models, their common feature being that all animals in a population are evaluated jointly. Animal models take into account all known relationships among animals in order to predict each animal's genetic merit, rather than predicting breeding values only for sires (Henderson, 1988). The information from these relationships is accounted for through the inclusion of the numerator relationship matrix (A) computed from the pedigree information (Henderson, 1976).

The basic formula to describe the observation under the animal model is:

$$y = Xb + Za + e.$$

The terms are as defined in (4).

The mixed model equation under the animal model can be written as:

$$\begin{bmatrix} X'X & X'Z \\ ZX & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}.$$

In this equation $\alpha = \sigma^2_e / \sigma^2_a$, A^{-1} is the inverse of the numerator relationship matrix among all animals in *a*, which has nonzero off-diagonals only for the animal's parents, progeny, and mates (Henderson, 1976). The elements of *a* can contain additive genetic effects, non-additive genetic effects, maternal effects, and permanent environmental effects (Henderson, 1988).

Under this model the animal's own performance is presented as Yield Deviation (YD) which is the weighted average of animal yield adjusted for all effects other than the genetic and error effects. Each animal's breeding value contains information from its own performance (YD), from its parents as Parent Average evaluation (PA, average breeding value of parents), and its Progeny Contribution adjusted for the genetic merit of the mate (PC, which is the weighted average of twice the progeny's breeding value minus the mate's breeding value). So for animal *i* the breeding value can be written as:

$$a_i = n_1(PA) + n_2(YD) + n_3(PC).$$

Here n1, n2, and n3 are weights that sum to one (VanRaden & Wiggans, 1991).

2.5.5 Multiple traits model

As with the selection index model, BLUP multiple traits can be applied for more accurate animal evaluation as the method accounts for the relationship between traits. This model was first introduced by Henderson and Quaas (1976). Multiple trait analysis involves simultaneous evaluation of animals for two or more traits, and accounts for the phenotypic and genetic correlations between these traits. With information on correlated traits the accuracy of the estimated breeding value increases (Henderson and Quaas, 1976). A second advantage is that multiple trait analysis is the only way to obtain unbiased estimates for a trait which is observed only on animals which have been selected based on values of a correlated trait. For example, in dairy cattle when selection is practised based on first lactation records, only cows that survived the first lactation will have the chance to be measured for second lactation and those are usually the better cows. Therefore, a model including information on the correlated trait, on which selection was based, can be used to correct for this type of selection, and thus it accounts for culling selection bias (Mrode, 2005).

The mixed model for multivariate analysis is similar to univariate models for each of the traits. With two traits for example, the observation vector *y* can be partitioned

into two parts, one for each trait. The same can be done with the random and fixed matrices. That is for trait 1:

$$y_1 = X_1 b_1 + Z_1 a_1 + e_1$$

and for trait 2:

$$y_2 = X_2 b_2 + Z_2 a_2 + e_2.$$

By ordering the vectors of *y* and *a* by animals within traits, the multivariate model equation for the two traits can be written as:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}.$$

The terms are as defined before, for example: y_1 is the vector of observation for trait 1, b_1 is the vector of the fixed effects for the trait 1 and so on.

The mixed model equation is of the form:

$$\begin{bmatrix} \grave{X}R^{-1}X & \grave{X}R^{-1}Z \\ \grave{Z}R^{-1}X & \grave{Z}R^{-1}Z + A^{-1} \otimes G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} \grave{X}R^{-1}y \\ \grave{X}R^{-1}y \end{bmatrix}$$

where: $X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}, \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix}, \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix}$ *and* $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$,

$A^{-1} \otimes G^{-1} = \begin{bmatrix} g_{11}^{-1}A^{-1} & g_{12}^{-1}A^{-1} \\ g_{21}^{-1}A^{-1} & g_{22}^{-1}A^{-1} \end{bmatrix}$ is the direct product of the two matrices which is obtained by multiplying A^{-1} by the diagonal and off-diagonal elements of the G^{-1}.

This model however, has a major impact on the number of equations to be solved by significantly increasing computing time. In addition, this model requires the genetic and phenotypic correlations among traits, and these may not be readily available (Mrode, 2005).

2.5.6 Repeatability model

The BLUP repeatability model is applied where animals have more than one record on the same trait, such as milk yield in successive lactations or litter size in

successive pregnancies. The phenotypic correlation between records is equal to the repeatability, while the genetic correlation between records is assumed to be unity. The principles and the assumptions which comprise this model have been covered in section 2.2.2. The aim of this approach to genetic evaluation is not only to estimate an animal's breeding value but also to derive its permanent environmental effects. The repeatability model can be written as:

$$y = Xb + Za + W_{pe} + e,$$

where *y* is the vector of observations, *b* is a vector of fixed effects, *a* is a vector of random animal effects, *pe* is the vector of permanent environmental effects and non-genetic effects, *e* is a vector of random residual effect, while *X, Z*, and *W* are incidence matrices relating records to fixed, animal, and permanent environmental effects respectively. Moreover, the permanent environmental effects and residual effects are assumed to be independently distributed with means of zero and variance σ_{pe}^2 and σ_e^2 respectively. Therefore:

$$var\begin{bmatrix} pe \\ e \\ a \end{bmatrix} = \begin{bmatrix} I\sigma_{pe}^2 & 0 & 0 \\ 0 & I\sigma_e^2 & 0 \\ 0 & 0 & A\sigma_a^2 \end{bmatrix},$$

where:

$$var(e) = I\sigma_e^2 = R,$$

$$var(y) = ZA\grave{Z}\sigma_a^2 + WI\sigma_{pe}^2\grave{W} + R$$

The mixed model equation for the repeated model is formed as:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{pe} \end{bmatrix} = \begin{bmatrix} \grave{X}R^{-1}X & \grave{X}R^{-1}Z & \grave{X}R^{-1}W \\ \grave{Z}R^{-1}X & \grave{Z}R^{-1}Z + A^{-1}\,1/\sigma_a^2 & \grave{Z}R^{-1}W \\ \grave{W}R^{-1}X & \grave{W}R^{-1}Z & \grave{W}R^{-1}W + I(1/\sigma_{pe}^2) \end{bmatrix}^{-1} \begin{bmatrix} \grave{X}R^{-1}y \\ \grave{Z}R^{-1}y \\ \grave{W}R^{-1}y \end{bmatrix}$$

By dropping R^{-1} from both sides of the equation, the MME can be written as:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{pe} \end{bmatrix} = \begin{bmatrix} \grave{X}X & \grave{X}Z & \grave{X}W \\ \grave{Z}X & \grave{Z}Z + A^{-1}\alpha_1 & \grave{Z}W \\ \grave{W}X & \grave{W}Z & \grave{W}W + I\alpha_2 \end{bmatrix}^{-1} \begin{bmatrix} \grave{X}y \\ \grave{Z}y \\ \grave{W}y \end{bmatrix},$$

where: $\alpha_1 = \sigma_e^2/\sigma_a^2$ and $\alpha_2 = \sigma_e^2/\sigma_{pe}^2$ (Mrode, 2005).

2.5.7 Conclusion

Selection indices (BLP) have been applied in order to obtain the optimum weightings of records combining information on several traits and individuals. However, prediction with BLP is not free of bias, as a result of the need for the records to be pre-adjusted for fixed effects. In addition, selection for multiple traits using relatives' records has been computationally difficult particularly with a large number of data sets. Henderson (1973) pointed out how the genetic relationships among all the animals in a population can be incorporated in best linear unbiased prediction (BLUP) or MME of their merit. MME allows simultaneous estimation of the fixed effects and prediction of the random effects. This method requires the inverse of the relationship matrix (A^{-1}) among the animals in the model. If the number of animals is large, finding A then inverting it becomes impossible. To solve this problem, Henderson (1976) introduced a simple method to compute the inverse of A from pedigree file without the need to be constructed itself. There are many more different BLUP models, which are not considered here, such as the common environmental effect model and the maternal genetic effects model. The application of these models depends on the traits of interest and the structure of the data.

The previous sections have dealt with the estimation of the breeding values using information on phenotype and pedigree. However, it has been suggested that breeding values can be estimated more accurately using information on variation between individuals at the DNA level (the so called genomic selection, Goddard and Hayes, 2007).

2.6 Genomic selection

Genomic selection (GS) is defined as the simultaneous selection for many (tens or hundreds of thousands of) markers, which cover the entire genome in a dense manner so that all quantitative trait loci (QTL) are expected to be in linkage disequilibrium with at least some of the markers (Meuwissen, 2007). In fact, genomic selection was demonstrated to be a form of marker assisted selection in which information about genetic markers that cover the whole genome is used to estimate the genomic breeding value (GEBV) upon which a selection decision is made (Goddard and

Hayes, 2007). The methodology for GS was first presented by Meuwissen et al. (2001). This revolution in genomic selection was facilitated by the recent sequencing of the bovine genome resulting in the discovery of many thousands of DNA markers in the form of SNPs (around 50,000 SNPs; Schaeffer, 2006). In parallel with the discovery of numerous SNP markers throughout the livestock genome, the cost of genotyping has been significantly reduced (with genotyping of a SNP now costing as little as 1US cent per animal (Goddarad and Hayes, 2007; Hays et al., 2009; Meuwissen, 2007).

To calculate GEBV, a prediction equation based on the SNP is first derived. The entire genome is divided into small segments, and then the QTL effects, deduced from either haplotype or SNPs markers, are calculated in a large reference population in which animals have both phenotypic and genotypic records. In this way, the effects of all loci that contribute to genetic variation of the trait are detected, even if the effects of the individual loci are small (Hayes, 2009). In subsequent generations, animals (called the selection population) need only be genotyped for the markers to determine which chromosome segments they carry, and the estimated effects of the segments the animal carry can then be summed across the whole genome to predict their GEBV (Goddard and Hayes, 2007). Meuwissen et al. (2001) demonstrated in simulations that using marker information alone, breeding values can be predicted with accuracies up to 0.85. A variety of methods have been suggested for the calculation of GS-estimated breeding values (EBV), ranging from BLUP, Bayesian methods such as BayesB, and machine learning techniques. These methods differ in their assumptions about the underlying genetic model. BLUP assumes the infinitesimal model (traits are determined by a large number of genes, each with small effects at many infinitely unlinked loci). Therefore, the QTL effects are drawn from a normal distribution with every gene across the entire genome having the same variance (Meuwissen, 2003). The machine learning technique assumes that the traits are affected by a limited number of genes, and that there is therefore a limited number of SNPs that are worth fitting (Meuwissen, 2007). BayesB is in between the previous two methods, i.e. it assumes that many genes are likely to have small effects on the trait and only a few genes will have moderate to large effects. It is similar to BLUP-GS in the sense that QTL effects are also drawn from a normal distribution but with the variance of the allelic effects assumed to be different for every gene

across the whole genome (Meuwissen et al, 2001; Meuwissen, 2003; Goddard and Hayes, 2007). The assumption on which the Bayesian method was based has resulted in slightly more accurate estimates of the GEBV than the BLUP method (Goddard and Hayes, 2007). However, BLUP methodology has been an attractive method in estimating marker effects as the only pre-requisite information is the additive genetic variance of the trait. In addition, for most traits in dairy cattle, the assumption by the BLUP method of an infinitesimal model was demonstrated to be closer to reality (Hayes et al., 2009). The BLUP-GS model, where the marker effects are estimated by BLUP, is equivalent to the traditional BLUP model, where the usual pedigree-based relationship matrix is replaced by a relationship matrix estimated by the markers (Meuwissen, 2007).

The application of genomic selection allows prediction of accurate GEBVs for young bulls that are not progeny tested. The reliabilities (square accuracy) of GEBV for young bulls fall in the range between 20 and 67% which is significantly greater than the reliabilities of parental average breeding values, the current criteria for selection of bull calves to enter progeny test. (Hayes et al, 2009). In the UK, GEBVs are speculated to be around 40-60% which is higher than the reliability of a pedigree index that is based on parental average (35%) but lower than the reliability of progeny proof (>80%; DairyCo, 2008). The accuracy of GEBV was demonstrated to depend on the following factors:

1- The degree of linkage disequilibrium (LD) between the marker and the trait: in order for the markers to be able to predict the effect of the QTL across the population and across generation, they must be in sufficient LD. The level of LD between the QTL and a single marker, or a linear combination of markers, is measured in terms of r^2 parameter which was defined as the proportion of variation in the trait that is caused by alleles at the QTL which is explained by the marker. The average value of r^2 decreases as the distance between the QTL and the markers increases. The accuracy of GEBV was found to increase dramatically as r^2 increased with value of ≥ 0.2 is sufficient to achieve accuracy of 0.8 (Goddard and Hayes, 2007; Meuwissen et al., 2001).

2- The number of animals with phenotypic and genotypic records in the reference population from which the SNP effects are estimated: the more phenotypic records the more observation will be per SNP and therefore the greater accuracy of the genomic selection (Meuwissen et al., 2001; Hayes et al., 2009).
3- The heritability of the trait: for traits with low heritability such as fertility, a large number of records are required to achieve high accuracies in GEBV (Meuwissen et al., 2001, Hayes et al., 2009).
4- The distribution of the QTL effects: in the case where many QTL with small effects were assumed to contribute to the variation in the trait, many phenotypic records are required to achieve high accuracy (Hayes et al., 2009).

Because the response to selection is related directly to the accuracy of estimated parameters (i.e. heritability and breeding value), it is imperative in practical breeding programs to achieve an accuracy as high as possible. This will depend on the experimental design including the number of animals evaluated and environmental effects in addition to the effect of the models applied in the estimation. This topic is covered in the following section.

2.7 Precision of the estimates of genetic parameters

Selection experiments can be used to estimate heritabilities or other genetic parameters in a population and to compare responses under different selection programmes. It is therefore essential to have some information on the precision of the estimates obtained (Hill, 1971). Precision of the estimate is determined by the standard error which is equal to the square root of sampling variance (Falconer, 1960). One of the most important properties of quantitative traits is their heritability which has a predictive role in term of expressing the reliability of the phenotypic values as a guide to breeding values (Cameron, 1997). Thus the success rate in genetic improvement of any trait can be predicted from the knowledge of its heritability (Falconer, 1960). In general, when an experiment aimed at estimating heritability, the design and the methods of the experiment must be chosen such that the highest possible accuracy will be obtained given the limitation imposed by the

scale of the experiment (Falconer, 1960). The accuracy of the estimate depends on its sampling variance, the lower the sampling variance the greater the accuracy (Robertson, 1959). The magnitude of sampling variance is associated with the experimental design and the value of the heritability being estimated. The heritability is estimated as:

$$h^2 = \frac{\sigma_a^2}{\sigma_P^2}$$

where: σ_a^2, σ_P^2 are the additive genetic variance and phenotypic variance respectively.

Therefore, the estimate of the heritability depends on the magnitude of all the components of variance and the change in any one of them will affect it (Cameron, 1997). All the genetic components are influenced by gene frequency and this may change from one generation to another. In particular, in small populations they are expected to have smaller heritabilities than in larger populations due to the loss of genetic variation (Falconer, 1960). The environmental variance which includes all non-genetic variation is dependent on the management and conditions in which individuals are kept: more variable conditions reduce the heritability, more consistent conditions increase it. In general, environmental variance is a source of error that reduces precision of the estimate in genetic studies and therefore consideration must be given to reducing it as much as possible by careful experimental design, i.e. nutrition and climatic factors are among the most common causes of environmental variances. Other factors include the error of measurements. For example, when a trait can be measured in units in length or weight, it is usually measured accurately. Some traits cannot be measured directly but can be graded such as carcass quality, and these traits will have substantial variance that is due to measurement (Falconer, 1960). Furthermore, variations can be caused from unknown factors and these usually cannot be eliminated by experiment design; these are called intangible variations (Falconer, 1960).

Furthermore, when designing an experiment, the class of relatives which will be recorded must be considered (these were discussed in section 2.1) as this will affect the sampling variance of the estimated parameters (Hill, 1971). In general, experiments contain information on the performance of both the parents and several

of their progeny. It is then possible to estimate heritability in two ways, firstly from the regression of progeny's phenotype on the phenotype of one parent or on the phenotype of both parents, and secondly from the half-sib or full-sib correlations (Hill, 1974). In the regression method, the estimate makes no use of the variance between the members of the same family or of the variance between the family means. In the intra-class correlation method, no use is made of parental performance (Hill, 1974).

However, major issues with experiment design are the choice of method and the decision of how many individuals in each family are to be measured (Robertson, 1959). As the number of individuals measured is limited by the scale of the experiment, an increase in the number of individuals per family necessitates reducing the number of families. The problem is to find a balance between the number of families and the number of individuals per family (Robertson, 1959, Falconer, 1960). Therefore, it is necessarily to consider all the conditions that might impose limitations on the scale of the experiment, such as labour or cost (Falconer, 1960). For example, if labour is the limiting factor, such as in visual heat detection, then the limitation is in the total number measured.

In 1971, Hill showed that the optimal number of individuals that must be recorded (*T*) in order to obtain the minimum sampling variance for heritability estimates ($V(h^2)$), was:

$$T = \frac{v}{100V(h^2)}$$

where v is a tabulated coefficient provided by Hill (1971).

The optimal family size that must be considered in order to obtain the highest precision of estimates of genetic parameters will be discussed in regard to the above mentioned offspring-parent regression method and sib analysis.

Offspring-parent regression

Let N be the number of families, thus the number of parents measured is N or 2N depending whether one or both parents are measured, and n be the number of offspring per family, therefore, the total number of individuals measured *T* will be with one parent: *T = N(n+1)* while with both parents: *T = N(n+2)* (Hill, 1974). Let *b* denotes the regression coefficient which is a measure of $\frac{1}{2}h^2$ when estimated by regression of offspring phenotype on one parent phenotype, while when using the regression on both parents, *b* is an estimate of h^2 (Cameron 1997). Using these symbols, the variance of the regression coefficient (σ_b^2) on one parent can be presented as:

$$\sigma_b^2 = \frac{1 + (n-1)t}{nN} \qquad (6)$$

where *t* is the intra-class correlation between offspring which is equal to $\frac{1}{4}h^2$ and $\frac{1}{2}h^2$ for half-sib and full-sib offspring respectively assuming that there is no additional covariances between the full-sibs (see section 2.3.2; Falconer, 1960). It has been shown that sampling variance in (6) is minimal when:

$n = \sqrt{\frac{(1+t)}{t}}$ when one parent is measured, or $n = \sqrt{\frac{2(1+t)}{t}}$ when both parents are measured.

Furthermore, as the regression coefficient *b* is equal to $\frac{1}{2}h^2$ or h^2 when one or both parents are measured respectively, the sampling variance for the heritability can be expressed relative to *b* as follows:

By regression on one parent: $h^2 = 2b$, thus $\sigma_{h^2}^2 = 4\,\sigma_b^2$
By regression on both parents: $h^2 = b$, thus $\sigma_{h^2}^2 = \sigma_b^2$, where $\sigma_{h^2}^2$ is the variance of the heritability

Therefore, an estimate of heritability based on both parents has a considerably lower sampling variance than that based on measurement of one parent (Falconer, 1960).

Sib analysis

When estimating heritability using sib information it is usually assumed that the only covariance between family members is that from additive genetic variance. However, with full-sib measurement there are additional covariances that are due to maternal, common environmental effects and non-additive genetic effects, especially dominance, and these will introduce bias to the estimate of heritability (Cameron, 1997). Therefore intra-class correlation estimates of heritability are normally made from the covariance of half-sibs.

Let *N* be the number of families and *n* be the number of individuals per family, so that the total number of individuals measured will be *T* = *nN*. So for more precise estimates of the genetic parameters one needs to take into account the optimal family size that will results in a minimum sampling variance (Robertson, 1959). Given *t* is the intra-class correlation between sibs records, it has been demonstrated that the sampling variance of *t* can be expressed as:

$$\sigma_t^2 = \frac{2[1 + (n-1)t]^2(1-t)^2}{n(n-1)(N-1)}$$

When the total number of measurements *T* is limited by the size of the experiments it has been demonstrated that the sampling variance of *t* is minimal when *nt =1* or when *n = 1/t* (where n is the number of individuals per sib family). As discussed before, in the case of half-sib measurements $t = \frac{1}{4}h^2$ while in the case of full-sib $t = \frac{1}{2}h^2$. Therefore, the optimal family size depends on the heritability and the most efficient design for half-sib and full-sib analyses has a family size *n* that is equal to $\frac{4}{h^2}$ and $\frac{2}{h^2}$ respectively (Robertson, 1959; Falconer, 1960).

Robertson (1959) has summarized the general rules that outline the optimal family size for sib analysis as follows:

1-when the heritability of a trait is known:

- In the case of half-sib analysis: the expected intra-class correlation for a half-sib analysis is $t = \frac{1}{4}h^2$, therefore the optimal family size is given by $n = \frac{4}{h^2}$.
- In the case when half-sib families are subdivided into full sib families such as there are s sires each mated to d dams and each dam has n progeny the heritability can be estimated using the correlations from both dam and sire components but with the assumption that there are no common environmental variances. It is desired to obtain precise estimates of both the dam and the sire intra-class correlations. In this case it has been suggested that the optimal dam family size should be given by $n = \frac{2}{h^2}$ with 3 or 4 dams per sire. Even if this family size cannot be achieved, it was still recommended to use 3 to 4 dams per sire. If the family size is below the optimum, the sire correlation will be estimated more accurately than the dam correlation and vice versa.
- A structure involving small groups of 2 or 3 animals per family is most inefficient, as variances in the half-sib intra-class correlations become extremely high.

2- If there is no previous evidence on the heritability, then the optimum family size for a half-sib analysis is 20 to 30, while for full-sib analysis, the optimum dam family size is 10 with 3 or 4 dams per sire.

By comparing the above two methods, it has been demonstrated that with low heritability traits the sib analysis is more preferable, while the regression method is more suitable for moderately high heritability traits as this methods will result in lower sampling variance (Hill, 1971; Falconer, 1960).

In the previous sections, molecular markers for fertility in different species have been reviewed. As infertility is one of the most serious problems facing dairy cattle industry, the current work will focus on finding molecular markers for fertility in dairy cattle along with exploring the possibility of genetically improving fertility through introducing new measures of fertility.

2.8 Objectives of the project

The objective of this study was to quantify the associations between different novel and known SNPs in many genes (*GnRHR, LHR, FSHR, activin receptor, neuropeptide Y receptorY2, oestrogen receptor α)* and fertility in dairy cattle, using predicted transmitting abilities (PTAs) for fertility traits available through the UK fertility index (Wall et al., 2003). In addition, this project aimed to test whether these SNPs can provide useful molecular markers for traits of oestrus expression which are used as different measures of fertility. This study also aimed to provide genetic parameters and breeding values for oestrous behaviour traits.

2.8.1 Hypothesis

- Polymorphisms at the different genes explain some of the variation associated with fertility traits PTAs and oestrous behaviour traits.
- Automated heat detection using electronic devices are not only of value in oestrus detection but could also be of particular importance in recording fertility traits for genetic evaluation.

Chapter 3: Materials and Methods

The methods described in this Chapter are general methods applicable to all the later Chapters in the thesis. The theoretical basis of the statistical methods used here have been covered in detail in Chapter 2, and are not described again here.

3.1 Animals

Numbers of animals in each of the groups studied, together with the sources of the data related to them, are given in Table 3.1. The animals analysed in Chapter 4 (group a) included a number of bulls for which semen samples were available through Prof John Woolliams at the Roslin Institute, Edinburgh, together with other animals (dams and sires) held at the Nottingham University Farm. Those treated with GnRH were part of a study carried out by Dr M. Royal, Prof. APF Flint and others on pre-pubertal LH secretion in response to GnRH, and these animals were held at the farms of Cogent UK Ltd, in Cheshire. The data on these animals were made available to me by Dr Royal and Professors Woolliams and Flint. The animals in Chapters 6 and 7 were held at the Nottingham University Dairy Centre. These animals were studied at different times; thereby ensuring different animals comprised the two groups.

Table 3.1 Numbers of animals used in each study, and sources of phenotypic data and information on PTAs from the UK Fertility Index.

Chapter	Number of animals studied	Source of phenotypic and PTA data
4	963 (animals genotyped) For PTA analysis, 408 (group a) For GnRH treatment, 431	Prof John Woolliams Dr Melissa Royal Dr Mike Coffey
5	427 Including 407 from group (a) in Chapter 4	Dr Melissa Royal Dr Mike Coffey
6	103	Farm staff; Prof Garnsworthy
7	**189**	Farm staff; Prof Garnsworthy

3.2 Detection of single nucleotide polymorphisms

In nearly all cases the genotyping was carried out commercially by KBiosciences Ltd (Hoddesdon. Herts, UK) under the supervision of Dr K. Derecka. In a limited number of animals and genes, where SNP locations were first established, genotypes which had been obtained by sequencing at the University of Nottingham were included in the analyses. In the case of the work described in Chapters 4 and 5, this had been done before my studies commenced, in order to generate the data that was available to me for genetic analysis. To identify the SNPs analysed in the present study, DNA was isolated from semen by phenol-chloroform extraction (Sambrook et al., 1989) and from whole blood using a modification of the Puregene DNA Purification Kit (Gentra, Flowgen, Nottingham) based on sequential precipitation of protein and DNA with an added phenol/chloroform extraction step. This work was carried out by Dr Kamila Derecka, and had been completed before my association studies began. The coding sequence of the bovine *GnRH receptor* spanning all 3 exons (GenBank® accession no. NM_177514.2) and a fragment of the GnRH receptor gene promoter region containing a putative gonadotroph-specific element

(Ngan et al., 1999; GenBank® accession no. AF034950) were amplified from 100 ng aliquots of genomic DNA in 35 PCR cycles with high fidelity DNA polymerase (Accuzyme Mix, Syngenta Bioline, Little Clacton, Essex). Primers and PCR conditions are given in Table 3.2 and primer positions are given relative to the original numbering used in GenBank® reports. PCR products obtained were purified from gels after electrophoresis (Qiagen, Crawley, Sussex) and sequenced (Beckman CEQ8000 Sequencer). Sizes of the products generated are given in Table 3.2.

Table 3.2 Details of PCR primers and reaction conditions.

Fragment	Sequence of primers	Position of primers[a] (bp)	Length of PCR product	Annealing temperature (°C)
Promoter	F tgctcagcattgtgtgattg R gaggctgcctggtgttagag	752-773 1361-1388	626	59
Exon 1	F aagtgccagaaacacgag R tgtggtccagcaaagatg	36-53 585-600	565	56
Exon 2	F tacatctttgggatgatcc R gatcctgatgaaggaccc	605-623 797-802	220	56
Exon 3	F agtccaagaacaatataccacg R gcctttctttgaactttctatgc	837-858 1077-1098	262	57

- [a]Numbering relates to the GenBank sequences (see Fig. 4.1).

3.3 Statistical analyses

Unless otherwise stated statistical analyses were carried out using REML (linear mixed models) in Genstat (release 8.1). With the GnRH response data: age, sex, date of treatment and genotype were fitted as fixed effects while sire as a random effect. In the SNP-PTAs association analyses, SNPs and PIN were fitted as fixed effects while sires as random effects.

SNP associations and the breeding value of oestrous traits were also evaluated by applying different animal models in ASREML (Gilmour et al., 2009). In chapter 6, the model included the fixed effects of the SNPs, phenotypic calving interval, days in milk, PIN and oestrous month, while the random term included the animal additive genetic effect. The mixed model included the effects of different factors on oestrous activity traits (ACTIVITY, ACTIVITY% and BASE) which were modelled as

dependent on: SNPs (3 levels), parity (3 levels), heat season(1 = winter (January-March),2 = Spring (April-June), 3 = Summer (July-September), 4 = Autum (October-December), oestrus number (15 levels), average daily milk since calving (MYa) and average milk yield over the baseline period (MY10), which were fitted as fixed effects. The random model included the additive genetic variance and the permanent environmental variance of the animals. The theoretical background to these analyses has been discussed in detail in Chapter 2.

3.3.1 Pedigree files

Pedigree information on the preceding three generations was generated for the 103 cows for which oestrous behaviour information was collected and analysed in Chapter 6, while a four generation-structured pedigree was created for the 189 cows in Chapter 7 (this was carried out by myself through searching in the Holstein UK animal data). These files were constructed by including the herd book number (HBN) of the cows, its sire and its dam. Then, the HBN of the parents of the sire and the dam were listed and so on.

The pedigree files were created according to the following principles:

- they had three columns in which the first column includes the animal; the second column includes the sire of the animal while the third includes the dam.
- they were sorted so that the line giving the pedigree of an individual appears before any line where that individual appears as a parent
- they were in a read free format; they were created in an excel file and saved as an ASCII-text file in order to be read by ASREML.
- they were specified on the line immediately preceding the data file line in the .as programme file
- identities of 0 were used for unknown parents (Gilmour et al., 2009).

For example, if we have a cow with a pedigree as illustrated in the figure below:

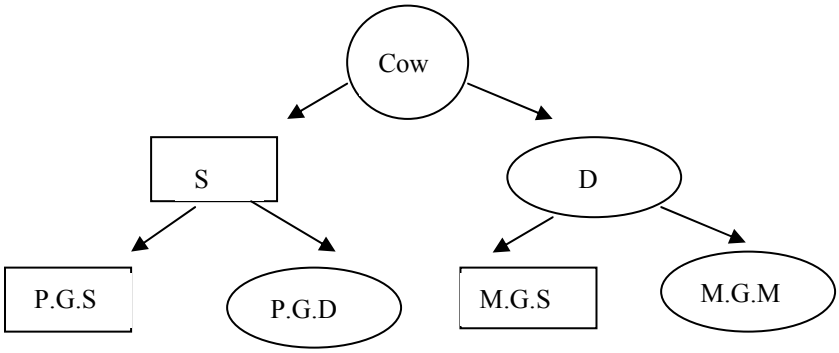


Figure 3.1 Three generation pedigree for a given cow, S: sire, D: dam, P.G.S: paternal grandsire, P.G.D: paternal granddame, M.G.S: maternal grandsire, M.G.D: maternal granddame,

then its pedigree can be organized as follow:

Animal	Male parent	Female parent
Cow	Sire	Dam
Sire	P.G.S	P.G.D
Dam	M.G.S	M.G.D

The information from pedigree files were then matched to SNPs and pedometer information in the data files. Further information on these files is given in more detail in Chapters 6 and 7.

3.3.2 SNPs in genes investigated in relation fertility PTAs

Different genes were selected as candidate gens in term of identification of molecular markers for fertility in dairy cattle. These genes were selected on the basis of their involvement in the hypothalamic/ovarian/uterine axis (*GnRH receptor*, *LH receptor, FSH receptor, oestrogen receptors, activin receptor*) and their known roles in central nervous pathways controlling reproduction (*neuropeptide Y and its receptor*).

3.3.3 SNPs in genes investigated in relation to oestrous behaviour

Most of the SNPs investigated here for associations with phenotypic data had been identified before this project started. However, a small number of cows were

analysed for SNPs specifically for this project, in order to test associations with genes involved in the control of oestrous behaviour. Candidate genes identified for this purpose were: *leptin, neuropeptide Y, oestrogen receptor β* and *ghrelin*. The rationale for investigating these genes has been discussed in Chapter 1 (Section 1.4.2).

SNPs were already known in some of these candidate genes (e.g. 2 SNPs shown to be related to fertility in the leptin gene, see Chapter 1; Liefers et al., 2005). The other 3 genes (*neuropeptide Y, oestrogen receptor β* and *ghrelin*) were studied for the first time in cattle in an extensive PCR/sequencing program involving 15 exons and associated flanking (intronic) sequences (Table 3.3). In these genes, previously unreported SNPs were identified in exon 1 in the *neuropeptide Y* gene and in exons 4 and 7 of the *oestrogen receptor β* gene. No SNPs were identified in exons in the bovine *ghrelin* gene. Therefore this project identified, in exons of candidate genes, 3 SNPs not previously reported in cattle. SNPs were also identified in introns in these genes, but these were not investigated further as they are less likely to be responsible for functional effects than SNPs in exons.

Table 3.3 Novel candidate gene sequences investigated in this part of the work.

Gene	**Neuropeptide Y**							
EXON	1	2	3					
Length	188	81	99					
Addition	166/18	220/14	125/28					
1 SNP identified in exon 1								
Gene	**Oestrogen receptor β**							
EXON	1	2	3	4	5	6	7	8
Length	353	173	117	300	139	134	181	426
Addition	251/20	105/25	160/10	137/18	168/17	130/15	63/18	338/12
2 SNPs identified in exon 4 and 7								
Gene	**Ghrelin**							
EXON	1	2	3	4				
Length	111	114	109	154				
Addition	260/17	186/18	37/315	250/12				
No SNP identified in exons								

Addition N/N: indicate to the 5´ and 3´ additional sequence that are added outside the sequence of exon in order to find all SNPs that might be located on the whole exon.

- A full list of the genes studied is given in Table 3.4, and Table 3.5 gives the primer sequences used in genotyping by primer extension
- The names of novel identified SNPs are given in bold while the names of previously known SNPs are given in normal font.
- The names given to the different SNPs were laboratory short hand.

.

Table 3.4 List of the names, locations, chromosome and nucleotide positions of all the SNPs studied in this thesis.

SNPs (KBiosciences' identifier)	Name of gene	Chromosome	Location of SNP	Position of SNP	Amino acid change (where applicable)
ACT_IIB_45	*Activin receptor 2B*	22	Intron 3	C>T	
ACT_IIB_46	*Activin receptor 2B*	22	Intron 3	T>G	
ACT_IIB_86_END	*Activin receptor 2B*	22	Intron 9	G>A	
ACT_IIB_95	*Activin receptor 2B*	22	Intron 1	G>A	
ACT_IIB_503	*Activin receptor 2B*	22	Intron 4	C>T	
FSHR_L502L	*FSH receptor*	11	Exon 10	C>T	
FSHR_N669N	*FSH receptor*	11	Exon 10	C>T	
FSHR_S596S	*FSH receptor*	11	Exon 10	C>T	
FSHR_T658S	*FSH receptor*	11	Exon 10	C>G	Thr > Ser
FSHR_T685T	*FSH receptor*	11	Exon 10	C>A	
bGNRHR_prom_SNP_-331	*GnRH receptor*	6	Promoter	-331 A>G	
bGNRHR_prom_SNP_-108	*GnRH receptor*	6	Promoter	- 108 T>C[3]	
bGNRHR_ex1_SNP_206	*GnRH receptor*	6	Exon 1	206 G>A[4]	

bGNRHE_ex1_SNP_260	*GnRH receptor*	6	Exon 1	260 C>T[3]	
bGNRHR_ex1_SNP_341	*GnRH receptor*	6	Exon 1	341 C>T[3]	
bGNRHR_ex1_SNP_383	*GnRH receptor*	6	Exon 1	383 C>T[4]	
bGNRHR_ex1_SNP_410	*GnRH receptor*	6	Exon 1	410 C>T[3]	
bLHB SNP1588	*LH beta*	18	Exon 1	1588 - >C	
LHR_W467C	*LH receptor*	11	Exon 11	1401 G>T	Tryp > Cyst
LHR_L490L	*LH receptor*	11	Exon 11	1470 C>T	
LHR_Q527H	*LH receptor*	11	Exon 11	1581 G>T	Gln > His
npy_ex1	*Neuropeptide Y*	4	Exon 1	T>C	
NPYRY2	*Neuropeptide Y receptor Y2*	17	Exon 1	72 G>A	
bERA_prom_SNP173	*Oestrogen receptor alpha*	9	Exon C promoter	173 G>A	
ESR1 ex1 A503C	*Oestrogen receptor alpha*	9	Exon 1	C>A	
bERA_ex8_SNP1820	*Oestrogen receptor alpha*	9	Exon 8	1820 T>C	
bERB_ex4_SNP421	*Oestrogen receptor beta*	10	Exon 2	421 C>A	Leu > Ile

bERB_ex7_SNP168	*Oestrogen receptor beta*	10	Exon 2	168 C>G	
Leptin_promoter -963	*Leptin*	No blast	Promoter	-963 T>C	
Leptin_promoter_1_-1457	*Leptin*	No blast	Promoter	-1457 G>A	
STAT1 c3141t	*Signal transducer and activator of transcription 1*	2		3141 T>C	
STAT5A g12195c	*Signal transducer and activator of transcription5A*	19	Exon 8	12195 G>C	
GHR Phe279Tyr	*Growth hormone receptor*	20	No blast	T>A	Phe > Tyr
GHRA257G ex10	*Growth hormone receptor*	No blast	Exon 10	A>G	
PRL 89398 g/a R	*Prolactin*	23	No blast	89398 G>A	
PRLR Ser18Asn	*Prolactin receptor*	20	No blast	G>A	Ser > Asn

- BLAST: is the Basic Local Alignment Search Tool which is an online tool that is used to find regions of local similarity between sequences.
- No blast: the sequence wasn't available online.

Table 3.5 List of the SNP IDs, allele substitutions, base pair positions, and the forward and reverse primer pairs used for genotyping by primer extension. Note K and Y indicated unspecified purine and pyrimidine residues respectively.

SNP ID	AlleleY	AlleleX	Sequence
ACT_IIB_45	C	T	GGGTAGATCCTATTA[C/T]KGCCAGATCCTGTTA
ACT_IIB_46	T	G	GGTAGATCCTATTAY[G/T]GCCAGATCCTGTTAA
ACT_IIB_503	C	T	GGGGTTCAGCCGCGC[C/T]TCCCTGCTTCAGGAC
ACT_IIB_86_END	G	A	CAGGCCTGTGGGCTC[A/G]GTCCTCAGGAACATC
ACT_IIB_95	G	A	CCCAGTGATGCTGGA[A/G]GGGTTGTCCCCGCCT
bERA_prom_SNP173	G	A	GGCTGGGGCCAGCAA[A/G]GCATCTGATCCAAGT
bERB_ex4	A	C	CAGCTGGTGCTTACG[A/C]TCCTGGAGGCCGAGC
bERB_ex7	G	C	GGTCTGGGTGATTGC[C/G]AAGAGTGGCATGTCC
bGNRHE_ex1_SNP_260	C	T	TTTGACTTTAGCCAA[C/T]CTGCTGGAGACTCTG
bGNRHR_ex1_SNP_206	G	A	TCAAAGGAAAGAGAA[A/G]AGGAAAAAACTCTCG
bGNRHR_ex1_SNP_341	C	T	TGGAGAGCTCCTTTG[C/T]AAAGTCCTCAGCTAT
bGNRHR_ex1_SNP_410	C	T	GGTGGTGATCAGCCT[C/T]GACCGCTCGCTGGCG

bGNRHR_prom_SNP_-108	C	T	AATATAAACCTGTGA[C/T]GTTATCAGCCAAAGA
bGNRHR_prom_SNP_-331	G	A	AAGAGAATTTTAATT[A/G]CAAAATTAAGTTTCA
bLHB SNP1588	-	C	CCATGAGCTGCGCTT[/C]GCCTCCGTTCGGCTCC
ESR1 ex1 A503C	C	A	CCCGCCTCCGCAGCC[A/C]CTCTCGCCCTTCCTG
ESR1	C	T	CGCCCCAGCCAACTT[C/T]GGGAGCGCACCTCCA
FSHR_L502L	C	T	TTTTGCAGTTGCCCT[C/T]TTTCCCATCTTTGGC
FSHR_N669N	C	T	CTTTCATCCAAGGAA[C/T]GGCCACTGCCCCCCA
FSHR_S596S	C	T	TGCCATCTCTGCCTC[C/T]CTCAAGGTGCCCCTC
FSHR_T658S	C	G	CCTATAGGTCAGAAA[C/G]CTCATCCACTGCCCA
FSHR_T685T	C	A	TGGTTCCAATTACAC[A/C]CTTATCCCCCTAAGA
GHR Phe279Tyr	T	A	TAGCAGTGACATTAT[A/T]TTTACTCATATTTTC
GHRA257G ex10	A	G	TCACACGTAGAGCCA[A/G]GCTTTAACCAGGAAG
leptin_promoter -963	T	C	TATCCTTCCTTTCTT[C/T]AATAGATAATTATTA
leptin_promoter_1 -1457	G	A	GCATGAGAACTCTTA[A/G]CTGCAGCATGTGGGA

LHR_L490L	C	T	CATTCCAGTCATGCT[C/T]GGAGGATGGCTCTTT
LHR_Q527H	T	G	AAGCACTCTCTCACA[G/T]GTCTACATCTTAACC
LHR_W467C	T	G	CACACTAGAAAGATG[G/T]CACACCATCACCTAT
NPYRY2	G	A	GGCTTTCCTCTCAGC[A/G]TTTCGCTGTGAGCAG
npy_ex1	T	C	CGTGTGCCTGGGCGC[C/T]CTGGCCGAGGCGTAC
PRL 89398 g/a R	G	A	CCTAGTCACCGAGGT[A/G]CGGGGTATGAAAGGA
PRLR Ser18Asn	G	A	TGCTACTTTTTCTCA[A/G]TGTCAGCCTTCTGAA
STAT1 c3141t	T	C	TAAACTTTACAAATT[C/T]ATGAGTAGTATCTTC
STAT5A g12195c	G	C	CTCAGCCCTGGTGAC[C/G]AGGTGACTCCTGGCC

Chapter 4: Single nucleotide polymorphisms in the bovine gonadotrophin releasing hormone receptor gene and their associations with fertility

4.1 Introduction

As discussed in Chapter 1, subfertility is an increasingly important problem in dairy cattle. As a result of an unfavourable genetic correlation between fertility traits and milk yield (Wall et al., 2003, Royal et al., 2002a; 2002b), and because dairy cattle have principally been selected for yield during the past 50 years, the gains in breeding values for yield have been accompanied by unfavourable changes in breeding values for fertility traits such as calving interval, days in milk to first service, non-return rate at 56 days after first service and number of inseminations required per conception (Wall et al., 2003). The drawbacks of using selection indices for genetic improvement have resulted in the search for pre-pubertal indicators of fertility which will help in estimating a cow's potential fertility, or the fertility of a sire's daughters early in life (Flint et al., 2004). Genotyping represents one such approach.

On the basis of the major impact that the GnRH receptor (GnRHR) has on reproduction (discussed in section 1.4.2.1) the *GnRHR* was chosen as a candidate gene in terms of identification of DNA markers for fertility. The potential importance of the *GnRHR* is emphasized by the identification of deleterious polymorphisms in the human *GnRHR* gene, which are associated with clinically significant reproductive pathologies in men and women (Huhtaniemi, 2002). Moreover, single nucleotide polymorphisms (SNPs) associated with fertility have already been identified elsewhere in the bovine hypothalamo-pituitary-gonadal system (Hastings et al., 2006).

The aim of the study presented in this Chapter was to determine whether SNPs identified in the *GnRHR* gene are associated with subfertility in dairy cattle, by using the predicted transmitting abilities (PTAs) for fertility traits available through the UK fertility index. Data were also available from a study in which *GnRHR* gene function

was assessed by measuring circulating luteinising hormone (LH) concentrations after administration of GnRH in pubertal animals. Associations between SNPs and LH responses were analysed at the level of both genotype and haplotype, the latter having the advantage of accounting for linkage disequilibrium between alleles (see sections 1.7.1, 1.7.2). The results of this analysis have been published as Derecka et al. (2009) *Animal Genetics* **41**, 329-331.

4.2 Materials and Methods

4.2.1 Analysis of data

Genotype associations with PTAs for fertility traits were tested by multivariate linear mixed model in Genstat, in 408 animals which was fitted as:

$$y_{ijkl} = u_i + A_{ik} + B_{ij} + C_{ijkl}$$

y_{ijkl} = PTAs for trait *i* for son *l* of sire *j* inheriting genotype k, with effects as follow:

u_i = Overall mean

A_{ik} = Fixed effect of the genotype *k* for trait *i*

B_{ij} = Random effect of sire *j* for trait *i*

C_{ijkl} = Random error term

PTAs for CI, DFS, NR56, CINS, and PIN were calculated in 2004; more recent values for CI, NR56, and 305 MY only were derived again in 2009 (courtesy of Dr R. Mrode, SAC, Edinburgh). As PTAs were not selected for reliability, they were de-regressed by multiplication by 1/(1-re) to account for variation in reliabilities of estimates.

Associations of genotypes with LH levels in animals treated with GnRH were analysed in 431 animals on log transformed LH concentrations for each of the times after treatment (pre-treatment basal level and levels at 15, 30, 60, 90, 120 and 150 min), the difference between 15 and 30 min (30 – 15), for the maximum LH level

reached at any time after GnRH, for the sum of LH concentrations after treatment, the area under the response curve (subtracting basal levels) and the time at which the maximum LH level was reached. Single polymorphism genotype associations were carried out using REML (linear mixed models) in Genstat (release 8.1) with age, sex, date of treatment, and genotype as fixed effects and sire as a random effect. Hence, the statistical model used to investigate the LH response to GnRH challenge was:

$$y_{ijkl} = u_i + A_i + B_i + C_i + D_{ik} + E_{ij} + F_{ikjl}$$

y_{ijkl} = LH response variable *i* for calve *l* of sire *j* inheriting genotype *k*, with effects as follow:

u_i = Overall mean

A_i = Fixed effect of age

B_i = Fixed effect of sex

C_i = Fixed effect of date of treatment

D_{ik} = Fixed effect of the genotype *k* for trait *i*

E_{ij} = Random effect of sire *j* for trait *i*

F_{ijkl} = Random error term

Frequencies of haplotypes and their associations were estimated using a maximum likelihood model based on the stochastic-EM algorithm (Tregouet et al., 2004a) applied in the program THESIAS (http://www.genecanvas.org; Tregouet et al., 2002; 2004b; 2007). This program tests for deviation from Hardy-Weinberg equilibrium using the X^2 goodness-of-fit test. Haplotypes were deduced from observed data, no assumptions being made based on partial genotypes and pedigrees. The program allowed for missing data. Haplotype associations with LH secretion parameters were tested with sex, age and date of treatment as covariates. For testing haplotype associations with PTAs, the production index PIN was used as a covariate.

4.2.2 Bioinformatics

Nucleosome residence times were estimated using the Recon Program (DNA nucleosomal organisation: http://wwwmgs. bionet.nsc.ru/mgs/programs/recon/;

Levitsky et al., 2001; Levitsky 2004). Transcription factor binding sites were evaluated using Matrix Search for Transcription Factor Binding Sites, public version 1.0 (BIOBASE Gmbh, Wolfenbüttel, Germany).

4.3 Results

4.3.1 Identified polymorphisms

Seven single nucleotide polymorphisms (SNPs) were identified in the *GnRH receptor* gene (Figure 4.1). These were in the promoter and coding regions at positions -331, -108, 206, 260, 341, 383 and 410 relative to the translation start site. All these polymorphisms were silent. The polymorphisms 206G>A and 383C>T were inherited together, as were those at -108T>C, 260C>T, 341C>T and 410C>T (Table 4.1). Of the $2^3 = 8$ potential haplotypes, 6 were present in the animals studied, and of these, 4 haplotypes accounted for 96.7% of those observed (Table 4.2). There was a departure from Hardy-Weinberg equilibrium at locus 2 (-108T>C; $p < 0.001$); the other loci were in equilibrium. Unless specified individually, these groups of polymorphisms are henceforward referred to here by the first SNP listed (i.e. -331A>G, -108T>C and 206G>A) and are identified by the nucleotides at each position (i.e. A-T-G).

Table 4.1 Genotypes observed at the three group of SNPs in the wild homozygous, heterozygous, and mutant homozygous carriers.

	Position	Wild type (0)	Heterozygous mutation (1)	Homozygous mutation (2)
Group 1 (locus 1):	-331	AA	AG	GG
Group 2 (locus 2):	-108	TT	CT	CC
	260	CC	CT	TT
	341	CC	CT	TT
	410	CC	CT	TT
Group 3 (locus 3):	206	GG	AG	AA
	383	CC	CT	TT

- Genotypes are given as nucleotides observed at each of the positions indicated (base pairs relative to the translation start site). Within groups 2 and 3 the polymorphisms given in each column (0, 1 and 2) were inherited together. Those in column 0 occurred most frequently and these were considered the 'wild type' alleles.

Table 4.2 Haplotype frequencies estimates in the whole population studied (group for PTAs analysis + group for GnRH treatment).

Haplotypes	Freq (%)	s.e.	t
A – T – G	60.0	-	-
g – t – g	1.5	0.3	5.18
A – C – G	14.1	0.8	18.11
G – C – G	14.7	0.9	15.92
a – t – a	1.8	0.5	3.82
A – C – A	7.9	0.7	10.73
Locus	**Percent**	**pHWE**	
1 (-331A>G)	83.7/16.3	0.440	
2 (-108T>C)	63.3/36.7	<0.001	
3 (206G>A)	89.0/11.0	0.495	

- Haplotype frequencies for the 963 animals studied. Haplotypes are given in the form of the nucleotides at positions -331 (A>G), -108 (T>C) and 206 (G>A) respectively. Principal haplotypes are given in upper case bold, minor haplotypes in lower case. There were 507 animals with no missing data, 318 with one missing SNP, 138 with 2 and 0 with 3, s.e., standard error; t, Student's t test.
- For each of the 3 co-segregating loci the percentage of the 2 alleles is given, together with the probability of departure from Hardy-Weinberg equilibrium (PHWE). Data derived using THESIAS (Tregouet and Varelle, 2007), assuming linkage disequilibrium.

```
 1  aacaaatttt ttgagaaatt acaaatttgt gattctttaa attacaaatc -388

51  ttaactggat caagtttttt tttagtattt tcttcttttt cattttgcat -338

101  gtattgaaca gaacacttaa ggtttttttt ttagaaaaca tcttcactaa -288
                                        -331g
151  atttacttga ataatatcct aagagaattt taattacaaa attaagtttc -238

201  aaacgtcctt cttcacttag gaaaaatgtt gtgaaaacca ggccatctgc -188

251  tgagatacta cagttacatt tggccctcag aaagtgtttg tctgctttgc -138
                                          (+)GSE
301  tttagcaccc tgctggatta caagcccaca aaacaagttc accttgatct  -88
                                  (-)GSE
351  ttcactttaa atacaaaata tctcagggac aaaatttgac ataaatataa  -38
        -108t                                  TSS
401  acctgtgacg ttatcagcca aagaaggctg aaataaaaca ggactccaag  +13

451  tgcaattaca ataaaatatc agaagtgcca gaaacacgag tcttgaagct  +63
                        ATG
501  gcatcagcca taaaggATGg caaacagtga ctctcctgaa cagaatgaaa +113

551  accactgttc agcgatcaac agcagcatcc ctctaacacc aggcagcctc +163

601  cccaccctga ccctatctgg aaagatccga gtgacagtta ctttcttcct +213

651  ttttctactc tccacaattt tcaacacttc tttcttgttg aaacttcaga +263
                           206a
701  attggactca aaggaaagag aagaggaaaa aactctcgag aatgaagttg +313
                                260t
751  cttttaaaac atttgacttt agccaacctg ctggagactc tgattgttat +363

801  gccactggat ggaatgtgga acataactgt tcaatggtat gctggagagc +413
        341t                                              383t
851  tcctttgcaa agtcctcagc tatctgaagc ttttctccat gtacgccccc +463
                                  410t
901  gccttcatga tggtggtgat cagcctcgac cgctcgctgg cgatcaccaa +513

951  gcctctagca gtgaaaagca acagcaagct tggacagttc atgattggct +563

1001  tggcctggct cctcagtagc atctttgctg gaccaca             +600
```

Figure 4.1 DNA sequence of the bovine *GnRH receptor* gene assembled from GenBank® accession nos. NM_177514.2 and AF034950 (Ensembl Cow Exon View: GnRHR_Bovin) including approximately 430 bp upstream (5') and 590 bp downstream (3') of the transcription start site (TSS, arrowed), showing polymorphisms identified at positions -108 and -331 upstream and 206, 260, 341, 383 and 410 downstream of the translation start site (upper case ATG, arrowed), the positions of the SNPs are shadowed grey. The positions of the GSE (gonadotroph-specific element) sites are also shown in grey shadow. Numbering on left side, as in GenBank®; on right side, relative to TSS.

4.3.2 Postulated effects of polymorphisms on gene function

Although the SNPs identified do not alter the primary sequence of the GnRH receptor protein, they may affect gene function in other ways. Examination of possible changes to transcription factor binding sites (Figure 4.2) showed that the SNP -331A>G destroyed Cart-1 and HLF sites and introduced a Nkx-6.1 site. The Nkx-2.5 site at this position was not affected. The -108T>C substitution destroyed CREB, CREBP/cJUN and Lmo2 complex motifs. The GATA-1 motif at -108 was unaffected. The SNPs downstream of the translation start site at 206G>A, 260C>T, 341C>T and 410C>T potentially generated new HNF-4, CCAAT and NFY sites, and resulted in the loss of 2 Oct-1 motifs.

The locations of the SNPs were then assessed for their potential to be wrapped in nucleosomes much of the time (Fig.4.3). The results showed that 206G>A and 260C>T were at a regions where the DNA was relatively accessible. In contrast -331A>G, -108T>C, 341C>T, 383C>T and 410C>T were in regions where nucleosomes were likely to remain for longer periods. This suggests that the SNPs at positions 206 and 260 were more likely to affect gene function than those elsewhere. As these positions represented two of the co-segregating SNP groups (Table 4.2), it was concluded that six of the seven polymorphisms observed potentially affect gene function.

SNPs in GNRHR exon 1

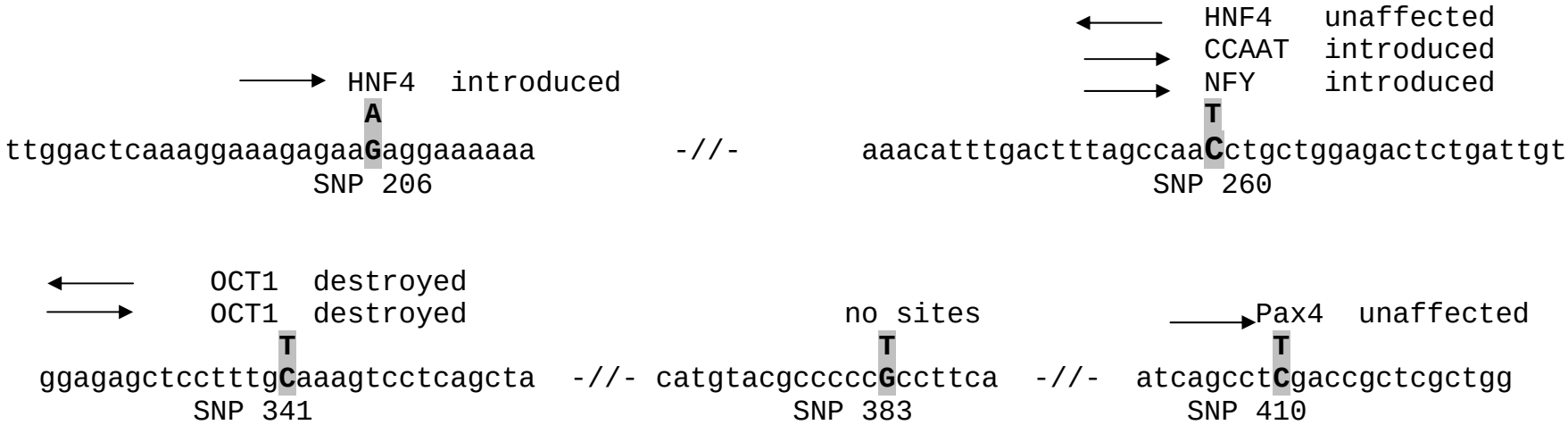


Figure 4.2 Effects of single nucleotide polymorphisms (SNPs) in the *GnRH receptor* gene on regulatory sequences. The SNPs are shown in the context of neighbouring nucleotides, arrows indicate direction and positioning of transcription factor sequence. The codon context of SNPs are shadowed grey.

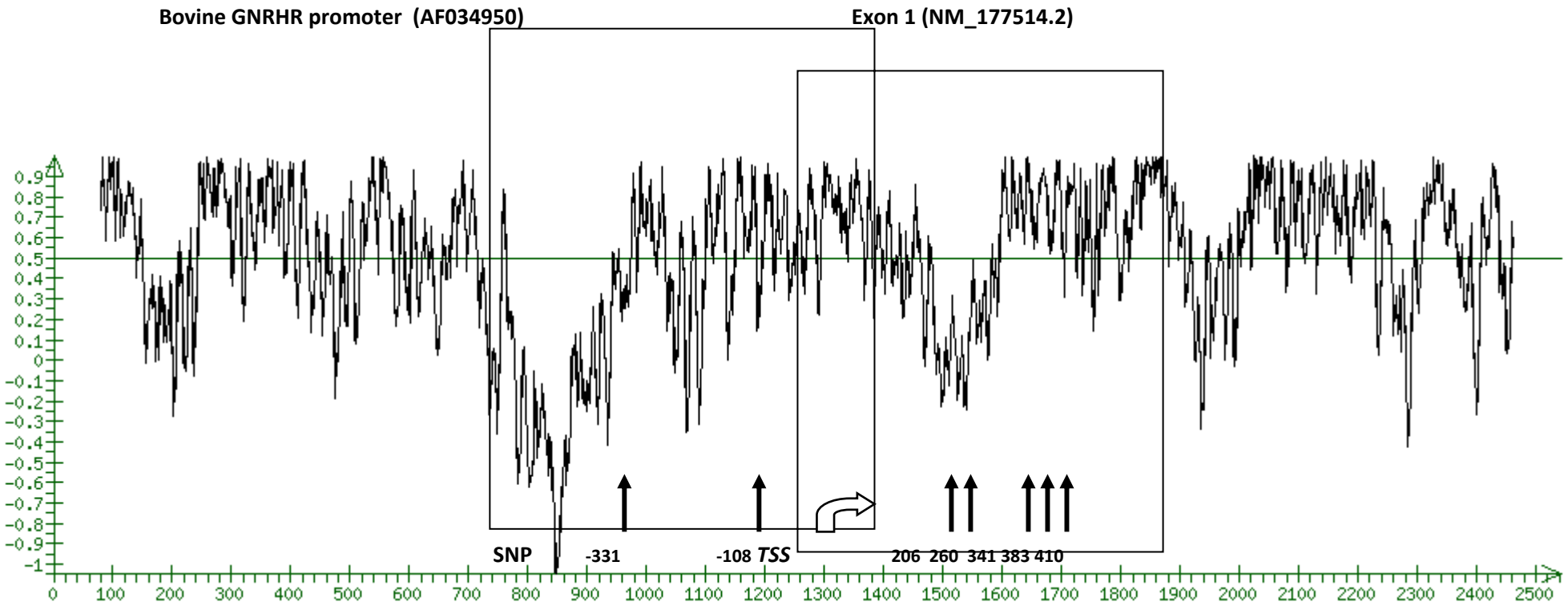


Figure 4.3 Nucleosomal residence-time plot of DNA coding promoter (GenBank® accession no. AF034950) and GnRH receptor cDNA (GenBank® accession no. NM_177514.2). Boxes mark positions of DNA fragments analyzed in the present study. SNP numbers are given on the basis of their position on each sequence individually. Open arrows indicate start of the *GnRH receptor* coding sequence (ATG). Black arrows indicate the positions of SNPs. The vertical axis indicates the preference of histone cores for this location (arbitrary units). The higher the peak the stronger the preference, with nucleosomes tending to reside where there are peaks (approaching 100 bases) mostly above the horizontal line. Likewise troughs indicate regions that nucleosomes make the DNA available to bind other proteins, such as transcription factors.

4.3.3 Effects of polymorphisms on responses to GnRH administration

To determine whether GnRH receptor polymorphisms affected pituitary sensitivity to GnRH, plasma LH was measured before and after GnRH treatment (Royal, 1999). As observed previously (Barnes et al., 1980; McLeod et al., 1984) circulating LH levels rose within 15 min following treatment with GnRH to peak between 15 and 150 min (Royal, 1999). Values for LH levels, and the time at which the peak plasma concentration of LH was observed, were determined for each animal and these and other parameters (see Methods section) were tested for association with genotypes.

Haplotype analysis of the 431 animals treated with GnRH showed that 6 haplotypes accounted for the genotypes identified, varying in frequency between 2 and 69% in this group of calves. Three of these haplotypes (G – C – A, A – C – A and A – T – A), which were present at low frequency (< 5%) and/or were not present in the other groups, were excluded from association analysis. After accounting for sex, age and date of test, there was a tendency for haplotype A – C – G to be associated with a small reduction in $\log_{10}$ maximum LH level ($p = 0.053$; Table 4.3), and with a reduction in $\log_{10}$ time to maximum level ($p = 0.050$) but there were no associations with other LH concentration parameters (e.g. $\log_{10}$ area under the curve; other data not shown). There was no effect of the haplotype G – C – G. There was a significant effect of the sex of the calf on all parameters except the basal LH concentration. Peak LH levels were raised in bull calves by approximately 10 ng/ml ($p<0.001$), and occurred 25 minutes earlier ($p<0.025$) than in heifers.

Table 4.3 Effects of haplotypes on luteinising hormone secretion following GnRH administration.

Haplotype:	**A – C – G**	**G – C – G**
Frequency (%):	14.1	7.8
Log_{10} Max LH level (ng/ml)		
Effect	-0.048	-0.030
s.e.	0.025	0.058
t	-1.933	-0.51
p	0.053	0.607
Log_{10} Time to peak LH (min)		
Effect	-0.050	0.014
s.e.	0.026	0.089
t	-1.957	0.162
p	0.050	0.871

- Effects are relative to the most frequent haplotype (A – T – G; frequency, 68.9% in this population). Values for 431 individuals (191 bull and 240 heifer calves). Frequency of haplotype A – C – A (1.9 ± 0.9 percent) was too low to be included in this analysis, s.e., standard error; t, Student's t test; p, probability.

4.3.4 Association of genotypes with fertility traits

Haplotype associations were used initially to investigate relationships between genotype and fertility PTAs this will allow identifying the minimal segregating set of haplotypes in this population. The haplotype A – C – A, which was present at 11% frequency in this group of animals, was included in this analysis. The haplotypes G – C – G, A – C – A and A – C – G reduced PTA for days in milk to first service by between 0.48 and 0.74 days relative to A – T – G ($p = 0.013$, 0.046 and 0.009 respectively) when PIN was taken into account. The inclusion of PIN in the analysis will exclude the possibility that the significant effects associated with these haplotypes were actually on PIN rather than on fertility PTAs. There was a tendency for haplotype G – C – G to be associated with a lower PTA for calving interval ($p < 0.1$). There were no associations of haplotype with PTAs for non-return rate or number of inseminations, but there were reductions in PIN in the case of G – C – G and A – C – A (Table 4.4).

Table 4.4 Associations of haplotypes with fertility PTAs and PIN.

Haplotype:	**G – C – G**	**A – C – A**	**A – C – G**
Frequency (%):	16.7	11.4	6.9
CI			
Effect	-0.520	-0.480	-0.560
s.e.	0.317	0.448	0.450
t	-1.65	-1.08	-1.247
p	0.098	0.280	0.212
LL	**-1520.28 X^2 = 4.76 p > 0.05**		
DFS			
Effect	-0.542	-0.480	-0.740
s.e.	0.220	0.241	0.285
t	-2.476	-1.995	-2.600
p	0.013	0.046	0.009
LL	**-1360.88 X^2 = 13.96 p < 0.005**		
NR 56			
Effect	-0.181	0.023	-0.055
s.e.	0.198	0.217	0.304
t	-0.906	-0.105	-0.182
p	0.364	0.916	0.856
LL	**-1339.71 X^2 = 1.00 p > 0.05**		
CINS			
Effect	0.005	0.0003	-0.0046
s.e.	0.005	0.0053	0.0061
t	1.031	0.0527	-0.752
p	0.302	0.958	0.451
LL	**10.31 X^2 = 3.00 p > 0.05**		
PIN			
Effect	-4.216	-7.98	3.96
s.e.	1.993	2.63	3.52
t	-2.115	-3.03	1.12
p	0.0344	0.002	0.261
LL	**-2201.20 X^2 = 14.86 p < 0.005**		

- Effects are relative to the most frequent haplotype (A – T – G; frequency, 63.5% in this group of animals). Haplotypes are indicated in the form -331 – -108 – 206. Values for 408 sires (318 with no missing genotype data; 38 with 1, 52 with 2, and 0 with 3 missing SNPs). For analysis of the fertility traits, PIN was used as a covariate. CI, calving interval PTA (days); DFS, days to first service PTA (days); NR56, PTA for non-return rate at 56 days after insemination (percent); CINS, PTA for number of inseminations required to establish pregnancy (number); PIN, production index (£). LL, loglikelihood; X^2, chi-squared (3 degrees of freedom in each case), s.e., standard error; t, Student's t test; p, probability.

An association study was subsequently carried out between the SNPs and fertility PTAs. Fertility PTAs were available for 408 sires, and this group was in Hardy-Weinberg equilibrium at all 3 loci. In the case of the -108T>C mutation, the

relationship with PTA for days in milk to first service (-0.43 days; p = 0.002) remained after taking the effects of sire and PIN into account (Table 4.5). Although there was a significant effect of PIN, Table 4.5 suggests this effect was not mediated solely through PIN. There was no effect of percent Holstein (data not shown). There were no significant effects on the other fertility traits analysed, and no significant associations with SNPs at positions -331 or 206/383.

Table 4.5 Effects of allelic substitution (B) at -108/260/341/410 on fertility PTAs.

PTAs	Beneficial effects	Effects of SNPs (means ± s.e.) at		
		-331A>G	**-108T>C**	**206G>A**
CI (days)	Decrease	-0.13 ± 0.28	-0.43 ± 0.20	-0.295 ± 0.32
DFS(days)	Decrease	-0.28±0.18	-0.43 ± 0.13	-0.415 ± 0.21
NR56(percent)	Increase	-0.05±0.16	0.056 ± 0.12	0.072 ± 0.182
CINS (number)	Decrease	0.004±0.0035	0.00 ± 0.003	0.001 ± 0.004
PIN (£)	Increase	-1.80±1.903	-2.55 ± 1.37	-5.88 ± 2.13
n		360	406	349
P value		>0.1	0.002	0.026

- Analysis of genotypes of 408 bulls using REML (multivariate linear mixed model) with sires accounted for as a random effect. Statistical significance for each SNP was estimated by Wald test (Chi probabilities) with 5 degrees of freedom. PTAs derived from the UK Fertility Index were: CI, calving interval; DFS, days to first service after calving; NR56, non-return rate at 56 days after calving; CINS, number of inseminations required to establish pregnancy. PIN, Production index. n=numbers of bulls with genotypes available at each locus; P, probability.

4.4 Discussion

The data show that although they do not alter the amino acid sequence of the GnRH receptor, the SNPs at positions -108/260/341/410 do affect fertility, as reflected by an association with sire fertility PTAs. This is consistent with the effects of polymorphisms in the human *GnRH receptor* gene, which are associated with hypogonadotrophic hypogonadism and a range of consequent conditions affecting

fertility (Huhtaniemi, 2002; Lin et al., 2006). The substitutions reported here in the promoter region are not at the same locations as those associated with human pathology, but mutations in the *GnRH* receptor gene promoter have been shown to have functional consequences (Cheng et al., 2002). The number of SNPs identified in the region upstream of the GnRH receptor gene (2 in 800 bp) is consistent with the reported frequency of polymorphisms in non-repetitive, non-coding regions (Nickerson et al., 1998).

Three haplotypes with -108T>C (A – C – G, A – C – A and G – C – G) were associated with reductions in PTA for days to first service, relative to the haplotype A – T – G. Multivariate linear mixed model analysis, which accounted for the effects of sire, confirmed the -108T>C polymorphisms to be responsible for this effect (those at -331A>G and 206G>A being ineffective). However it cannot be concluded that the effect on PTA for days to first service was a consequence of the effect on LH secretion, as the haplotype A – C – A was not associated with any change in LH secretion characteristics.

The various possible mechanisms by which polymorphisms in nucleotide sequence could affect gene function have been covered in section 1.6.2. For the purpose of this study, we have examined the SNPs described here for the possible introduction or deletion of transcription factor binding sites (Figure 4.2). Effects on gene transcription resulting from substitutions affecting transcription factor binding have been described previously (Shan et al., 2005; Pottier et al., 2007). However nucleotide substitutions affecting transcription factor binding sites are likely to be significant when they occur in regions of genomic DNA which is accessible proteins, and therefore we have analysed the *GnRH receptor* gene sequence for reduced nucleosome residence.

Furthermore, analysis of nucleosome residence times provides useful information on the potential accessibility of genomic DNA to transcription factors, and has been used previously to identify functional transcription factor binding sites. None of the SNPs observed here altered the plots to any great extent (data not shown), but this is expected as nucleosomal residence is based on the distribution of sequences in a relatively large (~100-base) window. The severe trough at around position 850 is

probably due to the oligo-AT tracts between bases 7 – 132 (relative to the transcription start site) of the sequence in Figure 4.1. The narrow peaks at sequence positions 1060-1300 in Fig. 4.3 (equivalent to -170 - +70bp relative to the transcription start site) suggest sites where nucleosomes can bind but not well, which is unsurprising in non-constitutive, highly-regulated genes. As the substitutions associated with the -108T>C polymorphisms have been identified as those responsible for the effects on fertility traits observed here, nucleosomal residence times at positions -108, 260, 341 and 410 all are potentially significant. Of these only the SNP 260C>T coincides with a decrease in nucleosomal residence, which implicates substitutions at this site.

As shown in Fig. 4.2, the 260C>T substitution introduces binding sites for the transcription factors CCAAT and NFY. Of these CCAAT is particularly significant, because this site has been implicated in control of the human GnRH receptor promoter in murine gonadotroph-derived αT3-1 cells, such that mutations at this site reduce promoter function (Cheng et al., 2002). As in the human gene, this site is in exon 1 of the bovine sequence, although the bovine 5' untranslated region is short compared with the human (78 versus 769bp).

The SNPs reported here also introduce a large number of other changes to transcription factor binding sites, and although they are located in regions of limited nucleosome slippage, these will also be potentially important regulatory sites if they are accessible. Comparison of bovine and ovine GnRH receptor 5' fragments suggests that about 660 bp upstream of the translation start site is highly conserved. This region is important in the regulation of *GnRH* receptor gene expression in both the human and sheep, and due to the conservation of the 5' UTR probably in the cow as well. Within it is located a gonadotroph-specific element (GSE), which is crucial for tissue specificity and the fine tuning of gene expression by steroidogenic factor-1 (SF-1) in gonadotrophs (Ngan at al., 1999). The two SNPs in the promoter, at -331A>G and -108T>C, were both located within this region, and both affected potential transcription factor binding sites. The substitution -108T>C destroyed binding sites for CREBP1/c-Jun, Lmo2 complex and CREB. The GATA-1 site was unaffected. The importance of Lmo2 lies in its ability to form a bridge in a higher order complex with GATA-1 (present on the minus strand) and some other factors, to

influence transcription. Although it was not associated with effects on either LH secretion or fertility PTAs, the substitution -331A>G had significant effects on transcription factor binding sites. It left unaffected an Nkx-2.5 site (which is important for myocardial development; Lints et al., 1993; Chen & Schwartz, 1995) and introduced a site for Nkx-6.1, a factor possibly important in control of islet development and/or regulation of insulin biosynthesis (Rudnick at al. 1994). The -331 SNP also eliminated a cartilage homeoprotein-1 (Cart-1) site and a hepatic leukaemia factor (HLF) site. HLF is a general transcription factor. Cart-1, which is a transcriptional modulator with both repressive and activatory activities, is a transcriptional repressor in non-pituitary cells and has been suggested to play a role in development of the reproductive tract (Gordon at al., 1996). The primary site of *GnRH receptor* expression is the pituitary gonadotrophs, but *GnRH* may also be involved in regulation of human ovary function (Cheng et al., 2002) and in the placenta (Lin at al., 1995; Wolfahnt et al., 1998) and it is possible that this site is used for tissue specific transcription regulation.

Systemic administration of GnRH has been widely used to test the sensitivity of pituitary gonadotrophs to GnRH. An increase in response, for example before puberty (Lacroix & Pelletier, 1979) and at the onset of the breeding season (Curlewis et al., 1991), corresponds to an increase in fertility, and the magnitude of the response is heritable (Royal, 1999; Haley et al., 1989). Therefore there is a good deal of information suggesting that factors controlling the response to GnRH may control reproductive processes. In the present work pre-pubertal calves were used to measure responses to GnRH, in order to reduce variation due to maturation of the LH secretory response with approaching puberty (Nakada et al., 2002), and used age as a covariate in the analysis of response to GnRH. Although there was a tendency towards a minor reduction in area under the curve with one of the haplotypes, there were no major effects on the magnitude of the LH response to GnRH.

The effects of the -108/260/341/410 SNPs on estimated breeding values for fertility traits were large relative to population variances in these PTAs. In the UK fertility index 95% of calving interval PTAs are within a 10-day range (-5 to +5 days either side of the mean; Wall et al., 2003). For days in milk to first service this range is 7 days. The effects observed here may account for up to 10 percent of these values

(Table 4.5) and are therefore potentially quantitatively important. Regression analysis suggested the effects of these SNPs were not additive (data not shown), possibly reflecting the relatively small number of homozygous C substitutions at -108 (n = 50, compared with 171 and 186 of the wild type and heterozygotes).

Haplotype frequencies were similar between the 3 groups of animals genotyped (Table 3.1). An exception was the appearance of the rare haplotype G – C – A in the GnRH-treated group, and the relatively low frequencies in this group of the haplotypes G – C – G and A – C – A, and the high frequency of A-C-G. These animals were studied at the premises of a major UK breeder, and this difference in genotype may have resulted from unintentional selection for these alleles. This was consistent with the departure from Hardy-Weinberg equilibrium of the -331A>G and -108T>C groups of SNPs in the GnRH-treated animals.

There was a major effect of sex of calf on the response to GnRH, which was unexpected (Schams et al., 1981). This effect, and potential effects of sire and age were taken into account as covariates in both genotype and haplotype association analysis. Furthermore, the -108T>C substitutions were associated with a reduction in PIN (Table 4.5), and it is possible therefore that this polymorphism has been selected against while the principal breeding goal in dairy cattle has been to increase yield. This may explain in part the accompanying decline in fertility, as these SNPs are beneficial in terms of fertility (i.e. reduce PTA for days to first service). It would be interesting to examine the frequency of the -108T>C polymorphisms in cattle that have not been subjected to extreme selection for yield.

In conclusion, among the mutations described here, the SNPs at positions -108/260/341/410 were associated with significant effects on fertility in dairy cattle. Because PTAs rather than breeding values are used, the effect of allelic substitution may represent a gain of nearly a day in DFS. Therefore, selection against the A-T-G and A-T-A haplotypes should improve fertility by this margin. However, the absence of an effect on pituitary sensitivity to GnRH in pre-pubertal animals raises the possibility of an effect exerted through an ectopic site of receptor expression (Wu et al., 2009).

Chapter 5: Investigation of the effects of other candidate genes on fertility in dairy cattle

5.1 Introduction

In Chapter 4, SNPs associations with fertility PTAs were analysed for the GnRH receptor gene. This led to the identification of a significant effect of one of the SNPs at postion -108T>C in the bovine GnRH receptor, on fertility. This approach is extended in the present Chapter, by consideration of other candidate genes. Further candidate genes were selected on the basis of their involvement in the hypothalamic/ovarian/uterine axis (*LH receptor, FSH receptor, oestrogen receptors, activin receptor*) and their known roles in central nervous pathways controlling reproduction (*neuropeptide Y and its receptor*). The choice of genes has been discussed in detail in Chapter 1. As in Chapter 4, an association study was carried out to quantify the possible effects of the identified SNPs within the genes of interest on fertility PTAs.

5.2 Materials and Methods

Details of the names and locations on the genes, and chromosome and nucleotide positions for all the SNPs studied in this chapter and the subsequent chapters are listed in Table 3.4, and are identified elsewhere in the thesis by the nomenclature given in this Table. In addition, information about base pair positions, and the forward and reverse genotyping primers are given in Table 3.5.

For the association studies of the SNPs on the *activin receptor, oestrogen receptor, FSHR, LHR,* and *NPYR* with fertility index PTAs, a total of 427 dairy cows, principally Holsteins, had been genotyped using DNA extracted from semen or blood. This comprised the group of animals genotyped for the SNPs on the *GnRHR* gene plus an extra 20 animals which have been genotyped for the *activin receptor, oestrogen receptor, FSHR, LHR*, and *NPYR* genes. However, there were variations in the number of animals genotyped for each locus. The association between genotypes

and the PTAs for CI, DFS, NR56, CINS, and PIN were assessed using a multivariate linear mixed model in GENSTAT with SNPs and PIN fitted as fixed effects while sires were included as random effects. PTAs for the fertility traits were calculated in 2004. This analysis was also performed using de-regressed values for PTAs to account for variation in the reliabilities of the PTA estimates (PTAs being de-regressed using multiplication by (1/(1 - reliability)). The multivariate linear mixed model was fitted as:

$$y_{ijkl} = u_i + A_{ik} + B_{ijl} + C_{ij} + D_{ijkl}$$

y_{ijkl} = PTAs for trait i for son l of sire j inheriting genotype k, with effects as follow:

u_i = Overall mean

A_{ik} = Fixed effect of the genotype k for trait i

B_{ijl} = Fixed effect of PIN for son l of sire j for trait i

C_{ij} = Random effect of sire j for trait i

D_{ijkl} = Random error term

SNP effects were also confirmed by applying a unvariate linear mixed model.

5.3 Results

SNPs identified in the genes under investigation included: a heterozygous T>C base change at nucleotide position 1820 of the *oestrogen receptor*, four SNPs in the *activin receptor* at nucleotide45, 86, 503, and 95, five SNPs in the *FSH receptor* at residues L502L, S596S, T658S, N669N, and T685T, and one SNP (A>G) in the *NPY receptor*. Furthermore three known SNPs in the *LHR* gene (Hastings et al., 2006) were also studied at positions W467C, L490L and Q527H. In describing these SNPs the most frequent allele is given first.

The initial multivariate linear mixed model analysis of the effects of SNPs at the *activin receptor* showed that the SNPs at nucleotide45 and 86 were associated with beneficial effects on fertility PTAs (reducing CI and CINS, increasing NR56) with P=0.014 and 0.012 respectively; Table 5.1). However, with the de-regressed PTAs

these effects were not significant (Table 5.2), indicating that the significant effect found without de-regression was related to the variation in the reliabilities of the PTA values used. The SNPs at residues 502, 658, 669, and 685 (A>C) in the *FSHR* were associated with significant unfavourable effects (P=0.007, 0.016, 0.003 and <0.001 respectively) on fertility PTAs (increasing CI, DFS, CINS, and decreasing NR56; Table 5.1). These effects became more significant (P≤0.001) with de-regression (Table 5.2). No significant associations were found between SNPs in the *LHR*, *NPYR* or *ERA* and any of the fertility PTAs, either before or after de-regression of the PTA values.

Table 5.1 Results of multivariate analysis between the SNPs in different genes and fertility PTAs: calving interval (CI, days), days to first service (DFS, days), nonreturne rate (NR56, percent), and the number of insemination per conception (CINS, count). Production index (PIN, £)

SNPs	CI	DFS	NR56	CINS	PIN	P value
ACT_IIB_45	-0.2410±0.2020	0.0506±0.1332	0.2803±0.1202	-0.0069±0.0026	0.7176±1.33821	**0.014**
ACT_IIB_86_END	-0.2655±0.2058	0.0160±0.1357	0.2933±0.1226	-0.0068±0.0026	1.0941±1.3635	**0.012**
ACT_IIB_95	-0.1867±0.2124	0.0631±0.1399	0.2031±0.1267	-0.0056±0.0027	0.8342±1.4059	0.149
ACT_IIB_503	-0.1809±0.2264	0.0470±0.1492	0.2020±0.1351	-0.0054±0.0029	2.2327±1.4942	0.124
FSHR_L502L	0.4505±0.2587	0.0864±0.1708	-0.3534±0.1542	0.0084±0.0033	-1.8045±1.7173	**0.007**
FSHR_N669N	0.5734±0.2544	0.1835±0.1682	-0.2878±0.1524	0.0079±0.0033	-2.5429±1.6911	**0.003**
FSHR_S596S	0.0048±0.3761	-0.1501±0.2471	-0.2572±0.2248	0.0051±0.0049	3.3628±2.4800	0.468
FSHR_T658S	0.4778±0.2552	0.1368±0.1686	-0.3006±0.1525	0.0074±0.0033	-1.6630±1.6958	**0.016**
FSHR_T685T	1.470±0.540	1.062±0.659	-0.649±0.331	0.019±0.007	-1.891±1.628	**<0.001**
LHR_L490L	0.3039± 0.3011	0.3029±0.1974	0.2379±0.1805	-0.0034±0.0039	1.7255±1.9904	0.255
LHR_Q527H	-0.0511±0.3672	-0.1286±0.2411	-0.0205±0.2201	0.0022±0.0048	-2.0222±2.4244	0.943
LHR_W467C	-0.1085±0.2304	0.0906±0.1515	0.1729±0.1375	-0.0052±0.0030	0.3506±1.5244	0.385
NPYRY2	0.0405±0.3087	-0.0434±0.2029	-0.0153±0.1848	0.0024±0.0040	-1.1615±2.0403	0.980
bERA_ex8_SNP1820	0.0336±0.3522	-0.0623±0.2317	-0.1228±0.2105	0.0043±0.0045	-3.5187±2.3224	0.609

- Analysis of genotypes of 427 bulls using REML (multivariate linear mixed model) with sires accounted for as a random effect. Statistical significance for each SNP was estimated by Wald test (Chi probabilities); P, probability.

Table 5.2 Results of multivariate analysis between the SNPs in different genes and de-regressed fertility PTAs: calving interval (CI, days), days to first service (DFS, days), nonreturne rate (NR56, percent), and the number of insemination per conception (CINS, count). Production index (PIN, £)

SNPs	dereg-CI	dereg-DFS	dereg-NR56	dereg-CINS	PIN	P value
ACT_IIB_45	8.78±2.86	11.39±5.95	-2.55±0.86	0.11±0.01	32.66±1.71	0.212
ACT_IIB_86_END	0.4111±0.4569	-0.1394±0.5549	-0.6040±0.2771	0.0068±0.0057	-1.1557±1.3614	0.173
ACT_IIB_95	8.77±2.86	11.40±5.95	-2.53±0.87	0.11±0.01	32.68± 1.71	0.517
ACT_IIB_503	0.4089±0.5030	-0.2382±0.6108	-0.4774±0.3061	0.0051±0.0063	-2.2327±1.4942	0.297
FSHR_L502L	1.3226±0.5763	0.7181±0.7029	-0.8813±0.3499	0.0224±0.0072	-1.8045±1.7173	**<0.001**
FSHR_N669N	1.541±0.565	0.991±0.691	-0.721±0.345	0.020±0.007	-2.543±1.691	**<0.001**
FSHR_S596S	-0.5386±0.8175	-0.0156±0.9932	-0.4248±0.4950	0.0091±0.01	3.3628±2.48	0.577
FSHR_T658S	1.3235±0.5655	0.8707±0.6895	-0.7331±0.3445	0.0197±0.0071	-1.663±1.6958	**0.001**
FSHR_T685T	1.470± 0.540	1.062± 0.659	-0.649±0.331	0.019±0.007	-1.89± 1.628	**<0.001**
LHR_L490L	0.7259±0.6710	0.6076±0.8151	0.4609±0.4068	-0.0046 ±0.0083	1.7255±1.9904	0.541
LHR_Q527H	0.0336±0.8498	-0.8328±1.0307	-0.5376±0.5144	0.0151±0.0104	-2.0222±2.4244	0.474
LHR_W467C	-0.0206±0.5054	-0.0382±0.6130	-0.3703±0.3073	0.0078±0.0063	-0.3506±1.5244	0.694
NPYRY2	0.0083±0.6917	0.2947±0.8387	-0.1003±0.4212	0.0036±0.0086	1.1615±2.0403	0.984
bERA_ex8_SNP1820	-0.7231±0.7641	-1.0284±0.9261	-0.1121±0.4658	-0.0050±0.0096	0.1777±2.2834	0.782

- Analysis of genotypes of 427 bulls using REML (multivariate linear mixed model) with sires accounted for as a random effect. Statistical significance for each SNP was estimated by Wald test (Chi probabilities) with 5 degrees of freedom; p, probability.

5.4 Discussion

Five SNPs were identified within the *FSHR* at amino acid residues 502, 669, 596, 658, and 685. All of these SNPs are located on exon 10 of the gene, which encodes the seven transmembrane (TM) regions and the COOH-terminal tail of the protein (see section 1.4.2.3 and Figurs. 5.1 and 5.2). Of these SNPs, only the SNP at residue 658 changes the amino acid sequence, in this case from threonine to serine in the C-terminal intracellular tail of the *FSHR*. This SNP had deleterious effects on fertility PTAs, increasing CI and DFS, while decreasing NR56. The other SNPs at positions 502, 669, and 685, although silent (i.e. not changing the amino acid sequence of the receptor protein), had similar significant effects, while the SNP at position 596 was without effect.

The effects of mutations in the human *FSHR* have been studied in detail, and the findings of this study are in agreement with the phenotype-genotype relationship reported with the different inactivating mutations in the human gene. These effects range from hypergonadotrophic primary amenorrhea with hypoplastic ovaries to secondary amenorrhea with normal sized ovaries and follicular development up to the antral stage. Few inactivating mutations have been identified to date on the human *FSHR* gene, which might reflect selection mechanisms operating against mutations of this kind, based on their strong antifertility effects preventing the inheritance of the unfavourable allele.

A mis-sense Ala189Val mutation in the extracellular domain of the human FSHR leads to female carriers having primordial stage follicles in their ovaries, but few follicles in more advanced stages. It has been suggested that this mutation impairs receptor protein folding accompanied by a decline in hormone binding activity (UIIoa-Aguirre & Timossi, 1998). Another two pairs of mutations were detected in women with fertility problems. The first of these, Ile160Thr and Asp224Val, are in the extracellular domain of the receptor, and cause near-complete loss of FSH binding. The other two mutations, Arg573Cys and Leu601Val, are in a transmembrane domain, causing less complete receptor inhibition. Furthermore,

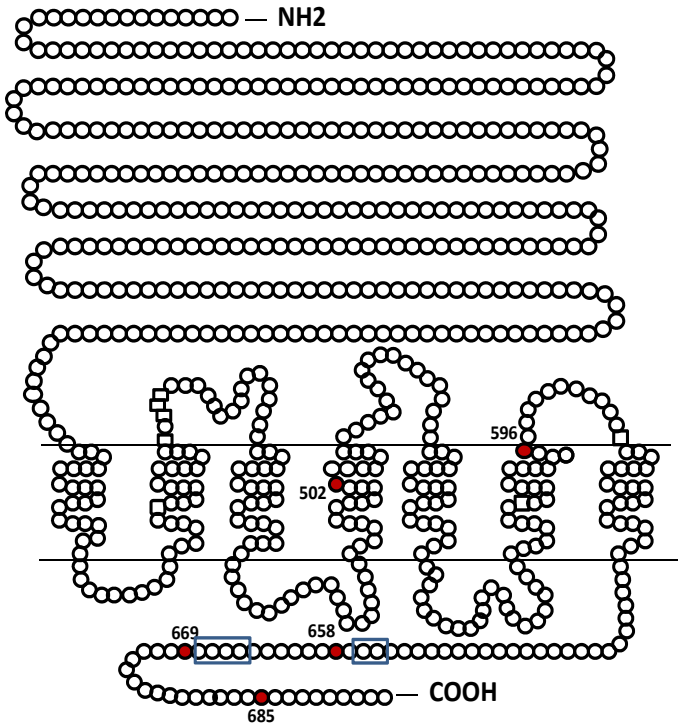
mutations at positions Ala419Thr in a transmembrane loop, and Pro346Arg in the extracellular domain, result in impaired FSH signal transduction, while polymorphisms in the FSHR at positions Thr 307Ala and Asn680Ser were associated with variation in ovarian sensitivity to FSH and the amount of the FSH required to induce similar ovarian stimulation (Huhtaniemi et al., 2001).

The *FSHR* mutations reported in this study are located in regions of the receptor molecule which in the human have been reported to impair fertility. Amino acid sequence alignment between human and bovine *FSHR* revealed a 90% overall sequence identity between both proteins (Figure 5.1). Therefore, it is instructive to compare the positions of the polymorphisms in the two species. The SNPs at residues 658 and 669 are located in the intracellular domain of the receptor. This region is rich in serine and threonine residues in the human FSHR, which may act as potential phosphorylation sites (Simoni et al., 1997). Similarly, the amino acid sequence of the bovine FSHR revealed a high level of serine and threonine residues in this region. Therefore, these SNPs may interfere with receptor coupling to its G protein, blocking signal transduction (Figure 5.2; Ulloa & Timossi, 1998). The SNP at position 502 is located in TM4, which is highly conserved among G-protein coupled receptors. In this region the proline residues contribute to the appropriate insertion of the protein into the membrane. In the bovine FSHR a proline residue is located at position 504, and it can be speculated therefore that the SNP at residue 502 may affect hormone binding (Simoni et al., 1997).

Figure 5.1 Amino acid sequence of bovine and human FSHR alignment (identical amino acids are indicated by stars).

```
Human           MALLLVSLLAFLSLGSGCHHRICHCSNRVFLCQESKVTEIPSDLPRNAIELRFVLTKLRV
60
Cow             MALLLVALLAFLSLGSGCHHRLCHCSNGVFLCQESKVTEIPSDLPRDAVELRFVLTKLRV
60              ******:**************:***** ******************:*:***********

Human           IQKGAFSGFGDLEKIEISQNDVLEVIEADVFSNLPKLHEIRIEKANNLLYINPEAFQNLP
120
Cow             IPKGAFSGFGDLEKIEISQNDVLEVIEANVFSNLPKLHEIRIEKANNLLYIDADAFQNLP
120             * ***************************:**********************:.:******

Human           NLQYLLISNTGIKHLPDVHKIHSLQKVLLDIQDNINIHTIERNSFVGLSFESVILWLNKN
180
Cow             NLRYLLISNTGIKHLPAVHKIQSLQKVLLDIQDNINIHTVERNSFMGLSFESMTVWLSKN
180             **:************* ****:*****************:*****:******: :**.**

Human           GIQEIHNCAFNGTQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSYGLEN
240
Cow             GIQEIHNCAFNGTQLDELNLSDNSNLEELPNDVFQGASGPVILDISRTRIRSLPSYGLEN
240             ***********************.**********:***************:*********

Human           LKKLRARSTYNLKKLPTLEKLVALMEASLTYPSHCCAFANWRRQISELHPICNKSILRQE
300
Cow             LKKLRAKSTYRLKKLPSLEKFVTLVEASLTYPSHCCAFANWRRQTSDLHPICNKSILRQE
300             ******:***.*****:***:*:*:****************** *:*************

Human           VDYMTQARGQRSSLAEDNESSYSRGFDMTYTEFDYDLCNEVVDVTCSPKPDAFNPCEDIM
360
Cow             VDDMTQARGQRVSLAEDDEPSYAKGFDVMYSEFDYDLCNEVVDVTCSPEPDAFNPCEDIM
360             ** ******** *****:*.**::***: *:*****************:***********

Human           GYNILRVLIWFISILAITGNIIVLVILTTSQYKLTVPRFLMCNLAFADLCIGIYLLLIAS
420
Cow             GDDILRVLIWFISILAITGNILVLVILITSQYKLTVPRFLMCNLAFADLCIGIYLLLIAS
420             * :******************:***** ********************************

Human           VDIHTKSQYHNYAIDWQTGAGCDAAGFFTVFASELSVYTLTAITLERWHTITHAMQLDCK
480
Cow             VDVHTKSQYHNYAIDWQTGAGCDAAGFFTVFASELSVYTLTAITLERWHTITHAMQLECK
480             **:*********************************************************:**

Human           VQLRHAASVMVMGWIFAFAAALFPIFGISSYMKVSICLPMDIDSPLSQLYVMSLLVLNVL
540
Cow             VQLRHAASIMLVGWIFAFAVALFPIFGISSYMKVSICLPMDIDSPLSQLYVMSLLVLNVL
540             ********:*::*******.****************************************

Human           AFVVICGCYIHIYLTVRNPNIVSSSSDTRIAKRMAMLIFTDFLCMAPISFFAISASLKVP
600
Cow             AFVVICGCYTHIYLTVRNPNITSSSSDTKIAKRMAMLIFTDFLCMAPISFFAISASLKVP
600             ********* ***********.******:*******************************

Human           LITVSKAKILLVLFHPINSCANPFLYAIFTKNFRRDFFILLSKCGCYEMQAQIYRTETSS
660
Cow             LITVSKSKILLVLFYPINSCANPFLYAIFTKNFRRDFFILLSKFGCYEVQAQTYRSETSS
660             ******:*******:**************************** ****:*** **:****

Human           TVHNTHPRNGHCSSAPRVTSGSTYILVPLSHLAQN 695
Cow             TAHNFHPRNGHCPPAPRVTNGSNYTLIPLRHLAKN 695
                *.** *******..*****.**.* *:** ***:*
```

Figure 5.2 Diagrammatic representation of the secondary structure of the bovine FSHR (redrawn after Ulloa & Timossi, 1998) showing positions of the bovine FSHR SNPs. Potential phosphorylation sites are outlined in rectangles.



Three previously reported SNPs on the *LHR* located at residues 467, 490 and 527 (Hastings et al., 2006) were also tested for possible associations with fertility PTAs in dairy cattle. Although two of these SNPs changed the amino acid sequence at residues Trp467Cys and Gln527His, none of them was significantly associated with effects on any of the fertility PTAs in the present study. These results are in contrast with the previously reported significant effects of these SNPs on fertility PTAs (Hastings et al., 2006). This discrepancy between the two studies may be due to the different genetic background of the animals being genotyped, which might lead to the effects of these SNPs on the fertility PTAs being obscured in the present work. In addition, the previous results were based on the analysis of only two heterozygous sire families, and furthermore PTAs and not de-regressed PTAs were used, either or both of which might have led to an initial false positive result. The use of the PTAs has yielded different results from those of the de-regressed PTAs when analysed in respect to the SNPs in the *activin receptor*, for example (Tables 5.1, 5.2). By

accounting for the reliabilities of PTA estimates the effects of the SNPs did not change in the same direction: with *FSHR* the SNPs became more significant while in the *activin receptor* SNPs lost their significant effects after de-regression. This was unexpected, because the same de-regressed PTA values were used in analysing the effects of both genes. At present, the best explanation for this discrepancy appears to be that the mutations in the 2 genes were not co-inherited (i.e. were not linked), and the distribution of the animals between the homozygotes and heterozygotes differed for these genes.

Moreover, none of the SNPs that were identified in the *activin receptor IIB, NPY*, and *ESR* showed any significant effects on the fertility of dairy cattle. Few studies have been carried out to investigate their effects, as candidate genes, on reproduction; these were covered in section 1.4.2. These SNPs will also be considered later (see Chapters 6 and 7) for further analysis in order to test their effects on a different measure of the fertility (oestrous behaviour) in dairy cattle.

From this study, it can be concluded that the SNPs at residues Leu502Leu, Asn669Asn, Thr658Ser, and Thr685Thr in the *FSHR* have unfavourable effects on the fertility of dairy cattle. Further work is needed to determine the actual effects of these SNPs on the gene or protein functions which will help to understand their effects on fertility.

Chapter 6: Analysis of oestrous behaviour: dataset 1

6.1 Introduction

As discussed before, the dairy cattle industry in the UK is currently experiencing subfertility problems which have been associated with efforts to increase milk production. Therefore, many studies have highlighted the need for genetic improvement of the fertility of dairy cattle (Royal et al., 2002a, Wall et al., 2003). However, the genetic improvement of fertility in dairy cattle has been limited by the low heritabilities of fertility traits currently used in the UK (Wall et al., 2003). Poor oestrous expression is one aspect of subfertility in dairy cattle (discussed in section 1.5.4) which can result in cows being inappropriately inseminated at a time when conception cannot occur, resulting in longer CI (López-Gatius et al., 2005). Standing heat, the standard symptom of oestrus, was observed in only 58 % of cows (Roelofs et al., 2005b). Therefore, oestrous activity measured by electronic tag has been suggested to provide a good indication of fertility in dairy cows (Lopez-Gatius et al, 2005). Various electronic devices have been shown to have potential as tools for achieving more precise detection of oestrus and ovulation time and consequently improving fertility (Roelofs et al., 2005). Detection of oestrus by means of electronic pedometers depends on the behavioural changes associated with oestrus (López-Gatius et al., 2005).

As the need for genetic improvement of fertility has been emphasized in several studies (Royal et al. 2002b; Derecka et al., 2009), the aim of the work reported in this Chapter was to employ SNP association studies with candidate genes to find molecular markers for genes associated with the expression of oestrus. These markers can then be incorporated in selection programmes for improved oestrous behaviour which will accelerate the rate of genetic gain. This study will also help to explore some other factors which might have an impact on oestrous strength, such as milk production. In addition, it will allow the investigation of the possible genetic associations between the traits of oestrous behaviour and other measures of fertility such as calving interval. Furthermore, in order for oestrous detection devices to be

effectively used in the genetic improvement of oestrous expression, genetic variation associated with oestrous traits measured by pedometers has also been assessed. A BLUP (best linear unbiased prediction) animal model was used to estimate the breeding values associated with oestrous traits for all cows in this study. The BLUP animal model utilizes all known relationships among animals (through the inclusion of the numerator relationship matrix A) in order to predict each animal's breeding value (see section 2.5.4).

6.2 Materials and Methods

6.2.1 Animals

Caudal venous blood samples were taken from 103 cows in their first lactation, for which pedometer, milk yield and other reproductive data were available at or preceding a successful insemination, and whose sires were known. DNA isolated from these samples was genotyped at 27 loci in ten genes for SNP identification. Genotyping was carried out commercially (KBiosciences Ltd) by primer extension. Candidate genes were selected as described in Chapter 1, on the basis of a) their involvement in hypothalamic/ovarian/uterine function (*GnRH receptor, LH receptor, FSH receptor, oestrogen receptors α* and *β, activin receptor*), b) known roles in energy metabolism (*leptin, ghrelin*) and c) roles in central nervous pathways directly controlling oestrous behaviour (*neuropeptide Y* and *its receptor*). In some cases the same gene product was expected to be involved in more than one of these processes through functions in different organs (e.g. the oestrogen receptors in the ovary, uterus and hypothalamus). Allele frequencies at each SNP locus in the animals studied here, as determined by simple gene counting, are given in Table 6.1.

Table 6.1 Genotype and allele frequency (%) in the gene loci determined by simple gene counting.

SNP	Allele		genotype frequency %			Allele frequency	
	W	+	WW	W+	++	W	+
ESR-1	T	C	56	38	6	75	25
bGNRHR_-331	A	G	54	43	3	76	24
bGnRHR_-108	T	C	36	55	9	63	37
bGNRHR_206	G	A	84	16	0	92	8
bGnRHR_260	C	T	36	55	9	64	36
bGNRHR_341	C	T	35	56	9	63	37
bGNRHR_410	C	T	36	55	9	63	37
bERA_prom_173	G	A	97	3	0	98	2
LHR_W467C	G	T	30	50	20	55	45
LHR_L490L	C	T	80	18	2	89	11
LHR_Q527H	G	T	69	29	2	83	17
NPYRY2	A	G	76	22	2	87	13
FSHR_L502L	T	C	64	31	5	79	21
FSHR_S596S	C	T	76	24	0	88	12
FSHR_T658S	C	G	64	31	5	79	21
FSHR_N669N	C	T	64	31	5	79	21
FSHR_T685T	A	C	64	31	5	79	21
ACT_IIB_95	A	G	28	52	20	54	46
ACT_IIB_503	C	T	34	45	20	57	43
ACT_IIB_86_END	A	G	34	45	20	57	43
ACT_IIB_45	T	C	28	49	23	53	47
ACT_IIB_46	G	T	34	45	21	56	44
Leptin_963	C	T	45	39	15	65	35
Leptin_1457	G	A	29	45	25	52	48
npy_ex1	C	T	69	28	3	83	17
bERB_ex7	C	G	73	25	2	86	15

- W= the wild type genotype, + = mutant genotype.

6.2.2 Experiment design

For STEPS measurements, animals were equipped with Fullwood pedometers. The animals were held at the dairy cattle research centre in Nottingham University, Sutton Bonington campus. The pedometers recorded the number of STEPS a cow made during the period since the last milking. Receivers were placed at the entrance and exit of the automatic milking system and the data were downloaded automatically to a computer, at each milking (data recording was carried out by Neil Saunders).

6.2.3 Traits

The increase in the number of steps during oestrus was expressed in terms of two different traits. Briefly, the daily mean number of steps (steps/hour) was calculated. An episode of high activity characteristic of oestrous behaviour was established when an increase of locomotor activity measured as number of steps per hour (STEPS) was recorded. The strength of oestrus (STEPS %) was measured as the percentage increase in the number of steps at oestrus over the mean number of steps (steps per hour) for the 10 days preceding oestrus (BASE).

6.2.4 Pedigree file

Pedigree information for the 103 cows was used in order to test the genetic basis for these traits, and to estimate breeding values for the two oestrus traits (STEPS and STEPS%) applying a BLUP animal model. All known pedigree information for the preceding three generations for each cow was recorded. This information, which was used to test the possible associations between the SNPs and oestrous traits in addition to the estimate of the animals' breeding values for these traits, included a total of only 297 animals which is extremely inefficient for such kind of genetic analysis. This file was created by including the herd book number (HBN), extracted from the Holstein UK database, for the animals, sires and dams of the 103 cows for which the SNPs and pedometer information was available in the data file. The pedigree file was

used to define the genetic relationships for fitting a genetic animal model in ASREML.

6.2.5 Statistical analysis

STEPS trait data were not normally distributed (Figures 6.1; 6.2) and because of the importance of normally distributed data when estimating variance components using linear models, oestrous traits were normalised as log to base 10. Mixed model association analysis was used to determine any significant associations between the SNPs and the log transformed data for both the number of STEPS and the percentage increase of STEPS during oestrus, applying a BLUP animal model.

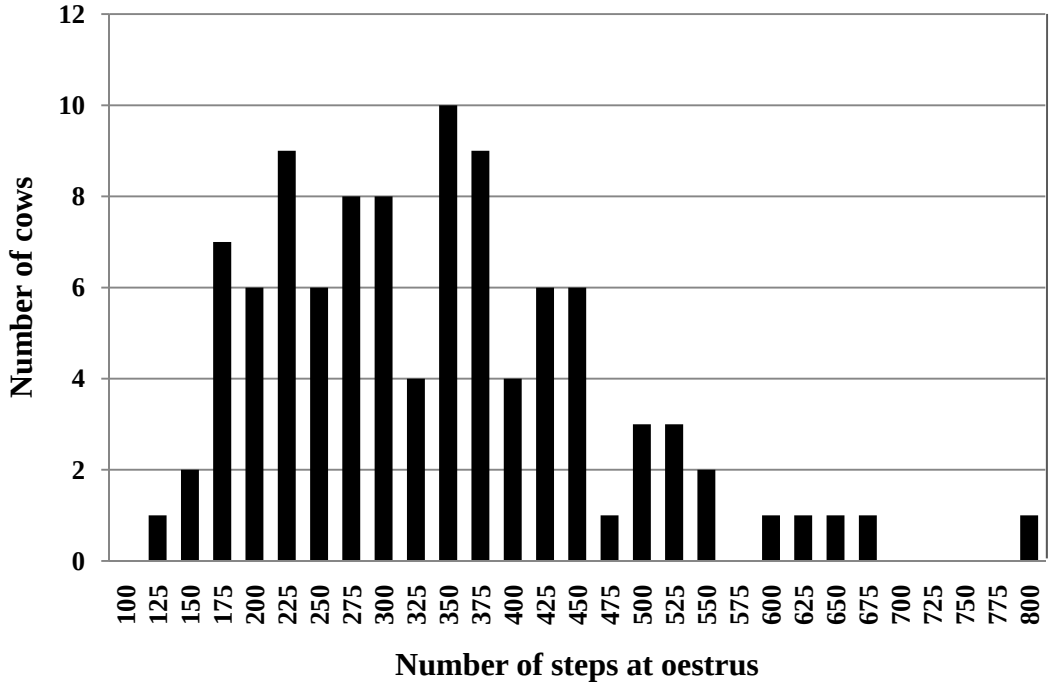


Figure 6.1 Frequency distribution of STEPS records of 103 dairy cows.

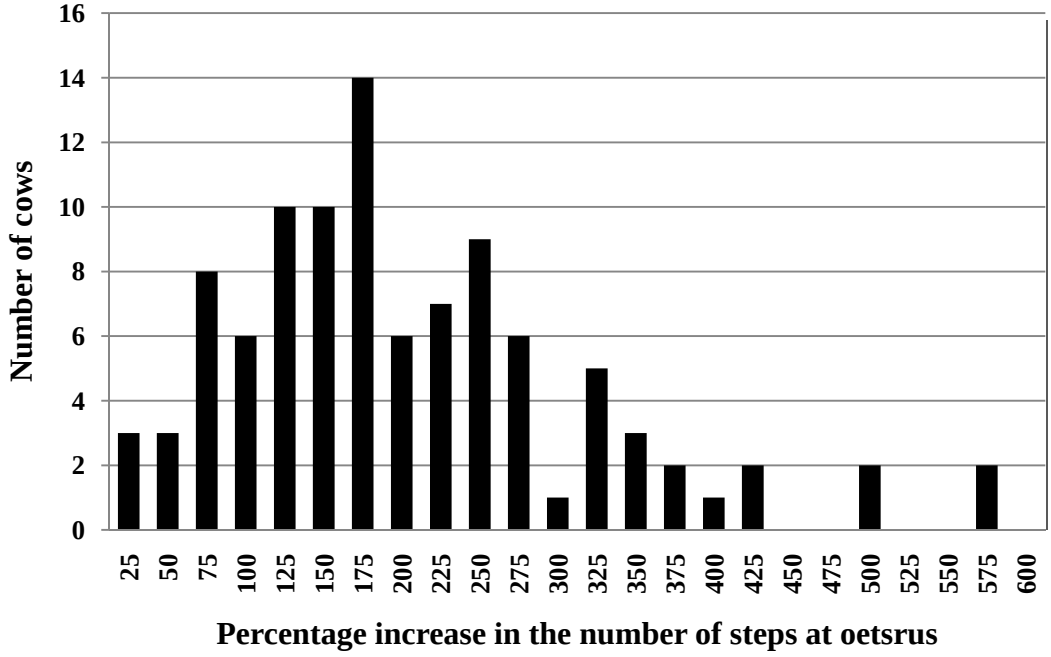


Figure 6.2 Frequency distribution of STEPS% records of 103 dairy cows.

The additive effects of the SNPs were tested by fitting each SNP individually as a fixed effect in the model. The effects of other factors such as phenotypic calving interval, days in milk, PIN and oestrous month were also fitted as covariates. All known pedigree information on the preceding three generations for each cow was also included (n = 297) to account for the additive genetic effects for each animal in the data using the pedigree information. Thus the random model estimated the additive genetic variance (σ_a^2) utilizing known pedigree. The additive effect were assumed to be multivariate normally distributed with a mean of 0 and (co)variance of $A\sigma_a^2$, where A is the numerator relationship matrix of the animals derived from the pedigree information available in the data. Hence the statistical model used to investigate the SNPs effects on behaviour-derived oestrus traits was:

$$y_{ijkl} = u_i + A_{ik} + B_{ic} + C_{id} + D_{ijl} + E_{im} + F_{il} + J_{ijkl}$$

y_{ijkl} = Value for trait *i* for animal *l* of sire *j* carrying genotype k, with effects as follow:

u_i = Overall mean

A_{ik} = Fixed effect of the genotype *k* for trait *i*

B_{ic} = Fixed effect of calving interval *c* for trait *i*

C_{id} = Fixed effect of days in milk *d* for trait *i*

D_{ijl} = Fixed effect of PIN for son *l* of sire *j* for trait *i*

E_{im} = Fixed effect of month *m* for trait *i*

F_{il} = Random genetic effect of animal *l* for trait *i* (N(0, σ_a^2)

J_{ijkl} = Random error term, assumed to be normally distributed with mean of 0 and variance σ_e^2 (N(0, σ_e^2)

The genetic variations associated with oestrous traits were tested in similar model to the one mentioned above, but with excluding the SNPs effects. For illustration this model can be expressed in matrix notation as:

$$y = Xb + Za + e$$

y = n x 1 vector of observations; n = 103 is the number of records;
b = p x 1 is an unknown fixed vector; p = 18 is the number of levels for fixed effects;
a = q x 1 vector of random animal effects; q = 297 is the number of levels for random effects;
e = n x 1 vector of random residual effects;
X is the design matrix of order n x p (103 x 18), which relates records to fixed effects; and Z is a design matrix of order n x q (103 x 297), which relates records to random animal effects.
Both X and Z are termed incidence matrices in which each element consists of either a zero or one, depending on which level of the fixed or random effect each animal is classified.

6.3 Results and discussion

6.3.1 Characteristics of phenotypic data

A total of 103 oestrous records were analysed. Over the 10 days preceding oestrus, the maximum value for the mean number of STEPS (BASE; n=209) was nearly three times higher than the minimum value (n=65). However, on the day of oestrus this difference became sevenfold (n=777 vs 115). On average, the number of STEPS on the day of oestrus (STEPS) was approximately double the basal number of STEPS (BASE; 192%) with variation from almost zero difference (102%) to approximately 658% (Table 6.2).

Table 6.2 STEPS measurement at time of oestrus

Trait	Minimum	Mean	Maximum	s.d.
BASE	64.80	116.4	208.9	32.24
STEPS	115.0	327.6	777.0	128.2
STEPS%	2.764	192.0	657.9	115.6

- BASE = mean STEPS (STEPS/hour) for 10 days preceding oestrus
- STEPS = STEPS (STEPS/hour) on day of oestrus
- STEPS % = STEPS a s percentage of BASE

6.3.2 SNPs effects on oestrous expression

Tables 6.3 and 6.4 show the genetic effects on STEPS and STEPS% when the additive effect for each SNP was fitted singly into the model. These effects represent the effects of allelic substitution at the corresponding SNP. Twenty SNPs did not show any association with locomotor traits and will not be considered further. These SNPs were: GnRHR_-331, GnRHR_206, LHR_467, LHR_490, LHR_527, NPYRY2, FSHR_502, FSHR_596, FSHR_658, FSHR_669, FSHR_685, ACT_95, ACT_503, ACT_86, ACT_45, ACT_46, leptin_1457, leptin_963, bERB_ex4, and bERB_ex7. The remaining seven SNPs had significant effects, and these are described below (the results will be presented per 100 STEPS and per 100% increase in the number of STEPS for STEPS and STEPS% respectively):

GnRHR

Four of the SNPs in the *GnRHR*, at positions -108, 260, 341, and 410 had significant effects on both STEPS and STEPS%. The impacts of allelic substitution at these positions were of similar magnitude, due to the linkage disequilibrium between them demonstrated previously (Chapter 4). These SNPs were associated with unfavourable effects on oestrous expression, reducing both STEPS and STEPS% by 16 and 34% per 100 STEPS and 100% respectively. Therefore, cows carrying the relatively rare T allele at positions 260, 341 and 410 (with a C at -108) are less likely to show strong oestrous behaviour than cows carrying C alleles at 260, 341 and 410, and they will therefore experience poorer fertility. These results are contradictory to the previous findings of the beneficial effect of the C allele at -108 on DFS (Tables 4.4 and 4.5).

ERα

Two of the SNPs in the *ERα* were found to be significantly associated with oestrous expression, and this is not unexpected, as oestradiol level is highly correlated with oestrous behaviour in dairy cattle (Lyimo et al., 1999). The SNP at position 503 was associated with reductions in both STEPS and STEPS% by 20 and 35 %

respectively. In addition, allelic substitution in the promoter region of ERα at position 173 also had unfavourable effects on oestrous expression, reducing both STEPS and STEPS% by 39 and 82 % respectively. Therefore, oestrous expression tends to be weaker in cows carrying the C or A allele at *ER1* and ERprom173 respectively than in cows carrying T or G alleles.

Neuropeptide Y

One SNP in *NPY* located in exon 1 was associated with a significant effect on oestrous expression. Cows carrying this SNP tended to show fewer STEPS at oestrus. On the other hand, this SNP did not have any significant effect on STEPS%. This result is in agreement with the reported effects of NPY on reproduction: ovariectomized cows treated with NPY experience a decreased circulating LH level, associated with an increase in growth hormone (GH). In addition, an elevated level of NPY expression was reported to be accompanied by low LH and high GH levels during feed restriction and poor body condition in other species (Thomas et al., 1999).

Other factors such as calving interval, days in milk, PIN and oestrous month had no significant effects on oestrous behaviour in the present study.

Table 6.3 Statistically significant effects of different allelic substitutions on STEPS.

Variate	STEPS	s.e.	P value
ESR1	-20	7	**0.002**
GnRHR-29	-16	7	**0.014**
GnRHR-340	-16	7	**0.011**
GnRHR-421	-16	7	**0.011**
GnRHR-490	-16	7	**0.014**
ERprom173	-39	29	0.054
npy_ex1	-16	9	**0.043**

Table 6.4 Statistically significant effects of different allelic substitutions on STEPS%.

Variate	STEPS%	s.e.	P value
ESR1	-35	15	0.004
GnRHR-29	-34	15	0.003
GnRHR-340	-34	14	0.002
GnRHR-421	-34	14	0.003
GnRHR-490	-34	15	0.003
GnRHR-286	-42	27	0.027
ERprom173	-82	64	<.001

From the different mixed models applied, it can be concluded that ESR1, GnRHR-108, and ERprom173 were significantly associated with poorer expression of oestrus. However the database used here was small, and more data are needed to confirm these effects.

6.3.3 Estimation of breeding values for oestrous strength

In general, breeding values are taken as the estimated random effects for the animals when fitted in an animal model. With the data set analysed in this Chapter the additive variance was effectively zero, so the animal model BLUPs or BVs were also effectively zero. The BVs for STEPS and STEPS% of 10 of the animals in the pedigree file are summarized in table 6.5-6.6:

Table 6.5 BVs for STEPS of 10 of the animals in the pedigree.

Animal HBN	BVs	s.e
UK60000340969084	0	0.0183
UK65000002249055	0	0.0183
UK108805234	0.0004	0.0183
UK72000601031850	0	0.0183
UK65000014689001	0.001275	0.01828
UK65000002119526	0.001804	0.01828
UK108690659	0	0.0183
UK341005854	0	0.0183
UK60001020888555	0.0008729	0.01829
UK64000000392457	0.001372	0.01827

Table 6.6 BVs for STEPS% of 10 of the animals in the pedigree.

Animal HBN	BVs STEPS%	s.e.
UK60000340969084	0	0.0004592
UK65000002249055	0	0.0004592
UK108805234	2.157E-08	0.0004592
UK72000601031850	0	0.0004592
UK65000014689001	0.000000208	0.0004592
UK65000002119526	2.222E-07	0.0004592
UK108690659	0	0.0004592
UK341005854	0	0.0004592
UK60001020888555	0.000000186	0.0004592
UK64000000392457	2.785E-07	0.0004592

This study was not planned for the evaluation of the genetic variation of oestrous traits. As such, in the current study with a small number of records (103) it was not possible to estimate breeding values or heritability. In practice, this kind of study requires large numbers of related animals with records, and sufficient numbers of daughters per sire (at least 10 daughters, Mrode personal comunication) are needed to eliminate environmental effects so as to estimate BVs for sires. In general, if the genetic variations could not be estimated this might be due to:

- The trait was not influenced by genetics
- It was strongly influenced by something else which was not recorded or recognised and which increased the residual variance in this data set. However, there was evidence of some SNP effects, which require further collection of independent data for confirmation.
- From the tabulation file it could be seen that most sires had a single record and that they were unrelated. Therefore, this might be the reason why the traits have shown no genetic variance.

6.4 Conclusion

Three previously unreported SNPs have been identified in genes involved in fertility, and it has been demonstrated for the first time that genes and polymorphisms can be identified as genetic markers for improved oestrous expression. These new findings may provide the opportunity for animals to be selected for improved oestrous STEPS, but without possible negative effects on other fertility traits. Despite the significant effects of some of the SNPs on oestrous traits, the genetic variation associated with these traits was not estimable. Therefore, more data are needed in order to be able to further investigate the genetic variation and the predictive values for oestrous traits. This approach was further investigated in the following Chapter, which includes further analysis of oestrous traits and the factors that may have an impact on them.

Chapter 7: Application of BLUP in the analysis of oestrous behaviour: dataset 2

7.1 Introduction

Many studies have highlighted the need for genetic improvement of dairy cattle in the UK, and this was covered in Chapter 1. However, the fertility traits currently employed in the genetic improvement of fertility have the drawback of low heritabilities, which slows the rate of genetic improvement (Wall, 2003). The low heritability of these traits reflects the fact that they are influenced by management decisions. In contrast, traits based on physiological function, which are less affected by management procedures, in general have higher heritabilities, and therefore allow more rapid genetic progress. Commencement of luteal activity after calving using milk progesterone levels is one such physiological trait, which has been identified as a promising tool for improving fertility (Royal et al., 2002b; Petersen et al., 2007).

In dairy herds, low oestrous detection is considered to be one of the major factors contributing to poor fertility (low conception rates; Lyimo et al., 1999). Currently, only a limited proportion of cows is reported to display standing heat (50%; Van Eerdenburg et al., 2002). Therefore, oestrous activity measured by electronic tag has been suggested to be a trait which could provide physiological information about fertility (Løvendahl and Chagunda, 2009). Despite the widespread use of the electronic devices, the genetic characteristics of the associated traits have yet to be established. In order for information on locomotor activity to be an effective tool in breeding programmes, the genetic variation associated with these traits, and the breeding values of animals for these traits need to be evaluated.

The objective of this study was to evaluate the genetic and phenotypic parameters associated with oestrous behaviour traits in more detail. These parameters will allow the identification of animals that are genetically superior in expressing oestrus, so that choosing these animals in breeding programmes will improve oestrous expression and fertility in dairy cattle. Molecular markers for oestrous activity strength discussed in the previous chapter were also investigated further.

7.2 Methods

7.2.1 Animals

Data were recorded in 189 Holstein cows held at the dairy cattle research centre in Nottingham University, Sutton Bonington campus. Caudal venous blood samples were drawn from these cows to determine their genotypes at 16 genes which were chosen on the basis of: their involvement in the hypothalamic/ovarian/uterine function (*GnRHR, FSHR, LHR, oestrogen receptor α,* and *β,* and *activin receptor*), their known roles in energy metabolism and milk production (*ghrelin, leptin, signal transducer and activator of transcription 1 (STAT1), signal transducer and activator of transcription 5A (STAT5A), growth hormone receptor (GHR), prolactin (PRL),* and *prolactin receptor (PRLR))* and their roles in central nervous system function directly controlling oestrous behaviour (*neuropeptide Y (NPY)* and *neuropeptide Y receptorY2 (NPYRY2)).* As in Chapter 6, the cows' genotypes were then matched to records of their locomotor activity. These records were repeatedly made for each cow until pregnancy was confirmed.

The database also contained individual average daily milk yields between calving and oestrus (MYa), average milk yield over the 10 days around oestrus (MY10), heat season, calving season, parity and days in milk after calving. All of the cows showed oestrus at least once. The number of oestrous periods per successful pregnancy ranged from 1 to 15. The parity of the cows varied between one and three, and activity data were obtained from them between 8 and 638 days in milk (mean ± s.d., 130 ± 101). Average daily milk yield since calving ranged between 1.5 and 54 kg/day. Average milk yield over the 10 days around oestrus ranged between 9 and 56 kg/day. The voluntary waiting period for artificial insemination (AI) was 40 days. AI was repeatedly performed until the cows either became pregnant or were culled.

7.2.2 Measurement of oestrous activities

Cows were milked between 1 and 3 (mean 2.8) times daily through a Lely Astronaut automatic (robotic) milking system. Each cow was fitted with a neck collar that

provided identification and also activity data per each two hour period. In the Lely system, activity information is reported as an index that quantifies all animal movements (walking, running, laying, standing up, head movement etc). These data are downloaded at milking and split into activity units for 2 hourly intervals. This can then be plotted against date and time; with peaks in the graph denoting increased activity, indicating an oestrous period (Figure 7.1). Locomotor activities stored were: the actual level of activity in each 2 hours, average baseline activity (BASE, BA) determined over a 4 days rolling average (48 activity episodes) and activity change, increase or decrease from the baseline. This allowed the definition of the following oestrous traits: ACTIVITY% (AU%) which is a measure of the strength of oestrous expressed as percentage increase in activity units, calculated as largest increase in activity on the day of oestrus divided by baseline activity (BA), multiplied by 100 and ACTIVITY: highest actual number of activity units (AU) at the day of oestrus. The episodes were numbered starting from 1 (indicating the first oestrus after calving). Data were recorded from March 2008 (when the new Lely robotic milking system was installed) until April 2010, which covered a total of 995 oestruses (data recording was carried out by Neil Saunders).

.

For management purposes, oestrus was detected using a combination of visual observation and locomotor activity records. Therefore, a locomotor oestrous alert was defined as correct when it was confirmed by behavioural (recorded) oestrus; otherwise it was regarded as false oestrus. In addition, any oestrous events (recorded or unrecorded) that occurred under and including 25 days postpartum were removed from analysis to eliminate any false recordings that could be mistaken for an oestrous event. Activity increases before 25 days postpartum could be due to cows re-entering the herd after solitary calving and establishing hierarchies, mixing with new herd mates and entering a new environment.

The mean values for ACTIVITY, ACTIVITY% and BASE at each oestrus are presented in Table 7.1. The range of the number of days between calving and oestrus, number of inseminations applied, number of successful pregnancies and number of oestruses confirmed by visual observation are presented in Table 7.2.

Figure 7.1: Example of activity changes 10 days befor and 10 days after the day of oestrus

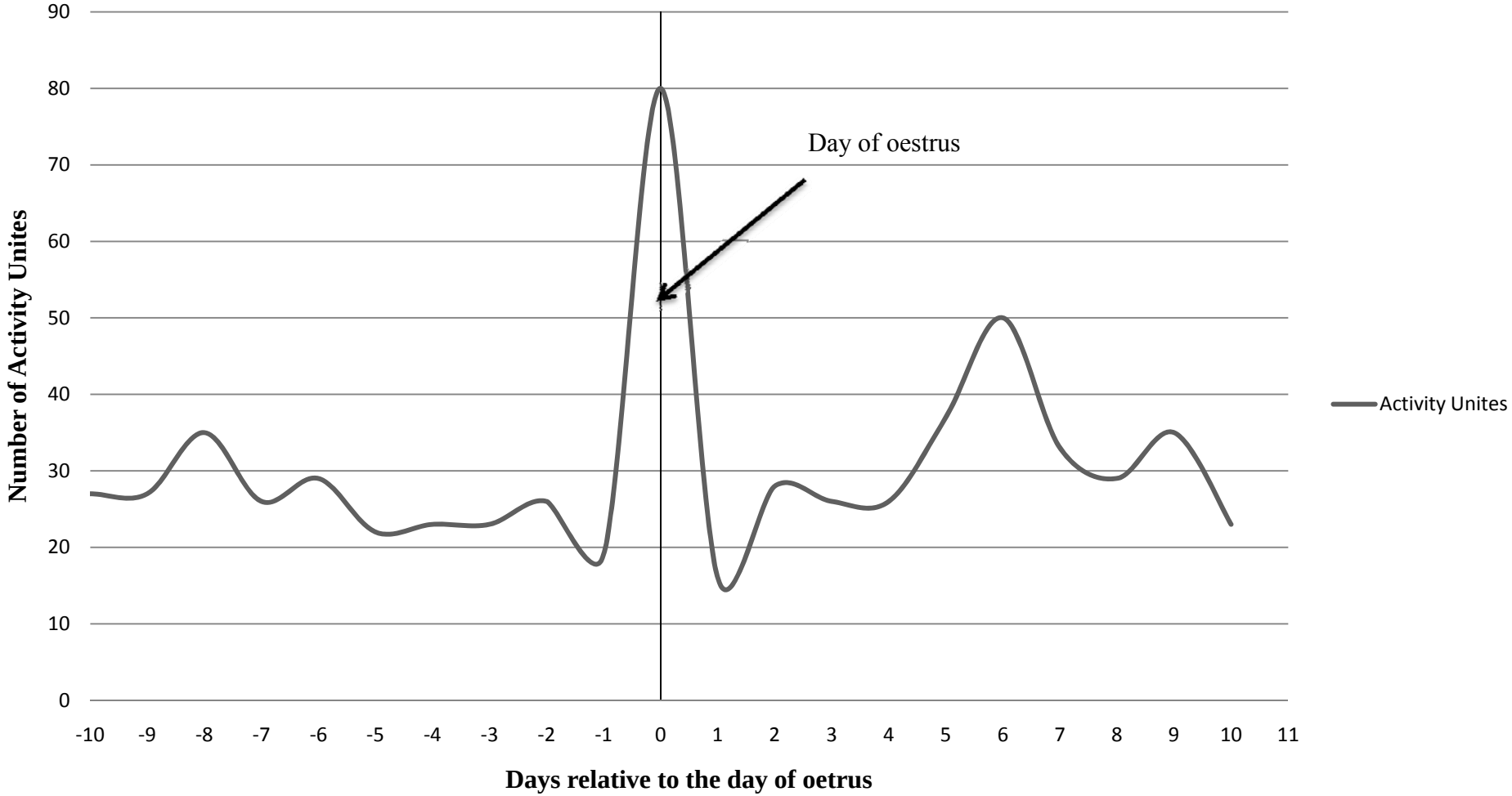


Table 7.1 Mean values for ACTIVITY, ACTIVITY%, and BASE for oestruses 1 to 10.

Oestrus No	No of cows	Mean BASE	Mean ACTIVITY	Mean ACTIVITY%
Oestrus1	261	35.15	68	94
Oestrus2	220	35.36	70	98
Oestrus3	165	34.5	67.5	95
Oestrus4	110	35.82	66.82	90
Oestrus5	76	34.2	68	99
Oestrus6	50	35	67	91
Oestrus7	35	34.2	67	96
Oestrus8	29	34	63	86
Oestrus9	16	34	69	101
Oestrus10	13	32.2	64.3	98

Table 7.2 The range of days from calving, numbers of recorded oestruses inseminations, and pregnancy rates, for oestruses 1 to 10.

Oestrus No	Days from calving	Recorded	Insemination	Pregnant
Oestrus1	8-230	194	139	38
Oestrus2	22-450	185	169	56
Oestrus3	42-429	152	151	52
Oestrus4	65-414	90	8	29
Oestrus5	72-434	67	65	24
Oestrus6	85-448	43	43	13
Oestrus7	110-479	32	31	3
Oestrus8	134-502	25	25	9
Oestrus9	175-442	14	14	3
Oestrus10	183-464	11	11	3

7.2.3 Genetic analysis

DNA isolated from the blood samples was genotyped at 37 loci in 16 genes for SNP identification. Genotyping was carried out commercially (KBioscience Ltd) by primer extension. The list of SNPs studied, their chromosomal locations, base pair

positions and genotypes are summarized in Table 3.4. Allele frequencies at each SNP locus determined by simple gene counting are given in Table 7.3.

Table 7.3 Genotype and allele frequency (%) in the gene loci studied determined by simple gene counting.

SNP	Allele		genotype frequency			Allele frequency	
	W	+	WW	W+	++	W	+
STAT1_c3141t	C	T	44	46	10	67	33
GHR_Phe279Tyr	T	A	60	37	3	78	22
PRL_89398_g_a_R	G	A	75	22	3	87	13
STAT5A_g12195c	C	G	53	35	12	70	30
ESR1	T	C	62	35	3	79	21
bGNRHR_206	G	A	86	13.5	0.5	93	7
bGNRHR_341	C	T	37	46	17	60	40
bGnRHR_410	C	T	38	46	16	61	39
bGNRHR_-331	A	G	62	34	4	79	21
bGNRHR_-108	T	C	38	46	16	61	39
bERA_prom_SNP173	G	A	92	8	0	96	4
LHR_W467C	G	T	39	48	13	63	37
LHR_L490L	C	T	76	23	1	87	13
LHR_Q527H	G	T	74	25	1	86	14
NPYRY2	A	G	74	25	1	86	14
FSHR_L502L	T	C	58	36	6	76	24
FSHR_S596S	C	T	79	21	0	89	11
FSHR_T658S	C	G	58	36	6	76	24
FSHR_N669N	C	T	58	36	6	76	24
FSHR_T685T	A	C	58	36	6	76	24
ACT_IIB_95	A	G	31	52	17	57	43
ACT_IIB_503	C	T	36	49	15	60	40
ACT_IIB_86_END	A	G	35	48	17	59	41
ACT_IIB_45	T	C	31	52	17	57	43
ACT_IIB_46	G	T	35	48	17	59	41
GHRA257G_ex10	A	G	91	9		95	5
leptin963	C	T	40	48	12	64	36
npy_ex1	C	T	59	38	3	78	22
bERB_ex7	C	G	68	26	6	81	19
leptin1457	G	A	27	50	22	52	47
PRLR_Ser18Asn	A	G	81	0	19	81	19
ESR1_ex1_A503C	A	C	42	45	13	64	36

- W= the wild type genotype, + = mutant genotype

7.2.4 Statistical analysis

The pedigree of all cows was traced back 4 generations (using the Holstein UK database) in order to construct the relationship matrix for genetic evaluation; the pedigree file included 893 animals. Normality tests showed that oestrous traits ACTIVITY and ACTIVITY % were not normally distributed; therefore these traits were log transformed before analysis. Variance components and fixed effect solutions were calculated using the ASREML software package (Gilmour et al., 2009). A single mixed model was initially used to determine the effects of different factors on oestrous activity traits. Oestrous activity traits were modelled as dependent on: SNPs (3 levels), parity (3 levels), heat season(1 = winter (January-March),2 = Spring (April-June), 3 = Summer (July-September), 4 = Autum (October-December), oestrus number (15 levels), average daily milk since calving (MYa) and average milk yield over the baseline period (MY10), which were fitted as fixed effects. The random model estimated the additive genetic variance (σ_a^2) utilizing known pedigree and the permanent environmental variance (σ_{pe}^2) utilizing the repeated records on cows. The additive effect were assumed to be multivariate normally distributed with a mean of 0 and (co)variance of $A\sigma_a^2$, where A is the numerator relationship matrix of the animals derived from the pedigree information available in the data. Thus the statistical model applied to estimate SNPs effects was fitted as:

$$y_{ijkl} = u_i + A_{ik} + B_{im1} + C_{im2} + D_{im} + E_{is} + F_{il} + J_{il} + H_{ijkl}$$

y_{ijkl} = Value for trait *i* for animal *l* of sire *j* carrying genotype *k*, depends on:

u_i = Overall mean

A_{ik} = Fixed effect of the genotype *k* for trait *i*

B_{im1} = Fixed effect of MY10 *m1* for trait *i*

C_{im2} = Fixed effect of MYa *m2* for trait *i*

D_{im} = Fixed effect of oestrous number *m* for trait *i*

E_{is} = Fixed effect of heat season *s* for trait *i*

F_{il} = Random genetic effect of animal *l* for trait *i* (j=1...893, N(0, Aσ_a^2)

J_{il} = Random permanent environmental effects of animal *l* for trait *i* (N(0, σ_{pe}^2)

H_{ijkl} = Random error term (N(0, σ_e^2)

where J_{il} and H_{ijkl} were assumed to be independent and normally distributed with a mean of 0 and variances of σ^2_{pe} and σ^2_e respectively.

Initially SNPs were fitted as variates in order to test their additive genetic effects on oestrous traits. Subsequently, the SNP effects were fitted as factors in which case ASREML assigned values -1, 0, +1 corresponding to the wild type homozygous, heterozygous, and mutant homozygous respectively. The analysis was repeated for each locus and trait.

Using the model described above a series of univariate analyses were undertaken for both ACTIVITY and ACTIVITY% traits using different datasets as follows:

- **Dataset 1:** In the first analysis, all of the 995 oestruses in the database were fitted in the analysis. This test allowed the identification of SNPs that had significant effects on the expression of oestrus. To account for repeated measures (permanent environmental effects), a BLUP repeatability model was fitted.
- **Dataset 2:** In the second analysis, only oestruses that were confirmed by visual observation were considered as true oestruses and included in the analysis.
- **Dataset 3:** In the third analysis, data were limited to the first occurring oestrus in order to examine the possible effects of SNPs on oestrous activity associated with the onset of cyclic activity after calving. In addition, this identified genetic variation associated with the first oestrus.
- **Dataset 4:** In the fourth analysis, only oestruses associated with insemination and a successful pregnancy were included in the analysis. This allowed the estimation of the genetic variation for oestruses associated with successful pregnancy.

In a separate series of analyses the additive effects of the SNPs were tested in a bivariate analysis which included ACTIVITY and % ACTIVITY or BASE and ACTIVITY.

Hence the statistical model applied to estimate SNPs effect in biavraite analysis was fitted as:

$$y_{jkl} = u + A_k + B_{m1} + C_{m2} + D_n + E_s + F_l + J_l + H_{jkl}$$

y_{jkl} = ACTIVITY, ACTIVITY%, or ACTIVITY, BASE (depending on the traits fitted) for animal *l* of sire *j* carrying genotype *k*, depends on:

u = Overall mean

A_k = Fixed effect of the genotype *k* for both traits

B_{m1} = Fixed effect of MY10 *m1* for both traits

C_{m2} = Fixed effect of MYa *m2* for both traits

D_n = Fixed effect of oestrous number *n* for both traits

E_s = Fixed effect of heat season *s* for both traits

F_l = Random genetic effect of animal *l* for both traits (j=1...893, N(0, Aσ_a^2)

J_l = Random permanent environmental effects of animal *l* for both traits (N(0, σ_{pe}^2)

H_{jkl} = Random error term (N(0, σ_e^2)

The genetic variations associated with these traits were also tested in a series of univariate and bivariate analyses (similar to the above described models) that excluded the SNP effect. The heritability, repeatability, and the genetic and phenotypic correlations among these traits were also estimated in this series of analyses. The following equations were used to calculate a number of parameters from the variance components:

$$\sigma_P^2 = \sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2$$

$$h^2 = \frac{\sigma_a^2}{\sigma_P^2}$$

$$re = \frac{\sigma_a^2 + \sigma_{pe}^2}{\sigma_P^2}$$

where: h^2 is the heritability, *re* is the repeatability, σ_a^2 is the animal additive genetic variance, σ_P^2 is the total phenotypic variance, $\sigma_{pe}^2$is the animal permanent environmental variance.

In addition to the above described parameters, the variance components of the bivariate analyses including traits X_1 and X_2 were used to estimate the phenotypic and genotypic correlation between the two traits as follows:

$$r_P = \frac{cov_P(X_1, X_1)}{\sigma_P(X_1)\sigma_P(X_{2)}}$$

$$r_A = \frac{cov_A(X_1, X_1)}{\sigma_A(X_1)\sigma_A(X_{2)}}$$

where: $cov_P(X_1, X_1)$, $cov_A(X_1, X_1)$ are the phenotypic and genotypic covariances between the two traits respectively. $\sigma_P(X_1)$, $\sigma_A(X_1)\sigma_A$ are the the phenotypic and genotypic standard deviation for trait X_1; $\sigma_P(X_2)$, $\sigma_A(X_2)\sigma_A$ are the the phenotypic and genotypic standard deviation for trait X_2,

7.3 Results

7.3.1 Characteristics of phenotypic data

A total of 995 occurrences of oestrus were detected by locomotor activity (Table 7.1), 832 of which were confirmed by visual observation. Of the confirmed oestruses 754 were followed by insemination (91%), 237 of which resulted in pregnancy (31%). However, the pregnancy status of 32 inseminated cows was unknown at the time of data collection. Detection efficiency can therefore be measured as a percentage of the number of confirmed pregnancies relative to the total number of inseminations (33%; excluding unknown pregnancy status). The error rate can be measured as the proportion of high activity oestruses that were not associated with visual confirmation of oestrus (16%).

During the 10 day period before oestrus there was a fivefold difference between the minimum and maximum activity levels; on the day of oestrus there was sevenfold difference. Descriptive statistics for the traits considered in the present study are shown in Table 7.4.

Table 7.4 Descriptive statistics for the traits ACTIVITY, ACTIVITY%, and BASE.

Trait	Minimum	Mean	Maximum	s.d.
Base	15	34	72	8.3
ACTIVITY	26	67	150	19.4
ACTIVITY%	0.31	0.95	5.6	0.33

On average, the number of AU at oestrus was approximately double the mean number of AU during the preceding 4 days, but the difference ranged between 0 and sixfold. Individual cows expressed up to 15 oestrous events, with the time of oestrus varying from 8 to 638 days in milk. There were no differences in the traits BASE, ACTIVITY and ACTIVITY % with successive oestrous events (see Tables 7. 5-7.7).

Table 7.5 Descriptive statistics for of ACTIVITY expressed per oestrus number.

Oestrus number	Mean	s.d.	Minimum	Maximum	Count
1	67	19	29	122	173
2	69	20	33	126	177
3	67	20	26	150	151
4	68	19	37	112	90
5	69	17	36	109	67
6	68	20	39	129	43
7	68	20	28	115	32
8	62	18	40	117	25
9	70	19	43	117	14
10	66	21	38	104	11
11	64	20	39	109	9
12	53	3	50	57	5
13	66	18	52	87	3
14	49	0	49	49	1
15	87	0	87	87	1

Table 7.6 Descriptive statistics for of ACTIVITY% expressed per oestrus number.

Oestrus number	Mean	s.d.	Minimum	Maximum	Count
1	98	46	36	560	173
2	97	32	40	200	177
3	95	29	31	194	151
4	92	30	42	181	90
5	102	32	38	177	67
6	91	30	44	163	43
7	97	28	40	161	32
8	86	28	45	142	25
9	102	24	59	141	14
10	1	20	68	129	11
11	95	28	56	142	9
12	91	19	67	119	5
13	102	29	79	135	3
14	96	0	96	96	1
15	132	0	132	132	1

Table 7.7 Descriptive statistics for of BASE expressed per oestrus number.

Oestrus Number	Mean	s.d.	Minimum	Maximum	Count
1	34	8	7	56	173
2	35	9	17	70	177
3	34	8	17	63	151
4	35	9	23	68	90
5	35	7	22	49	67
6	36	8	23	52	43
7	34	11	19	66	32
8	33	9	20	62	25
9	35	7	26	53	14
10	33	10	21	51	11
11	32	7	25	45	9
12	28	2	26	30	5
13	32	4	29	37	3
14	25	0	25	25	1
15	38	0	38	38	1

7.3.2 Effects of SNPs on oestrous traits

Preliminary analyses of the entire dataset revealed that oestrous traits ACTIVITY and ACTIVITY %, although related, are controlled by different SNPs. Therefore they are discussed here separately. SNP effects are presented per 100 AU or 100% increase in activity at oestrus.

7.3.2.1 Analysis of additive effects of SNPs

This analysis allowed the identification of SNPs having additive genetic effects on oestrous traits. The results are described below. For the sake of brevity, only those effects which were statistically significant are presented in the following Tables.

ACTIVITY

Only two SNPs were associated with statistically significant additive genetic effects on ACTIVITY. The SNP in the *GHR*, phe279tyr, and the SNP in the *PRLR*, ser18asn were associated with a reduction in the number of AU by 5 and 7 per 100 AU at oestrus. Other effects associated with two SNPs located on the *activin receptor* at positions 95 and 86 were close to significance, suggesting a possible association with ACTIVITY. Heat season, parity and MY10 were also associated with significant effects on ACTIVITY. Animals at season three and parity 1 tended to show the highest level of ACTIVITY. In addition, every 1kg increase in MY10 was associated with a decline in ACTIVITY by 1 AU per 100 AU at oestrus (Table 7.8).

Table 7.8 Additive allele substitution effect on ACTIVITY.

Factor	Genotype	Effect	s.e.	P value
GHR_phe279tyr	T:A	-5	3	0.045
ACT_IIB _95	A:G	-4	2	0.055
ACT_IIB _86	A:G	-4	2	0.059
PRLR_ser18asn	A:G	-7	3	**0.033**
Heat-season	1	0	0	
Heat-season	2	3	1	**<0.001**
Heat-season	3	6	1	
Heat-season	4	2	1	
Parity	1	0	0	
Parity	2	-1	2	**0.002**
Parity	3	-6	2	
MY10	1	-1	0.4	**0.027**

ACTIVITY%

Table 7.9 shows the effects of allele substitutions on ACTIVITY% when SNPs were fitted individually in the additive allelic effect model. Only, STAT1 SNP and ESR1 SNP had significant effects, of -5% and -6% (percent change) on ACTIVITY%, respectively. MY10, MYa and parity were also associated with significant effects on ACTIVITY%. Parity 2 was associated with the highest values for ACTIVITY%. MY10 was associated with a decrease of 1%, while MYa was associated with an increase of 1% in ACTIVITY% per 1kg increase in milk yield.

Table 7.9 Additive allelic substitution effects on ACTIVITY%.

Factor	Genotype	Effect	s.e.	P value
STAT1_c3141t	C:T	-5	2	**0.024**
ESR1		-6	2	
MY10	1	-1	0.3	**<0.001**
MYa	1	1	0.3	**0.007**
Parity	1	0	0	
Parity	2	4	2	**0.031**
Parity	3	1	2	

7.3.2.2 Testing SNPs for their possible additive and dominance effects

Under this model SNPs were fitted singly as factors, to account for both additive and dominance effects, and were tested with the four datasets described in Section 7.2.4.

ACTIVITY

Dataset 1 (all oestruses)

Table 7.10 shows the significant genetic effects on ACTIVITY when each SNP was fitted individually as a factor. Of the 37 SNPs studied only the W467C SNP in the *LHR* was associated with significant effects on the number of AU. A G>T allelic substitution at this position was associated with a reduction in the number of AU by 8 per 100 AU at oestrus. The results suggest an over dominance effect of this SNP on ACTIVITY, meaning that the heterozygous T:G genotype expressed the lowest number of ACTIVITY in comparison with the homozygous T:T and G:G genotypes.

Heat season, parity, and MY10 also had significant effects on ACTIVITY. The highest number of ACTIVITY was observed in season 3 and parity 1. Higher MY10 is associated with a reduction in ACTIVITY by 1 AU per kg increase in MY10.

Table 7.10 Additive and dominance effect of the SNPs on ACTIVITY.

Factor	Genotype	Effect	s.e.	P value
LHR_W467C	T:G	-8[d]	-1	**0.011**
heat_season	1	0	0	
heat_season	2	3	1	
heat_season	3	5	1	**<.001**
heat_season	4	1	1	
Parity	1	0	0	
Parity	2	-2	1	**0.003**
Parity	3	-7	2	
MY10	1	-1	0	**0.014**

- d=The dominance effect was estimated as the non-additive genetic effect or the deviation of the heterozygote from the mean of the 2 homozygotes.

Dataset 2 (visually confirmed oestruses)

This dataset generated similar results to dataset 1. The W467C SNP in the *LHR* was associated with significant non-additive effects on ACTIVITY. This effect is illustrated by a reduction in ACTIVITY at oestruses that were confirmed by visual observations by about 7 AU. Furthermore, heat season, parity, and MY10 also had significant effects on ACTIVITY (Table 7.11).

Table 7.11 Additive and dominance effects of the SNPs on ACTIVITY.

Factor	Genotype	Effect	s.e.	P value
LHR_W467C	T:G	-7[d]	-1	**0.027**
heat_season	1	0	0	
heat_season	2	3	1	
heat_season	3	5	1	**<.001**
heat_season	4	2	1	
Parity	1	0	0	
Parity	2	-2	2	**0.001**
Parity	3	-7	2	
MY10	1	-1	0.2	**0.03**

- d=The dominance effect was estimated as the non-additive genetic effect or the deviation of the heterozygote from the mean of the 2 homozygotes.

Dataset 3 (first oestrus)

The trait ACTIVITY was not affected by any of the SNPs or factors fitted using this dataset.

Dataset 4 (successful pregnancy)

When only oestruses that were followed by pregnancy were considered, the W467C SNP was again associated with a significant non-additive effect on ACTIVITY (-8 AU; Table 7.12). In addition, the SNP at position 86 of the *ACT_IIB receptor* tended to have a significant additive effect (-7 AU). MY10 was associated with a reduction in the number of AU at oestrus by 1 step per kg increase in milk yield. None of the other factors had significant effects.

Table 7.12 Additive and dominance effect of the SNPs on ACTIVITY.

Factor	Genotype	Effect	s.e.	P value
LHR_W467C	T:G	-8[d]	-1	**0.048**
ACT_IIB_86_END	G:A	-7[a]	-1	0.056
MY10	1	-1	-1	**0.03**

- d=The dominance effect was estimated as the non-additive genetic effect or the deviation of the heterozygote from the mean of the 2 homozygotes.
- a= The additive effect was estimated as the difference between the 2 homozygous means divided by 2.

ACTIVITY%

All the above mentioned analyses were repeated with the trait ACTIVITY% as follows.

Dataset 1

Using this dataset four SNPs were associated with significant effects on ACTIVITY%. The STAT1 SNP c3141t and ESR1 SNP had significant additive effects while the GnRHR T410C and LHR W467C SNPs had predominantly non-additive effects. For each SNP these effects were estimated to be -2%, +2%, +6%, and -4% respectively. There was an association between parity and ACTIVITY% with the highest ACTIVITY% value shown by animals in their second parity. Furthermore, every 1 kg increase in MY10 was associated with a decrease in ACTIVITY% of 1% at oestrus. On the other hand, every 1 kg increase in MYa was associated with an increase in ACTIVITY% by 1% (Table 7.13).

Table 7.13 Additive and dominance effects of SNPs on ACTIVITY%.

Factor	Genotype	Effect	s.e.	P value
STAT1_c3141t	C:T	-2[a]	-1	**0.005**
ESR1	C:T	2[a]	1	**0.01**
GnRHR_T410C	C:T	6[d]	-1	**0.041**
LHR_W467C	T:G	-4[d]	-1	**0.018**
Parity	1	0	0	
Parity	2	4	2	**0.048**
Parity	3	1	2	
MY10	1	-1	0.2	**<.001**
MYa	1	1	0.2	**0.003**

- a,d=additive and dominance effects were defined as given in Table 7.12

Dataset 2

The genetic analysis revealed that STAT1 SNP c3141t, ESR1 and LHR SNP W467C all had significant effects on ACTIVITY%. These effects were similar to those from dataset 1 except for the ESR1 SNP, which showed an unfavourable correlation with ACTIVITY% at visually confirmed oestruses. Furthermore, the STAT5 SNP g12195c also showed a significant non-additive genetic effect of -8%. Again, the highest value of ACTIVITY% was observed in parity 2. MY10 and MYa also showed significant effects similar to dataset 1 (Table 7.14).

Table 7.14 Additive and dominance effects of SNPs on ACTIVITY%.

Factor	Genotype	Effect	s.e.	P value
STAT1_c3141t	T:C	-6[a]	-1	**0.026**
STAT5A_g12195c	C:G	-8[d]	-1	**0.012**
ESR1	C:T	-2[a]	1	**0.014**
LHR_W467C	T:G	-4[d]	-1	**0.021**
Parity	1	0	0	
Parity	2	4	2	**0.045**
Parity	3	1	2	
MY10	1	1	0.3	**<.001**
MYa	1	1	0.3	**0.006**

- a,d=additive and dominance effects were defined as given in Table 7.12

Dataset 3

None of the SNPs had statistically significant effects on ACTIVITY% using this dataset.

Dataset 4

When only oestruses resulting in pregnancies were considered, significant allelic substitution effects on ACTIVITY% were found for STAT1 SNP c3141t, LHR SNP W467C, and FSHR T658S SNP. The effects of the SNPs in *STAT1* and *FSHR* were additive, while the effect of the SNP in *LHR* was non-additive. MY10 was associated with a reduction in ACTIVITY% of 1% per kg increase in milk yield (Table, 7.15).

Table 7.15 Additive and dominance effects of SNPs on ACTIVITY%.

Factor	Genotype	Effect	s.e.	P
STAT1_c3141t	T:C	-6[a]	-1	**0.025**
LHR_W467C	T:G	-6[d]	-1	**0.05**
FSHR_T658S	C:G	-4a	-1	**0.05**
MY10	1	-1	0.3	**0.018**

- a,d=additive and dominance effects were defined as given in Table 7.12.

As the data for oestruses confirmed by visual observation best represented locomotor ACTIVITY information, the subsequent analyses were carried out on this dataset only.

7.3.2.3 Bivariate analyses

7.3.2.3.1 ACTIVITY and ACTIVITY%

In this study, ACTIVITY and ACTIVITY% are different traits, but are genetically and phenotypically correlated. Therefore, in order to improve the accuracy of the genetic analysis, information about both traits was used in a further bivariate genetic analysis. Moreover, the additive effects of the SNPs represent the effects of the allelic substitution of the SNPs on breeding values (Falconer, 1996). Therefore from

the breeding point of view, the additive-only model, which generates the additive effects of the SNPs, is the most appropriate model when considering animals carrying these SNPs for selection. Eighteen SNPs were found to have significant additive effects ($p<0.05$) on ACTIVITY and ACTIVITY%. Of these only four SNPs were found to have had significant effects in the previous univariate additive-only model (Table 7.16).

The STAT1 SNP was associated with a reduction in both ACTIVITY and ACTIVITY% by -4 and -5% respectively. The ESR1 and GHR phe279tyr SNPs were also associated with significant negative effects on ACTIVITY and ACTIVITY%. Furthermore, the PRLR ser18asn SNP had a significant effect on ACTIVITY but not on ACTIVITY%. The estimated effects of these SNPs under this model were of similar magnitude to those estimated under the corresponding univariate analysis except for the PRLR ser18asn SNP which showed a smaller effect under this model in comparison with the univariate analysis.

The STAT5 SNP tended to have a significant effect on both ACTIVITY and ACTIVITY%, increasing both traits by 2 and 3% respectively, although this effect did not reach significance. This SNP was also found to be associated with significant dominance effects on ACTIVITY% (section 7.3.2.2).

Two of the SNPs in the *LHR* at positions W467C and Q527H were associated with reductions in ACTIVITY and ACTIVITY%. Furthermore, three SNPs in the *FSHR* had significant effects. The SNP at position 596 was associated with an increase in ACTIVITY and ACTIVITY% estimated to be about 5 and 6% respectively. The other two SNPs at positions 502 and 658 were associated with effects of +3%, -3% on ACTIVITY% but with no significant effects on ACTIVITY. Four SNPs in the *activin receptor* were associated with negative effects on ACTIVITY and ACTIVITY%. Finally, the GHR A257G and PRLR ser18asn SNPs were associated with reductions in ACTIVITY by 7. However, their effects on ACTIVITY% were not significantly different from zero.

Table 7.16 Allelic substitution effects on ACTIVITY and ACTIVITY%.

SNPs	Genotype	AU	s.e.	AU%	s.e.	P Value
STAT1_c3141t	C:T	-4	1	-5	1	**<.001**
GHR_Phe279Ty	T:A	-5	1	-2	2	**0.002**
PRL_89398_g_	G:A	-4	2	0.5	2	**0.004**
STAT5A_g1219	C:G	2	1	3	1	0.074
ESR1	T:C	-2	1	-6	2	**<.001**
bGNRHR_-331	A:G	-3	1	-4	2	**0.013**
LHR_W467C	G:T	-3	1	-3	1	**0.013**
LHR_Q527H	G:T	-4	2	-2	2	0.055
FSHR_L502L	T:C	0	2	3	2	**0.043**
FSHR_S596S	C:T	5	2	6	2	**0.009**
FSHR_T658S	C:G	0	2	-3	2	**0.039**
ACT_95	A:G	-4	1	-2	1	**0.006**
ACT_86	A:G	-4	1	-2	1	**0.005**
ACT_45	T:C	-3	1	-1	1	0.057
ACT_46	G:T	-3	1	-1	1	0.058
GHRA257G_ex10	A:G	-7	3	-1	3	**0.012**
PRLR_Ser18Asn	A:G	-7	2	1	2	**<.001**

7.3.2.3.2 ACTIVITY and BASE

To confirm the results of the previous analysis, a bivariate analysis including both ACTIVITY and BASE was carried out. 14 SNPs were found to have significant effects on both traits (Table 7.17). Allelic substitution at position c3141t in *STAT1* was associated with a reduction in ACTIVITY and BASE by 4 and 2 respectively. These results are consistent with the results of the previous analyses as this SNP had a negative effect on ACTIVITY% (Table 7.16). When the trait BASE was included in the model, the effect of the STAT5 SNP on ACTIVITY was no longer significant.

The effect of the GHR SNP at position phe279tyr was also in agreement with the results of the previous analyses. This SNP had negative effects on both ACTIVITY and BASE. The ESR1 SNP was associated with negative effects on ACTIVITY but

had no significant effects on BASE. Its negative effect on ACTIVITY contributed to its negative effect on ACTIVITY% demonstrated in the previous analysis. In addition, two of the SNPs in the *LHR* at positions W467C and Q527C had significant effects on both traits. The effects of allelic substitution at these positions were associated with negative effects on both traits. However, their effects on ACTIVITY were greater than on BASE, confirming its negative effect on ACTIVITY%.

In the *FSHR* the SNP at position S596S was associated with significant effects on both traits, increasing ACTIVITY by higher rate than BASE resulting in a positive effect on ACTIVITY%, which was demonstrated in the previous analysis.

Four SNPs in the *ACT_IIB receptor* at positions 45, 46, 86 and 95 had significant effects on ACTIVITY and BASE. Allelic substitutions at these positions were associated with reductions in both traits. In addition, the SNPs in the *GHR* and *PRLR* at positions A257G and ser18asn respectively were associated with reductions in both traits.

Table 7.17 Allelic substitution effects of the SNPs on ACTIVITY and BASE.

SNP	Genotype	AU	s.e.	BA	s.e.	P value
STAT1_c3141t	C:T	-4	1	-2	1	**0.003**
GHR_Phe279Ty	T:A	-5	1	-3	1	**<.001**
PRL_89398_g_	G:A	-4	2	-5	1	**<.001**
ESR1	T:C	-2	1	1	1	**0.014**
LHR_W467C	G:T	-3	1	-2	1	**0.026**
LHR_Q527H	G:T	-4	2	-3	2	**0.035**
FSHR_S596S	C:T	5	2	2	2	**0.05**
ACT_95	A:G	-4	1	-3	1	**0.003**
ACT_86	A:G	-4	1	-3	1	**0.003**
ACT_45	T:C	-3	1	-2	1	**0.03**
ACT_46	G:T	-3	1	-2	1	**0.041**
GHRA257G_ex10	A:G	-7	3	-6	2	**0.003**
PRLR_Ser18Asn	A:G	-6	2	-7	2	**<.001**

7.3.2.4 Heritabilities, repeatabilities and genetic correlations

The heritabilities, repeatabilities, and genetic correlations between the 3 oestrous behaviour traits, ACTIVITY, ACTIVITY% and BASE, were estimated applying the models discussed previously, but with the exclusion of the SNP effects (see section 7.2.4).

7.3.2.4.1 Univariate analyses

A single-trait mixed model was used to evaluate the effects of covariance components for individual animals, solving for both permanent environmental effects and genetic effects on ACTIVITY and ACTIVITY%. The model also included the fixed effects of oestrus number, MY10, MYa, parity and heat season as described in section 7.2.4

Preliminary, univariate analysis for ACTIVITY revealed animal, permanent environmental and residual variances of 0.0019, 0.0052 and 0.0089 respectively. The heritability and repeatability of this trait were 0.07 (s.e.=0.101) and 0.41 (s.e.=0.04) respectively. Similar analysis for ACTIVITY% resulted in animal, permanent environmental and residual variances of 0.0019, 0.00085 and 0.017 respectively. These values give estimates of the heritability and repeatability of 0.096 (s.e. = 0.07) and 0.14 (s.e. = 0.04).

MY10, MYa and parity had significant effects on both ACTIVITY and ACTIVITY%. The highest number of ACTIVITY was associated with parity 1. In addition, the number of ACTIVITY at oestrus decreased by 0.5 AU per kg increase in MY10, and by 0.3 AU per kg increase in MYa. On the other hand, the highest value of ACTIVITY% was achieved by cows within their parity 2. In addition, MY10 was associated with a reduction in ACTIVITY% by 0.9% per 1kg increase in milk, while MYa was associated with an increase of 0.7% per 1kg increase.

7.3.2.4.2 Bivariate analyses:

A series of three bivariate analyses were undertaken to estimate the genetic and phenotypic correlations between the three traits ACTIVITY, ACTIVITY% and BASE. The first bivariate analysis included ACTIVITY and ACTIVITY%, the second analysis included ACTIVITY and BASE, and the third included ACTIVITY% and BASE. The results of the associated heritabilities, genetic correlations, phenotypic correlations and repeatabilities are presented in Tables 17.18 - 7.20 respectively. The trait BASE was strongly positively correlated both phenotypically (0.7 ± 0.01) and genetically (0.82 ± 0.053) with ACTIVITY. However, the genetic and phenotypic correlations between BASE and ACTIVITY% were low (0.05 ± 0.08 and 0.05 ± 0.02 respectively). The genetic and phenotypic correlations between ACTIVITY and ACTIVITY% were moderate and estimated to be 0.35 ± 0.07 and 0.49 ± 0.02 respectively. There were no differences in the values of the heritability and the repeatability of BASE when fitted with either ACTIVITY or ACTIVITY%.

Table 7.18 Estimates of heritabilities (diagonal), phenotypic correlations (above diagonal), genetic correlations (below diagonal) and repeatabilities (*re,* bold italics) for ACTIVITY and ACTIVITY%.

	ACTIVITY	**ACTIVITY%**
ACTIVITY	0.24 ± 0.02	0.49 ± 0.017
ACTIVITY%	0.35 ± 0.07	0.16 ± 0.02
re	***0.26 ± 0.02***	***0.18 ± 0.02***

Table 7.19 Estimates of heritabilities (diagonal), phenotypic correlations (above diagonal), genetic correlations (below diagonal) and repeatabilities (*re,* bold italics) for ACTIVITY and BASE.

	ACTIVITY	**BASE**
ACTIVITY	0.19 ± 0.02	0.7 ± 0.01
BASE	0.82 ± 0.05	0.25 ± 0.02
re	***0.22 ± 0.02***	***0.28 ± 0.02***

Table 7.20. Estimates of heritabilities (diagonal), phenotypic correlations (above diagonal), genetic correlations (below diagonal) and repeatabilities (*re,* bold italics) for ACTIVITY% and BASE.

	ACTIVITY%	**BASE**
ACTIVITY%	0.16 ± 0.02	0.05 ± 0.02
BASE	0.05 ± 0.08	0.25 ± 0.02
re	***0.18 ± 0.02***	***0.27 ± 0.02***

7.3.2.4.3 Multivariate analysis (three traits together)

Basically, this model was similar to the previous bivariate analysis, but it fitted the three traits ACTIVITY, ACTIVITY% and BASE simultaneously in one multivariate analysis. This analysis allowed the estimation of the parameters relating to each trait, yielding the covariance components between the traits (Table 7.21). This included the additive genetic covariance, which then can be used to estimate the genetic correlations. In this model, BASE and ACTIVITY were strongly correlated genetically and phenotypically (0.65 ± 0.37 and 0.76 ± 0.04 respectively). On the other hand, there was no correlation between BASE and ACTIVITY%. The genetic correlation between ACTIVITY and ACTIVITY% was of similar magnitude to that derived from the bivariate analysis for both traits only. The heritabilities and repeatabilities dropped significantly with this model in comparison with the previous bivariate analysis, the highest values being for the trait BASE (estimated to be 0.107 and 0.125 respectively).

Table 7.21 Estimates of heritabilities (diagonal), phenotypic correlations (above diagonal), genetic correlations (below diagonal) and repeatabilities (*re,* bold italics) for ACTIVITY, ACTIVITY% and BASE.

	ACTIVITY	**ACTIVITY%**	**BASE**
ACTIVITY	0.09 ± 0.03	0.62 ± 0.03	0.76 ± 0.04
ACTIVITY%	0.37 ± 0.25	0.06 ± 0.013	0.06 ± 0.016
BASE	0.65 ± 0.37	-0.004 ± 0.16	0.107 ± 0.021
re	***0.105 ±0.03***	***0.08 ± 0.013***	***0.125 ± 0.021***

The above mentioned models also produced the animal effects (i.e breeding values) and the results are presented in the following section.

7.3.2.5 Breeding values

7.3.2.5.1 Univariate analysis

The means, standard deviations, and ranges of animal BVs for the three traits are presented in Table 7.22 with distributions of BVs shown in Figure 7.2-7.3. The animals' BVs for ACTIVITY, ACTIVITY% and BASE lay in the ranges -5 to 8, -14 to 14, and -3 to 4 respectively. The corresponding permanent environmental effects associated with ACTIVITY, ACTIVITY% and BASE fell in the ranges -27 to 37, -7 to 6 and -29 to 43 respectively.

Table 7.22 Descriptive statistics of the BVs of ACTIVITY, ACTIVITY% and BASE.

	ACTIVITY	**ACTIVITY%**	**BASE**
Mean	0	1	1
Standard deviation	1.7	3.1	0.8
Minimum	-5	-13	-3
Maximum	8	14	4

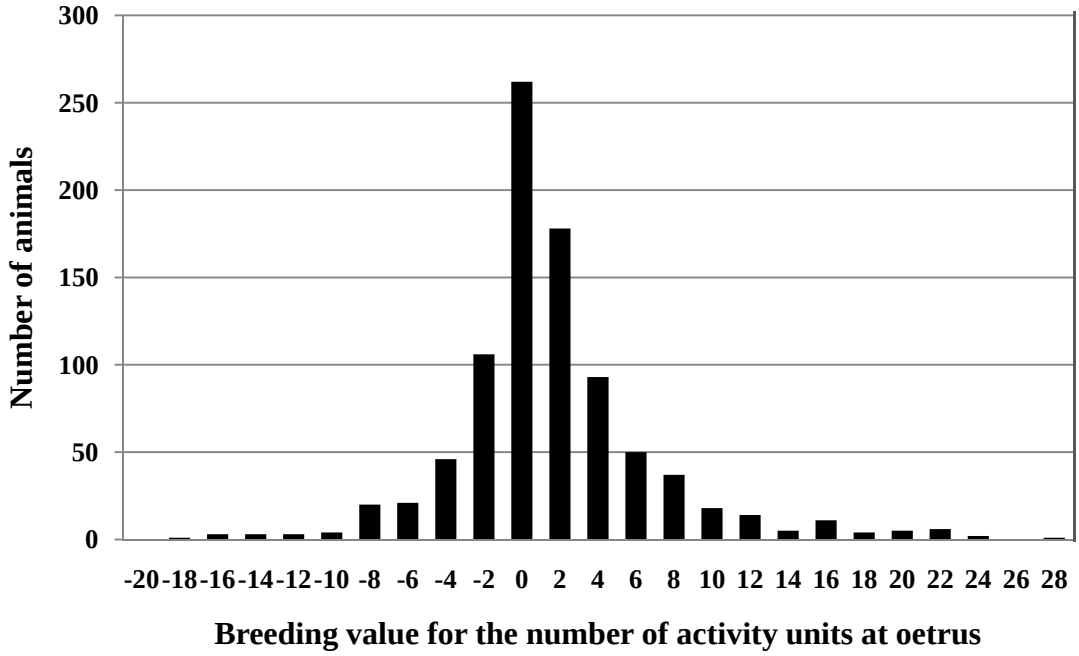


Figure 7.2 Frequency distribution of animals estimated breeding values for ACTIVITY.

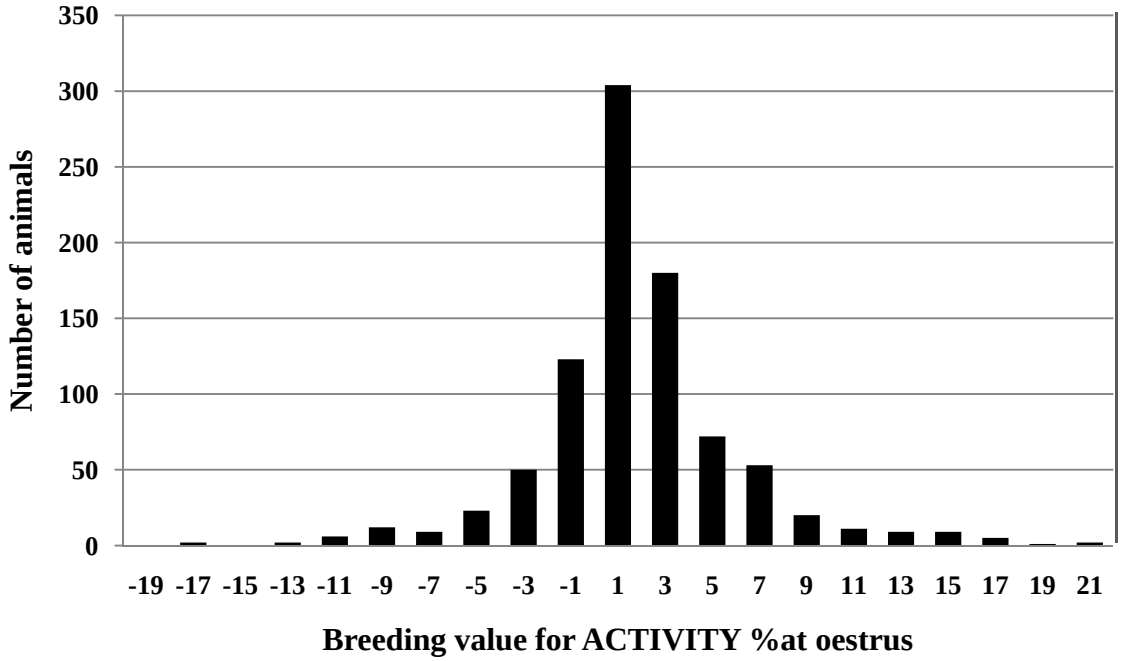


Figure 7.3 Frequency distribution of animals estimated breeding values for ACTIVITY%.

7.3.2.5.2 Bivariate analysis

The above bivariate analyses (Section 7.3.2.4.2), in addition to generating genetic and phenotypic correlations, also produced BLUP BVs for all the animals in the pedigree file. The bivariate analyses between the traits accounted for the genetic and phenotypic correlations between them and therefore improved the results of the BVs and the accuracy of their estimation. Due to the large number of animals in the database, the results will be presented here for the animals with the 10 highest and 10 lowest BVs for the traits. Each BV was adjusted to account for the animal effect per 100 AU or per 100 % increase in AU%, for example animal with BV for ACTIVITY of 23 AU will increase the number of ACTIVITY by 23 per 100 AU at oestrus, i.e. every 100 will become 123 AU at oestrus.

7.3.2.5.2.1 Bivariate analysis (ACTIVITY and ACTIVITY%)

The means, standard deviations, and ranges of BVs for the two traits are presented in Table 7.23. The full range of BVs for ACTIVITY and ACTIVITY% were 46 (-19 to 27) and 37 (-18 to 19) respectively. With this model ACTIVITY and ACTIVITY% had wider ranges and higher standard deviations compared with the univariate analysis (Table 7.22).

Table 7.23 Descriptive statistics for BVs of ACTIVITY and ACTIVITY% (bivariate analysis).

	ACTIVITY	**ACTIVITY%**
Mean	1	1
Standard deviation	5	4
Minimum	-19	-18
Maximum	27	19

The animals with the highest and lowest BVs for ACTIVITY and ACTIVITY% are given in Tables 7.24 and 7.25, respectively. In Table 7.24 some of the animals with high ACTIVITY BVs had negative ACTIVITY% BVs and this can be explained by the high BVs for BASE that were characteristic of these animals. On the other hand,

there were animals with negative ACTIVITY BVs but with positive ACTIVITY% which might be due to their negative BASE BVs. These two analyses confirm the high genetic and phenotypic correlations between ACTIVITY and BASE demonstrated earlier.

All the animals with the highest ACTIVITY% BVs had positive ACTIVITY BVs except one animal which had a negative ACTIVITY BV (Table 7.25). In this case it can speculated that despite having a high ACTIVITY% BV this animal tended to have lower ACTIVITY and BASE BVs than average, resulting in positive ACTIVITY% BVs, while retaining negative BVs for the other two traits. Moreover, some of the animals with negative ACTIVITY% BVs had positive ACTIVITY BVs and this can be explained by the high BASE BVs of these animals that exceed their BVs for ACTIVITY.

Table 7.24 Highest and lowest BVs of ACTIVITY and their corresponding ACTIVITY%.

Animal HBN	Rank of AU BVs	s.e.	AU% BVs	s.e
X110919620	27	6	11	7
UK141797700667	23	5	1	6
UK164137700274	23	6	19	6
UK141797600491	22	6	18	6
X111057295	22	5	-12	6
UK141797600666	22	6	-6	7
UK141797400405	22	7	16	7
X110919619	21	5	-3	6
UK141797400545	21	6	5	7
UK141797200802	20	6	10	6
UK141797600540	-12	7	1	7
X110841321	-12	5	-7	6
UK141797400797	-12	6	-11	6
UK200484700791	-14	6	-13	7
UK141797300845	-14	6	3	7
X101784320	-15	6	-18	6
X110820390	-16	6	4	7
UK141797300530	-16	6	1	7
X110515125	-16	7	-5	7
UK598400830	-19	6	-18	6

Table 7.25 Highest and lowest BVs of ACTIVITY% and their corresponding ACTIVITY.

Animal HBN	AU BVs	s.e.	Rank of AU% BVs	s.e.
UK164137700274	23	6	19	6
UK141797500630	15	5	19	6
UK141797600491	22	6	18	6
UK141797500847	10	7	16	7
UK141797400405	22	7	16	7
UK141797600736	11	6	16	6
UK141797200837	15	7	15	7
X110998615	14	6	15	7
X110530731	-1	6	15	7
UK141797100759	3	6	14	7
UK141797400797	-12	6	-11	6
UK141797500336	11	6	-11	6
UK141797400818	8	6	-12	7
X111057295	22	5	-12	6
UK141797300635	7	7	-13	7
UK200484700791	-14	6	-13	7
X110450345	-7	5	-13	6
UK141797600820	2	6	-14	7
UK598400830	-19	6	-18	6
X101784320	-15	6	-18	6

7.3.2.5.2.2 Bivariate analysis (ACTIVITY and BASE)

The means, standard deviations, and ranges of BVs for the two traits are presented in Table 7.26. The full ranges of the ACTIVITY and BASE BVs were 44 (-19 to 25) and 48 (-20 to 28) respectively.

Table 7.26 Descriptive statistics for BVs of ACTIVITY and BASE (bivariate analysis).

	ACTIVITY	BASE
Mean	1	1
Standard deviation	5	5
Minimum	-19	-20
Maximum	25	28

Table 7.27 shows that animals with high ACTIVITY BVs tended to have high and positive BASE BVs. In addition, animals with negative ACTIVITY BVs had negative BVs for BASE. These results confirmed the previous findings of high correlations between the two traits.

Table 7.27 Highest and lowest BVs for ACTIVITY and their corresponding BVs for BASE.

Animal HBN	Rank of AU BVs	s.e.	BA BVs	s.e.
X110919620	25	6	21	5
UK141797700667	22	5	22	4
UK164137700274	22	6	12	5
X111057295	21	5	28	4
UK141797600491	21	6	12	5
UK141797600666	21	6	25	5
X110919619	21	5	23	4
UK141797400405	20	7	12	6
UK141797400545	20	6	18	5
UK141797200802	20	6	15	5
UK141797600540	-11	6	-11	6
UK141797400797	-12	5	-7	5
X110841321	-12	5	-10	4
UK200484700791	-13	6	-9	5
UK141797300845	-13	6	-14	5
X101784320	-14	5	-8	5
UK141797300530	-15	6	-15	5
X110515125	-16	7	-14	6
X110820390	-16	6	-20	5
UK598400830	-19	6	-12	5

Table 7.28 illustrates the consistency between the BVs of ACTIVITY and BASE with animals of high BVs for BASE having high BVs for ACTIVITY. In addition, all animals with negative BVs for BASE had negative BVs for ACTIVITY.

Table 7.28 Highest and lowest BVs for BASE and their corresponding BV for ACTIVITY.

Animal HBN	AU BVs	s.e.	Rank of BA BVs	s.e.
X111057295	21	5	28	4
UK141797600666	21	6	25	5
X110919619	21	5	23	4
UK141797700667	22	5	22	4
UK200484300801	16	5	22	4
X110919620	25	6	21	5
UK141797400545	20	6	18	5
UK141797500336	10	6	16	5
X65000002234121	17	6	15	5
UK141797200802	20	6	15	5
X110057561	-9	7	-11	6
UK141797600540	-11	6	-11	6
UK141797400552	-9	5	-11	4
X110683005	-10	5	-12	4
UK598400830	-19	6	-12	5
X110361573	-11	5	-14	4
X110515125	-16	7	-14	6
UK141797300845	-13	6	-14	5
UK141797300530	-15	6	-15	5
X110820390	-16	6	-20	5

7.3.2.5.2.3 Bivariate analysis (ACTIVITY% and BASE)

The means, standard deviations, and ranges of BVs for the two traits are presented in Table 7.29. When ACTIVITY% and BASE were fitted together they have expressed similar ranges of variations to those obtained from their bivariate analyses with ACTIVITY.

Table 7.29 Descriptive statistics of BVs for ACTIVITY% and BASE (bivariate analysis).

	ACTIVITY%	BASE
Mean	1	1
Standard deviation	4	5
Minimum	-18	-20
Maximum	20	29

Table 7.30 demonstrates the weak correlation between ACTIVITY% and BASE; for example three animals had the same BV for BASE but with three different BVs for ACTIVITY%. On the other hand, there were many animals with similar BVs for ACTIVITY% but with differing BASE BVs. These results reflect the previous finding of a weak correlation between the two traits.

Table 7.30 Highest and lowest BV for ACTIVITY% and corresponding BVs for BASE.

Animal HBN	Rank of AU% BVs	s.e.	BA BVs	s.e.
UK164137700274	20	6	12	5
UK141797500630	19	6	5	4
UK141797600491	18	6	12	5
UK141797500847	17	7	1	6
UK141797400405	16	7	12	6
UK141797600736	16	6	3	5
UK141797200837	15	7	7	5
X110998615	15	7	6	5
X110530731	15	7	-8	5
UK141797200774	14	6	-2	4
UK141797400797	-11	6	-7	5
UK141797500336	-11	6	16	5
UK141797400818	-11	7	14	5
X111057295	-11	6	29	4
UK141797300635	-12	7	14	6
UK200484700791	-13	7	-8	5
X110450345	-13	6	-1	4
UK141797600820	-13	7	8	5
UK598400830	-18	6	-12	5
X101784320	-18	6	-8	5

Table 7.31 shows the negative and weak correlation between BASE and ACTIVITY% traits. Two animals with the same ACTIVITY% BVs (of -11) had highly differing BASE BVs (of 29 and 16). In addition, there were many animals with similar BASE BVs but with markedly different BVs for ACTIVITY%.

Table 7.31 Highest and lowest BV for BASE and corresponding BV for ACTIVITY%.

Animal HBN	AU % BVs	s.e.	Rank of BA BVs	s.e.
X111057295	-11	6	29	4
UK141797600666	-6	7	25	5
X110919619	-2	6	23	4
UK141797700667	1	6	22	4
UK200484300801	-10	6	22	4
X110919620	11	7	20	5
UK141797400545	6	7	17	5
UK141797500336	-11	6	16	5
X65000002234121	5	7	15	5
UK141797200802	11	6	15	5
X110057561	1	8	-11	6
UK141797600540	0	7	-11	6
UK141797400552	6	6	-11	4
UK598400830	-18	6	-12	5
X110683005	4	6	-12	4
X110361573	6	6	-14	4
X110515125	-5	7	-14	6
UK141797300845	2	7	-14	5
UK141797300530	1	7	-15	5
X110820390	4	7	-20	5

7.3.2.5.2 .4 Multivariate analysis

This model accounts for the covariance between the three traits simultaneously and further clarifies the results of the previous analyses (Table 7.32). When considering the animals with the highest ACTIVITY BVs, these animals also had high and positive BASE BVs. However, their BVs for ACTIVITY% depended on the difference between their BVs for the other two traits. If the BV for ACTIVITY was higher than the BV for BASE an animal had a positive BV for ACTIVITY%, and the size of this BV depended on the value of this difference. The same conclusion can be drawn when considering animals with the lowest values for ACTIVITY BVs.

Table 7.32 Highest and lowest BVs for ACTIVITY and their corresponding BVs for ACTIVITY% and BASE.

Animal HBN	Rank of AU BVs	s.e.	AU % BVs	s.e.	BA BVs	s.e.
X110919620	23	5	11	6	18	4
UK164137700274	21	5	19	6	11	4
UK141797600491	20	5	18	6	10	4
UK141797700667	19	5	1	5	19	4
UK141797400405	19	6	16	7	10	5
X110919619	19	5	-2	5	21	4
X111079422	18	5	13	5	12	4
X111057295	18	5	-11	5	25	4
UK141797200802	18	5	10	6	13	4
UK141797400545	18	5	6	6	15	4
UK141797600540	-10	6	0	7	-10	5
X110841321	-11	5	-7	5	-9	4
UK141797400797	-11	5	-11	6	-6	4
UK141797300845	-12	5	2	6	-13	4
UK200484700791	-13	5	-13	6	-7	4
X110820390	-14	5	4	6	-18	4
UK141797300530	-14	5	1	6	-14	4
X110515125	-14	6	-5	7	-12	5
X101784320	-14	5	-18	6	-7	4
UK598400830	-18	5	-18	6	-11	4

Table 7.33 shows that animals with high ACTIVITY% BVs had higher BVs for ACTIVITY than the BASE BVs. On the other hand, the lowest ACTIVITY% BVs of the animals were associated with higher BASE BVs than ACTIVITY BVs.

Table 7.33 Highest and lowest BV for ACTIVITY% and their corresponding BVs for ACTIVITY and BASE.

Animal HBN	AU BVs	s.e.	Rank of AU % BVs	s.e.	BA BVs	s.e.
UK164137700274	21	5	19	6	11	4
UK141797500630	14	5	19	5	5	4
UK141797600491	20	5	18	6	10	4
UK141797500847	10	6	16	7	1	5
UK141797400405	19	6	16	7	10	5
UK141797600736	10	5	15	6	3	4
UK141797200837	14	6	15	7	6	5
X110998615	13	6	15	6	5	5
X110530731	0	5	15	6	-7	4
UK141797200774	5	5	14	5	-1	4
UK141797400797	-11	5	-11	6	-6	4
UK141797500336	9	5	-11	6	14	4
UK141797400818	6	5	-11	6	12	4
X111057295	18	5	-11	5	25	4
UK141797300635	5	6	-12	7	12	5
UK200484700791	-13	5	-13	6	-7	4
X110450345	-7	5	-13	5	-1	4
UK141797600820	1	5	-13	6	7	4
UK598400830	-18	5	-18	6	-11	4
X101784320	-14	5	-18	6	-7	4

Table 7.34 indicates that animals with the highest BVs for BASE tended to have high BVs for ACTIVITY, resulting in negative BVs for ACTIVITY%. Furthermore, there were animals with high and positive BVs for the three traits together. On the other hand, some animals had negative BVs for the three traits.

Table 7.34 Highest and lowest BVs for BASE and their corresponding BVs for ACTIVITY% and ACTIVITY.

Animal HBN	AU BVs	s.e.	AU % BVs	s.e.	Rank of BA BVs	s.e.
X111057295	18	5	-11	5	25	4
UK141797600666	18	5	-5	6	21	4
X110919619	19	5	-2	5	21	4
UK200484300801	14	5	-10	5	19	4
UK141797700667	19	5	1	5	19	4
X110919620	23	5	11	6	18	4
UK141797400545	18	5	6	6	15	4
UK141797500336	9	5	-11	6	14	4
X65000002234121	16	6	5	6	14	5
UK141797200802	18	5	10	6	13	4
UK141797600540	-10	6	0	7	-10	5
X110057561	-8	7	1	7	-10	5
UK141797400552	-8	5	6	5	-10	4
UK598400830	-18	5	-18	6	-11	4
X110683005	-9	5	4	5	-11	4
X110515125	-14	6	-5	7	-12	5
X110361573	-10	5	6	5	-12	4
UK141797300845	-12	5	2	6	-13	4
UK141797300530	-14	5	1	6	-14	4
X110820390	-14	5	4	6	-18	4

7.4 Discussion

The limited number of cows available was a major problem with this study. The dataset included a total of 189 recorded cows and 65 sires. Even after taking into account the availability of pedigree information, the number of genetic relationships was low, with an average of ~2.9 daughters per sire. This deficiency in genetic relationships means that there isn't sufficient data to allow a reliable estimate of the genetic parameters and the breeding values (this has been dealt with in more details in the following chapter; see section 8.4). However, the availability of repeated measurements and the application of bivariate analyses (facilitated by the moderate to high genetic and phenotypic correlations between different oestrous-derived traits) might have improved the reliability of the estimates. Thus, the results of this study although not precise, give indications of the estimated parameters.

The first aim of this study was to investigate the effects of different SNPs on three locomotor traits associated with oestrous behaviour (BASE, ACTIVITY and ACTIVITY%). 37 SNPs were analysed across 16 genes. The results identified certain significant effects on locomotor traits, but the significance of these effects depended on the allele substitution effect models fitted to different datasets, and on the use of pedigree information. Preliminary analysis, fitting the SNPs as variates, revealed that the GHR_phe279tyr and PRLR_ser18asn SNPs were associated with significant negative effects on ACTIVITY. In addition, STAT1_c3141t and ESR1 were found to have significant negative effects on ACTIVITY%.

Subsequently a series of analyses were undertaken applying increasingly stringent models. Four different datasets were used in which the SNPs were treated as factors, thereby accounting for both additive and dominance effects. In this case, the additive regression coefficients represented half the difference between the homozygotes while the dominance coefficient was the associated deviation of the heterozygote from the average of the two homozygotes (Falconer, 1996). Many published studies have estimated the additive effects of SNPs, but fewer have quantified their dominance effects. In this study, statistically significant dominance effects were associated with different SNPs. It was found that some of the SNPs with significant effects under the additive-only model had become not significant. On the other hand, other SNPs had significant effects, some which were non-additive. The different analyses showed that the LHR_W467C SNP was the only SNP that was consistently associated with dominance genetic effects on ACTIVITY. The effect of this SNP was -8 per 100 AU. Furthermore, two SNPs were found to affect ACTIVITY% in the four models applied. The STAT1_c3141t SNP was associated with an additive effect on ACTIVITY% that varied between -2 and -6%. The LHR_W467C SNP was also found to have a significant dominance effect on ACTIVITY% with animals inheriting the G allele tending to show lower values for ACTIVITY% by between 4 and 6%.

It is only when the additive effect of the SNPs is considered in the analysis, that the estimated effect reflects the effect of the allelic substitution on breeding value. In addition dominance effects were always considered to be less important in the genetic evaluations. Therefore, from the breeding point of view, the effects of SNPs

derived from additive-only models are most relevant in considering the value of animals carrying these SNPs for selection (Mrode, 2005). To further improve the accuracy of the evaluation, only the additive effects of the SNPs were tested in two further bivariate analyses that included ACTIVITY and ACTIVITY%, and subsequently ACTIVITY and BASE. In these analyses, more SNPs had significant effects on the traits of interest (Table 7.16). This can be explained by the availability to these models of additional information, notably the genetic and phenotypic correlations between the traits.

The SNP in *ESR1* was associated with significant unfavourable effects on both ACTIVITY and ACTIVITY%. However, when considering the univariate analysis, this SNP had no effect on ACTIVITY. Oestrogen plays an important role in the regulation of the behavioural changes associated with oestrus (Salisbury et al., 1978; see section 1.5.4). Therefore, allelic substitution in this gene might be expected to have a significant effect on the expression of oestrus. In the previous study (Chapter 6), this SNP was demonstrated to have significant effects on both STEPS and STEPS% although the magnitude of these effects was higher than that reported in this study. In comparison with the previous study, the dataset is larger in this study, therefore the effect of the ESR1 SNP is considered to be in the range of -2 to -6% on ACTIVITY%.

The LHR SNP at position 467 was associated with unfavourable effects on fertility, as illustrated in the negative effects on oestrous expression. The *LHR* plays a key role in the ovaries and its expression in the granulosa cells is critical to the establishment and maintenance of the dominant follicles (Xu et al., 1995). In addition, the activation of the LHR through binding to its ligand LH stimulates the production of androgen, oestrogen and progesterone (Themmen & Huhtaniemi, 2000). However, the results of this study are inconsistent with the reported beneficial effects of three SNPs in the *LHR* at positions 467, 490 and 527 which were associated with significant favourable effects on fertility traits in dairy cattle (Hastings et al., 2006). The FSHR SNP at position 596 was associated with stronger expression of oestrus. In the ovaries, FSH stimulates follicular maturation and oestrogen biosynthesis in granulosa cells (Simoni et al., 1997). The results of this study are in disagreement with the effects of the allelic substitutions identified in the

FSHR which were associated with unfavourable effects on fertility traits. The FSHR SNP at position 596 did not previously show any association with PTAs for fertility traits (Chapter 5).

Activin is implicated in a variety of reproductive processes, with roles in granulosa cells including the potentiation of FSH action, differentiation and steroidogenesis (Knight, 1996). Four SNPs in the *activin receptor* were associated with reductions in the strength of oestrous expression.

GnRH plays a key role in the oestrous cycle through its direct effects on release of LH and FSH. Only one SNP in the *GnRHR* at position -331 was associated with a significant effect on oestrous expression in the current study. Allelic substitution at this position was associated with a reduction in ACTIVITY. This effect was consistent with the reported negative effects of the GnRHR SNPs on oestrous behaviour reported in the previous Chapter. Four of the SNPs in the *GnRHR* at positions -108, 260, 341, and 410 had significant unfavourable effects on ACTIVITY and ACTIVITY%. However, these SNPs were previously found to be positively correlated with fertility traits (Chapter 4).

The GHR SNP at position 279 was associated with a reduction in both ACTIVITY and ACTIVITY% at oestrus. In addition, two SNPs in *prolactin* and its receptor (*PRLR*) had significant effects on oestrous traits. The SNP in *PRL* was associated with negative effects on ACTIVITY and ACTIVITY%, while the SNP in the *PRLR* was associated with a negative effect on ACTIVITY. Its effect on ACTIVITY% was positive, which can be explained by its negative effect on BASE. During lactation, prolactin reduces the level of ovarian steroid synthesis therefore polymorphisms in the *prolactin* gene or its receptor might reduce circulating oestradiol levels and subsequently reduce the expression of oestrus. The two SNPs in the *PRLR* and *GHR* are associated with favourable effects on milk yield traits in dairy cattle (Viitala et al., 2006).

STAT proteins are involved in cytokine signalling pathways, and regulate gene transcription in response to cytokines and growth factors (i.e. act as signalling molecules). In addition, STAT5A is involved in lactation and the establishment of

pregnancy. In this study a SNP in the *STAT5* gene was associated with favourable effects on both ACTIVITY and ACTIVITY%. This SNP was reported to be associated with a decrease in milk protein and fat percentage as well as lower embryonic survival (Khatib et al., 2008). The SNP in *STAT1* also had significant effects on oestrous expression. Allelic substitution at this position was associated with weaker oestrous expression. This is in agreement with the findings of another study which demonstrated that this SNP is associated with a significant increase in milk fat and protein percentage and therefore negative effect on fertility (Cobanoglu et al., 2006).

In the present study SNP effects were corrected for MY10, MYa, parity, heat season, and oestrus number. This eliminates the possibility that the significant effects associated with some of these SNPs were actually on milk yield rather than on oestrous behaviour. Of these factors, heat season, parity, and MY10 had significant effects on ACTIVITY under the different models applied. The highest number of ACTIVITY was observed during the summer (July to September). During the summer cows showed on average 11 AU more than during the winter (January to March). This is in contrast to the study of López-Gatius et al. (2005), who found stronger oestrous expression during cool weather (October-April). In addition, it has been reported that conception rate in dairy cattle tends to be lower during summer than in winter, and increased conception rates occur after turning cows out during the spring (Salisbury et al., 1978). One possibility is that the animals were relatively unstressed by ambient temperature in the current study, and as a result fertility was not negatively affected during the summer months. Moreover, in the current study, cows were housed throughout the year, and so were not directly exposed to radiant sunlight. Another reason might be the climatic differences between the areas where the studies were conducted.

Oestrous expression was also affected by parity. The activity of primiparous cows was approximately 6 AU higher than in multiparous cows. This is in agreement with studies that found oestrous strength measured by electronic activity tags is highest in primiparous cows (Roelofs et al., 2005b). In addition, primiparous cows were also reported to express stronger behavioural signs than multiparous cows (Roelofs et al., 2005a).

Several studies have demonstrated a negative correlation between milk yield and fertility in dairy cattle (Wall, 2003; Royal et al., 2002b; 2002a). In addition, Lóvendahl et al. (2010) have described a negative correlation between BCS and the onset of cyclic activity, as illustrated by the days from calving to first episode of high activity. In the present study, milk yield 10 days before oestrus was negatively associated with locomotor activity at oestrus. Each 1kg increase in milk yield was associated with a decrease of 1AU at oestrus. These results are in agreement with the findings of a previous study demonstrating the negative relationship between milk yield and the strength of oestrus (López-Gatius et al., 2005; Yaniz et al (2006). Harrison et al (1990) have also reported the antagonistic relationship between high milk yield and the expression of oestrus. Furthermore, López-Gatius et al (2004) demonstrated a reduced serum oestradiol level on the day of oestrus in high producing cows. A lower level of oestradiol at oestrus could therefore explain the lower activity at oestrus associated with higher milk yield found in this study, as oestrogen is the main hormone inducing oestrous expression. In contrast, average milk yield since calving (MYa) was associated with an increase in ACTIVITY% of 1% per 1kg increase. As this trait is derived from both ACTIVITY and BASE, it can be speculated that MYa is negatively associated with BASE. The genetic analysis revealed that, although the effect was not statistically significant, there was a tendency for MYa to be associated with lower values of BASE and, unexpectedly, with higher values of ACTIVITY.

However, the antagonistic relationship between oestrous expression and milk yield was expressed at the phenotypic level (milk yield being fitted as a fixed effect). This might affect the reliability of the results of this study in terms of both SNP effects and genetic parameter estimation, as some of the detected effects of the SNPs on oestrous traits might be confounded with their effects on milk yield. To clarify this confusion further analysis was carried out in which the effect of milk yield on oestrous expression was considered at the genetic level (fitting milk yield in bivariate analysis). As the dataset included repeated measurements per cow (1-15 records), considering all oestruses will result in the effect of milk yield being confounded with the high range of days in milk (8-638) and this will further increase within cow variation for milk yield (i.e. through variations in lactation curves). Therefore, only the first three oestruses were considered for this analysis, and this has further

contributed to the problem of a small dataset as a substantial number of records have then been excluded. The results further confirmed the negative phenotypic correlation of (-0.126) and the negative genetic correlation of (-0.094) between milk yield and oestrous expression. However, fewer SNPs effects were detected. This might be due either to the smaller datasets or to the previous concern that the effects of the SNPs were actually due to their effects on milk yield rather than on oestrous expression. Therefore, more data, together with different experimental designs that include only oestruses within the first two months of location (as this is the critical period in term of deleterious effect of milk yield on resumption of ovarian activity after calving) are essential to elucidate the findings of this study.

The second aim of this study was to estimate the genetic and phenotypic variations associated with oestrous behaviour traits. These parameters were estimated under the different models discussed above without including any SNP effects. In the univariate analysis, both ACTIVITY and ACTIVITY% had low heritabilities of 0.07 and 0.10 respectively, with repeatability estimates of 0.41 and 0.14 respectively. When bivariate models were fitted, estimates of the heritability of ACTIVITY ranged from 0.19 to 0.24, while the estimates of the heritabilities of both ACTIVITY% and BASE were similar under the different bivariate models, and were estimated to be 0.16 and 0.25 respectively. In addition, the trait BASE was strongly correlated both phenotypically and genetically with ACTIVITY, while its genetic and phenotypic correlations with ACTIVITY% were low. Furthermore, the genetic and phenotypic correlations between ACTIVITY and ACTIVITY% were moderate. There were no differences in the heritability and the repeatability estimates of BASE when fitted with either ACTIVITY or ACTIVITY%. On the other hand, ACTIVITY and ACTIVITY% had lower heritabilities when fitted with BASE than when fitted together in the model, suggesting that part of the variation associated with ACTIVITY and ACTIVITY% reflected variation between cows in the base number of activity. The repeatabilities for ACTIVITY, ACTIVITY% and BASE in the different univariate analyses fell in the ranges of 0.22 to 0.26, 0.18, and 0.27 to 0.28 respectively showing that these traits are markedly influenced by genetic mechanisms.

When the three traits were fitted in one multivariate analysis, the results of the previous bivariate analyses were confirmed. However, the heritabilities and repeatabilities for the oestrous traits dropped significantly. In this model the highest values for h^2 and *re* were observed with the trait BASE, and these were estimated to be 0.11 and 0.125 respectively. The accuracy of estimates produced by multivariate BLUP (MBLUP) depends on the difference in the genetic and phenotypic correlations between the different traits. In this analysis, there were no genetic and phenotypic correlations between ACTIVITY% and BASE, and this was also the case in the bivariate analysis that included both traits. Therefore, there was little to be gained from including ACTIVITY% and BASE together in one analysis. On the other hand, the genetic and phenotypic correlations between ACTIVITY and ACTIVITY% were moderate (0.35 and 0.49 respectively). Therefore, including both of these traits in one analysis will improve the accuracy particularity for ACTIVITY% which has a lower heritability than ACTIVITY. The same argument applies to ACTIVITY and BASE which also had strong genetic and phenotypic correlations (0.82 and 0.7 respectively).

The high genetic and phenotypic correlations of ACTIVITY% and BASE with ACTIVITY suggests that cows that are normally more active tend to show higher numbers of ACTIVITY at oestrus in comparison with less normally active cows. On the other hand, the low genetic and phenotypic correlations between ACTIVITY% and BASE means that the variation associated with ACTIVITY% does not reflect variation in BASE. Therefore, as ACTIVITY% is derived from both ACTIVITY and BASE, the values of ACTIVITY% that cows expressed is more closely linked to their number of ACTIVITY at oestrus than to their base number of ACTIVITY before oestrus. For example, there were many cows that have similar values of BASE but different values for ACTIVITY%.

Initially, the phenotypic and genetic variances associated with each trait in the univariate BLUP were used to estimate BVs for oestrous traits. Subsequently, in order to improve the accuracy of the BV estimates, a series of bivaraite analyses were applied making use of the genetic and phenotypic correlation between the traits. Under MBLUP, BVs had wider ranges of variations and wider ranges of values compared with the univariate analyses. In addition, the estimates of the BLUP

BVs confirm the high genetic and phenotypic correlations between ACTIVITY and BASE. From these results it can be concluded that ACTIVITY and ACTIVITY% are different but related traits as animals with the highest BVs for ACTIVITY are not necessarily those with the highest BVs for ACTIVITY%. Furthermore, the results of the BVs demonstrated the weak correlation between ACTIVITY% and BASE. For example, three animals had the same BVs for BASE, but different BVs for ACTIVITY%. On the other hand there were many animals with similar BVs for ACTIVITY% but with very different BVs for BASE.

We conclude therefore that animals that are normally active (high BASE BVs) tend to express high numbers of ACTIVITY at oestrus, and this resulted in a low value for ACTIVITY%. Furthermore, there are animals that expressed high and positive BVs for the three traits together. These animals are normally active (high BASE value) and tend to show high numbers of AU at oestrus (high ACTIVITY) resulting in a high value for ACTIVITY%. On the other hand, all the animals with the lowest values for BASE BVs had negative ACTIVITY BVs, resulting in positive ACTIVITY% BVs. Some animals had negative BVs for all three traits. These cows were normally less active than other cows and the number of AU at oestrus, although indicating oestrus, was still low, resulting in low value of ACTIVITY%. Depending on the weak correlations both genetically and phenotypically between ACTIVITY% and BASE, the BVs estimated from any model including both traits will be less accurate than those including either ACTIVITY and BASE or ACTIVITY and ACTIVITY%, which are associated with strong and moderate correlations respectively. In addition, running the bivariate analyses has the advantage of selecting animals that have favourable BVs for all of the traits. For example, the best expression of oestrous behaviour is demonstrated by cows that have attained high BVs for ACTIVITY but low BVs for BASE or high BVs for both ACTIVITY and ACTIVITY%.

Based on the strong genetic correlations between oestrous traits it is therefore recommended that a repeatability animal model should be adopted, using the parameters estimated in the current study, for estimating breeding values for oestrous traits.

Chapter 8: General discussion

1.8 Introduction

With the exception of traits reflecting physiological processes (e.g. milk progesterone patterns) fertility traits in dairy cattle currently recognized in the UK generally have low heritabilities (Wall et al., 2003). Therefore, one aim of this study was to identify molecular markers associated with improved reproductive performance, in order to make available to the industry information on beneficial alleles that can be used for improving fertility in breeding programmes. Molecular markers for fertility have already been identified in the bovine genome (Hastings et al., 2006; Khatib et al., 2008). Furthermore, although the use of electronic tags in oestrous detection is already widespread, information about their possible use in genetic improvement is yet to be established. Therefore, one of the main aims of this project was to estimate the genetic parameters associated with oestrous-derived traits. All the aims of this work were achieved and the principal outcomes will be summarized and their impacts on fertility in dairy cattle will be highlighted.

8.2 Molecular markers for fertility

Association studies were carried out between SNPs in various candidate genes well known to play important roles in reproduction, and a number of fertility traits (CI, DFS, NR56 and CINS). This project has identified one SNP in the *GnRHR* to be associated with a reduction of approximately 0.4 days in the PTAs for DFS (Chapter 4). This effect is relatively large considering the population variances in DFS PTAs in the UK fertility index (with 95% of DFS PTAs are within a 7-day range; Wall et al., 2003). Furthermore, four SNPs in the *FSHR* were found to be associated with significant unfavourable effects on fertility PTAs (increasing CI, DFS, CINS, and decreasing NR56; Chapter 5). The findings of this project can support the development of breeding programs that use molecular information for the genetic evaluation of animals for better fertility. However, the results were based on only 408 animals (Chapter 4) and 427 animals (Chapter 5), meaning that they should be validated with independent studies. In addition, it was not possible from this study to

confirm that the SNPs with significant effects are the causal effects as they might be in linkage disequilibrium with mutations in other genes.

Furthermore, one aspect of poor fertility in dairy cattle is the low rate of oestrous detection (Dobson et al., 2008; Yaniz et al., 2005; Lyimo et al., 1999). Therefore, other measures of fertility STEPS and STEPS% or ACTIVITY and ACTIVITY%, reflecting strength of oestrous expression, were studied as part of this work. Association studies were also carried out to find molecular markers for better oestrous expression, and this approach was tested in two separate datasets. Many SNPs were found to be associated with oestrous expression.

The SNP in the *ESR1* was associated with significant unfavourable effects on both oestrous behaviour traits. These effects were consistent in both oestrous datasets. Four SNPs in the *activin receptor* at nucleotide positions 45, 46, 86, and 95 were also associated with reductions in the strength of oestrous expression. However these SNPs did not show any significant effects when analysed with dataset 1. Furthermore, four of the SNPs in the *GnRHR* at positions -108, 260, 341, and 410 were associated with significant and unfavourable effects on oestrous expression (chapter 6). However, these results were in disagreement with those reported in chapter 4 (beneficial effects on fertility. On the other hand, these SNPs did not show any association with oestrous expression when analysed with the second oestrous dataset.

The FSHR SNP at position 596 was associated with stronger expression of oestrus. These results are in disagreement with the findings of unfavourable effects of the SNPs in the *FSHR* on fertility traits in addition to that the SNP at position 596 did not show any association with fertility PTAs (chapter 5). The LHR SNP at position 467 was associated with unfavourable effects on fertility, as illustrated in the negative effects on oestrous expression (dataset 2). In contrast, none of the SNPs in the *LHR* had any significant effects on oestrous strength in dataset 1. These results are inconsistent with the reported beneficial effects of the allelic substitutions in the *LHR* on fertility traits PTAs in dairy cattle (Hastings et al., 2006). These findings

were also confirmed by subsequent analysis which did not include pedigree information (Garnsworthy, personal communication).

The inconsistency between analyses of PTA and oestrous behaviour associations was surprising, since many of the SNPs with favourable effects on fertility PTAs had unfavourable effects on oestrous behaviour traits and vice versa. This inconsistency might be due to different groups of animals being used in the analyses, although this seems unlikely. Smaller numbers of animals with oestrous records were genotyped for these SNPs raising the need for more data in order to either confirm or reject these results. Another possibility is that some of these SNPs might be in linkage disequilibrium with SNPs in other genes which might have conflicting effects on fertility PTAs and oestrous behaviour traits. Furthermore, these SNPs might also have different effects at different stages of oestrus. There is evidence of potential conflict between effects on the resumption of ovarian activity postpartum and production of good quality oocytes, insulin being an example of an agent with contradictory effects: it acts as a stimulator of early ovarian cyclicity but has detrimental effects on oocyte developmental competence (Garnsworthy et al., 2008; 2009). In addition, there is the possibility of an abnormal pattern of feedback between oestradiol and gonadotrophins, such as applies in the case of ovarian cysts where a high level of oestradiol is associated with a low level of GnRH (Kesler & Garverick, 1982).

When interpreting the above mentioned results we must take into account that fertility is one of the most difficult and complex traits due, to its low heritability caused by the polygenic nature of reproductive traits and the predominant effects of environment on reproduction, and the long generation intervals in cattle. The impact of molecular markers studies on selection schemes can be considered in the context of the response to selection *R* which is expressed as:

$$R = ir_{ay}\sigma_a$$

where: i is the intensity of selection, r is the accuracy of the estimated breeding value, σ_a is the genetic standard deviation (see section 2.2.2).

Therefore, information from genetic markers will increase genetic gain primarily through increasing the accuracy of selection. This is particularly important in traits with low heritability, sex-limited or in traits that are measured late in life, such as reproductive traits (Dekkers, 2004).

The SNPs considered in the different analyses are listed in Table 8.1.

Table 8.1 Summary list of the SNPs analysed for their possible effects on fertility traits and behaviour-based oestrous traits.

SNPs	Mutation	Fertility PTAs				Oestrous data 2		Oestrous data 2		
		CI	DFS	NR56	CINS	STEPS	STEP	ACTIVITY	ACTIVITY%	BASE
STAT1_c3141t	T>C							**NF**	**NF**	**F**
GHR_Phe279Tyr	T>A							**NF**	**NF**	**F**
PRL_89398_g_a_R	G>A							**NF**	**F**	**F**
STAT5A_g12195c	G>C							**F**	**F**	NS
ESR1						**NF**	**NF**	**NF**	**NF**	**NF**
bGNRHR_-331	A>G	NS	NS	NS	NS	NS	NS	**NF**	**NF**	NS
bGNRHR_-108	T>C	**F**	**F**	NS	NS	**NF**	**NF**	NS	NS	NS
bGNRHR_206	G>A	NS	NS	NS	NS	NS	NS	NS	NS	NS
bGNRHR_260	C>T	**F**	**F**	NS	NS	**NF**	**NF**	NS	NS	NS
bGnRHR_341	C>T	**F**	**F**	NS	NS	**NF**	**NF**	NS	NS	NS
bGNRHR_383	C>T	NS	NS	NS	NS	NS	NS	NS	NS	NS
bGNRHR_410	C>T	**F**	**F**	NS	NS	**NF**	**NF**	NS	NS	NS
bERA_prom_SNP173	G>A							NS	NS	NS
LHR_W467C	G>T	**F**	**F**	NS	NS	NS	NS	NS	NS	NS
LHR_L490L	C>T	**F**	**F**	NS	NS	NS	NS	NS	NS	NS
LHR_Q527H	G>T	**F**	**F**	NS	NS	NS	NS	NS	NS	NS
NPYRY2	G>A							NS	NS	NS

FSHR_L502L	C>T	**NF**	**NF**	**NF**	NS	NS	NS	NF	F	NS
FSHR_S596S	C>T	NS	NS	NS	NS	NS	NS	**F**	**F**	NS
FSHR_T658S	C>G	**NF**	**NF**	**NF**	NS	NS	NS	NS	**NF**	NS
FSHR_N669N	C>T	**NF**	**NF**	**NF**	NS	NS	NS	NS	NS	NS
FSHR_T685T	C>A	**NF**	**NF**	**NF**	NS	NS	NS	NS	NS	NS
ACT_IIB_95	G>A					NS	NS	**NF**	**NF**	**NF**
ACT_IIB_503	C>T					NS	NS	NS	NS	NS
ACT_IIB_86_END	G>A					NS	NS	**NF**	**NF**	**NF**
ACT_IIB_45	C>T					NS	NS	**NF**	**NF**	**NF**
ACT_IIB_46	T>G					NS	NS	**NF**	**NF**	**NF**
GHRA257G_ex10	A>G							NS	NS	NS
leptin963	T>C					NS	NS	NS	NS	NS
npy_ex1	T>C					**NF**	**NF**	NS	NS	NS
bERB_ex7	C>G					NS	NS	NS	NS	NS
leptin promoter 1457	G>A					NS	NS	NS	NS	NS
PRLR_Ser18Asn	G>A							**NF**	NS	**NF**
ESR1_ex1_A503C	C>A					**NF**	**NF**	**NF**	**NF**	**NF**

- NS=Not significant; F= favourable; NF= not favourable; Blank = where data not available.

8.3 Genetic parameters and breeding values for oestrous traits

One of the main outcomes of this project was the estimation of the genetic and phenotypic parameters and the breeding values associated with oestrous-derived traits. The findings confirm the genetic basis of these traits and prove that the expression of oestrus is amenable to genetic improvement. The heritabilities of oestrous traits were found to be moderate (0.16 and 0.20 for ACTIVITY and ACTIVITY% respectively). These values are comparable to the heritabilities of other physiological measures of fertility (for instance for CLA h^2 ranges between 0.16 and 0.23; Royal et al., 2002b). However, in a different study the heritabilities of oestrous characteristics were found to be lower, and were estimated for oestrus duration to be between 0.02 and 0.08 and for the strength of oestrus to range between 0.04 and 0.06 (Løvendahl and Chagunda, 2009). On the other hand, the heritability of oestrous traits in the present work is comparable with the heritability of another trait derived from electronic tags, days to first oestrus, which was estimated to be between 0.12 and 0.18 (Løvendahl and Chagunda, 2009).

Fullwood and Lely data obtained by fitting the tags on legs and necks respectively are not fully comparable, and therefore the data of Chapter 6 could not be analysed together with those of Chapter 7, and hence separate analyses had to be carried out. Due to the small dataset and the poor pedigree information in Chapter 6, BVs for oestrous strength were not estimable with the Fullwood dataset. With Lely records, the availability of more records in the dataset together with repeated measurements and a well-structured pedigree allowed the application of an animal repeatability model in order to estimate the BVs. Furthermore, due to the established phenotypic and genetic correlations between various behaviour-based oestrous traits, bivariate genetic analyses were also applicable and this partially helped in overcoming the problem of the small datasets frequently available for genetic analysis of physiological data. It would be beneficial to combine these two datasets in one genetic analysis, to obtain more accurate results. It should be recognized that although in this study BV was used to express the additive genetic value of an individual, many studies of genetic evaluations (e.g. the UK fertility index) use predicted transmitting ability (PTA) rather than the BV, where PTA is one-half of the predicted breeding value.

8.4 Validation of the results

As the data used in this project was not originally designed for the purpose of genetic evaluation, the major problem with both datasets (Chapters 6 and 7) was the lack of data. For these results to be meaningful and to be of value for breeding programmes, the precision by which the genetic parameters and the heritabilities were estimated have to be evaluated. Precision of the estimate is indicated to by the standard error which is equal to the square root of sampling variance. In general, when an experiment aims to estimate heritability, the design of the experiment must be chosen such that the highest possible accuracy will be obtained given the limitation imposed by the scale of the experiment (Falconer, 1960). The accuracy of the estimate depends on its sampling variance, the lower the sampling variance the greater the accuracy (Robertson, 1959).

The total number of animals and the number of individuals per family that must be recorded to achieve the optimal accuracy in the estimation of genetic parameters will be discussed in relation to half sib analysis (which is the primary structure of the relationship between cows in this study). Let *N* be the number of half sib families, and *n* be the number of individuals per family giving a total number of recorded individuals $T = nN$. Based on the equation introduced by Hill in 1971:

$$T = \frac{v}{100V(h^2)}$$

where: $V(h^2)$ is the sampling variance of the heritability, and v is a tabulated coefficient provided by Hill (1971).

The total number of cows that must be recorded for oestrous expression in order to obtain the optimal precision of a heritability estimate (minimum sampling variance) can be calculated on this basis. This will be discussed in terms of two values of the heritability: 0.4 and 0.1 (as the value of v in Hill's (1971) equation was given only in relation to these two values of the heritability). Thus, for example, in the case of a trait with a heritability of 0.4 and a small standard error of 10 percent of the mean = 0.04, the total number that must be recorded can be given as:

$$T = \frac{1037}{100(0.04)^2} = 6481$$

While for a trait with a heritability of 0.1 and standard error of 0.01 the total number will be:

$$T = \frac{304}{100(0.01)^2} = 30400$$

Therefore, for a given trait, a total number of ~ 6000 and 30.000 recorded individuals will produce the minimum sampling variance in estimating heritability with an expected value of 0.4 and 0.1 respectively. As the heritability of oestrus traits is expected to be within the range of 0.2 - 0.3 (as these are physiological traits), a total number of around 10.000 recorded animals might be satisfactory in term of obtaining heritability estimates with minimum standard error. However, in practice this number of cows might not be available in experimental herds for such kind of analysis (given that oestrous records are difficult and expensive to collect). Therefore, by recording around 2000 cows for a trait of heritability of 0.1 the sampling variance for the heritability will be:

$$V(h^2) = \frac{v}{100T} = \frac{304}{100(2000)} = 0.0015$$

thus the standard error is equal to 0.04 and this will give an estimate of the heritability which is significantly different from zero (for a trait of a heritability of 0.2 the s.e. will be even smaller considering this number of cows).

In addition to the total number of individuals that must be recorded, one needs to take into account the optimal family size that will deliver the minimum sampling variance (Robertson, 1959). As shown in section 2.7, in the case of half sib measurements, the intra-class correlation $t = \frac{1}{4}h^2$. Therefore, the optimal family size depends on the heritability, and the most efficient design for a half sib analysis has a family size *n* that is equal to $\frac{4}{h^2}$ (Robertson, 1959; Falconer, 1960). Although oestrous expression is affected by managerial and environmental factors (i.e. season, age, which were accounted for in the analysis), oestrous expression is predominantly affected by the physiological status of the cow. Therefore, it is expected that the heritability of oestrous traits will be in the range of 0.2 to 0.3 (of similar range to

CLA trait; see section 1.5.3.1). Given these expectations of the value of the heritability, the optimal half sib family size will be in the range of between 13 and 20 daughters per sire. Considering the total number of recorded cows *T* of 189 in the current study, this data must have consisted of around 10-14 sire families. Therefore, family size of 2-3 half-sibs is considered to be extremely insufficient to get reliable estimates of the genetic parameters (as it is the case in this study; Robertson, 1959).

Despite the lack of recorded cows, however, the availability of multiple records per cow might have increased the accuracy of estimation through reducing the within cow environmental variance (see section 2.2.2). Furthermore, the use of a pedigree file covering four generations might also have improved the results through accounting for more genetic links between animals in the relationship matrix (see sections 2.2.2 and 2.5.4).

8.5 Future of genetic research

Nowadays, new techniques for assessing the effects of genotypes on phenotypic traits, such as those related to reproduction in dairy cattle are becoming available at apparently ever increasing speed. Microarray high-throughput technology which provides a tool for the analysis of tens of thousands of genes simultaneously will generate enormous amounts of quickly accessible information on static aspects of the genome (Veerkamp and Beerda, 2007). This subsequently will help in identifying the transcriptional and genomic changes in a genome-wide fashion. Gene expression profiling using this technology for functional genomics has produced an abundance of information regarding possible genes and physiological processes underlying phenotypic fertility traits, such as identifying centrally expressed genes that are involved in oestrous behaviour (Beerda et al., 2008). Studies on epigenetics deals with the change in phenotype that is due to change in cellular properties (DNA function) that can be inherited without alternation in the genotype, and this is likely to prove relevant in specific areas of fertility, such as early embryonic development (Beerda et al., 2008).

Most likely, the speed in genotyping large numbers of animals for many tens of thousands of SNPs will facilitate the inclusion of genome-wide marker information

in the prediction of breeding values (Meuwissen et al., 2001). It has been demonstrated that the accuracy of GEBV for a bull calf can be as high as the EBV after progeny testing (Schaeffer, 2006). Therefore, genomic selection could potentially double the rate of genetic improvement through selection and breeding from bulls at two year of age rather than five year of age. So that conventional progeny testing will ultimately become obsolete with enormous cost savings up to 97% (Schaeffer, 2006). Genomic selection of young bulls is ongoing in many countries, (The Netherlands, United States, New Zealand and Australia; Hayes et al., 2009). Genome-wide selection is expected to have profound effects on the dairy industry as it will shift the structure of the dairy cattle breeding industry to a model in which AI organization will maintain a nucleus herd or consortium of about 10,000 animals (Schaeffer, 2006). Genome-wide association analysis would provide the knowledge regarding the most important genetic factors that influence fertility, such as genes involved in the regulation of nutrition, lactation and stress, and these can be exploited to improve management regardless of whether there is or is not genetic variation in the genes. Therefore, the accumulation of a large body of information is expected to increase the effectiveness of the genetic improvement of reproduction traits in the future (Veerkamp and Beerda, 2007).

8.6 Recommendations for the industry

Throughout this project there was a serious problem with the number of records available for genetic analysis and this will affect the reliability of the results in term of both detection of SNPs effects and estimation of genetic parameters. Furthermore, there were some difficulties in constructing pedigree files for oestrus data analysis (Chapters 6 and 7) in association with incomplete sires and dams names in addition to some missing pedigree information. Therefore, in order to be able to find the pedigree for the cows in the online UK Holstein database, all the sires and dams names had to be checked manually and corrections made where needed, with some cows being left without an identified pedigree. Correct and well identified pedigree information is one of the most important elements to get an accurate estimation of the breeding values and various genetic parameters, this being highlighted in sections 2.2.2 and 2.5.4. Another problem was the difficulty in comparing the results from

this study with similar kinds of studies either due to different analytical methods, different traits, different SNP nomenclature and different experiment design. This was even more obvious while reviewing literature on molecular markers for fertility illustrated in the varying magnitude of the established effects or even conflicted results. Therefore, with regard to the results from this project recommendations to the dairy cattle industry are:

For marker-assisted selection to be an effective tool in genetic improvement, accurate recording of phenotypic data on a large scale must be available, to facilitate more precise detection, estimation and confirmation of quantitative trait loci effects and thereafter using these estimates in selection. It is important to adopt a uniform kind of measurement method for traits, a well-described experimental design, and to apply the recommendations of the Human genome organization gene nomenclature committee in naming SNPs and genes tested in the individual experiments. This will help in comparing the results of experiments carried out in different areas and in making international comparisons.

Furthermore, despite the great potential of genome-wide selection (discussed in section 2.6) to govern the breeding programme for improving fertility in dairy cattle, conventional breeding programmes based on phenotypic data recording are likely to predominate for some years. Therefore, in the near future information on DNA markers, phenotypes and pedigree information will need to be integrated in order to estimate breeding values for potential selection candidates for the different traits. Thus, extensive phenotypic recordings of the traits for individuals and pedigree information are still a prerequisite. If this is not taken into consideration, there is a risk that information on animal physiology, behaviour and pathology will lag behind the DNA molecular approach thus rendering it less beneficial than expected. A challenge in doing this is the data collection and the quality of data. Therefore, new ways of measuring (such as physiological measures), recording (automated heat detection), and analysing data, early indicator traits (juvenile predictors of fertility), and increased knowledge of genes and their regulation are of profound importance. This will ensure proper measurements of the traits at a low cost and guarantee that these records can be effectively incorporated into breeding programmes alongside the ongoing progress in genomic information.

Heat detection (which has been covered extensively in this project) plays an important role in the economy of the dairy industry (López-Gatius et al., 2005). Therefore, extensive use of electronic tags for oestrous detection in dairy farming will help in higher oestrus detection rates. It will also make large datasets available for genetic studies, which will facilitate the exploration of the genetic background of oestrus behaviour, thus integrating information in breeding programmes for improving fertility in dairy cattle. Although they were limited in scope, the data analysed here indicated that electronic tags used for oestrous detection are reliable tools in identifying cows in oestrus: with Lely equipment 91% of electronic tag-based oestruses were confirmed by visual observation. Use of these devices will help identify cows in oestrus, and therefore eliminate the bias resulted from involving farmer decisions and management, such as delaying AI in high yielding cows.

8.7 Future work

Veerkamp and Beerda (2007) have argued that the results from SNP association studies might lack reliability as some of the discovered significant effects might be due to closely linked genes rather than to the variation in the gene itself or directly result from the experimental design or heterogeneous structure of the population. Within the context of the results of this project, the effects of the different SNPs on fertility PTAs and oestrus traits were in different directions which was surprising. Therefore, in order to provide more comprehensive definitions of the actions of SNPs, one approach will be to introduce these SNPs into the genes of interest and then to find what impact they have on proteins function. If the effects of these SNPs on oestrous strength and fertility PTAs are confirmed to be in different directions, the opportunity for genetic improvement in fertility will be severely hindered. Accordingly, in the meantime careful decisions have to be made when choosing animals carrying these SNPs for breeding for better fertility in dairy cattle.

As AI in the oestrous study (Chapter 7) was only applied when oestrus was confirmed by both electronic tag and visual observation, it would be beneficial in further studies to determine the conception rate associated with electronically recorded oestruses only. In addition, due to the fact that not all of the cows were inseminated when showing high episodes of activity and as DFS is one of the key

fertility traits in the UK, it will be interesting to establish the conception rate associated with the first episode of high activity. Measuring progesterone level in cows fitted with electronic tags for oestrous detection would help in further exploring the physiological background of behaviour-based oestrous traits. Another approach might be to limit the recording of oestrous behaviour to the first two months of lactation, as this period is critical in terms of re-establishment of ovarian function and the recovery from the impact of negative energy balance on fertility and oestrus expression. This will also help in exploring the impact of milk yield and the resultant NEB during this period on fertility in association with oestrous expression at the genetic level.

Currently, only management traits are included in the fertility index in the UK, and this may affect the reliability of fertility proofs. Therefore, it will be advantageous if traits with less management bias (CLA, oestrous traits) can be included in the selection index for better fertility. However, as selecting for better fertility might involve an additional cost in terms of reducing milk yield, it is fundamentally important to estimate the economic costs of adopting oestrous traits in breeding programmes before including it in PLI. The relative economic weight for each of these traits will also determine the economic impact of each of the SNPs discussed above. As behaviour-based oestrous traits are expected to be correlated with other fertility traits it is necessary to establish the correlation between them when considering oestrous traits to be included in PLI. This will help avoid double counting of the economic costs associated with these traits. For instance, improving heat detection is expected to reduce CI and DFS while increasing NR56 therefore some of the cost of low heat detection is already included in the cost of CI and DFS.

8.8 Summary

All the aims of this project were met. Molecular markers for fertility have been identified. Genetic and phenotypic parameters for oestrous behaviour traits were calculated. The heritabilities of oestrous-derived traits were calculated and these estimates were within the expected range of physiological traits. A novel aspect of this project was the estimation of breeding values for oestrus traits. These results

provide a basis for further exploring the genetic improvement of oestrus behaviour in dairy cattle.

Bibliography

Ascoli, M., Fanelli, F. and Segaloff, D. L. (2002). The lutropin/ choriogonadotropin receptor, a 2002 perspective. *Endocrine Reviews.* **23**, 141-174.

Barnes, M.A., Bierley, S.T., Halman, R.D. and Henricks, D.M. (1980). Follicle stimulating hormone, luteinizing hormone and oestradiol-17β response in GnRH-treated prepubertal Holstein heifers. *Biology of Reproduction*. **22**, 459-465.

Bauer-Dantoin, A.C., McDonald, J.K. and Levine, J.E. (1992). Neuropeptide Y potentiates luteinizing hormone (LH)-releasing hormone-induced LH secretion only under conditions leading to preovulatory LH surges. *Endocrinology.* **131**, 2946-92952.

Berglund, B. (2008). Genetic improvement of dairy cow reproductive performance. *Reproduction in Domestic Animals.* **43**, 89-95 (Supplement 2).

Beuzen, N.D., Stear, M. and Chang, K.C.(2000). Molecular markers and their use in animal breeding. *The Veterinary Journal*. **160**, 42-52.

Blomqvist, A.G. and Herzog, H. (1997). Y-receptor subtypes: how many more? *Trends in Neuroscience*. **20**, 294-298.

Bradford, B. J. and Allen, M. S. (2008). Negative energy balance increases periprandial ghrelin and growth hormone concentrations in lactating dairy cows. *Domestic Animal Endocrinology*. **34**, 196-203.

Brotherston, S., Banos.G. and Coffey, M.P. (2002). Evaluation of yield traits for the development of a UK fertility index for dairy cattle. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Butler, W.R. and Smith, R.D. (1989). Interrelationship between energy balance and postpartum reproductive function in dairy cattle. *Journal of Dairy Science.* **72**, 767-783.

Butler, W.R. (2000). Nutritional interactions with reproductive performance in dairy cattle. *Animal Reproduction Science.* **60-61**, 449-457.

Cameron, N.D. (1997). *Selection indics and Prediction of Genetic Merit in Animal Breeding*. CAB International, Wallingford, UK.

Caminos, J.E., Tena-Sempere, M. Gaytan, F. Sanchez-Criado, J.E., Barreiro, M.L., Nogueiras, R., Casanueva, F.F., Aguilar, E. and Dieguez, C. (2003). Expression of ghrelin in the cyclic and pregnant rat ovary. *Endocrinology*. **144**, 1594-1602.

Cárdenas, H. and Pope, W.F. (2002). Control of ovulation rate in swine. *Journal of Animal Science*. **80**, 36-46.

Carlini, D.B., Chen, Y. and Stephan, W. (2001). The relationship between third codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. *Genetics Society of America*. **159**, 623-633.

Cassady, J. P., Johnson, R. K. and Ford, J. J. (2000). Comparison of plasma FSH concentration in boars and gilts from lines selected for ovulation rate and embryonal survival, and litter size and estimation of (co)variance components for FSH and ovulation rate. *Journal of Animal Science*. **78**, 1430-1435.

Chen, C.Y. and Schwartz, R. J. (1996). Recruitment of the tinman homolog Nkx-2.5 by serum response factor activates cardiac α-actin gene transcription. *Molecular and Cellular Biology*. **16**, 6372-6384.

Chen, Y.G., Wang, Q., Lin, S.L., Chang, C.D., Chuang, J. and Ying, S.Y. (2006). Activin signalling and its role in regulation of cell proliferation, apoptosis, and carcinogenesis. Experimental Biology and Medicine (Maywood). **231**, 534-544.

Cheng, C.K., Yeung, C.M., Chow, B.K.C. and Leung, P.C.K. (2002). Characterization of a new upstream GnRH receptor promoter in human ovarian granulosa-luteal cells. Molecular Endocrinology. **16**, 1552-1564.

Chillon M., Dork, T., Casals, T., Gimenez, J., Fonknechten, N., Will, K., Ramos, D., et al. (1995). A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811 + 1.6kbA-G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. *The American Journal of Human Genetics*. **56,** 623-629

Clarke, I.J., Backholer, K. and Tilbrook, A.J. (2005). Y2 receptor-selective agonist delays the estrogen-induced luteinizing hormone surge in ovariectomized ewes, but y1-receptor selective agonist stimulates voluntary food intake. *Endocrinology*. **146**, 769-775.

Cobanoglu, O., Zaitoun, I., Chang, Y. M., Shook, G. E. and Khatib, H. (2006). Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. *Journal of Dairy Science*. **89**, 4433-4437.

Corp, E.S., Gréco, B., Powers, J.B., Marín Bivens, C.L. and Wade, G.N. (2001). Neuropeptide Y inhibits oestrous behaviour and stimulates feeding via separate receptors in Syrian hamsters. *American Journal of Physiology, Regulatory, Integrative and Comparative Physiology*. **280**, 1061-1068.

Crowley, W.R. and Kalra, S.P. (1988). Regulation of luteinizing hormone secretion by neuropeptide Y in rats: hypothalamic and pituitary actions. *Synapse.* **2**, 276-281

Curlewis, J.D., McLeod, B.J. and Loudon, A.S. (1991). LH secretion and response to GnRH during seasonal anoestrus of the Père David's deer hind (Elaphurus davidianus). *Journal of Reproduction and Fertility*. **91**, 131-138.

Darwash, A.O., Lamming, G.E. and Wooliams, J.A. (1997). Estimation of the genetic variation of the interval from calving to postpartum ovulation of dairy cows. *Journal of Diary Science*. **80**, 1227-1234.

DairyCo Datum. (2010). Farm Data- Cow numbers. Date accessed 31 March 2010. http://www.dairyco.org.uk/datum/on-farm-data/cow-numbers/uk-cow-numbers.aspx

DairyCo Farming information centre (2010). Farming information centre-fertility-breeding. Date accessed 1 April 2010. http://www.dairyco.org.uk/farming-info-centre/fertility-/breedingplus.aspx

Davis, G.H., Montgomery, G.W., Allison, A.J., Kelly, R.W. and Bray, A.R. (1982). Segregation of a major gene influencing fecundity in progeny of Booroola sheep. *New Zealand Journal of Agricultural Research*. **25**, 525-529.

Davis, G.H., McEwan, J.C., Fennessy, P.F., Dodds, K.G. and Farquhar, P.A. (1991). Evidence for the presence of a major gene influencing ovulation rate on the X

chromosome of sheep. *Biology of Reproduction*. **44**, 620-624 (quoted from Davis, 2005).

Davis, G.H., McEwan, J.C., Fennessy, P.F., Dodds, K.G., McNatty, K.P. and Wai-Sum, O. (1992). Infertility due to bilateral ovarian hypoplasia in sheep homozygous ($FecX^{I}$ $FecX^{I}$) for the Inverdale prolificacy gene located on the X chromosome. *Biology of Reproduction*. **46**, 636-640.

Davis, G.H., Galloway, S.M., Ross, I.K., Gregan, S.M., Ward, J, et al. (2002). DNA tests in prolific sheep from eight countries provide new evidence on origin of the Booroola (FecB) mutation. *Biology of Reproduction*. **66**, 1869-1874.

Davis, G. H. (2004). Fecundity genes in sheep. *Animal reproduction Science*. **82-83**, 247-253.

Davis, G.H. (2005). Major genes affecting ovulation rate in sheep. *Genetics Selection. Evolution*. **37**, 11-23.

den Dunnen, J.T. and Antonarakis, S.E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Human Mutation*. **15**, 7-12.

Defra, 2009. Defra Economics and Statistics website. Retrieved 31 March 2011. http://www.defra.gov.uk/

Dekkers, J. C. M. (2004). Commercial application of marker- and geneassisted selection in livestock: Strategies and lessons. *Journal of Animal Science*. **82,** E313-E328 ((E. Supplemnt).

Derecka, H., Ahmad, S., Hodgman, C., Royal, M.D., Woolliams, J.A. and Flint, A.P.F. (2009). Single nucleotide polymorphisms in the bovine gonadotrophin releasing hormone receptor and their effect on fertility. *Animal Genetics*. **41**, 329-331.

De Vivo, I., Huggins, G.S., Hankinson, S.E., Lescault, P.J., Boezen, M., Colditz, G.A. and Hunter, D.J. (2002). A functional polymorphism in the promoter of the progesterone receptor gene associated with endometrial cancer risk. Proceedings of the National Academy of Science. **99**, 12263-12268.

Dobson, H., Walker, S. L., Morris, M. J., Routly, J. E. and Smith, R. F. (2008). Why is it getting more difficult to successfully artificially inseminate dairy cows? *Animal*. **2**(8), 1104-1111.

Dobson, H., Smith, R. F., Royal, M. D., Knight, C. H. and Sheldon, I.M. (2007).The high-producing dairy cow and its reproductive performance. *Reproduction in Domestic Animals*. **42**, 17-23 (Supplement 2).

Drogemuller, C., Hamann, H. and Distl, O. (2001). Candidate gene markers for litter size in different German pig lines. *Journal of Animal Science*. **79**, 2565-2570.

Duan, J., Wainwright, M.S., Comeron, J.M., Saitou, N., Sanders, A.R., et al. (2003). Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Human Moleclua Genetics*. **12**: 205-16.

den Dunnen, J.T. and Antonararkis, S.E. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Human Mutation*. **15**, 7-12.

Espejo, L. A., Endres, M. I. and Salfer, J. A. (2006). Prevalence of lameness in high-producing Holstein cows housed in freestall barns in Minnesota. *Journal of Dairy Science*. **89**, 3052-3058.

Estrada, K.M., Pompolo, S., Morris, M.J., Tilbrook, A.J. and Clarke, I.J. (2003). Neuropeptide Y (NPY) delays the oestrogen-induced luteinizing hormone (LH) surge in the ovariectomized ewe: further evidence that NPY has a predominant negative effect on LH secretion in the ewe. *Journal of Neuroendocrinology*. **15**, 1011-1020.

Evans, R.D., Wallace, M., Shalloo, L., Garrick, D.J. and Dillon, P. (2006). Financial implications of recent declines in reproduction and survival of Holstein-Friesian cows in spring-calving Irish dairy herds. *Agricultural Systems*. **89**, 165-183.

Fabre, S., Pierre, A., Mulsant, P., Bodin, L., Di Pasquale E., Persani, L., Monget, P. and Monniaux, D. (2006). Regulation of ovulation rate in mammals: contribution of sheep genetic models. *Reproductive Biology and Endocrinology.* **4**, 20.

Falconer, D.S. (1960). *Introduction to Quantitative Genetics*. Oliver and Body Ltd, Glasgow, UK.

Falkenstein, E., Tillmann, H.C., Christ, M,, Feuring, M. and Wehling, M. (2000). Multiple actions of steroid hormones—a focus on rapid, nongenomic effects. *Pharmacological Review.* **52**, 513-556.

Fan, N.C., Peng, C., Krisinger, J. and Leung, P.C. (1995). The human gonadotropin-releasing hormone receptor gene: complete structure including multiple promoters, transcription initiation sites, and polyadenylation signals. *Molecular and Cellular Endocrinology*. **107**, 18.

Flint, A.P.F., Wall, E., Coffey, M. P., Simm, G., Brotherstone, S., Stott, A.W., Santarossa, J., Royal, M.D. and Wooliams, J.A.F. (2002). Introducing a UK Fertility Index. *Cattle Practice*. Volume 10 Part 4.

Flint., A.P.F., Royal, M.D., Coffey, M., Wall, E., Simm, G., Brtherstone, S., Stott, A.W. and Wooliams, J.A. (2003). Developing a UK fertility index. Cattle Fertility Conference, 2003.

Flint., A.P.F., Brtherstone, S., Coffey, M.P., Royal, M.D., Simm, G., Santarossa, G., Simm, J., Stott, A. W., and J.A.F. Wooliams J. (2004). The UK fertility Index. The British Cattle Conference. The British Cattle Breeders Club. 59-62.

Flint, A.P.F. (2006). Dairy cow fertility: an inherited disease, *Cattle Practice*. **14**, 29-32.

Friedman, J.M. and Halaas, J.L. (1998). Leptin and the regulation of body weight in mammals. *Nature.* **395**, 763-770

Galloway, S.M., McNatty, K.P., Cambridge, L.M., Laitinen, M.P.E., Juengel, J.L., Jokiranta, T.S., McLaren, R.J., Luiro, K., Dodds, K.G., Montgomery, G.W., Beattie, A.E., Davis, G.H. and Ritvos, O. (2000). Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genetics*. **25**, 279-283.

Garcia, M.C., Lopez, M., Alvarez, C.V., Casanueva, F., Tena-Sempere, M. and Dieguez, C. (2007). Role of ghrelin in reproduction. *Reproduction.* **133**, 531-540.

Giblin, L., Butler, S.T., Kearney, B.M., Waters, S.M., and Berry, D.P. (2010). Association between leptin polymorphisms and performance traits in Holstein-Friesian dairy cattle. BSAS annual conference, Belfast, 2010.

Gilmour, A.R., Gogel, B. J., Cullis, B R. and Thompson, R. (2009). *ASREML User Quid, release 3*.

Goddard, M. E. and Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*. **124**, 323-330.

Gordon, D.F., Wagner, J., Atkinson, B.L., Chiono, M., Berry, R., Sikela, J. and Gutierrez-Hartmann, A. (1996). Human Cart-1: Structural organization, chromosomal localization, and functional analysis of cartilage-specific homeodomain cDNA. *DNA & Cell Biology*. **15**, 531-541.

Garnsworthy, P.C., Sinclair, K.D. and Webb, R. (2008). Integration of physiological mechanisms that influence fertility in dairy cows. *Animal*. **2**, 1144-1152.

Garnsworthy, P.C.,Fouladi-Nashta, A.A., Mann, G.E., Sinclair, K.D. and Webb, R. (2009). Effect ofvdietary-induced changes in plasma insulin concentrations during the early post partum period on pregnancy rate in dairy cows. *Reproduction*. **137**, 759-768.

Gustafsson J. A. (1999). Estrogen receptor β a new dimension in oestrogen mechanism of action. *Journal of Endocrinology*. **163**, 379-383.

Haley, C.S., Lee, G.J., Fordyce, M., Baxter, G., Land, R.B. and Webb, R. (1989). Study of LH response to GnRH in the young male as a criterion of genetic merit for female reproduction in sheep. *Journal of Reproduction and Fertility*. **86**, 119-133.

Harney, J. P., Mirando, M. A., Smith, L. C. and Bazer, F. W. (1990). Retinol-binding protein: A major secretory product of the pig conceptus. *Biology of Reproduction*. **42**, 523-532.

Harrison, R.O., Ford, S.P., Young, J.W. and Conley, A.J. (1990). Increased milk production versus reproductive and energy status of high producing dairy cows. *Journal of Dairy Science.* **73**, 2749-2758.

Hastings, N., Donn, S., Derecka, K., Flint, A.P.F. and Wooliams, J.A. (2006). Mutations within the coden regions of the bovine luteinising hormone receptor and their association with fertility traits. *Animal Genetics*. **37**, 583-585.

Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Goddard, M.E. (2009). Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science.* **92**, 433-443.

Hazel, L. N. (1943). The genetic basis for constructing selection indices. *Genetics.* **38**, 476-490.

Hazel, L. N., Dickerson, G. E. and Freeman, A. E. (1994). The selection index: Then, now, and for the future. *Journal of Dairy Science*. **77**, 3236-3251.

Havez, E.S.E. (1993). *Reproduction in Farm Animals, 6th Edition*. Lea and Febiger, Malven, USA.

Henderson, C. R. (1949). Estimation of changes in herd environment. *Journal of Dairy Science*. **32**, 706-711 (quoted from Henderson, 1973).

Henderson, C. R. (1973). Sire evaluation and genetic trends. In Proceedings of the Animal Breeding and Genetics Symposium in Honor of Jay L. Lush. ASAS and *American Dairy Science Association*, Champaign, IL, 10-41.

Henderson, C. R. (1975). Comparison of alternative sire evaluation methods. *Journal of Dairy Science*. **41**, 760-770.

Henderson, C.R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*. **32**, 69-83.

Henderson, C. R. and R. L. Quaas. (1976). Multiple trait evaluation using relatives' records. *Journal of Dairy Science*. **43**, 1188-1197.

Henderson, C. R. (1984). *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph, Ontario.
Henderson, C. R. (1988). Theoretical basis and computational methods for a number of different animal models. *Journal of Dairy Science*. **71**, 1-16 (supplement 2).
Herynk, M.H. and Fuqua, S.A. (2004). Oestrogen receptor mutations in human disease. *Endocrine Review*. **25**, 869-898.
Hill, W. G. (1971). Design and efficiency of selection experiments for estimating genetic parameters. *Biometrics*. **27**, 293-311.
Hill, W. G. (1974). Design and efficiency of selection experiments for estimating genetic parameters. *Biometrics*. **30**, 447-468.
Huhtaniemi, I.T. and Themmen, A.P.N. (2001). Mutations of gonadotropins and gonadotropin receptor genes: the clinical perspective. *Endocrinology and Diabetes*. **8**, 291-295.
Huhataniemi, I.T. (2002). The role of mutation affecting gonadotrophin secretion and action in disorders of pubertal development. *Best Practice and Research Clinical Endocrinology. Metabolism*.**16**, 123-138.
Inskeep, E.K., Braden, T.D., Lewis, P.E., Garcia-Winder, M. and Niswender, G.D. (1988). Receptors for luteinizing hormone and follicle-stimulating hormone in largest follicles of postpartum beef cows. *Biology of Reproduction*. **38**, 587-591.
Isler B.J., Irvin K.M., Neal S.M., Moeller S.J. and Davis M.E. (2002). Examination of the relationship between the oestrogen receptor gene and reproductive traits in swine. *Journal of Animal Science*. **80**, 2334-2339.
Johnson, R. K., Nielsen, M. K. and Casey, D. S. (1999). Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. *Journal of Animal Science*. **77**, 541-557.
Kadarmindeen, H.N. and Coffey, M. (2001). Quality and validation of insemination data for national genetic evaluation for dairy cow fertility in the United Kingdom. Proceeding of the INTERBULL meeting, Budapest, Hungary, INTERBULL Bulletin. **27**, 133-138.
Kakar, S.S., Rahe, C.H., and Neill, J.D. (1993). Molecular cloning, sequencing, and characterizing the bovine receptor for gonadotropin releasing hormone (GnRH). *Domestic Animal Endocrinology*. **10**, 335-342.
Kennedy, B.W., Schaeffer, L.R. and Sorensen, D.A. (1988). Genetic properties of animal models. *Journal of Dairy Science*. **71**, 17-26 (Supplement 2).
Kesler, D.J. and Garverick, H.A. (1982). Ovarian cysts in dairy cattle: A review. *Journal of Animal Science*. **55**, 1147-1159.
Khatib, H., Monson, R. L., Schutzkus, V., Kohl, D. M., Rosa, G. J. and Rutledge, J. J. (2008). Mutations in the STAT5A gene are associated with embryonic survival and milk composition in cattle. *Journal of Dairy Science*. **91**, 784-793.
Knight, P.G. (1996). Roles of inhibins, activins, and follistatin in the female reproductive system. Front Neuroendocrinology. **17**, 476-509
Kojima, M. and Kangawa, K. (2005). Ghrelin: Structure and function. *Physiology Review*. **85**, 495-522.

Kuiper, G.G., Enmark, E., Pelto-Huikko, M., Nilsson, S. and Gustafsson, J. Å. (1996). Cloning of a novel oestrogen receptor expressed in rat prostate and ovary. Proceeding of the National Academy of Science. USA. **93**, 5925-5930.

Lacroix, A. and Pelletier, J. (1979). LH and testosterone release in developing bulls following LH-RH treatment. Effect of gonadectomy and chronic testosterone propionate pre-treatment. Acta Endocrinologica, Copenhagen. **91**, 719-729.

Leroy, J.L., Van Soom, A., Opsomer, G., Goovaerts, I.G. and Bols, P.E. **(**2008). Reduced fertility in high-yielding dairy cows: are the oocyte and embryo in danger? Part II. Mechanisms linking nutrition and reduced oocyte and embryo quality in high-yielding dairy cows. *Reproduction of Domestic Animals*. **43**, 623-632.

Levitsky, V.G., Podkolodnaya, O.A., Kolchanov, N.A. and Podkolodny, N.L. (2001). Nucleosome formation potential of eukaryotic DNA: Calculation and promoters analysis. *Bioinformatics*. **17**, 998-1010.

Levitsky, V.G. (2004). RECON: A program for prediction of nucleosome formation potential. *Nucleic Acids Research*. **32**, 346-349.

Liefers, S.C., Veerkamp, R.F., Te Pas, M.F.W., Delavaud, C., Chilliard, Y. and Van der Lende, T. (2003). Leptin concentrations in relation to energy balance, milk yield, intake, live weight, and estrus in dairy cows. *Journal of Dairy Science*. **86**(3), 799-807.

Liefers, S. C., Veerkamp, R. F., Te Pas, M. F., Delavaud, C., Chilliard, Y. and Van der Lende. T. (2004). A missense mutation in the bovine leptin receptor gene is associated with leptin concentrations during late pregnancy. *Animal Genetics*. **35**, 138-141.

Liefers, S. C., Veerkamp, R. F., Te Pas, M. F., Chilliard, Y. and Van der Lende, T. (2005a). Genetics and physiology of leptin in periparturient dairy cows. Domestic *Animal Endocrinology*. **29**, 227-238.

Liefers, S.C., Veerkamp, R.F., Te Pas, M.F., Delavaud, C., Chilliard, Y., Platje, M. and Van der Lende, T. (2005b). Leptin promoter mutations affect leptin levels and performance traits in dairy cows. *Animal Genetics*. **36**, 111-118.

Lin, L.S., Roberts, V.J. and Yen, S.S. (1995). Expression of human gonadotropin-realising receptor gene in the placenta and its functional relationship to human chorionic gonagotropin secretion. *Journal of Clinical Endocrinology and Metabolism*. **80**, 580-585.

Lin, L., Conway, G.S., Hill, N.R. , Dattani, M.T., Hindmarsh, P.C. and Achermann, J.C. (2006). A homozygous R26Q mutation in the gonadotropin-releasing hormone receptor (GnRHR) presenting as constitutional delay of growth and puberty with subsequent borderline oligospermia. *Journal of Clinical Endocrinology and Metabolism*. **91**, 5117-5121.

Lints, T.J., Parsons, L.M., Hartley, L. Lyons, I. and Harvey, R.P. (1993). Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development*. **119**, 419-431.

Linville, R. C., Pomp, D. Johnson, R. K. and Rothschild, M. F. (2001). Candidate gene analysis for loci affecting litter size and ovulation rate in swine. *Journal of Animal Science*. **79**, 60-67.

Liu, N., Zhang, K., Zhao, H. (2008). Haplotype association analysis. *Advances in Genetics*. **60**, 335-405.
Lopez, H., L. D. Satter, and M. C. Wiltbank. (2004). Relationship between level of milk production and estrous behavior of lactating dairy cows. *Animal Reproduction Science*. **81**, 209-223.
López-Gatius, F., Santolaria, P., Mundet, I. Yániz, J.L. (2005). Walking activity at estrus and subsequent fertility in dairy cows. *Theriogenology*. **63**, 1419-1429
Løvendahl, P. and Chagunda, M.G. (2009). Short communication: Genetic variation in oestrus STEPS traits. *Journal of Dairy Science*. **92**, 4683-4688.
Løvendahl, P. and Chagunda, M.G. (2010). On the use of physical STEPS monitoring for estrus detection in dairy cows. *Journal of Dairy Science*. **93**, 249-259.
Lubahn, D.B., Moyer, J.S., Golding, T.S., Couse, J.F., Korach, K.S. and Smithies, O. (1993). Alteration of reproductive function but not prenatal sexual development after insertional disruption of the mouse oestrogen receptor gene. . Proceedings of the National Academy of Sciences . USA. **90**, 11162-11166.
Lydon, J.P., DeMayo, F.J., Funk, C.R., Mani, S.K., Hughes, A.R., Montgomery, C. A., Jr., Shyamala, G., Conneely, O. M. And O'Malley, B. W. (1995). Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes and Development*. **9**, 2266-2278.
Lyimo, Z.C., Nielen, M., Ouweltjes, W., Kruip, T.M.A. and van Eerdenburg, F.J.C.M. (2000). Relationship among oestradiol, cortisol and intensity of oestrous behaviour in dairy cattle. *Theriogenology*. **53**, 1783-1795.
Lucy, M. C. (2001). Reproductive loss in high producing dairy cattle: Where will it end? *Journal of Dairy Science*. **84**, 1277-1293.
Maciel, M. N., Zieba, D. A., Amstalden, M., Keisler, D. H. Neves, J. P. and Williams, G. L. (2004). Leptin prevents fasting-mediated reductions in pulsatile secretion of luteinizing hormone and enhances its gonadotropin-releasing hormone-mediated release in heifers. *Biology of Reproduction*. **70**, 229-235.
Margetic, S., Gazzola, C., Pegg, G.G. and Hill, R.A. (2002). Leptin: a review of its peripheral actions and interactions. *International Journal of Obesity and Related Metabolic Disorders*. **26**, 1407-1433.
McLeod, B.J., Haresign, W., Peters, A.R. and Lamming, G.E. (1984). Plasma LH and FSH concentrations in prepubertal beef heifers before and in response to repeated injections of low doses of GnRH. *Journal of Reproduction and Fertility*. **70**, 137-144.
McNatty, K.P., Juengel, J.L., Wilson, T., Galloway, S.M. and Davis,G.H. (2001). Genetic mutations influencing ovulation rate in sheep. Reproduction, Fertility and Development. **13**, 549-555.
McNatty, K,P., Galloway, S.M., Wilson, T., Smith, P., Hudson, N.L., O'Connell, A., Bibby, A,H,, Heath, D.A., Davis, G.H., Hanrahan, J.P. and Juengel, J.L. (2005). Physiological effects of major genes affecting ovulation rate in sheep. Genetic Selection Evolution. **37**, 25-38 (Supplement 1).
Meuwissen, T. (2003). Genomic selection: The future of marker assisted selection and animal breeding. In: Marker Assisted selection: a fast track to increase genetic gain in plants and animal breeding? Setion II: MAS in animals. FAO Electronic

forum on biotechnology in food and agriculture, http://www.fao.org/Biotech/docs/Meuwissen.pdf
Meuwissen T. (2007). Genomic selection: Marker-assisted selection on a genome wide scale. *Journal of Animal Breeding and Genetics.* **124**, 321-322.
Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. **157**, 1819-1829.
Millar, R.P., Lu, Z.L., Pawson, A.J., Flanagan, C.A., Morgan, K. and Maudsley, S.R. (2004). Gonadotropin-releasing hormone receptors. *Endocrine Reviews.* **25**, 235-275.
Miner, J. L. (1992). Recent advances in the central control of intake in ruminants. *Journal of Animal Science*. **70**, 1283-1289.
Montgomery, G.W., McNatty, K.P. and Davis, G.H. (1992). Physiology and molecular genetics of mutations that increase ovulation rate in sheep. *Endocrine Reviews*.**13**, 309-328.
Montgomery, G.W., Crawford, A. M., Pentyk, M., Dodds, G., Ede, A. J., Henry, H.M., Pierson, C.A., Lord, E.A., Galloway, S.M., Schmack, A.E., Sise, J.A., Swarbrick, P.A., Hanrahan, V., Buchanan, F.C. and Hill, D.F. (1993). The ovine Booroola fecundity gene *(FecB)* is linked to markers from a region of human chromosome *4q*. *Nature Genetics*. **4**, 410-414.
Montgomery, G.W., Galloway, S.M., Davis, G.H. and McNatty, K.P. (2001). Genes controlling ovulation rate in sheep. *Reproduction*. **121**, 843-852.
Mooney, S. (2005). Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. *Briefings in Bioinformatics*. **6**, 44-56.
Muñoz, M., Fernández, A, I., Óvilo, C., Muñoz, G., Rodriguez, C., Fernández, A., Alves, E. and Silió, L. (2010). Non-additive effects of *RBP4, ESR1* and *IGF2* polymorphisms on litter size at different parities in a Chinese-European porcine line. *Genetics Selection Evolution.* **42,** 23.
Mrode, R.A. (2005). *Linear Models for the Prediction of Animal Breeding Values*, 2nd Edition. CAB International, Wallingford, UK.
Mulac-Jericevic, B., Lydon, J.P., DeMayo, F.J. and Conneely, O.M. (2003). Defective mammary gland morphogenesis in mice lacking the progesterone receptor β isoform. Proceedings of the National Academy of Sciences. USA.**100**, 9744–9749.
Mulac-Jericevic, B. and Conneely, O.M. (2004). Reproductive tissue selective actions of progesterone receptors. *Reproduction*.**128**, 139–146.
Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., Pisselet, C., Riquet. J., Monniaux, D., Callebaut, I., Cribiu, E., Thimonier, J., Teyssier, J., Bodin, L., Cognie, Y., Chitour, N. and Elsen, J.M.l. (2001). Mutation in bone morphogenetic protein receptor-1B is associated with increased ovulation rate in Booroola Merino ewes. Proceedings of the National Academy of Sciences. USA. **98**, 5104-5109.
Nakada, K., Ishikawa, Y., Nakao, T. and Sawamuka, Y. (2002). Changes in responses to GnRH on luteinising hormone and follicle stimulating hormone secretion in prepubertal heifers. *Journal of Reproduction and Development*. **48**, 545-551.
Ngan, A.E., Cheng, P.K., Leung, P.C.K. and Chow, B.K.C. (1999). Steroidogenic Factor-1 interacts with Gonadotrope-Specific Element within the first exon of the

human gonadotropin-releasing hormone receptor gene to mediate gonadotrope specific expression. *Endocrinology*. **140**, 2452-2462.

Nickerson, D.A., Taylor, S.L., Weiss, K.M., Clark, A.G., Hutchinson, R.G., Stengard, J. Salomaa, V., Vartiainen, E., Boerwinkle, E. and Sing, C.F. (1998). DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nature Genetics*. **19**, 233-240.

Nordhoff, V., Gromoll, J. and Simoni, M. (1999). Constitutively active mutations of G protein-coupled receptors: the case of the human luteinizing hormone and follicle-stimulating hormone receptors. *Archives of Medical Research*. **30**, 501-509.

Pangas, S.A., Jorgez, C.J., Tran, M., Agno, J., Li, X., Brown, C.W., Kumar, T.R. and Matzuk, M,M. (2007). Intraovarian activins are required for female fertility. *Molecular Endocrinology*. **21**, 2458-2471.

Petersson, K.J., Berglund, B., Strandberg, E., Gustafsson, H., Flint, A.P.F., Woolliams, J.A. and Royal, M.D. (2007). Genetic analysis of postpartum measures of luteal STEPS in dairy cows. *Journal of Dairy Science*. **90**, 427-434.

Piper, L.R. and Bindon, B.M. (1982). The Booroola Merino and the performance of medium non-Peppin crosses at Armidale. In the Booroola Merino (eds Piper, LR. *et al*.) 9−20. CSIRO, Melbourne.

Pottier, N., Cheok, M.H., Yang, W., Assem, M., Tracey, L., Obenauer, J.C., Panetta, J.C., Relling, M.V. and Evans, W.E. (2007). Expression of SMARCB1 modulates steroid sensitivity in human lymphoblastoid cells: identifi cation of a promoter SNP that alters PARP1 binding and SMARCB1 expression .*Human Molecular Genetics*. **16** (19), 2261- 2271.

Pryce, J.E., Coffey, M.P. and Brtherstone, S. (2000). The genetic relationship between calving interval, Condition score and linear type and management traits in pedigree registered Holstein Cows. *Journal of Dairy Science*. **83**, 2664-2671.

Pryce, J.E., Royal, M. D., Garnsworthy, P.C and Mao, I.L. (2004). Fertility in the high producing dairy cow. *Livestock Production Science*. **86,** 125–135.

Rahal, P., Latronico, A.C., Kohek, M.B.F., de Lucia, R.F.S., Milazzotto, M.P., Wheeler, M.B., Ferraz, J.B.S., Eler, J.P. and Garcia, F. (2000). Polymorphisms in the bovine follicle-stimulating hormone receptor gene. *Animal Gentics*. **31**, 280-281.

Revelli, A., Massobrio, M. and Tesarik, J. (1998) Nongenomic actions of steroid hormones in reproductive tissues. *Endocrine Reveiews.* **19,** 3-17.

Robertsona, A. (1959). Experimental design in the evaluation of genetic parameters. *Biometrics*. **15,** 219-226.

Roche, J. R., Sheahan, A. J. Chagas, L. M. and Berry, D. P. (2006). Short communication: Genetic selection for milk production increases plasma ghrelin concentration in pasture-based dairy cows. *Journal of Dairy Science*. **89**, 3471-3475.

Roelofs, J.B., van Eerdenburg, F.J.C.M, Soede, N.M. and Kemp, B. (2005a). Various behavioural signs of oestrus and their relationship with time of ovulation in dairy cattle. *Theriogenology*. **63**, 1366-1377.

Roelofs, J.B., van Eerdenburg, F.J.C.M., Soede, N.M. and Kemp, B. (2005b). Pedometer readings for estrus detection and as predictor for time of ovulation in dairy cattle. *Theriogenology.* **64**, 1690-1703.

Rothschild, M. F., Jacobson, C., Vaske, D. A., Tuggle, C. K., Wang, L., Short, T., Eckardt, G., Sasaki, S., Vincent, A., Mclaren, D., Southwood, O., Van Der Steen, H.,

Mileham, A. and Plastow, G. (1996). The oestrogen receptor locus is associated with a major gene influencing litter size in pigs. Proceedings of the National Academy of Sciences. USA. **93**, 201-205.
Rothschild, M. F., Messer, L., Day, A., Wales, R., Short, T., Southwood, O. and Plastow, G. (2000). Investigation of the retinol-binding protein 4 (RBP4) gene as a candidate gene for increased litter size in pigs. *Mammalian Genome*. **11**, 75-77.
Rothschild, M.F, Hu Z-L, Jiang Z. (2007). Advances in QTL Mapping in Pigs. *International Journal of Biological Science*. **3**, 192-197.
Royal, M. D. (1999). *Genetic variation in endocrine parameters of fertility*. PhD thesis. University of Nottingham.
Royal, M.A.D., A.O. Darwash, A.P.F. Flint, R. Webb, J.A. Wooliams, and G.E. Lamming. (2000a). Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility. *Animal Science*. **70**, 487-501.
Royal, M.A.D., Mann, G.E. and Flint, A.P.F. (2000b). Strategies for reversing the trend towards subfertility in dairy cattle. *The Veterinary Journal*. **160**, 53-60.
Royal, M.A.D., Flint, A.P.F. and Woolliams, J.A. (2002a). Genetic and phenotypic relationships among endocrine and traditional fertility traits and production traits in Holstein-Frisian dairy cows. *Journal of Dairy Science*. **85**, 958-967.
Royal, M.A.D., Pryce, J.E. Woolliams, J.A. and Flint, A.P.F. (2002b). The genetic relationship between commencement of luteal STEPS and calving interval, body condition score, production, and linear type traits in Holstein-Frisian dairy cows. *Journal of Dairy Science*. **85**, 3071-3080.
Roxström, A., Strandberg, E., Berglund, B., Emanuelson, U. and Philipsson. J. (2001). Genetic and environmental correlations among female fertility traits and milk production in different parities of Swedish Red and White dairy cattle. *Acta Agriculturae Scandinavica Section A Animal Science*. **51**, 7-14.
Rudnick, A., Ling, T.Y., Odagiri, H., Rutter, W.J. and German, M.S. (1994). Pancreatic beta cells express a diverse set of homeobox genes. Proceedings of the National Academy of Sciences, USA. **91**, 12203-12207.
Salisbury, G.W., VanDemark, N.L and Lodge, J.R. (1978). *Physiology of Reproduction and Artificial Insemination of Cattle, Second Edition*. W. H. Freeman and Company, San Francisco, USA.
Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Jouranl of Animal Breeding and Genetics*. **123,** 218-223.
Schams, D., Schallenberger, E., Gombe, S. and Karg, H. (1981). Endocrine patterns associated with puberty in male and female cattle. *Journal of Reproduction and Fertility*. **30**, 103-110 (Supplement).
Shackell G.H., Hudson, N.L., Heath, D.A., Lun, S., Shaw, L., Condell, L., Blay, L.R. and McNatty, K.P. (1993). Plasma gonadotropin concentrations and ovarian characteristics in Inverdale ewes which are heterozygous for a major gene (FecX') on the X chromosome that influences ovulation rate. *Biology of Reproduction*. **48**, 1150-1156.
Shan, K., Ying, W., Jian-Hui, Z., Wei, G. and Yan, L. (2005). The function of the SNP in the MMP1 and MMP3 promoter in susceptibility to endometriosis in China. *Molecular Human Reproduction*. **11**, 423-427.

Sherman, E. L., Nkrumah, J. D., Murdoch, B. M., Li, C., Wang, Z., Fu, A. and Moore, S. S. (2008). Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. *Journal of Animal Science*. **86**, 1-16.

Shimasaki, S., Moore, R.K., Otsuka, F. and Erickson, G.F. (2004). The bone morphogenetic protein system in mammalian reproduction. *Endocrine Reviews*. **25**, 72-101.

Short, T. H., Rothschild, M. F., Southwood, O. I., McLaren, D. G., de Vries, A., Van Der Steen, H., Eckardt, G.R., Tuggle, C.K., Helm, J., Vaske, D.A., Mileham, A.J. and Plastow, G.S. (1997). Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. *Journal of Animal Science*. **75**, 3138-3142.

Simm, G. (1998). *Genetic Improvement of Cattle and Sheep*. Farming Press, Tonbridge, UK.

Simoncini, T. And Genazzani, A.R. (2003). Non-genomic actions of sex steroid hormones. *European Journal of Endocrinology*. **148**, 281-292

Simoni, M., Gromoll, J. and Nieschlag, E. (1997). The follicle-stimulating hormone receptor: biochemistry, molecular biology and pathophysiology. *Endocrine Reviews*. **18**, 739-773

Sørensen, A.C., Lawlor, T. and Ruiz, F. (2007). A survey on fertility in the Holstein populations of the world. In: Proceeding of International Conference: Fertility in Dairy Cows bridging the gaps. British Society of Animal Science and University of Liverpool. 1- 17

Souza, C.J., MacDougall, C., Campbell, B.K., McNeilly, A.S. and Baird, D.T. (2001). The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. *Journal of Endocrinology*. **169**, 1-6.

Stojilkovic, S.S., Reinhart, J. and Catt, K.J. (1994). Gonadotropin-releasing hormone receptors: structure and signal transduction pathways. *Endocrine Review*s. **15**, 462-499.

Terman, A. (2005). Effect of the polymorphism of prolactin receptor (PRLR) and leptin (LEP) genes on litter size in Polish pigs. *Journal of Animal Breeding and Genetics.* **122**, 400-404.

Themmen, A.P.N. and Huhtaniemi, I.T. (2000). Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary–gonadal function. *Endocrine Reviews*. **21**, 551-583.

Thomas, M.G., Gazal, O.S., Williams, G.L., Stanko, R.L. and Keisler, D.H. (1999). Injection of neuropeptide Y into the third cerebroventricle differentially influences pituitary secretion of luteinizing hormone and growth hormone in ovariectomised cows. *Domestic Animal Endocrinology*. **16**, 159-169.

Tomás, A., Casellas, J., Ramirez, O., Munoz, G., Noguera, J.L. and Sanchez, A. (2006). High amino acid variation in the intracellular domain of the pig prolactin receptor (PRLR) and its relation to ovulation rate and piglet survival traits. *Journal of Animal Science.* **84**, 1991-1998

Tregouet, D.A., Barbaux, S., Escolano, S., Tahri, N., Golmard, J.L., Tiret, L. and Cambien, F. (2002). Specific haplotypes of the P-selectin gene are associated with myocardial infarction. *Human Molecular Genetics*. **11**, 2015-2023.

Tregouet, D.A., Escolano, S., Tiret, L., Mallet, A. and Golmard, J.L. (2004a). A new algorithm for haplotype-based association analysis: the SEM algorithm. *Annals of Human Genetics*. **68**, 165-177.

Tregouet, D-A., Ricard, S., Nicaud, V., Arnould, L., Soubigou, S., Rosier, M., Duverger, N., Poirier, O., Mace, S., Kee, F., Morrison, C., Denefle, P., Tiret, L., Evans, A., Deleuze, J.F. and Cambien, F. (2004b). In-depth haplotype analysis of ABCA1 gene polymorphisms in relation to plasma ApoA1 levels and myocardial infarction. Arteriosclerosis, *Thrombosis and Vascular Biology*. **24**, 775-781.

Tregouet, D.A. and Garelle, V. (2007). A new JAVA interface implementation of THESIAS: Testing haplotype effects in association studies. *Bioinformatics*. **23**, 1038-1039.

Tschöp, M., Smiley, D. L. and Heiman, M. L. (2000). Ghrelin induces adiposity in rodents. *Nature*. **407**, 908-913

Van Eerdenburg, F.J.C.M., Karthaus, D., Taverne, M.A.M., Merics, I. and Scenzi, O. (2002). The relationship between estrous behavioral score and time of ovulation in dairy cattle. *Journal of Dairy Science*. **85**, 1150-1156.

VanRaden, P. M. and Wiggans, G. R. (1991). Derivation, Calculation and use of national animal model information. *Journal of Dairy Science*. **74**, 2737-2746.

Van Rens, B. T. T. M. and Van der Lende, T. (2002). Litter size and piglet traits of gilts with different prolactin receptor genotypes. *Theriogenology*. **57**, 883-893

Veerkamp, R.F., Oldenbroek, J.K. and van der Lede, T. (1997). The use of milk progesterone measurement for genetic improvement of fertility traits in dairy cattle. Proceedings of the International Workshop on Genetic Improvement of Functional Traits In Cattle; Fertility and Reproduction, GRUB, Germany. Interbull. **18**, 62-65.

Vegeto, E., Shahbaz, M.M., Wen, D.X., Goldman, M.E., O'Malley, B.W. and McDonnell, D.P. (1993). Human progesterone receptor A form is a cell- and promoter-specific repressor of human progesterone receptor B function. *Molecular Endocrinology*. **7**, 1244-1255.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al (2001).The sequence of the human genome. *Science*. **291**, 130-1351.

Veerkamp, R. F. and Beerda. B. (2007). Genetics and genomics to improve fertility in high producing dairy cows. *Theriogenology*. **68S**, S266-S273

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*. **34**(3), 275-305.

Viitala, S., Szyda, J., Blott, S., Schulman, N., Lidauer, M., Maki-Tanila, A., Georges, M. and Vilkki, J. (2006): The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. *Genetics*. **173**, 2151-2164.

Vincent, A. L., Evans, G., Short, T. H., Southwood, O. I., Plastow, G. S., Tuggle, C. K. and Rothschild, M. F. (1998). The prolactin receptor gene is associated with

increased litter size in pigs. In: Proc. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production. Armidale, NSW, Australia. **27**, 15-18.

Wain, H.M., Bruford, E.A., Lovering, R.C., Lush, M.J., Wright, M.W. and Povey, S. (2002)Guidelines for Human Gene Nomenclature**.** *Genomics*. **79,** 464-470.

Wall, E., Brotherstone, S., Wooliams, J.A., Banos, G. and Coffey, M.P. (2003). Genetic evaluation of fertility using direct and correlated traits. *Journal of Dairy Science*. **86**, 4093-4102.

Wathes, D.C., Brickell, J.S., Bourne, N.E., Swalia, A. and Cheng, Z. (2008). Factors influencing heifer survival and fertility on commercial dairy farms. *Animal*. **2**, 1135-1143.

Wells, S. J., Trent, A. M., Marsh, W. E. and Robinson, R. A. (1993). Prevalence and severity of lameness in lactating dairy cows in a sample of Minnesota and Wisconsin herds. *Journal of the American Veterinary Medical Association*. **202**, 78-82.

Welt, C., Sidis, Y., Keutmann, H. and Schneyer, A. (2002). Activins, inhibins, and follistatins: from endocrinology to signaling. A paradigm for the new millennium. *Expermintal Biology and Medicine*. (Maywood). **227**, 724-752.

Wertz-Lutz, A.E., Knight, T.J., Pritchard, R.H., Daniel, J.A., Clapper, J.A., Smart, A.J., Trenkle, A. and Beitz, D.C. (2006). Circulating ghrelin concentrations fluctuate relative to nutritional status and influence feeding behavior in cattle. *Journal of Animal Science*. **84**, 3285-3300.

Wilson, T., Wu Xi-Yang, Juengel, J.L., Ross, I.K., Lumsden J.M., Lord, E.A., Dodds, K,G., Walling, G.A., McEwan, J.C., O'Connell, AR., McNatty, K.P. and Montgomery, G.W. (2001). Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK- 6) that is expressed in both oocytes and granulosa cells. *Biology of Reproduction*. **64**, 1225-1235.

Wolfahnt, S., Kliene, B. and Rossmanith, W.G. (1998). Detection of gonadotropin-releasing hormone and its receptor mRNA in human placental trophoblasts using in situ reverse transcription-polymerase chain reaction. Molecular *Human Reproduction*. **4**, 999-1006.

Wren, A.M., Small, C.J., Ward, H.L., Murphy, K.G., Dakin, C.L., Taheri, S., Kennedy, A.R., Roberts, G.H., Morgan, D.G., Ghatei, M.A. and Bloom, S.R. (2000). The novel hypothalamic peptide ghrelin stimulates food intake and growth hormone secretion. *Endocrinology*. **141**, 4325-4328.

Wu H.-M., Wang H.-S., Huang H.-Y., Soong Y.-K., MacCalman C.D. and Leung P.C.K. (2009). GnRH signaling in intrauterine tissues. *Reproduction* **137**, 769-77.

Xu, Z., Garverick, H.A., Smith, G.W., Smith, M.F., Hamilton, S.A. and Youngquist, R.S. (1995). Expression of follicle-stimulating hormone and luteinizing hormone receptor messenger ribonucleic acids in bovine follicles during the first follicular wave. *Biology of Reproduction*. **53**, 951-957.

Yaniz, J. L., Santolaria, Giribet, A. and Lopez-Gatius, F. (2006). Factors affecting walking activity at estrus during postpartum period and subsequent fertility in dairy cows. *Theriogenology*. **66** (8), 1943-1950.

Yoshioka, K., Suzuki, C. and Iwamura, S. (1998). Activin A and follistatin regulate developmental competence of in vitro-produced bovine embryos. *Biology of Reproduction*. **59**, 1017-1022.

Zhao, Y., Ning, L., Lu, X. Gengsheng, C., Yizhen, C., Shun, Z., Yongfu, C., Changxin, W., Jiansheng, Z., Shiquan, S. and Xueqing, X. (1998). FSH*β* subunit gene is associated with major gene controlling litter size in commercial pig breeds. *Science in China Series C: Life Science*. **41**, 664-668.

Zhao, H., Pfeiffer, R. and Gail, M.H. (2003) Haplotype analysis in population genetics and association studies. *Pharmacogenomics*. **4**, 171-178.