Technische Universität München
Lehrstuhl für Tierzucht

Genomic Characterization of Genes Encoding Diacylglycerol Acyltransferase Activity in Cattle and Swine

Andreas Winter

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. sc. agr., Dr. agr. habil. Leo Dempfle

Prüfer der Dissertation:
1. Univ.-Prof. Dr. sc. techn. ETH, Dr. agr. habil. Hans-Rudolf Fries
2. Univ.-Prof. Dr. med. Thomas Alois Meitinger
3. Univ.-Prof. Dr. rer. nat., Dr. rer. nat. habil. Gerhard Wenzel

Die Dissertation wurde am 21.05.2003 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 15.07.2003 angenommen.

Be warned, my son, of anything in addition to them. Of making many books there is no end, and much study wearies the body. Now all has been heard; here is the conclusion of the matter:

Fear God and keep his commandments, for this is the whole duty of man.

Ecclesiastes 12:12-13

Publications arising from this thesis

Winter, A., W. Kramer, F. A. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G. Thaller and R. Fries (2002). "Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk fat content." Proceedings of the National Academy of Sciences U S A **20**: 20.

Winter, A., A. Alzinger and R. Fries (in press). "Assessment of the gene content of the chromosomal regions flanking bovine *DGAT1*." Genomics.

Winter, A., M. van Eckeveld, O. R. P. Bininda-Emonds, F. Habermann and R. Fries (in press). "Genomic organization of the *DGAT2*/*MGAT* gene family in cattle (*Bos taurus*) and other mammals." Cytogenetics and Genome Research.

Winter, A., F. Habermann and R. Fries (submitted). "Physical Mapping of *DGAT2*, *MGAT2* and *DGAT2* homologous gene in pig." Animal Genetics.

Kaupe, B., A. Winter, R. Fries and G. Erhardt (in press). "*DGAT1* polymorphism in *Bos indicus* and *Bos taurus* cattle breeds." Journal of Dairy Research.

Thaller, G., W. Kramer, A. Winter, B. Kaupe, G. Erhardt and R. Fries (2003). "Effects of *DGAT1* variants on milk production traits in German cattle breeds." Journal of Animal Science. **81**(8): 1911-1918.

Thaller, G., C. Kühn, A. Winter, G. Ewald, O. Bellmann, J. Wegner, H. Zühlke and R. Fries (in press). "*DGAT1*, a new positional and functional candidate gene for intramuscular fat deposition in cattle." Animal Genetics.

Contents

List of tables and figures

List of tables

List of figures

Abbreviations

A	alanine
A	adenosine
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
bp	base pair
BV	German Brown (Braunvieh)
C	cytosine
cDNA	copy deoxyribonucleic acid (cloned copies of mRNA)
cM	centi Morgan
cR_{3000}	centi Ray (number refer to radiation dose: 3 000 rads)
DC	DGAT candidate
DGAT	diacylglycerol acyltransferase
DNA	deoxyribonucleic acid
dNTP	nucleotides
DTT	dithiothreitol
EDTA	ethylendiamintetraacetat
EMBL	European Molecular Biology Laboratory
EST	expressed sequence tag
FV	German Simmental (Fleckvieh)
G	guanine
HF	German Holstein (Holstein-Friesian)
kb	kilo base pairs
mRNA	messenger ribonucleic acid
N	A, C, G, T, U
NCBI	National Center for Biotechnology Information
PCR	polymerase chain reaction
QTL	quantitative trait loci
RFLP	restriction fragment length polymorphism
RH	radiation hybrid
rpm	rounds per minute
SBE	single-base extension
SDS	sodium dodecylsulfat
SINE	short interspersed element
SNP	single nucleotide polymorphism
SSC	saline sodium citrate buffer
STS	sequence-tagged sites
TE	Tris EDTA buffer
TEMED	N', N', N', N', tetramethylethylendiamin
Tris	Tris (hydroxymethyl) aminomethane
VNTR	variable number of tandem repeats

1 Introduction and goals

In western countries, milk is an important agricultural product. In 1997, German dairies produced 27.7 million tons of milk, with milk and milk products representing 19% of consumed food (Bundeslandwirtschaftsministerium 1998). The profit for the dairy farmer is mainly a function of milk yield and, to a far lesser extent, of protein and fat content. Milk yield as well as milk components are subject to considerable interindividual variation within particular cattle breeds. In the German Simmental (Fleckvieh) population, milk fat percentage ranges between 2.8 and 5.6%. Milk fat percentage is a quantitative trait that is determined by the collective effect of multiple genes and environmental factors. The heritability (genetic contribution to the variation) of the milk fat percentage was estimated to be between 0.45 and 0.5 (Goddard *et al.* 1999). It is this genetic variability that is the basis for breeding. The aim is to concentrate as many of the positive gene variants as possible in one animal to improve its genetic potential. So far, selection has been based on observable phenotypes and applying sophisticated statistical analyses. However, selection based on phenotypic information is limited in this utility, particularly when the phenotype is expressed subsequent to reproductive age and confounded by long generation intervals, when individual has to be sacrificed to score its phenotype, or when the traits are expressed in only one gender (such as lactation traits). Present efforts aim to reveal the genetic architecture of quantitative traits through molecular genetics. Knowledge of the genes causing variation within a trait enable marker assisted selection (MAS): selection based on the presence or absence of genetic markers that are linked to desired or undesired characteristics. By applying MAS, animals can be tested for their genetic potential early in their development and independently from their gender.

Quantitative trait loci (QTL) are chromosomal positions delimited by genetic markers, with the marker alleles being associated with a measurable effect on a quantitative characteristic. Mapping of QTLs is a first step towards identifying genes that contribute to variation in quantitative traits. A second approach identifies functional candidate genes based on metabolic pathways. Diacylglycerol acyltransferase (DGAT) catalyzes the final step in triglyceride synthesis and was presumed to be rate limiting with respect to lipid metabolism (Mayorek *et al.* 1989). A study with knock-out mice lacking *diacylglycerol acyltransferase 1* (*DGAT1*) gene (Smith *et al.* 2000) emphasized *DGAT1* as a strong candidate gene for milk fat percentage. Surprisingly, *DGAT1*-deficient mice were viable, indicating the existence of alternative mechanisms and/or further genes for triglyceride synthesis. However, the crucial point was that the mice were not able to produce milk. This observation highlights the determining role of *DGAT1* in milk fat synthesis and milk production in general. *DGAT1* had been mapped to chromosome 15 in mice (Cases *et al.* 1998). Comparative mapping allowed the prediction that *DGAT1* was located on chromosome 8 in human and on chromosome 14 in cattle. The latter fell approximately within the region of a QTL for milk fat percentage and

other milk performance traits (e.g., Riquet *et al.* 1999). Thus, both functional and positional data made *DGAT1* a promising candidate gene for milk fat percentage in cattle.

The goals of this thesis were:

1. Investigation of the gene encoding *DGAT1* and the encompassing chromosomal region using:
- BAC clone isolation, physical mapping, sequence and structure analysis of bovine *DGAT1*
- Screening for sequence variation and testing genetic variances associated with the milk fat percentage by genotyping bulls with extremely high and low breeding values for milk fat percentage in cattle
- Generation of a bovine BAC contig of the *DGAT1* region
- Identification and mapping of *DGAT1* neighboring genes and screening them in cattle for sequence variations
- BAC clone isolation and physical mapping of porcine *DGAT1*

2. Optimization of the single base extension (SBE) assay for simultaneous genotyping of multiple single nucleotide polymorphisms (SNPs) at medium throughput for haplotype studies

3. Investigation of *DGAT2* gene family members using:
- BAC clone isolation, physical mapping, sequence and structure analysis of bovine *DGAT2* gene family members
- Screening for sequence variation and testing genetic variances associated with the milk fat percentage in cattle
- BAC-cloning and physical mapping of porcine *DGAT2* gene family members

2 Literature review

2.1 Synthesis of triglycerides

2.1.1 Pathways and enzymes of lipid synthesis

Biosynthesis of triglycerides occurs in the membrane of the endoplasmic reticulum and was reviewed recently by Lehner *et al.* (1996) and Coleman *et al.* (2000). Two pathways were described in mammals for *de novo* biosynthesis of triglycerides: the glycerol-3-phosphate pathway and the monoacylglycerol pathway (Figure 2.1). In both, diacylglycerol is synthesized and subsequently converted to triacylglycerol by diacylglycerol acyltransferase (DGAT). Additionally, triglycerides can be synthesized in mammals by the diacylglycerol transacylase pathway (Lehner *et al.* 1993). A third mechanism for the final step in triglyceride synthesis was described in plants and yeast using the enzyme phospholipid:diacylglycerol acyltransferase (PDAT), which can synthesize triacylglycerol from phospholipids and diacylglycerol (Dahlqvist *et al.* 2000). The pathways for triglyceride synthesis in mammals are described briefly.

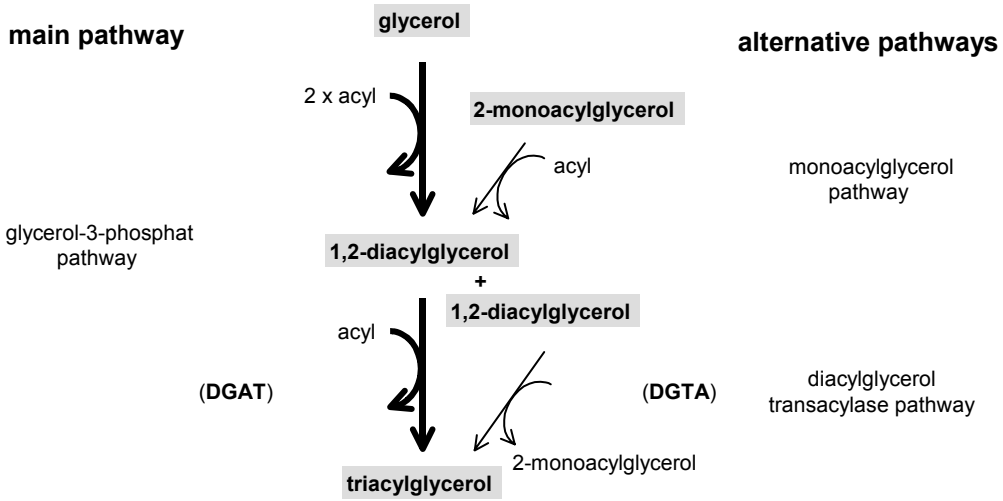


Figure 2.1: *De novo* biosynthesis of triglycerides in mammals.

Glycerol-3-phosphate pathway

Two major sources for glycerol-3-phosphate exists: (1) glycerol, either endogenous or from an extra-cellular source, is phosphorylated by glycerol kinase or (2) dihydroxyacetone phosphate, an intermediate of glycolysis, is reduced by glycerol-3-phosphate dehydrogenase. Glycerol-3-phosphate is stepwise acylated by glycerol-3-phosphate acyltransferase to 1-acyl-sn-glycero-3-phosphate (lysophosphatidate) and by acylglycerol-3-phosphate acyltransferase to 1,2-diacylglycerol-3-phosphate (phosphatidate). Phosphatidate is converted to 1,2-diacylglycerol in a reaction catalyzed by phosphatidate phosphatase (Figure 2.2).

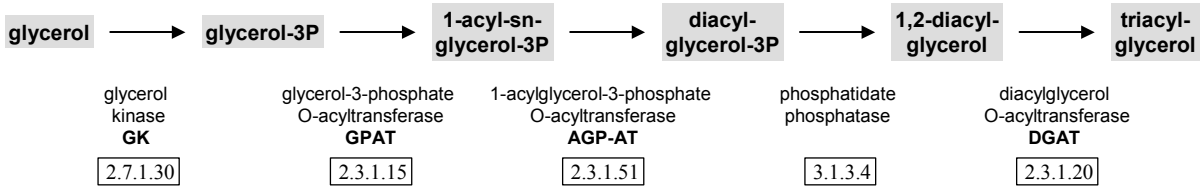


Figure 2.2: Glycerol-3-phosphate pathway.
Source: Kyoto Encyclopedia of Genes and Genomes, KEGG (Ogata *et al.* 1999).

Monoacylglycerol pathway

Dietary triglycerides are hydrolyzed in the intestinal lumen by pancreatic lipase to 2-monoacylglycerol and free fatty acids. Monoacylglycerol acyltransferase (MGAT) was reported in epithelial cells of intestine and liver (Coleman *et al.* 1985). This enzyme produces diacylglycerol by reacylation of diet-derived 2-monoacylglycerols.

Diacylglycerol acyltransferase (DGAT)

DGAT catalyses the final step in the triglyceride synthesis and may be the rate-limiting step in triglyceride synthesis (Mayorek *et al.* 1989). DGAT activity was identified in 1960 (Weiss *et al.* 1960). Two genes are known encoding DGAT activity at this time, *DGAT1* and *DGAT2* (see chapter 2.1.3).

Diacylglycerol transacylase pathway

Diacylglycerol transacylase synthesizes triacylglycerol by acyl-CoA independent transacylation between two 1,2-diacylglycerol molecules, resulting in one triacylglycerol and one 2-monoacylglycerol molecule. The 52 kDa enzyme was purified to homogeneity from intestinal microsomes. The activity of diacylglycerol transacylase was determined to be 15% of that of DGAT (Lehner *et al.* 1993).

2.1.2 Fat synthesis in the mammary gland

Lipid synthesis and secretion

Lipid secretion is one of five major types of secretion across the mammary secretory epithelium from the blood side to milk (Shennan *et al.* 2000):

- transmembrane secretion of directly from blood derived components like water and ions;
- exocytosis of components processed by the golgi apparatus like casein, whey proteins, lactose, citrate, and calcium;
- transcytosis of extra-alveolar proteins such as immunoglobulins, hormones and albumin from the interstitial space;
- paracellular route, the direct passage from interstitial fluid to milk (immunoglobulins in the presence of mastitis and during involution);
- milk fat route.

Milk fat secretion was reviewed by Mather *et al.* (1998). Lipids, synthesized in the endoplasmic reticulum, are packed into very small microlipid droplets (MLD, diameters of <0.5 µm). These micro droplets can fuse with one other and form cytoplasmic lipid droplets (CLDs, diameters of 1-5 µm). After moving to the apical pole of the cell, both MLDs and CLDs are released to the alveolar lumen by an exocytosis process. The mean diameter of lipid droplets in cow milk is approximately 4 µm (Mulder *et al.* 1974).

Milk fat composition

Total content of lipids in milk varies among species, ranging from 0% in rhinoceros, to 4% in humans and ruminants, to as much as 50% in pinnipeds and whales (Neville *et al.* 1997). Over 99% of the lipids in cow milk were found within droplets and triacylglycerol account for at least 97% (Mather *et al.* 1998). In addition to varying among species, the rate of secretion of milk lipids and fatty acid composition of milk varies within species, depending on factors such as lactating state, dietary composition, fasting and body lipid content (Neville *et al.* 1997). A characteristic feature of the triacylglycerols in the milk from ruminants is that they contain the short-chain acids butyric and hexanoic acid. Both short-chain acids are not found in the triacylglycerols from other tissues of ruminants and normally not in the milk of non-ruminants (Marshall *et al.* 1977). Marshall *et al.* (1977) present evidence that DGAT from a lactating cow mammary gland can utilize butyryl-CoA and hexanoyl-CoA. The hypothesis of a specific DGAT enzyme present exclusively in the mammary gland of ruminants was disproved by a subsequent study (Marshall *et al.* 1979), showing that DGAT isolated from bovine tissue other than mammary gland can also utilize short-chain fatty acids for triglyceride synthesis. It was assumed (Marshall *et al.* 1979) that the fatty acid composition at the sn-3 position of triacylglycerol (catalyzed by DGAT) is primarily a function of the composition of the intracellular acyl-CoA pool. It was further assumed that the intracellular concentration of butyryl-CoA and hexanoyl-CoA, relative to medium- and long-chain acyl-CoA, is much higher in ruminant mammary gland than in other mammals.
A study with *DGAT1* deficient mice (Smith *et al.* 2000) showed that the expression of *DGAT1* is not only necessary for lipid synthesis within the mammary glands, it is generally crucial for lactation, as *DGAT1*$^{-/-}$ mice produce no milk at all.

Milk fat precursor

Triglycerides are synthesized *de novo* in the endoplasmic reticulum of mammary secretory epithelia cells as described in chapter 2.1.1 from the precursors glycerol-3-phosphate, long-chain fatty acids drawn from the plasma, and middle- and short-chain fatty acids synthesized *de novo* within the mammary epithelial cell.

- Glycerol-3-phosphate is derived from either glucose or glycerol, both of which entered from the plasma. Glyceraldehyde-3-phosphate generated in the glycolytic chain can be converted to Glycerol-3-phosphate. Glycerol is phosphorylated to Glycerol-3-phosphate.
- Fatty acids drawn from the plasma contribute to approximately 50% of the total milk lipid synthesis in cows and exclusively for long-chain and unsaturated fatty acids. There are two

main extra cellular sources for fatty acids. First, triacylglycerol-rich lipoproteins (chylomicra or very low density lipoprotein (VLDL), coming mainly from the intestine) are hydrolyzed within the capillary lumen by the enzyme lipoprotein lipase (LPL) (Eckel 1989). The second source of fatty acids is important in fasting states: non-esterified fatty acids (NEFA), bound to albumin, are generated during lipolysis of adipose tissue (Neville *et al.* 1997). For the uptake of fatty acids, a protein that is located in the plasma membrane and termed as fatty acid translocator (FAT) may be important (Abumrad *et al.* 1993). Recombinant expression of FAT in fibroblasts enhances their ability to take up exogenous long-chain fatty acids (Ibrahimi *et al.* 1996). After transport into the cell, fatty acids are bound to fatty acid binding proteins (FABP, Glatz *et al.* 1996), which are likely responsible for maintaining a readily available fatty acid pool for TAG synthesis (Neville *et al.* 1997).

- *De novo* synthesis of fatty acids within the mammary epithelial cell produces chain lengths of less than 16 carbons (Barber *et al.* 1997). The major substrates are glucose, glycerol and ketone bodies, which are converted to acetyl-coenzyme A (CoA) by pyruvate and citrate synthesis and to malonyl-CoA by acetyl-CoA carboxylase (ACC). Ketones come primarily from ß-hydroxybutyrate that is produced in the rumen of cows. In addition, during early lactation, ketones are produced in excess by abnormal carbohydrate metabolism as a consequence of the activation of fatty acid oxidation during lipolysis, resulting in acetone, acetoacetate and ß-hydroxybutyrate. Fatty acids are formed by the stepwise elongation of acetyl-CoA by two carbons, which are derived from malonyl-CoA. One step represents one cycle comprised of seven reactions that are all catalyzed by a single multidomain enzyme, the fatty acid synthase (FAS, reviewed in Smith 1994).

2.1.3 DGAT1 and DGAT2 gene families

ACAT encoding genes

Diacylglycerol O-acyltransferase 1 (*DGAT1*) belongs to a gene family of three known members (Table 2.1). The other two represent acyl-CoA:cholesterol acyltransferases. Acyl-CoA:cholesterol acyltransferase (ACAT) forms cholesterol esters by joining cholesterol and fatty acyl-CoA. ACAT is an integral membrane enzyme located in the plasmatic reticulum and was discovered in 1950s (reviewed in Buhman *et al.* 2001). Chang and colleagues cloned *ACAT1* by functional complementation of mutant cells lacking ACAT activity (Chang *et al.* 1993). Murine *DGAT1* (Cases *et al.* 1998), human *ACAT2* and *DGAT1* (Oelkers *et al.* 1998) were isolated using the human *ACAT1* sequence to search EST databases.

Table 2.1: ACAT and DGAT1 encoding genes in human.

Symbol	Enzyme	Alternate symbols in human	Accession nr.	Chromosomal position
ACAT1	acyl-CoA:cholesterol acyltransferase	*SOAT1*	NM_003101	1q25
ACAT2	acyl-CoA:cholesterol acyltransferase	*SOAT2, ARGP2*	NM_003578	12q12
DGAT1	diacylglycerol O-acyltransferase	*DGAT1, ARGP1*	NM_012079	8qter

DGAT1 encoding gene

DGAT1 was predicted to be a membrane-bound protein with nine putative transmembrane domains, a diacylglycerol-binding signature sequence (at amino acids 382-392), one N-linked glycosylation site (amino acids 246-248) and one putative tyrosine phosphorylation site (amino acids 309-316) (Oelkers *et al.* 1998). In insect cells, DGAT1 utilized only diacylglycerol as an acyl acceptor (Cases *et al.* 1998). Expression of mRNA and activity of DGAT1 were ubiquitous in mouse and human tissues, with the highest levels in liver, small intestine, and adipose tissue (Cases *et al.* 1998; Oelkers *et al.* 1998; Farese *et al.* 2000; Smith *et al.* 2000). DGAT1 (Cheng *et al.* 2001) and ACAT1 (Yu *et al.* 1999) form homotetramers, with the subunits catalyzing the triglyceride synthesis independently (Cheng *et al.* 2001).
At this time, *DGAT1* has been identified in several species (Table 2.2).

Table 2.2: Species with isolated *DGAT1* cDNA.

Species			GenBank accession no.	Reference	Release date
Homo sapiens	human	mammal	AF059202	(Oelkers *et al.* 1998)	15.10.1998
Mus musculus	house mouse	mammal	AF078752	(Cases *et al.* 1998)	11.11.1998
Arabidopsis thaliana	thale cress	plant	AJ238008	(Hobbs *et al.* 1999; Zou *et al.* 1999)	18.06.1999
Nicotiana tabacum	tobacco	plant	AF129003	(Bouvier-Nave *et al.* 2000b)	22.12.1999
Brassica napus	rape seed	plant	AF251794	(Nykiforuk *et al.* 2002)	16.04.2000
Cercopithecus aethiops	african green monkey	mammal	AF236018	unpublished	12.08.2000
Rattus norvegicus	norway rat	mammal	AF296131	unpublished	03.09.2000
Perilla frutescens	shiso-zoku	plant	AF298815	unpublished	16.10.2000
Tropaeolum majus	nasturtium	plant	AY084052	unpublished	08.04.2002
Sus scrofa	pig	mammal	AY093657	(Nonneman *et al.* 2002)	30.04.2002
Drosophila melanogaster	fruit fly	insect	AF468649	(Buszczak *et al.* 2002)	07.05.2002
Caenorhabditis elegans	roundworm	nematode	AF221132	(Bouvier-Nave *et al.* 2000a)	22.05.2002
Olea europaea	olive	plant	no entry	(Giannoulia *et al.* 2000)	

DGAT2 gene family

Two polypeptides showing DGAT activity were isolated from the lipid bodies of the oleaginous fungus *Mortierella ramanniana*. The two polypeptides belonged to a new class of DGAT genes and were referred as *MrDGAT2A (*AF391089) and *MrDGAT2B (*AF391090) (Lardizabal *et al.* 2001). Orthologues were found in *Caenorhabditis elegans* (*U64852*, AF003384 and Z81557), *Arabidopsis thaliana* (AL133452) and *Saccharomyces cerevisiae* (*DGA1*, Z75153). In yeast, *DGA1* was the sole representative of the *DGAT2* gene family (Oelkers *et al.* 2002). In mammals, *DGAT2* and further genes with high sequence identity to *DGAT2* (referred as *DGAT candidates*, *DC*) were isolated in human and in mice (Cases *et al.* 2001). *DGAT2* was demonstrated to be a DGAT-encoding gene by expression studies in insect cells. *DGAT2* is expressed in many tissues with high expression levels in liver, white adipose tissue, mammary gland, testis, and peripheral blood leukocytes (Cases *et al.* 2001). Analysis of intestine from *DGAT1*$^{-/-}$ mice revealed that activity of DGAT2 apparently helps to compensate for the absence of DGAT1 (Buhman *et al.* 2002), with a residual DGAT activ-

ity of 10-15%. In adipose tissue membranes of *DGAT1*$^{-/-}$ mice, a residual DGAT activity of approximately 50% was observed (Cases *et al.* 2001). Recently, *DC2* and *DC5* were identified as genes encoding monoacylglycerol acyltransferase 1 (MGAT1, Yen *et al.* 2002) and monoacylglycerol acyltransferase 2 (MGAT2, Cao *et al.* 2003; Yen *et al.* 2003), respectively.

Excursion: Gene families

A gene family is a group of genes showing similarity in their nucleotide sequences. Members of gene families arise from an ancestral gene by gene duplication. Genes arising in this manner are called paralogues, whereas the term orthologues is used for homologous genes in different species descended from a common ancestor (Fitch 1970).

Duplicated genes contribute to genetic buffering, a robustness against genetic mutations, as shown in a simulation study (Lenski *et al.* 1999). However, more important for genetic buffering is the existence of redundant metabolic pathways (Kitami *et al.* 2002). An example of genetic buffering was *DGAT1* deficient mice (Smith *et al.* 2000), which were still able to produce triglycerides. Gene duplication is an important source of evolutionary novelty and adaptation (Ohno 1970). While one copy fulfills its function, the other gene copy is free for modifications by mutations, which occasionally leads to a gene with a new function (e.g., Zhang *et al.* 2002). In most cases, duplicated genes degenerate to pseudogenes: gene copies (including exons and introns) without function. In contrast, pseudogenes without introns are the result of retrotransposition, where a copy of the mRNA is integrated back to a random position to the genome. Three possible gene duplication processes have been described:

- duplication of the whole genome (polyploidization);
- tandem duplication of a single gene or of a chromosome segment, caused by unequal crossing over between homologous chromosomes in meiosis (Shen *et al.* 1981; Fitch *et al.* 1991) or by unequal exchanges between sister chromatids during mitosis;
- duplication by retrotransposons, which are DNA segments that can move from one place to another in the genome.

The hypothesis of two rounds of whole genome duplication (2R hypothesis) was first suggested by Ohno (1970). His hypothesis was supported by the presence of unlinked duplicated genes and differences in the amount of DNA per cell in vertebrates compared to their ancestors. As reviewed in Wolfe (2001), there are two approaches to prove the polyploidization hypothesis in vertebrates: the map-based approach and the tree-based approach. Comparisons of genetic maps among mammals (e.g. human and mouse, Nadeau *et al.* 1984)) confirmed that they contain large segments of conserved synteny groups. Support for the hypothesis is currently limited to the mapping of a small number of gene duplicates. Investigated gene families (Skrabanek *et al.* 1998) with known substitutes on four human chromosomes, thereby supporting the hypothesis, are the major histocompatibility complex (MHC) on chromosomes 1/6/9/19, the homeobox (hox) genes on chromosomes 2/7/12/17 and the fibroblast growth factor receptor (FGFR) genes on chromosomes 4/5/8/10. Progress in whole genome sequencing will launch comprehensive possibilities for further investigations. So far,

studies of the chromosomal locations of duplicated genes generally support the polyploidization hypothesis (Wolfe 2001). The second approach using phylogenetic trees would expect by assuming two rounds of genome duplications a gene tree topology of [(A,B)(C,D)], with similar distances for the branch points for different gene families. Hughes (1999) performed phylogenetic studies on seven protein families. The results indicated strong evidence against the 2R hypothesis.

Loss of duplicated genes and tandem duplication, as well as inter- and intrachromosomal rearrangements were thought to remove evidence for whole chromosome duplication (Nadeau *et al.* 1997). The rate of chromosomal rearrangements was estimated to be approximately one per one million years (Burt *et al.* 1999; Stanyon *et al.* 1999).

2.2 Analysis of quantitative traits

Most characters of individuals within a species are quantitative in nature, which is evident in the continuous distribution of their phenotypic measurement. The infinitesimal model (Fisher 1918) postulates an infinitesimal number of genes with small effects, so called polygenes, to explain the genetic contribution onto the phenotypic variation. Polygenes that are traceable by markers are "quantitative trait loci" (QTL, Geldermann 1975).

2.2.1 DNA markers

Searching for the genetic factors responsible for traits starts with the association of the trait with a chromosomal position represented by one or more polymorphic DNA markers. Alleles of the marker are co-inherited with alleles of the target gene (linkage makes recombination less likely). The first DNA markers were RFLPs (restriction enzyme length polymorphisms, Botstein *et al.* 1980). RFLPs are caused by single nucleotide polymorphisms (SNPs), which were difficult to genotype in the 80ies. Till these days, the markers of choice are extended DNA sequences such as minisatellites or VNTR (variable number of tandem repeats) markers (Wyman *et al.* 1980) and microsatellites or SSR (simple sequence repeats) markers (Litt *et al.* 1989; Weber *et al.* 1989). The number of tandem repeats in either can be up to 10 to 40 bp (minisatellites) and 1 to 4 bp (microsatellites). SNP are ideal DNA markers due to their relative high frequency and even distribution over the genome. However, only recently improvements in their detections have allowed their use as DNA markers. SNPs are bi-allelic marker; even though four alleles are conceivable, virtually only two alleles are present at a given SNP locus. In contrast, microsatellite markers appear in multiple alleles with variable numbers of the repeat unit.

2.2.2 Physical and genetic mapping

Gene maps display the locations of genes or markers by their chromosomal position. There are two different mapping methods: physical mapping and genetic (linkage) mapping.

Physical mapping

Physical mapping assigns genes or genetic markers to chromosomal regions either by:

- fluorescence *in situ* hybridization (FISH, Pinkel *et al.* 1986), utilizing fluorescence labeled DNA probes of target sequence that hybridize to complementary DNA on chromosomal spreads;
- radiation hybrid mapping (Cox *et al.* 1990), see chapter Methods 3.2;

and finally with the highest resolution by

- the localization within the genomic sequence, which is now feasible due to progress in whole genome sequencing (e.g. human, Lander *et al.* 2001; Venter *et al.* 2001).

Genetic mapping

A linkage or genetic map provides the relative positions of markers or genes on a chromosome, determined based on how often alleles are inherited together depending on the recombination frequencies between the corresponding loci. To detect recombination between two loci, the parents must have distinguishable alleles at each locus and the offspring must be informative, i.e. alleles present in the offspring can be assigned to either of the parents. The recombination rate depends on the physical distance between the loci, as the distances increase, so does the recombination rate. The genetic distance between two loci is the expected number of recombinations between them. A standard distance of 1 Morgan (M) means an average of one crossover per meiosis between the loci. Map functions are used to transform recombination rates into genetic distances. For closely linked loci, the map function of Morgan can be used (Sturtevant 1913; Morgan 1928), which sets the genetic distance equal to the recombination rate. For larger distances, there is no linear relationship between recombination rate and genetic distance because multiple crossovers can occur between distant loci. The map function of Haldane (1919) allows double crossovers between two loci and the map function of Kosambi (1944) adjusts the map distance based on interference, which changes the proportion of double crossovers.

Genetic distances can be added up in contrast to recombination rates. The total length of the bovine genetic map is estimated to be approximately 30 Morgan (Kappes *et al.* 1997). Comprehensive cattle genetic linkage maps based on microsatellite markers were constructed by Barendse *et al.* (1994), where 36 linkage groups are related to 30 chromosomes, and by Bishop *et al.* (1994), where 30 linkage groups are related to 25 chromosomes. Maps that are more detailed were reported by Ma *et al.* (1996), Barendse *et al.* (1997) and Kappes *et al.* (1997).

Comparative gene maps

Comparative mapping is based on conservation of synteny: genes found together in one species within a chromosomal region are also found together in another species. Results of comparative mapping can be used for evolutionary studies (O'Brien *et al.* 1999) and for using well-characterized model organisms to predict the chromosomal gene position within non-

model species. The latter one is used to identify candidate genes within QTL regions. Several approaches have been used to establish comparative maps between species. In interspecies chromosome painting (Zoo-FISH), chromosome-specific DNA probes from one species are hybridized to the chromosomes of another species. Whole genome comparative maps have been generated by this method for human - cattle (Hayes 1995; Solinas-Toldo *et al.* 1995; Chowdhary *et al.* 1996) and for human - pig (Goureau *et al.* 1996; Fronicke *et al.* 2001). Further, radiation hybrid (RH) mapping is used to map orthologous genes within two species; For example, the whole genome comparative map for human - cattle (Band *et al.* 2000). Progress in whole genome sequencing of different species (e.g. human, Lander *et al.* 2001; Venter *et al.* 2001; fugu, Aparicio *et al.* 2002; and mouse, Waterston *et al.* 2002) will allow direct comparison of the genomic sequences.

2.2.3 Approaches to the mapping of quantitative trait loci (QTL)

The first step in QTL mapping is a genome scan by linkage analysis. QTLs are detected by co-segregation of linked polymorphic markers in a well-characterized pedigree. Parents are assumed to be heterozygous at the QTL and offspring inherit alternate alleles traceable by linked markers. The procedure is to test for differences in trait means between groups of offspring having inherited the same marker alleles for each marker in turn (Mackay 2001a). Because the recombination rate increases with the distance between marker and QTL, lower differences in the trait means will result between offspring groups that have inherited the same marker.

Several statistical methods have been developed to test if a chromosomal region is associated with an effect on the trait (e.g. interval mapping, Lander *et al.* 1989; or maximum likelihood method, Mackinnon *et al.* 1995). In dairy cattle, two experimental designs are mainly employed: the daughter design and the granddaughter design (Weller *et al.* 1990). Both designs use the large half-sib family structures in cattle populations, created through the use of artificial insemination, and also the large amount of phenotypic information in the form of production traits, routinely collected in form of the breeding value or the daughter yield deviation (DYD, VanRaden *et al.* 1991).

The daughter design requires marker information and phenotypic data from daughters having a common sire. In the granddaughter design, sons of a common sire are genotyped and the phenotypic information of the sons is provided by large number of daughters (granddaughters of the sire). The advantage of the granddaughter design compared to the daughter design is (a) higher efficiency in genotyping (i.e. fewer individuals need to be sampled) and (b) the availability of DNA samples (i.e. females are not normally genotyped). Disadvantage is the loss of a direct connection between geno- and phenotype.

Linkage mapping is used as an initial step to reveal chromosome regions harboring QTLs, but the resolution is limited (e.g. to an average marker interval of 19 cM, Coppieters *et al.* 1998). A reason for the limited resolution is that there is only one generation for recombination to occur between closely linked markers during meiosis (Darvasi *et al.* 1993). Linkage disequi-

librium (LD) or association mapping is used for fine mapping, based on inherited historical recombinations that are accumulated over a long time, namely since the causal mutation happened.

The final goal is to identify the gene representing the QTL and subsequently to identify the polymorphic site within the gene causal for the differences in the trait phenotype - the quantitative trait nucleotides (QTNs, Mackay 2001a). The resolution of QTL mapping in natural populations is limited to intervals resistant to recombination, namely haplotype blocks, which can contain several genes. Daly *et al.* (2001) identified haplotype blocks in humans with sizes ranging from 3 to 92 kb. Functional candidate genes located within the QTL interval, mapped directly within the species or predicted through comparative mapping, are investigated for putative polymorphisms causing the phenotypic effect.

2.2.4 Quantitative trait loci (QTL) for milk fat in cattle

A QTL mapping strategy in combination with the daughter and the granddaughter designs supplies a powerful tool to identify trait-associated chromosomal locations in cattle. The first genome scan QTL mapping study applying a granddaughter design was conducted by Georges *et al.* (1995) in US Holstein cattle. Results of whole genome scans using the granddaughter design for milk fat traits (milk fat yield and milk fat percentage) are shown in Table 2.3.

Table 2.3: QTL mapping results for milk fat traits in cattle.

Reference	Population	Investigated traits	Results for milk fat[a]	Chromosome	Position	Marker
(Georges *et al.* 1995)	US Holstein	production traits	FY	1		
			FY	9		
			FY	10		
(Zhang *et al.* 1998a)	US Holstein	production traits	FP	6	11 cM	TGLA37
			FP	20	28 cM	TGLA153
			less significance:			
			FY	9	89 cM	TGLA73
			FY + FP	14	0 cM	ILSTS11
			FY	23	66 cM	MGTG7
			FP	26	15 cM	TGLA22
			FP	28	12 cM	TGLA82
(Coppieters *et al.* 1998)	Dutch and New Zealand Holstein	milk production traits	FP[b]	14	0 cM	CSSM66
(Heyen *et al.* 1999)	US Holstein	production and health traits	FP	3	22 cM	ILSTS96
			FP	14	2 cM	ILSTS39
(Ashwell *et al.* 2001)	US Holstein	production and health traits	FY+FP	14	6.2 cM	BMS1678

[a]FY, milk fat yield; FP, milk fat percentage

[b]Additional for milk yield and composition

As shown in Table 2.3, several genome scans revealed a QTL on the centromeric end of bovine chromosome 14 with a strong effect on milk fat yield and percentage, as well as for milk yield and milk composition (data not shown in the table). The centromeric end of bovine chromosome 14 is known to correspond to the telomeric end of human chromosome 8 (Barendse *et al.* 1994; Solinas-Toldo *et al.* 1995).

Following the genome scan, fine mapping studies of bovine chromosome 14 were continued. In Riquet *et al.* (1999), the chromosome segment harboring the QTL was narrowed to less than 9.5 cM (5 cM) flanked by the closest non-identical-by-state markers ILSTS039 and BULGE004. In a second study (Farnir *et al.* 2002), the interval was refined to a 3 cM segment flanked by the markers BULGE09 and BULGE11, proximal to the interval described in Riquet *et al.* (1999). In another study (Looft *et al.* 2001), 12 ESTs derived from mammary gland tissue of lactating cows were used as candidate genes for QTLs affecting milk production traits. For KIEL_E8, mapped to the centromere of bovine chromosome 14 and homologous to the CRH gene in mouse, linkage disequilibrium was observed between the positional candidate KIEL_E8 and the segregating QTL-alleles.

2.3 SNP genotyping

Various technologies have been established to genotype SNPs. The different methods, of which there is no clear favorite, have different properties. The most appropriate method has to be ascertained on a case by case basis. Decision criteria are the expected throughput and the experimental design: for example, a definite or variable set of SNPs, or the relative number of SNPs (few SNPs and high number of samples, or a high number of SNPs and few samples). Further decision criteria are initial and running costs for a method. In most cases, the existing facilities are deciding: the methods are adapted to the devices already present in the lab. SNP detection starts with the amplification of a target sequence by PCR to increase the copy number of the target for more sensitive and specific detection. The core of SNP genotyping is the analytical biochemical techniques to distinguish between the two alleles: allele-specific enzyme reactions, hybridization probes that rely on differences in hybridization, and combinations of both. The product of the analytical techniques is detected by appropriate device. A short overview about SNP detection is given by Grant and Phillips (2001), a comprehensive review by Kwok (2001) and a review about automated genotyping is given by Gut (2001).

Principles of analytical biochemical techniques

Differential hybridization of PCR fragments and DNA probes without enzymatic reaction.
Allele-specific amplification (ASA) or amplification refractory mutation system (ARMS).
Primer extension with allele-specific nucleotide incorporation: sequencing, minisequencing (Pyrosequencing, Ronaghi *et al.* 1996), or single base extension (SBE) also called template-directed dye-terminator incorporation (TDI)
Allele-specific DNA cleavage: restriction enzymes in RFLP assay, glycosylase mediated polymorphism detection (GMPD), flap endonuclease in Invader assay (Harrington *et al.* 1994) or 5'-3' exonuclease activity of Taq DNA polymerase (Holland *et al.* 1991) in TaqMan assay (termed TaqMan because of the analogy to the video game PacMan).
Allele-specific ligation assays: oligos (Landegren *et al.* 1988) or padlock probes (Nilsson *et al.* 1994).

Detection mechanisms

The analytical biochemical techniques have to convert the genetic information in the two alleles to a technically detectable value, in general in an allele-specific mass (mass spectrometry of DNA products) or an allele-specific emission of light (fluorescence or luminescence of molecular labels). **Direct fluorescence detection** of fluorophores linked to nucleotides or probes need a separation step such as gel separation or wash steps by solid-phase reactions. Formation of product can be monitored without separation using intercalating dyes like syber green, which emits fluorescence only in the presence of double-stranded DNA. Two further physical phenomena are utilized for real-time monitoring without separation of products from unincorporated dyes (homogeneous assay): fluorescence resonance energy transfer and fluorescence polarization. By **Fluorescence resonance energy transfer** (FRET, Foster 1965), the electrically excited donor molecule (reporter) transfers energy to an acceptor molecule (quencher) without emission of a photon. Quenching is highly efficient within the Förster radius of the donor/acceptor pair (which is often in the 50 - 60 Å range). Outside of this distance, quenching efficiency falls off rapidly, decreasing by the inverse sixth power of the intermolecular separation.
Designs of probes using FRET vary from linear probes in 5'-3' exonuclease assays, oligo ligation assays, circular padlock probes, hairpin-forming molecules like scorpion primers for self-probing amplicons (Whitcombe *et al.* 1999), and molecular beacons (Tyagi *et al.* 1996; Kostrikis *et al.* 1998; Tyagi *et al.* 1998), which are hybridization probes with a fluorescent reporter group at one end and a fluorescence quencher group at the other end. In absence of a target, the molecule forms an internal hairpin resulting in quenched reporter fluorescence. In the presence of target, the probe molecule unfolds and hybridizes with it. Reporter and quencher are now separated and the reporter dye will emit fluorescence signal upon stimulation.

Secondly, **fluorescence polarization** (FP, Perrin 1926) enables the distinction between unincorporated fluorescence labeled small molecules like single nucleotides, and fluorescence labeled larger molecules like single nucleotides linked to oligo-nucleotides. Fluorophores are excited by plane-polarized light. A fluorophore linked to a small molecule has faster motion. This leads to a rotation of the molecule between excitation and emission and subsequently to a depolarized emission. A fluorophore linked to a large molecule has slower motion and emission remains polarized in a fixed plane.
In Pyrosequencing (Ronaghi *et al.* 1996), **luminescence** is emitted in an ATP-dependent luciferase reaction. The incorporation of a deoxynucleotide triphosphate is accompanied by the release of pyrophosphate (PPi). ATP sulfurylase converts PPi to ATP in the presence of adenosine 5´ phosphosulfate. This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light (Pyrosequencing).

Methods for distinction and detection of assay products

Gel separation by product size and fluorescence detection (horizontal gel electrophoresis, plate and capillary sequencer).
Mass spectrometry, distinction by product mass using MALDI-TOF (matrix-assisted laser desorption/ionization time-of-flight).
Microarrays are DNA or oligo nucleotides spotted on a glass surface, distinction by position and fluorescence detection.
Flow cytometry, similar to microarrays, with DNA or oligo nucleotides bound to fluorescence-labeled beads. Two unique fluorescent dyes at 10 different concentrations provide a set of 100 distinguishable entities. A third dye is used to identify the allele information of a SNP.
Plate-reader and integrated in thermocycler for fluorescent detection.

Currently there are several commercially available methods for SNP detection by single base extension:

- Capillary DNA sequencing platforms (SNaPshot by Applied Biosystems)
- MALDI – TOF mass spectrometry (MassARRAY by Sequenom)
- DNA microarray (e.g. GenFlex by Affymetrix)
- Bead-based technology (Luminex)
- ELISA-style microtiter plate formats with colorimetric detection (SNP-IT by Orchid BioSciences)
- Fluorescence polarization detection systems (HEFP: High Efficiency Fluorescence Polarization by Molecular Devices, formerly LJL Biosystems)

3 Materials and Methods

3.1 Acquiring and processing sequence information using online resources

3.1.1 Searching NCBI resources for existing sequence information

Human mRNA sequence of *DGAT1* was used to BLAST (Basic Local Alignment Search Tool) search (Altschul *et al.* 1990) the EST division (dbEST) of GenBank (Boguski *et al.* 1993). ESTs (expressed sequence tags) are generated by partial sequencing of randomly selected cDNA clones (Adams *et al.* 1991). For the *DGAT2* gene family, in addition to the standard nucleotide BLAST (blastn), translated BLAST (tblastx) using human mRNA sequence of *DGAT2* was employed to detect members of the gene family not described in humans and mouse (Cases *et al.* 2001). For a description of BLAST services at the NCBI (National Center for Biotechnology Information), see http://www.ncbi.nlm.nih.gov/blast/blast_help.html. Briefly, standard nucleotide BLAST compares a given nucleotide sequence against the nucleotide database, whereas translated BLAST converts the nucleotide query sequence into protein sequences for all reading frames and compares them to the nucleotide database that has similarly been translated in all reading frames.

To obtain bovine sequence information for genes neighboring *DGAT1*, BLAST searches were performed for all human genes that were listed in the human draft sequence (NCBI MapView build 28) to fall in a range of about 320 kb before and after *DGAT1* on human chromosome 8q24.3. NCBI MapView build 28 of the human draft was based on sequence information available on December 24, 2001.

3.1.2 Editing of obtained sequences for primer design

Bovine EST sequences of each identified gene were assembled to yield consensus cDNA sequences via pairwise BLAST search of NCBI (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html). The consensus cDNA sequences were aligned with the corresponding human mRNA sequences using PileUp and PrettyBox of the GCG software package (Genetics Computer Group 2001). Putative splice sites were derived from human and mouse, assuming conserved exon/intron structure between the orthologous genes. Human splice sites and human intron sizes were used to design PCR primers. Sources of human splice sites were the NCBI EvidenceViewer and the Ensemble Genome Browser of EMBL (http://www.ensembl.org/Homo_sapiens/). The Evidence Viewer was accessed over the NCBI LocusLink interface: http://www.ncbi.nlm.nih.gov/LocusLink/ (Pruitt *et al.* 2001). Before the Evidence Viewer and Ensemble Genome Browser became available in 2001, the human mRNA sequence was compared with the human draft sequence by pairwise BLAST.

No bovine EST sequence was available for exon 1 of *DGAT1*, exon 4 and 5 of *DGAT candidate 2* (*DC2*), and entire *DGAT candidate 5* (*DC5*). In these cases, primers for cattle were

designed using human mRNA or porcine EST sequences by considering highly conserved regions between species such as *Homo sapiens, Mus musculus, Rattus norwegicus and Caenorhabditis elegans*. Primers that were not designed from bovine sequence are labeled with "h" for human and "p" for porcine in Appendices 9.4 (Primers used for direct sequencing BAC DNA) and 9.5 (PCR primer).

3.1.3 Primer design

Primers were designed using the Primer3 software available at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi (Rozen *et al.* 1998). The estimated melting temperatures of primers were between 59 and 62°C with the optimum at 60°C. The optimal primer size was 20 bases, with a range from 15 to 29 bases. Primers for direct sequencing of BAC DNA preferably contained a GC-clamp of one base at the primer 3'end.

3.2 RH mapping of DGAT1 in human (Genbridge 4)

Radiation hybrid (RH) mapping provides a way to localize a known sequence to a map position in the genome by performing PCR with DNA from a collection of hybrid cell lines (RH panel) as a template. Hybrid cells are generated by fusing X-irradiated cells of the species of interest with rodent cells. X-irradiation causes chromosome fragmentation, whereupon hybrid cells lose most of these chromosome fragments until they carry a stable random set of fragments. A PCR pattern displays the presence or absence of a locus within each hybrid cell line of the panel, with the similarity across the PCR pattern for two loci being a measure of the physical distance between the. The unit of radiation mapping is centiRay (cR), which depends on the intensity of the irradiation. A distance of 1 cR_{5000} represents 1% frequency of breakage between two markers after exposure to a dose of 5000 rad.

DGAT1 was located by RH mapping within the human GENMAP '98 RH map (Deloukas *et al.* 1998). PCR was performed on 84 clones of Genbridge 4 RH panel (Gyapay *et al.* 1996) (HGMP Resource Center, Hinxton, UK) with primers specific to human *DGAT1* (5'-GAG GCCTCTCTGCCCTATG-3', 5'-TTTATTGACACCCTCGGACC-3') under the following conditions: 10 µL total volume containing 25 ng DNA, 0.5 µM of each primer, 200 µM of each dNTP, 1 µL 10x PCR reaction buffer, 1.5 mM $MgCl_2$ and 0.5 units AmpliTaq polymerase (Perkin Elmer Applied Biosystems Division, Foster City, CA, USA). The reactions were amplified in a T-Gradient Thermocycler (Biometra, Göttingen, Germany) under following conditions: initial denaturing at 94°C for 3 minutes; followed by 30 cycles at 95°C for 30 seconds, 60°C for 1 minute, 72°C for 1 minute; and final extension at 72°C for 10 minutes. PCR products were separated on a 2% agarose gel and PCR results were analyzed using the Sanger Center RH server (http://www.sanger.ac.uk/Software/RHserver/Rhserver.shtml).

3.3 Screening of BAC-Libraries

Two BAC libraries were screened, which were supplied by the Children's Hospital Oakland Research Institute (BACPAC Resources, Oakland, CA, USA): male bovine BAC library

RPCI-42 (Warren *et al.* 2000) and male porcine BAC library RPCI-44 (Fahrenkrug *et al.* 2001). RPCI-44 BAC library was constructed from Holstein bull white blood cells and RPCI-44 BAC Library was constructed from pooled pig white blood cells. Partially *Eco*RI digested and size-selected DNA was cloned between the *Eco*RI sites of the pBACe3.6 vector (bovine) and the pTARBAC2 vector (porcine). Ligation products were transformed into DH10B electrocompetent cells. For the characterization of used libraries, see Table 3.1. The BAC libraries were gridded onto 22 x 22 cm positively charged nylon filters for hybridization screening with radiolabeled DNA probes.

Table 3.1: BAC libraries used in this study.

Library	Segment	Cloning vector	DNA	Total plates	Total clones	Average insert size	Genomic coverage
RPCI-42	1	pBACe3.6	Holstein Bull White Blood Cell	288	108 776	165 kb	6.0x
	2	pBACe3.6	Holstein Bull White Blood Cell	288	107 663	163 kb	5.9x
	1 + 2			576	216 439	164 kb	11.9x
RPCI-44	1	pTARBAC2	Pig[a] White Blood Cell	240	83 946	157 kb	4.4x
	2	pTARBAC2	Pig[a] White Blood Cell	288	101 443	171 kb	5.8x
	1 + 2			528	185 389	165 kb	10.2x

[a]From four male pigs that were 1/4 Meishan, 3/8 Yorkshire, and 3/8 Landrace.

Generation of radiolabeled PCR probes

PCR probes were used for BAC library screening (Table 3.2), with PCR performed as described in chapter 3.6.1. PCR products were purified from primer and nucleotides using QIAquick PCR purification kit (28106; Qiagen, Hilden, Germany).

Table 3.2: BAC library screening.

Probe specific to	Forward primer[a]	Reverse primer[a]	Product [bp]	Library	Segment	Signals	Positive clones[b]
Bovine *DGAT1*	1599	1601	565	RPCI-42	1	4	4
Bovine *DGAT2*	1897	1898	807	RPCI-42	1	6	6
Bovine *DC2*	1904	1905	347	RPCI-42	2	6	6
Bovine *DC5*	1906	1908	422	RPCI-42	1	7	5
Bovine BAC ends of clones containing *DGAT1*				RPCI-42	2	20	10
56F1-T7	1691	2405	988				
240A1-SP6	1688	2407	885				
56F1-SP6	1686	2404	854				
240A1-T7	1689	2406	834				
269H17-SP6	1963	1964	ca. 800				
Bovine genes neighboring *DGAT1*				RPCI-42	1	12	5
RECQL4	2430	2432	ca. 800				
KIAA0496	2454	2457	ca. 1100				
FOXH1	2450	2453	ca. 1000				
MGC13010	2438	2441	ca. 1100				
GPT	2442	2445	ca. 1000				
Porcine *DGAT1*	1915	1916	ca. 550	RPCI-44	2	4	3
Porcine *DGAT2*	1897	1898	ca. 850	RPCI-44	1	9	8
Porcine *DC5*	1906	1908	ca. 400	RPCI-44	1	9	3
Porcine *DC7*	1909	1911	ca. 400	RPCI-44	2	6	2

[a]For primer sequences see Appendix 9.5 (PCR primers).
[b]Confirmed by Colony-PCR

To isolate BAC clones for the BAC contig, a protocol with pooled PCR probes was used to reduce the screening effort. One pool was composed of five probes specific to the BAC ends of clones containing *DGAT1* that were isolated up to that time, and a second pool was composed of probes specific to five genes neighboring *DGAT1* (Table 3.2).
PCR products were radiolabeled using Megaprime DNA labeling system (RPN1604; Amersham Biosciences, Freiburg, Germany) and deoxyadenosine 5'-(α-^{32}P) triphosphate (AA0004-250µCi; Amersham Biosciences, Freiburg, Germany). First, 40 ng of PCR product (pooled probes: 5 x 40 ng) was adjusted to a volume of 21 µL and then denatured at 99°C for five minutes together with 5 µL of primer mix (random nanomer primers). After chilling on ice, 5 µL 10x reaction buffer; 4 µL each of dGTP, dCTP and dTTP; 2 µL Klenow fragment (1 unit/µL); and 5 µL dATP^{32} (1.85 kBq) were added to the reaction. The reaction was incubated at 37°C for 15 minutes and then stopped by adding 10 µL of 0.2 M EDTA (pH 8.0). Radiolabeled probes were denatured at 95°C for seven minutes.

Dot-Blot as positive control

Two µL of each PCR product, both undiluted and ten-fold diluted, were blotted onto nylon membranes (Hybond-N+, PRN303B; Amersham Biosciences, Freiburg, Germany) that were saturated with 0.4 N NaOH. After five minutes, the dot-plot was shacked moderately for one minute in 5x SSC buffer.

Hybridization

The filters of the BAC library were applied to roller bottles together with dot-blot controls and were prehybridized at 67°C: two filters per bottle with 10-20 mL Church buffer (Church *et al.* 1984): 5% SDS, 1 mM EDTA, 0.341 M Na_2HPO_4 and 0.159 M NaH_2PO_4. After 30 minutes, 17 µL of radiolabeled probe were added to each bottle and hybridized in 10 mL Church buffer at 67°C overnight (14-16 hours). Filters were washed twice in 2x SSC for 20 minutes at 63°C and a third time in 0.5x SSC + 0.1% SDS (20 minutes at 63°C). Finally, each filter was rinsed in 2x SSC, wrapped in household plastic film and placed together with the medical X-ray film NewRX (03E220; FUJIFILM Medical Systems, Stamford, CT, USA) in cassettes. X-ray films were exposed for five hours at –80°C. Following the library documentation, positive clones present on the developed autoradiograms (Figure 3.1) were traced back to the location on the original 384 well plate from which they were gridded. Assigned clone names represent the plate number, row and column in the respective library. Colony-PCR with primers used for screening was applied to test if the clones contained the right insert (see chapter 3.8.1).

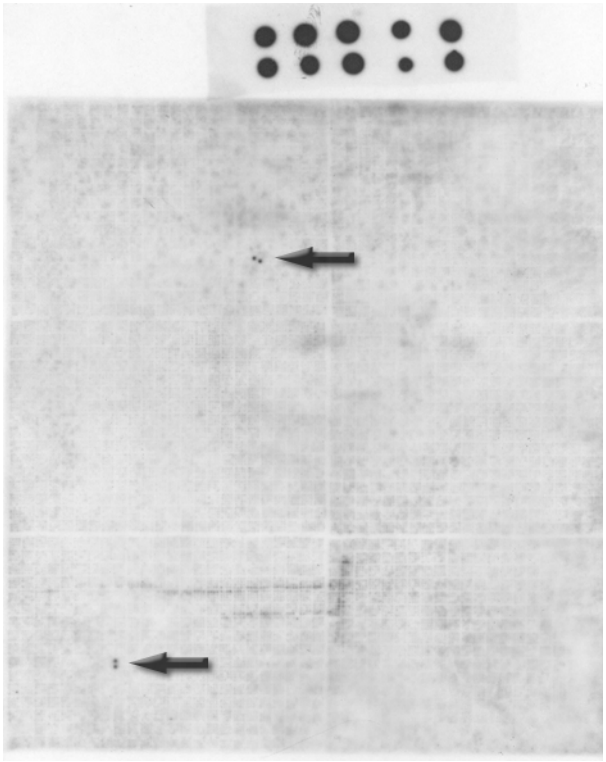


Figure 3.1: Autoradiogram of BAC library screening.
Pooled BAC ends specific probes hybridized to filter 7L of RPCI-42, segment 2 and dot-blot (on top). Arrows indicate positive signals (upper: 334E6, lower: 293G16).

3.4 Preparation of BAC DNA

Each clone was streaked out on luria broth (LB) agar plates containing 12.5 µg/ml chloramphenicol (0634433; Roche Diagnostics, Mannheim, Germany) and incubated overnight at 37°C. A starter culture of 3 mL LB medium (12.5 µg/mL chloramphenicol) was inoculated with a single colony and incubated for 8 hours at 37°C with vigorous shaking (300 rpm). A culture of 500 mL LB medium (12.5 µg/mL chloramphenicol) was inoculated with 500 µL of the starter culture and incubated for 14 h at 37°C with shaking at 300 rpm. After centrifugation, BAC DNA was prepared from the cell pellet using Qiagen Plasmid Midi Kit (12145; Qiagen, Hilden, Germany) or Qiagen Large-Construct Kit (12462; Qiagen, Hilden, Germany). The latter is designed to isolate large DNA constructs as BAC DNA with low copy number from clones. The protocols start with alkaline lysis (Birnboim *et al.* 1979) of the cells. The released BAC DNA is bound to Qiagen resins under low salt and pH conditions. After a wash step with medium-salt, the pure BAC DNA is eluted in high-salt buffer. The Large-Construct Kit was used to isolate BAC DNA from clones containing bovine *DGAT1* following the supplied protocol. For all other clones, the Qiagen Plasmid Midi Kit was used with a modified protocol: a starter culture of 500 µL was transferred into 100 mL LB medium containing chloramphenicol (12.5 µL/mL) and incubated at 37°C. After 14 hours, the pellet of 50 mL of the culture medium was lysed following the supplied protocol with 12 mL of each of the buffer P1, P2 and P3. After centrifugation of the lysate (20 000 g, 30 min) to remove precipitated material, the supernatant was filtered over a prewetted folded filter. DNA was precipitated by adding 20 mL of isopropanol and centrifuging at 5290 G for 60 minutes. After the DNA pellet was redissolved in 500 µL of 10 mM Tris-Cl buffer (pH 8.0), 4.5 mL of

buffer QBT was added and the whole sample was applied to an equilibrated Qiagen resin. Samples were washed twice by adding 10 mL of buffer QC. BAC DNA was eluted with 5 x 1 mL of buffer QF, which were preheated to 70°C for better recovering large DNA molecules. DNA was precipitated in 3.5 mL of isopropanol, centrifuged at 5290 G for 60 minutes and redissolved in 500 µL of TE buffer. The redissolved BAC DNA was desalted using Microcon YM-100 filters (42413, Millipore, Eschborn, Germany). Samples were centrifuged at 450 G for 15 minutes, and for another 10 minutes after adding 250 µL Tris buffer (10 mM, pH 8). Before recovering the BAC DNA by upside down centrifugation of the filters for one minute, 50 µL of TE buffer was added.
The concentration of BAC DNA was measured with a fluorometer (DyNA Quant 200; Hoefer Pharmacia Biotech Inc., San Francisco, CA, USA). To test both the quality and the quantity of the BAC DNA, 1µL of it was applied to 0.8% agarose gel and compared against lambda DNA (SD0011; MBI Fermentas, St. Leon-Rot, Germany).

3.5 Selection and preparation of DNA samples

3.5.1 Selection of bulls with extreme breeding values

To test whether the variance of milk fat content is associated with alleles of the *DGAT1* gene, bulls used for artificial insemination (German Holstein, German Simmental, and German Brown) with high (+) and low (-) breeding values for milk fat percentage were selected (Table 3.3). Breeding values were evaluated by INTERBULL (http://www.interbull.org) using Best Linear Unbiased Prediction (Henderson 1974). For the average milking merit of all German cows belonging to the breeds German Holstein, German Simmental, German Brown and Jersey, see Table 3.4.

Table 3.3: DNA pools containing bulls used for artificial insemination with extreme breeding values for milk fat percentage (BVF).

	Population			Pools				
Breed	Year of birth	Number	BVF mean	Name	Number	BVF min	BVF max	BVF mean
German Holstein	1988 or later	2857	-0.148 (±0.284)	HF32+	32	0.48	1.08	0.622 ± 0.125
				HF32-	32	-0.92	-0.68	-0.771 ± 0.063
German Simmental	1990 or later	4070	+0.089 (±0.217)	FV32+	32	0.56	0.83	0.683 ± 0.153
				FV32-	32	-0.51	-0.31	-0.454 ± 0.061
German Brown	1990 or later	656	+0.006 (±0.185)	BV20+	20	0.29	0.73	0.424 ± 0.156
				BV20-	20	-0.40	-0.22	-0.317 ± 0.096

For composition of the pools see appendix 9.2.

Table 3.4: Average milk merit of four cattle breeds in Germany (ADR 2002).

Breed	Number of cows	Age	Milk	Fat		Protein	
		[Years]	[kg]	[%]	[kg]	[%]	[kg]
German Holstein	1 509 457	4.8	7 988	4.20	336	3.41	272
German Simmental	657 277	4.9	6 430	4.15	267	2.50	225
German Brown	164 521	5.4	6 626	4.20	278	3.60	239
Jersey	2 169	5.3	5 125	5.98	306	4.12	211

To reduce the amount of genotyping, an approach of selective DNA pooling was used (Darvasi *et al.* 1994), with the pooled DNA samples representing the extreme high and low phenotypic groups of the population (see Appendix 9.2). In addition to samples with extreme breeding values, DNA samples of randomly selected individuals of different *Bos taurus* and *Bos indicus* cattle breeds were used for SNP detection and SNP genotyping to determine allele frequencies in those breeds (Table 3.5). Additional DNA samples were from a German Simmental granddaughter design (progeny of 20 sires) and two each from yak and water buffalo.

Table 3.5: Breeds and numbers of investigated individual animals.

Species		Breeds	Numbers of DNA samples
Bos taurus taurus	(cattle)	Anatolian Black	50
		Angus	1
		German Brown	56
		German Simmental	115
		German Simmental progeny of 20 sires	800
		German Simmental sires[a]	16
		German Yellow	3
		German Holstein	47
		Jersey	7
		Kerry	1
		Original Brown	8
		Pinzgauer	43
		Original Simmental	50
Bos taurus indicus	(zebu cattle)	Hariana	7
		Sahival	5
		Tharparkar	4
Bos grunniens	(yak)		2
Bubalus bubalus	(water buffalo)		2

[a]No DNA available for four sires.

3.5.2 Preparation of genomic DNA

Preparation of genomic DNA from bull sperm

Preparation was based on a protocol described in Buitkamp *et al.* (1999). Semen samples in plastic straws provided from artificial insemination stations were transferred to 1.5 mL microfuge tubes. Each semen sample which contained glycerol and other additives, was washed twice by adding 1 mL of TE buffer, followed by centrifugation at 1000 G for 10 minutes and removal of the supernatant. Pellets were resuspended in 600 µL of PKS buffer (20 mM Tris-HCl, 4 mM EDTA, 100 mM NaCl, 2% SDS, pH 7.4). For cell lysis, 25 µL of DDT (1 M) and 60 µL of proteinase K (20 mg/mL) were added and incubated overnight at 37°C with moderate shaking. The lysate was transferred to VACUTAINER Blood Collection Tubes (BD-368510; Becton, Dickinson and Company, Franklin Lakes, NJ), which are 9.5 mL glass tubes with a serum separation gel that forms a physical barrier between the upper aqueous phase and lower organic phase during centrifugation. In the first purification step, 800 µL of phenol/chloroform/isoamylalcohol (25:24:1) was added to the lysate, followed by mixing by

inverting the tube for 10 minutes and then centrifugation at 2000 G for 15 minutes. The procedure was repeated with 800 µL of chloroform/isoamylalcohol (24:1). The aqueous phase containing DNA was transferred to a 15 mL Falcon tube and mixed with 1600 µL ethanol by shaking. Precipitated DNA was transferred to a 1.5 mL microfuge tube, washed with 800 µL of ethanol (70%) and redissolved in TE buffer.

Preparation of genomic DNA from blood

For isolation of DNA from yaks and water buffalo, one male and one female blood sample were obtained from each species. Three mL of blood sample diluted with 3 mL NaCl (0.9%) was centrifuged at 1000 G for 12 minutes. The precipitate was resuspended in 1 mL NaCl, centrifuged at 1000 G for five minutes, resuspended again in 1 mL of 10 mM Tris (10 mM, pH 8.0) and centrifuged at 2800 G for 10 minutes. Cell lysis and purification of DNA were done in the same way as for the semen samples.

Evaluation of isolated DNA and compilation of DNA Pools

The concentration of DNA in a 1:5 dilution of the isolated DNA samples was measured with a fluorometer (DyNA Quant 200; Hoefer Pharmacia Biotech Inc., San Francisco, CA, USA) and adjusted with TE buffer to a concentration of 25 ng/µL. To control the concentration and the quality of the DNA, it was applied to a 0.8% agarose gel in TAE buffer together with lambda DNA (SD0011; MBI Fermentas, St. Leon-Rot, Germany) (Figure 3.2 A). Quality of DNA samples determined from DNA pools was tested by performing PCR (Figure 3.2 B). Only DNA samples with consistent results in both concentration and quality were included in the DNA pools. Preparing of the DNA pools comply with suggestion that have been published recently (Sham *et al.* 2002).

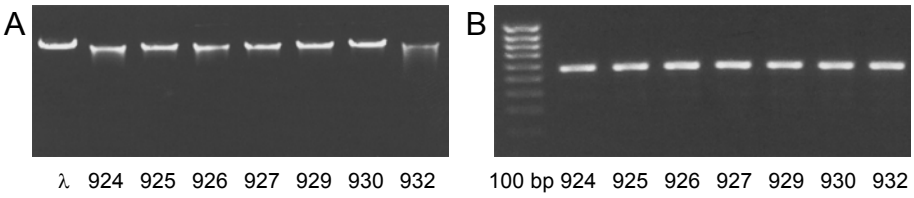


Figure 3.2: Evaluation of individual samples for DNA pools.
A. 100 ng λ DNA, 4 µL of 25 ng/µL genomic DNA, (0.8% agarose in TAE) and **B.** 4 µL of PCR (1.5% agarose in TAE).

3.6 Polymerase chain reaction (PCR)

3.6.1 Standard PCR

A standard PCR reaction of 20 µL containing 0.5 units Qiagen HotStar Polymerase (203203; Qiagen, Hilden, Germany), a hot start polymerase, 1x Qiagen PCR buffer, 1.5 mM MgCl2, 200 µM of each nucleotide and 0.5 µM each of forward and reverse primer (Thermo Hybaid, Ulm, Germany) was used to amplify 50 ng genomic DNA in a T-Gradient Thermocycler (Biometra, Göttingen, Germany) under following conditions: initial denaturing at 95°C for

15 minutes; followed by 35 cycles at 94°C for one minute, 60°C for one minute and 72°C for one minute; and final extension at 72°C for three minutes.

3.6.2 PCR optimization

New PCR primers were tested with bovine genomic DNA using an annealing temperature of 60°C in three reactions (Figure 3.3): (1) standard reaction without any additives; (2) standard reaction with 1x Qiagen Q-solution; and (3) standard reaction with 5% DMSO (dimethyl sulfoxide). The PCR conditions supplying the best result were used in further experiments. In cases of weak or unspecific PCR results in all three reactions, gradient PCR was applied with annealing temperatures varying from 54°C to 66°C. For some primer combinations, additional PCR additives with varying concentrations were tested: glycerol (5%, 10%, 15%), formamide (1.25%, 2.5%, 5%), TMAC (tetramethylammonium chloride, 15 mM, 50 mM, 100 mM), or Tween20 (nonionic detergents, 0.01%, 0.05%, 0.1%). Use of appropriate PCR additives proved to be more effective than varying the $MgCl_2$ concentration. PCR products together with a DNA size marker (Genruler 100 bp ladder: SM024; MBI Fermentas, St. Leon-Rot, Germany) were separated on agarose gels in TAE buffer.

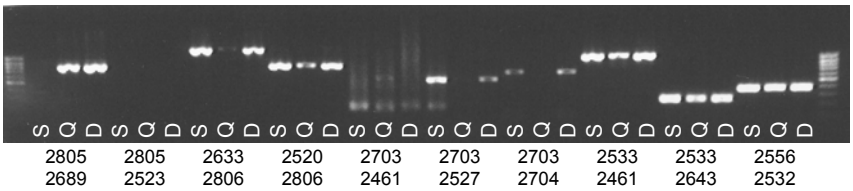


Figure 3.3: PCR optimization of 10 primer combinations with PCR additives.
S, standard PCR; Q, 1x Q-solution; D, 5% DMSO (1.2 % agarose gel in TAE).

3.6.3 Use of DMSO to avoid unequal amplification of the two alleles

A PCR fragment within *DGAT1* (primer numbers 1532, 1636) that contained a double nucleotide substitution (AA to GC) showed unequal amplification of the two alleles under standard conditions. The AA-allele was preferred in PCR amplification, although the degree varied stochastically from PCR to PCR. Addition of 5% DMSO enforced the equal amplification of the two alleles. Figure 3.4A shows the result of an RFLP assay (3.9.4, p. 31) of eight individuals using PCR with and without 5% DMSO. The effect of adding DMSO could also be observed in sequence traces. Figure 3.4B shows sequencing results for eight repeats of one heterozygous individual, the first four with DMSO and the second four without DMSO. Allele-specific amplification was not observed in other fragments.

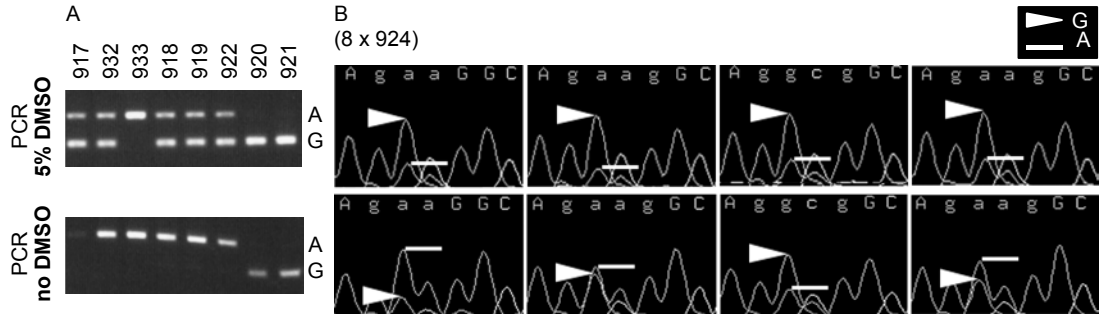


Figure 3.4: Effect of DMSO on PCR with unequal amplification of the two alleles.
A and G allele within *DGAT1* (PCR primer no. 1532 and 1636). **A.** RFLP assay of eight individuals with different genotypes; upper row with 5% DMSO in PCR, lower row without DMSO. **B.** Sequence trace views of eight PCR repeats of one individual, upper four repeats with 5% DMSO in PCR and lower four without DMSO.

3.6.4 Long range PCR.

PCR amplification of fragments up to 15 kb in size was achieved by reactions containing a combination of Qiagen AmpliTaq polymerase (201207; Qiagen, Hilden, Germany) as the main polymerase and ProofStart DNA polymerase (202203; Qiagen, Hilden, Germany) as a "proofreading" polymerase at low concentration. Proofreading is achieved by a 3' to 5' exonuclease activity where the polymerase recognizes and removes incorrectly incorporated deoxynucleotides. The exonuclease activity can lead to primer degradation during PCR setup, which result in unspecific amplification products. Qiagen ProofStart polymerase has been chemically modified for initial temporally inactivation. The enzyme is activated by the initial denaturing step at 95°C.

Conditions for long range PCR were: one reaction of 20 µL volume containing two units of Qiagen AmpliTaq Polymerase, 0.1 units of ProofStart DNA polymerase, 1x Qiagen PCR buffer, 1.5 mM $MgCl_2$, 300 µM of each nucleotide, 0.5 µM each of forward and reverse primer, 4 µL of Qiagen Q-solution, 2% DMSO and 20 ng of BAC DNA. Each reaction was overlaid with mineral oil and amplified in a T-Gradient Thermocycler under following conditions: initial denaturing and activation of the proofreading polymerase at 95°C for two minutes; followed by 35 cycles at 94°C for 10 seconds, 61°C for one minute and 68°C for 20 minutes (one minute per 1 kb); and final extension at 68°C for three minutes. Long range PCR fragments were separated on 0.7% agarose gels in TAE buffer and compared to a Gene Ruler 1 kb ladder (SM0311; MBI Fermentas, St. Leon-Rot, Germany) and Lambda Mix Marker 19 (SM0231; MBI Fermentas, St. Leon-Rot, Germany).

3.7 DNA Sequencing

DNA sequencing was performed according Sanger (Sanger *et al.* 1977) on an ABI 377 automated sequencer (Applied Biosystems, Foster City, CA, USA).

3.7.1 Primer walking and BAC end sequencing

To obtain sequence information transcending known sequences, iterative direct sequencing of BAC DNA, also known as primer walking, was performed: the priming site for new sequencing runs were selected from the most distant reliable sequence obtained in the previous cycle. In addition to internal sequencing of the BAC inserts starting from gene sequences, BAC ends were sequenced to generate STS markers. Initial primers for BAC end sequencing were derived from T7 and SP6 promoter sites (T7: 5'-CCGCTAATACGACTCACTATAGGG-3'; SP6, 5'-TTTGCGATCTGCCGTTTC-3'), which were located on the pBACe3.6 vector flanking the insert (Figure 3.5). Obtained sequences were compared by using BLAST algorithms (1) to each other to identify identical BAC ends, and (2) against the NCBI sequence database to identify genes or repetitive sequences. BAC end sequences showing no similarity to database entries were used as STS markers.

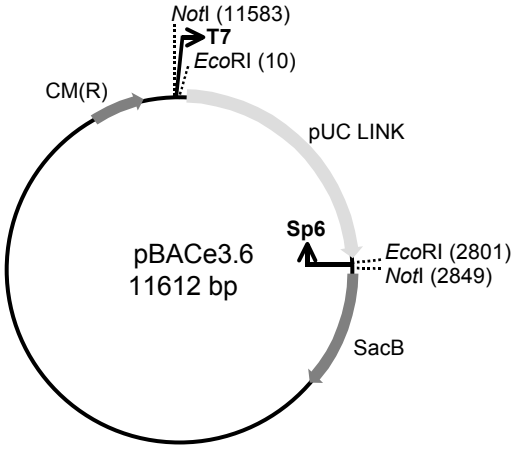


Figure 3.5: Map of pBACe3.6 vector.
pUC LINK is replaced by insert over EcoRI sites. CM(R): chloramphenicol resistance. T7 and SP6 promoter sequence was used for primer design for BAC end sequencing.

3.7.2 Sequencing reactions

BAC DNA

Long reads of up to 600 bp of high quality sequence from BAC DNA were achieved by reducing the amount of template and increasing the number of cycles from the recommended 40 to 100. The annealing temperature was increased from the recommended 51°C to 57°C to deplete the background signals from unspecific primer annealing. For primer information used for direct sequencing of BAC DNA, see appendix 9.3. Conditions for 10 µL reactions were 150 ng of BAC DNA, 0.5 pM of primer and 2 µL of BigDye terminator cycle sequencing ready reaction kit v2.0 (4314419; Perkin Elmer Applied Biosystems Division, Foster City, CA, USA). Conditions for temperature cycling were initial denaturing at 96°C for five minutes; followed by 100 cycles at 96°C for 20 seconds, 57°C for 10 seconds, and 60°C for four minutes; temperature ramping for all steps was 1°C/ second.

PCR products

PCR products were purified using MultiScreen PCR filtration plates (MANU03010; Millipore, Eschborn, Germany) in combination with a Millipore vacuum manifold (MAVM0960R; Eschborn, Germany). Samples were resuspended in 35 µL of Tris buffer (10 mM, pH 8.0) and separated together with a DNA marker (Genruler 100 bp ladder: SM0241; MBI Fermentas, St. Leon-Rot, Germany) on 1.5% agarose gels for quantification.
Cycle sequencing was carried out in 5 µL reactions containing 10-20 ng of purified PCR product, 0.5 pM of either forward or reverse primer, and 2 µL of BigDye terminator cycle sequencing ready reaction kit v2.0 (4314419; Perkin Elmer Applied Biosystems Division, Foster City, CA, USA). Conditions for temperature cycling were initial denaturation at 96°C for 15 seconds; followed by 50 cycles at 96°C for 10 seconds, 51°C for five seconds and 60°C for four minutes; temperature ramping for all steps was 1°C/ second.

3.7.3 Sequencing on an automated sequencer

Cleanup of sequencing reaction

Unincorporated dye terminators were removed by applying gel filtration. Columns of MultiScreen filtration plates (MAHVN4510; Millipore, Eschborn, Germany) were loaded with 45 µL of Sephadex G-50 Fine (G-50-50; Sigma-Aldrich Chemie GmbH, Deisenhofen, Germany) and with 300 µL of double distilled water. Column resins were ready for use after three hours at room temperature and centrifugation at 894 G for five minutes. The sequencing reactions, adjusted to 20 µL with double distilled water were applied to the columns and centrifuged at 894 G for five minutes. The samples were dried in a vacuum centrifuge (Speed Vac Plus, SC110A; Thermo Savant, Holbrook, NY, USA).

Electrophoresis

Sequencing reactions were dissolved in 2 µL of loading buffer (formamide colored with dextran blue, 47670 and 31393, respectively; Fluka, Buchs, Switzerland). After denaturing at 95°C for two minutes, 1 µL was loaded to 36 cm WTR (well to read) polyacrylamid sequencing gel with a composition of 20 mL of water (HPLC grade), 21.0 g of urea, 8.4 mL of 30% acrylamide/bisacrylamide (29:1), 6.0 mL of 10x TBE buffer, 20 µL of TEMED, and 300 µL of 10% ammonium persulfate. For electrophoresis and data collection, an automated ABI 377 sequencer (Applied Biosystems, Foster City, CA, USA) was used with the run module *Seq Run 36E-1200* and a run time of nine hours.

3.7.4 Analysis of sequencing data

Base calling, sequence assembly and polymorphism detection were performed using the Phred/Phrap/Polyphred software suite (Nickerson *et al.* 1997; Ewing *et al.* 1998a; Ewing *et al.* 1998b) and editing of the sequencers was done by the Consed software (Gordon *et al.* 1998).

3.8 BAC contig assembly for bovine DGAT1 region

3.8.1 Mapping of loci by colony PCR of overlapping BAC clones

The content of genes and STS markers neighboring *DGAT1* was assessed for each clone by performing colony PCR. Presence of PCR products was used to identify overlaps between the clones and to assemble them into a contig. A single colony of each clone was resolved in 60 µL of Tris buffer (10 mM, pH 8) and incubated at 99°C for one minute. Two µL of the lysed clones was used as template in a PCR reaction of 10 µL containing 0.5 units of Qiagen HotStar Polymerase (Qiagen, Hilden, Germany), 1x Qiagen PCR buffer, 1.5 mM $MgCl_2$, 200 µM of each nucleotide, and 0.5 µM each of forward and reverse primer. Bovine genomic DNA (25 ng) was used as a positive control. The reactions were amplified in a T-Gradient Thermocycler under following conditions: initial denaturing at 95°C for 15 min; followed by 35 cycles at 94°C for one min, 60°C for one min, and 72°C for one min; and final extension at 72°C for three min. PCR reactions were analyzed on a 1.5% agarose gel.

3.8.2 Fingerprint analysis

In addition to PCR analysis, insert size and overlap of clones were determined by restriction enzyme fingerprinting and comparison of clone banding patterns. DNA aliquots of each clone (50-100 ng) were digested by two units of *Not*I (ER0592; MBI Fermentas, St. Leon-Rot, Germany) and 0.6 µL of 10x reaction buffer (B11; MBI Fermentas, St. Leon-Rot, Germany) in a final volume of 6 µL at 37°C for eight hours. Pulsed-field gel electrophoresis (PFGE) with a CHEFF-DR II system (Bio-Rad, Hercules, CA, USA) was used to separate *Not*I fragments of clones together with following size standards: MidRange I and MidRange II PFG Markers (3551-1 and 355-2; New England BioLabs, Frankfurt am Main, Germany), Gene Ruler 1 kb ladder (SM0311; MBI Fermentas, St. Leon-Rot, Germany) and Lambda Mix Marker 19 (SM0231; MBI Fermentas, St. Leon-Rot, Germany). Conditions were 1.0% Large DNA low Melt agarose (Biozyme, Hessisch Oldendorf, Germany) in 0.5% TBE buffer, with the temperature maintained at 14°C, an electric field of 6.0 V/cm and a pulse time of 1-15 seconds over a total run time of 16 hours. In a second approach, the total run time was 20 hours. Gels were stained in 1x TBE containing 0.5 µg/ml of ethidium bromide (X-328; Amresco, Solon, Ohio, USA).

3.9 Detection and genotyping of polymorphisms

3.9.1 Detection and genotyping of polymorphisms by re-sequencing

Polymorphisms were detected by re-sequencing selected individual animals and pooled DNA samples (see chapter 3.5.1). DNA samples and primer systems used for re-sequencing are listed in appendices 9.3 and 9.5, respectively.

In bovine *DGAT1*, all exons and smaller introns were re-sequenced using the six extreme DNA pools of German Holstein, German Simmental and German Brown, as well as a set of 10 randomly selected individual animals: three unrelated German Simmental animals, three

unrelated German Holstein bulls, and one bull from each of Kerry, Angus, Hariana and Sahival. For the larger introns and the flanking regions of *DGAT1*, only the six DNA pools were used. Parts of *DGAT1*-neighboring genes were re-sequenced by using the four extreme DNA pools of the German Simmental and German Holstein breeds, as well as four DNA samples from individual animals that were selected by their genotype for the lysine^{232}-alanine substitution within *DGAT1* (see results 4.1.4, page 43 and 47): two German Holstein bulls, one homozygous for alanine and the other homozygous for lysine; and two German Simmental bulls, one heterozygous and the other homozygous for lysine.

All exons and smaller introns of bovine *DGAT2*, *DC2* and *DC5* were re-sequenced using the six extreme DNA pools of German Holstein, German Simmental and German Brown, as well as 12 (10 for DC5) individual DNA samples from German Holstein and German Simmental bulls, which were selected partly randomly and partly for extreme breeding values.

Obtained sequences were analyzed for polymorphisms using Phred/Phrap/Polyphred/Consed software suite (Nickerson *et al.* 1997; Ewing *et al.* 1998a; Ewing *et al.* 1998b; Gordon *et al.* 1998) (Figure 3.6).

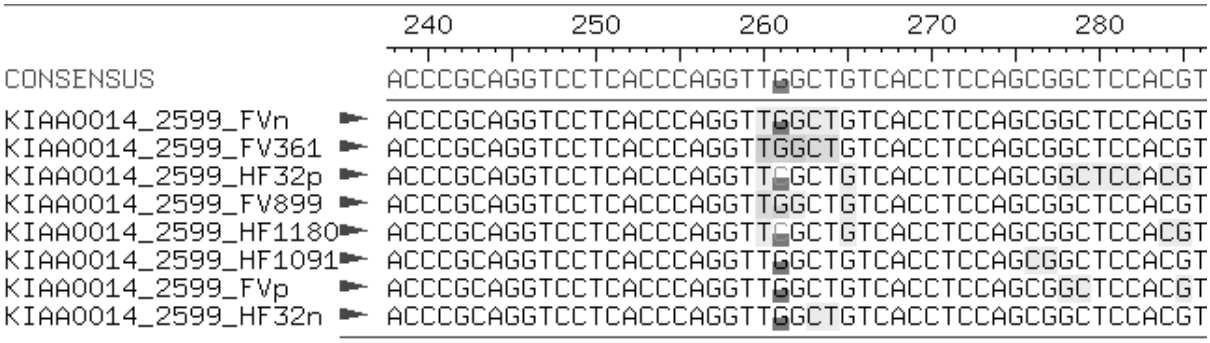


Figure 3.6: Consed view of aligned sequence traces.
Sequence alignment of eight samples indicating a SNP at position 261 (G/C substitution).

3.9.2 Allele frequency estimation from pooled DNA sequence traces

Allele frequencies were estimated by comparing sequencing traces of pooled DNA samples with sequencing traces of individual animals (Kwok *et al.* 1994). Peak heights were derived from data files with the extension ".poly" that were created by the base calling program Phred (Ewing *et al.* 1998a; Ewing *et al.* 1998b). Allele frequencies were calculated by a python script (Durstewitz *et al.* 2002). Chi-square 2x2 tests of homogeneity were performed to test the significance level of differences in allele frequencies between the DNA pools with high and low breeding value for milk fat percentage (Table 3.6).

Table 3.6: Observed number of alleles for DNA pools.

	Allele 1	Allele 2	Row totals
+ pool	O^{1+}	O^{2+}	total +pool
- pool	O^{1-}	O^{2-}	total -pool
Column totals	total allele 1	total allele 2	grand total

Observed number of alleles O^{ji} were calculated as

$$O^{1i} = f^{1i} \cdot 2 \cdot N^i \text{ and } O^{2i} = (1 - f^{1i}) \cdot 2 \cdot N^i$$

where i represents the + pool and – pool, j the allele 1 and allele 2, f^{1i} the estimated frequencies for allele 1 and N^i the numbers of individuals in the respective DNA pool i. The expected number of alleles for each cell was computed by multiplying the total for the respective column with the total of the respective row, and was divided by the grand total of rows. To assess the significance of differences in the allele frequencies between the extreme pools of each breed, the test statistic G was calculated using the standard formula

$$G = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O^{ji} - E^{ji})^2}{E^{ji}} .$$

When testing the null hypothesis H_0 (no association between SNP alleles and the breeding value for milk fat percentage) the test statistic G follows a χ^2-distribution with one degree of freedom.

3.9.3 VNTR genotyping by PCR with fluorescence-labeled primer

The number of repetitions of the 18 bp repeat unit upstream to bovine *DGAT1* was determined on an ABI 377 sequencer (Applied Biosystems, Foster City, CA) based on the size dependent electrophoretic mobility of fluorescence-labeled PCR products (146 bp by five repeat units; primers: 5'-6-Fam-TCAGGATCCAGAGGTACCAG-3' and 5'-GGGGTCCAA GGTTGATACAG-3'). PCR reactions of 10 µL contained 0.25 units of Qiagen HotStar Polymerase (203203; Qiagen, Hilden, Germany), 1x of Qiagen PCR buffer, 1.5 mM of $MgCl_2$, 200 µM of each nucleotide, 0.5 µM each of forward and reverse primer, 1x of Qiagen Q-solution and 25 ng of genomic DNA. The reactions were amplified in a T-Gradient Thermocycler under following conditions: initial denaturing at 95°C for 15 minutes; followed by 35 cycles at 94°C for 30 seconds, 59°C for 45 seconds, and 72°C for 1.5 minutes; and a final extension at 72°C for 10 minutes. Reactions were diluted 1:30 after thermal cycling and 1.2 µL of the dilution were mixed with 1.3 µL of loading buffer (formamide colored with dextran blue, 47670 and 31393, respectively; Fluka, Buchs, Switzerland) and 0.2 µL of GenScan-500 TAMRA size standard (401733; Applied Biosystems, Foster City, CA, USA). One µL was loaded and electrophoresed using the run module *GS 36C-240* with a run time of

1.8 hours. The number of repeat units was assessed using Genotyper software version 2.5 (Applied Biosystems, Foster City, CA, USA).

3.9.4 SNP genotyping by RFLP analysis

To find restriction enzymes that specifically cut one of the two alleles, the "Map" program of the GCG software package (Genetics Computer Group 2001) was used. The restriction enzyme *Cfr*I (ER0162; MBI Fermentas, St. Leon-Rot, Germany) was selected for SNPs in both *DGAT1* (Figure 3.7A) and *DGAT2* (Figure 3.7B). Details about the SNPs and RFLP systems are listed in Table 3.7.

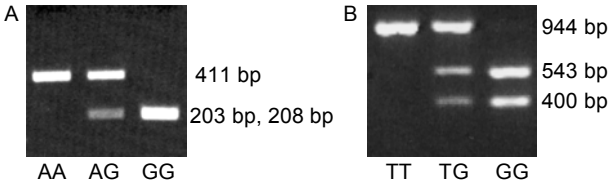


Figure 3.7: RFLP assay for SNPs in *DGAT1* (snp_id 252) and *DGAT2* (snp_id 303).

Table 3.7: Two RFLP systems for genotyping SNPs in *DGAT1* and *DGAT2*.

Locus	***DGAT1***	***DGAT2***
snp_id	252	303
Position	exon 8, base 15 and 16	intron 6, base 617
PCR	standard PCR including 5% DMSO[a]	standard PCR
Forward primer	1532 (5'-GCACCATCCTCTTCCTCAAG-3')	2093 (5'-AGCAGCTCCTTGGCTCCT-3')
Reverse primer	1636 (5'-GGAAGCGCTTTCGGATG-3')	1900 (5'-TGGTGATGGGCTTGGAGTAG-3')
Restriction enzyme	*Cfr*I	*Cfr*I
Allele 1	**AA**: 411 bp	**T**: 944 bp
Allele 2	**GC**: 203 bp + 208 bp	**G**: 543 bp + 400 bp[b]

[a] See chapter 3.6.3.
[b] Deletion in PCR-fragment of g-allele.

Buffer B+ (BB5; MBI Fermentas, St. Leon-Rot, Germany) was used instead of the recommended buffer Y+ because it yielded more precise bands on the agarose gel. Four µL of PCR reaction were digested by two units of *Cfr*I in a volume of 20 µL for four hours. Ten µL of the reaction were separated on a 2.0% agarose gel.

3.9.5 Multiplex SNP genotyping by single base extension (SBE)

Single base extension (SBE) in combination with gel separation by product size and fluorescence detection was established for SNP genotyping based on a method reported in Lindblad-Toh *et al.* (2000). Excess primers and nucleotides remaining after PCR amplification were removed enzymatically. The SBE primer binds to the PCR amplified fragment that terminates on the 5' base immediately preceding thee SNP (Figure 3.8). Primers are extended by a single fluorescently labeled didesoxynucleotide, the color of which represents the base at the SNP position. Multiplexed SBE primers were discriminated on polyacrylamid sequencing gel by SNP-specific primer length that was adjusted by adding a 5' polyA tail.

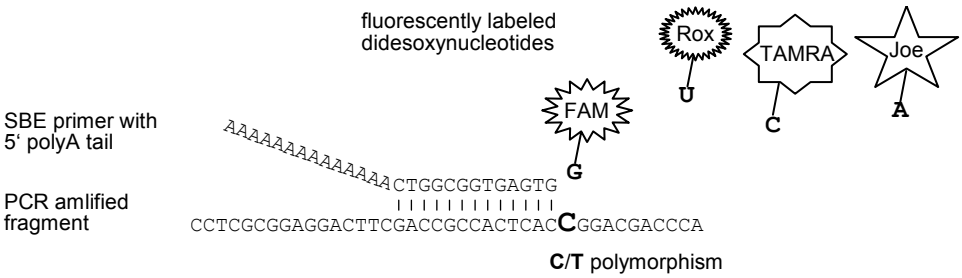


Figure 3.8 Principle of single base extension (SBE).

Primer design

PCR primers were designed to generate short fragments between 60 and 150 bp. PCR fragments in multiplex reactions have to differ in size by at least five bp to allow sufficient discrimination on agarose gels to permit the presence of all fragments to be checked. Primer sets for multiplex genotyping were checked for primer dimer formation by the software "oligos v.9.4" by Ruslan Kalendar, University of Helsinki, Finland (http://www.biocenter.helsinki.fi/bi/bare-1_html/oligos.htm). Primers for SBE reactions were designed to have a melting temperature between 60 and 64°C and to terminate on the 5' base preceding the SNP. SBE primers could not contain any neighboring SNPs and could be derived from either the sense or antisense strand. SBE primer lengths were adjusted to 18, 22, 26, 30, 34, 38, 42, 46 and 50 nucleotides by adding a polyA tail to the 5'-end.

Multiplex PCR

In multiplex PCR reactions, the concentration of polymerase was doubled and the concentrations of each primer were halved compared to a standard PCR. Weak PCR products could be compensated for in some cases by increasing the primer concentration from 0.25 µM up to 1.0 µM. A multiplex PCR with a final volume of 10 µL contained 0.5 units of Qiagen HotStar Polymerase (203203; Qiagen, Hilden, Germany), 1x of Qiagen PCR buffer, 1.5 mM of $MgCl_2$, 200 µM of each nucleotide, 0.25 µM of each primer and 25 ng of genomic DNA. The reactions were amplified in a T-Gradient Thermocycler under following conditions: initial denaturing at 95°C for 15 minutes; followed by 35 cycles at 94°C for 40 seconds, 60°C for 50 seconds, and 72°C for 30 seconds; and final extension at 72°C for three minutes.

The presence and quantity of each product in a multiplex PCR was assessed by analyzing the multiplex reaction and reactions for each PCR product on a 3% BMA MetaPhor agarose gel (850180; Biozyme, Hessisch Oldendorf, Germany) in 1x TAE buffer. Higher resolution was archived using a 12% polyacrylamid gel (14.8 mL of double distilled H_2O, 12 mL 30% of acrylamide/bisacrylamide 29:1, 3.0 mL of 10x TBE buffer, 105 µL of TEMED, 210 µL of 10% ammonium persulfate, and 9 mg of ethidium bromide) in 1x TBE buffer (Figure 3.9). The size standard used was pUC19 DNA/MspI Marker, 23 (SM0221; MBI Fermentas, St. Leon-Rot, Germany).

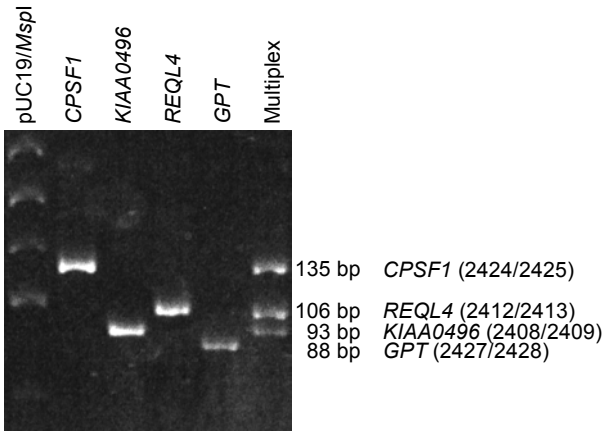


Figure 3.9: Multiplex PCR on polyacrylamid gel. Four single reactions compared with one multiplex reaction (12% polyacrylamid gel in 1x TBE). The PCR fragment for GPT is absent in the multiplex PCR.

Cleanup of PCR reaction

Enzymatic cleanup of PCR reactions (Werle *et al.* 1994) was done using Exonuclease I (ExoI, 162110; Biozyme, Hessisch Oldendorf, Germany) to hydrolyze single-stranded DNA (primers) and Shrimp Alkaline Phosphatase (SAP, EF0511; MBI Fermentas, St. Leon-Rot, Germany) to remove the 5'-phosphate group. A reaction of 5 µL included 3 µL the PCR reaction, 0.5 units SAP and 0.4 units of ExoI. Reactions were performed at 37°C for 1.5 hours and terminated by heat inactivation at 80°C for 15 minutes.

Optimization of SBE reaction

Initial concentrations for the four didesoxynucleotides within the SBE reaction were derived from Lindblad-Toh *et al.* (2000). 15 SNPs were genotyped by SBE in three individuals (one of each homozygous and one heterozygous). The concentrations for each fluorescent dye were calibrated within the SBE reaction based on the obtained average peak heights. In a second experiment, four SNPs were genotyped by SBE to investigate for any possible influence of SAP in the PCR purification reaction on the SBE results, and to optimize the amounts of the SBE reagents (Table 3.8). The four SNPs showed differences in peak heights of up to ten fold compared to each other. A third experiment tested the effect of SBE primer concentration by SBE genotyping of two SNPs. The SNPs showed very low and very high signals when the primer concentrations were increased and reduced, respectively (Table 3.8).

Table 3.8: Optimization of SBE reaction.

	Reduced			Standard	Increased				
	0.25x	0.5x	0.75x	1.0x	1.5x	2.0x	3.5x	4.0x	5.0x
Concentration of									
SAP in the PCR purification reaction		0.5x	0.75x	0.015 u/µL	1.5x				
Fluorescent didesoxynucleotides		0.5x		[a]					
Thermosequenase (from two lots)				0.05u/µL		2.0x		4.0x	
Volume of SBE reaction applied to the gel		0.5x		1 µL		2.0x			
Concentration of SBE primer									
SNP with low signal				0.2 µM		2.0x	3.5x		5.0x
SNP with high signal	0.25x	0.5x	0.75x	0.2 µM					

[a] 0.06 µM of ddGTP, 0.074 µM of ddATP, 0.3 µM of ddCTP and 0.37 µM of ddUTP

SBE reaction

The optimized SBE reactions were carried out in a final volume of 10 µL containing 1 µL of PCR cleanup reaction, 0.4 units of Thermosequenase (US78500; Amersham Biosciences, Freiburg, Germany), 1x of Thermosequenase reaction buffer, 0.2 µM of each SBE primer, 0.03 µM of FAM-ddGTP (NEL483; PerkinElmer Life Sciences, Boston, MA, USA), 0.03 µM of JOE-ddATP (NEL486; PerkinElmer Life Sciences, Boston, MA, USA), 0.20 µM of ROX-ddUTP (NEL476; PerkinElmer Life Sciences, Boston, MA, USA), 0.14 µM of TAMRA-ddCTP (NEL473; PerkinElmer Life Sciences, Boston, MA, USA). Temperature cycling used initial denaturing at 96°C for two minutes; followed by 30 cycles at 96°C for 20 seconds, 55°C for 20 seconds and 60°C for 30 seconds. SBE reactions were incubated at 37°C for one hour together with one unit of SAP. Dephosphorylation of the unincorporated florescent didesoxynucleotides helped to keep them out of the critical gel region.

Electrophoresis

Composition of the polyacrylamid gels for SBE was the same as for sequencing (see chapter 3.7.3 on page 27). One µL was applied to the gel from a mix containing 2 µL SBE reaction, 0.3 µL GeneScan-500 Rox size standard (401734; Applied Biosystems, Foster City, CA, USA) and 2.7 µL loading buffer (formamide colored with dextran blue, 47670 and 31393, respectively; Fluka, Buchs, Switzerland). Electrophoresis and data collection were carried out by means of an ABI 377 sequencer (Applied Biosystems, Foster City, CA) using run module *GS Run 36A-2400*. Run time was one hour. The sizes of the oligos were assigned to the raw signals using GeneScan software version 3.1 (Perkin Elmer Applied Biosystems Division, Foster City, CA, USA) and the internal GeenScan-500 Rox size standard. The GeneScan output was transferred into Genotyper software version 2.5 (Perkin Elmer Applied Biosystems Division, Foster City, CA, USA) for allele calling based on defined categories for each SNP allele. Each category is defined by the color and the size range in which the peak appears. Called SNP alleles were called and labeled automatically (Figure 3.10) and appended to a table. The table was exported and the genotypes were derived using a python script.

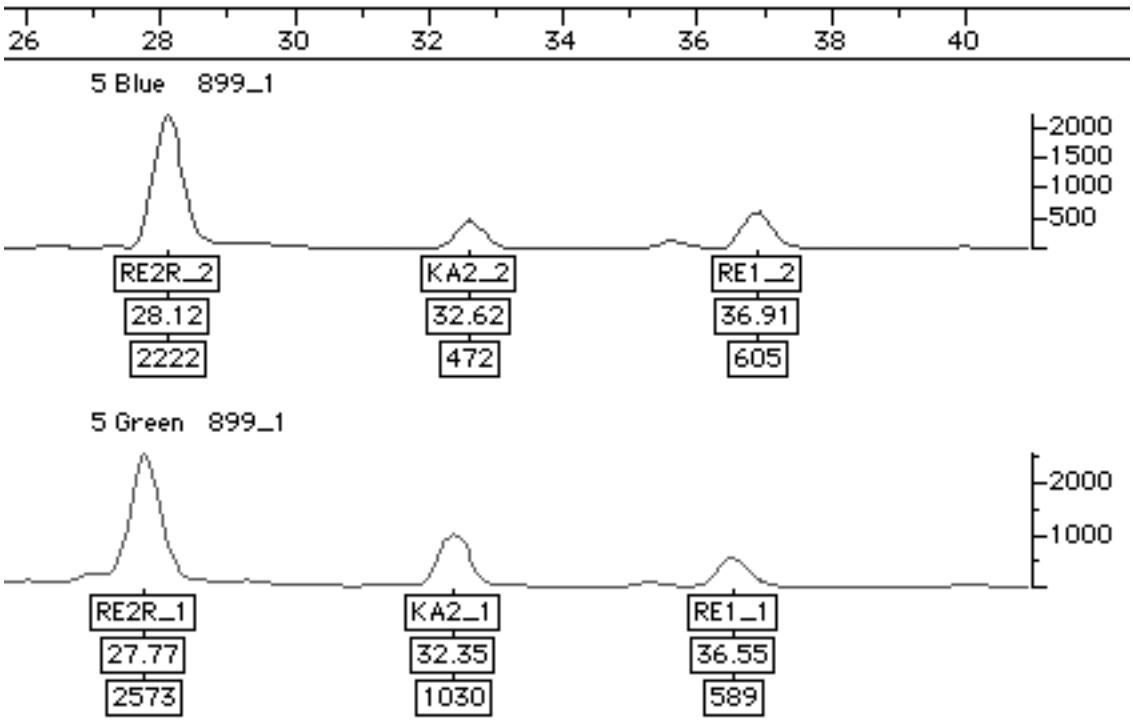


Figure 3.10: Plot view of Genotyper showing SBE results.

SBE results of an individual (sample id 899) at three heterozygous SNPs (RE2, KA2 and RE1). SBE primer size is indicated along the top (24 bp to 40 bp). The upper curve represents the blue fluorescent signal, the lower curve the green fluorescent signal. Peaks showing the presence of the respective fluorescently labeled SBE primer are labeled with the name for the SNP allele, the size and the peak height.

4 Results

4.1 Association of DGAT1 with milk fat percentage

4.1.1 Cloning of bovine DGAT1

The cDNA sequence of *DGAT1* in mouse (Cases *et al.* 1998) and in human (Oelkers *et al.* 1998) allowed me to identify 12 homologous bovine EST sequences spanning from exon 2 to exon 17 (Table 4.1). Five of the EST sequences differ from the human cDNA due to inclusion intronic sequences or the lack of exonic sequences. Four putative transcripts would lead to truncated enzymes or to frameshifts. One bovine EST (AW446985), covering *DGAT1* exon 7 to 11, lacks 66 bp from exon 8, which includes the missense mutation responsible for the lysine^{232} – alanine substitution (see 4.1.4).

Table 4.1: Bovine EST sequences for *DGAT1*.

Locus	GenBank Accession	Size [bp]	Source[a]	Position[b]		Discrepancy to consensus sequence + inclusion, - deletion
DGAT1	AW483961	205	1	1594-1745	3'UTR	
	AW486026	385	1	1336-1720	exon 17 - 3'UTR	
	BE664357	456	1	1321-1745	exon 17 - 3'UTR	
	BE664362	415	1	1321-1735	exon 17 - 3'UTR	
	AW326076	141	2	703-772	exon 8 - exon 9	+ 76 bp of intron 8 (frameshift)
	AW446908	479	2	256-780	exon 2 - exon9	- 47 bp in exon 7 (frameshift, stop codon in exon 8)
	AW446985	485	2	594-1143	exon 7 - exon 11	- 66 bp of exon 8
	AW652329	542	2	990-1530	exon 13 - 3'UTR	
	BE751071	475	3	1087-1560	exon 14 - 3'UTR	
	BE753833	422	3	1369-1745	exon 17 - 3'UTR	
	BE900091	527	4	1097-1561	exon 14 - 3'UTR	+ 37 bp of intron 13 and + 28 bp unknown sequence
	BE486748	174	5	906-986	exon 11 - exon 12	+ 76 bp of intron 11 and + 20 bp of intron 12 (frameshift)

[a]Source:
1: pooled tissue from day 20 and day 40 embryos
2: pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary
3: pooled tissue from testis, thymus, semitendonosus muscle, longissimus muscle, pancreas, adrenal gland, and endometrium
4: adipose tissue
5: mammary tissues from eight physiological, developmental, and disease states
[b]Base 1 = first base of start codon

Four bovine BAC clones were isolated by means of *DGAT1* specific probes from the RPCI-42 BAC library: 240A1, 258E13, 269H17 and 56F1.

4.1.2 Physical mapping of DGAT1

RH-mapping in human

To assess the positional candidate gene status of *DGAT1*, its precise position on human chromosome 8 was determined. RH mapping (Figure 4.1) placed *DGAT1* next to marker AFMa082wh9 at cR_{3000} = 552.63 (lod score = 9.5) on the GB4 map (http:// www.ncbi.nlm.nih.gov/genemap99/), 3.87 centirads from human *KIAA0278*, a marker that is 16.6 centirads

from *CSSM066* on the bovine radiation hybrid map (Figure 2 in Riquet *et al.* 1999). Riquet *et al.* (1999) showed *CSSM066* to be part of the identical-by-descent segment that indicated the minimal mapping interval of the QTL for milk fat percentage. This provided indirect support of the positional candidate status of *DGAT1* with regard to the QTL.

```
         10          20          30          40          50          60
70          80          90
       .       .       .       .       .       .       .       .       .
02102020012210000010100011000001001001000012011000100001020001000000121020000012100000000100
```

Figure 4.1: Vector of PCR scores of human Genbridge 4 radiation hybrid panel.
Results of the PCR reactions are recorded in a string with definite positions for each cell line (= vector): 0, no PCR product; 1, PCR product; and 2, ambiguous or not typed.

RH and FISH mapping in cattle

RH mapping in cattle using the BovRH5 panel (Womack *et al.* 1997), placed *DGAT1* proximal to *ILSTS039* (Table 4.2), a marker that indicated the proximal boundary of the chromosomal segment predicted to contain the QTL (Riquet *et al.* 1999). More recently, however, the QTL interval was revised to the centromere (Farnir *et al.* 2002), thereby supporting *DGAT1* as a positional candidate gene for milk fat percentage.

Table 4.2: RH mapping results (BovRH5 panel) of bovine *DGAT1*.

Locus	Break frequency	cR_{5000}
BE217466	-	-
DGAT1	0.147	15.9
PTK2	0.184	20.3
CYC1	0.089	9.3
CACNB3	0.078	8.1
ILST039	0.099	10.4
CSSM66	0.086	9.0
Map length		73.0

RH mapping in cattle was done by Johannes Buitkamp (Landesanstalt für Tierzucht in Grub, Germany).

Mapping of BAC-DNA (clone 56F1) by FISH assigned *DGAT1* to bovine chromosome 14q12-14 (Flqter = 0.79±0.05, mean± SD; Flqter: relative fractional length from the long arm telomere to the hybridization signal), the approximate physical location of the QTL for milk fat percentage. FISH mapping of bovine *DGAT1* was performed by Felix Habermann (Lehrstuhl für Tierzucht, Technische Universität München, Germany).

4.1.3 Characterization of bovine DGAT1

Sequence and gene structure

Bovine *DGAT1* gene was sequenced completely; including all introns, as well as 3500 bp of the upstream and 1900 bp of the downstream sequence. Sequence information for bovine

DGAT1 has been deposited in the EMBL database under the accession number AJ318490. The coding sequence in both cattle and human is 1470 bp and is organized in 17 exons separated by 16 introns (Figure 4.2 and Table 4.) .

Figure 4.2: Exon/intron structure of bovine *DGAT1*.
Boxes represent exons; white boxes are untranslated regions and gray, numbered boxes are coding regions. The horizontal lines represent the introns. Triangles indicate polymorphic positions.

Table 4.3: Exon/intron organization of bovine *DGAT1*.

	Exon					Intron		
	No.	Position in cDNA[a]	Size [bp]	3'-splice acceptor[b]	5'-splice donor[b]	No	Size [bp]	Size [bp] (human)
DGAT1	1	1-191	191	.....	CCTGAG**gt**agcg	1	3414	5.0 k
	2	192-279	88	ctcc**ag**GTGTCA	..... ATGCTG**gt**acgt	2	1944	2.3 k
	3	280-320	41	tcgc**ag**ATCTTA	..... CATCAA**gt**gagt	3	79	107
	4	321-406	86	ctgc**ag**GTATGG	..... TCATTG**gt**gagc	4	92	83
	5	407-459	53	cctc**ag**TGGCCA	..... GCCGTG**gt**aagc	5	215	132
	6	460-565	106	cccc**ag**GGAGCT	..... CTCCAG**gt**gggc	6	89	98
	7	566-679	114	ccac**ag**TGGGCT	..... AGGCTG**gt**gagg	7	100	91
	8	680-754	75	tcgt**ag**CTTTGG	..... ACCGCG**gt**gagg	8	70	77
	9	755-858	104	ttcc**ag**ATCTCT	..... GAGATG**gt**gagg	9	90	91
	10	859-897	39	cccc**ag**CTGTTC	..... CAGCAG**gt**acgt	10	60[c]	71
	11	898-939	42	ttgc**ag**TGGATG	..... TTCAAG**gt**gagc	11	73	82
	12	940-984	45	ccac**ag**GACATG	..... CTGGCG**gt**gagt	12	74	98
	13	985-1097	113	ccac**ag**GTCCCC	..... CTGGTG**gt**gggt	13	87	80
	14	1098-1163	66	ccgc**ag**GAACTC	..... CATCAG**gt**gggt	14	86	77
	15	1164-1251	88	ccgc**ag**ACACTT	..... CACGAG**gt**cagt	15	81	106
	16	1252-1314	63	cctc**ag**TACCTG	..... GCGCAG**gt**gagc	16	72	143
	17	1315-1470	156	cccc**ag**ATCCCG	.....			

[a]Base 1 = first base of start codon
[b]Exon sequences are indicated in uppercase letters, intron sequences in lowercase letters. The consensus splice site sequences are in boldface.
[c]Intron 10 contains a (G)n stretch that could not be resolved by sequencing.

All intron/exon splice junctions conformed to the GT/AG rule (Breathnach *et al.* 1978). The structure of the bovine genes is highly conserved compared to the human orthologues (Table 4.). The entire gene spans 8.7 kb in cattle and 10.6 kb in human. The increase in size of 1.9 kb in human derives from the larger introns 1 and 2. The coding sequence of human and bovine *DGAT1* show 88.8% identity, as determined by the software program "gap" of the GCG package (Genetics Computer Group 2001).

To identify conserved domains in DGAT1, the derived peptide sequences of 12 species were aligned (Figure 4.3). In addition, the human ACAT1 and ACAT2 peptide sequence were included in the alignment. The species included five mammals (two primates including human,

cattle, pig and two rodents), one fly, one nematode and five plants. Species names are given in the legend of Figure 4.3. Transmembrane domains were predicted using the program "Residue-based Diagram editor" (Campagne *et al.* 1999) and additional motifs as identified in human (Oelkers *et al.* 1998) are indicated above the aligned sequences:

- one N-linked glycosylation site (N in human and cattle at position 246)
- two putative tyrosine phosphorylation sites (Y in human and cattle at position 316 and 361)
- one putative diacylglycerol-binding motif (HKWCIRHFYKP in human and cattle at positions 382-392).
- an HSF motif (residues 268-270), the central serine of which was found to be essential to the activity and stability of Chinese hamster ovary ACAT1 (Cao *et al.* 1996).

Partial views of these motifs are displayed enlarged in Figure 4.4.

Description Figure 4.3, page 40 and Figure 4.4, page 41.

Compete and partial views of aligned peptide sequences of DGAT1 for following species: *Arabidopsis thaliana* (At), *Brassica napus* (Bn), *Tropaeolum majus* (Tm), *Nicotiana tabacum* (Nt), *Perilla frutescens* (Pf), *Homo sapiens* (h), *Cercopithecus aethiops* (c), *Sus scrofa* (p), *Olea europaea* (b), *Mus musculus* (m), *Rattus norvegicus* (r), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce) and for human ACAT1 and ACAT2. Numbers on the left indicate amino acid positions. Black and gray backgrounds of letters indicate identical and conserved amino acids, respectively. The triangle indicates the position of the lysine-alanine substitution in *Bos taurus*. (Alignment was done using ClustalX, with shading of alignment by BOXSHADE 3.21, http://www.ch.embnet.org/software/BOX_form.html). A putative N-linked glycosylation site, two putative tyrosine phosphorylation sites and a putative diacylglycerol-binding motif are indicated as suggested in Oelkers *et al.* (1998).

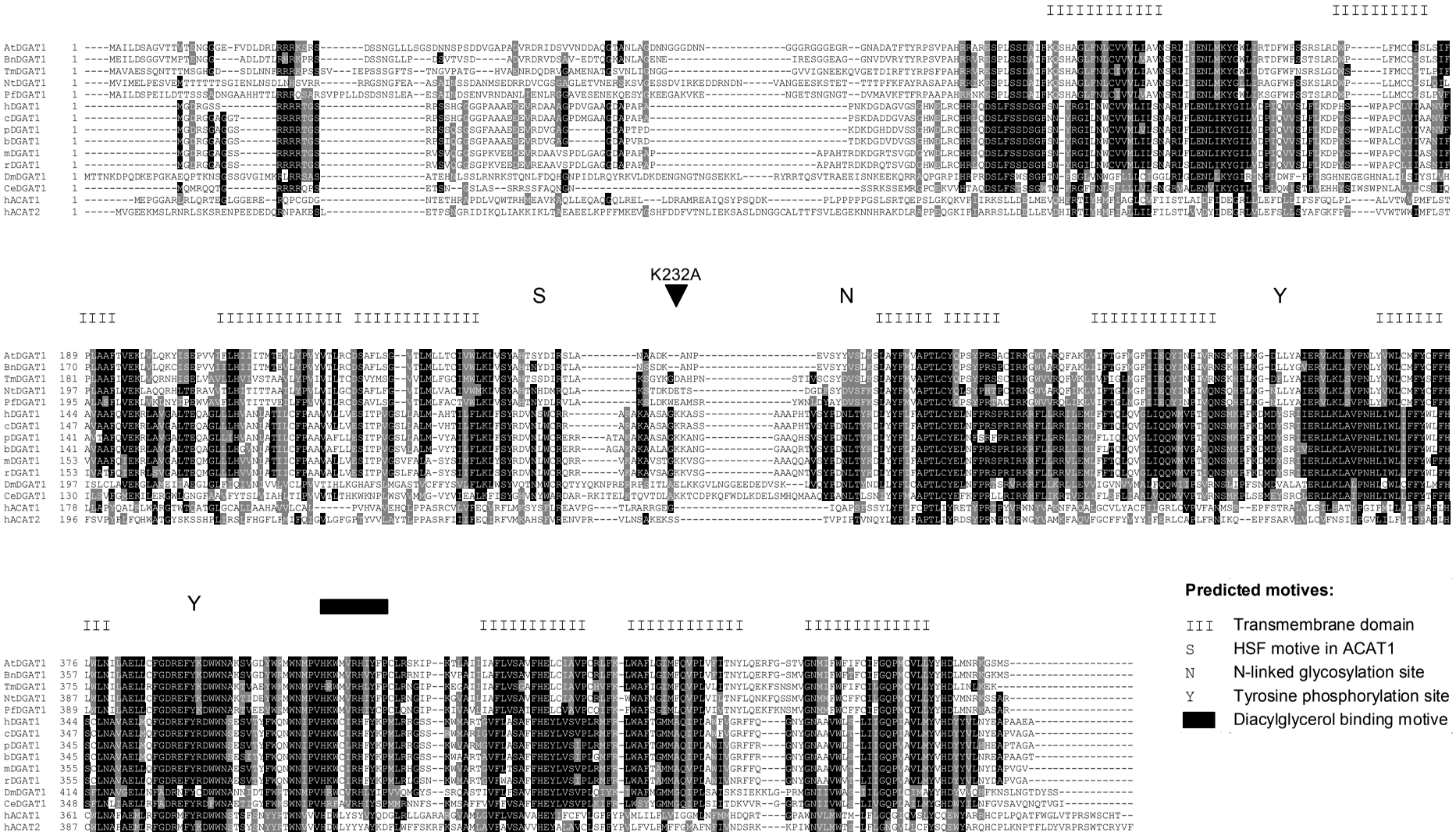


Figure 4.3: Alignment of DGAT1-derived peptide sequences of different species.

For description see page 39.

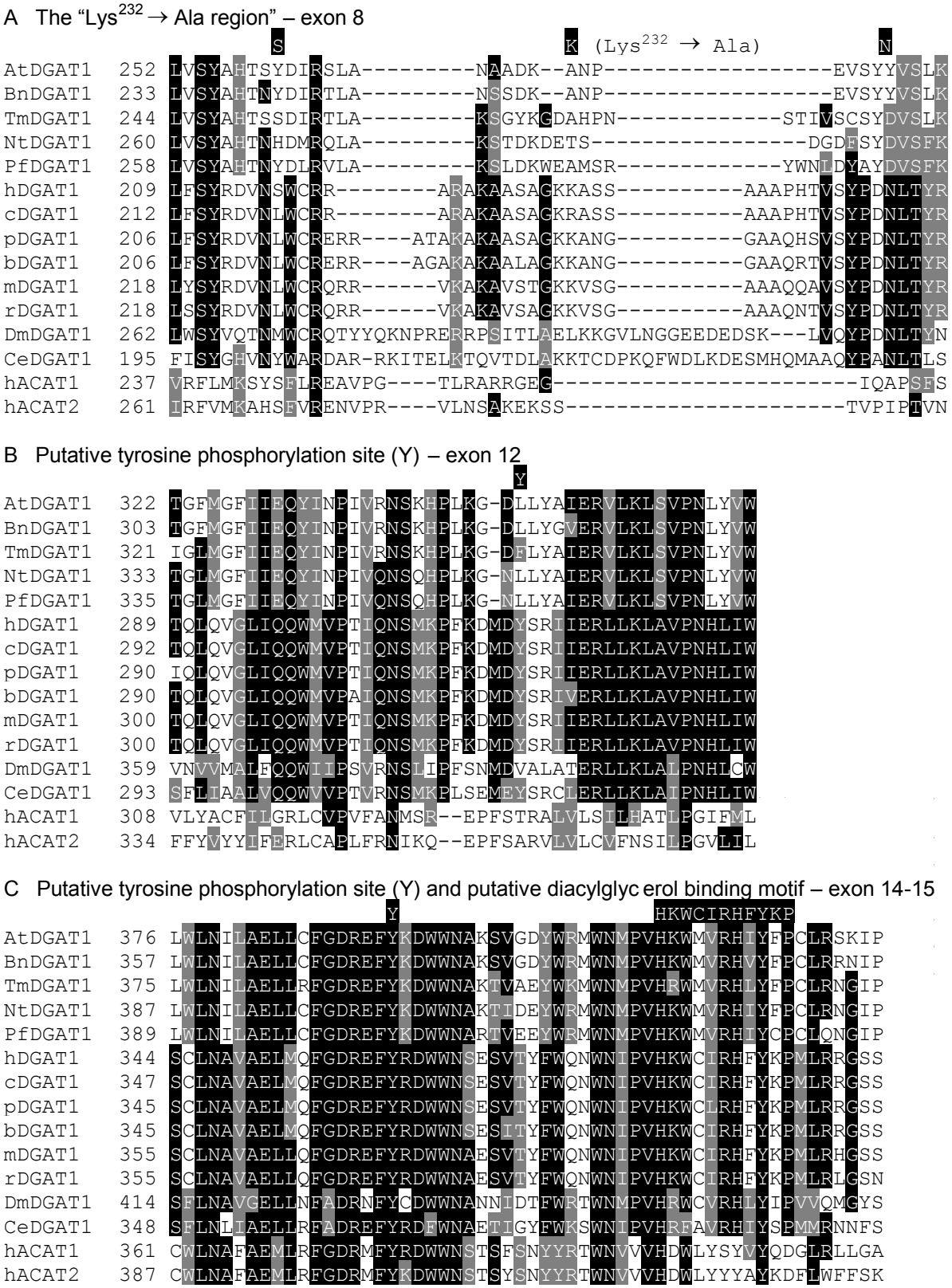


Figure 4.4: Motifs within DGAT1 peptide sequence in different species.
Partial views of **Figure 4.3**. For description see page 39.

Transcription start

No bovine EST sequence was available for 5'end of *DGAT1* including exons 1 and 2. Rapid Amplification of cDNA Ends (RACE) was applied to obtain the first exons and the transcription start. However, this resulted in no product for the 5'end. After elucidating the full *DGAT1* sequence, including up- and downstream regions, by direct sequencing of BAC DNA, the upstream sequence was used to predict a putative transcription start (Figure 4.5) using the software program "Promoter Prediction" (Reese *et al.* 1995; Reese *et al.* 1996). The highest value was obtained for guanine at position 3249 in the obtained sequence (accession number: AJ318490).

Start	End	Score	Promoter Sequence
3209	3259	0.99	AAATCCTGTGTTTATAGAGCGGGACAAGGGGCAGGCAGCG**G**TCAGCAGAG

Figure 4.5: Predicted transcription start for bovine *DGAT1*.
Rectangle indicates a putative TATA-Box and the emphasized letter (G) indicates the putative transcription start at position 3249 in the sequence with the GenBank accession number AJ318490 as determined using "Neural Network Promoter Prediction" (http://www.fruitfly.org/seq_tools/promoter.html).

The positional accuracy of the transcription start was specified within +/- 3 bp because promoter elements may appear at different relative positions (help of web interface http://www.fruitfly.org/seq_tools/promoter.html). The prediction program was evaluated using 429 eukaryotic RNA Polymerase II promoters and 305 unrelated genes: 10% of promoters were recognized with no false positives using a threshold (score) of 0.99.

CpG island

CpG islands are short, dispersed regions of unmethylated DNA with a high frequency of CpG dinucleotides relative to the whole genome. CpG islands are defined as regions greater than 200 bp in length with a G+C content of more than 50% and a value of at least 0.6 for the ratio of observed CpG content / expected CpG content (Gardiner-Garden *et al.* 1987). A CpG rich area with a length of 208 bp (3187-3394 in AJ318490) fulfills the CpG island criteria and occurs immediately before the 5'UTR of bovine *DGAT1* (Figure 4.6). This result is consistent with the observation that CpG islands are often associated with the 5' end of genes (Larsen *et al.* 1992).

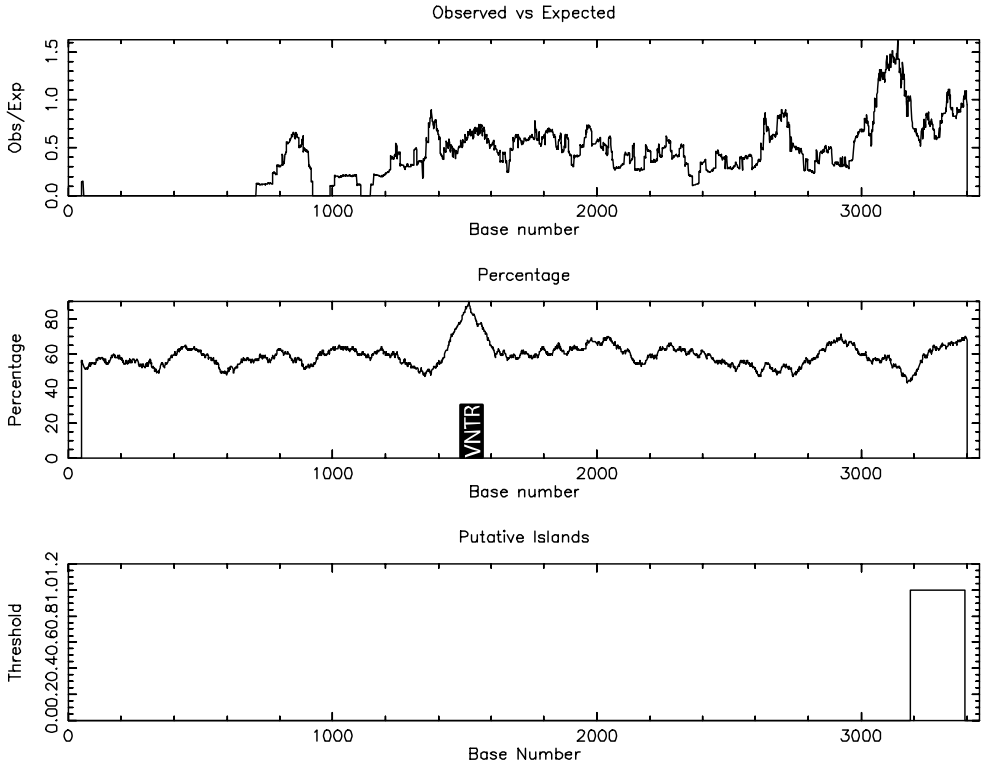


Figure 4.6: CpG islands mapping of the 5'end of DGAT1.

The upstream sequence of *DGAT1* (1-3450 of AJ318490) was used as input for CPGPLOT program (by Alan Bleasby, http://bioweb.pasteur.fr/seqanal/interfaces/cpgplot.html). The original program was described in Larsen *et al.* (1992)). One putative CpG island was detected with a length of 208 bp (3187-3394 in AJ318490) immediately before 5'UTR of *DGAT1*. The position of the 5'VNTR is indicated as black box. Criteria were: Observed/Expected ratio > 0.60, C+G > 50%, Length > 200 bp).

4.1.4 Polymorphisms in DGAT1

Screening for polymorphisms

Re-sequencing of *DGAT1* in a panel including 10 individuals and seven DNA pools (see appendix 9.3) revealed 21 SNPs and a variable number of tandem repeats (VNTR: DG1) in the upstream sequence (Table 4.4). Eighteen SNPs were found in noncoding regions and one SNP in exon 8 was silent (SNP 254). The double substitution (SNPs 252 and 253) located in exon 8 represents a missense mutation, resulting in a substitution of lysine to alanine at residue 232 ($Lys^{232} \rightarrow Ala$). In addition, variable PCR amplification (primer numbers 1618 and 1678) was observed in the region of intron 10. Good PCR amplification was associated with the $lysine^{232}$ allele, while weak or no PCR amplification was associated with the $alanine^{232}$ allele. The region could not be resolved by sequencing. Fragmentary sequence information suggested a longer G-stretch within intron 10 that could be variable in length.

Table 4.4: Polymorphisms in bovine *DGAT1*.

Locus	SNP		Region	Accession	Position	Allele		Effect
	snp_id[a]	lab name				1	2	
DGAT1	-	DG1	5'end	AJ318490	1465	Repeat[b]		
	28	DG2	5'end	AJ318490	3343	C	G	
	33	DG3	5'UTR	AJ318490	3399	T	G	
	305	DG21	intron 1	AJ318490	7115	T	G	
	62	DG4	intron 1	AJ318490	7233	A	G	
	63	DG5	intron 2	AJ318490	8491	T	C	
	169	DG6	intron 2	AJ318490	8567	A	G	
	237	DG7	intron 2	AJ318490	8607	G	A	
	230	DG8	intron 2	AJ318490	9284	C	T	
	250	DG9	exon 6	AJ318490	10034	C	T	silent
	251	DG10	intron 6	AJ318490	10147	A	C	
	252, 253	**DG11**	**exon 8**	**AJ318490**	**10433-4**	**GC**	**AA**	**Ala-Lys**
	254	DG12	exon 8	AJ318490	10486	C	T	silent
	255	DG13	intron 8	AJ318490	10515	G	del	
	258	DG14	intron 12	AJ318490	11030	G	A	
	259	DG15	intron 12	AJ318490	11048	C	T	
	260	DG16	3'UTR	AJ318490	11993	T	C	
	261	DG17	3'UTR	AJ318490	12005	A	C	
	262	DG18	3'UTR	AJ318490	12036	T	C	
	263	DG19	3'UTR	AJ318490	12056	A	G	
	264	DG20	3'UTR	AJ318490	12136	G	A	

[a]SNP_id refers to SNPZoo entry; http://www.snpzoo.de/ (Fries *et al.* 2001).
[b]Observed number of repeat unit (AGGCCCCGCCCTCCCCGG): 1 to 7 times.
[c]Pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicate high and low breeding values for milk fat percentage, respectively.

Genotyping of polymorphisms in pooled DNA samples and individual animals

Allele frequencies were estimated based on sequencing traces for pooled DNA samples (Figure 4.7 A and Table 4.5). RFLP genotyping of each individual in the pools for SNP 252, which causes the $Lys^{232} \rightarrow$ Ala substitution, resulted in an observed allele frequency that deviated from the estimated values by less than 10% (Figure 4.7 B). For both the German Simmental and German Holstein breeds, significant ($\alpha = 0.001$) differences in frequency existed for SNPs in the *DGAT1* gene between the pooled DNA samples with high and low breeding values. The German Holstein breed showed the extreme differences. Allele 1 of SNP 28 was fixed in the German Holstein breed and was the predominant allele in German Simmental and German Brown. In both breeds, the $lysine^{232}$-encoding allele was more frequent in animals with high breeding values. The $lysine^{232}$-encoding allele was also present in German Brown animals from the high end of the distribution of the milk fat percentage breeding values, although to a lesser extent than in the other breeds ($\alpha = 0.05$). This is remarkable considering the low frequency of the lysine allele (2%) in German Brown (Figure 4.8). Allele frequencies for the $lysine^{232}$ allele vary within *Bos taurus* breeds between 0% (Pinzgauer) and 71% (Jersey). Only a few animals of the *Bos indicus* breeds and from yak and water buffalo were genotyped; all showed the $lysine^{232}$ allele exclusively.

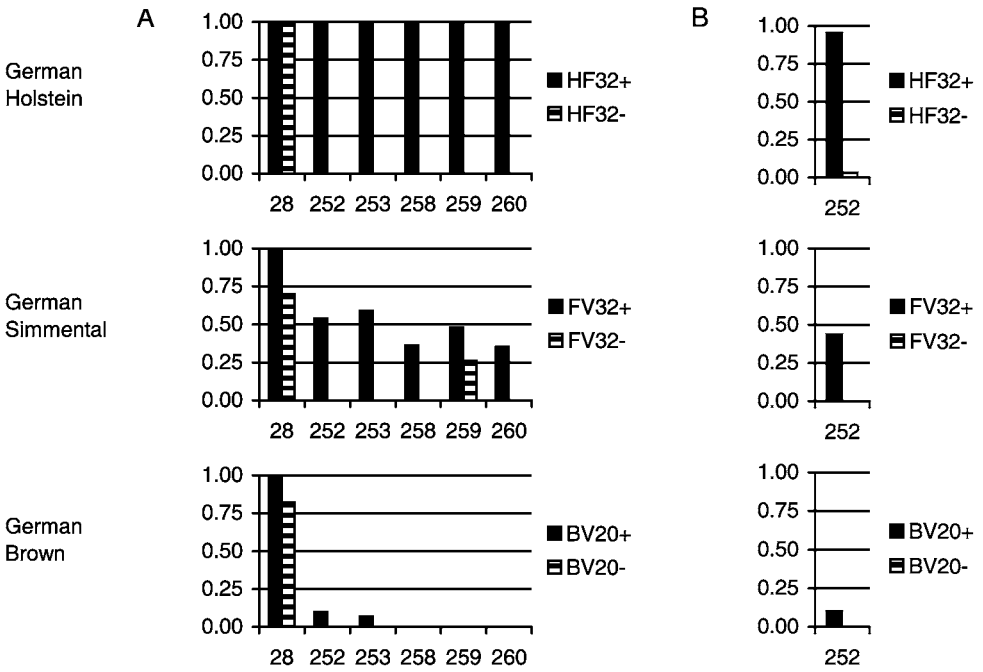


Figure 4.7: Allele frequencies of *DGAT1* SNPs in pooled DNA samples.

A: Allele frequencies estimated based on sequence traces of pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicate high and low breeding values for milk fat percentage, respectively; **B**: Observed allele frequencies of SNP 252 obtained by RFLP genotyping of each individual in the pools.

Table 4.5: Allele frequencies of *DGAT1* SNPs in pooled DNA samples.

			German Holstein				German Simmental				German Brown			
SNP		Allele	Frequency HF32+	HF32-	G^a	α-value	Frequency FV32+	FV32-	G^a	α-value	Frequency BV20+	BV20-	G^a	α-value
28	5'end	C	1.00	1.00			1.00	0.70	22.59	0.001	1.00	0.82	7.91	0.01
230	intron 2	C					0.46	0.08	23.44	0.001	0.10	0.00	4.21	0.05
252	exon 8	G^b	1.00	0.00	128.00	0.001	0.54	0.00	47.34	0.001	0.10	0.00	4.21	0.05
258	intron 12	G	1.00	0.00	128.00	0.001	0.36	0.00	28.10	0.001	0.00	0.00		
259	intron 12	C	1.00	0.00	128.00	0.001	0.48	0.26	6.64	0.01	0.00	0.00		
260	3'UTR	T	1.00	0.00	128.00	0.001	0.35	0.00	27.15	0.001	0.00	0.00		

[a] When testing the null hypothesis H_0 (no association between SNP alleles and the breeding value for milk fat percentage) the test statistic G follows a χ^2-distribution with one degree of freedom..

[b] $Lys^{232} \rightarrow Ala$

Species	Breed	no.	Lys [%]	Ala [%]
Bos taurus taurus	German Holstein	40	34	66
	German Simmental	66	7	93
	German Brown	48	2	98
	Original Brown	8	6	94
	Orginal Simmental	35	13	87
	Pinzgauer	42	0	100
	Anatolian Black	39	28	72
	Jersey	7	71	29
Bos taurus indicus	Hariana	7	100	0
	Sahival	5	100	0
	Tharparkar	4	100	0
Bos grunniens	Yak	2	100	0
Bubalus bubalus	Water buffalo	2	100	0

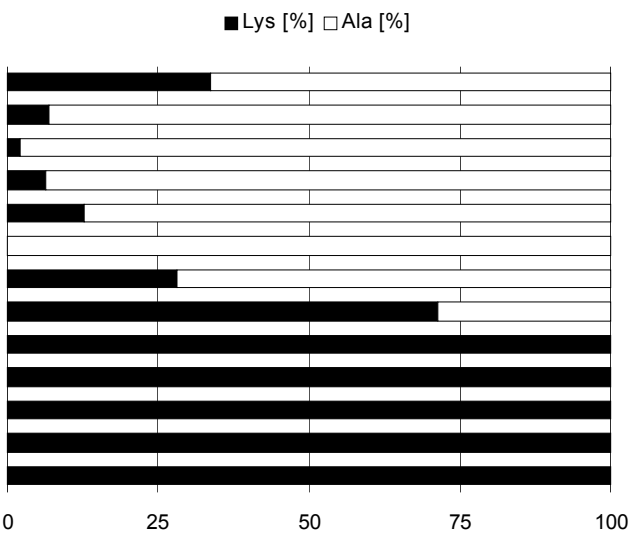


Figure 4.8: Allele frequencies of DGAT1 lysine^{232} allele in different species and breeds.
Observed allele frequencies obtained by RFLP genotyping of each individual for SNP 252, which is responsible for the Lys^{232} → Ala substitution.

Individual animals of the different breeds and from the pooled DNA samples were genotyped for the VNTR in the upstream sequence of *DGAT1* (Table 4.6). The number of repeat units ranges in *Bos taurus* from 3 to 7, which more than 80% of the alleles contain the 4, 5 or 6 units. The four alleles of water buffalo had only one repeat unit.

Table 4.6: Allele frequencies of the VNTR in the upstream sequence of bovine *DGAT1*.

Category[a]	Animals	Alleles typed	Number of repeat units 1	2	3	4	5	6	7	n.d.
HF32+	32	62	-	-	-	-	0.95	0.02	0.03	-
HF32-	32	60	-	-	0.05	0.25	0.17	0.42	0.12	-
FV32+	32	64	-	-	0.02	0.23	0.33	0.42	-	-
FV32-	32	64	-	-	-	0.70	0.14	0.16	-	-
BV20+	20	38	-	-	-	0.08	0.74	0.18	-	-
BV20-	20	40	-	-	-	0.20	0.78	0.03	-	-
Angus	1	2	-	-	-	-	0.50	0.50	-	-
Kerry	1	2	-	-	-	0.50	-	0.50	-	-
Hariana	1	2	-	-	-	-	0.50	0.50	-	-
Sahival	1	2	-	-	-	1.00	-	-	-	-
Jersey	7	12	-	-	-	0.83	0.17	-	-	-
Yak	2	4	-	-	-	1.00	-	-	-	-
Water buffalo	2	4	1.00	-	-	-	-	-	-	-

[a]Pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicate high and low breeding values for milk fat percentage, respectively.

Polymorphisms in noncoding regions

Allele frequencies of most detected SNPs within *DGAT1* were correlated with variation in milk fat percentage (Figure 4.7 A). Most of the SNPs were located within noncoding regions without a direct effect on the peptide sequence. However, they might still influence gene expression. First, polymorphisms in promoter motifs might affect transcription. In addition to the promoter region in the upstream sequence of a gene, promoter regions within introns have

also been reported. For example, the human p53 oncogene contains a promoter within intron 1 (Reisman *et al.* 1988). Second, polymorphism within introns can affect RNA splicing. Besides the conserved sites at both the beginning (GT) and end (AG) of each intron, a third feature, termed the "branch point", is necessary for correct pre-mRNA splicing (e.g. Hastings *et al.* 2001)), with the 5' end of the intron looping and binding to the adenine of the branch point sequences (BPS) approximately 25 nucleotides upstream of the 3' end of the intron. The consensus sequence for branch points in mammals is YNCURAY (where Y is a pyrimidine, R is a purine, and N is any base, Keller *et al.* 1984). None of the detected SNPs within bovine *DGAT1* were located within one of these features known to be necessary for pre-mRNA splicing.

VNTR in the DGAT1 upstream sequence

Results of CpG island mapping with GenBuilder (www.itba.mi.cnr.it/webgene, Milanesi *et al.* 1999), revealed an additional CpG-rich region with a ratio of Observed/Expected = 0.59 (2571-2837 in AJ318490) containing the VNTR (Figure 4.6). Since CpG islands are involved in gene regulation (Attwood *et al.* 2002), the VNTR before bovine *DGAT1* gene might influence the expression of *DGAT1* depending on the number of repeat units in it. However, preliminary examination of German Holstein, German Simmental and German Brown bulls showed no correlation between any of the different VNTR alleles in the upstream sequence of *DGAT1* with the breeding value for milk fat percentage.

Lysine^{232} → Alanine substitution in exon 8

The polymorphism (SNP 252 and 253) in exon 8 represents a missense mutation, resulting in a substitution of lysine (a basic residue) to alanine (a nonpolar residue). Conservation at this position of lysine was observed over the different orders of mammals (Figure 4.4A). The Lys^{232} → Ala substitution in cattle is located within a region of *DGAT1* showing overall a lower degree of conservation across the species in Figure 4.3. Therefore, it seems unlikely that an essential motif for the DGAT activity is directly affected by this substitution. In close proximity to the Lys^{232} → Ala substitution is the HSF motif. Substitution of the central serine in hamster ACAT1 results in complete loss of activity. The substitution of Lysine^{232} in DGAT1 could not have the same dramatic effect, since individuals homozygous for lysine as well as for alanine are present. However, the substitution of lysine^{232} by alanine may still reduce DGAT1 activity.

4.1.5 Lys^{232} → Ala substitution is associated with milk fat percentage variation

Alternative hypotheses to the one that Lys^{232} → Ala substitution in DGAT1 represents the milk fat QTL on chromosome 14 are

- the observed association is a relict of population admixture (Lander *et al.* 1994)
- or the causal variant is located in another gene neighboring *DGAT1* and in linkage disequilibrium with the *DGAT1* variants.

Is the observed association a relict of population admixture?

Population admixture cannot be ruled out for all three breeds investigated when considering the frequencies of the lysine^{232}-encoding allele (Figure 4.8 on page 46) and the history of the breeds. Admixture in German Holstein and German Brown could have resulted from cross-breeding with Jersey, a breed displaying both high milk fat percentage and high frequency of the lysine^{232} allele. Admixture in German Simmental could have resulted from the introgression of German Holstein. However, pedigree analysis revealed that animals with German Holstein ancestry could be found in both the "positive" and the "negative" pools. This argues against admixture being responsible exclusively for the observed association.

Additionally, the presumption that the association with the breeding value for milk fat percentage arises through a number of rare alleles identical by descent can be ruled out since the lysine allele was found in unrelated individuals of the positive DNA pools of German Simmental and German Holstein.

Another argument for *DGAT1* (or linked loci) being responsible for the QTL-variation on chromosome 14 was provided by QTL mapping in two different breeds. QTL for milk fat percentage was detected initially in the Holstein breed. However, the QTL for milk fat percentage was also observed in German Simmental using a granddaughter design. The result indicates the most likely position of the QTL close to marker *ILSTS039* on chromosome 14 (Figure 4.9).

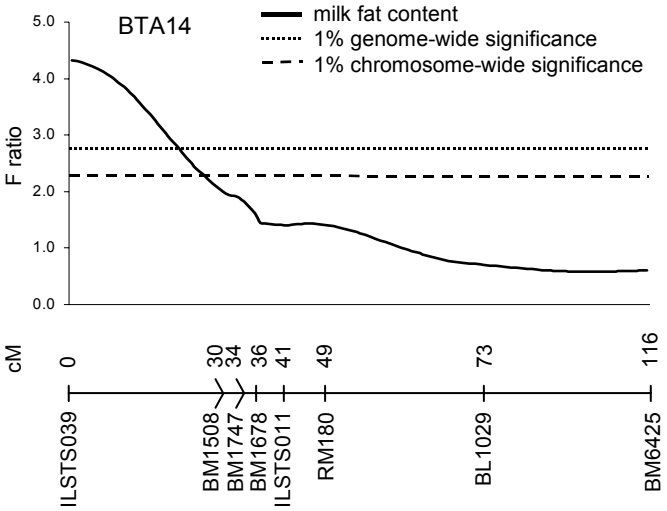


Figure 4.9: Result of QTL mapping on bovine chromosome 14.

Across family test statistic curve for QTL analyses of milk fat percentage on chromosome 14 for a German Simmental granddaughter design. F-ratios testing for the presence of a segregating QTL are plotted for given positions along the chromosome. The marker map with distances in centimorgans (cM) between markers is shown on the x-axis. Empirical chromosome-wide and genome-wide 1% significance levels achieved via 10 000 permutations are indicated as horizontal lines.

Evidence was highly significant for segregation of the QTL in 2 out of 20 families (Figure 4.10 A) with estimates of QTL effects for milk fat percentage being 0.313 ± 0.070 and 0.409 ± 0.064, respectively. These allele substitution effects greatly exceed the genetic standard

deviation of about 0.2 in the German Simmental population and account for about 10% of the phenotypic variation in this breed (based on a frequency of 0.07). The genotypes for the predicted Lys^{232} → Ala substitution determined by an RFLP assay were compatible with the heterozygous status of the segregating (Qq) sires and homozygosity of the alanine-encoding variant of the non-segregating (most likely qq) sires (Figure 4.10 B).

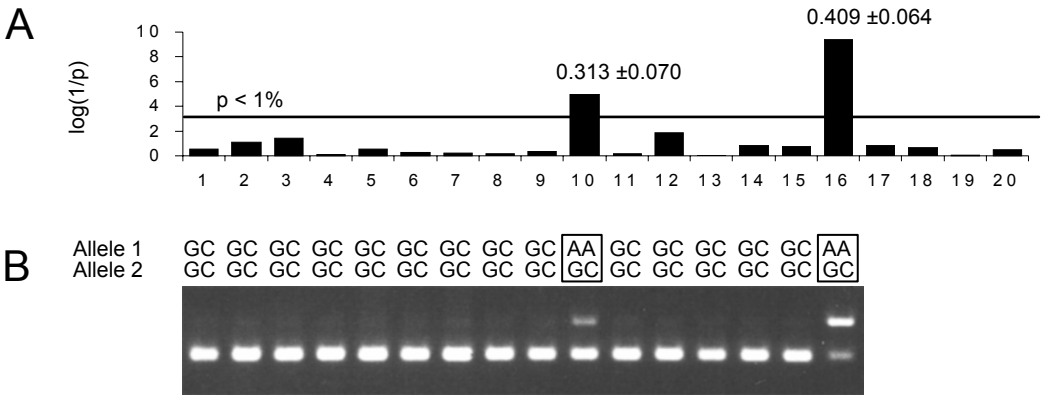


Figure 4.10: Segregating sires tested by RFLP at Lys^{232} → Ala position.
A. Bars show transformed significance levels (log (1/p)) of the test statistic for a segregating QTL present at 0 cM within each family (x-axis). The horizontal line indicates the transformed 1% significance level for a single family after correcting for multiple testing of 20 families. QTL-effects for milk fat percentage and their respective standard errors are shown on top of the bars for significantly segregating sires. **B.** RFLP genotyping of SNP 252 (Lys^{232} → Ala) of the DGAT1 gene by *Cfr*I-cleavage in a 411 bp PCR product from bovine genomic DNA of sires 1 to 16. Cleavage by *Cfr*I is diagnostic for the allele encoding alanine^{232} (GC). No DNA samples were available for sires 17 to 20.

Genotyping by direct sequencing of *DGAT1* from DNA and determining the repeat number of the 5'-VNTR in the two segregating German Simmental bulls (sample-id 705 and 899) and some of their progeny (appendix 9.8) allowed the derivation of their haplotypes based on the genotypes of homozygous progeny. The lysine-encoding variant was present on two different haplotypes. German Simmental bull 16 (sample-id: 899) carried the only lysine^{232}-encoding haplotype found in German Holstein. German Simmental bull 10 (sample-id: 705) carried a lysine-encoding haplotype found in German Simmental, Anatolian Black and Sahival (Table 4.7 and Figure 4.11). This could indicate that a lysine^{232}-encoding allele had been introduced into German Simmental from German Holstein. Pedigree analysis indeed showed that the great-grandfather of bull 16 was a purebred Holstein-Frisian animal, while there was no indication of Holstein ancestry for bull 10. Three (SNP 28, 258 and 260) of the seven variable positions that make up the haplotypes were homozygous for the QTL in the heterozygous (Qq) bull 10 (Figure 4.11). Thus, they can be excluded as being causal. The locus representing the Lys^{232} → Ala polymorphism, however, is heterozygous in both Qq bulls.

Table 4.7: SNP haplotypes within *DGAT1*.

	28	252	258	259	260
	T	A	G	C	C
	C	A	G	C	C
	C	A	G	C	T
	C	A	A	C	C
	C	A	A	C	T
	C	G	G	C	T
	C	G	G	T	T
	G	G	G	T	T
German Simmental		x	x	x	x	x	x
German Holstein			x			x	
German Brown						x	x
Anatolian Black		x	x			x	
Jersey		x		x		x	
Hariana					x		
Sahival			x				
Yak		x					
Water buffalo	x						

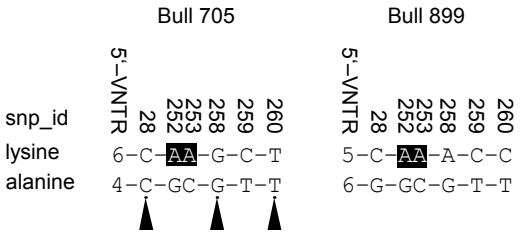


Figure 4.11: Haplotypes of two heterozygous bulls. Arrows indicate homozygous SNP positions within bull 705, which segregates for the milk fat percentage QTL. Bull 705 and bull 899 represent sire 10 and 16, respectively, in the German Simmental granddaughter design.

Is the observed association due to linkage with the causal variation in a neighboring gene?

Frequency shifts support this hypothesis although there was no complete correlation within the German Simmental pools and only some animals carried the lysine^{232} allele in the positive German Brown pools. To reject the hypothesis that *DGAT1* Lys^{232} → Ala substitution is only closely linked with the causal mutation, genes neighboring *DGAT1* were examined.

4.1.6 Genes neighboring DGAT1

Genes identified from the corresponding human genomic draft sequence

Bovine EST sequences were available for 23 out of the 31 human genes that fall within a 640 kb region centered on *DGAT1* (Table 4.8) as indicated by the draft sequence available as of December 2001 (built 28). Together with human genes that were listed in previous draft versions (builds 26 and 27) and with genes identified by BLAST using BAC ends, 29 genes (Table 4.9) were tested as to whether they were located in the bovine BAC contig (see next chapter).

Table 4.8: Summary statistic of database search for genes neighboring *DGAT1*.

Region relative to human *DGAT1* (NCBI MapView build 28)	Total number of genes[a]	Genes used for BLAST search	Genes with bovine ESTs	Genes within bovine contig
-330 kb to -200 kb	10	6[b]	6	0
-200 kb to 200 kb	32	22[c]	14	14
200 kb to 350 kb	7	3[b]	3	0

[a]Numbers based on MapView build 28 (December 2001).
[b]Only mRNA of confirmed gene models (model based on alignment of mRNA to the genomic sequence) were used for BLAST search.
[c]Only gene models containing introns with several hundreds of bps were considered because of the higher likelihood of SNP in intronic sequences.

Table 4.9: GenBank accession numbers of bovine ESTs for genes neighboring *DGAT1*.

Symbol (* not in contig)	Alternative Symbols	Locus type[a]	Product	Accession no.	
				Human mRNA	Bovine EST or mRNA**
BOP1	*KIAA0124*	1	block of proliferation 1	NM_015201	AV605047, AV606209, BF602301, BF655249, BE682192, AV605046
CPSF1	*CPSF160, HSU37012*	1	cleavage and polyadenylation specific factor 1, 160kD subunit	NM_013291	X83097**
CYC1		1	cytochrome c-1	NM_001916	U97172**, BF600225, BF602111
CYHR1	*CHRP, KIAA0496*	1	cysteine and histidine rich 1	XM_035349	AV605047, BF776617, BF706051, AW653372, BI776026, AW776992, BF601268, BF076560
FBXL6	*FBL6, FLJ22888*	1	F-box and leucine-rich repeat protein 6, isoform 1 and 2	NM_012162	AV610804, AV611064, BF072969, BE588919
FOXH1	*FAST1, FAST-1*	1	forkhead box H1	NM_003923	BE664973
GPT	*AAT1, ALT1, GPT1*	1	glutamic-pyruvate transaminase (alanine aminotransferase)	NM_005309	BE752933, BF072844, BG834947
HSF1	*HSTF1*	1	heat shock transcription factor 1	NM_005526	AW655211, BE487647, AW655600, BE757045, BF191918, BI340115, BI340425, BF441279, BM256130
*LOC58500**		1	zinc finger protein (clone 647)	XM_035324	AW659101, BF230592
NFKBIL2	*IKBR*	1	I-kappa-B-related protein	NM_013432	BF076286, BE756860, BE752196, BE756348 BF776617
PPP1R16A	*MYPT3, MGC14333*	1	protein phosphatase 1, regulatory (inhibitor) subunit	NM_032902	BF045030, BF046230
RECQL4	*RTS, RECQ4*	1	RecQ protein-like 4	NM_004260	BE756255, BI535313, BE683314, BE723182
SLC39A4	*ZIP4, FLJ20327*	1	solute carrier family 39 (zinc transporter), member 4	NM_017767	AW477707
TSTA3	*FX*	1	tissue specific transplantation antigen P35B	NM_003313	AV604540, BM030330, BE685454
VPS28	*LOC51160*	1	VPS28: vacuolar protein sorting 28 (yeast)	NM_016208	AV611485, BM286370
DKFZp547 F072		2	hypothetical protein DKFZp547F072	NM_032274	AV663367
FLJ11856	*GPCR*	2	hypothetical protein FLJ11856, putative G-protein coupled receptor GPCR41	NM_024531	BF041390, BI775510, AW657006
*FLJ12150**		2	hypothetical protein FLJ12150	NM_024736	BF706408, BM030137, AV617378, BF654873
*FLJ13852**		2	hypothetical protein FLJ13852	NM_023078	BF707143, BI849502, BE808164
*FLJ20897**		2	hypothetical protein FLJ20897	NM_032378	AV610684, BM362121
KIAA0014		2	KIAA0014 gene product	NM_014665	BE588417, BF076371, BF603009, AW483901, BE684636, BF776619, BI849721
*KIAA0628**		2	KIAA0628 gene product	NM_014789	BF776617, BM253919, BF776617
KIAA1833	*LOC84500*	2	KIAA1833 protein	AB058736	BE751545, BE750349, BF074883, BG689571, BM089242, BF776617, BI774984
MGC10520		2	hypothetical protein MGC10520	NM_030580	BF774052, BF230746, AW462983
MGC13010		2	hypothetical protein MGC13010	NM_032687	BE236645, BI682991
*PP3856**	*LOC93100*	2	similar to CG3714 gene product	XM_049247	BM251577, BM258954, BF599727, BM431222, BM431271
RRP41	*FLJ20591*	2	exosome component Rrp41	NM_019037	AV588786, BI536922
*LOC90979**		3	similar to hypothetical protein FLJ14855	XM_035323	BF776617, BE236854
LOC157534		0	similar to CG7616 gene product	XM_088320	BF073939, BF773808, BM107277, BM255011

[a]Locus Type: 1 gene with protein product of function known or inferred
2 gene with protein product of function unknown
3 model, supported by mRNA and EST alignments
0 This record was removed at the submitters request

Protein tyrosine kinase 2 (PTK2)

DGAT1 may be post-translationally regulated by tyrosine kinase (Rohlfs *et al.* 1993). *Protein tyrosine kinase 2* (*PTK2*) mapped 20.3 $cRad_{5000}$ next to *DGAT1* on BovRH5 panel (Table 4.2, page 37). The distance between *DGAT1* and *PTK2* in the human draft sequence (NCBI MapView build 30) is 2.3 Mb. *PTK2* was found not to be in the bovine contig (see below), indicating that *PTK2* is at least 250 kb away from *DGAT1*. One PCR fragment was sequenced from *PTK2* (Table 4.15, page 59). The high number of small exons and large introns of several kb complicated direct sequencing of PCR products. One SNP was identified in intron 7 (Table 4.16, page 61), but allele frequencies of this SNP (Figure 4.12 and Table 4.10) show only weak correlation with the variation in milk fat percentage in German Holstein $\alpha = 0.05$). However, several SNPs in form of haplotypes would be necessary to draw final conclusions in this regard.

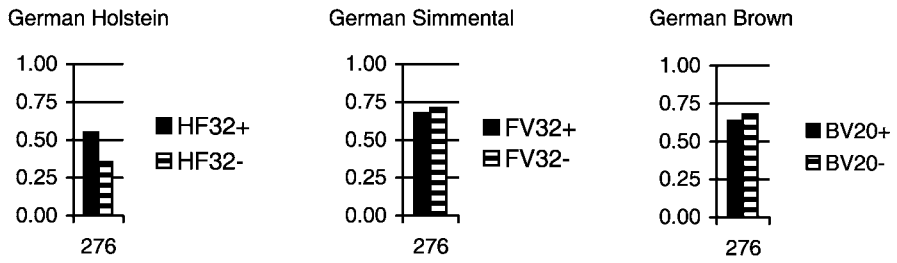


Figure 4.12: Allele frequencies of *PTK2* polymorphism in pooled DNA samples.
Allele frequencies were estimated based on sequence traces of pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicate high and low breeding values for milk fat percentage, respectively.

Table 4.10: Allele frequencies of *PTK2* SNP 276 in pooled DNA samples.

			German Holstein				German Simmental				German Brown			
SNP		Allele	Frequency				Frequency				Frequency			
			HF32+	HF32-	G^a	α-value	FV32+	FV32-	G^a	α-value	BV20+	BV20-	G^a	α-value
276	intron 7	T	0.55	0.355	4.91	0.05	0.675	0.71	0.18	no	0.635	0.68	0.18	no

[a]When testing the null hypothesis H_0 (no association between SNP alleles and the breeding value for milk fat percentage) the test statistic G follows a χ^2-distribution with one degree of freedom.

Further candidate genes

Ashrafi *et al.* (2003) used RNA-mediated interference (RNAi) to disrupt the expression of genes and to screen the *Caenorhabditis elegans* genome for fat regulatory genes. Inactivation of 305 and 112 genes caused reduced and increased fat storage, respectively. A subset of these 417 genes were chosen according to their function and checked to see if one of them was located in the human chromosomal region (8q24.3) corresponding to the region of bovine milk fat percentage QTL on chromosome 14:

- protein and tyrosine kinases (WormBase accessions: F46G11.3, K10D3.5, M01B12.5, T04B2.2, T04B2.2, W08D2.1, W08D2.1 and ZC504.4),

- zinc finger proteins (WormBase accessions: C46E10.9, C47C12.3, C56E10.4, F16B4.9, F55B11.4, F56F3.4, F56F3.4, H12C20.3, T09F3.1, T23F11.4, ZK666.1 and ZK686.4)
- further genes (WormBase accessions: AH10.1, C01C10.3, C01G6.5, C02F4.2, C04G2.4, C04G2.4, C24F3.2, C33A12.6, F11E6.5, F46C5.6, F49C5.6, F54C9.9, K05F1.3, K10B3.7, T12B5.8, T12B5.8 and Y44A6B.2).

The gene encoding zinc finger C2H2 type domain (WormBase accession: F55B11.4) was found to reduce the fat content in the *Caenorhabditis elegans* test strain. Human orthologous gene (secreted Ly-6/uPAR related protein 1 precursor, SLURP-1) was located 1.1 Mb proximal to *DGAT1* on the human gene map (ENSEMBL, January 2003). No bovine EST sequences were found in NCBI EST database for this gene.

4.1.7 BAC contig of the DGAT1 region

BAC clones

Screening of the bovine BAC library RPCI-42 resulted in 19 clones (Table 4.11) that were positively tested by colony-PCR to be specific for *DGAT1* and neighboring loci.

Table 4.11: Bovine BAC clones covering *DGAT1* region.

Probe	Clone	Internal lab number	Position of BAC insert
DGAT1	240A1	762	DGAT1
	258E13	767	*DGAT1*
	269H17	762	*DGAT1*
	56F1	761	*DGAT1*
BAC ends of clones containing *DGAT1*	293G16	886	distal to *DGAT1*
	334E6	885	*DGAT1*
	352D2	887	proximal to *DGAT1*
	360L24	888	*DGAT1*
	410E24	893	proximal to *DGAT1*
	414O23	895	*DGAT1*
	428F15	892	*DGAT1*
	428P15	891	distal to *DGAT1*
	521O21	901	proximal to *DGAT1*
	557K4	904	distal to *DGAT1*
Genes neighboring *DGAT1*	100P18	876	proximal to *DGAT1*
	111I3	878	proximal to *DGAT1*
	156I10	881	proximal to *DGAT1*
	3O1	874	proximal to *DGAT1*
	78M13	875	proximal to *DGAT1*

BAC-insert size and fingerprinting

As determined by digestion with *Not*I (Figure 4.13), the average insert size was 162 kb (range 64–214 kb), in good agreement with published estimates for this BAC library of 163 kb for segment 1 and 165 kb for segment 2 (Warren *et al.* 2000). Overall, 21 genomic *Not*I sites were identified within the BAC contig, which spans 576 kb (Table 4.12).

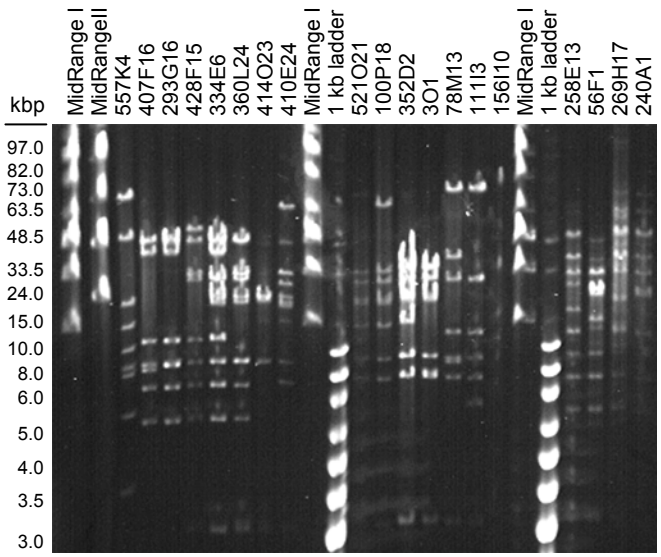


Figure 4.13: *Not*I digested BAC DNA separated by pulsed-field gel electrophoresis.

Table 4.12: Content of *Not*I fragments of 18 bovine BAC clones covering *DGAT1* region.

Clone	Internal lab no.	*Not*I fragments [kb]																					Insert size
		4	55	24	11	16	8	6	76	12	53	6	8	34	27	3	38	24	34	8	76	12	[kb]
557K4	904	4	55	24	11	16	8	6	76														200
428F15	892								59	12	53	6	8	34	35	3							209
407F16	886								49	12	53	6	8	8									135
293G16	891								49	12	53	6	8										127
334E6	885								49	12	53	6	8	34	27	3	24						214
269H17	763								41	12	53	6	8	34	22								175
56F1	761								41	12	53	6	8	34	27	3	18						201
360L24	888										53	6	8	34	27	3	38	24					192
240A1	762										15	6	8	34	27	3	24						116
414O23	895													34	27	3							64
410E24	893														27	3	38	24	34	8	73		206
100P18	876														15	3	38	24	34	8	73		194
521O21	901														15	3	38	24	34	8	76		197
352D2	887														15	3	38	24	34	8			121
3O1	874															3	38	24	34	8			106
156I10	881																	24	34	8	76		141
111I3	878																	3	34	8	76	12	6	138
78M13	875																	3	34	8	76	12	41	173

Numbers with black backgrounds represent actual *Not*I fragments; numbers with gray backgrounds indicate fragments assumed to be partial *Not*I fragments resulting from *Eco*RI cloning.

The number of *Not*I recognition sites found within the BAC clones is remarkably high, with 5.6 *Not*I sites per clone. Previously reported values for *Not*I sites within the bovine genome are much lower: 0.15 per clone (189 clones with an average insert size of 105 kb, Zhu *et al.* 1999) and 0.19 per clone (32 clones with an average insert size of 146 kb, Cai *et al.* 1995). Since *Not*I has a recognition sequence of eight nucleotides (GCGGCCGC), the average fragment size in a random sequence would be 4^8 = 65 536 bp. However, the average fragment size within the BAC contig was found to be 27 kb. The higher than expected number of cut-

ting sites in the contig indicates a high GC content, and might be associated with a high density of CpG islands and therefore high gene content in the investigated chromosomal region. Rare-cutting (C-G) restriction enzymes were suggested to detect CpG islands within BAC DNA (Lindsay *et al.* 1987).

BAC-end sequencing

Sequencing and BLAST analysis of the 38 BAC ends revealed 11 BAC ends to be identical, six to be located within genes, and two to be within bovine cDNAs that had no corresponding human sequences available in the NCBI sequence database (Table 4.13). Seven BAC ends contained repetitive sequences, mainly bovine SINE sequences, while 10 BAC-end sequences did not show similarity to sequence entries in GenBank and were used as STS markers. From one BAC clone, no BAC end sequences were obtained.

Table 4.13: BAC-end sequences and BLAST results.

		T7 cloning site			SP6 cloning site		
Clone	Orient.	Identical with BAC-end	Gene	Repeat	Identical with BAC-end	Gene	Repeat
557K4	?			SINE Bov-2			Repeat
428F15	-						
407F16	-	*293G16-T7*			*293G16-SP6, 334E6-SP6*		
293G16	-	*407F16-T7*		LINE BovB-B	*407F16-SP6, 334E6-SP6*	*BF077085*	SINE Bov-2
334E6	-				*407F16-SP6, 293G16-SP6*		
269H17	+	*56F1-T7*				*FBXL6*	
56F1	+	*269H17-T7*	*RRP41*				
360L24	-		*BE667943*	SINE Bov-tA			Repeat
240A1	-			SINE Bov-2		*KIAA1833*	
414O23	+			Repeat			
410E24	+				*100P18-T7*		
100P18	-	*410E24-SP6*			*352D2-T7, 521O21-SP6*	*FBXL6*	
521O21	-	*156I10-T7*			*100P18-SP6, 352D2-T7*		
352D2	+	*100P18-SP6, 521O21-SP6*	*FBXL6*		*3O1-SP6*		
3O1	+				*352D2-SP6*		
156I10	-	*521O21-T7*		SINE Bov-tA		*NFKBIL2*	
111I3	-			Repeat	*78M13-SP6*		
78M13	-		*MGC10520*		*111I3-SP6*		

Gene and STS content of BAC clones

The gene and STS contents of 18 BAC clones were assessed by locus specific PCR (Figure 4.14 and Table 4.14).

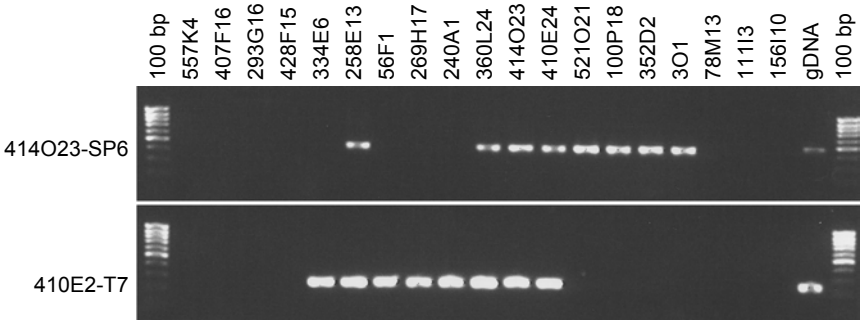


Figure 4.14: Colony PCR results for two loci.

Table 4.14: Gene and STS content of 18 BAC clones.

	428F15-SP6	BF077085	Rrp41	CYC1	KIAA1833	BOP1	DKFZp547F072	DGAT1	HSF1	410E24-T7	FBXL6	FLJ1856	56F1-SP6	3O1-T7	FLJ35454	CPSF1 (3')	334E6-T7	CPSF1 (5')	240A1-T7	SLC39A4	414O23-SP6	NFKBIL2	BE667343	VPS28	PPP1R16A	FOXH1	KIAA0014	MGC13010	KIAA0496	GPT	RECQL4	3O1-SP6	100P18-SP6	111I3-T7	MGC10520
Orient.[a]	+	+	+		+					+	-		-	+		-	-	-	-		-		-									-	-	-	+
557K4	x	x	x	x																														
428F15	x	x	x	x	x	x	x	x	x																									
407F16		x	x	x	x	x																												
293G16		x	x	x	x	x																												
334E6		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x																	
269H17			x	x	x	x	x	x	x	x	x																							
56F1			x	x	x	x	x	x	x	x	x	x	x																					
360L24					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x											
240A1					x	x	x	x	x	x	x	x	x	x	x	x	x	x																
414O23							x	x	x	x	x	x	x	x	x	x	x	x	x	x	x													
410E24										x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
100P18												x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
521O21												x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
352D2												x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
3O1														x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
156I10																						x	x	x	x	x	x	x	x	x	x	x		
111I3																								x	x	x	x	x	x	x	x	x	x	
78M13																								x	x	x	x	x	x	x	x	x	x	x

[a]Orientation of loci within the BAC contig: +, sense strand; - antisense strand.

Assembly of the BAC contig

Results of the fingerprint analysis, BAC end sequencing, and gene and STS content analysis were used to assemble 18 clones to a contiguous BAC map (Figure 4.16). This BAC contig spans a region of 576 kb and contains 11 STS markers and 24 genes including two novel bovine cDNAs. The orientation of one clone (557K4) within the contig could not be determined since both BAC-ends contained repetitive sequences.

The order of two gene clusters (*HSF1*, *DGAT1*, *DKFZp547F072*) and (*VPS28*, *PPP1R16A*, *FOXH1*, *GPT*, *RECQL4*, *KIAA0014*, *MGC13010* and *CYHR1*) could not be distinguished because they were present in the same BAC clones. The relative order of *DGAT1* and two neighboring genes was assessed by long range PCR (Figure 4.15); however, the orientation of these three genes within the BAC contig could not be determined experimentally. Instead, the three genes were arranged based on the human draft sequence, and sequencing results in cattle (showing *HSF1* (+) – *DGAT1* (-)) and mouse (showing *BOP* (-) – *HSF1* (+) (Zhang *et al.* 1998b)), and assuming conserved gene order.

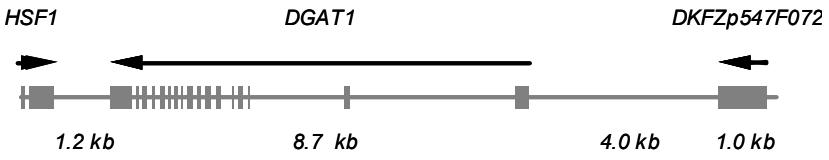


Figure 4.15: Order of *DGAT1*, *HSF1* and *DKFZp547F072*.

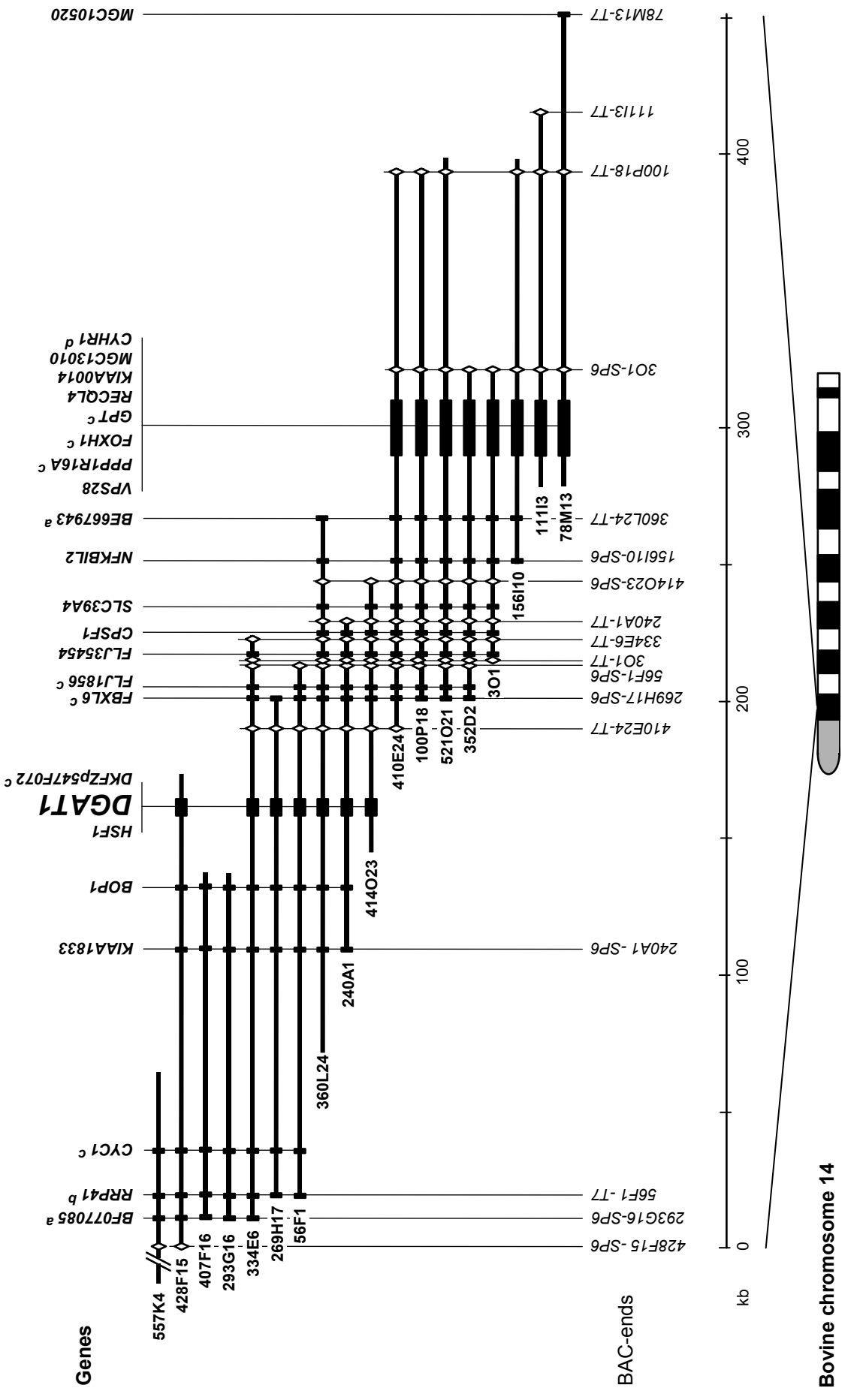


Figure 4.16: BAC contig encompassing the bovine *DGAT1* region.

Legend to Figure 4.16**.**
BAC contig from the centromeric region of bovine chromosome 14 containing *DGAT1* with the actual map of human draft of the chromosomal region (8q24.3) on the top. BAC inserts are shown as bold horizontal lines, with the length of the lines reflecting their size. The name of each clone corresponds to the plate address in the bovine RPCI-42 library. Names of genes are written above the BAC contig and names of STS markers are written below. The positions within the BAC contig of genes and STS markers are indicated as vertical lines. Symbols indicate the gene and STS content of each clone: black rectangles represent genes and diamonds represent STS markers. The horizontal line below the BAC contig indicates distances in relation to the leftmost STS marker (28F15-SP6). Bovine genes without homologous genes (indicated by lines) within the human sequence map: a. no homologous sequence entries in human sequence databases; b. human cDNA sequence available, but not mapped; c. human cDNA sequence available, mapped cytogenically to the human chromosome region 8q24.3; and d. human cDNA sequence available, mapped cytogenically to human chromosome 8.

BAC-end *334E6-T7* was located within intron 2 of *CPSF1* (0.8 kb after exon 1, and 7-8 kb before exon 3, as determined by long range PCR). *Not*I fragment content in combination with gene content analysis and BAC end sequencing to identify identical BAC ends allowed the establishment of the *Not*I fragment assembly in Table 4.12. Overall, the assembly was coherent, except for the following observations:

- Clones 407F16 and 293G16 contained identical inserts (identical BAC-ends at SP6 and overlapping BAC-ends at T7, differing by 335 bp), but 407F16 contained an additional *Not*I fragment of 8 kb. BLAST analysis identified both T7 sequences as repeat sequences (LINE BovB), which may account for the apparent overlap.
- The BAC insert from clone 414O23 could have more than the two larger fragments observed of 27 and 34 kb, which could not be resolved properly.
- BAC insert 428F15 showed a 35 kb fragment that could not be integrated in the *Not*I fragment assembly.
- BAC-ends *269H17-SP6* and *100P18-SP6* were adjacent and were located within intron 2 of FBXL6. Together, the assumed partial *Not*I fragments (22 kb and 15 kb) were 37 kb in length; compared with the *Not*I fragment assembly it should have been 27 kb.
- BAC-ends *56F1-SP6* and *3O1-T7* are adjacent, as assessed by PCR. A lack of overlap was also supported by colony PCR. In contrast, *Not*I fragment analysis indicated an overlap via a 18 kb fragment of *56F1-SP6*.

The BAC contig overlaps with a recently published contig (Figure 1 in Grisart *et al.* 2002a) from the locus *CYC1* (middle of the marker interval BULGE11-BULGE9) on the left side to the identical clone 78M13 (contains BULGE9 and also isolated in this study) on the right side.

Genes within the BAC contig

Twenty-three genes were mapped within the BAC contig next to *DGAT1* (Table 4.9 on page 51). Thirteen genes encoded for proteins with known or inferred function; seven genes were assigned as hypothetical proteins or protein products with unknown function. One mapped

gene represented a gene model, supported by mRNA and EST alignments in human, and two genes that were supported by bovine cDNA alignment have no homologous sequences within NCBI sequence databases. The genes with known function have not been implicated directly with lipid metabolism; however, they could be involved indirectly.

4.1.8 Sequences and Polymorphisms of neighboring genes

On average, 2015 bp of sequence was obtained for each of the 23 genes neighboring *DGAT1* by direct sequencing of BAC DNA and PCR products in a panel including four individuals and four DNA pools (see appendix 9.3). Sequence information for the bovine genes and STS markers neighboring *DGAT1* have been deposited in EMBL (Table 4.15).

Table 4.15: Entries for genes neighboring *DGAT1* as deposited in the EMBL nucleotide sequence database.

Locus	Accession no.	Description
BOP1	AJ518948	*Bos taurus* partial bop1 gene for block of proliferation 1, exons 4-16
CPSF1	AJ518949	*Bos taurus* partial cpsf1 gene for cleavage and polyadenylation specificity factor 1, exons 7-28
CPSF1	AJ518950	*Bos taurus* partial cpsf1 gene for cleavage and polyadenylation specificity factor 1, exons 33-37
CYC1	AJ518951	*Bos taurus* partial cyc1 gene for cytochrome c-1, exons 3-6
CYHR1	AJ518952	*Bos taurus* partial cyhr1 gene for cysteine and histidine rich 1, exons 1-2
CYHR1	AJ518967	*Bos taurus* partial cyhr1 gene for cysteine and histidine rich 1
FBXL6	AJ518953	*Bos taurus* partial fbxl6 gene for F-box and leucine-rich repeat protein 6, exons 2-8
FLJ11856	AJ518954	*Bos taurus* partial gpcr41 gene for putative G-protein coupled receptor 41, exons 2-3
FLJ35454	AJ518956	*Bos taurus* partial ORF FLJ35454 DNA for hypothetical protein
FOXH1	AJ518957	*Bos taurus* foxh1 gene for forkhead box H1, exons 1-3
GPT	AJ518958	*Bos taurus* partial gpt gene for glutamic-pyruvate transaminase
HSF1	AJ518959	*Bos taurus* partial hsf1 gene for heat shock transcription factor 1, exons 2-4
HSF1	AJ518960	*Bos taurus* partial hsf1 gene for heat shock transcription factor 1, exons 5-9
HSF1	AJ518961	*Bos taurus* partial hsf1 gene for heat shock transcription factor 1, exons 12-13
KIAA0014	AJ518962	*Bos taurus* partial ORF KIAA0014 DNA for hypothetical protein, exons 3-4
KIAA0014	AJ518963	*Bos taurus* partial ORF KIAA0014 DNA for hypothetical protein, exon 5
KIAA1833	AJ518964	*Bos taurus* partial ORF KIAA1833 DNA for hypothetical protein, exons 11-12
KIAA1833	AJ518965	*Bos taurus* partial ORF KIAA1833 DNA for hypothetical protein, exons 16-19
KIAA1833	AJ518966	*Bos taurus* partial ORF KIAA1833 DNA for hypothetical protein, exon 20
MGC10520	AJ518968	*Bos taurus* partial ORF MGC10520 DNA for hypothetical protein
MGC13010	AJ518969	*Bos taurus* partial ORF MGC13010 DNA for hypothetical protein
NFKBIL2	AJ518970	*Bos taurus* partial nfkbil2 gene for I-kappa-B-related protein, exons 5-6
NFKBIL2	AJ518971	*Bos taurus* partial nfkbil2 gene for I-kappa-B-related protein, 3' end
PPP1R16A	AJ518972	*Bos taurus* partial ppp1R16A gene for protein phosphatase 1, regulatory inhibitor subunit
PTK2	AJ519780	*Bos taurus* partial ptk2 gene for protein tyrosine kinase 2, exons 7 and 8
RECQL4	AJ518973	*Bos taurus* partial recql4 gene for RecQ protein-like 4, exons 13-16
RRP41	AJ518955	*Bos taurus* rrp41 gene for putative exosome complex exonuclease RRP41, exons 1-3
VPS28	AJ518974	*Bos taurus* partial vps28 gene for putative vacuolar protein sorting 28, exons 6-8
100P18-T7	AJ519351	*Bos taurus* STS RPCI42-100P18-T7
111I3-SP6	AJ519352	*Bos taurus* STS RPCI42-111I3-SP6
111I3-T7	AJ519353	*Bos taurus* STS RPCI42-111I3-T7
240A1-T7	AJ519354	*Bos taurus* STS RPCI42-240A1-T7
293G16-SP6	AJ519355	*Bos taurus* STS RPCI42-293G16-SP6
334E6-T7	AJ519356	*Bos taurus* STS RPCI42-334E6-T7
360L24-T7	AJ519357	*Bos taurus* STS RPCI42-360L24-T7
3O1-SP6	AJ519358	*Bos taurus* STS RPCI42-3O1-SP6
3O1-T7	AJ519359	*Bos taurus* STS RPCI42-3O1-T7
410E24-T7	AJ519360	*Bos taurus* STS RPCI42-410E24-T7
414O23-SP6	AJ519361	*Bos taurus* STS RPCI42-414O23-SP6
428F15-SP6	AJ519362	*Bos taurus* STS RPCI42-428F15-SP6
56F1-SP6	AJ519363	*Bos taurus* STS RPCI42-56F1-SP6

In loci neighboring *DGAT1*, 55 polymorphic positions were found (Table 4.16, for genotypes see appendix 9.8). Thirty-nine lay within 15 genes and 16 lay within seven STS markers. Within exons, 11 SNPs were discovered. Four of them resulted in missense mutations: two conservative substitutions in *CPSF1* (Pro-Ala) and *RECQL4* (Met-Val) and two non-conservative substitutions in *CPSF1* (Thr-Ile) and in *FLJ11856* (Gly-Lys).

Bovine genes with identified missense mutations

Cleavage and polyadenylation specificity factor (CPSF) is a part of a multicomponent complex that is responsible for adding the polyA tail to mRNA (Samiotaki *et al.* 2000). The 160 kDa subunit CPSF1 (Jenny *et al.* 1995) has been shown to interact with AAUAAA motif (Keller *et al.* 1991).

RECQL4 is member of human DNA helicase RecQ gene family (Kitao *et al.* 1999a). DNA helicase unwinds double-stranded DNA into single strands. In cattle, the detected missense mutation (SNP 286) was located in the helicase domain (position 383 in human mRNA sequence, accession number AB026546). Human polymorphisms in this region are associated with the Rothmans-Thomson syndrome (Kitao *et al.* 1999b), a rare autosomal recessive genetic disorder characterized by skin and skeletal abnormalities, short stature, manifestations of premature aging, and increased risk of mesenchymal tumors (Kitao *et al.* 1999a).

The products of *CPSF1* and *RECQL4* have essential functions within the organism. Altering the activities of these genes would affect overall viability. Thus, it is not likely that detected mutations will have major effects on the variation in milk fat percentage.

FLJ11856 (*GPCR41*) belongs to the gene family of G-protein coupled receptors (GPCR), which currently has approximately 2000 known members. GPCRs are involved in the recognition and transduction of messages as diverse as light, Ca^{2+}, odorants, and small molecules including amino-acid residues, nucleotides and peptides. They also control the activity of enzymes, ion channels and the transport of vesicles.

To determine the relevance of *GPCR41* to lipid metabolism, the G protein-coupled receptor database system (GPCRDB, http://www.gpcr.org/7tm/, Horn *et al.* 1998) was searched to reveal the receptor families and ligand specificity of this gene. *GPCR41* could not be found in the database. Nor could it be assigned to a subfamily by the GPCR Subfamily Classifier (Karchin *et al.* 2002).

Since genes were not completely sequenced and screened for polymorphisms, additional polymorphisms in coding regions may exist. To exclude them as causal mutations, genotyping of individuals of the German Simmental granddaughter design at known SNPs will be the basis for linkage disequilibrium studies. To do this effectively, SBE was established for SNP genotyping.

Table 4.16: Polymorphisms in genes and STS markers neighboring *DGAT1*.

Locus	SNP		Region	Position		PCR primer		Allele		Effect
	snp_id[a]	lab name		bases in	accession no.	up	down	1	2	
BOP1	320	KC1	intron 5	131	AJ518948	2776	2779	C	T	
	321	KC2	exon 10	909	AJ518948	2776	2779	C	T	silent
	322	KC3	intron 11	1183	AJ518948	2780	2781	C	G	
	-	KC4	intron 14	1826-37	AJ518948	2780	2781	N0	N12	
	323	KC5	exon 15	1915	AJ518948	2780	2781	T	C	silent
CPSF1	268	CP1	intron 7	232	AJ518949	1994	1995	T	C	
	269	CP2	intron 8	429	AJ518949	1994	1995	A	G	
	270	CP3	exon 9	546	AJ518949	1994	1995	C	T	silent
	271	CP4	intron 9	619	AJ518949	1994	1995	A	G	
	274	CP5	exon 23	3596	AJ518949	2000	2001	G	A	silent
	277, 278	CP6	intron 26	4327-8	AJ518949	2004	2005	TG	C-	
	279	CP7	intron 26	4376	AJ518949	2004	2005	T	C	
	280	CP8	intron 26	4536	AJ518949	2004	2005	A	G	
	281	CP9	exon 11	928	AJ518949	1996	1997	C	G	Pro-Ala
	282	CP10	exon 13	1249	AJ518949	1996	1997	C	T	Thr-Ile
	283	CP11	intron 20	2906	AJ518949	2000	2001	C	T	
CYHR1	284	KA1	intron	764	AJ518952	2454	2457	A	G	
	285	KA2	intron	793	AJ518952	2454	2457	G	A	
	307	LA1	intron	128	AJ518967	2587	2588	T	G	
	308	LA2	3'UTR	795	AJ518967	2589	2590	A	G	
	309	LA3	3'UTR	1038	AJ518967	2589	2590	C	A	
FBXL6	324	FX1	intron 1	669	AJ518953	2669	1960	G	C	
FLJ11856	325	FA1	exon 3	834	AJ518954	2502	2586	G	A	Gly-Lys
GPT	288	GP1	exon	800	AJ518958	2442	2445	G	A	silent
	289	GP2	intron	1269	AJ518958	2442	2445	C	T	
HSF1	265	HS1	3'end	809	AJ518961	1756	1729	C	G	
	266	HS2	intron	310	AJ518959	1965	1966	G	A	
	267	HS3	intron	329	AJ518959	1965	1966	C	T	
KIAA1833	318	KB1	intron 14	593	AJ518965	2732	2733	G	C	
	319	KB2	exon 15	1592	AJ518965	2732	2733	G	T	silent
	317	KB3	intron 14	1760	AJ518965	2746	2747	C	T	
KIAA0014	331, 332	KD1	3'UTR	521-2	AJ518963	2479	2480	TG	del	
	333	KD2	3'UTR	1804	AJ518963	2599	2600	C	G	
	334	KD3	3'UTR	1957	AJ518963	2599	2600	C	T	
	-	KD4	3'UTR	2187-99	AJ518963	2599	2600	(AC)7	(AC)6	
MGC10520	306	MG1	3'UTR	448	AJ518968	2651	2652	A	G	
PTK2	276	PT2	intron 7	751	AJ519780	2874	2875	T	C	
RECQL4	286	RE1	exon 13	227	AJ518973	2430	2432	A	G	Met-Val
	287	RE2	exon 13	250	AJ518973	2430	2432	T	C	silent
VPS28	-	VP1	intron 6	471-82	AJ518974	2591	2592	Gn	C	
100P18-T7	336	BG1		279	AJ519351	2782	2766	G	C	
293G16-SP6	316	BC1		367	AJ519355	2486	2487	G	C	
	315	BC2		423	AJ519355	2486	2487	G	A	
	314	BC3		507	AJ519355	2486	2487	A	G	
	313	BC4		593	AJ519355	2486	2487	T	G	
360L24-T7	330	BE1		386	AJ519357	2737	2740	C	T	
	-	BE2		471	AJ519357	2737	2740	(C)6A(C)4	(C)9A(C)3	
3O1-SP6	335	BF1	repeat	40	AJ519358	2674	2671	del	C	
3O1-T7	326	BD1		731	AJ519359	2771	2783	G	T	
	327	BD2		778	AJ519359	2771	2783	C	T	
	328	BD3		806	AJ519359	2771	2783	C	G	
	329	BD4		1045	AJ519359	2771	2783	G	C	
428F15-SP6	310	BB1		451	AJ519362	2697	2506	T	G	
	311	BB2		453	AJ519362	2697	2506	G	T	
56F1-SP6	290	BA1		142	AJ519363	1686	2404	G	T	
	-	BA2		450-9	AJ519363	1686	2404	GATACAACT	del	

[a]SNP_id refers to SNPZoo entry; http://www.snpzoo.de/ (Fries *et al.* 2001).

4.2 SNP genotyping – optimization of single base extension (SBE) assay

Multiplex SBE assay results of four SNPs (individual heterozygous at all four SNPs) are shown in Figure 4.17 and in Table 4.17 subject to

- different concentrations of SAP in the PCR purification step,
- different concentrations of fluorescently labeled didesoxynucleotides (dye mix),
- different concentrations of thermosequenase,
- and the volume of SBE reaction applied to the gel.

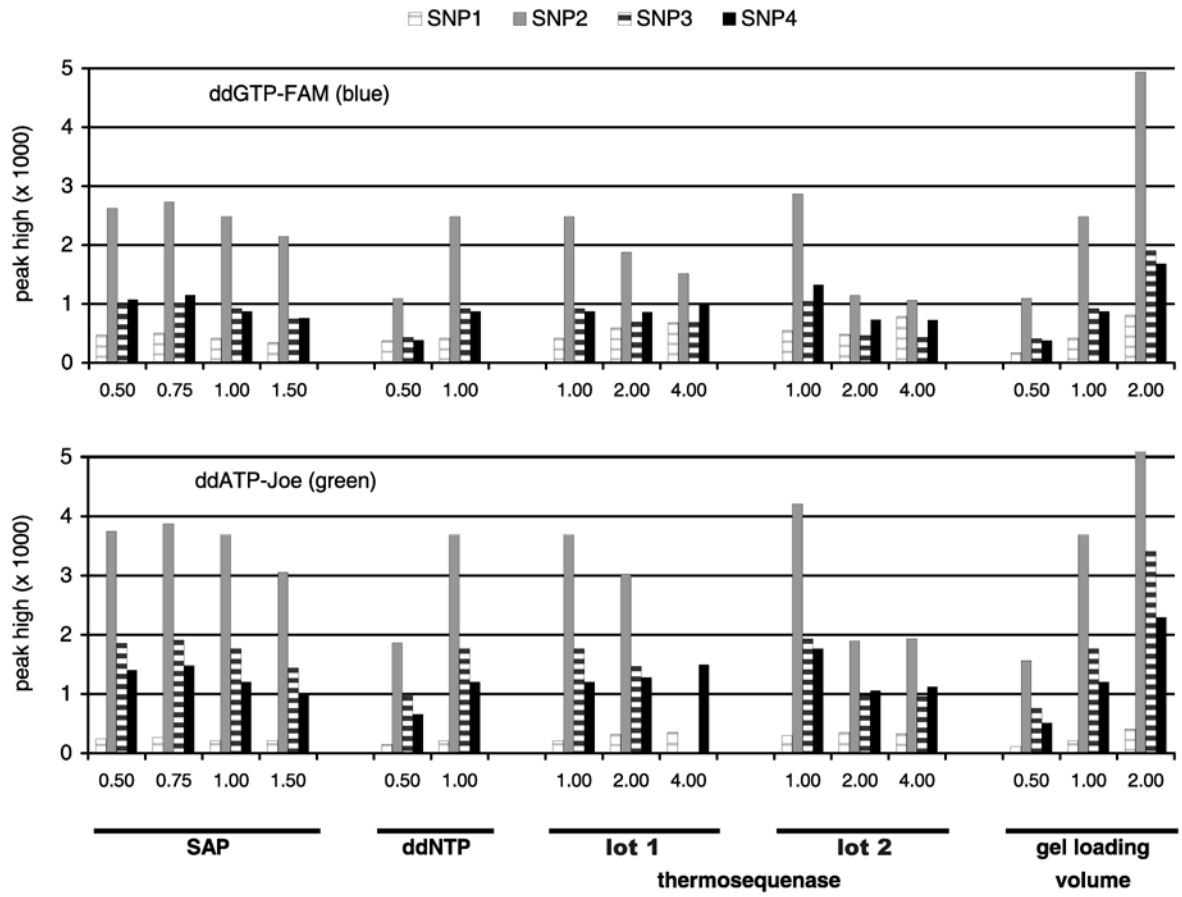


Figure 4.17: SBE signal intensity depending on the concentration of SAP, dye mix, thermosequenase and loading volume.

For experiment setup, see Table 3.8 in Materials and Methods, page 33. SNP id of SNP1 to SNP4 in plot were (SBE primer number in parentheses): 274 (2020), 285 (2411), 286 (2414) and 287 (2465).

Reducing the amount of SAP and using less SAP treated PCR product in the SBE reaction led to higher SBE signals, presumably by reducing the glycerol concentration, which may reduce the activity of thermosequenase. Applying more than 0.05 u/µL of thermosequenase led to an increase in signal intensity.

The effect of increasing or decreasing the concentration of particular SBE primer was minimal (Figure 4.18), despite recommended to compensate for consistently low or high signals of the respective primer.

SBE concentrations described in Materials and Methods were based on these optimization results.

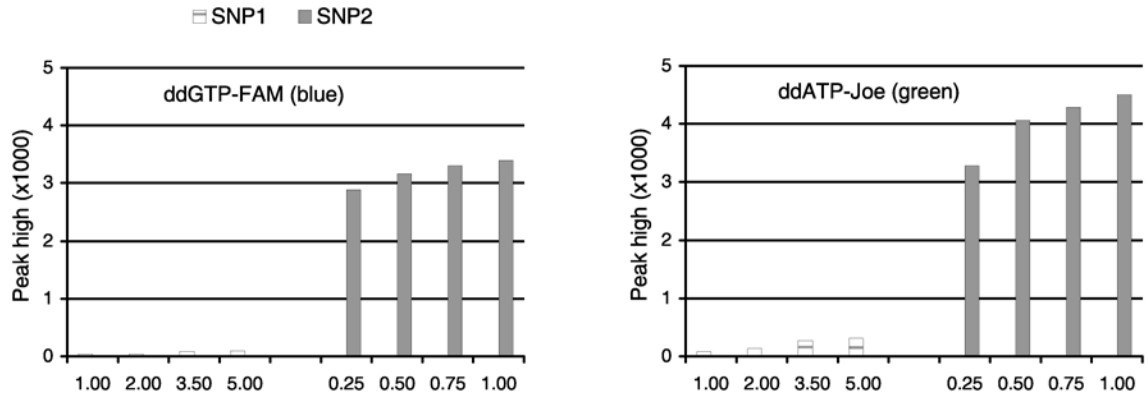


Figure 4.18: Influence of different SBE primer concentration on SBE signal intensity.
For experiment setup, see Table 3.8 in Materials and Methods, page 33. SNP id of SNP1 to SNP4 in plot were (SBE primer number in parentheses): 274 (2020) and 285 (2411).

Table 4.17: SBE optimization.

Reagent	Amount[a]	Allele 1 (ddGTP-FAM)				Allele 2 (ddATP-Joe)			
		SNP1	SNP2	SNP3	SNP4	SNP1	SNP2	SNP3	SNP4
SAP in the PCR purification reaction	0.50	473	2625	1006	1074	244	3754	1861	1408
	0.75	504	2732	1022	1153	266	3876	1915	1480
	1.00	**416**	**2485**	**923**	**873**	**215**	**3686**	**1769**	**1204**
	1.50	349	2147	751	763	216	3062	1444	1022
Fluorescent didesoxynucleotides	0.50	383	1092	437	384	149	1865	974	657
	1.00	**416**	**2485**	**923**	**873**	**215**	**3686**	**1769**	**1204**
Thermosequenase (charge 1)	**1.00**	**416**	**2485**	**923**	**873**	**215**	**3686**	**1769**	**1204**
	2.00	590	1879	698	862	318	3025	1472	1279
	4.00	684	1517	690	995	355	n.d.	n.d.	1500
Thermosequenase (charge 2)	**1.00**	**551**	**2870**	**1049**	**1325**	**304**	**4213**	**1936**	**1767**
	2.00	487	1146	469	731	352	1897	981	1057
	4.00	786	1067	437	730	332	1933	964	1126
Volume of SBE reaction applied to the gel	0.50	176	1097	415	383	112	1565	763	513
	1.00	**416**	**2485**	**923**	**873**	**215**	**3686**	**1769**	**1204**
	2.00	808	4938	1907	1686	412	5274	3412	2298
Concentration of SBE primer (SNP with low signal)	**1.00**	**43**				**84**			
	2.00	43				137			
	3.50	90				278			
	5.00	95				310			
Concentration of SBE primer (SNP with high signal)	0.25		2879				3275		
	0.50		3158				4057		
	0.75		3298				4281		
	1.00		**3389**				**4503**		

[a]See Table 3.8 in Materials and Methods, page 33. Bold values refer to the initial concentration (1.00).

4.3 DGAT1 in pig

One porcine EST sequence spanning exon 1 to exon 7 of *DGAT1* was found within the EST division of NCBI sequence database (accession number: BI340705). Three porcine BAC clones containing *DGAT1* were isolated from BAC library RPCI-44 (334O4, 370M1 and 494L16). FISH of BAC-DNA (clone 334O4) assigned *DGAT1* to porcine chromosome

4pter-p15. FISH mapping was done by Felix Habermann (Lehrstuhl für Tierzucht, Technische Universität München, Germany). Recently, the porcine *DGAT1* gene was published (accession number AY093657, Nonneman *et al.* 2002).

4.4 DGAT2 gene family in cattle and in pig

4.4.1 Cloning of the genes in cattle and in pig

In addition to the five known members of the human *DGAT2* gene family (Cases *et al.* 2001), one new member was identified in the NCBI sequence database, which was termed *DC6* (Table 4.18). BLAST of the NCBI sequence database revealed homologous sequences in mouse for all six members, in cattle for two members (*DGAT2* and *DC2*) and in pig for two members (*DGAT2* and *DC5*). Further, one new member (*DC7*) was identified in pig by three ESTs without a homologous sequence in human. Nomenclature of *DGAT2* gene family members was adapted from Cases *et al.* (2001).

Table 4.18: *DGAT2* gene family in human, mouse, cattle and swine.

Symbol[a]	Symbol NCBI	Name	Homo sapiens Accession no.[b]	*Mus musculus* Accession no.[b]	*Bos taurus* no. ESTs	*Sus scrofa* no. ESTs
DGAT2	*DGAT2*	diacylglycerol O-acyltransferase homolog 2	BC015234	AF384160	10	3
DC2[c]	*DGAT2L1*	diacylglycerol O-acyltransferase 2 like 1	AF384163	AF384162	2	-
DC3	*LOC158833*	similar to bA351K23.5 (novel protein)	XM_088691	XM_113091	-	-
DC4	*LOC158835*	similar to bA351K23.5 (novel protein)	XM_088683	XM_141969	-	-
DC5[d]	*FLJ22644*	hypothetical protein	AK026297	BF607517	-	1
DC6 (new)	*LOC170132*	similar to bA351K23.5 (novel protein)	XM_093119	XM_141971	-	-
DC7 (new)	-	not described in human	-	-	-	3

[a]Adapted from Cases *et al.* (2001)
[b]GenBank or NCBI RefSeq record
[c]DC2 = MGAT1 (Yen *et al.* 2002)
[d]DC5 = MGAT2 (Cao *et al.* 2003; Yen *et al.* 2003)

A cladogram of the DGAT2 gene family was generated using human, mouse, cattle and pig cDNA sequences and two *Mortierella ramanniana DGAT2* sequences as an outgroup (Figure 4.19). The tree showed the typical structure for gene families: orthologous genes in different species had higher identity with one another than did the paralogous genes in any single species. The newly identified porcine *DC7* was in the *DC5* cluster (identity of 61.6%). Identities ranged between 80.2 and 90.2% for the cattle and pig *DGAT2* gene family members compared to the human orthologous genes (Table 4.19). Cattle and pig *DGAT2* had an identity of 91.8% to one another. Identities were calculated using the software program "gap" of the GCG package (Genetics Computer Group 2001). For alignments of human, cattle and pig *DGAT2* gene family members see appendix 9.7.

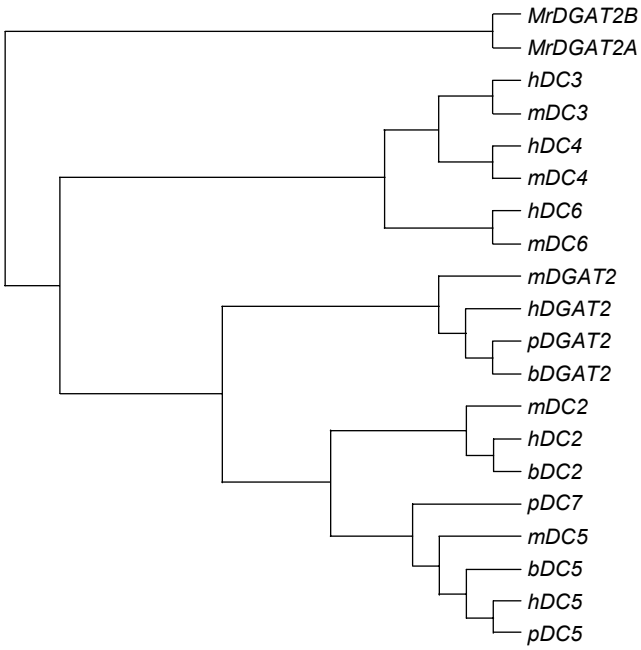


Figure 4.19: Cladogram of *DGAT2* gene family in four mammals.
Species were *Homo sapiens* (*h*), *Bos taurus* (*b*), *Mus musculus* (*m*) and *Sus scrofa* (*p*). *DGAT2* sequences of *Mortierella ramanniana* (Mr) were used as an outgroup (accession number: AF391089 and AF391090). For accession numbers for human, mouse and cattle see Table 4.18 and Table 4.24. Partial cDNA sequences for the pig were derived from EST sequences (Table 4.21). The tree was obtained using a neighbor-joining distance analysis in ClustalX version 1.81.

Table 4.19: Coding sequence comparison of mouse, cattle and pig *DGAT2* gene family members with their respective human orthologues.

	Identity [%]		
	Mus musculus	*Bos taurus*	*Sus scrofa*
DGAT2	89.6	90.2	85.7
DC2	77.0	84.1	-
DC3	83.8	-	-
DC4	86.0	-	-
DC5	83.0	80.2	84.1
DC6	73.8	-	-

EST sequences for cattle and pig

EST sequences were found for bovine *DGAT2* and *DC2* (Table 4.20) and for porcine *DGAT2*, *DC5* and DC7 (Table 4.21). The 10 bovine EST sequences for *DGAT2* represent the complete mRNA sequence from exon 1 to exon 8 and the three identified porcine EST sequences span exon 2 to exon 6. The consensus sequence of bovine ESTs for *DC2* covers exon 1 to exon 3 and exon 6, lacking exon 4 and 5. The EST with the accession number BE754760 contains 236 bp of intron 6. For *DC5*, an EST sequence was only available for the pig. However, PCR primers derived from porcine *DC5* sequence were used to screen the bovine BAC library.

Table 4.20: Bovine EST sequences for *DGAT2* and *DC2*.

Locus	GenBank Accession no.	Size [bp]	Source[a]	Position[b]		Discrepancy to consensus sequence + inclusion, - deletion
DGAT2	BE724193	335	1	901-1235	exon 7 - 3'UTR	
	BI536057	569	1	667-1235	exon 6 - 3'UTR	
	AW326247	432	2	103-534	exon 2 - exon 5	
	BI681948	567	2	679-1230	exon 6 - 3'UTR	
	BE482224	502	5	136-637	exon 2 - exon 6	
	BE479873	508	5	454-961	exon 6 - 3'UTR	
	BF868335	679	5	-198-481	exon 1 - exon 5	
	BG694175	343	5	425-767	exon 5 - exon 7	
	BG687855	291	5	380-669	exon 5 - exon 6	
	BF430191	421	6	487-907	exon 5 - exon 7	
DC2	AW429404	376	1	6-377	exon 1 - exon 3	
	BE754760	501	3	851-1115	exon 6 - 3'UTR	+ 236 bp of intron 5

[a]Source:
1: pooled tissue from day 20 and day 40 embryos
2: pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary gland
3: pooled tissue from testis, thymus, semitendonosus muscle, longissimus muscle, pancreas, adrenal gland , and endometrium
4: adipose tissue
5: mammary tissues at eight physiological, developmental, and disease states
6: library obtained from Stratagene, catalog #937721. Library made from skeletal muscle of a two year old Holstein cow.
[b]Base 1 = first base of start codon

Table 4.21: Porcine EST sequences for *DGAT2*, *DC5* and *DC7*.

Locus	GenBank Accession no.	Size [bp]	Source[a]	Position[b]		Discrepancy to consensus sequence + inclusion, - deletion
DGAT1	BI340705	422	1	180-602	exon 1 - exon 7	
DGAT2	BF189320	543	1	42-502	exon 2 - exon 5	
	BE232328	177	2	433-610	exon 5 - exon 6	
	BE014044	540	2	81-620	exon 2 - exon 6	
DC5	BE030672	540	2	11-550	exon 1 - exon 4	
DC7	BI345601	570	1			
	BE032482	520	2			
	BE031168	549	2			+ 112 bp and + 147 bp of introns

[a]Source:
1: library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary gland, and placenta.
2: library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos.
[b]Base 1 = first base of start codon

Isolating BAC clones

BAC clones were isolated for cattle containing *DGAT2*, *DC2* and *DC5*, and for pig containing *DGAT2*, *DC5* and *DC7* (Table 4.22).

Physical gene mapping

FISH mapping (Table 4.23) of bovine *DGAT2* and *DC5* (clone 5L16), *DC2* (clone 307A24), and porcine *DGAT1* (clone 334O4), *DGAT2* and *DC5* (clone 156H14) and *DC7* (clone 376D18) was done by Felix Habermann (Lehrstuhl für Tierzucht, Technische Universität München, Germany).

Table 4.22: Bovine and porcine BAC clones containing *DGAT1* and *DGAT2* genes.

Bovine BAC library **RPC-42**			
Probe	Clone[a]	Positive colony PCR	Internal lab no.
DGAT2	107N13	*DGAT2*	844
	269A17	*DGAT2, DC5*	846
	288A16	*DGAT2*	845
	5L16	***DGAT2***	**841**
	74F6	*DGAT2*	843
	80P6	*DGAT2*	842
DC2	307A24	*DC2*	847
	350D19	*DC2*	849
	362M12	***DC2***	**848**
	470C12	*DC2*	850
	484K21	*DC2*	851
	503L4	*DC2*	852
DC5	143N13	*DC5*	857
	167J22	*DC5*	859
	187D16	*DC5*	858
	20B12	***DC5***	**854**

Porcine BAC library **RPCI-44**			
Probe	Clone	Positive colony PCR	Internal lab no.
DGAT2	139G11	*DGAT2*	832
	148P24	*DGAT2*	828
	156H14	*DGAT2, DC5*	829
	185A13	*DGAT2*	830
	226N3	*DGAT2, DC5*	831
	41K3	*DGAT2*	824
	60A22	*DGAT2*	827
	86A12	*DGAT2*	826
DC5	218D11	*DC5*	834
DC7	376D18	*DC7*	839
	516N20	*DC7*	840

[a]Clones in bold were used for direct sequencing of BAC DNA.

Table 4.23: Chromosomal positions of DGAT1 and *DGAT2* gene family members.

Locus	Homo sapiens[a]	Mus musculus[a]	Bos taurus[b]	Sus scrofa[b]
DGAT2	11q13.3	2	15q23-25	9pter-p23
DC2	2q36.2	1	2q42-44	-
DC3	Xq12	-	-	-
DC4	Xq12	X	-	-
DC5	11q13.3	7	15q23-25	9pter-p23
DC6 (new)	Xq12	X	-	-
DC7 (new)	-	5	-	3pter-p15

[a]As noted in NCBI.
[b]Determined by FISH by Felix Habermann (Lehrstuhl für Tierzucht, Technische Universität München, Germany).

4.4.2 Characterization of bovine DGAT2, DC2 and DC5

Sequence

For *DGAT2*, the complete coding sequence was available in form of EST sequences. Smaller introns (intron 3, 5 and 6) were sequenced completely after PCR amplification, whereas larger introns (intron 1, 2, 4 and 7) and the *DGAT2* flanking regions were partially sequenced by primer walking. The size of the larger introns was determined by long range PCR (Figure 4.20). Bovine *DGAT2* encodes a 361 residue protein (human DGAT2: 388 residues).

The complete coding sequence of bovine *DC2* was obtained as well as the smaller introns (intron 2 and 4). The larger introns (intron 1, 3 and 5), as well as the *DC2* flanking regions were partially sequenced by primer walking. The size of the large introns was determined by long range PCR (Figure 4.20). Bovine *DC2* encodes a 334 residue protein (human DC2: 334 residues).

For bovine *DC5*, the complete coding sequence was determined for the truncated splice variant ($DC5^{trunc}$: exon 1 to exon 5 including intron 4). Exon 6 and the larger introns (introns 1, 2

and 5) were not sequenced. The sizes of introns 1 and 2 could not be determined by long range PCR. Bovine *DC5*^{trunc} encodes a 226 residue protein (human DC5: 284 residues). Sequence information for bovine *DGAT2*, *DC2* and *DC5*^{trunc} have been deposited in EMBL (Table 4.24).

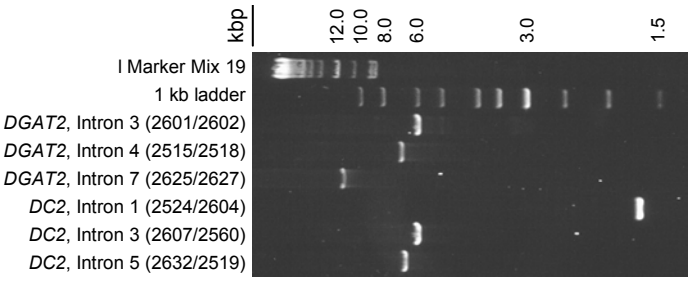


Figure 4.20: Long range PCR across larger introns in *DGAT2* and *DC2* in cattle. (0.7% agarose in TAE)

Table 4.24: Entries for *DGAT2* genes in the EMBL nucleotide sequence database.

Locus	Accession no.	Description
DGAT2	AJ519787	*Bos taurus* mRNA for putative diacylglycerol O-acyltransferase (DGAT2 gene)
	AJ534368	*Bos taurus* DGAT2 gene for diacylglycerol O-acyltransferase 2, exon 1 and joined CDS
	AJ534369	*Bos taurus* DGAT2 gene for diacylglycerol O-acyltransferase 2, exon 2
	AJ534370	*Bos taurus* DGAT2 gene for diacylglycerol O-acyltransferase 2, exons 3 and 4
	AJ534371	*Bos taurus* DGAT2 gene for diacylglycerol O-acyltransferase 2, exons 5 to 7
	AJ534372	*Bos taurus* DGAT2 gene for diacylglycerol O-acyltransferase 2, exon 8
DC2	AJ519785	*Bos taurus* mRNA for putative diacylglycerol O-acyltransferase 2 (DC2 gene)
	AJ534373	*Bos taurus* DC2 gene for putative diacylglycerol O-acyltransferase 2, exon 1 and joined CDS
	AJ534374	*Bos taurus* DC2 gene for putative diacylglycerol O-acyltransferase 2, exons 2 and 3
	AJ534375	*Bos taurus* DC2 gene for putative diacylglycerol O-acyltransferase 2, exons 4 and 5
	AJ534376	*Bos taurus* DC2 gene for putative diacylglycerol O-acyltransferase 2, exon 6
DC5	AJ519786	*Bos taurus* mRNA for putative diacylglycerol O-acyltransferase 2 (DC5 gene)
	AJ534377	*Bos taurus* DC5 gene for putative diacylglycerol O-acyltransferase 2, exon 1 and joined CDS
	AJ534378	*Bos taurus* DC5 gene for putative diacylglycerol O-acyltransferase 2, exon 2
	AJ534379	*Bos taurus* DC5 gene for putative diacylglycerol O-acyltransferase 2, exons 3 and 4

Structure

The structure of bovine *DGAT2*, *DC2* and *DC5* (Figure 4.21 and Table 4.25) was highly conserved with respect to their respective human genes. The exon sizes in cattle were identical to those in human except for two differences: bovine exon 1 in *DGAT2* was 81 bp or 27 residues shorter (indicated by the light gray box in Figure 4.21) and bovine exon 4 in *DC5*^{trunc} was 174 bp or 58 residues shorter. The sizes of the bovine introns were only relatively conserved compared to human (Table 4.25). The exon structure between bovine *DGAT2*, *DC2* and *DC5* was partly conserved (Figure 4.21). All exon/intron splice sites of *DGAT2*, *DC2* and *DC5* were in agreement with the GT-AG rule (Breathnach *et al.* 1978).

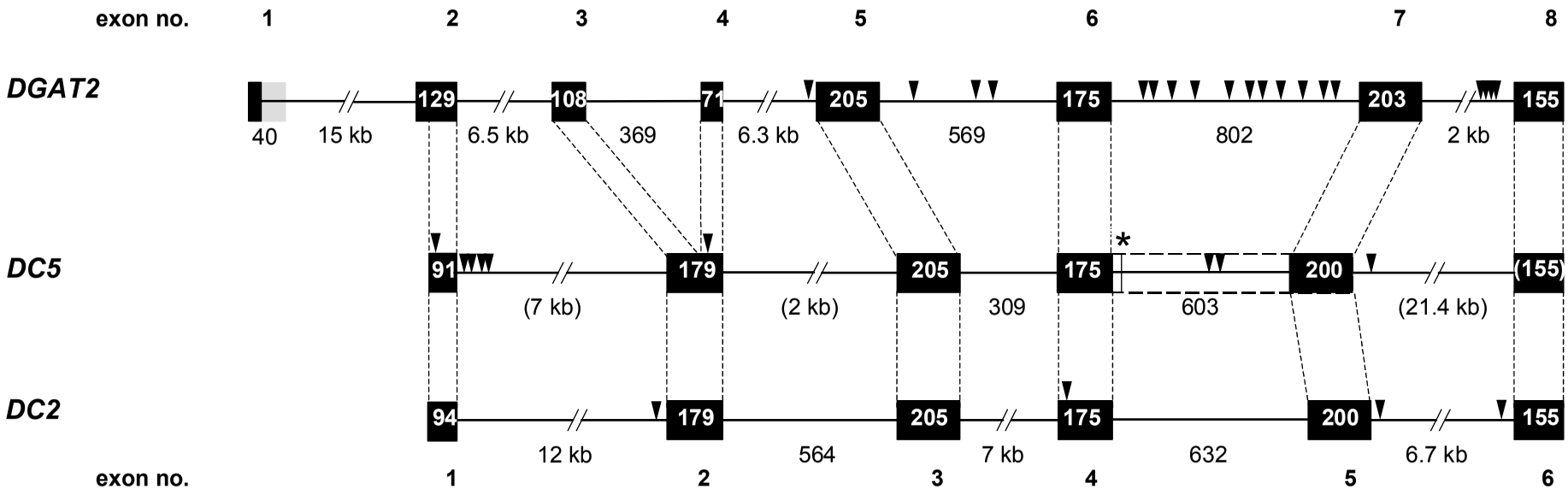


Figure 4.21: Exon/intron structure of bovine *DGAT2* gene family members.

Black boxes represent exons with the sizes in bp within the boxes. The sizes of the introns, in bp (or in kb if indicated), are placed below the horizontal lines, which represent the introns. Sizes in brackets were adopted from human. Triangles represent polymorphic positions. Vertical lines indicate conserved exon structure. Exon 1 of bovine *DGAT2* is shortened (gray box) compared to human. For *DC5* exist a splice variant in human containing intron 4 (hatched lines) including a stop codon (asterisk).

Table 4.25: Exon/intron organization of bovine *DGAT2*, *DC2* and *DC5*.

	Exon					Intron		
	No.	Position in cDNA[a]	Size [bp]	3'-splice acceptor[b]	5'-splice donor[b]	No	Size [bp]	Size [bp] (human)
DGAT2	5'UTR		198					
	1	1-40	40		..... TGCGAG**gt**gagc	1	15 k[c]	15.5 k
	2	41-169	129	tcac**ag**GCACTG	..... TGCTGG**gt**aagc	2	6.5 k[c]	5.4 k
	3	170-277	108	cccc**ag**GAGTGG	..... AGAAAG**gt**agga	3	369	365
	4	278-348	71	tgga**ag**GTGGCA	..... ATTCAG**gt**aaaa	4	6.3 k[c]	5.6 k
	5	349-553	205	ctgc**ag**CTGGTG	..... CTGGAG**gt**gaga	5	569	625
	6	554-728	175	cccc**ag**GCATCT	..... CCATGG**gt**gagt	6	802	895
	7	729-931	203	ttcc**ag**AGCCGA	..... CTGTCG**gt**aagc	7	2 k[c]	1.9 k
	8	932-1086	155	ctgc**ag**TGGGCG	.....			
	3'UTR	1087-1235	149					
DC2	5'UTR		?					
	1	1-94	94		..... TGTTCG**gt**aagg	1	12 k[c]	16.5 k
	2	95-273	179	ttac**ag**CACAGG	..... ATTCAT**gt**gagt	2	564	742
	3	274-478	205	ctgt**ag**CTCATC	..... GCAGTG**gt**aagt	3	7 k[c]	4.9 k
	4	479-653	175	acta**ag**GGCCAG	..... CCATGG**gt**aagg	4	632	552
	5	654-853	200	tgcc**ag**TGCTTA	..... CTGTTG**gt**atgt	5	6.7 k[c]	14.5 k
	6	854-1008	155	ttgc**ag**TTGGCC	.....			
	3'UTR	1006-1115	100					
DC5	5'UTR		68					
	1	1-91	91		..... TCCTGG**gt**aaga	1	[d]	2.0 k
	2	92-270	179	cccc**ag**GCGTTG	..... ATCTCA**gt**gaat	2	[d]	7.3 k
	3	271-475	205	ctgc**ag**CTGGTC	..... CAGGGG**gt**gagt	3	309	330
	4	476-650	175	cccc**ag**GGCTGG	..... GCATGG**gt**attg	4	603	645
	5	651-850	200	ctct**ag**GGCAGC	..... CCGTGG**gt**gagc	5	?	21.4 k
	6	not sequenced	?					
	3'UTR	not sequenced	?					
DC5^{trunc}	4	476-681	206	cccc**ag**GGCTGG	.....			

[a]Base 1 = first base of start codon
[b]Exon sequences are indicated in uppercase letters, intron sequences in lowercase letters. The consensus splice site sequences are in boldface.
[c]Determined by LR-PCR.
[d]Could not be determined by PCR

4.4.3 Polymorphisms in bovine DGAT2, DC2 and DC5

Re-sequencing in a panel including 12 individuals and six DNA pools (see appendix 9.3) revealed 23 polymorphic positions in *DGAT2*, four polymorphic positions in *DC2* and 18 polymorphic positions in *DC5* (Table 4.26). Most were in non-coding regions. Only SNP 347 in exon 4 of *DC2* represented a non-conservative substitution of a cysteine by a lysine.

Table 4.26: Polymorphisms in bovine *DGAT2*, *DC2* and *DC5*.

Locus	SNP		Region	Accession	Position	Allele		Effect
	snp_id[a]	lab name				1	2	
DGAT2	338	DH16	intron 4	AJ534371	433	A	G	
	291	DH1	intron 5	AJ534371	755	A	G	
	-	DH2	intron 5	AJ534371	959-66	N8	T	[N8 = CCCTGGCA]
	292	DH3	intron 5	AJ534371	1004	A	G	
	293	DH4	intron 6	AJ534371	1501	G	T	
	294	DH5	intron 6	AJ534371	1514	T	C	
	295	DH6	intron 6	AJ534371	1541	C	G	
	296	DH7	intron 6	AJ534371	1578	C	T	
	297	DH8	intron 6	AJ534371	1614	A	G	
	298	DH9	intron 6	AJ534371	1637	A	G	
	299	DH10	intron 6	AJ534371	1694	A	C	
	300	DH11	intron 6	AJ534371	1740	C	T	
	301	DH12	intron 6	AJ534371	1766	A	del	
	302	DH13	intron 6	AJ534371	1927	G	A	
	303	DH14	intron 6	AJ534371	2012	T	G	
	304	DH15	intron 6	AJ534371	2065	T	C	
	339	DH17	intron 7	AJ534372	349	A	G	
	340	DH18	intron 7	AJ534372	357	A	G	
	341	DH19	intron 7	AJ534372	396	A	G	
	342	DH20	intron 7	AJ534372	448	A	G	
	343	DH21	intron 7	AJ534372	481	A	G	
	344	DH22	intron 7	AJ534372	668	C	G	
	345	DH23	3'UTR	AJ534372	975	A	G	
DC2	346	DI1	intron 1	AJ534374	154	G	del	
	347	DI2	exon 4	AJ534374	426	G	C	Cys-Ser
	348	DI3	intron 5	AJ534374	1485	C	T	
	349	DI4	intron 5	AJ534376	502	C	T	
DC5	350	DJ1	5'end	AJ534377	109	G	A	
	351	DJ2	5'end	AJ534377	156	G	C	
	352	DJ3	5'end	AJ534377	361	G	A	
	353	DJ4	5'end	AJ534377	421	A	C	
	354	DJ5	5'end	AJ534377	463	C	T	
	355	DJ6	5'end	AJ534377	467	G	C	
	356	DJ7	5'end	AJ534377	513	G	C	
	357	DJ8	5'UTR	AJ534377	578	G	A	
	358	DJ9	exon 1	AJ534377	618	G	A	silent
	359	DJ10	intron 1	AJ534377	724	C	T	
	360	DJ11	intron 1	AJ534377	727	C	A	
	361	DJ12	intron 1	AJ534377	757	C	T	
	362	DJ13	intron 1	AJ534377	760	T	C	
	363	DJ14	exon 2	AJ534378	290	G	A	silent
	365	DJ16	intron 4	AJ534379	1188	A	G	
	366	DJ17	intron 4	AJ534379	1212	C	A	
	367	DJ18	intron 5	AJ534379	1725	T	C	

[a]SNP_id refers to SNPZoo entry; http://www.snpzoo.de/ (Fries *et al.* 2001).
[b]Pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicate high and low breeding values for milk fat percentage, respectively.

Allele frequencies were estimated using pooled DNA samples (Table 4.27) and plotted for *DGAT2* (Figure 4.22A), *DC2* (Figure 4.23) and *DC5* (Figure 4.24). Some of the differences between pools with high and low breeding values for milk fat percentage were significant at the nominal significance level of $\alpha = 0.05$ (Table 4.27). However, when a Bonferroni correction for multiple comparisons was applied, none of the differences of allele frequency were significant at a global significance level of $\alpha = 0.05$. To confirm results for pooled values, individuals were genotyped for the arbitrarily chosen SNP 303 of *DGAT2*. RFLP genotyping of each individual in the pools for SNP 303 likewise did not show significant differences in

allele frequency between the pools (Figure 4.22 B). SNP 303 is tightly linked with SNP 296, which shows significant differences in estimated allele frequencies between German Holstein DNA pools ($\alpha = 0.05$). SNP 303 was one of 12 SNPs (SNP 293 to 304) within a 634 bp PCR fragment. Genotyping of the 12 SNPs by direct sequencing in 37 individuals indicated the presence of only two haplotypes (for genotypes see appendix 9.8). The individuals belonged to the breeds German Holstein (4), German Simmental (16), German Brown (6), Jersey (7), Anatolian Black (2) and Yak (2). Observed differences in allele frequency may arise from errors caused by pooling, by unequal PCR amplification of the two alleles, or by inadequate frequency estimation based on sequence traces. For example, unequal amplification of the two alleles was observed for the polymorphism responsible for the $Lys^{232} \rightarrow$ Ala substitution (See chapter 3.6.3 page 24).

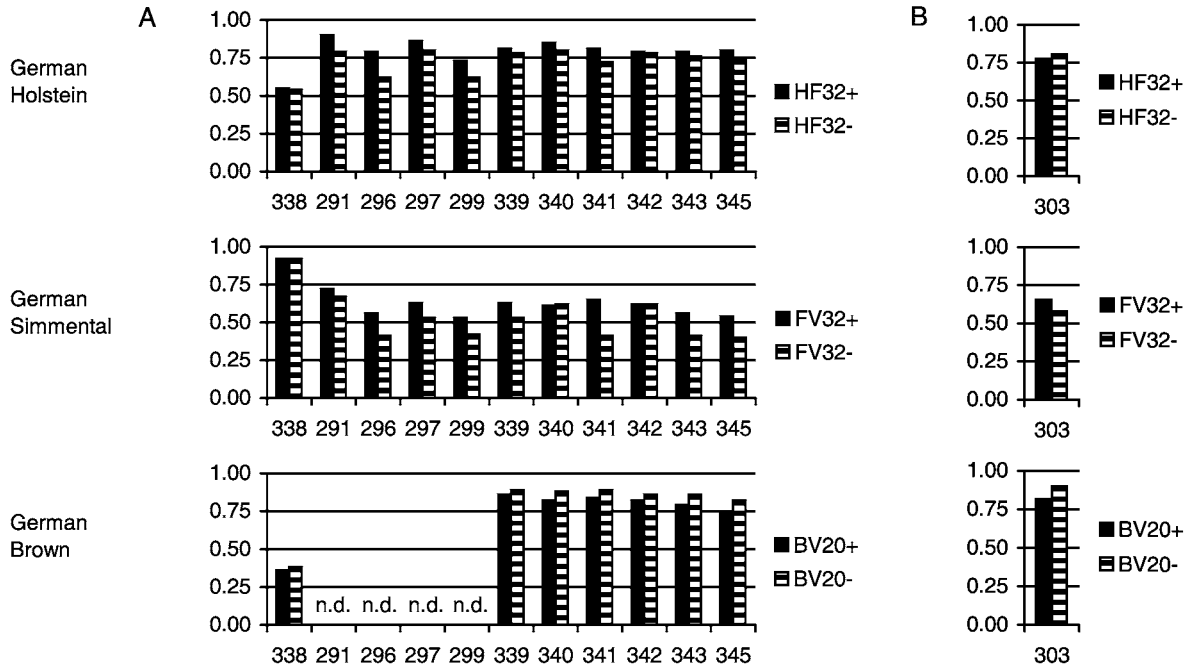


Figure 4.22: Allele frequencies of *DGAT2* polymorphisms in pooled DNA samples.
A: Allele frequencies estimated based on sequence traces of pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicated high and low breeding values for milk fat percentage, respectively; n.d., not determined; **B**: Observed allele frequencies of SNP 303 obtained by RFLP genotyping each individual in the pools.

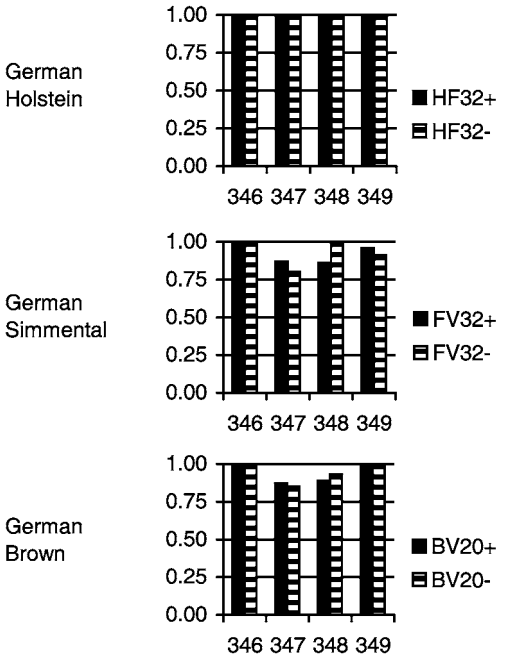


Figure 4.23: Allele frequencies of *DC2* polymorphisms in pooled DNA samples.
Allele frequencies estimated based on sequence traces of pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicate high and low breeding values for milk fat percentage, respectively.

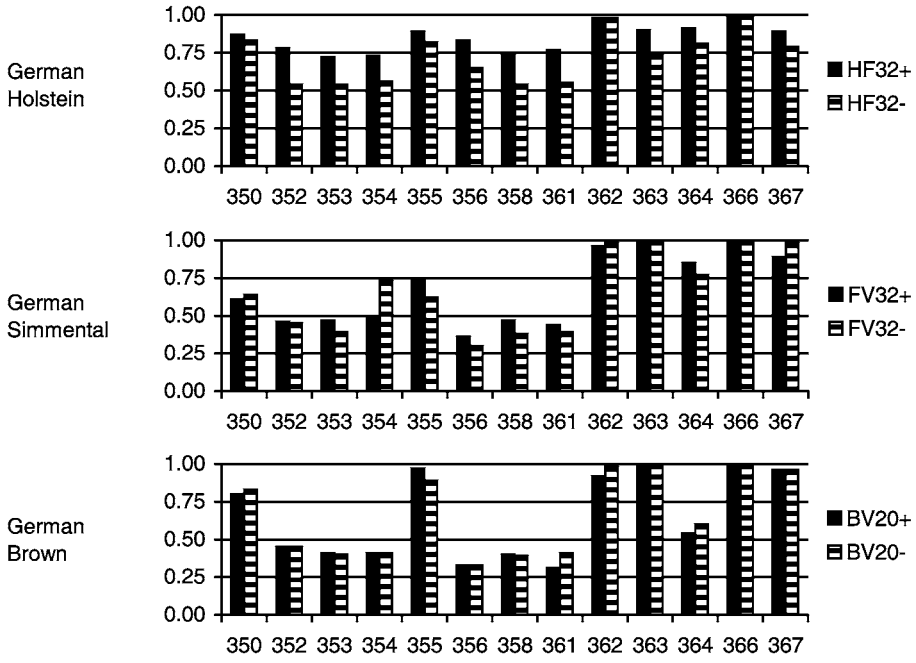


Figure 4.24: Allele frequencies of *DC5* polymorphisms in pooled DNA samples.
Allele frequencies estimated based on sequence traces of pooled DNA samples of the breeds German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-). Number of individuals per pool is indicated by the name, + and – indicate high and low breeding values for milk fat percentage, respectively.

Table 4.27: Allele frequencies of SNPs in pooled DNA samples.

SNP		Allele	German Holstein Frequency HF32+	HF32-	G[a]	α-value	German Simmental Frequency FV32+	FV32-	G[a]	α-value	German Brown Frequency BV20+	BV20-	G[a]	α-value
	DGAT2													
338	intron 4	A	0.55	0.54	0.01	no	0.92	0.92	0.00	no	0.36	0.38	0.03	no
291	intron 5	A	0.90	0.79	2.96	no	0.72	0.67	0.38	no				
296	intron 6	C	0.79	0.62	4.45	0.05	0.56	0.41	2.88	no				
297	intron 6	A	0.86	0.80	0.82	no	0.63	0.53	1.31	no				
298	intron 6	A	0.84	0.80	0.35	no	0.62	0.50	1.87	no				
299	intron 6	A	0.73	0.62	1.77	no	0.53	0.42	1.55	no				
339	intron 7	A	0.81	0.78	0.18	no	0.63	0.53	1.31	no	0.86	0.89	0.16	no
340	intron 7	A	0.85	0.80	0.55	no	0.61	0.62	0.01	no	0.82	0.88	0.56	no
341	intron 7	A	0.81	0.72	1.44	no	0.65	0.41	7.40	0.01	0.84	0.89	0.43	no
342	intron 7	A	0.79	0.78	0.02	no	0.62	0.62	0.00	no	0.82	0.86	0.24	no
343	intron 7	A	0.79	0.76	0.17	no	0.56	0.41	2.88	no	0.79	0.86	0.68	no
345	3'UTR	A	0.80	0.74	0.65	no	0.54	0.40	2.52	no	0.75	0.82	0.58	no
	DC2													
346	intron 1	G	1.00	1.00			1.00	1.00			1.00	1.00		
347	exon 4	G	1.00	1.00			0.87	0.80	1.14	no	0.87	0.85	0.07	no
348	intron 5	C	1.00	1.00			0.86	1.00	9.63	0.01	0.89	0.93	0.39	no
349	intron 5	C	1.00	1.00			0.96	0.91	1.32	no	1.00	1.00		
	DC5													
350	5'end	G	0.87	0.83	0.40	no	0.61	0.64	0.12	no	0.80	0.83	0.12	no
352	5'end	G	0.78	0.54	8.21	0.01	0.46	0.45	0.01	no	0.45	0.45	0.00	no
353	5'end	A	0.72	0.54	4.45	0.05	0.47	0.39	0.84	no	0.41	0.40	0.01	no
354	5'end	C	0.73	0.56	4.04	0.05	0.50	0.73	7.15	0.01	0.41	0.41	0.00	no
355	5'end	G	0.89	0.82	1.26	no	0.73	0.62	1.77	no	0.97	0.89	1.97	no
356	5'end	G	0.83	0.65	5.39	0.05	0.36	0.30	0.52	no	0.33	0.33	0.00	no
358	exon 1	G	0.74	0.54	5.56	0.05	0.47	0.38	1.06	no	0.40	0.39	0.01	no
361	intron 1	C	0.77	0.55	6.90	0.01	0.44	0.39	0.33	no	0.31	0.41	0.87	no
362	intron 1	T	0.98	0.98	0.00	no	0.96	0.98	0.44	no	0.92	0.98	1.52	no
363	exon 2	G	0.90	0.74	5.55	0.05	1.00	1.00			1.00	1.00		
366	intron 4	C	1.00	1.00			1.00	1.00			1.00	1.00		
367	intron 5	T	0.89	0.79	2.38	no	0.89	0.98	4.26	0.05	0.96	0.96	0	no

[a]When testing the null hypothesis H_0 (no association between SNP alleles and the breeding value for milk fat percentage) the test statistic G follows a χ^2-distribution with one degree of freedom.

5 Discussion

5.1 DGAT1

5.1.1 Bovine DAGT1 and association with milk fat percentage breeding value

It has been shown in this dissertation that a missense mutation ($Lys^{232} \rightarrow Ala$) in the bovine *DGAT1* gene is associated with variation in the breeding value for milk fat percentage. *DGAT1* encodes diacylglycerol O-acyltransferase (EC 2.3.1.20), a microsomal enzyme that catalyzes the final, and presumably rate-limiting, step of triglyceride synthesis (Mayorek *et al.* 1989). In parallel with Grisart and colleagues (2002a), our research represents the first successful positional cloning of a quantitative trait locus (QTL) in a species other than a model organism or a plant (see Table 1 in Glazier *et al.* 2002). Glazier *et al*. (2002) propose four criteria for the successful identification of the causal mutation of a QTL. Three criteria have been fulfilled: linkage and association, fine-mapping and sequence analysis. The last, but most difficult criteria, that of functional evidence, has not been provided yet. However, several lines of evidence support the proposition that $Lys^{232} \rightarrow Ala$ in DGAT1 is causal for variation of milk fat percentage:

- *DGAT1* is a functional candidate gene due to its role in fat metabolism (Cases *et al.* 1998) and evidence from mouse knockout studies indicate that *DGAT1* is crucial for lactation (Smith *et al.* 2000).
- Mapping in cattle placed *DGAT1* proximal to the microsatellite marker *ILST039* on chromosome 14, and therefore within a QTL interval of 3 cM for milk fat percentage (Farnir *et al.* 2002). The QTL for milk fat percentage was identified in two cattle breeds: in Holstein (Coppieters *et al.* 1998; Zhang *et al.* 1998a) and in German Simmental (Winter *et al.* 2002).
- Two animals of the German Simmental granddaughter-design (Winter *et al.* 2002) that were genotyped heterozygous (Qq) at the QTL based on marker-assisted QTL-genotyping were heterozygous for the $Lys^{232} \rightarrow Ala$ substitution, whereas 14 animals that are most likely qq at the QTL were homozygous for the alanine-encoding allele.
- The lysine variant is associated with high milk fat percentage and the alanine variant with low milk fat percentage in different breeds ($\alpha = 0.001$ in German Holstein and German Simmental, $\alpha = 0.05$ in German Brown). However, not all individuals in the German Simmental breed with high breeding value carry the $lysine^{232}$ allele. This is also true for German Brown bulls, where only two out of 20 carry the $lysine^{232}$ allele. Both breeds have a noticeably lower $lysine^{232}$ allele frequency than the German Holstein breed. This observation is compatible with *DGAT1* belonging to a "polygene consortium" influencing milk fat percentage (polygene model of quantitative genetics).

- It is readily plausible that the identified mutation in *DGAT1* has an effect on DGAT1 activity by altering the peptide sequence. In contrast, the consequence of a noncoding sequence variant would be more complicated to interpret, as the relationship between promoter sequences, gene expression and trait phenotype is less well understood (Glazier *et al.* 2002). Moreover, it has been speculated that complex traits (like milk fat percentage) result more often from noncoding regulatory variants than from coding sequence variants (Mackay 2001b). This may not be the case here.
- Lysine^{232} is located within a region of the DGAT1 peptide sequence that is less conserved across species (mammals, fly, nematode and plants; Figure 4.3). Therefore, it appears that changes in DGAT1 activity do not result from changes in an essential motif. However, lysine^{232} is conserved among mammals (primates including human, cattle, pig and rodents; Figure 4.4), indicating a possible functional importance of a positively charged, hydrophilic residue at that position. Alanine has an uncharged hydrophobic residue.

One putative bovine *DGAT1* transcript (EST accession number AW446985) lacks 66 bp of exon 8 including position of Lys^{232} → Ala. The alternative transcript (without Lys^{232} → Ala) and the DGAT1 alanine^{232} variant had similar DGAT activities in insect cells (Grisart *et al.* 2002b). The activities of both variants were between that of insect cells without bovine *DGAT1* and the lysine^{232} variant. The results indicate that lysine^{232} effects higher activity, but is not crucial for enzyme function in general.
Four further bovine EST sequences (see Table 4.1 page 36) out of 12 identified ESTs would imply the existence of alternative transcripts. Such transcripts could encode for proteins with alternative sequences, but may represent artifacts due to EST sequence generation, or may represent unspliced transcripts. The proportion of unspliced transcripts depends on the transcription rate of a gene (Wolfsberg *et al.* 1997). Alternative transcripts for *DGAT1* have been also reported in human (Cheng *et al.* 2001) and in olive tissue (Giannoulia *et al.* 2000).

The lysine^{232} alanine substitution probably took place early in the history of domesticated cattle or even before domestication as surmised by the existence of the alanine^{232} variant in the Anatolian Black breed (Table 4.7), which is indigenous to a region known as the site of domestication of the European *Bos taurus* (Medjugorac *et al.* 1994). Genotyping of 35 cattle breeds for Lys^{232} → Ala revealed the alanine^{232} variant to be present in other old breeds, for example in Chianina and N'Dama (Kaupe *et al.* in press).

All other identified sequence variants in *DGAT1* were located in noncoding sequences. One of them, the variable number of tandem repeat (VNTR) upstream of *DGAT1* may also affect the expression of *DGAT1*. Although, different VNTR alleles were obviously not associated with the breeding value for milk fat percentage, the effect of different numbers of repeat units on the expression of *DGAT1* should be investigated by employing luciferase reporter gene analysis (e.g. Minagawa *et al.* 2002).

Sequence variants in *DGAT1* associated with different phenotypes were also reported in *Homo sapience* and in *Arabidopsis thaliana*. The human *DGAT1* promoter allele 79T was found to be associated with leanness, higher HDL-C levels and lower diastolic blood pressures in a population of Turkish women (Ludwig *et al.* 2002). In *Arabidopsis thaliana*, a sequence variant exists for the *TGA1* gene (homologous to mammalian *DGAT1*) in form of a duplication of exon 2 (Zou *et al.* 1999). The sequence variant had a reduced DGAT activity of 40-70% (Katavic *et al.* 1995).

Arguments against the hypothesis of $Lys^{232} \rightarrow Ala$ being causal

There are two possible arguments against the hypothesis of $Lys^{232} \rightarrow$ Ala being causal for variation of the milk fat percentage.

Recent admixture of populations can lead to spurious associations between a phenotype and unlinked candidate loci (Pritchard *et al.* 1999). A well known example of this is human population with different ethnic subgroups (e.g. in North America) that shows a significant correlation between alleles of a gene and the use of chopstick. The gene has nothing to do with chopstick use but just happens to have different allele frequencies in Asians and Caucasians, who differ in chopstick use for purely cultural rather than biological reasons (Hamer *et al.* 2000). In contrast, introgression in the investigated cattle breeds has occurred several generations ago. The association of the breeding value for milk fat percentage with lysine and alanine alleles in *DGAT1* was observed in all three investigated breeds and the degree of association depended on the frequency of the lysine allele in the respective population. One lysine-carrying allele was introgressed into the German Simmental population through Red Holstein animals as shown by the segregation bull 899. However, in addition to the haplotype found in German Holstein, an additional $lysine^{232}$-containing haplotype was found in the second segregating bull of the German Simmental granddaughter-design (Table 4.7 page 50). The results of preliminary haplotype studies and pedigree analysis indicate that the frequency of the lysine allele was increased through introgression, whereas spurious association through population admixture is unlikely.

The second argument against $Lys^{232} \rightarrow$ Ala being the causal variant is the possibility of variants in neighboring genes, which are in linkage disequilibrium with *DGAT1*. Countering the argument requires excluding *DGAT1* neighboring genes as putative candidate genes. Therefore, genes next to *DGAT1* were identified and investigated as a first step towards this goal. Future work includes genotyping of SNPs in genes and BAC ends neighboring *DGAT1*, together with deducing haplotype structure using family pedigree analysis for subsequent linkage disequilibrium studies.

DGAT1 neighboring genes within the BAC contig

A BAC-based contig map of the bovine *DGAT1* region was established. Instead of stepwise BAC end sequencing and library screening for elongation of the BAC contig, species-specific EST sequence information revealed by human mRNA sequences of the chromosomal region

were used to screen the bovine BAC library in one pass. In addition, pooled radiolabeled PCR probes for five loci were used to screen one filter set to reduce the effort further. This protocol enables straightforward mapping of genes in another species.

The fact that 24 genes were identified within the BAC contig indicates a gene rich chromosome region. The number of identified genes is three times as high as we would expect in a region of 576 kb assuming a total of 30 000-35 000 genes for a mammalian genome according to latest estimates in humans (Lander *et al.* 2001). As far as it is known, none of these identified neighboring genes is involved directly in lipid metabolism. The 55 identified polymorphisms within loci neighboring *DGAT1* will form the basis for subsequent linkage disequilibrium studies to positively exclude these other genes as being causal.

Further candidate genes

Lipid secretion from mammary cells has been shown to be regulated by protein kinases (PTK, Rohlfs *et al.* 1993). Altogether, 90 unique protein tyrosine kinase genes have been identified in the human genome (Robinson *et al.* 2000). *PTK2* is a candidate gene for the milk performance QTL on chromosome 14, since it has been mapped in cattle 20.3 cR_{5000} next to *DGAT1*. Two reports also indicate that the DGAT enzyme might be regulated post-translationally by a tyrosine kinase (Haagsman *et al.* 1982; Lau *et al.* 1996). However, no significant functional changes to DGAT activity have been observed when the conserved tyrosine phosphorylation site in human *DGAT1* was mutated by a single base pair substitution (Yu *et al.* 2002). Additional evidence for the regulation of DGAT1 by post-translational modification arises from the observation that there is no correlation between DGAT1 and DGAT2 activities and the mRNA content for either (Waterman *et al.* 2002). At this time, the regulation mechanisms for DGAT enzymes are unknown.

The human gene encoding secreted Ly-6/uPAR related protein 1 precursor (SLURP-1) is located 1.1 Mb proximal to *DGAT1*. Disruption in the expression of the orthologous gene in *Caenorhabditis elegans* (WormBase accession number F55B11.4) by RNA-mediated interference resulted in a reduced fat content in the test strain (Ashrafi *et al.* 2003). *SLURP-1* is therefore a potential candidate gene for milk fat percentage in cattle if it also maps close to *DGAT1*.

Direct evidence of causality of the lysine^{232} → alanine substitution

Direct evidence that the lysine^{232} and alanine^{232} variants of *DGAT1* give rise to phenotypic differences might include:

- Demonstration of differences in the enzymatic activity of the DGAT1 variants by cloning respective transcripts into insect cells (was done in Grisart *et al.* 2002b).
- Demonstration of different transcript levels for the variant lysine^{232} and alanine^{232} encoding alleles. Transcript levels in tissue-specific mRNA can be analyzed using a single base extension assay specific to both alleles (Cowles *et al.* 2002).
- Demonstration of different phenotypes arising from lysine^{232} and alanine^{232} DGAT1 variants using transgenesis. This would be the most conclusive evidence. Transgenesis is a

standard practice in mice, but not in cattle. However, bovine *DGAT1* alleles can be cloned in *DGAT1*-deficient mice to investigate if the lysine^{232} and alanine^{232} variants have an effect on milk performance in mice.

However, final proof for causality is not feasible in most cases. Instead, sufficient evidence may consist of collecting multiple corroborating pieces of evidence, no single one of which is convincing, but which together consistently point to a single candidate gene (Mackay 2001a).

Effects of lysine ^{232} alanine substitution in cattle

The average effects of the gene substitution - the difference of the average effect of the lysine^{232} allele compared to the alanine^{232} allele (Falconer *et al.* 1996) - on milk performance traits were analyzed by Thaller *et al.* (in press-a) using a German Simmental and German Holstein granddaughter design (Table 5.1)

Table 5.1 Estimated gene substitution effects of the lysine allele from first to third lactation.

	German Simmental	German Holstein
fat percentage	0.35%	0.28%
protein percentage	0.10%	0.06%
fat yield	7.5 to 14.8 kg	7.6 to 10.7 kg
protein yield	-3.6 to 0.2 kg	-4.8 to -5.2 kg
milk yield	-242 to -180 kg	-260 to -320 kg

Correlation of gene substitution effects of Lys^{232} → Ala on milk performance traits seems to be negative: compared to the alanine^{232} variant, the lysine^{232} variant increases fat yield while decreasing milk and protein yield (Table 5.1). This observation might be because the available glucose in the mammary gland is a precursor for the synthesis for all lipids, amino acids and particularly for lactose (Kronfeld 1982). Increased synthesis of one component would result in less precursor being available for the other components. The volume of milk secreted is closely related to the rate of lactose synthesis, since the apical membranes of the mammary secretory cells are impermeable to lactose, but freely permeable to water. Lactose is synthesized within the lumen of the Golgi apparatus by an enzyme complex collectively known as lactose synthase and transferred by exocytosis to the luminal side of the cell (Shennan *et al.* 2000). Finally, lactose draws water into the milk by osmosis and thereby determines the milk yield.

Additive and dominant effects could not be estimated because no lactating cows were genotyped. However, Grisart *et al.* (2002a) calculated the additive effect of Lys^{232} → Ala in a daughter design. The dominance derivation – arising from the property of dominance among the alleles at a locus (Falconer *et al.* 1996) – proved not to be significantly different from zero (see Table 2 in Grisart *et al.* 2002a).

Intramuscular fat (IMF)

In addition to the QTL for milk fat percentage on bovine chromosome 14, a QTL for intramuscular fat (% lipid content of muscle) was identified in close proximity to it (about 8.3 cM

away, Figure 2 in Riquet *et al.* 1999) through association with the microsatellite locus *CSSM66* (Barendse 1999). A SNP in 5' region of *tyroglobuline* (*TG*), in close linkage with *CSSM66*, was associated with enhanced intramuscular fat (Barendse 1999). A preliminary study indicates that *TG* and *DGAT1* have significant effects on intramuscular fat in German Holstein and Charolais cattle (Thaller *et al.* in press-b). These effects seem to be independent of one another and both intramuscular fat enhancing effects seem to be recessive. The lysine^{232} allele also seemed to be associated with high intramuscular fat. Studies involving a larger number of animals and additional breeds are presently underway.

Applications arising from the observed association in DGAT1

Natural variation through the effects of the different *DGAT1* alleles can be used for marker-assisted selection (MAS). Introduction of transgenes into breeding populations are conceivable, but will not be applicable in the near future in any large scale. Even if functional proof is lacking that Lys^{232} → Ala is the causal mutation, the locus can be used as marker (in linkage with the causal mutation) in selection programs.

MAS may support breeding in four ways (Georges 2001):

- increasing genetic variance by marker assisted introgression,
- increasing selection accuracy by genotyping phenotype-associated loci,
- reducing generation interval, as genotyping at the DNA level can be achieved at very early stages of development and independently of the sex of the animal, and
- increasing selection intensity, as more animals can be genotyped.

The use of molecular data for genetic improvement will be more effective, when the genetic architecture of a quantitative trait is completely transparent, in terms of the number, the positions and the effects of all the genes involved (Dekkers *et al.* 2002). Unless genetic markers explain most of the genetic variation of a trait, which is far from the case at present, selection must be based on a combination of genetic marker and conventional phenotypic data (Dekkers *et al.* 2002). The use of MAS depends on the expected benefit in relation to costs for DNA collection, genotyping and analysis.

The recently founded boviQuest joint venture already offers a test for the *DGAT1* locus to facilitate changing the milk composition in a breeding population (http://www.boviquest.com/Index.asp).

5.1.2 Porcine DGAT1

In this work, porcine *DGAT1* was cloned and mapped to a region on chromosome 4. Recently, the sequence of porcine *DGAT1* was published and the gene was suggested as a candidate for both the growth and fatness QTLs (Nonneman *et al.* 2002). At the chromosomal position of *DGAT1*, QTLs for back fat and abdominal fat (map positions 3 and 7 cM, respectively) were reported (Andersson *et al.* 1994). However, these QTLs were not confirmed in following studies (Rattink *et al.* 2000; Walling *et al.* 2000). Major QTLs were located at the long arm of chromosome 4, whereas *DGAT1* mapped to the telomeric end of the short arm. *DGAT1* may be therefore not be a candidate gene for fatness QTLs.

5.2 SNP genotyping by SBE

Aside from DNA sampling and preparation, genotyping is the limiting factor in genetic studies. SNP genotyping by single base extension (SBE) is a commonly used method for reliable allele discrimination, with several methods for detection of SBE reaction products being available. Detection of SBE products with an automated sequencer enables reasonable throughput even in a small lab: multiplexing up to 18 SNPs per reaction and parallel detection of 96 samples has been reported (Lindblad-Toh *et al.* 2000). All steps can be done by a pipetting robot. The cost of one SBE reaction as described in this thesis was 0.30 Euro; for complete genotyping (PCR, cleanup, SBE, size marker, gel), the cost is 1.2 Euro. Multiplexing reduces the cost considerably, so by fourfold multiplexing, the cost per genotype could be reduced to 0.30 Euro. SBE genotyping is flexible through the use of regular primers without expensive labeling. However, multiplex systems need some optimization effort.

5.3 DGAT2 gene family

5.3.1 Bovine members of DGAT2 gene family

Characterization

DGAT2 and DGAT1 share no sequence similarity and their membrane topologies differ considerably. Hydrophobic analysis of the DGAT2 peptide sequence revealed two putative transmembrane domains (Cases *et al.* 2001), whereas DGAT1 has nine (Oelkers *et al.* 1998). Chromosomal localizations of human DGAT2 gene family members coincide with paralogous blocks on human chromosome 2, 11 and X (http://wolfe.gen.tcd.ie/dup/human5.28/, McLysaght *et al.* 2002). The newly identified *DC6* definitely belongs to the *DGAT2* gene family, but differs clearly from all other known members. Enzyme functions have been identified in mice for *DGAT2* (Cases *et al.* 2001) and *DC2* (monoacylglycerol acyltransferase, MGAT1, (Yen *et al.* 2002)). Recently, *DC5* was identified in human and mice as gene encoding monoacylglycerol acyltransferase 2 (MGAT2, Cao *et al.* 2003; Yen *et al.* 2003). Functions for the other members are not known. They might have similar activities or they might be pseudogenes.

DGAT2 variants and variation in milk fat percentage

A missense mutation in *DC2* exon 4 results in a non-conservative substitution of cysteine^{170} (uncharged, hydrophobic residue) to lysine (positively charged, hydrophilic residue). Together with the fact that paired cysteines can form disulfide bonds in proteins, the substitution may alter the activity of DC2, which encodes monoacylglycerol acyltransferase. However, no significant association was found between the polymorphisms identified in bovine *DGAT2*, *DC2* and *DC5* and the breeding value for milk fat percentage.

Effects on other traits

The literature was searched for reported QTL regions that match the chromosomal positions of bovine *DGAT2*, *DC2* and *DC5*. *DGAT2* and *DC5* were mapped cytogenetically to the same

region on bovine chromosome 15 as the microsatellite marker *INRA50* (Vaiman *et al.* 1993). Microsatellite marker *INRA50* is within the 95% confidence interval of a QTL for beef longissimus tenderness in steers (Keele *et al.* 1999). Tenderness is the mechanical strength of intramuscular connective tissue. The main determinant of tenderness appears to be the extent of postmortem proteolysis of key target proteins within muscle fibers (Taylor *et al.* 1995). The calpain proteolytic enzyme family (CAPN, calcium-activated neutral proteases) and calpastatin (CAST, endogenous protease inhibitor that acts specifically on calpain, bovine chromosome 7) are considered to be candidate genes for meat tenderness. However, the development of adipose tissues in longissimus muscle appears to disorganize the structure of the intramuscular connective tissue and contributes to tenderization of highly marbled beef as shown in Japanese Black cattle during the late fattening period (Nishimura *et al.* 1999). Therefore, *DGAT2* and *DC5* may also be candidate genes for tenderness. No SNP with an effect to the peptide sequence was found in the coding region of *DGAT2* and *DC5*, but only samples from the *Bos taurus* breeds German Holstein, Simmental and Brown were investigated in this thesis, whereas a Brahman × Hereford cross bull was used for QTL mapping.
No QTLs were found in the regions containing bovine *DC2*.

5.3.2 Porcine members of DGAT2 gene family

Porcine *DGAT2*, *DC5* and *DC7* were cloned and mapped and the literature was searched for reported QTL regions that match the chromosomal positions of these genes (Table 5.2). Porcine *DGAT2* and *DC5* were located in a QTL region for intramuscular fat content.
DC3, *DC4* and *DC6* were not isolated in the pig. Human *DC3* and *DC4* are 2 Mb distal to *Androgen receptor* (*AR*). *AR* is located in the centromere region of chromosome X in human and pig. Several porcine QTL for intramuscular fat content and backfat thickness were reported in this chromosomal region. Assuming *DC3* and *DC4* also map in the pig to this region, they are candidate genes for the QTLs in that region.

Table 5.2: Porcine DGAT2 gene family members and known QTL.

Gene	Position of gene	QTL[a]	Position of QTL	Reference
DGAT2, DC5	9pter-p23	IMF	11 cM	(Gossner 2002)
DC7	3pter-p15	-		
DC3, DC4, DC6	(X)[b]	BFT	60 cM	(Harlizius *et al.* 2000)
		IMF	69 cM	(Rohrer *et al.* 1998)

[a]BFT, backfat thickness; IMF, intramuscular fat content.
[b]Genes not isolated in pips. However comparative mapping would assign these genes to the centromere region of chromosome X.

6 Summary

It has been shown in this dissertation that a missense mutation ($Lys^{232} \rightarrow Ala$) in the bovine *DGAT1* gene is associated with the variation in the breeding value for milk fat percentage. In parallel with Grisart and colleagues (2002a), this thesis project resulted in the first successful positional cloning of a quantitative trait locus (QTL) in a species other than a model organism or a plant.

Milk fat percentage is an important performance feature in cattle breeding. As a quantitative trait, milk fat percentage is determined by the collective effect of multiple genes and environmental factors. The genetic variability is the basis for breeding. Knowledge of the genes causing variation in a trait enables the testing of breeding animals for their genetic potential early in their development and independently from their gender. *DGAT1* encodes diacylglycerol acyltransferase (EC 2.3.1.20), which catalyzes the final step in triglyceride synthesis. *DGAT1* became a prime candidate gene for milk fat percentage after it was reported that *DGAT1* knock out mice were viable but unable to produce milk.

Screening of a bovine BAC library identified four BAC clones containing *DGAT1*. Physical mapping with fluorescence *in situ* hybridization and a radiation hybrid cell panel placed *DGAT1* centromeric on bovine chromosome 14. A quantitative trait locus (QTL) for milk fat percentage was reported in the same region in several studies. The nucleotide sequence and gene structure of bovine *DGAT1* was determined. The coding region spans 8.7 kb and the gene is predicted to encode a 490 residue protein. Re-sequencing revealed 21 single nucleotide polymorphisms (SNPs) and a variable number of tandem repeats (VNTR) in the upstream sequence of *DGAT1*. For an association study, bulls having extreme high and low breeding values for milk fat percentage were used. The bulls belonged to the breeds German Holstein (n: 2 x 32), German Simmental (n: 2 x 32) and German Brown (n: 2 x 20). Allele frequency estimations for most SNPs based on sequence traces of pooled DNA samples revealed significant associations with the breeding value for milk fat percentage. One polymorphism in exon 8 results in a lysine to alanine substitution at residue 232. The lysine variant was associated with high milk fat percentage and the alanine variant with low milk fat percentage in different breeds ($\alpha = 0.001$ in German Holstein and German Simmental, $\alpha = 0.05$ in German Brown). Two animals that were genotyped heterozygous (Qq) at the QTL based on marker-assisted QTL-genotyping were heterozygous for the $Lys^{232} \rightarrow Ala$ substitution, whereas 14 animals that are most likely qq at the QTL were homozygous for the alanine-encoding allele. However, not all individuals with high breeding values in the German Simmental breed carry the $lysine^{232}$ allele. This is also true for German Brown bulls, where only two out of 20 carry the $lysine^{232}$ allele. Both breeds have noticeably lower $lysine^{232}$ allele frequencies (7% and 2%, respectively) than the German Holstein breed (34%).

These observations are compatible with a polygene model for quantitative traits. Lysine seems to be the ancestral allele, since lysine was fixed in the investigated *Bos indicus* breeds and was present in *DGAT1* GenBank entries for other mammals including human.
The contra argument that the association might be spurious due to population admixture was not considered as likely because (i) the association was observed in all three investigated cattle breeds, (ii) the history of the breeds, and (iii) the results of preliminary haplotype studies. A more valid argument against the hypothesis of $Lys^{232} \rightarrow$ Ala being causal for variation in milk fat percentage is that the observed association is due to the linkage with a causal mutation in a gene close to *DGAT1* or other *DGAT1* variants.
As first step towards the goal of excluding genes close to *DGAT1* as being causal, a bovine BAC contig was constructed spanning 576 kb of the chromosomal region containing *DGAT1* and twenty-three neighboring genes. Annotated human genes were used to search for homologous bovine EST sequences. PCR based on bovine EST sequence information was applied to first screen the BAC library in one pass and subsequently for mapping genes within the BAC contig. BAC ends of 18 isolated clones and genes mapped in the contig were partly sequenced and screened for sequence variants. After genotyping the SNPs in *DGAT1*-neighboring genes and BAC ends, haplotypes can now be deduced by family pedigree analysis for subsequent linkage disequilibrium studies.
To enhance SNP genotyping, a multiplex single base extension (SBE) assay was optimized based on a self-composed SBE reaction including four fluorescent labeled didesoxynucleotides, thermosequenase and length adjusted primers for separation on an automated DNA sequencer.
Bovine BAC clones were isolated for *DGAT2*, *DC2* and *DC5*. Genes were physically mapped to bovine chromosomes 15q23-25, 2q42-44 and 15q23-25, respectively. All exons and parts of introns were sequenced. Bovine *DGAT2*, *DC2* and *DC5* genes encode for proteins with 361, 334 and 284 residues, respectively. Re-sequencing revealed sequence variants in all three genes. Allele frequency estimates based on sequence traces of pooled DNA samples revealed no significant association with breeding values for milk fat percentage. Recently, *DC2* and *DC5* were identified as genes encoding monoacylglycerol acyltransferase 1 (MGAT1) and monoacylglycerol acyltransferase 2 (MGAT2), respectively.
Porcine BAC clones were isolated for *DGAT1*, *DGAT2*, *DC5* and *DC7*. Genes were physically mapped to porcine chromosomes 4pter-p15, 9pter-p23, 9pter-p23 and 3pter-p15, respectively.
Bovine *DGAT2* and porcine *DGAT2* and *DC5* (*MGAT2*) are candidate genes for fatness QTL, which were reported in the respective chromosome positions.

Milk performance of cows can be influenced by selection based on genotyping results for *DGAT1* variants (marker assisted selection). Gene substitution effects of milk performance traits were negatively correlated: compared to the $alanine^{232}$ variant, the $lysine^{232}$ variant increases fat yield while decreasing milk and protein yield.

7 Acknowledgments

First, I am most grateful to my supervisor Prof. Dr. Ruedi Fries for giving me the opportunity and encouragement to work on this exciting subject, and for providing me with the facilities needed to do this research.

I also thank the Arbeitsgemeinschaft Deutscher Rinderzüchter (German cattle breeders federation, ADR) and the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research, BMBF) for supporting this project financially.

Special thanks to the various German artificial insemination stations for donating semen samples and in particular to Christa Kühn (FBM Dummersdorf) for assembling semen samples for German Holstein pools and to Prof. Dr. med. Thomas Meitinger (Institut für Humangenetik, TUM) for supplying DNA samples of radiation hybrid panel GB4.

I owe my gratitude to Prof. Dr. Dr. Heinrich H.D. Meyer, Prof. Dr. Dieter Schams and PD Dr. Ralf Einspanier (Institute of Physiology) for providing me with an introduction to lactation physiology and RACE PCR.

I also want to express my deepest gratitude Dr. Georg Thaller for numerous discussions and especially for introduction to and support in statistics and population genetics.

For discussions and guidance regarding the laboratory work, I thank Dr. Gregor Durstewitz, Sonja Kollers, Dr. Johannes Buitkamp and Rudolf Antes. I especially thank Dr. Felix Habermann for his critical view on several aspects of my work.

I warmly thank Fabian A. O. Werner for his valuable assistance during his diploma work and Wolfgang Krämer for discussions and his help in many matters. Both were colleagues who I enjoyed sharing time and the office with.

I also warmly thank the student apprentices Ariane Alzinger and Miranda van Eckefeld for their great help and active interest.

I deeply appreciate the comments that Dr. Olaf Bininda-Emonds made on this dissertation as well as on other manuscripts.

Also, I wish to thank Rita Popken for her invaluable and friendly help in administrative matters.

I warmly thank all colleagues not mentioned by name for their help, encouragement and friendship during these years.

I want to extend my deepest thanks to Dipl.-Ing. Hermann Heiler (FH Weihenstephan); his support was essential for me to accomplish this work.

Finally, I would like to thank my family, friends and roommates who helped me to relax and showed me that life is still more than a Ph.D. thesis.

Andreas Winter,
Freising, April 2003

8 Bibliography

Abumrad, N. A., M. R. el-Maghrabi, E. Z. Amri, E. Lopez and P. A. Grimaldi (1993). "Cloning of a rat adipocyte membrane protein implicated in binding or transport of long-chain fatty acids that is induced during preadipocyte differentiation. Homology with human CD36." J Biol Chem **268**(24): 17665-8.

Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno and et al. (1991). "Complementary DNA sequencing: expressed sequence tags and human genome project." Science **252**(5013): 1651-6.

ADR (2002). Rinderproduktion in der Bundesrepublik Deutschland 2001. Bonn, Germany, Arbeitsgemeinschaft Deutscher Rinderzüchter e.V.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.

Andersson, L., C. S. Haley, H. Ellegren, S. A. Knott, M. Johansson, K. Andersson, L. Andersson-Eklund, I. Edfors-Lilja, M. Fredholm, I. Hansson and et al. (1994). "Genetic mapping of quantitative trait loci for growth and fatness in pigs." Science **263**(5154): 1771-4.

Aparicio, S., J. Chapman, E. Stupka, N. Putnam, J. M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M. D. Gelpke, J. Roach, T. Oh, I. Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas*, et al.* (2002). "Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes." Science **297**(5585): 1301-10.

Ashrafi, K., F. Y. Chang, J. L. Watts, A. G. Fraser, R. S. Kamath, J. Ahringer and G. Ruvkun (2003). "Genome-wide RNAi analysis of Caenorhabditis elegans fat regulatory genes." Nature **421**(6920): 268-72.

Ashwell, M. S., C. P. Van Tassell and T. S. Sonstegard (2001). "A genome scan to identify quantitative trait loci affecting economically important traits in a US Holstein population." J Dairy Sci **84**(11): 2535-42.

Attwood, J. T., R. L. Yung and B. C. Richardson (2002). "DNA methylation and the regulation of gene transcription." Cell Mol Life Sci **59**(2): 241-57.

Band, M. R., J. H. Larson, M. Rebeiz, C. A. Green, D. W. Heyen, J. Donovan, R. Windish, C. Steining, P. Mahyuddin, J. E. Womack and H. A. Lewin (2000). "An ordered comparative map of the cattle and human genomes." Genome Res **10**(9): 1359-68.

Barber, M. C., R. A. Clegg, M. T. Travers and R. G. Vernon (1997). "Lipid metabolism in the lactating mammary gland." Biochim Biophys Acta **1347**(2-3): 101-26.

Barendse, W., S. M. Armitage, L. M. Kossarek, A. Shalom, B. W. Kirkpatrick, A. M. Ryan, D. Clayton, L. Li, H. L. Neibergs, N. Zhang and et al. (1994). "A genetic linkage map of the bovine genome." Nat Genet **6**(3): 227-35.

Barendse, W., D. Vaiman, S. J. Kemp, Y. Sugimoto, S. M. Armitage, J. L. Williams, H. S. Sun, A. Eggen, M. Agaba, S. A. Aleyasin, M. Band, M. D. Bishop, J. Buitkamp, K. Byrne, F. Collins, L. Cooper, W. Coppettiers, B. Denys, R. D. Drinkwater, K. Easterday*, et al.* (1997). "A medium-density genetic linkage map of the bovine genome." Mamm Genome **8**(1): 21-8.

Barendse, W. J. (1999). "Assessing lipid metabolism. International Publication Number W0 99/23248.".

Birnboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**(6): 1513-23.

Bishop, M. D., S. M. Kappes, J. W. Keele, R. T. Stone, S. L. Sunden, G. A. Hawkins, S. S. Toldo, R. Fries, M. D. Grosz, J. Yoo and et al. (1994). "A genetic linkage map for cattle." Genetics **136**(2): 619-39.

Boguski, M. S., T. M. Lowe and C. M. Tolstoshev (1993). "dbEST--database for "expressed sequence tags"." Nat Genet **4**(4): 332-3.

Botstein, D., R. L. White, M. Skolnick and R. W. Davis (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." Am J Hum Genet **32**(3): 314-31.

Bouvier-Nave, P., P. Benveniste, A. Noiriel and H. Schaller (2000a). "Expression in yeast of an acyl-CoA:diacylglycerol acyltransferase cDNA from Caenorhabditis elegans." Biochem Soc Trans **28**(6): 692-5.

Bouvier-Nave, P., P. Benveniste, P. Oelkers, S. L. Sturley and H. Schaller (2000b). "Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase." Eur J Biochem **267**(1): 85-96.

Breathnach, R., C. Benoist, K. O'Hare, F. Gannon and P. Chambon (1978). "Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries." Proc Natl Acad Sci U S A **75**(10): 4853-7.

Buhman, K. K., H. C. Chen and R. V. Farese, Jr. (2001). "The enzymes of neutral lipid synthesis." J Biol Chem **276**(44): 40369-72.

Buhman, K. K., S. J. Smith, S. J. Stone, J. J. Repa, J. S. Wong, F. F. Knapp, Jr., B. J. Burri, R. L. Hamilton, N. A. Abumrad and R. V. Farese, Jr. (2002). "DGAT1 is not essential for intestinal triacylglycerol absorption or chylomicron synthesis." J Biol Chem **277**(28): 25474-9.

Buitkamp, J., R. Antes and V. Wagner (1999). DNA profiling in veterinary medicine. DNA profiling and DNA fingerprinting. J. T. Epplen and T. Lubjuhn. Basel, Birkhäuser**:** 53-70.

Bundeslandwirtschaftsministerium (1998). Statistisches Jahrbuch über Ernährung, Landwirtschaft und Forsten der Bundesrepublik Deutschland. Münster-Hiltrup, Landwirtschaftsverlag.

Burt, D. W., C. Bruley, I. C. Dunn, C. T. Jones, A. Ramage, A. S. Law, D. R. Morrice, I. R. Paton, J. Smith, D. Windsor, A. Sazanov, R. Fries and D. Waddington (1999). "The dynamics of chromosome evolution in birds and mammals." Nature **402**(6760): 411-3.

Buszczak, M., X. Lu, W. A. Segraves, T. Y. Chang and L. Cooley (2002). "Mutations in the midway Gene Disrupt a Drosophila Acyl Coenzyme A. Diacylglycerol acyltransferase." Genetics **160**(4): 1511-8.

Cai, L., J. F. Taylor, R. A. Wing, D. S. Gallagher, S. S. Woo and S. K. Davis (1995). "Construction and characterization of a bovine bacterial artificial chromosome library." Genomics **29**(2): 413-25.

Campagne, F. and H. Weinstein (1999). "Schematic representation of residue-based protein context-dependent data: an application to transmembrane proteins." J Mol Graph Model **17**(3-4): 207-13.

Cao, G., J. L. Goldstein and M. S. Brown (1996). "Complementation of mutation in acyl-CoA:cholesterol acyltransferase (ACAT) fails to restore sterol regulation in ACAT-defective sterol- resistant hamster cells." J Biol Chem **271**(24): 14642-8.

Cao, J., J. Lockwood, P. Burn and Y. Shi (2003). "Cloning and functional characterization of a mouse intestinal Acyl- CoA:monoacylglycerol acyltransferase, MGAT2." J Biol Chem **7**: 7.

Cases, S., S. J. Smith, Y. W. Zheng, H. M. Myers, S. R. Lear, E. Sande, S. Novak, C. Collins, C. B. Welch, A. J. Lusis, S. K. Erickson and R. V. Farese, Jr. (1998). "Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis." Proc Natl Acad Sci U S A **95**(22): 13018-23.

Cases, S., S. Stone, P. Zhou, E. Yen, B. Tow, K. D. Lardizabal, T. Voelker and R. V. Farese, Jr. (2001). "Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members." J Biol Chem **31**: 31.

Chang, C. C., H. Y. Huh, K. M. Cadigan and T. Y. Chang (1993). "Molecular cloning and functional expression of human acyl-coenzyme A:cholesterol acyltransferase cDNA in mutant Chinese hamster ovary cells." J Biol Chem **268**(28): 20747-55.

Cheng, D., R. L. Meegalla, B. He, D. A. Cromley, J. T. Billheimer and P. R. Young (2001). "Human acyl-CoA:diacylglycerol acyltransferase is a tetrameric protein." Biochem J **359**(Pt 3): 707-14.

Chowdhary, B. P., L. Fronicke, I. Gustavsson and H. Scherthan (1996). "Comparative analysis of the cattle and human genomes: detection of ZOO- FISH and gene mapping-based chromosomal homologies." Mamm Genome **7**(4): 297-302.

Church, G. M. and W. Gilbert (1984). "Genomic sequencing." Proc Natl Acad Sci U S A **81**(7): 1991-5.

Coleman, R. A. and E. B. Haynes (1985). "Subcellular location and topography of rat hepatic monoacylglycerol acyltransferase activity." Biochim Biophys Acta **834**(2): 180-7.

Coleman, R. A., T. M. Lewin and D. M. Muoio (2000). "Physiological and nutritional regulation of enzymes of triacylglycerol synthesis." Annu Rev Nutr **20**: 77-103.

Coppieters, W., J. Riquet, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar and M. Georges (1998). "A QTL with major effect on milk yield and composition maps to bovine chromosome 14." Mamm Genome **9**(7): 540-4.

Cowles, C. R., N. H. Joel, D. Altshuler and E. S. Lander (2002). "Detection of regulatory variation in mouse genes." Nat Genet **32**(3): 432-7.

Cox, D. R., M. Burmeister, E. R. Price, S. Kim and R. M. Myers (1990). "Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes." Science **250**(4978): 245-50.

Dahlqvist, A., U. Stahl, M. Lenman, A. Banas, M. Lee, L. Sandager, H. Ronne and S. Stymne (2000). "Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants." Proc Natl Acad Sci U S A **97**(12): 6487-92.

Daly, M. J., J. D. Rioux, S. F. Schaffner, T. J. Hudson and E. S. Lander (2001). "High-resolution haplotype structure in the human genome." Nat Genet **29**(2): 229-32.

Darvasi, A. and M. Soller (1994). "Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus." Genetics **138**(4): 1365-73.

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller and M. Soller (1993). "Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map." Genetics **134**(3): 943-51.

Dekkers, J. C. and F. Hospital (2002). "The use of molecular genetics in the improvement of agricultural populations." Nat Rev Genet **3**(1): 22-32.

Deloukas, P., G. D. Schuler, G. Gyapay, E. M. Beasley, C. Soderlund, P. Rodriguez-Tome, L. Hui, T. C. Matise, K. B. McKusick, J. S. Beckmann, S. Bentolila, M. Bihoreau, B. B. Birren, J. Browne, A. Butler, A. B. Castle, N. Chiannilkulchai, C. Clee, P. J. Day, A. Dehejia*, et al.* (1998). "A physical map of 30,000 human genes." Science **282**(5389): 744-6.

Durstewitz, G., A. Winter and R. Fries (2002). "SNP allele frequency estimation based on the analysis of swquencing traces." 28th International Society of Animal Genetics meeting: Göttingen, Germany.

Eckel, R. (1989). "Lipoprotein lipase. A multifunctional enzyme relevant to common metabolic diseases." N Engl J Med **320**(16): 1060-1068.

Ewing, B. and P. Green (1998a). "Base-calling of automated sequencer traces using phred. II. Error probabilities." Genome Res **8**(3): 186-94.

Ewing, B., L. Hillier, M. C. Wendl and P. Green (1998b). "Base-calling of automated sequencer traces using phred. I. Accuracy assessment." Genome Res **8**(3): 175-85.

Fahrenkrug, S. C., G. A. Rohrer, B. A. Freking, T. P. Smith, K. Osoegawa, C. L. Shu, J. J. Catanese and P. J. de Jong (2001). "A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration." Mamm Genome **12**(6): 472-4.

Falconer, D. S. and T. F. C. Makay (1996). Introduction to Quantitative genetics. Harlow, Essex, England, Longman Group Ltd.

Farese, R. V., Jr., S. Cases and S. J. Smith (2000). "Triglyceride synthesis: insights from the cloning of diacylglycerol acyltransferase." Curr Opin Lipidol **11**(3): 229-34.

Farnir, F., B. Grisart, W. Coppieters, J. Riquet, P. Berzi, N. Cambisano, L. Karim, M. Mni, S. Moisio, P. Simon, D. Wagenaar, J. Vilkki and M. Georges (2002). "Simultaneous Mining of Linkage and Linkage Disequilibrium to Fine Map Quantitative Trait Loci in Outbred Half-Sib Pedigrees. Revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14." Genetics **161**(1): 275-87.

Fisher, R. A. (1918). "The correlation between relatives on the supposition of Mendelian inheritance." Transactions of the Royal Society of Edinburgh **52**: 399-433.

Fitch, D. H., W. J. Bailey, D. A. Tagle, M. Goodman, L. Sieu and J. L. Slightom (1991). "Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates." Proc Natl Acad Sci U S A **88**(16): 7396-400.

Fitch, W. M. (1970). "Distinguishing homologous from analogous proteins." Syst Zool **19**(2): 99-113.

Foster, T. (1965). Modern quantum chemistry, Istanbul lectures, part III. New York, Academic Press.

Fries, R. and G. Durstewitz (2001). "Digital DNA signatures for animal tagging." Nat Biotechnol **19**(6): 508.

Fronicke, L. and J. Wienberg (2001). "Comparative chromosome painting defines the high rate of karyotype changes between pigs and bovids." Mamm Genome **12**(6): 442-9.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-82.

Geldermann, H. (1975). "Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods." Theor. Appl. Genet.(46): 319–330.

Genetics Computer Group (2001). Wisconsin Package Version 10.2. Madison, Wisc.

Georges, M. (2001). "Recent progress in livestock genomics and potential impact on breeding programs." Theriogenology **55**(1): 15-21.

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao and et al. (1995). "Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing." Genetics **139**(2): 907-20.

Giannoulia, K., K. Haralampidis, Z. Poghosyan, D. J. Murphy and P. Hatzopoulos (2000). "Differential expression of diacylglycerol acyltransferase (DGAT) genes in olive tissues." Biochem Soc Trans **28**(6): 695-7.

Glatz, J. F. C. and G. J. van der Vusse (1996). "Cellular fatty acid-binding proteins: their function and physiological significance." Progress in Lipid Research **35**(3): 243-282.

Glazier, A. M., J. H. Nadeau and T. J. Aitman (2002). "Finding genes that underlie complex traits." Science **298**(5602): 2345-9.

Goddard, M. E. and G. R. Wiggans (1999). Genetic improvement of dairy cattle. The Genetics of Cattle. R. Fries and A. Ruvinsky. Wallingford, CABI Publishing**:** 511-37.

Gordon, D., C. Abajian and P. Green (1998). "Consed: a graphical tool for sequence finishing." Genome Res **8**(3): 195-202.

Gossner, H. (2002). Kartierung von QTL in einer F2-Kreuzungspopulation des Schweins unter Verwendung der QTL-Wxpress-Applikation. Department für Tierwissenschaften. Freising-Weihenstephan, Germany, Technische Universität München**:** 86-91.

Goureau, A., M. Yerle, A. Schmitz, J. Riquet, D. Milan, P. Pinton, G. Frelat and J. Gellin (1996). "Human and porcine correspondence of chromosome segments using bidirectional chromosome painting." Genomics **36**(2): 252-62.

Grant, D. M. and M. S. Phillips (2001). Technologies for the Analysis of Single-Nucleotide Polymorphisms: An Overview. Pharmacogenomics. Kalow, Meyer and Tyndale. Inc. New York, Dekker, Marcel. **113:** 183-190.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges and R. Snell (2002a). "Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition." Genome Res **12**(2): 222-31.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges and R. Snell (2002b). "Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition." 28th International Society of Animal Genetics meeting: Göttingen, Germany.

Gut, I. G. (2001). "Automation in genotyping of single nucleotide polymorphisms." Hum Mutat **17**(6): 475-92.

Gyapay, G., K. Schmitt, C. Fizames, H. Jones, N. Vega-Czarny, D. Spillett, D. Muselet, J. F. Prud'Homme, C. Dib, C. Auffray, J. Morissette, J. Weissenbach and P. N. Goodfellow (1996). "A radiation hybrid map of the human genome." Hum Mol Genet **5**(3): 339-46.

Haagsman, H. P., C. G. de Haas, M. J. Geelen and L. M. van Golde (1982). "Regulation of triacylglycerol synthesis in the liver. Modulation of diacylglycerol acyltransferase activity in vitro." J Biol Chem **257**(18): 10593-8.

Haldane, J. B. S. (1919). "The combination of linkage values, and the calculation of distances between the loci of linked factors." J. of Genetics **8**: 299-309.

Hamer, D. and L. Sirota (2000). "Beware the chopsticks gene." Mol Psychiatry **5**(1): 11-3.

Harlizius, B., A. P. Rattink, D. J. de Koning, M. Faivre, R. G. Joosten, J. A. van Arendonk and M. A. Groenen (2000). "The X chromosome harbors quantitative trait loci for backfat thickness and intramuscular fat content in pigs." Mamm Genome **11**(9): 800-2.

Harrington, J. J. and M. R. Lieber (1994). "Functional domains within FEN-1 and RAD2 define a family of structure- specific endonucleases: implications for nucleotide excision repair." Genes Dev **8**(11): 1344-55.

Hastings, M. L. and A. R. Krainer (2001). "Pre-mRNA splicing in the new millennium." Curr Opin Cell Biol **13**(3): 302-9.

Hayes, H. (1995). "Chromosome painting with human chromosome-specific DNA libraries reveals the extent and distribution of conserved segments in bovine chromosomes." Cytogenet Cell Genet **71**(2): 168-74.

Henderson, C. R. (1974). "General flexibility of linear model technique for sire evaluation." Dairy Sci. **57**: 963-972.

Heyen, D. W., J. I. Weller, M. Ron, M. Band, J. E. Beever, E. Feldmesser, Y. Da, G. R. Wiggans, P. M. VanRaden and H. A. Lewin (1999). "A genome scan for QTL influencing milk production and health traits in dairy cattle." Physiol Genomics **1**(3): 165-75.

Hobbs, D. H., C. Lu and M. J. Hills (1999). "Cloning of a cDNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana and its functional expression." FEBS Lett **452**(3): 145-9.

Holland, P. M., R. D. Abramson, R. Watson and D. H. Gelfand (1991). "Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase." Proc Natl Acad Sci U S A **88**(16): 7276-80.

Horn, F., J. Weare, M. W. Beukers, S. Horsch, A. Bairoch, W. Chen, O. Edvardsen, F. Campagne and G. Vriend (1998). "GPCRDB: an information system for G protein-coupled receptors." Nucleic Acids Res **26**(1): 275-9.

Hughes, A. L. (1999). "Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history." J Mol Evol **48**(5): 565-76.

Ibrahimi, A., Z. Sfeir, H. Magharaie, E. Z. Amri, P. Grimaldi and N. A. Abumrad (1996). "Expression of the CD36 homolog (FAT) in fibroblast cells: effects on fatty acid transport." Proc Natl Acad Sci U S A **93**(7): 2646-51.

Jenny, A. and W. Keller (1995). "Cloning of cDNAs encoding the 160 kDa subunit of the bovine cleavage and polyadenylation specificity factor." Nucleic Acids Res **23**(14): 2629-35.

Kappes, S. M., J. W. Keele, R. T. Stone, R. A. McGraw, T. S. Sonstegard, T. P. Smith, N. L. Lopez-Corrales and C. W. Beattie (1997). "A second-generation linkage map of the bovine genome." Genome Res **7**(3): 235-49.

Karchin, R., K. Karplus and D. Haussler (2002). "Classifying G-protein coupled receptors with support vector machines." Bioinformatics **18**(1): 147-59.

Katavic, V., D. W. Reed, D. C. Taylor, E. M. Giblin, D. L. Barton, J. Zou, S. L. Mackenzie, P. S. Covello and L. Kunst (1995). "Alteration of seed fatty acid composition by an ethyl methanesulfonate- induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity." Plant Physiol **108**(1): 399-409.

Kaupe, B., A. Winter, R. Fries and G. Erhardt (in press). "*DGAT1* polymorphism in Bos indicus and Bos taurus cattle breeds and effect of breeding.".

Keele, J. W., S. D. Shackelford, S. M. Kappes, M. Koohmaraie and R. T. Stone (1999). "A region on bovine chromosome 15 influences beef longissimus tenderness in steers." J Anim Sci **77**(6): 1364-71.

Keller, E. B. and W. A. Noon (1984). "Intron splicing: a conserved internal signal in introns of animal pre- mRNAs." Proc Natl Acad Sci U S A **81**(23): 7417-20.

Keller, W., S. Bienroth, K. M. Lang and G. Christofori (1991). "Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA." Embo J **10**(13): 4241-9.

Kitami, T. and J. H. Nadeau (2002). "Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication." Nat Genet **32**(1): 191-4.

Kitao, S., N. M. Lindor, M. Shiratori, Y. Furuichi and A. Shimamoto (1999a). "Rothmund-thomson syndrome responsible gene, RECQL4: genomic structure and products." Genomics **61**(3): 268-76.

Kitao, S., A. Shimamoto, M. Goto, R. W. Miller, W. A. Smithson, N. M. Lindor and Y. Furuichi (1999b). "Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome." Nat Genet **22**(1): 82-4.

Kosambi, D. D. (1944). "The estimation of map distances from recombination values." Ann. Eugenics **12**: 172-175.

Kostrikis, L. G., S. Tyagi, M. M. Mhlanga, D. D. Ho and F. R. Kramer (1998). "Spectral genotyping of human alleles." Science **279**(5354): 1228-9.

Kronfeld, D. S. (1982). "Major metabolic determinants of milk volume, mammary efficiency, and spontaneous ketosis in dairy cows." J Dairy Sci **65**(11): 2204-12.

Kwok, P. Y. (2001). "Methods for genotyping single nucleotide polymorphisms." Annu Rev Genomics Hum Genet **2**: 235-58.

Kwok, P. Y., C. Carlson, T. D. Yager, W. Ankener and D. A. Nickerson (1994). "Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products." Genomics **23**(1): 138-44.

Landegren, U., R. Kaiser, J. Sanders and L. Hood (1988). "A ligase-mediated gene detection technique." Science **241**(4869): 1077-80.

Lander, E. S. and D. Botstein (1989). "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps." Genetics **121**(1): 185-99.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan*, et al.* (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-48.

Lardizabal, K. D., J. T. Mai, N. W. Wagner, A. Wyrick, T. Voelker and D. J. Hawkins (2001). "DGAT2: A new diacylglycerol acyltransferase gene family: Purification, cloning and expression in insect cells of two polypeptides from Mortierella ramannniana with diacylglycerol acyltransferase activity." J Biol Chem **31**: 31.

Larsen, F., G. Gundersen, R. Lopez and H. Prydz (1992). "CpG islands as gene markers in the human genome." Genomics **13**(4): 1095-107.

Lau, T. E. and M. A. Rodriguez (1996). "A protein tyrosine kinase associated with the ATP-dependent inactivation of adipose diacylglycerol acyltransferase." Lipids **31**(3): 277-83.

Lehner, R. and A. Kuksis (1993). "Triacylglycerol synthesis by an sn-1,2(2,3)-diacylglycerol transacylase from rat intestinal microsomes." J Biol Chem **268**(12): 8781-6.

Lehner, R. and A. Kuksis (1996). "Biosynthesis of triacylglycerols." Prog Lipid Res **35**(2): 169-201.

Lenski, R. E., C. Ofria, T. C. Collier and C. Adami (1999). "Genome complexity, robustness and genetic interactions in digital organisms." Nature **400**(6745): 661-4.

Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Laviolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J. B. Fan, T. Gingeras, J. Warrington, N. Patil, T. J. Hudson and E. S. Lander (2000). "Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse." Nat Genet **24**(4): 381-6.

Lindsay, S. and A. P. Bird (1987). "Use of restriction enzymes to detect potential gene sequences in mammalian DNA." Nature **327**(6120): 336-8.

Litt, M. and J. A. Luty (1989). "A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene." Am J Hum Genet **44**(3): 397-401.

Looft, C., N. Reinsch, C. Karall-Albrecht, S. Paul, M. Brink, H. Thomsen, G. Brockmann, C. Kuhn, M. Schwerin and E. Kalm (2001). "A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine Chromosome 14." Mamm Genome **12**(8): 646-50.

Ludwig, E., R. Mahley, E. Palaoglu, S. Ozbayrakci, M. Balestra, I. Borecki, T. Innerarity and R. Farese (2002). "DGAT1 promoter polymorphism associated with alterations in body mass index, high density lipoprotein levels and blood pressure in Turkish women." Clin Genet **62**(1): 68-73.

Ma, R. Z., J. E. Beever, Y. Da, C. A. Green, I. Russ, C. Park, D. W. Heyen, R. E. Everts, S. R. Fisher, K. M. Overton, A. J. Teale, S. J. Kemp, H. C. Hines, G. Guerin and H. A. Lewin (1996). "A male linkage map of the cattle (Bos taurus) genome." J Hered **87**(4): 261-71.

Mackay, T. F. (2001a). "The genetic architecture of quantitative traits." Annu Rev Genet **35**: 303-39.

Mackay, T. F. (2001b). "Quantitative trait loci in Drosophila." Nat Rev Genet **2**(1): 11-20.

Mackinnon, M. J. and J. I. Weller (1995). "Methodology and accuracy of estimation of quantitative trait loci parameters in a half-sib design using maximum likelihood." Genetics **141**(2): 755-70.

Marshall, M. O. and J. Knudsen (1977). "Biosynthesis of triacylglycerols containing short-chain fatty acids in lactating cow mammary gland. Activity of diacylglycerol acyltransferase towards short-chain acyl-CoA esters." Eur J Biochem **81**(2): 259-66.

Marshall, M. O. and J. Knudsen (1979). "Specificity of diacylglycerol acyltransferase from bovine mammary gland, liver and adipose tissue towards acyl-CoA esters." Eur J Biochem **94**(1): 93-8.

Mather, I. H. and T. W. Keenan (1998). "Origin and secretion of milk lipids." J Mammary Gland Biol Neoplasia **3**(3): 259-73.

Mayorek, N., I. Grinstein and J. Bar-Tana (1989). "Triacylglycerol synthesis in cultured rat hepatocytes. The rate- limiting role of diacylglycerol acyltransferase." Eur J Biochem **182**(2): 395-400.

McLysaght, A., K. Hokamp and K. H. Wolfe (2002). "Extensive genomic duplication during early chordate evolution." Nat Genet **31**(2): 200-4.

Medjugorac, I., W. Kustermann, P. Lazar, I. Russ and F. Pirchner (1994). "Marker-derived phylogeny of European cattle supports demic expansion of agriculture." Anim Genet **25 Suppl 1**: 19-27.

Milanesi, L., D. D'Angelo and I. B. Rogozin (1999). "GeneBuilder: interactive in silico prediction of gene structure." Bioinformatics **15**(7-8): 612-21.

Minagawa, M., T. Yasuda, T. Watanabe, K. Minamitani, Y. Takahashi, D. Goltzman, J. H. White, G. N. Hendy and Y. Kohno (2002). "Association between AAAG repeat polymorphism in the P3 promoter of the human parathyroid hormone (PTH)/PTH-related peptide receptor gene and adult height, urinary pyridinoline excretion, and promoter activity." J Clin Endocrinol Metab **87**(4): 1791-6.

Morgan, T. H. (1928). The theory of genes, New Haven, Conn.: Yale University Press.

Mulder, H. and P. Walstra (1974). The Milk Fat Globule: Emulsion as Applied to Milk Products and Comparable Foods. Wageningen, The Netherlands, Centre for Agricultural publishing and documentation.

Nadeau, J. H. and D. Sankoff (1997). "Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution." Genetics **147**(3): 1259-66.

Nadeau, J. H. and B. A. Taylor (1984). "Lengths of chromosomal segments conserved since divergence of man and mouse." Proc Natl Acad Sci U S A **81**(3): 814-8.

Neville, M. C. and M. F. Picciano (1997). "Regulation of milk lipid secretion and composition." Annu Rev Nutr **17**: 159-83.

Nickerson, D. A., V. O. Tobe and S. L. Taylor (1997). "PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing." Nucleic Acids Res **25**(14): 2745-51.

Nilsson, M., H. Malmgren, M. Samiotaki, M. Kwiatkowski, B. P. Chowdhary and U. Landegren (1994). "Padlock probes: circularizing oligonucleotides for localized DNA detection." Science **265**(5181): 2085-8.

Nishimura, T., A. Hattori and K. Takahashi (1999). "Structural changes in intramuscular connective tissue during the fattening of Japanese black cattle: effect of marbling on beef tenderization." J Anim Sci **77**(1): 93-104.

Nonneman, D. and G. A. Rohrer (2002). "Linkage mapping of porcine DGAT1 to a region of chromosome 4 that contains QTL for growth and fatness." Anim Genet **33**(6): 472-3.

Nykiforuk, C. L., T. L. Furukawa-Stoffer, P. W. Huff, M. Sarna, A. Laroche, M. M. Moloney and R. J. Weselake (2002). "Characterization of cDNAs encoding diacylglycerol acyltransferase from cultures of Brassica napus and sucrose-mediated induction of enzyme biosynthesis." Biochim Biophys Acta **1580**(2-3): 95-109.

O'Brien, S. J., J. F. Eisenberg, M. Miyamoto, S. B. Hedges, S. Kumar, D. E. Wilson, M. Menotti-Raymond, W. J. Murphy, W. G. Nash, L. A. Lyons, J. C. Menninger, R. Stanyon, J. Wienberg, N. G. Copeland, N. A. Jenkins, J. Gellin, M. Yerle, L. Andersson, J. Womack, T. Broad*, et al.* (1999). "Genome maps 10. Comparative genomics. Mammalian radiations. Wall chart." Science **286**(5439): 463-78.

Oelkers, P., A. Behari, D. Cromley, J. T. Billheimer and S. L. Sturley (1998). "Characterization of two human genes encoding acyl coenzyme A:cholesterol acyltransferase-related enzymes." J Biol Chem **273**(41): 26765-71.

Oelkers, P., D. Cromley, M. Padamsee, J. T. Billheimer and S. L. Sturley (2002). "The DGA1 gene determines a second triglyceride synthetic pathway in yeast." J Biol Chem **277**(11): 8877-81.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono and M. Kanehisa (1999). "KEGG: Kyoto Encyclopedia of Genes and Genomes." Nucleic Acids Res **27**(1): 29-34.

Ohno, S. (1970). Evolution by Gene Duplication. Heidelberg, Springer-Verlag.

Perrin, F. (1926). "Polarization de la lumiere de fluorescence. Vie moyenne de molecules dans l'etat excite." J. Phys. Radium **7**: 390-401.

Pinkel, D., T. Straume and J. W. Gray (1986). "Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization." Proc Natl Acad Sci U S A **83**(9): 2934-8.

Pritchard, J. K. and N. A. Rosenberg (1999). "Use of unlinked genetic markers to detect population stratification in association studies." Am J Hum Genet **65**(1): 220-8.

Pruitt, K. D. and D. R. Maglott (2001). "RefSeq and LocusLink: NCBI gene-centered resources." Nucleic Acids Res **29**(1): 137-40.

Pyrosequencing http://www.pyrosequencing.com/pages/technology.html. **2002**.

Rattink, A. P., D. J. De Koning, M. Faivre, B. Harlizius, J. A. van Arendonk and M. A. Groenen (2000). "Fine mapping and imprinting analysis for fatness trait QTLs in pigs." Mamm Genome **11**(8): 656-61.

Reese, M. G. and F. H. Eeckman (1995). Novel Neural Network Algorithms for Improved Eukaryotic Promoter Site Recognition. The Seventh International Genome Sequencing and Analysis Conference, Hilton Head Island, South Carolina.

Reese, M. G., N. L. Harris and F. H. Eeckman (1996). Large Scale Sequencing Specific Neural Networks for Promoter and Splice Site Recognition. Biocomputing: Proceedings of the 1996 Pacific Symposium, Singapore, World Scientific Publishing Co, Singapore.

Reisman, D., M. Greenberg and V. Rotter (1988). "Human p53 oncogene contains one promoter upstream of exon 1 and a second, stronger promoter within intron 1." Proc Natl Acad Sci U S A **85**(14): 5146-50.

Riquet, J., W. Coppieters, N. Cambisano, J. J. Arranz, P. Berzi, S. K. Davis, B. Grisart, F. Farnir, L. Karim, M. Mni, P. Simon, J. F. Taylor, P. Vanmanshoven, D. Wagenaar, J. E. Womack and M. Georges (1999). "Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle." Proc Natl Acad Sci U S A **96**(16): 9252-7.

Robinson, D. R., Y. M. Wu and S. F. Lin (2000). "The protein tyrosine kinase family of the human genome." Oncogene **19**(49): 5548-57.

Rohlfs, E. M., D. S. Louie and S. H. Zeisel (1993). "Lipid synthesis and secretion by primary cultures of rat mammary epithelial cells." J Cell Physiol **157**(3): 469-80.

Rohrer, G. A. and J. W. Keele (1998). "Identification of quantitative trait loci affecting carcass composition in swine: I. Fat deposition traits." J Anim Sci **76**(9): 2247-54.

Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlen and P. Nyren (1996). "Real-time DNA sequencing using detection of pyrophosphate release." Anal Biochem **242**(1): 84-9.

Rozen, S. and H. J. Skaletsky (1998). Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.htm.

Samiotaki, M., N. A. Balatsos, N. Courtis and C. M. Tsiapalis (2000). "Assignment of the 100-kDa subunit of cleavage and polyadenylation specificity factor (CPSF2) to human chromosome 14q31.3 by radiation hybrid mapping." Cytogenet Cell Genet **90**(3-4): 328-9.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-7.

Sham, P., J. S. Bader, I. Craig, M. O'Donovan and M. Owen (2002). "DNA Pooling: a tool for large-scale association studies." Nat Rev Genet **3**(11): 862-71.

Shen, S. H., J. L. Slightom and O. Smithies (1981). "A history of the human fetal globin gene duplication." Cell **26**(2 Pt 2): 191-203.

Shennan, D. B. and M. Peaker (2000). "Transport of milk constituents by the mammary gland." Physiol Rev **80**(3): 925-51.

Skrabanek, L. and K. H. Wolfe (1998). "Eukaryote genome duplication - where's the evidence?" Curr Opin Genet Dev **8**(6): 694-700.

Smith, S. (1994). "The animal fatty acid synthase: one gene, one polypeptide, seven enzymes." FASEB J. **8**(15): 1248-1259.

Smith, S. J., S. Cases, D. R. Jensen, H. C. Chen, E. Sande, B. Tow, D. A. Sanan, J. Raber, R. H. Eckel and R. V. Farese, Jr. (2000). "Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat." Nat Genet **25**(1): 87-90.

Solinas-Toldo, S., C. Lengauer and R. Fries (1995). "Comparative genome map of human and cattle." Genomics **27**(3): 489-96.

Stanyon, R., F. Yang, P. Cavagna, P. C. O'Brien, M. Bagga, M. A. Ferguson-Smith and J. Wienberg (1999). "Reciprocal chromosome painting shows that genomic rearrangement between rat and mouse proceeds ten times faster than between humans and cats." Cytogenet Cell Genet **84**(3-4): 150-5.

Sturtevant, A. H. (1913). "The linear arrangement of six sex-linked factors in Drosophila, as shown by

their mode of association." J. Exp. Zool. **14**: 43-59.

Taylor, R. G., G. H. Geesink, V. F. Thompson, M. Koohmaraie and D. E. Goll (1995). "Is Z-disk degradation responsible for postmortem tenderization?" J Anim Sci **73**(5): 1351-67.

Thaller, G., W. Krämer, A. Winter, B. Kaupe, G. Erhardt and R. Fries (in press-a). "Effects of DGAT1 variants on milk production traits in German cattle breeds." Journal of Animal Science.

Thaller, G., C. Kühn, A. Winter, G. Ewald, O. Bellmann, J. Wegner, H. Zühlke and R. Fries (in press-b). "DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle." Animal Genetics.

Tyagi, S., D. P. Bratu and F. R. Kramer (1998). "Multicolor molecular beacons for allele discrimination." Nat Biotechnol **16**(1): 49-53.

Tyagi, S. and F. R. Kramer (1996). "Molecular beacons: probes that fluoresce upon hybridization." Nat Biotechnol **14**(3): 303-8.

Vaiman, D., I. Bahri-Darwich, D. Mercier, M. Yerle, A. Eggen, H. Leveziel, G. Guerin, J. Gellin and E. P. Cribiu (1993). "Mapping of new bovine microsatellites on cattle chromosome 15 with somatic cell hybrids, linkage analysis, and fluorescence in situ hybridization." Mamm Genome **4**(11): 676-9.

VanRaden, P. M. and G. R. Wiggans (1991). "Derivation, calculation, and use of national animal model information." J Dairy Sci **74**(8): 2737-46.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, *et al.* (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Walling, G. A., P. M. Visscher, L. Andersson, M. F. Rothschild, L. Wang, G. Moser, M. A. Groenen, J. P. Bidanel, S. Cepica, A. L. Archibald, H. Geldermann, D. J. de Koning, D. Milan and C. S. Haley (2000). "Combined analyses of data from quantitative trait loci

mapping studies. Chromosome 4 effects on porcine growth and fatness." Genetics **155**(3): 1369-78.

Warren, W., T. P. Smith, C. E. Rexroad, 3rd, S. C. Fahrenkrug, T. Allison, C. L. Shu, J. Catanese and P. J. de Jong (2000). "Construction and characterization of a new bovine bacterial artificial chromosome library with 10 genome-equivalent coverage." Mamm Genome **11**(8): 662-3.

Waterman, I. J., N. T. Price and V. A. Zammit (2002). "Distinct ontogenic patterns of overt and latent DGAT activities of rat liver microsomes." J. Lipid Res. **43**(9): 1555-1562.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork*, et al.* (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Weber, J. L. and P. E. May (1989). "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction." Am J Hum Genet **44**(3): 388-96.

Weiss, S. B., E. P. Kennedy and J. Y. Kiyasu (1960). "The enzymatic synthesis of triglycerides." J Biol Chem(235): 40-44.

Weller, J. I., Y. Kashi and M. Soller (1990). "Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle." J Dairy Sci **73**(9): 2525-37.

Werle, E., C. Schneider, M. Renner, M. Volker and W. Fiehn (1994). "Convenient single-step, one tube purification of PCR products for direct sequencing." Nucleic Acids Res **22**(20): 4354-5.

Whitcombe, D., J. Theaker, S. P. Guy, T. Brown and S. Little (1999). "Detection of PCR products using self-probing amplicons and fluorescence." Nat Biotechnol **17**(8): 804-7.

Winter, A., W. Kramer, F. A. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G. Thaller and R. Fries (2002). "Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content." Proc Natl Acad Sci U S A **20**: 20.

Wolfe, K. H. (2001). "Yesterday's polyploids and the mystery of diploidization." Nat Rev Genet **2**(5): 333-41.

Wolfsberg, T. G. and D. Landsman (1997). "A comparison of expressed sequence tags (ESTs) to human genomic sequences." Nucleic Acids Res **25**(8): 1626-32.

Womack, J. E., J. S. Johnson, E. K. Owens, C. E. Rexroad, 3rd, J. Schlapfer and Y. P. Yang (1997). "A whole-genome radiation hybrid panel for bovine gene mapping." Mamm Genome **8**(11): 854-6.

Wyman, A. R. and R. White (1980). "A highly polymorphic locus in human DNA." Proc Natl Acad Sci U S A **77**(11): 6754-8.

Yen, C. L. and R. V. Farese, Jr. (2003). "MGAT2, a monoacylglycerol acyltransferase expressed in the small intestine." J Biol Chem **5**: 5.

Yen, C. L., S. J. Stone, S. Cases, P. Zhou and R. V. Farese, Jr. (2002). "Identification of a gene encoding MGAT1, a monoacylglycerol acyltransferase." Proc Natl Acad Sci U S A **99**(13): 8512-7.

Yu, C., J. Chen, S. Lin, J. Liu, C. C. Chang and T. Y. Chang (1999). "Human acyl-CoA:cholesterol acyltransferase-1 is a homotetrameric enzyme in intact cells and in vitro." J Biol Chem **274**(51): 36139-45.

Yu, Y. H., Y. Zhang, P. Oelkers, S. L. Sturley, D. J. Rader and H. N. Ginsberg (2002). "Post-transcriptional control of the expression and function of acyl- CoA:diacylglycerol acyltransferase-1 in mouse adipocytes." J Biol Chem **28**: 28.

Zhang, J., Y. P. Zhang and H. F. Rosenberg (2002). "Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey." Nat Genet **30**(4): 411-5.

Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller and M. D. Bishop (1998a). "Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree." Genetics **149**(4): 1959-73.

Zhang, Y., S. Koushik, R. Dai and N. F. Mivechi (1998b). "Structural organization and promoter analysis of murine heat shock transcription factor-1 gene." J Biol Chem **273**(49): 32514-21.

Zhu, B., J. A. Smith, S. M. Tracey, B. A. Konfortov, K. Welzel, L. C. Schalkwyk, H. Lehrach, S. Kollers, J. Masabanda, J. Buitkamp, R. Fries, J. L. Williams and J. R. Miller (1999). "A 5x genome coverage bovine BAC library: production, characterization, and distribution." Mamm Genome **10**(7): 706-9.

Zou, J., Y. Wei, C. Jako, A. Kumar, G. Selvaraj and D. C. Taylor (1999). "The Arabidopsis thaliana TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene." Plant J **19**(6): 645-53.

9 Appendices

9.1 Buffer

TE		(10 mM Tris-Cl, 1 mM EDTA, pH 8)
TAE	1x	(0.04 M Tris-acetate, 0.001 M EDTA)
TBE	1x	(0.09 M Tris-borate, 0.002 M EDTA)
SSC	20x	(3 M NaCl, 0.3 M Na-Citrate)

9.2 Composition of DNA pools

for German Holstein (HF32+, HF32-), German Simmental (FV32+, FV32-) and German Brown (BV20+, BV20-); bulls were selected according their breeding values for milk fat percentage (BVF).

HF32+			HF32-		FV32+			FV32-		BV20+			BV20-		
Lab no.	Herdbook no.	BVF	Lab no.	Herdbook no.	BVF	Lab no.	Herdbook no.	BVF	Lab no.	Herdbook no.	BVF	Lab no.	Herdbook no.	BVF	Lab no.	Herdbook no.	BVF
1143	840173	0.48	1082	629862	-0.92	901	194100	0.83	1019	45432	-0.31	909	78780	0.73	1004	78225	-0.22
1146	840129	0.48	1083	396172	-0.91	902	195260	0.78	1021	53381	-0.31	929	79030	0.63	1006	78200	-0.22
1147	800220	0.49	1086	136618	-0.87	903	50223	0.77	1023	178075	-0.31	943	340530	0.54	1007	348215	-0.24
1148	800257	0.50	1087	627893	-0.85	906	39910	0.75	1025	191283	-0.31	951	79195	0.50	1008	72625	-0.24
1150	560446	0.52	1089	830349	-0.84	907	169044	0.74	1026	39733	-0.31	952	79115	0.47	1009	348607	-0.24
1152	282913	0.53	1091	629367	-0.83	910	178317	0.71	1029	68130	-0.33	953	348544	0.46	1011	72680	-0.27
1153	395379	0.53	1092	840200	-0.83	911	165011	0.70	1032	27876	-0.34	954	78475	0.45	1012	72470	-0.27
1155	395053	0.55	1093	395123	-0.81	912	7889	0.70	1033	21971	-0.34	955	348105	0.44	1014	72930	-0.29
1156	800231	0.55	1094	398133	-0.81	1066	1146	0.70	1034	22043	-0.34	956	349447	0.42	1015	78090	-0.30
1157	394790	0.55	1096	397250	-0.80	913	34380	0.69	1035	60552	-0.34	957	78635	0.40	1017	78470	-0.31
1158	800274	0.56	1097	413509	-0.80	914	187217	0.69	1036	68030	-0.34	959	77888	0.38	1018	78840	-0.31
1159	136601	0.57	1098	810338	-0.79	916	60535	0.69	1038	22093	-0.35	961	348247	0.37	1024	78860	-0.31
1160	395924	0.57	1099	830319	-0.79	917	60250	0.69	1039	184256	-0.35	962	348591	0.37	1027	78560	-0.30
1161	396398	0.57	1100	397582	-0.77	918	54474	0.69	1040	187114	-0.35	964	349569	0.35	1028	72490	-0.30
1162	800258	0.57	1102	397223	-0.77	919	172162	0.68	1043	184280	-0.36	965	72695	0.34	1030	85550	-0.30
1164	399676	0.58	1103	810460	-0.76	920	184506	0.68	1046	53487	-0.37	966	340573	0.33	1037	78015	-0.40
1166	840172	0.59	1104	395139	-0.76	921	169042	0.67	1047	53493	-0.37	967	340015	0.32	1042	78695	-0.40
1167	134413	0.60	1105	398783	-0.75	922	172174	0.66	1048	68175	-0.37	968	78980	0.31	1044	348104	-0.40
1168	810325	0.60	1106	629399	-0.75	923	178308	0.66	1049	53607	-0.37	971	79080	0.31	1045	340010	-0.40
1169	665532	0.61	1108	561217	-0.74	924	165010	0.65	1050	53625	-0.38	972	78880	0.29	1057	78155	-0.40
1171	742512	0.62	1109	742898	-0.74	925	22153	0.65	1051	176156	-0.38						
1172	800195	0.63	1110	393043	-0.74	926	645073	0.65	1053	27848	-0.39						
1173	664186	0.67	1114	414239	-0.73	927	60527	0.64	1054	68040	-0.41						
1174	810294	0.67	1115	398749	-0.73	930	34554	0.63	1055	53517	-0.41						
1176	395369	0.68	1120	504355	-0.72	932	187049	0.62	1056	7787	-0.43						
1177	396406	0.69	1126	398440	-0.71	933	21784	0.62	1058	176009	-0.45						
1178	395383	0.69	1127	136785	-0.70	935	187138	0.59	1060	39860	-0.46						
1180	135515	0.72	1128	398158	-0.70	936	175061	0.59	1061	53460	-0.46						
1181	134428	0.77	1130	398010	-0.70	937	191053	0.59	1062	53293	-0.46						
1184	395051	0.83	1131	397550	-0.69	939	53535	0.58	1063	27847	-0.47						
1185	394435	0.84	1136	398405	-0.69	940	191045	0.57	1064	68195	-0.51						
1186	135500	1.08	1139	252100	-0.68	942	50246	0.56	1065	27851	-0.51						

9.3 DNA samples used for polymorphism detection

Locus	*DGAT1*				Genes and STS markers neighboring *DGAT1*		*DGAT2* and *DC2*		*DC5*	
Regions	exons, small introns		large introns, down-stream and up-stream sequence		parts of the genes		exons, small introns		exons, small introns	
	animal id	herdbook number	animal id	herdbook number	animal id	herdbook number	animal id	herdbook number	animal id	herdbook number
Pools	HF32+ [a]		HF32+ [a]		HF32+ [a]		HF32+ [a]		HF32+ [a]	
	HF32- [a]		HF32- [a]		HF32- [a]		HF32- [a]		HF32- [a]	
	FV32+ [a]		FV32+ [a]		FV32+ [a]		FV32+ [a]		FV32+ [a]	
	FV32- [a]		FV32- [a]		FV32- [a]		FV32- [a]		FV32- [a]	
	BV20+ [a]		BV20+ [a]				BV20+ [a]		BV20+ [a]	
	BV20- [a]		BV20- [a]				BV20- [a]		BV20- [a]	
	SB pool [b]									
German Holstein	SB26	790580			1091	629367	1176	395369	1176	395369
	SB37	102430			1180	135515	1170	283062	1170	283062
	SB45	252006					1103	810460	1091	629367
							SB26	790580	1180	135515
							SB37	102430	1087	627893
							SB45	252006		
German Simmental	FV19	7620			899	49704	FV19	7620	921	169042
	FV27	25100			361	175075	FV27	25100	923	178308
	FV28	50148					FV28	50148	924	165010
							924	165010	925	22153
							925	22153	1019	45432
							1019	45432		
Other breeds	AN1	(Angus)								
	KE2	(Kerry)								
	SA4	(Sahival)								
	HA8	(Hariana)								

[a]see Appendix 9.2;

[b]SB pool is composed of 10 German Holstein animals with the herdbook numbers: 102399, 790121, 790223, 790253, 790510, 790361, 790062, 790183, 102350, 102315;

9.4 Primers used for direct sequencing BAC DNA

Primers for direct sequencing of bovine DGAT1

Loci	Primer no.		Position	Primer sequence [5'-3']
DGAT1	1738	R	5'end	TGATGCCTACCTAAGCTCTACC
DGAT1	1739	R	5'end	TTTAGGGTCTGAGCCACCAG
DGAT1	1728	R	5'end	TCCCGACTCTTTGTGACTCC
DGAT1	1734	R	5'end	TGGATTGCAAAGTCCTGTCC
DGAT1	1717	R	5'end	CAGGAAGGGCCTCTGTACC
DGAT1	1716	R	5'end	ACAGCTGGAGTGAGGACACC
DGAT1	1710	R	5'end	CCCTCAGCGCTAGGACTC
DGAT1	1709	R	5'end	TGTCTTGGAGTAGCGTGTGG
DGAT1	1706	R	5'end	AGGCCCCCACAGTAGACAAG
DGAT1	1705	R	5'end	ACGGTCGTGCTCTGTGAAC
DGAT1	1699	R	5'end	CCCTTGTCCCGCTCTATAAAC
DGAT1	1698	R	5'end	CGCGCATACCTTTGTAGTCC
DGAT1	1697	R	5'end	CGCCTCTACTACGCCACTG
DGAT1	1632	F	exon 1	GCCACTGGGAGCTGAGG
DGAT1	1681	R	intron 1	ACAGCTGTGCACCAAGGTC
DGAT1	1680	F	intron 1	TGGCTGCTCTAGGGTCAAAG
DGAT1	1693	F	intron 1	ATCTTCACTGGGTGCTGTGG
DGAT1	1694	F	intron 1	CTGCTCCTGTCCTGTTGATG
DGAT1	1696	R	intron 1	AGCCACCTCATGCTACAACC
DGAT1	1695	R	intron 1	GCCCTCTTCTTCATGACTCTG
DGAT1	1679	R	intron 1	GGCCACCATTCAAACCAC
DGAT1	1602	F	exon 2	GAATTGGTGTGTGGTGATGC
DGAT1	1675	R	intron 2	GGTAGGGTCCCAGGGTACG
DGAT1	1673	F	intron 2	GCCACACTCTGCAGGACTC
DGAT1	1674	R	intron 2	CAGTCCTGCTCCCTCCAG
DGAT1	1671	R	intron 2	TGACAGGCTCAGAGATGCAG
DGAT1	1660	R	intron 2	AGCCCCAGTGAAGTCCAAG
DGAT1	1634	R	exon 3	TAGAAATAACCGTGCGTTGC
DGAT1	1633	R	exon 4	ACCTGGATGGGGTCCAC
DGAT1	1593	F	3'end	GTGGGTGTTGGACTGCTTTG
DGAT1	1711	F	3'end	CCATGCTCTGGAAACCCTAC
DGAT1	1729	F	3'end	TCAGCAGGTAGTTGGGTGTG
DGAT1	1730	F	3'end	GAAACCCTGAGGCTGTGC
DGAT1	1732	F	3'end	CCCACCTGGTCCTCTAGTGC
DGAT1	1733	F	3'end	CCAGGAGGCTCCAGTGTG
DGAT1	1737	F	3'end	GTTCTGAGCCCGTCAGCAG
DGAT1	1739	F	3'end	TTTAGGGTCTGAGCCACCAG

Primers for direct sequencing of bovine DGAT2, DC2 and DC5

Loci	Primer no.		Position	Primer sequence [5'-3']
DGAT2	1901	F	exon 8	GCGAGCCCATTACCATCC
DGAT2	1903	F	exon 7	GGTCGAGGCCTCTTCTCCT
DGAT2	2076	R	exon 1	GTAGGCGGCTATGAGGGTCT
DGAT2	2077	F	exon 1	CGACCTGTACTGGCTTCGTC
DGAT2	2078	R	exon 2	GCCCTATTGAGCCAAGTGAC
DGAT2	2082	R	exon 5	ACTTCTGTGGCCTCTGTGCT
DGAT2	2093	F	intron 6	AGCAGCTCCTTGGCTCCT
DGAT2	2099	R	intron 6	CCCTCAGGGCTGTACAAGAGT
DGAT2	2508	R	intron 1	TTCTCATTCCTCACCTCTACCC
DGAT2	2510	R	5'end	GGACTCTTGCTCCTCACAGC
DGAT2	2511	R	intron 1	TGTGCCAGCACACTCCTG
DGAT2	2513	F	intron 1	CTGGTGTGGGGTACTCTGC
DGAT2	2515	F	intron 3	TCCCCCACTCCTACTCCTTC
DGAT2	2516	R	exon 8	ACGTACATGGCGTGGTACAG
DGAT2	2517	F	exon 8	AAGCATCATGGGTGTCTGTG
DGAT2	2518	R	intron 4	GACTGCTCTAAAAGCCCAGTG
DGAT2	2553	R	5'end	GGGCTCCTAAATCCCTCAAG
DGAT2	2554	R	intron 7	AGGGGGATCCTTCCTTACAG
DGAT2	2601	F	intron 1	CGGACAGGCTGACATCTG
DGAT2	2602	R	intron 3	AAGGGGCAGTACCCACAAC
DC2	1904	F	intron 5	ACAATCCAGCATGTGCAGAG
DC2	1905	R	exon 6	CTGGAATACCATACTTCCCTTTG
DC2	2073	F	exon 3	GGATTTGGATCCGAGTCACA
DC2	2074	R	intron 5	CTCTGCACATGCTGGATTGT
DC2	2075	R	exon 6	CTGGAATACCATACTTCCCTTTG
DC2	2458	F	exon 1	AGTTTGCGCCACTCAACATC
DC2	2459	R	exon 2	CAAATGGCCCAGTTTCTGAC
DC2	2519	R	intron 5	CTTACTGCGGTGATCCTTTTAC
DC2	2560	R	intron 4	TAGCTTCCCTGACCCAGTTG
DC2	2561	F	intron 4	GAAAATGCTTACTCTTCCTCCTTG
DC2	2604	R	intron 1	GGGTCAAGGTGTTGTTGTTG
DC2	2606	R	intron 3	CCACTGCTCATCAGATATTCC
DC2	2607	F	intron 3	ATAGGCTGCAGTCCATGAGG
DC2	2608	F	intron 6	TCTTAGAAGTCATGCAAGAGAGC
DC2	2631	R	intron 3	AAACAAAAGGCGTTAACTAATTGC
DC2	2632	F	intron 5	GAAGCCCATTCACACTGTTG
DC2	2773	R	exon 1	ACAGCAACCAATGCAGCAC
DC2	2774	F	exon 6	GAAGAACACAAAGGGAAGTATGG
DC5	1906	F	exon 3 (p)	CCCCCATCTGATGATGCT
DC5	2530	F	exon 2 (p)	ATCCTGTATGCGACCTGGTG
DC5	2531	R	exon 2	GGGAAATAGTCCTTCATGTACTTCC
DC5	2532	R	intron 3	AGAACTGTTGACAGCCCTCTG
DC5	2533	F	intron 4	CATTCCCAAATAGCCAGAGAAG
DC5	2555	R	intron 2	CACAACCTCCCCCATTATTC
DC5	2559	F	intron 4	CACAGTGCAAGGCTGTGG
DC5	2610	F	intron 4	GAAGATTCTGTGTTGGTGATCG
DC5	2638	R	intron 1	TCAGAGGTATGAAGCACAAGC
DC5	2639	F	intron 2	TTGGTTTCTCTCTGGGTATGG
DC5	2664	F	3'end	TCTCCTTTGGGGAGAATGAC
DC5	2706	F	exon 1	TTGTCCGTGCCATGGG
DC5	2777	R	intron 1	AACCCCAGTCTAGTAGGGTCTG
DC5	2778	F	exon 4	AGAACTCCCCTGGCTCCTG
DC5	2803	R	exon 1	CGGTTGATTCCACAGGTTTC

(p) porcine

Primers for direct sequencing of bovine loci neighboring DGAT1

Locus	Primer no.		Primer sequence [5'-3']
T7 (vector pBACe3.6)	1658		CCGCTAATACGACTCACTATAGGG
SP6 (vector pBACe3.6)	1659		TTTGCGATCTGCCGTTTC
56F1-T7	1721	F	TGAGGCCCTGATCTCTCAAC
269H7-SP6	1724	F	CCTGCTTGGTTTTCTTTTCC
240A1-T7	1740	F	TCCTTCCGATGAACATTCAAG
428P15-SP6	2467	F	TGATGTTTGTGATATCGGTCAAC
428P15-SP6	2488	F	GCCCTGAGTCTCCAGTGAGC
56F1-T7	2493	F	AAGGGGGATGTCCCTGAAT
428F15-SP6	2506	R	ACAGGCCCCAGAACACAATA
428F15-SP6	2507	F	GTGGAGGGTGCTGGTAGTCA
100P18-SP6	2670	F	GGCTGCTGCCGTGGAC
3O1-T7	2673	F	AACTCACATGATGCAACCTCAG
3O1-SP6	2676	F	GGGATGATCGGCTGTGTTG
111I3-T7	2682	F	TGGGGAAACAATGGAAATAGTGA
156I10-SP6	2689	F	CCTCGCCTTTGAGGAAGC
557K4-T7	2696	F	CCTTCTCCTGCCTTCAATCTT
334E6-T7	2736	F	CAGCTGTACTAATCCACACATGATGA
100P18-T7	2743	F	AATGTACACGCATGCACCAG
56F1-T7	2422	R	CCCCTGGATCTCTCTCCTGA
240A1-SP6	2494	R	CCGGCTTCTGATCACTCCT
KIAA1833	1688	F	GGCAGCAGTGTCTGTGTGTT
KIAA1833	1729	F	TCAGCAGGTAGTTGGGTGTG
KIAA1833	2423	F	CCCATCTGCCCTTGACTCTAC
KIAA1833	2424	R	GTGCTCCTCTTGGGTCTCCT
KIAA1833	2425	F	CCCGAGATTGTGAGTGTGCT
RECQL4	2431	R	GACAGACACGCAAGTAACAAGG
GPT	2443	R	CATCTCCGTGAGCACCTTCT
PPP1R16A	2447	R	CCTCATCCTCTTCCGATGG
PPP1R16A	2448	F	CGTGCCCAGTGCATGTC
FOXH1	2451	R	GCAGGGGAAGCAGGAAAC
FOXH1	2452	F	CTGCGGCTGCAGAACAC
KIAA0496	2455	R	TGATCTCACAACGACAGTTGG
KIAA0496	2456	F	ACGAGCTGATGGAGATCCTG
Rrp41	2466	R	TGTTGCCCTGTTCGATGTAG
KIAA1833	2469	F	CAACATGCTGCAGGAGAAGG
KIAA1833	2470	R	CGTCCATATTGGACAGCTAGG
KIAA1833	2473	F	ACTGCACGAGGTCGCATC
KIAA1833	2494	R	CCGGCTTCTGATCACTCCT
FBXL6	2500	F	GGCTGACTCTAGCCAAGGAA
RECQL4	2505	F	CCCACGGTGAGGTGGAG
FBXL6	2670	R	GGCTGCTGCCGTGGAC

PCR primers for bovine DGAT1

Locus	Forward primer			Reverse primer			Product size [bp]	T [°C]	Additives
	No.	Position	Sequence [5'-3']	No.	Position	Sequence [5'-3']			
DGAT1	1755	5'end	AGAAATGGGAAGTGCAGACC	1738	5'end	TGATGCCTACCTAAGCTCTACC	550	60	
DGAT1	1754	5'end	CAGGGTGGGATCACCTGAG	1734	5'end	TGGATTGCAAAGTCCTGTCC	641	60	
DGAT1	1753	5'end	GGTGGATGACGGGTAGAGG	1716	5'end	ACAGCTGGAGTGAGGACACC	735	60	1x Q-Solution
DGAT1	1881	5'end	6-Fam-TCAGGATCCAGAGGTACCAG	1874	5'end	GGGGTCCAAGGTTGATACAG	147	60	1x Q-Solution
DGAT1	1721	5'end	TGAGGCCCTGATCTCTCAAC	1709	5'end	TGTCTTGGAGTAGCGTGTGG	641	60	
DGAT1	1722	5'end	AAGGGGATACTCCTGATCCAC	1706	5'end	AGGCCCCCACAGTAGACAAG	713	60	
DGAT1	1723	5'end	TCTGCAGATGAAGGCAGAAG	1698	5'end	CGCGCATACCTTTGTAGTCC	521	60	
DGAT1	1701	5'end	CGCGTTGGGTGTCAGC	1681	intron 1	ACAGCTGTGCACCAAGGTC	812	60	5% DMSO
DGAT1	1866	intron 1	GACACCTGGTGCGTCCTTC	1867	intron 1	GAGGGGAGCATTTCCCAATC	697	60	
DGAT1	1868	intron 1	TACCCCCACAGACTGTCCTC	1679	intron 1	GGCCACCATTCAAACCAC	742	60	
DGAT1	1702	intron 1	TGGCTTCTGCAGTGGACTC	1675	intron 2	GGTAGGGTCCCAGGGTACG	589	64	
DGAT1	1673	intron 2	GCCACACTCTGCAGGACTC	1671	intron 2	TGACAGGCTCAGAGATGCAG	736	63.5	
DGAT1	1672	intron 2	TGGTAAGCTGGCTGGTTAGG	1634	intron 2	TAGAAATAACCGTGCGTTGC	822	60	5% DMSO
DGAT1	1670	intron 2	GTGGCTGACAGCGTTATGTC	1676	intron 4	GTTCAGGCCCAGATCAGC	309	60	
DGAT1	1614	exon 4	TATGGCATCCTGGTGGAC	1617	exon 6	AGTGATAGACTCGAGGAGAAAGG	546	60	
DGAT1	1616	exon 6	GGAGCTCTGACGGAGCAG	1635	exon 7	GTTGACGTCCCGGTAGGAG	267	60	
DGAT1	1532	exon 7	GCACCATCCTCTTCCTCAAG	1636	exon 9	GGAAGCGCTTTCGGATG	411	60	5% DMSO
DGAT1	1618	exon 9	CCCTGTGCTACGAGCTCAAC	1678	intron 11	CACAGCTGGCTCCCTCAG	372	60	1% formamide
DGAT1	1638	exon 11	GCCATCCAGAACTCCATGA	1640	exon 14	CAGGGATGTTCCAGTTCTGC	469	60	
DGAT1	1599	exon 16	CGAGTACCTGGTGAGCATCC	1601	3'UTR	TGTGCACAGCACTTTATTGAC	565	60	
DGAT1	1711	3'end	CCATGCTCTGGAAACCCTAC	1718	3'end	GCGGCAGAGCCAGTAGAG	658	60	
DGAT1	1729	3'end	TCAGCAGGTAGTTGGGTGTG	1756	3'end	CTCCCTGTCTGTTCCTCCTG	763	60	1x Q-Solution

PCR primers for bovine DGAT2 and DC2

Locus	Forward primer			Reverse primer			Product	T	Additives
	No.	Position	Sequence [5'-3']	No.	Position	Sequence [5'-3']	size [bp]	[°C]	
DGAT2	2619	5'end	GACGGAAAGGTAGAGGCAAC	2620	intron 1	GGCATCATGGCTCCTTAATC	751	60	5% DMSO
DGAT2	1895	exon3	GCCTGCAGCGTCATCC	1896	exon4	TGGGAAAGTAGTCTCGAAAGTAGC	538	60	
DGAT2	2623	intron 4	ATGCTCTGTCCCTTCTCTGG	2624	intron 5	ATGAACAGTTCTCTCGCTCAG	758	60	
DGAT2	1897	exon 5	CAGGAACTACATCTTTGGGTACCA	1898	exon 6	ATTGCCACTCCCATTCTTTG	807	60	
DGAT2	1899	exon 6	TGAACCGGGACACCATAGAC	1900	exon 7	TGGTGATGGGCTTGGAGTAG	1180	60	
DGAT2	2093	intron 6	AGCAGCTCCTTGGCTCCT	2099	intron 6	CCCTCAGGGCTGTACAAGAGT	634	60	
DGAT2	2626	intron 7	GAGAGGAAGCAGGAGGATGAC	2628	intron 8	GCTTGAAATATTCCCTTTCTCC	823	60	
DGAT2	2619	5'end	GACGGAAAGGTAGAGGCAAC	2508	intron 1	TTCTCATTCCTCACCTCTACCC	15 kb	61	LR-PCR
DGAT2	2601	intron 1	CGGACAGGCTGACATCTG	2602	intron 3	AAGGGGCAGTACCCACAAC	6.8 kb	61	LR-PCR
DGAT2	2515	intron 3	TCCCCCACTCCTACTCCTTC	2518	intron 4	GACTGCTCTAAAAGCCCAGTG	6 kb	61	LR-PCR
DGAT2	2625	intron 6	AGACCCTGGCCTCTCTCTTC	2627	intron 7	TTTGTCATCCTCCTGCTTCC	1.7 kb	61	LR-PCR
DC2	2805	5'end	GGAGGAAGAGGTGGGTTTAAGG	2629	intron 1	AGTCGTTCGCTACGAGCTTC	784	60	1x Q-Solution
DC2	2660	intron 1	ACGAGATGGTTAGATTGCATTAC	2630	intron 2	ATGTTGCCCGATTAGTTTCC	891	60	5% DMSO
DC2	2605	intron 2	TCTGAATGGGTCAGAAACTGG	2630	intron 2	ATGTTGCCCGATTAGTTTCC	574	60	
DC2	2703	intron 2	TGAAGAAACTGCCTGCAATG	2704	intron 3	ATGGGATTTTCCAGCCAAG	1172	60	
DC2	2661	intron 3	CACCTTCATCAAGAGGCTTTTTAG	2560	intron 4	TAGCTTCCCTGACCCAGTTG	616	60	5% DMSO
DC2	2521	exon 4 (h)	CTCTGTTCATCCGCCAGC	2522	exon 5 (h)	TTTAAACAGTTCATTTTCACCAAAAG	647	60	5% DMSO
DC2	2561	intron 4	GAAAATGCTTACTCTTCCTCCTTG	2662	intron 5	TGCCACTTCTAAAATTGGATTG	483	60	5% DMSO
DC2	2633	intron 5	AGTGGAACCCTTGTGATTAGG	2659	exon6	TTGCAAATTTGATTCCATTGTC	736	60	5% DMSO
DC2	2633	intron 5	AGTGGAACCCTTGTGATTAGG	2806	3'end	GGGTTTCCATATGGTGATTCAT	1263	60	
DC2	1904	intron 5	ACAATCCAGCATGTGCAGAG	1905	exon6	CTGGAATACCATACTTCCCTTTG	347	60	
DC2	2524	intron 1	CATCCAAATGGAACCACGAC	2604	intron 1	GGGTCAAGGTGTTGTTGTTG	11.5 kb	61	LR-PCR
DC2	2607	intron 3	ATAGGCTGCAGTCCATGAGG	2560	intron 4	TAGCTTCCCTGACCCAGTTG	7 kb	61	LR-PCR
DC2	2632	intron 5	GAAGCCCATTCACACTGTTG	2519	intron 5	CTTACTGCGGTGATCCTTTTAC	6 kb	61	LR-PCR

(h) human

PCR primers for bovine DC5

Locus	Forward primer			Reverse primer			Product	T	Additives
	No.	Position	Sequence [5'-3']	No.	Position	Sequence [5'-3']	size [bp]	[°C]	
DC5	2807	5'end	AGGCCAAGTGGTTTCAAGG	2777	intron 1	AACCCCAGTCTAGTAGGGTCTG	802	60	
DC5	2753	intron 2	GGACTGTGCTGCTAGTGCTG	2555	intron 3	CACAACCTCCCCCATTATTC	287	60	
DC5	2753	intron 1	GGACTGTGCTGCTAGTGCTG	2641	intron 2	GACCAGCTGCAGGGACAG	ca. 2500	66	10% DMSO
DC5	2556	intron 2	GAGAGGGGGAGCAAGTTACC	2558	intron 4	ATGACCCTTGACCCTATTCC	941	60	
DC5	2642	intron 3	TTCCCAACCCTGGGTAAC	2643	intron 4	CCATGACAGGAGAAAGAATGC	831	60	
DC5	2533	3'UTR	CATTCCCAAATAGCCAGAGAAG	2804	3'UTR	CAAGGACAGGGTGATCTTTTG	1073	60	
DC5	1906	exon3 (p)	CCCCCATCTGATGATGCT	1908	exon4 (p)	TGCTCAGGATGTGAGCAGC	422	60	

(p) porcine.

PCR primers for porcine DGAT1 and DC7

Locus	Forward primer			Reverse primer			Product	T	Additives
	No.		Sequence [5'-3']	No.		Sequence [5'-3']	size [bp]	[°C]	
DGAT1	1915	(p)	GTACGGCATCCTGGTAGACC	1916	(p)	GATGGACTCCAGCAGGAAAG	ca. 500	60	
DC7	1909	(p)	AGCCTCCAAGTGTTTGCTGT	1911	(p)	CACGGAAATGGCTCCAGAT	ca. 400	60	
DC7	1910	(p)	CTGCCCTGATCCTGGTACTC	1911	(p)	CACGGAAATGGCTCCAGAT	150	60	
DC7	1910	(p)	CTGCCCTGATCCTGGTACTC	1913	(p)	GATGTAGTCCCGGAAGATGG	400	60	
DC7	1912	(p)	GAAAACCGCAGAGCTGGAT	1913	(p)	GATGTAGTCCCGGAAGATGG	200	60	
DC7	1912	(p)	GAAAACCGCAGAGCTGGAT	1914	(p)	CCGCGACAAGAGATAGGAAG	400	60	

(p) porcine.

PCR primers for bovine genes neighboring DGAT1

Locus	Forward primer			Reverse primer			Product	T	Additives
	No.	Position	Sequence [5'-3']	No.	Position	Sequence [5'-3']	size [bp]	[°C]	
CPSF1	1975	exon 32	ACGTCATGAAGAGCATCTCG	1953	exon 37	GGTACAGGTAGCGGTTGAGC	1000	60	
CPSF1	1990	exon 1	GTACAGGTCCCCCATCAGC	1991	exon 2	TTGAGGCGGTACACGTAGAG	250	60	
CPSF1	1994	exon 7	TCCTGCCTAGCTACATCATCG	1995	exon 11	CAGCCTTGTCGAAGTGGAAG	900	60	1x Q-Solution
CPSF1	1996	exon 10	CGGCCTTCATCTCCTATGAC	1997	exon 16	ATAGCCGGAGCACACCAC	1205	64	5% DMSO
CPSF1	1998	exon 16	GAGCCAGACCTGGAGATCG	1999	exon 20	CAGGGGTGGCTTATGCAG	973	60	1x Q-Solution
CPSF1	2000	exon 20	GACCCCTACGTGGTCATCAT	2001	exon 23	ACCAGCAGCACCTCCTTG	911	61	5% DMSO
CPSF1	2002	exon 23	CGTCCTGGTGGACAGCTC	2003	exon 26	GCAGTTGATGTTGTGGAACG	741	60	1x Q-Solution
CPSF1	2004	exon 26	CACTGGCTCCTGGTGACTG	2005	exon 28	GGTGTGCTGGTGCTAGTGG	1200	60	1x Q-Solution
CYC1	2611	exon 3	TCCAGGTGTACAAGCAGGTG	2612	exon 6	GATGGTCGTGTTCTGGTTCAG	900	60	
DKFZp547F072	1958	exon 1	ACAGCAGGGGCTCATGTC	1956	exon 1	GCATCTCGGGCCCTACTTAT	400	60	
FBXL6	2696	intron1	CCTTCTCCTGCCTTCAATCTT	1960	exon 4	CGAGGACAGGACTCGCTAAC	850	61	5% DMSO
FBXL6	1961	exon 3	CCACTGGAAGTCCCAGCTAC	1962	exon 6	CAGGACCTGGAGCTGTGG	800	64	1x Q-Solution
FBXL6	1963	exon 5	ACTCCCAGACAACAGCCATC	1964	exon 8	CGCAGACTGTGACACCACTT	700	60	
FLJ11856	2436	exon 2	GCTGTGACCTTTGCCCTGT	2437	exon 3	AGGAAAACGCTCTGGGAAGT	800	60	
FLJ11856	2502	exon 3	ACCAACCCCCACCAATG	2586	exon 4	ATCAGGTAGGCCCCAAAGAA	700	60	
FOXH1	2450	exon 1	CCTCCCCACACTACCACACT	2453	exon 3	GTAGGGGCCGAGATCCTTG	1200	60	
GPT	2442	exon 2	CAACGTGTATGCCGAGAGC	2445	exon 6	AGGAGTACTCGCGGGTGAA	1200	60	1x Q-Solution
HSF1	1729	3end	TCAGCAGGTAGTTGGGTGTG	1756	3'end	CTCCCTGTCTGTTCCTCCTG	650	60	
HSF1	1967	exon 2	AGCACCCGTGCTTCCTG	1968	exon 3	CTTCATGGCCAGCAGCTT	450	60	
HSF1	1969	exon 4	GAGAACGAGGCGCTGTG	1970	exon 8	GAGAACGAGGCGCTGTG	1000	60	1x Q-Solution
HSF1	1971	exon 9	CGAGCTCAGCGACCACTT	1972	exon 10	CAGGCTGCTGTCCAGGTC	300	62	
HSF1	1965	exon 1	AGCTTCCACGTGCTGGAC	1966	exon 2	CACTTTCCTCTTGATGTTCTCG	600	60	
KIAA0014	2479	exon 5	CTCTGGACCACAGACATCTACG	2480	3'UTR	GTTCGTCATCGACCGTTCC	1000	60	
KIAA0014	2599	3'UTR	GTTGACCAAGAGCTGGAAGG	2600	3'UTR	CTGCGCCTCACAGGTAATTC	1000	60	5% DMSO
KIAA0014	2434	exon 3	GACCTCCTGCCTGGCCTA	2435	exon 4	GTGGGATGATGACGGACAG	1200	60	

PCR primers for bovine genes neighboring DGAT1 (continued)

Locus	Forward primer			Reverse primer			Product	T	Additives
	No.	Position	Sequence [5'-3']	No.	Position	Sequence [5'-3']	size [bp]	[°C]	
KIAA0124	2615	exon 13	ACCCAGGTGCTGATCCAC	2616	exon 16	AGCCCCGGAGGAGAAGAC	800	60	
KIAA0124	2776	exon 4	CATGTGGGCTATGACCTGGA	2779	exon 8	CGTGGGCACAGGTAGAGGT	1000	60	5% DMSO
KIAA0124	2780	exon 8	CTTACGGCCGCTTCATCC	2781	exon 13	GGAAGGCTACACGCTGCAC	1000	60	1x Q-Solution
KIAA0496	2454	exon 1	ATGTGCGCTGGCTGTTTTAT	2457	exon 2	AGCAGGCTGAAGATGCTGTT	1200	60	
KIAA1833	2423	intron 16	CCCATCTGCCCTTGACTCTAC	2490	intron17	GGCTCAGTGCCCAATCAC	846	60	
KIAA1833	2491	intron 17	GACAAGAACCCAGCCACAGT	2424	exon 18	GTGCTCCTCTTGGGTCTCCT	1072	60	
KIAA1833	2732	intron 15	AGCAAGATGCTCGTTGGTTG	2733	intron15	ACGTGGGGTGGATGCAG	988	61	
KIAA1833	2744	exon 9	GCCAACAAAGTGAGGAGTGC	2745	exon 10	ATCTGCGCCACCAGCTC	1160	61	1x Q-Solution
KIAA1833	2746	exon 14	CTGAGCCCATGGATCAAGTC	2747	intron15	ACAGACGATGCCCTGAACAC	1035	61	
KIAA1833	1688	intron 15	GGCAGCAGTGTCTGTGTGTT	2407	intron16	GACCAGCGGGGTAGACTAGG	879	60	1x Q-Solution
LOC157534	2541	exon 7	CTCTGCTCCTTCAACCACCT	2542	3'UTR	CCAGCCCACCTTACTGGAC	900	60	5% DMSO
LOC157542	2535	exon 1	TGGAGAGGCACCAGAAAGAG	2536	exon 2	GCCGATCTTTTCAAAGCTGA	900	60	
LOC157542	2587	exon 2	AGTGCAAGTACAAGCGCATC	2588	exon 3	CCGAGGCGTCTCATAGTACA	2kb	60	5% DMSO
LOC157542	2589	exon 3	GTACCGCACAGACGACTTCA	2590	3'UTR	CACCCCAAAACTCTCCTCAT	1100	60	5% DMSO
MGC10520	2651	exon 6	TGCAAAGAGTGTGGCAAAGG	2652	3'UTR	GTGTCTGGACACAACCTACGC	800	60	5% DMSO
MGC13010	2438	exon 2	GGAACCTGGGACTCCTGAAG	2441	3'UTR	GAGACCAAGCCTCTCTCTGG	1100	60	
NFKBIL2	2772	intron 23	CTCCCCAAGGCTCACTTCTG	2785	3'UTR	CCATTTGTAGCCTGTCTTCACG	956	60	
NFKBIL2	2537	exon 23	GAGCTCCTGTCTACCCTCCA	2538	exon 24	ACCTTGTCCCAGAGGTCCAG	1300	60	
NFKBIL2	2539	exon 5	GCTCCCAGAAACCTTTGCAG	2540	exon 6	CTCCTCGGACTCCTCTAGTCG	650	60	
PPP1R16A	2446	exon 16	CTGTACCGCAGGGAGCAC	2449	exon 17	CTCCCAGGTACAGCTTCTGC	832	60	5% DMSO
RECQL4	2430	exon 11	GCCTAGATGAGGCCCACTG	2432	exon 14	GACACAGGGCAGTCAGGTG	605	60	
Rrp41	1691	intron	GGGAGAGGACGAGTCAAGAG	2405	intron	CGGACCCTTAGTCACTGCTG	988	60	1x Q-Solution
Rrp41	2421	exon	TCCTATCGGACCAGGGCTAC	2422	intron	CCCCTGGATCTCTCTCCTGA	957	60	
SLC39A4	2750	exon 11	GGCCTTCATCGGCCTCTAC	2751	exon 12	AATTTATTGAAGCTGGGAAGCAG	450	61	1x Q-Solution
VPS28	2543	exon 1	GGATCCCAGCCACTCCTG	2544	exon 3	GGGCGTTCTTGTACAGCTTC	700	60	
VSP28	2591	exon 5	AGGTCCAGGGCTCAGAAATC	2592	exon 7	GTCCATGGCTCGGATCTC	620	60	5% DMSO

PCR primers for bovine STS markers neighboring DGAT1

Locus	Forward primer		Reverse primer		Product size [bp]	T [°C]	Additives
	No.	Position Sequence [5'-3']	No.	Position Sequence [5'-3']			
100P18-T7	2743	CTGGTCTGAGGAACGCACTG	2766	TGGAGCACAGTTGGGAGTGT	565	61	
100P18-T7	2782	CTCGTGTGCACTGGAGTCTG	2766	TGGAGCACAGTTGGGAGTGT	950	60	
111I3-T7	2680	ACATTTACTTCTGCTTCATTGACTATGTG	2681	TTTTGAACCAGTCCGCTGTC	181	60	
156I10-SP6	2689	CCTCGCCTTTGAGGAAGC	2772	CTCCCCAAGGCTCACTTCTG	518	61	
240A1-T7	1689	GGCTTCCCTGTCCATCACTA	2406	GATCCCAGAATGGGGTCACT	834	60	1x Q-Solution
334E6-T7	2734	ACATTCCTGGCAAAGGGAAC	2735	CCACCCCTCCCTATCCTTG	223	61	
334E6-T7	2734	ACATTCCTGGCAAAGGGAAC	2784	AGATGCCCACACAAAACAGG	824	60	
360L24-T7	2737	GGGAAGCCCCATTTCATTAC	2740	CTGGTCTGAGGAACGCACTG	534	61	
3O1-SP6	2674	TCCAGGCCAGAATACTTTGC	2675	GCCCCAGGGAGTTGTGTG	374	60	
3O1-SP6	2674	TCCAGGCCAGAATACTTTGC	2771	GCCTGGACCCATGACCAC	985	60	
3O1-T7	2671	GGTGGCTACAAAACTACAGTAATCAA	2672	TGGAATGGGAAAGTACTCCAG	293	60	
3O1-T7	2671	GGTGGCTACAAAACTACAGTAATCAA	2783	CAGGGTTCTGTGGCTAATACTCC	1119	60	
410E24-T7	2769	CCACCTTCCGTTCCAACC	2770	CCTGCCTGCAGCTTGTCTC	237	61	
414O23-SP6	2686	CCTCGCCTTTGAGGAAGC	2675	GCCCCAGGGAGTTGTGTG	478	61	1x Q-Solution
428F15-SP6	2697	GCCCTCAGATGATGCTTCG	2698	GAGGGATCTGCCCAGTCTGT	881	61	
428P15-SP6	2486	TGCAGAGTTGGACATGACTTAG	2487	CCAGCAACAAATTGCAACA	842	57	1x Q-Solution
56F1-SP6	1686	TCAGCACTTTTACTGCCAAAGA	2404	AGTGGGGCAGGAGAACTGA	855	60	1x Q-Solution

PCR primers for bovine genes neighboring DGAT1 that were not in the bovine contig

Locus	Forward primer			Reverse primer			Product size [bp]	T [°C]	Additives
	No.	Position	Sequence [5'-3']	No.	Position	Sequence [5'-3']			
FLJ12150	2566	exon 3	GACAGGGCAGGCTCTCAG	2567	exon 4	CCTCCTGCTTGTGCGTTC	900	60	
FLJ13852	2646	exon 3	CGTCACTGCTGAGCACATCT	2647	exon 5	CCCATTTTGATGGCACCTT	800	60	5% DMSO
FLJ20897	2571	exon 5	CGATGATGCAGAAAGGAAGTT	2572	exon 6	CCAGAGACTTCTGGATGTTTTC	200	60	
FLJ20897	2665	exon 8	GATCTGCAGCAGGCTGTCTC	2666	exon 10	GAGCTCCCCCAGACCAAC	700	61	1x Q-Solution
KIAA0628	2575	exon 1	CTGACCGAGCACCAGAAGAT	2576	exon 1	CAAGCATGTTCTCCCAACG	500	60	
LOC58500	2564		GCAAGGCCTTCAACCACAG	2565		ACCTTCTGGTGCTGGATGAG	800	60	1x Q-Solution
LOC90979	2573	exon 1	GCCTATGAATGTAGCGAATGC	2574	exon 1	CCGGAATGGATTCTCTGATG	500	60	
LOC93100	2644	exon 8	ACATTGACGAGGAGGAGCTG	2645	exon 10	GCCACACTCTCAGCTCCTG	600	63	5% DMSO
TSTA3	2577	exon 2	GCCTGTTCCGGAATATCAAA	2578	exon 3	GGCCGAGTGTAGCACATTATC	1000	60	

9.6 Primers used for single base extension (SBE)

SNP_id	SBE primer							PCR primers				
	No.	5' mod.[b]	Sequence [5'-3']	Size [bases]	[c]	Allele[a] 1	2	Forward No.	Sequence [5'-3']	Reverse No.	Sequence [5'-3']	Product size [bp]
274	2020	(N)12	GGCGAGGCCCGGAA	26	F	G	A	2024	GGTGAAGAACTTCCCTGTGG	2025	ACCAGCAGCACCTCCTTG	135
277	2021	(N)22	ACAGGTAGGGAGGCCCCA	30	F	T	C	2014	CCCGAGGGTTCCTGTACTTC	2015	GGCAGACCAGCGACTGAA	304
279	2022	(N)14	CCGAGGACAACTAGGCAGGA	34	F	T	C					
280	2023	(N)15	GAGGACGAGGAGATTGGCA	34	F	A	G					
282	2046	(N)25	GAAGCGTGTGGACGCCA	42	F	C	T	2035	AGGAAGAGCCACCCTCCA	2036	GTTCAGCCCACACCCAAC	99
283	2064	(N)32	CTGCCCGCGTGCCC	46	F	C	T	2062	CAGGGCCTGCTGACCTC	2063	TGTGATCACCTTGGACTGCT	80
289	2429	(A)13	CGTCCTCCTCCCGAACC	30	F	C	T	2427	GGGACGTCCTCCCTGAG	2428	GCGCAGGCAGAAGAACAT	88
	2463	(A)10	CAGGCCCAGCTCCTGG	26	R							
265	2052	(N)18	GACACAGCACACTGGAGCCT	38	F	C	G	2055	TGGAGGGAGGCAGATGG	2056	CCTGGTCCCTTCATAAAGAACTC	88
284	2410	(N)13	TTCCTAGGCCCTCCCGA	30	F	A	G	2408	AATTCACACATGCCCTGCTC	2409	ACTGCTCCCTGAAGGACAAA	93
	2464	(A)15	CCCCGAGAAGGGGTCAGAG	34	R							
285	2534	(A)7	CCCCGAGAAGGGGTCAGAG	26	R							
	2411	(N)17	TCTCGGGGCCAACACAT	34	F	G	A					
286	2414	(N)21	GCAGCACTGCCCTCGAC	38	F	A	G	2412	ACTGCTTCCTGGGCCTTAC	2413	ATGGTAGCTGGCCCTCTGA	106
287	2415	(N)10	GCACCTGGGCGTGACC	26	F	T	C					
	2465	(A)13	CCTCTGAGCACGGACTCCTC	30	R							

[a]Beses of sense sequence. Note that by reverse primer bases of complementary sequence are detected.
[b]Number of bases of poly-a (A) or random sequence (N).
[c]F, forward strand; R, reverse strand.

9.7 Alignments of DGAT2 gene families in human, cattle and pig

Alignment was done using ClustalX version 1.81, with shading of alignment by BOXSHADE 3.21, http://www.ch.embnet.org/software/BOX_form.html

Predicted peptide sequence alignment of DGAT2 genes in human (h), mouse (m), pig (p) and cattle (b)

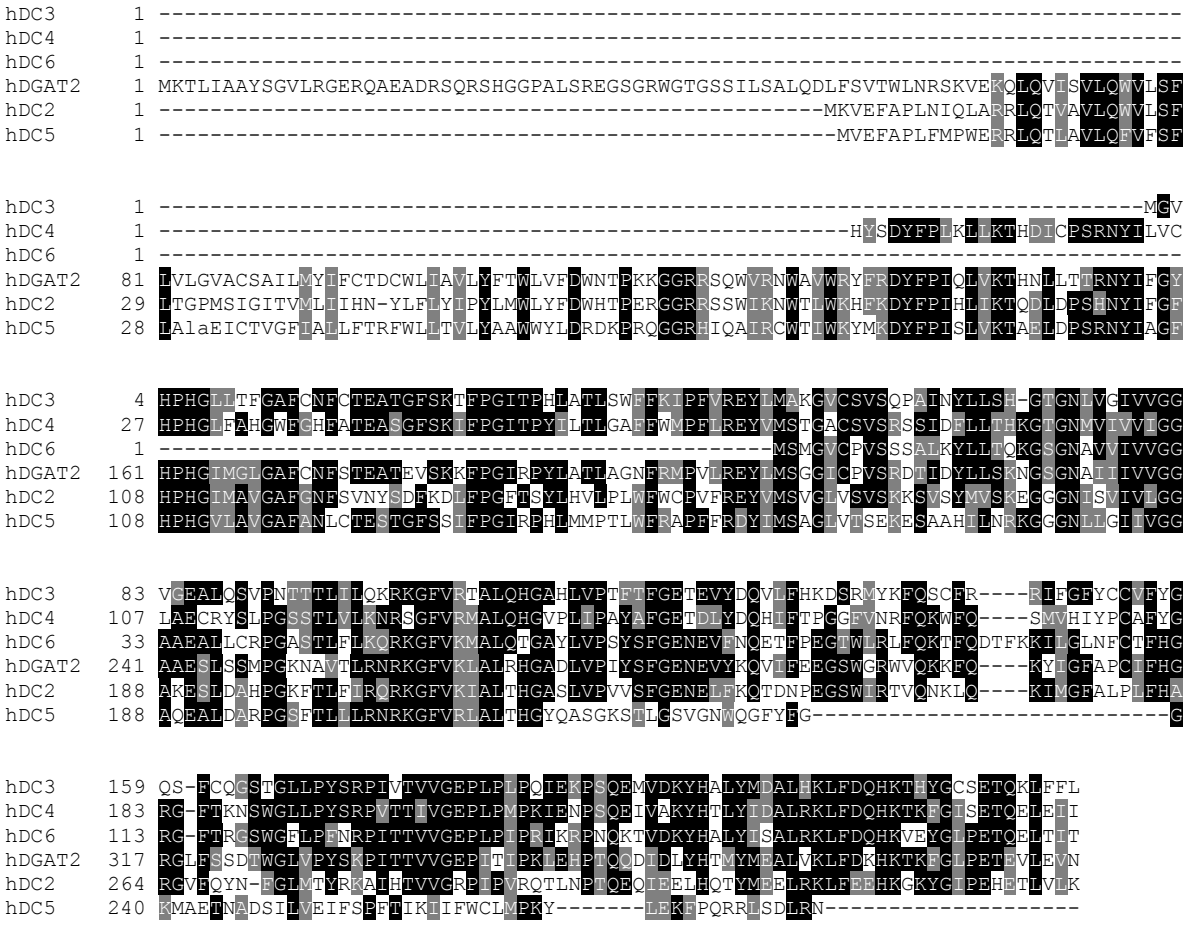
```
hDGAT2    1 MKTLIAAYSGVLRGERQAEADRSQRSHGGPALSREGSGRWGTGSSILSALQDLFSVTWLNRSKVEKQLQVISVLQWVLSF
mDGAT2    1 MKTLIAAYSGVLRGERRAEAARSENKNKGSALSREGSGRWGTGSSILSALQDIFSVTWLNRSKVEKQLQVISVLQWVLSF
pDGAT2    1 MKTLIAAYSGVLRG--------------------------TGSSILSALQDLSAITWLNRSKVEKQLQVISVLQWVLSF
bDGAT2    1 MKTLIAAYSGVLRG--------------------------TGSSILSALQDLFSVTWLNRAKVEKQLQVISVLQWVLSF

hDGAT2   81 LVLGVACSAILMYIFCTDCWLIAVLYFTWLVFDWNTPKKGGRRSQWVRNWAVWRYFRDYFPIQLVKTHNLLTTRNYIFGY
mDGAT2   81 LVLGVACSVILMYTFCTDCWLIAVLYFTWLAFDWNTPKKGGRRSQWVRNWAVWRYFRDYFPIQLVKTHNLLTTRNYIFGY
pDGAT2   54 LVLGVACSVILVYLICTDCWLITALYFTWLAFDWNTPKKGGRRSQWVRNWAVWRYFRDYFPIQLVKTHNLLTTRNYIFGY
bDGAT2   81 LVLGVACSVILMYTFCTDCWLIAVLYFTWLVFDWNTPKKGGRRSQWVRNWAVWRYFRDYFPIQLVKTHNLLTSRNYIFGY

hDGAT2  161 HPHGIMGLGAFCNFSTEATEVSKKFPGIRPYLATLAGNFRMPVLREYLMSGGICPVSRDTIDYLLSKNGSGNAIIIVVGG
mDGAT2  161 HPHGIMGLGAFCNFSTEATEVSKKFPGIRPYLATLAGNFRMPVLREYLMSGGICPVNRDTIDYLLSKNGSGNAIIIVVGG
pDGAT2  134 HPHGIMGLGAFCNFSTEATEVSKKFPGIKPYLATLAGNFRMPVLREYLMSGGICPVNRDTIDYLLSKNGSGNA-------
bDGAT2  161 HPHGIMGLGAFCNFSTEATEVSKKFPGIRPYLATLAGNFRMPVLREYLMSGGICPVNRDTIDYLLSKNGSGNAIIIVVGG

hDGAT2  241 AAESLSSMPGKNAVTLRNRKGFVKLALRHGADLVPIYSFGENEVYKQVIFEEGSWGRWVQKKFQKYIGFAPCIFHGRGLF
mDGAT2  241 AAESLSSMPGKNAVTLKNRKGFVKLALRHGADLVPTYSFGENEVYKQVIFEEGSWGRWVQKKFQKYIGFAPCIFHGRGLF
pDGAT2      ----------------------------------------------------------------------------------
bDGAT2  241 AAESLSSMPGKNAVTLRNRKGFVKLALRHGADLVPTYSFGENEVYKQVIFEEGSWGRWVQKKFQKYIGFAPCIFHGRGLF

hDGAT2  321 SSDTWGLVPYSKPITTVVGEPITIPKLEHPTQQDIDLYHTMYMEALVKLFDKHKTKFGLPETEVLEVN
mDGAT2  321 SSDTWGLVPYSKPITTVVGEPITVPKLEHPTQKDIDLYHAMYMEALVKLFDNHKTKFGLPETEVLEVN
pDGAT2      ------------------------------------------------------------
bDGAT2  321 SSDTWGLVPYSKPITTVVGEPITIPRLERPTQQDIDLYHAMYVQALVKLFDQHKTKFGLPETEVLEVN
```

Predicted peptide sequence alignment of DC2 (MGAT1) genes in human (h), cattle (b) and mouse (m)

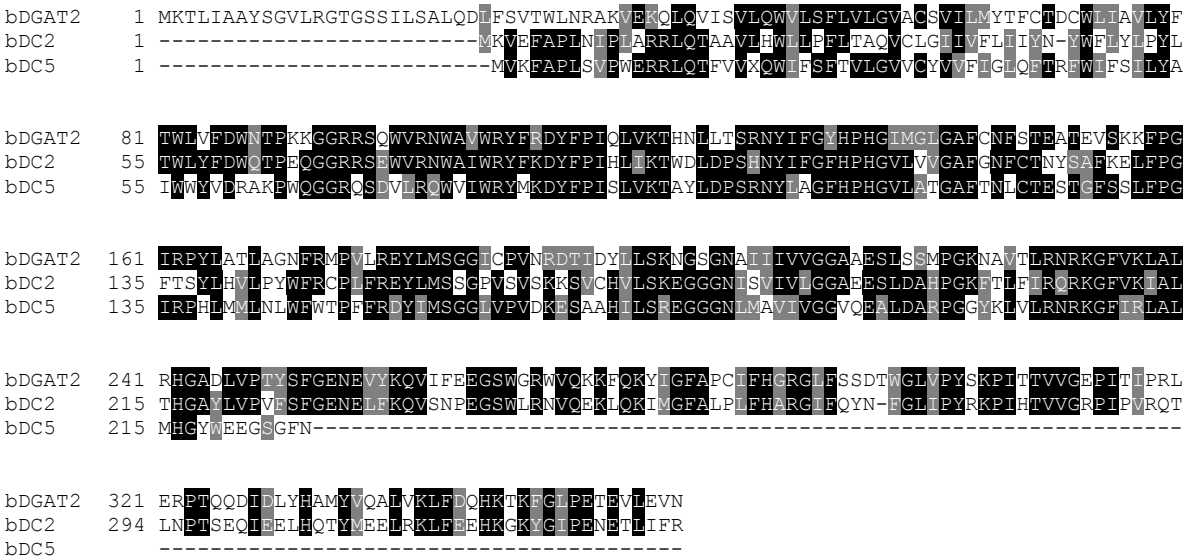
```
hDC2     1 MKVEFAPLNIQLARRLQTVAVLQWVLSFLT-GPMSIGITVMLIIHNYLFLYIPYLMWLYFDWHTPERGGRRSSWIKNWTL
bDC2     1 MKVEFAPLNIPLARRLQTAAVLHWLLPFLT-AQVCLGITVFLIIYNYWFLYIPYLTWLYFDWQTPEQGGRRSEWVRNWAI
mDC2     1 MMVEFAPLNTPLARCLQTAAVLQWVLSFLLLVQVCIGIMVMLVLYNYWFLYIPYLVWFYYDWRTPEQGGRRWNWVQSWPV

hDC2    80 WKHFKDYFPIHLIKTQDLDPSHNYIFGFHPHGIMAVGAFGNFSVNYSDFKDLFPGFTSYLHVLPLWFWCPVFREYVMSVG
bDC2    80 WRYFKDYFPIHLIKTWDLDPSHNYIFGFHPHGVLVVGAFGNFCTNYSAFKELFPGFTSYLHVLPYWFRCPLFREYLMSSG
mDC2    81 WKYFKEYFPICLVKTQDLDPGHNYIFGFHPHGIFVPGAFGNFCTKYSDFKKLFPGFTSYLHVAKIWFCFPLFREYLMSNG

hDC2   160 LVSVSKKSVSYMVSKEGGGNISVIVLGGAKESLDAHPGKFTLFIRQRKGFVKIALTHGASLVPVVSFGENELFKQTDNPE
bDC2   160 PVSVSKKSVCHVLSKEGGGNISVIVLGGAEESLDAHPGKFTLFIRQRKGFVKIALTHGAYLVPVFSFGENELFKQVSNPE
mDC2   161 PVSVSKESLSHVLSKDGGGNVSIIVLGGAKEALEAHPGTFTLCIRQRKGFVKMALTHGASLVPVFSFGENDLYKQINNPK

hDC2   240 GSWIRTVQNKLQKIMGFALPLFHARGVFQYNFGLMTYRKAIHTVVGRPIPVRQTLNPTQEQIEELHQTYMEELRKLFEEH
bDC2   240 GSWLRNVQEKLQKIMGFALPLFHARGIFQYNFGLIPYRKPIHTVVGRPIPVRQTLNPTSEQIEELHQTYMEELRKLFEEH
mDC2   241 GSWLRTIQDAMYDSMGVALPLIYARGIFQHYFGIMPYRKLIYTVVGRPIPVQQILNPTSEQIEELHQTYLEELKKLFNEH

hDC2   320 KGKYGIPEHETLVLK
bDC2   320 KGKYGIPENETLIFR
mDC2   321 KGKYGIPEHETLVFK
```

Predicted peptide sequence alignment of DC5 genes in human (h), pig (p) and cattle (b)

Predicted peptide sequence alignment of human DGAT2 gene family



Predicted peptide sequence alignment of bovine DGAT2, DC2 and DC5 genes



Predicted peptide sequence alignment of porcine DGAT2, DC5 and DC7 genes

9.8 Genotypes

Genotypes for SNPs within the DGAT1 gene

		snp_id		28	63	169	237	230	252	258	259	260
			Repeat									
		Allele 1	1 to 7	C	T	A	G	C	A	A	C	C
		Allele 2	units	G	C	G	A	T	G	G	T	T
Breed	Animal	BVF[a]										
German Simmental	906	0.75	56	CC	TC	AA	GG	CC	AA	AA	CC	CC
	916	0.69	45	CC	TT	AA	GG	CT	AA	AA	CC	CT
	933	0.62	56	CC	TC	AA	GG	CC	AA	AA	CC	CC
	902	0.78	44	CG	TT	GG		TT	GG	GG	TT	TT
	914	0.69	45	CG	TT	GG		TT	GG	GG	TT	TT
	920	0.68	56	CC	TT	GG		CT	GG	GG	CT	TT
	921	0.67	44	CC	TT	GG	AA	TT	GG	GG	TT	TT
	923	0.66	36	CC	TT	GG		CT	GG	GG	CT	TT
	917	0.69	66	CC				CT	AG	GG	CT	CT
	932	0.62	46	CC	TC	AG		CT	AG	GG	CT	CT
	705	0.80	46	CC	TC	AG	GA?	CT	AG	GG	CT	TT
	899	0.22	45	CG					AG	AG	CT	CT
	FV19		44	CC					GG	GG	TT	TT
	FV27		46	GC					GG	GG	TT	TT
	FV28		44	CC					GG	GG	TT	TT
German Holstein	SB26		56	CC	TT	AA	GG		AA	AA	CC	CC
	SB37		56	CC					AG	AG	CT	CT
	SB45		57	CC					AG	AG	CT	CT
Angus	AN1		56	CC					GG	GG	TT	TT
Kerry	KE2		46	n.d.					GG	GG	TT	TT
Jersey	JE1071		44	CC					AA	AG	CC	CC
	JE1072		44	CC					AA	AG	CC	CC
	JE1073		44	CC					GG	GG	TT	TT
	JE1074		44	CC					AG	GG	CT	TT
	JE1075		45	CC					AG	GG	CT	CT
	JE1076		45	CC					AA	GG	CC	CC
	JE1077		n.d.	CC					AA	AG	CC	CT
Sahival	SA4		44	CC					AA	GG	CC	TT
Hariana	HA8		56	CC					AA	AA	CC	TT
Yak	Yak1080		44	CC					AA	GG	CC	CC
	Yak1081		44	CC					AA	GG	CC	CC
Water buffalo	WB1078		11	TT					AA	GG	CC	
	WB1079		11	TT					AA	GG	CC	CC

[a]Breeding value for milk fat percentage

Genotypes for SNPs within the DGAT1 neighboring loci

Locus	PTK2	4281F15-SP6			2933G16-SP6				KIAA1833				BOP1				DGAT1				HSF1			FBXL6	FLJ11856	56F1-SP6		
SNP id	276	310	311	316	315	314	313	317	318	319	320	321	322	KC4	323	28	252	258	259	260	265	266	267	324	325	290	BA2	
Allele 1	A	T	G	G	G	A	T	C	G	G	C	C	C	N0	T	C	A	A	C	C	C	G	C	G	G	G	N0	
Allele 2	G	G	T	C	A	G	G	T	C	T	T	T	G	N1	C	G	G	G	T	T	G	A	T	C	A	T	N2	
HF32+	AG	TT	GG	GG	GG	AA	TT	CC	GG	GG	CC	CC	CC	N0N0	TT	CC	AA	AA	CC	CC				GG	GG	n.d.	N0N0	
HF32-	AG	GG	GG	GC	CA	AG	TG	CT	GC	GT	CT	CT	CC	N0N0	TT	CC	GG	GG	TT	TT				CC	AA	GG	N2N2	
FV32+	AG	GG	GG	GG	GG	AA	TT	CC	GC	GG	CT	CT	CG	N0N1	TT	CC	AG	AG	CT	CT	CG	GG	CT	GC	GA	GT	N0N2	
FV32-	AA	GG	GG	GG	GG	AA	TT	CC	CC	GG	TT	TT	CG	N0N1	TT	CG	GG	GG	CT	TT	GG	GG	TT	CC	AA	GG	N2N2	
899	AG	T?	GG	GG	GG	AA	TT	CC	?	GG	CT	CT	?	?	TT	CG	AG	AG	CT	CT				GC	GA			
361	AG	TT	GG	GG	GG	AA	TT	CC	GG	GG	CC	CC	CC	N0N0	TT	CC	AA	AA	CC	CC	GC			GG	GG			
1091	GG	T?	GG	GC	CA	AG	TG	CT	GC	GT	CT	CT	CC	N0N0	TT	CC	GG	GG	TT	TT				CC	AA			
1180	AG	TT	GG	GG	GG	AA	TT	CC	GG	GG	CC	CC	CC	N0N0	TT	CC	AA	AA	CC	CC				GG	GG			

N0 del
N1 GGGGCGGAGGGC
N2 GATACAACT
N3 $(C)_6A(C)_4$
N4 $(C)_9A(C)_3$
N5 $(G)_n$
N6 $(G)_7C(G)_3$

Locus	301-T7				CPSF1												360L24-T7		KIAA0014			VPS28	CYHR1					GPT			RECQL4			301-SP6	100P18-T7	MGC10520	
SNP id	326	327	328	329	268	269	270	271	281	282	283	274	277, 278	279	280	330	BE2	331, 332	333	334	KD4	VP1	284	285	307	308	309	288	289	286	287	335	336	306	
Allele 1	G	C	C	G	T	A	C	A	C	C	C	G	TG	T	A	C	N3	TG	C	C	(AC)7	N5	A	G	T	A	C	G	C	A	T	del	G	A	
Allele 2	T	T	G	C	C	G	T	G	G	T	T	A	C-	C	G	T	N4	N0	G	T	(AC)6	N6	G	A	G	G	A	A	T	G	C	C	C	G	
HF32+	GG	CC	CC	GG	CC	GG	TT	GG	GG	TT	CC	AA				C?	N3N4	TG	TG	CC	77	N5N5	AA	GG	TT	AA	CC	GG	CC	AA	TT	del	GG	AA	
HF32-	TT	TT	GG	CC	TT	AA	CC	AA	GG	CC	CC	GG				TT	N3N3	TG	N0	GG	56	N5N6	GG	AA		GG	AA	GA	TT	GG	CC	CC	CC	AG		
FV32+	TT	TT	GG	CC	TC	AG	CT	AG	CC	CT	CT	GA	TG	C- TC	AA	TT	N3N3	TG	N0	GG	CT	N6N6						GG	CT			del	CC			
FV32-	TT	TT	GG	CC	TT	AA	CC	AA	GG	CC	CC	GG	TG	TG TT	Ag	TT	N3N3	TG	N0	GG	CC	N6N6						GG	CT			CC	CC			
899	GT	CT	CG	GC									TG	C- TC	AA	TT	N3N3	TG	N0	GC	CC	67	N5N5	AG	GA		GG		CT	AG	TC	del	GC	AG		
361	GG	CC	CC	GG									C-	C- CC	AA	TT	N3N3	TG	N0	GC	CT	57	N5N5	AG	GA	GA		CC		AG	TC	CC	GC	AG		
1091	TT	TT	GG	CC									TG	TG TT	GG	TT	N3N3	N0	N0	GG	CC	56	N5N5	GG	AA		GG		TT	GG	CC	CC	CC	GG		
1180	GG	CC	CC	GG									C-	C- CC	AA	CT?	N4N4	TG	TG	CC	CC	77	N5N5	AA	GG		AA		CC	AA	TT	del	GG	AA		

Genotypes for 12 SNPs within the DGAT2 gene and one SNP within the DGAT1 gene (252)

	snp_id	293	294	295	296	297	298	299	300	301	302	303	304	252
	Allele 1	T	T	C	C	A	A	A	C	A	G	T	T	A
	Allele 2	G	C	G	T	G	G	C	T	del	A	G	C	G
Breed	Animal													
German Holstein	HF1180	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AA
	HF1184	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	AA
	HF1087	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	GG
	HF1091	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
German Simmental	FV902	TT	CC	GG	TT	GG	GG	CC	TT	--	AA	GG	CC	GG
	FV906	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	AA
	FV914	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
	FV916	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	AA
	FV920	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
	FV921	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	GG
	FV923	TT	CC	GG	TT	GG	GG	CC	TT	--	AA	GG	CC	GG
	FV932	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AG
	FV933	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AA
	FV1066	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AG
	FV1063	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
	FV1064	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	GG
	FV1065	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	GG
	FV361	GG	TT	CC	CC	AA	AA	AA	CC	AA	AA	TT	TT	AA
	FV705	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AG
	FV899	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	AG
German Brown	BV929	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AG
	BV909	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AG
	BV943	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	GG
	BV1044	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
	BV1045	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
	BV1057	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
Anatolian Black	AB27	TT	CC	GG	TT	GG	GG	CC	TT	--	AA	GG	CC	AA
	AB4	??	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	GG
Jersey	JE1071	TT	CC	GG	TT	GG	GG	CC	TT	--	AA	GG	CC	AA
	JE1072	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AA
	JE1073	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	GG
	JE1074	TT	CC	GG	TT	GG	GG	CC	TT	--	AA	GG	CC	AG
	JE1075	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	AG
	JE1076	GT	TC	CG	CT	AG	AG	AC	CT	A-	GA	TG	TC	AA
	JE1077	GG	TT	CC	CC	AA	AA	AA	CC	AA	GG	TT	TT	AA
Yak	Yak1080	GG	CC	GG	CC	GG	GG	CC	TT	--	AA	GG	CC	AA
	Yak1081	GG	CC	GG	CC	GG	GG	CC	TT	--	AA	GG	CC	AA

Locus	exons pred.	exons inv.	Sequence [bp] total	5'end	5'utr	coding	introns	3'utr	3'end	SNPs total	5'end	5'utr	exons silent	exons effect	introns	3'utr	3'end	[bp/SNP] total	5'end	5'utr	exons silent	exons effect	introns	3'utr	3'end
total			84549	5490	575	22910	35754	5910	3504	101	8	2	11	6	59	16	1	837	686	288	2083	5959	100	219	837
DGAT1 region			69414	3748	309	20138	25728	4496	2599	57	1	1	9	5	30	12	1	1218	3748	309	2238	5146	150	217	1218
DGAT2	8	8	7023	619	198	1086	4580	149	391	23					22	1		305					208	149	
DC2	6	6	6375	603	?	1005	4143	100	514	4				1	3			1594				1005	1381		
DC5	4	4	1737	520	68	681	1303	1165		17	7	1	2		4	3		102	74	68	341		326	388	
DGAT1	17	17	12800	3500	104	1470	6834	275	617	2	1	1	2	1	10	5		6400	3500	104	735	1470	683	55	
BOP1	16	13	2807			1860	947			5			2		3			561			930		316		
CPSF1	38	27	6283			3722	2561			11			2	2	7			571			1861	1861	366		
CYC1	7	4	968			548	420																		
CYHR1	>3	3	2960		18	825	1349	768		5					3	2		592					450	384	
DKFZp547F072	1	1	349			349																			
FBXL6	9	7	2494			1008	1486			1					1			2494					1486		
FLJ11856	5	2	1247			1013	234			1				1				1247				1013			
FLJ35454	10	3	815			262	223	330																	
FOXH1	3	3	1405		48	970	387																		
GPT	11	7	1525			926	599			2			1		1			763			926		599		
HSF1	12	10	3205			1461	932	185	627	3					2		1	1068					466		627
KIAA0014	5	3	3197			694	185	2318		4						4		799						580	
KIAA1833	18	7	6416			937	5479			3			1		2			2139			937		2740		
MGC10520	6	1	734			446		288		1						1		734						288	
MGC13010	3	2	1097		118	805	174																		
NFKBIL2	24	3	2699			305	1672	?	722																
PPP1R16A	7	2	1643			541	466	213	423																
RECQL4	21	4	1002			773	229			2			1	1				501			773	773			
RRP41	3	3	2285	248	21	738	1039	29	210																
SLC39A4	12	2	412			223	99	90																	
VPS28	9	3	675			262	413			1					1			675					413		
100P18-T7			1071							1								1071							
111I3-SP6			708																						
111I3-T7			773																						
240A1-T7			876																						
293G16-SP6			1357							4								339							
334E6-T7			933																						
360L24-T7			771							2								386							
3O1-SP6			1041							1								1041							
3O1-T7			1266							4								317							
410E24-T7			664																						
414O23-SP6			660																						
428F15-SP6			1399							2								700							
56F1-SP6			877							2								439							