

Molecular Coat Colour Genetics

**Dominant White in Pigs and
Greying with Age in Horses**

Gerli Pielberg

*Department of Animal Breeding and Genetics
Uppsala*

**Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2004**

Acta Universitatis Agriculturae Sueciae
Agraria 500

ISSN 1401-6249
ISBN 91-576-6779-9
© 2004 Gerli Pielberg, Uppsala
Tryck: SLU Service/Repro, Uppsala 2004

Abstract

Pielberg, G. 2004. *Molecular Coat Colour Genetics. Dominant White in Pigs and Greying with Age in Horses*. Doctor's dissertation.
ISSN 1401-6249, ISBN 91-576-6779-9.

Coat colour, being sufficiently easy to observe and score, is a model phenotype for studying gene action and interaction. Farm animals provide us a valuable resource for identifying genes behind such phenotypic traits, because they show a remarkably higher diversity in coat colour compared to natural populations.

This thesis concerns two different farm animal coat colour phenotypes. First, the dominant white phenotype in domestic pigs is studied. It has been previously determined that white coat colour in domestic pigs is caused by two mutations in the *KIT* gene – a gene duplication and a splice mutation in one of the copies. The genetic analysis of such a locus with many alleles is complicated. Pyrosequencing and minisequencing are the methods applied in present studies for quantitative analysis of the splice mutation and unexpectedly high allelic diversity at the porcine *KIT* locus was revealed. Furthermore, a method based on pyrosequencing for analysing gene copy numbers, using this locus as a model, was developed. This method is based on a competitive amplification of the breakpoint regions from duplicated and normal copies of *KIT* and quantification of the products. This strategy, analysing genes with multiple copies, could have a much broader use than just analysing porcine *KIT* locus. For instance, several human genetic diseases are caused by changes in gene copy number and analysis of these genes has been quite complicated.

The second part of the thesis is summarising the greying with age phenotype in horses. These horses are born coloured, turn progressively grey and most of them develop melanoma. The locus responsible for this phenotype has been mapped to horse chromosome 25, homologous to human chromosome 9. Two Swedish Warmblood horse families were used in present studies to perform a comparative linkage mapping of the *Grey* locus. The identified interval region has no obvious candidate genes affecting coat colour or melanoma development. Since this mutation occurred most probably only once, thousands of years ago, the identification of a minimum shared haplotype among individuals from different breeds should reveal the gene and the mutation behind this fascinating phenotype.

Keywords: coat colour, farm animals, dominant white in pigs, duplication, pyrosequencing, greying with age in horses, comparative linkage mapping.

Author's address: Gerli Pielberg, Department of Animal Breeding and Genetics, SLU. BMC, Box 597, S-751 24 UPPSALA, Sweden. Email: Gerli.Pielberg@bmc.uu.se.

*And then you wrap yourself into silence
Walk across the winds to find the door
Look into the keyhole
And what you see is not the world you saw before
It's changing*

*Sometimes you fly and tend to fall
Or lose the way at all
To find – you're in confusion
May be that goals had been too small
You gain them and it's all
An end of old illusion*

*And then you wrap yourself into silence
Walk across the winds to find the door
Look through doorless keyhole
And what it sees is not the you you knew before
You're changing*

Tarmo & Toomas Urb

Contents

Introduction, 7

Pig coat colour genetics, 9

Dominant white coat colour in pigs, 9

Gene duplications in mammalian genomes, 10

Horse coat colour genetics, 11

Greying with age in horses, 12

Objectives of the thesis, 14

Main methods used in the present studies, 15

Pyrosequencing, 15

Minisequencing, 15

Real-time PCR (TaqMan), 17

Invader assay, 18

Linkage analysis, 18

Results and discussion, 20

Unexpectedly high allelic diversity at the porcine *KIT* locus (Paper I), 20

A sensitive method for copy number analysis of duplicated genes (Paper II), 21

Comparative linkage mapping of the grey coat colour gene in horses (Paper III), 23

Future prospects, 26

KIT gene in pigs, 26

Grey coat colour gene in horses, 27

References, 29

Acknowledgements, 33

Appendix

Papers I-III

The present thesis is based on the following papers, which will be referred to by their Roman numerals:

- I. Pielberg, G., Olsson, C., Syvänen, A.C. & Andersson, L. 2002. Unexpectedly high allelic diversity at the *KIT* locus causing dominant white color in the domestic pig. *Genetics* 160, 305-311.
- II. Pielberg, G., Day, A.E., Plastow, G.S. & Andersson, L. 2003. A sensitive method for detecting variation in copy numbers of duplicated genes. *Genome Research* 13, 2171-2177.
- III. Pielberg, G., Mikko, S., Sandberg, K. & Andersson, L. Comparative linkage mapping of the grey coat colour gene in horses. (Manuscript).

Papers I-II are reproduced by the permission of the journals concerned.

Introduction

Coat colour in general has intrigued scientists all over the world for centuries. Such an interest could be partly explained by this phenotypic characteristic being easily recognisable, making it simple to follow the inheritance from generations to generations. Therefore, coat colour has become a model phenotype for studying gene action and gene interaction. In fact, the first genetic studies in mammals were conducted on coat colour genetics.

Farm animals provide a valuable resource of coat colour diversity compared to limited variation in natural populations. This could be explained by a release from natural selection against individuals with an odd appearance during mate selection and prey-predator interactions, and selection for particular phenotypes by humans (Andersson, 2001). Since coat colour is used as a trademark for different breeds, it plays an important role in animal breeding, and several DNA tests for coat colour genes are being used by industry.

Molecular coat colour genetics in farm animals relies strongly on mouse coat colour genetics. There is more than 100 genes correlated with coat colour phenotypes in mice, and approximately 50% of them are still unidentified (Oetting, W.S. & Bennett, D.C., <http://ifpcs.med.umn.edu/micemut.htm>; 20-Oct-2004). After identifying genes affecting coat colour in mice, different mutations in these genes in farm animals, not to mention humans, have been revealed. One of the examples is a gene identified as the *extension* (*e*) locus, called *MC1R* (melanocortin receptor 1). This receptor is a G-protein-coupled receptor with seven transmembrane domains, expressed at high levels in melanocytes and playing a key role in the melanin synthesis signalling pathway through regulating intracellular levels of cyclic AMP (cAMP). This receptor could be called a pigmentary switch – activation leads to synthesis of eumelanin (brown/black) at the expense of pheomelanin (yellow/red). Mutations in this gene were first shown in mice (Robbins *et al.*, 1993) and subsequently in cattle (Klungland *et al.*, 1995), horses (Marklund *et al.*, 1996), pigs (Kijas *et al.*, 1998), sheep (Vage *et al.*, 1999), dogs (Newton *et al.*, 2000) and chicken (Takeuchi *et al.*, 1996; Kerje *et al.*, 2003).

In addition to *MC1R* several other genes, known to affect coat colour in mice, have been and are being tested as potential candidates for farm animal coat colour phenotypes. Among these are the *MC1R* agonist (melanocyte stimulating hormone, α -MSH) and antagonist (*agouti* protein, *ASIP*), which are ligands regulating *MC1R* signalling as extracellular proteins by stimulating or inhibiting the *MC1R* receptor, respectively. Tyrosinase (*Tyr*), tyrosinase related protein 1 (*Tyrp1*) and tyrosinase related protein 2 (*Tyrp2*, *Dct*) are other genes known to influence pigmentation or pigment synthesis level. In addition, there are genes affecting melanocyte precursor migration from neural crest, melanocyte survival and development. Examples are the transcription factors *Pax3* and *Mitf*, receptor tyrosine kinase *Kit* and its ligand *Kitl* (*MGF*), a G-protein coupled receptor called endothelin receptor B (*Ednrb*) and its ligand endothelin 3 (*Edn3*). All this

provides us with a directional view of complex pigment cell and pigmentation development, wherein *Ednrb* signalling precedes *KIT* signalling, which in turn precedes *MC1R* signalling in a serial pathway (Jackson, 1997) (Fig. 1).

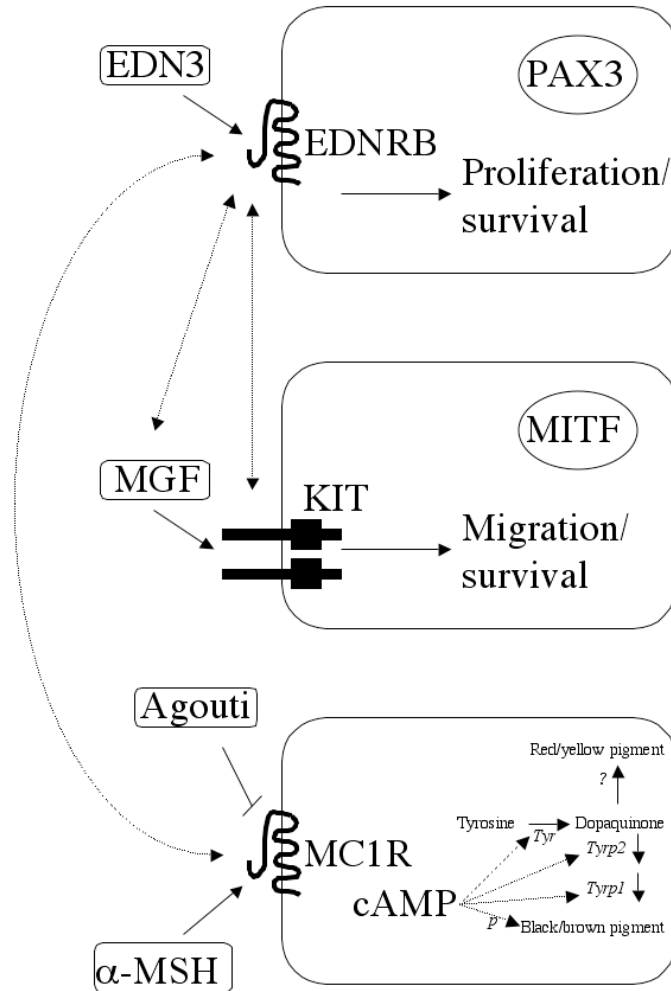


Fig. 1. Gene action and interaction during pigment cell development and melanogenesis. Shows how *Edn3* and *Ednrb* genes act relatively early, affecting the proliferation and/or survival of neural crest cells destined for the pigment and the enteric ganglia lineage. *MGF* and *KIT* genes act somewhat later to affect melanoblast migration and/or survival, whereas *agouti*, α -*MSH* and *MC1R* genes affect melanogenic control. The double-headed dashed arrows illustrate potential interactions between *Ednrb* and *MGF*, *KIT* or *MC1R* genes. Melanogenic control is mediated by intracellular levels of cAMP (modified from Barsh, 1996).

Pig coat colour genetics

Domestication of pigs has occurred independently from Wild Boar subspecies in Europe and Asia. Historical records and molecular evidence indicate that Asian pigs were introduced into Europe during the 18th and early 19th centuries (Giuffra *et al.*, 2000; Kijas *et al.*, 2001a).

There are several varieties of coat colour phenotypes in domesticated pigs. Most of the European breeds, such as Large White and the different Landrace pigs have a white coat colour, while the uniform black coat is the most common type in breeds from China (Legault, 1998). The main genes in pig coat colour genetics, controlling the relative amounts of melanin, eumelanin (black/brown) and pheomelanin (yellow/red), have also been identified as the *extension* and *agouti* loci. In a variety of mammals, dominant alleles at the *extension* act to produce uniform black coat colour, whereas recessive alleles at this locus extend the amount of red/yellow pigment. Dominant alleles at the *agouti*, on the other hand, cause a yellow coat whereas the homozygosity for the bottom recessive allele is associated with a uniform black coat (Jackson, 1997). The uniform black colour in pigs is caused by mutations in the *extension* locus, probably resulting in constitutively active melanocortin receptor 1 (Kijas *et al.*, 1998). There are other *MC1R* alleles associated with wild-type, black spotting and uniform red coat colour, the latter being probably caused by a disruption of *MC1R* function (Kijas *et al.*, 1998; Kijas *et al.*, 2001b). Several additional loci affecting coat colour in pigs have been mapped recently by whole-genome scan (Hirooka *et al.*, 2002). During this study additional loci affecting white and black phenotypes were located, one of them possibly representing the porcine equivalent to the *KIT* ligand called *MGF*.

Dominant white coat colour in pigs

Already in 1906, Spillman concluded that the white coat colour in domestic pigs is a dominant trait. Since then, the potential cause for white coat in domestic pigs has been unravelled primarily using an intercross between Large White (*II*) and Wild Boar (*i/i*) pigs, carried out in Uppsala. First it was shown that the responsible gene is located on pig chromosome 8 and *KIT* gene was proposed as a potential candidate gene (Johansson *et al.*, 1992). *KIT* encodes a mast/stem cell growth factor receptor, a key signal-transducing receptor in stem cell systems feeding into multiple lineages such as haematopoietic cells, germ cells, melanocytes, intestinal pacemaker cells and neuronal cells (reviewed in Bernstein *et al.*, 1990; Broudy, 1997). Therefore, the dominant white phenotype in pigs, being due to a mutation in *KIT* that leads to lack of melanocytes in the skin, is consistent. Mutations in this gene also cause pigmentation disorders in mice, called *Dominant white spotting/W* (Chabot *et al.*, 1988; Geissler, Ryan & Housman, 1988), and in humans, called piebald trait (Fleischman *et al.*, 1991; Giebel & Spritz, 1991). In the mouse, loss-of-function mutations are associated with limited white spotting in heterozygotes, but they are often lethal or sublethal in the homozygous condition due to their effects on haematopoiesis (Fleischman & Mintz, 1979; Waskow *et al.*, 2002).

In 1996 Johansson Moller and others presented genetic data demonstrating that the *KIT* gene in white pigs is duplicated. Furthermore, in 1998, the cause for dominant white phenotype in pigs was being presented as two mutations in the *KIT* gene – one gene duplication associated with a partially dominant phenotype and a splice mutation in one of the copies, leading to the fully dominant allele (Marklund *et al.*, 1998). The splice mutation is a G to A substitution in the first nucleotide of intron 17 leading to skipping of exon 17. Exon 17 is composed of 123 bp and the mutation is thus an in-frame mutation deleting 41 amino acids from the protein. Those 41 amino acids are part of a highly conserved region in tyrosine kinases, comprising the catalytic loop and part of the activation loop (Hubbard *et al.*, 1994). In contrast to the duplication, considered to be a regulatory mutation, the splice mutation is expected to cause a receptor with impaired or absent tyrosine kinase activity. But on the other hand, the existence of *KIT* duplication in domestic pigs allows the expression of at least one normal *KIT* receptor per chromosome, which is apparently sufficient to avoid severe pleiotropic effects on the tissues.

Another common phenotype in domesticated pigs, associated with *KIT* locus, is a white belt. It is a dominant phenotype and may occur against a solid black (Hampshire) or red background (Bavarian Landschwein). Many mouse mutants display a white belt of variable width and position across the body, for instance patch and rump-white. It is known that patched phenotype in mice is caused by a deletion located ~180 kb upstream of *KIT* gene and rump-white is caused by a large inversion involving the same region. In pigs, the locus responsible for belted phenotype has been mapped to chromosome 8, which harbours the *KIT* gene. Therefore, the belt mutation is probably caused by a regulatory mutation affecting *KIT* expression (Giuffra *et al.*, 1999).

Gene duplications in mammalian genomes

Large segmental duplications are very common in mammalian genomes. These duplications are abundant (5% of the genome), span large genomic distances (1-100 kb) and can have very high sequence identity (95-100%) (Eichler, 2001; Bailey *et al.*, 2002). The high level of sequence identity provides an ample substrate for recombination events. Furthermore, nearly identical sequence copies in the genome created by duplications may lead to large-scale chromosomal rearrangements, such as deletions, inversions, translocations and additional duplications.

The duplications can harbour large pieces of genetic material and contain both high-copy number repeats and gene sequences with exon-intron structures. An important feature of duplications is that if they harbour coding sequences, they may not include the whole coding region and regulatory elements of a gene. It has become apparent though, that many duplicated segments are expressed, although the transcripts frequently show tissue specific transcription pattern (germline, fetal, cancerous), contain premature stop codons and are non-functional (Horvath *et al.*, 2001). Several of such transcripts are chimeric, incorporating exons from different

genes, which raises the possibility of a mechanism akin to exon shuffling. Therefore, such chromosomal rearrangements as duplications are considered as exclusive contributors to the origin of reproductive isolation and the formation of new species (Lynch, 2002). Moreover, it has been proposed that imprinting may originally have evolved on a simple basis of dosage compensation required for some duplicated genes or chromosomes (Walter & Paulsen, 2003). This indicates that duplications may play an even more important role in the development of genomes than just gene innovation and speciation.

The distribution of duplication events on human chromosomes seems non-uniform and detailed analysis has revealed that larger blocks of duplicated material are, in fact, composed of smaller units of modules of duplications. There are several potential explanations to duplications being more abundant in subtelomeric and periscentromeric regions. One of them relies on their lower gene density, showing a greater tolerance of these regions for the incorporation of new genetic material without adverse effects to the organism. However, there are other parts of chromosomes that are devoid of genes and yet show no increase in segmental duplication (Horvath *et al.*, 2001). Another explanation to duplicated segments being positioned near centromeres could be a simple suppression of recombination in this region, resulting in a slower deletion of new genetic material.

Segmental duplications in human have been indicated as one type of chromosomal rearrangements causing genetic diseases. It has been estimated that 1 in every 1000 human births has a duplication-mediated germline rearrangement. For instance, human cytochromes and chemokines are known to have interindividual and interethnic variabilities in copy numbers. Since cytochromes are enzymes involved in metabolism of endogenous and exogenous compounds, the variability in copy number leads to differences in effects and toxicities of many drugs (antidepressants, neuroleptics, cardiovascular agents, *etc.*) and environmental compounds (nicotine, *etc.*) (Xu *et al.*, 2002; Lovlie *et al.*, 1996; Johansson *et al.*, 1996). Differences in copy number of chemokines are likely to affect responses to infection, and influence the cause of autoimmune and inflammatory diseases, as well as alter the patient reaction to treatment (Townson, Barcellos & Nibbs, 2002). Therefore, an established fast and easy-to-use method for analysis of gene copy numbers is of major interest.

Horse coat colour genetics

Horse genomics was recognised as essential to veterinary medicine and horse breeding several years after genome analysis programs for cattle, pig, sheep and chicken were initiated. Although horse is not an animal reared primarily for food and production, the industry has a strong economic impact through its role in sports and recreation.

Horse coat colour genetics is of great interest for horse breeders – different coat colour tests could be used for instance for parentage testing or just for selection of breeding animals with preferred coat colour genotype. During recent years,

comparative genomics and whole genome scanning approaches have been used to develop DNA tests for a variety of horse coat colours. For instance, a missense mutation at the *extension* locus (*MC1R*) has been associated with the recessive chestnut colour - described as entirely red coat lacking black or brown pigmentation (Marklund *et al.*, 1996). Substitutions in the corresponding transmembrane region of *MC1R* in humans have been associated with red hair and fair skin. Another example is the *agouti* locus, where an 11 bp deletion has been found to be completely associated with recessive black coat colour in horses (Rieder *et al.*, 2001). In addition, the *cremello* locus in horses causes a cream dilution seen in the golden body colour of palominos and buckskins, as well as the ivory body colour of cremellos and perlinos. This locus has been identified as *MATP* gene, encoding a transporter protein in which human and mouse mutations have been associated with hypopigmentation (Locke *et al.*, 2001; Mariat *et al.*, 2003). Alleles at this locus function in combination with alleles at the *extension* and *agouti* locus, creating interesting series of diluted colour phenotypes.

There are other genes that have been recently mapped in horses, but the mutations behind these colour phenotypes are yet to be identified. Some examples are genes behind grey (Henner *et al.*, 2002; Locke *et al.*, 2002; Swinburne, Hopkins & Binns, 2002), tobiano (Brooks, Terry & Bailey, 2002) and appaloosa (Terry *et al.*, 2004) coat colours. So far, the total amount of horse markers, as well as amount of physically mapped horse markers have been the major limiting factor for identifying genes behind different horse traits. The overall growing interest for animal genetics and its practical use in breeding has recently raised several initiatives for speeding up the development of horse genomics.

Greying with age in horses

Greying with age is a dominantly inherited coat colour that occurs in many different horse breeds. In some horse breeds, like Arabian and Lippizaner, grey is the most common coat colour, but grey horses exist also among Icelandic and Thoroughbred horses, for instance. These horses are born coloured and progressively turn grey, reaching a uniform white colour.

This colour phenotype has been closely linked to melanotic melanoma development – approximately 80% of aged grey horses have the melanoma, while melanomas in the non-grey horses are generally believed to be very rare. The true nature of these melanomas is still uncertain, but they are believed to be non-malignant, with a potential to develop into malignant ones over the years (Sutton & Coleman, 1997). Despite this fact, melanomas can affect the welfare of grey horses by causing discomfort when located in sensitive regions of body and reaching a big volume. The location of melanomas in grey horses is very distinctive, with the most common sites being the ventral surface of the tail and the perianal region, but they are also quite frequently found on the head, neck, parotid gland and external genitalia as well as affecting internal organs (Swinburne, Hopkins & Binns, 2002). Histologically, the grey melanomas are dermal accumulations of melanocytes and melanophages. Therefore, they are most commonly compared to human blue naevus – a type of benign mole. A simple

storage/transport disorder has also been proposed as one of the reasons behind the melanoma development (Sutton & Coleman, 1997).

The *Grey* locus has been mapped to horse chromosome 25 (ECA25) by three independent groups (Henner *et al.*, 2002; Locke *et al.*, 2002; Swinburne, Hopkins & Binns, 2002). By using whole-genome scan approach, they positioned the *Grey* gene close to the *TXN* gene. The respective chromosomal regions in the human (HSA9q) and the mouse (MMU2 and MMU4) harbour no obvious candidate genes with a function in pigmentation or melanoma development. Therefore, fine-mapping of the gene responsible for the greying with age phenotype in horses is necessary, but needs new markers in the candidate region. Identifying the gene and the mutation behind this fascinating phenotype could shed light on important aspects of melanocyte biology. It may also pave the way for the development of a cure for not only grey horses, but also for certain types of human melanomas.

Objectives of the thesis

The objectives of this thesis were to give an overview of coat colour genetics in farm animals and to study further two specific coat colour phenotypes – the dominant white phenotype in pigs and greying with age phenotype in horses. The specific aims of the projects were:

- To establish a sensitive and easy-to-use method for genotyping the *Dominant White/KIT* locus in domestic pigs and use it for analysing this locus in different domestic pig breeds
- To establish a method for analysis of gene copy numbers by using porcine *KIT* locus as a model and compare the established method with other similar available techniques
- To perform a comparative linkage mapping of greying with age phenotype in horses with the ultimate goal to identify the underlying mutation

Main methods used in the present studies

Pyrosequencing

PyrosequencingTM is a method mainly used for SNP (Single Nucleotide Polymorphism) analysis. The method was first described by Ronaghi and co-workers and has been developed further for several different applications (Ronaghi *et al.*, 1996; Ronaghi, Uhlen & Nyren, 1998).

Pyrosequencing is a simple primer extension reaction based on a sequencing-by-synthesis principle and is performed by the use of an enzymatic complex (Fig. 2). The first step is to PCR amplify the region of interest with one of the PCR primers being biotinylated. Single-stranded PCR fragment is obtained by a subsequent denaturation of the PCR-fragment while 'holding on' to the fragment by biotin-streptavidin bonding. The sequencing primer is then annealed to the vicinity of the SNP of interest and by adding nucleotides one by one, DNA polymerase performs the primer extension. Other enzymes in the enzymatic complex, like sulfurylase and luciferase are needed for turning the nucleotide incorporation signal into light emittance. Apyrase is needed for degradation of the intermediate ATP (adenosine triphosphate) and unincorporated nucleotides before adding the next nucleotide to the reaction (Fig. 2).

Pyrosequencing has been used as a method for SNP genotyping and quantitative analysis of SNPs in the papers included in this thesis. The first paper describes a use of pyrosequencing for quantifying an SNP position present in one, two or three copies of the gene. The second paper is presenting more advanced experimental design by using competitive PCR for amplifying end-regions of different copies of a gene and quantifying the first bases where the different copies show a sequence difference as SNPs. In the third paper, pyrosequencing is used as a method for SNP genotyping.

Minisequencing

Minisequencing is an SNP genotyping method described by Syvänen and co-workers (Syvanen *et al.*, 1990; Syvanen, Sajantila & Lukka, 1993). This method has previously been used to distinguish between one, two and three copies of an allele on human chromosome 4 (Laan *et al.*, 1995) and to accurately quantify alleles present at frequencies ranging from 1 to 99% in pooled DNA samples (Olsson *et al.*, 2000).

Minisequencing is also a primer extension reaction performed by DNA polymerase, but the sequencing primer should be positioned immediately adjacent to the SNP position. The single-stranded PCR fragment is obtained by biotin-streptavidin bonding and the signal is obtained by incorporating labelled nucleotides at the SNP position. It is possible to use either fluorescently labelled or tritium labelled nucleotides, depending on the purpose of the experiment. Fluorescently labelled nucleotides enable genotyping on the microarray-type

slides, where the genotype is visualized in a certain colour combination. Tritium labelled nucleotides are useful for accurate quantification experiments and, since in this case we only incorporate a single labelled nucleotide per sample well, different alleles are analysed in different wells as different samples (*i.e.* if the SNP is an A/G, nucleotide A is added to one well and G to another well with the same template and sequencing primer). The results are measured and obtained as numeric values directly reflecting the relative amounts of the sequences in the sample.

Minisequencing was used in the first paper of this thesis for evaluating the power of pyrosequencing for resolving small differences in the ratio of different alleles at the porcine *KIT* locus.

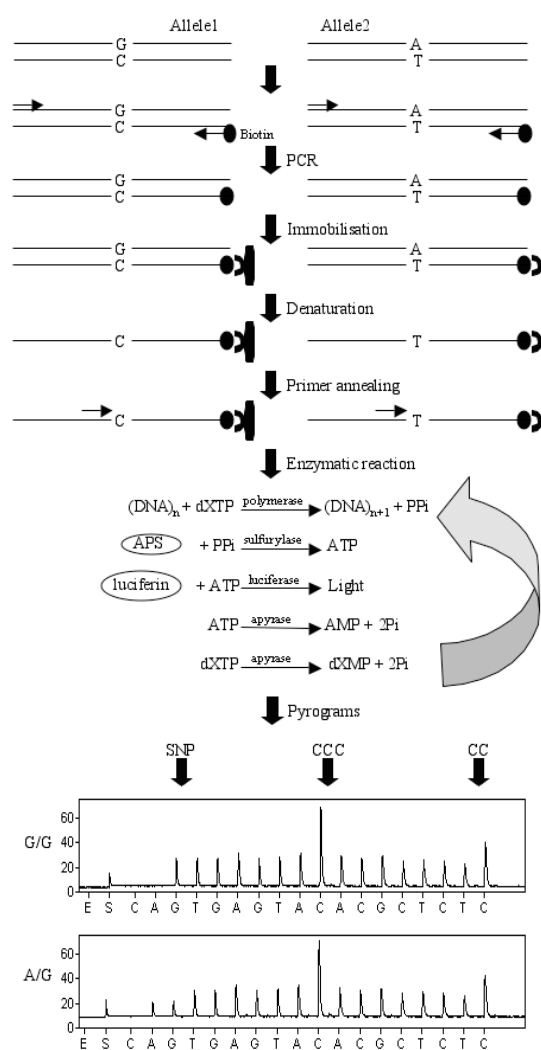


Fig. 2. Principle of pyrosequencing. The region including the SNP is amplified by PCR using one biotinylated primer. PCR products are immobilised to streptavidin-coated beads and single-stranded PCR product is obtained by denaturation. Sequencing primer is annealed to the vicinity of SNP and extended nucleotide-by-nucleotide in a series of enzymatic reactions. APS (adenosine 5'-phosphosulfate) and luciferin are substrates used by sulfurylase and luciferase, respectively. The SNP peaks (A and/or G), as well as three and two C-nucleotide stretches in the sequence are indicated with the arrows.

Real-time PCR (TaqMan)

The TaqMan kinetic PCR method is based on the FRET (Fluorescent Resonance Energy Transfer) technique, which exploits the ability of a quencher molecule to quench the fluorescent emission from a reporter molecule when they are in the proximity of each other. An oligonucleotide probe (TaqMan probe) labelled at the 3'-end with a quencher molecule and at the 5'-end with a reporter fluorescent molecule, hybridises to the DNA fragment that is being amplified. The nuclease activity of the DNA polymerase degrades the TaqMan probe during primer extension. The fluorescent emission from the reporter molecules that have been separated from the quenchers can be monitored in real time, and the accumulation of PCR product can be measured (Fig. 3A).

In the first paper of this thesis, TaqMan assay was used for comparative quantification of the porcine *KIT* gene copy number versus single copy control gene (in this case *ESR*, estrogen receptor gene).

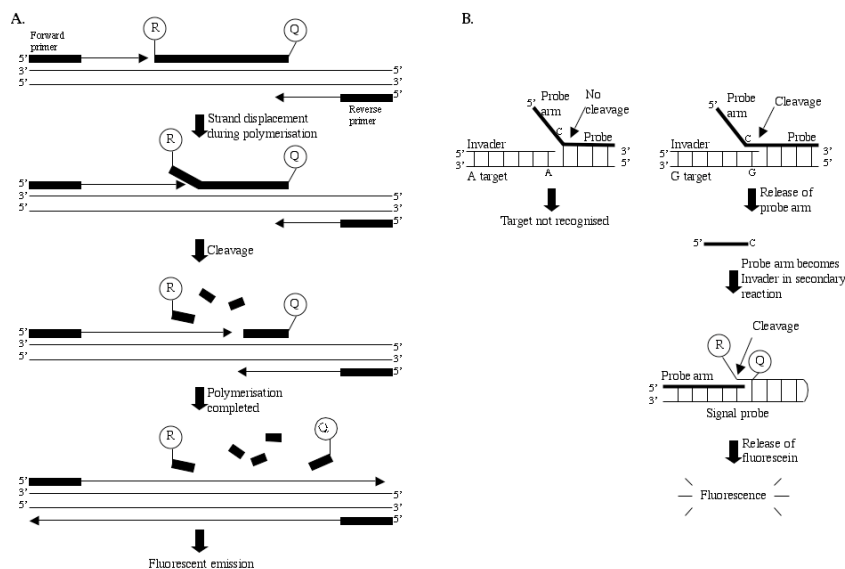


Fig. 3. A. Principle of TaqMan. B. Principle of Invader assay. R-reporter fluorescent molecule, Q-quencher molecule.

Invader assay

Invader assay is considered to be useful for gene copy number quantifications, because it involves a linear amplification of the target sequence, whereas a PCR reaction involves an exponential amplification (Hall *et al.*, 2000; Lyamichev *et al.*, 2000; Nevilie *et al.*, 2002; Olivier *et al.*, 2002). The assay is based on an “invading” oligonucleotide, upstream from the cleavage site, and a partially overlapping downstream probe forming a specific structure when bound to a complementary DNA template (Fig. 3B). This structure is recognised and cut at a specific site by Cleavase[®] enzymes, resulting in release of the 5’-arm of the probe-oligonucleotide. This fragment then serves as the “Invader” oligonucleotide in a secondary reaction, resulting in specific cleavage of the signal probe by Cleavase enzymes. Fluorescence signal is generated when the signal probe, FRET labelled with dye molecules, is cleaved. When the targeted sequence is not present in the sample tested, cleavage does not occur, the probe arm is not released, and a fluorescent signal is not generated. The signal generated by the variable signal probe is related to the signal generated by an internal control signal probe, thus allowing a copy number call to be made.

The Invader technology requires more genomic DNA and a careful quantification of the DNA template, but does not require prior knowledge of the duplication breakpoint. In a study of human *CYP2D6* copy numbers based on Invader technology, it was reported that the method could resolve individuals with one, two and three copies, but not individuals with more than three copies (Nevilie *et al.*, 2002).

Invader assay was used in the second paper of this thesis for evaluating the accuracy of the copy number test using pyrosequencing. Once again, *ESR* was used as a single copy control gene.

Linkage analysis

Linkage analysis is an approach for genetic localisation of genes of interest. It is the only method that can be used to provide the chromosomal localisation of a gene underlying a phenotypic trait before the molecular basis has been revealed. The principle of this analysis could shortly be the following: if we assume a parent being double heterozygous for a pair of loci, then this individual can transmit four types of gametes – “parentals” that correspond to the linkage phase of the grandparental gametes, and “recombinants” that are generated by reassortment of grandparental alleles. By estimating the proportion of parental versus recombinant gametes, we can estimate the recombination fraction between the analysed loci. If two loci are located on different chromosomes they will assort independently, whereas a recombinant gamete, with respect to two loci located on the same chromosome, requires a crossing-over event between these two loci. Since the occurrence of crossing-over event is more likely the farther the two loci are from each other, as a consequence - the closer the two loci are to each other the more drops the frequency of recombinant gametes. On the other hand, over long chromosomal distances it is likely that there can be multiple cross-overs taking

place, leading to double-recombinants. Therefore, a mapping function – the relationship between the observed recombination rate for a pair of loci and the actual genetic distance – might be useful for transforming the recombination rate between the two loci into a genetic distance. One of the most commonly used mapping functions for mammalian genomes is Kosambi's mapping function. In this way we can identify the groups of linked loci (*i.e.* two-point analysis), but to determine the most likely order and recombination rates between markers we need likelihood calculations using linked markers jointly (*i.e.* multipoint analysis). The likelihood of genotypic data is calculated for different possible marker orders. For each order, the recombination rates between adjacent markers that maximize the likelihood under the corresponding order are determined. The order and the associated recombination rates maximizing the likelihood are traditionally accepted as the correct order if the odds versus alternative orders is >1000 . Such calculations can be performed with, for instance, CRIMAP (Green, Falls & Crooks, 1990) and LINKAGE packages (Lathrop & Lalouel, 1988).

Linkage analysis was used in the third paper of this thesis for assigning new markers to horse chromosome 25, to reveal unlikely double recombinants given our dense map and to establish a linear order of markers in relation to *Grey*.

Results and discussion

Unexpectedly high allelic diversity at the porcine *KIT* locus (Paper I)

Despite a strong selection for white colour in several pig breeds for at least 100 years, breeders have not been able to completely fix the desired phenotype. It was known that the dominant white phenotype in pigs is caused by two mutations in the *KIT* gene – one duplication and one splice mutation. Before this study, four alleles had been identified at the porcine *Dominant white/KIT* locus: the recessive *i* allele for the normal colour, the semidominant I^P allele for the Patch phenotype, the fully dominant *I* allele for the white phenotype and I^{Be} for the dominant Belt phenotype (Fig. 4). This high allelic diversity at porcine *KIT* locus makes it hard to genotype – for instance, the difference between I/I^P , I/i and I/I genotypes is that the ratio between the splice mutation and the normal form at the first nucleotide of intron 17 is 25, 33 and 50%, respectively.

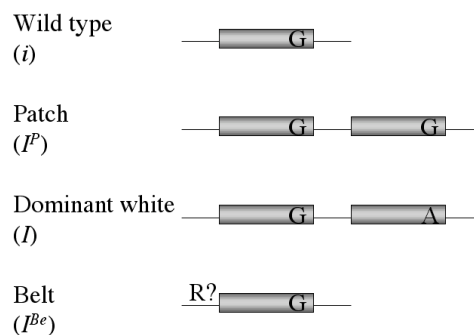


Fig. 4. Schematic description of *Dominant white/KIT* alleles in pigs. The duplication is ~450 kb. G and A reflect the normal and the splice mutation, respectively, at the first nucleotide of intron 17. R? indicates the postulated regulatory mutation in the *Belt* allele. The relative order of *KIT* copies (with and without splice mutation) has not been established yet.

In this study, minisequencing and pyrosequencing were applied for quantitative analysis of the number of copies with the splice form versus normal. By analysing the observed ratio and segregation data in a Large White x Wild Boar intercross pedigree, four different *KIT* alleles in eight white females used as founder animals were identified. The splice versus normal copy ratios among these females were: 25% splice for I/I^P genotype, 33% splice for I/I^{Be} genotype, 40% splice for animals heterozygous for a new allele with 3 *KIT* copies and 50% splice for I/I genotype. The females showing 40% ratio were identified by segregation analysis as

heterozygous for a new variant of dominant white allele designated I^2 with three *KIT* copies, only one of which carrying the splice mutation. The revealed extensive allelic diversity was further studied in commercial white populations and evidence for a sixth allele at the *Dominant white/KIT* locus was obtained. Four Large White animals showed a significantly higher ratio of the splice mutation (~60%) than any of the genotype combinations formed by the already described alleles. We postulated that these animals have one *KIT* allele with three gene copies of which two carry the splice mutation (I^3).

This variability in *KIT* copy number was confirmed by quantitative real-time PCR analysis (TaqMan). By comparing the amplification of *KIT* to a single-copy control gene *ESR* among 10 founder, 23 F_1 and 178 F_2 animals from the pedigree, as well as 4 Large White animals showing ~60% of splice ratio, a highly significant correlation between the *ESR/KIT* ratio and the predicted *KIT* copy number was found.

This study indicated clearly that the porcine *Dominant white/KIT* locus is genetically unstable. The reason for such a genetic instability is most likely the large size of the duplication (~450 kb) and a very high sequence identity (>99%) between the normal and duplicated copy of the gene. These characteristics of *KIT* locus give a substantial ground for generation of new alleles by unequal crossing-over and possibly by gene conversion. Furthermore, this study has shed some light to the reasons behind white colour in domestic breeds being still not fixed and causing a cost in pig breeding, despite of a strong selection over a long time.

One of the objectives of this study was to compare the power of minisequencing and pyrosequencing for resolving small differences in the splice ratio. Minisequencing has previously been reported as sufficient to distinguish between one, two and three copies of an allele (Laan *et al.*, 1995). Since our results show a very good agreement between results obtained with these two methods, pyrosequencing method also provides an excellent quantification of the incorporated nucleotides. Neither real-time PCR nor quantification of PCR-RFLP (Restriction Fragment Length Polymorphism) fragments, used in previous studies, have been able to resolve the allelic diversity at the porcine *KIT* locus to the same extent as reported in this paper.

A sensitive method for copy number analysis of duplicated genes (Paper II)

In this study, a sensitive method for analysing variation in gene copy number based on pyrosequencing and its application to genetic analysis of the duplicated *Dominant white/KIT* locus in pigs is described.

A tandem duplication has one unique DNA sequence not present on non-duplicated chromosomes, called the duplication breakpoint. This is the region where the 3'-end of the duplication is followed by the 5'-end of the next copy. By taking advantage of this feature, a competitive PCR assay amplifying breakpoint

regions from duplicated and normal copies of *KIT* was designed. For this assay we used one shared forward primer and two reverse primers specific for the two different *KIT* copies (Fig. 5). Pyrosequencing using a shared sequencing primer was then performed to quantify the two copies based on the relative proportion of diagnostic nucleotides after the breakpoint. For calibrating the test, two porcine BAC (**B**acterial **A**rtificial **C**hromosomes) clones – one representing the normal copy and the other the duplicated copy, were used. A standard curve run with dilution series of 0% to 100% of the duplicated copy mixed with the normal copy of *KIT* revealed an obvious nonlinear amplification. This nonlinear amplification was a consequence of the initial experimental design being an end-point analysis, where the amplification of the more common amplicon becomes retarded as the concentration of the specific reverse primer falls more rapidly than the concentration of the reverse primer for the rarer amplicon. Therefore, we designed a competitive PCR assay, using tailed-primer approach. In this experimental design, a specific tail-sequence was added to the 5'-ends of both reverse primers and a third shared reverse primer, corresponding to the tail-sequence, was included. By using a considerably lower amount of the copy-specific reverse primers compared to the tail-specific reverse primer, our idea was to limit the role of copy-specific primers to the initiation of PCR reaction only. During the rest of the reaction, the two amplicons would then compete for the same forward and reverse primers, still reflecting the actual ratio of normal versus duplicated copies in the template by the saturation of the reaction. The standard curve run with this competitive assay design showed a very good linearity.

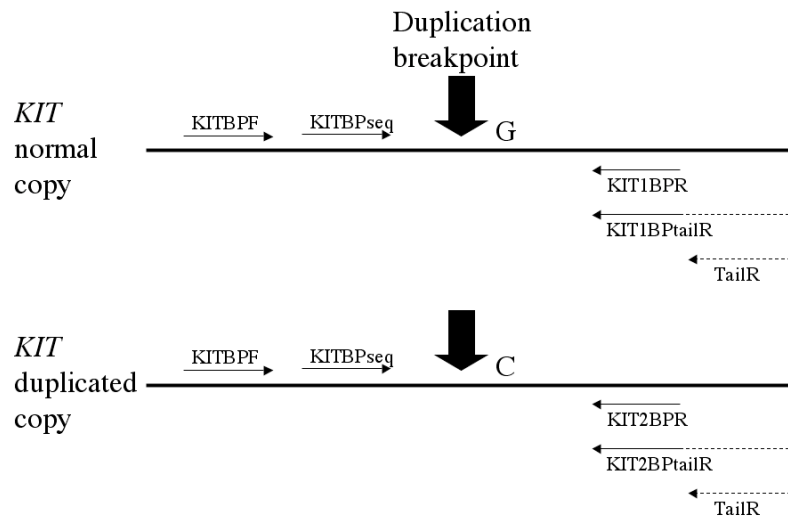


Fig. 5. Design of a DNA test for gene copy number quantification at the porcine *KIT* locus. The normal copy was PCR-amplified using the shared forward primer (KITBPF) and the specific reverse primer KIT1BPR, or alternatively a mix (1:100) of the two reverse primers KIT1BBtailR and TailR. Similarly, the duplicated copy was amplified using KITBPF and KIT2BPR or a mix (1:100) of KIT2BPtailR and TailR. Primer KITBPseq was used for pyrosequencing.

In order to evaluate the accuracy of the copy number test based on pyrosequencing, samples from previously described Large White x Wild Boar pedigree were analysed in parallel with Invader assay. There was an excellent correlation of the results obtained with pyrosequencing and Invader assay. Furthermore, our previous postulation of the existence of alleles with a *KIT* triplication was confirmed.

By combining the copy number test described in this paper with the splice test described in the previous paper, we were able to resolve more genotypes at the porcine *KIT* locus than before. For instance, analysing i/i , i/I^p and I^p/I^p genotypes for the splice test gives us exactly the same result – 0% splice, whereas the copy number test splits them into 2-, 3- and 4-copy groups, respectively. The combined results of splice and copy number tests for the Large White x Wild Boar pedigree were in good agreement with the expected theoretical outcome. This gave us confidence to analyse the *KIT* locus in commercial white populations, proving once again the high allelic diversity at the locus. Furthermore, an additional allele having a single copy of *KIT* with the splice mutation was revealed. This allele was denoted as I^s , since based on comparative data on null mutants in mice, it is assumed to be lethal in the homozygous condition.

This study provides an accurate and simple method for gene copy number analysis, while having prior sequence information of the duplication breakpoint. The combination of splice test and copy number test of porcine *Dominant white/KIT* locus described in the first two papers of this thesis provides an excellent aid for genotyping the *KIT* locus in pigs. Furthermore, we believe that the method described here can have a wide application in other species, as duplicated genes are common in many species and variation in copy numbers has important phenotypic effects.

Comparative linkage mapping of the grey coat colour gene in horses (Paper III)

In this study, a comparative linkage mapping approach was used to narrow down the localisation of the *Grey* locus in horses. This locus has been mapped to horse chromosome 25 (ECA25), close to the *TXN* gene. ECA25 is homologous to the human chromosome 9 (HSA9q) and mouse chromosomes 2 and 4 (MMU2, MMU4). Since there are not enough horse sequences and markers available for fine-mapping in the candidate region, we took advantage of human and mouse genomic sequence alignments. PCR primers were designed in conserved regions, usually exons, in order to amplify fragments with suitable size for sequencing (1-2 kb). Sequences obtained from these fragments were screened for SNPs showing heterozygosity in two grey Swedish Warmblood stallions. These stallions were known to be heterozygous *Grey* individuals (G/g), since they had both grey and non-grey offspring from matings to non-grey mares (g/g) and the grey phenotype is dominant. We identified eight new SNPs associated with genes on ECA25 and analysed them together with five previously identified microsatellite markers in more than 300 progeny from the two stallions. Since all the offspring had non-

grey mothers, it was possible to deduce their genotypes at the *Grey* locus from their phenotypes – grey being heterozygous (*G/g*), non-greys homozygous (*g/g*). A region between genes *NANS* and *ABCA1* was identified as an interval for *Grey* localisation. This region in humans is ~6.9 Mb and does not include any obvious candidate genes for pigmentation disorders or melanoma development. Furthermore, we observed no recombination between the *Grey*, *TGFBR1* and *TMEFF1* genes, the last two being 1.4 Mb apart from each other in the human genome. This indicates a quite large region without any recombination, considering the 168 informative meioses between *TGFBR1* and *TMEFF1* in the pedigree. Possible explanations for that could be either this region being physically smaller in the horse or lower recombination rate in this region of the genome.

A comparison between ECA25 and HSA9q revealed at least one interchromosomal rearrangement, since *FBP1* gene has been physically assigned to ECA23 instead of ECA25 (Fig. 6). Based on the physical assignment of *UCDEQ464* to ECA25q18-q19 (Lindgren *et al.*, 2001) and *XPA* to ECA25q12-q13 (Godard *et al.*, 2000) in combination with our linkage map, we could conclude that at least one inversion must have occurred in the horse or human lineages. Besides these rearrangements, there is a perfect linear gene order in the candidate region between ECA25 and HSA9q. Therefore, comparative mapping approach, for developing new markers, could be considered as successful in this case.

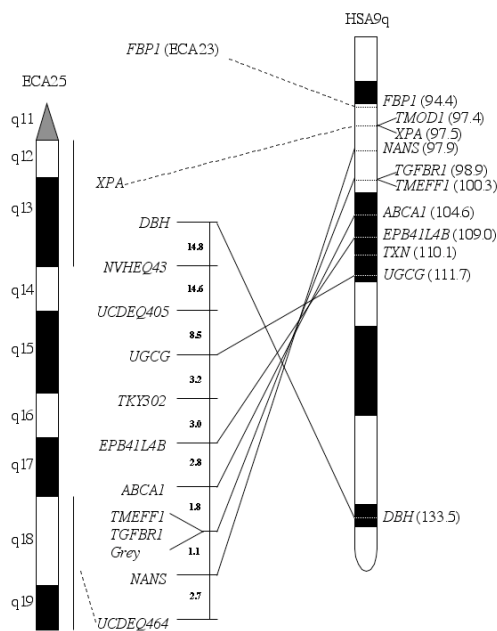


Fig. 6. Comparative map of horse (ECA25) and human (HSA9q) chromosomes. The bars next to ECA25 chromosome indicate the physical assignments of *UCDEQ464* and *XPA* based on fluorescent in situ hybridisation (Lindgren *et al.* 2001; Godard *et al.* 2000). Genetic distances between markers (in Kosambi cM) are shown to the left of the vertical bar with horse markers. Numbers in the brackets after locus designation in human indicate the position of these genes (Mb) in genomic contig sequences (UCSC, <http://genome.ucsc.edu>; 20-Oct-2004).

The defined region, assumed to harbour the human homologue for the *Grey*, does not reveal any obvious candidate genes for loss of hair pigmentation or melanoma development. Therefore, we expect the grey phenotype to be caused by

a novel gene that has previously not been associated with such phenotypes in any species. Interestingly, there are several publications reporting loss of heterozygosity on human chromosome 9q in melanomas and metastasis (Bogdan *et al.*, 2001; Morita *et al.*, 1998), indicating that HSA9q could comprise genes involved in melanoma development.

Future prospects

KIT gene in pigs

In this thesis I present accurate and easy-to-use methods for analysing copy number and the proportion of copies with and without splice mutation at porcine *Dominant white/KIT* locus. There has been a great interest among pig breeders for analysing the *KIT* locus in order to ensure homozygosity at this locus in white pig lines. This is important, since white pigs are usually crossed with coloured pigs like Hampshire (black with white belt) or Duroc (red), with the expectation of the offspring maintaining the white coat colour.

During these projects we have been interested to study pleiotropic effects of the *KIT* genotype on domesticated pigs. Since the product of the *KIT* gene is a key signal-transducing receptor with important roles in stem cell systems, it is possible that the genotype at the *Dominant white/KIT* locus influences several vital processes in the pig body. It has previously been reported that I^1/I^1 homozygotes among F_2 animals from the Large White x Wild Boar intercross pedigree have a lower number of white blood cells compared to I/i and i/i animals (Marklund *et al.*, 1998). We are currently proceeding more thorough studies for identifying the effects of *KIT* genotype on haematopoiesis, several blood parameters and fertility traits in different pig breeds (Johansson *et al.*, in preparation).

An interesting future aspect of porcine *Dominant white/KIT* project would be the determination of the order of copies with and without splice mutation. Since the splice mutation causes exon skipping, that is an in-frame mutation in this case, the protein encoded from this gene copy lacks parts of highly conserved catalytic loop and the activation loop. This mutation is expected to cause a receptor with normal ligand binding, but with impaired or absent tyrosine kinase activity that does not transmit *KIT* signalling. It is well-known that even if a duplication in a mammalian genome harbours the whole coding sequence of a gene, it usually does not include all the regulatory elements necessary for expression. Therefore, the order of the copies with and without splice mutation on the chromosome might be crucial for understanding the dominant white phenotype in pigs.

Another question that remained unsolved during this project was the difference between the wild allele (i) and the Belt allele (I^{Be}). Both of these alleles have only one copy of the gene and no splice mutation. With the two methods presented during these studies, the wild and the Belt allele remain undistinguished. By taking several mouse mutants (patched, rump-white) into consideration, it is most probable that the belted phenotype is caused by regulatory mutations between the *PDGFRA* and *KIT* genes, affecting *KIT* expression. We are currently applying AFLP (Amplified Fragment Length Polymorphism) together with long-range PCR strategy for screening this region for differences between pigs with wild and belted phenotypes.

From the evolutionary point of view it would be also interesting to use these methods presented here for a wider analysis of haplotype diversity at the porcine *KIT* locus in order to study the dynamic evolution of this locus. We are planning to screen a large number of progeny from matings between white (*I/I*) and coloured (*i/i*) parents for mutational events changing the copy number or the ratio of the splice mutation. Because this type of mating is very common in commercial pig production, it is possible to collect the sample size needed to detect mutations. This will allow us to estimate the frequency of unequal crossing-over as well as gene conversion in this system.

Concerning the methods developed during the studies presented here, it would be of great interest to apply them for analysing other loci in other model organisms. For instance, as mentioned before, there are several human genes known to exist in more than one copy and affecting a phenotype according to the copy number. Analysing a copy number of a gene affecting drug metabolism would be a big step towards more individualised treatment strategies. A major drawback of the methods presented here is the need for prior knowledge of the breakpoint sequence. Therefore, developing this pyrosequencing-based method further, possibly using a one-copy control gene for evaluating gene copy numbers, would be an interesting future task.

Grey coat colour gene in horses

Since it appears to be a fairly low recombination rate in the identified candidate region, it would be hard to achieve a high-resolution localisation by linkage analysis. Therefore, an **I**dentical **B**y **D**escent (IBD) mapping strategy is an attractive approach for refining the localisation of the *Grey* locus and eventually identifying the gene and the mutation behind this phenotype. This mapping strategy exploits the fact that the *Grey* mutation has probably occurred only once, thousands of years ago and thereafter spread to different breeds through many generations. Therefore, instead of obtaining information from recombination events during one generation, one could analyse recombination events that have accumulated during the transmission of the *Grey* from a common ancestor. Identifying the minimum shared haplotype among grey horses from different breeds is expected to provide a sufficient map resolution, which should lead to the identification of the gene and mutation responsible for this fascinating phenotype.

An important step forward during this project would be to utilise the existing horse BAC libraries – Texas A&M (TAMU), Institut de la National Recherche Agronomique (INRA) and Children’s Hospital of Oakland Research Institute (CHORI-241). It would be of great interest to screen these BAC libraries for the genes in the candidate interval and utilise the identified BAC clones for obtaining more horse sequence, like we have described before for porcine *PRKAG3* region (Amarger *et al.*, 2003). Horse sequence would enable us to compare grey and non-grey individuals for identifying new SNPs and to, ultimately, resequence the whole candidate region using long-range PCRs. Another important feature of

utilising BAC clones is the possibility to screen the candidate region for microsatellite markers. An informative microsatellite in the candidate interval would provide us a lot more information than biallelic SNPs.

Once the gene and the mutation behind the grey phenotype in horses are identified, one could take advantage of this knowledge and produce a transgenic mouse carrying the *Grey* mutation. This experiment would provide us important information about the development of this phenotype, in particular the mechanism behind such a difference between skin and hair pigmentation, as well as melanoma development. The true nature of grey horse melanomas is still unknown, but they are most often compared to human blue naevus. There are several types of inherited melanomas known in human population and families with high melanoma incidence would provide us an excellent genetic material for studying phenotype similar to the one in the grey horses. Furthermore, identifying the gene and the mechanism behind grey horse melanoma would hopefully provide information necessary for not only developing cure for the grey horses, but for some type of melanomas in general.

In general, studying coat colour genetics in farm animals is of great interest because of its agricultural, biological and medical significance. Using farm animals for such studies gives us an advantage of doing test-matings and thereby dissect complexes of gene action and interaction, for instance present in coat colour phenotypes, piece by piece. Comparative gene mapping is a successful key activity in farm animal genomics, considering the limited finances for whole genome sequencing projects. In conclusion, genome research in farm animals is adding to our basic understanding of the genetic control of studied traits and the results are being applied in breeding programmes to reduce the incidence of disease, to improve product quality and production efficiency. Furthermore, the obtained knowledge might lead us to a better understanding of the genetic basis of some human diseases, with the ultimate goal to develop a cure and helping mankind.

References

- Amarger, V., Erlandsson, R., Pielberg, G., Jeon, J.T. & Andersson, L. 2003. Comparative sequence analysis of the PRKAG3 region between human and pig: evolution of repetitive sequences and potential new exons. *Cytogenetic and Genome Research* 102, 163-172.
- Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews. Genetics* 2, 130-138.
- Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. & Eichler, E.E. 2002. Recent segmental duplications in the human genome. *Science* 297, 1003-1007.
- Barsh, G.S. 1996. The genetics of pigmentation: from fancy genes to complex traits. *Trends in Genetics* 12, 299-305.
- Bernstein, A., Chabot, B., Dubreuil, P., Reith, A., Nocka, K., Majumder, S., Ray, P. & Besmer, P. 1990. The mouse W/c-kit locus. *Ciba Foundation Symposium* 148, 158-172.
- Bogdan, I., Xin, H., Burg, G. & Boni, R. 2001. Heterogeneity of allelic deletions within melanoma metastases. *Melanoma Research* 11, 349-354.
- Brooks, S.A., Terry, R.B. & Bailey, E. 2002. A PCR-RFLP for KIT associated with tobiano spotting pattern in horses. *Animal Genetics* 33, 301-303.
- Broudy, V.C. 1997. Stem cell factor and hematopoiesis. *Blood* 90, 1345-1364.
- Chabot, B., Stephenson, D.A., Chapman, V.M., Besmer, P. & Bernstein, A. 1988. The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. *Nature* 335, 88-89.
- Eichler, E.E. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends in Genetics* 17, 661-669.
- Fleischman, R.A. & Mintz, B. 1979. Prevention of genetic anemias in mice by microinjection of normal hematopoietic stem cells into the fetal placenta. *Proceedings of the National Academy of Sciences of the United States of America* 76, 5736-5740.
- Fleischman, R.A., Saltman, D.L., Stastny, V. & Zneimer, S. 1991. Deletion of the c-kit protooncogene in the human developmental defect piebald trait. *Proceedings of the National Academy of Sciences of the United States of America* 88, 10885-10889.
- Geissler, E.N., Ryan, M.A. & Housman, D.E. 1988. The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. *Cell* 55, 185-192.
- Giebel, L.B. & Spritz, R.A. 1991. Mutation of the KIT (mast/stem cell growth factor receptor) protooncogene in human piebaldism. *Proceedings of the National Academy of Sciences of the United States of America* 88, 8696-8699.
- Giuffra, E., Evans, G., Tornsten, A., Wales, R., Day, A., Looft, H., Plastow, G. & Andersson, L. 1999. The Belt mutation in pigs is an allele at the Dominant white (I/KIT) locus. *Mammalian Genome* 10, 1132-1136.
- Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, O., Jeon, J.T. & Andersson, L. 2000. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154, 1785-1791.
- Godard, S., Vaiman, A., Schibler, L., Mariat, D., Vaiman, D., Cribsu, E.P. & Guerin, G. 2000. Cytogenetic localization of 44 new coding sequences in the horse. *Mammalian Genome* 11, 1093-1097.
- Green, P., Falls, K. & Crooks, S. 1990. *Documentation for CRI-MAP*. Version 2.4. Washington University School of Medicine, St Louis, MO.
- Hall, J.G., Eis, P.S., Law, S.M., Reynaldo, L.P., Prudent, J.R., Marshall, D.J., Allawi, H.T., Mast, A.L., Dahlberg, J.E., Kwiatkowski, R.W., de Arruda, M., Neri, B.P. & Lyamichev, V.I. 2000. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proceedings of the National Academy of Sciences of the United States of America* 97, 8272-8277.

- Henner, J., Poncet, P.A., Guerin, G., Hagger, C., Stranzinger, G. & Rieder, S. 2002. Genetic mapping of the (G)-locus, responsible for the coat color phenotype "progressive greying with age" in horses (*Equus caballus*). *Mammalian Genome* 13, 535-537.
- Hirooka, H., de Koning, D.J., van Arendonk, J.A., Harlizius, B., de Groot, P.N. & Bovenhuis, H. 2002. Genome scan reveals new coat color loci in exotic pig cross. *Journal of Heredity* 93, 1-8.
- Horvath, J.E., Bailey, J.A., Locke, D.P. & Eichler, E.E. 2001. Lessons from the human genome: transitions between euchromatin and heterochromatin. *Human Molecular Genetics* 10, 2215-2223.
- Hubbard, S.R., Wei, L., Ellis, L. & Hendrickson, W.A. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* 372, 746-754.
- Jackson, I.J. 1997. Homologous pigmentation mutations in human, mouse and other model organisms. *Human Molecular Genetics* 6, 1613-1624.
- Johansson, M., Ellegren, H., Marklund, L., Gustavsson, U., Ringmar-Cederberg, E., Andersson, K., Edfors-Lilja, I. & Andersson, L. 1992. The gene for dominant white color in the pig is closely linked to ALB and PDGRFRA on chromosome 8. *Genomics* 14, 965-969.
- Johansson, I., Lundqvist, E., Dahl, M.L. & Ingelman-Sundberg, M. 1996. PCR-based genotyping for duplicated and deleted CYP2D6 genes. *Pharmacogenetics* 6, 351-355.
- Johansson Moller, M., Chaudhary, R., Hellmen, E., Hoyheim, B., Chowdhary, B. & Andersson, L. 1996. Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* 7, 822-830.
- Kerje, S., Lind, J., Schutz, K., Jensen, P. & Andersson, L. 2003. Melanocortin 1-receptor (MC1R) mutations are associated with plumage colour in chicken. *Animal Genetics* 34, 241-248.
- Kijas, J.M. & Andersson, L. 2001a. A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *Journal of Molecular Evolution* 52, 302-308.
- Kijas, J.M., Moller, M., Plastow, G. & Andersson, L. 2001b. A frameshift mutation in MC1R and a high frequency of somatic reversions cause black spotting in pigs. *Genetics* 158, 779-785.
- Kijas, J.M., Wales, R., Tornsten, A., Chardon, P., Moller, M. & Andersson, L. 1998. Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* 150, 1177-1185.
- Klungland, H., Vage, D.I., Gomez-Raya, L., Adalsteinsson, S. & Lien, S. 1995. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mammalian Genome* 6, 636-639.
- Laan, M., Gron-Virta, K., Salo, A., Aula, P., Peltonen, L., Palotie, A. & Syvanen, A.C. 1995. Solid-phase minisequencing confirmed by FISH analysis in determination of gene copy number. *Human Genetics* 96, 275-280.
- Lathrop, G.M. & Lalouel, J.M. 1988. Efficient computations in multilocus linkage analysis. *American Journal of Human Genetics* 42, 498-505.
- Legault, C. 1998. *The Genetics of the Pig*. CAB International, Oxon, UK. 51-69 pp.
- Lindgren, G., Swinburne, J.E., Breen, M., Mariat, D., Sandberg, K., Guerin, G., Ellegren, H. & Binns, M.M. 2001. Physical anchorage and orientation of equine linkage groups by FISH mapping BAC clones containing microsatellite markers. *Animal Genetics* 32, 37-39.
- Locke, M.M., Penedo, M.C., Bricker, S.J., Millon, L.V. & Murray, J.D. 2002. Linkage of the grey coat colour locus to microsatellites on horse chromosome 25. *Animal Genetics* 33, 329-337.
- Locke, M.M., Ruth, L.S., Millon, L.V., Penedo, M.C., Murray, J.D. & Bowling, A.T. 2001. The cream dilution gene, responsible for the palomino and buckskin coat colours, maps to horse chromosome 21. *Animal Genetics* 32, 340-343.

- Lovlie, R., Daly, A.K., Molven, A., Idle, J.R. & Steen, V.M. 1996. Ultrarapid metabolizers of debrisoquine: characterization and PCR-based detection of alleles with duplication of the CYP2D6 gene. *FEBS Letters* 392, 30-34.
- Lyamichev, V.I., Kaiser, M.W., Lyamicheva, N.E., Vologodskii, A.V., Hall, J.G., Ma, W.P., Allawi, H.T. & Neri, B.P. 2000. Experimental and theoretical analysis of the invasive signal amplification reaction. *Biochemistry* 39, 9523-9532.
- Lynch, M. 2002. Genomics. Gene duplication and evolution. *Science* 297, 945-947.
- Mariat, D., Taourit, S. & Guerin, G. 2003. A mutation in the MATP gene causes the cream coat colour in the horse. *Genetics, selection, evolution* 35, 119-133.
- Marklund, L., Johansson Moller, M., Hoyheim, B., Davies, W., Fredholm, M., Juneja, R.K., Mariani, P., Coppieters, W., Ellegren, H. & Andersson, L. 1996. A comprehensive linkage map of the pig based on a wild pig-Large White intercross. *Animal Genetics* 27, 255-269.
- Marklund, S., Kijas, J., Rodriguez-Martinez, H., Ronnstrand, L., Funa, K., Moller, M., Lange, D., Edfors-Lilja, I. & Andersson, L. 1998. Molecular basis for the dominant white phenotype in the domestic pig. *Genome Research* 8, 826-833.
- Morita, R., Fujimoto, A., Hatta, N., Takehara, K. & Takata, M. 1998. Comparison of genetic profiles between primary melanomas and their metastases reveals genetic alterations and clonal evolution during progression. *Journal of Investigative Dermatology* 111, 919-924.
- Nevilie, M., Selzer, R., Aizenstein, B., Maguire, M., Hogan, K., Walton, R., Welsh, K., Neri, B. & de Arruda, M. 2002. Characterization of cytochrome P450 2D6 alleles using the Invader system. *Biotechniques* 34, 40-43.
- Newton, J.M., Wilkie, A.L., He, L., Jordan, S.A., Metallinos, D.L., Holmes, N.G., Jackson, I.J. & Barsh, G.S. 2000. Melanocortin 1 receptor variation in the domestic dog. *Mammalian Genome* 11, 24-30.
- Oetting, W.S. & Bennett, D.C. Mouse Coat Color Genes. International Federation of Pigment Cell Societies. <http://ifpcs.med.umn.edu/micemut.htm>; (accessed 20-Oct-2004).
- Olivier, M., Chuang, L.M., Chang, M.S., Chen, Y.T., Pei, D., Ranade, K., de Witte, A., Allen, J., Tran, N., Curb, D., Pratt, R., Neefs, H., de Arruda Indig, M., Law, S., Neri, B., Wang, L. & Cox, D.R. 2002. High-throughput genotyping of single nucleotide polymorphisms using new bplex invader technology. *Nucleic Acids Research* 30, e53.
- Olsson, C., Waldenstrom, E., Westermark, K., Landegren, U. & Syvanen, A.C. 2000. Determination of the frequencies of ten allelic variants of the Wilson disease gene (ATP7B), in pooled DNA samples. *European Journal of Human Genetics* 8, 933-938.
- Rieder, S., Taourit, S., Mariat, D., Langlois, B. & Guerin, G. 2001. Mutations in the agouti (*ASIP*), the extension (*MC1R*), and the brown (*TYRP1*) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mammalian Genome* 12, 450-455.
- Robbins, L.S., Nadeau, J.H., Johnson, K.R., Kelly, M.A., Roselli-Rehffuss, L., Baack, E., Mountjoy, K.G. & Cone, R.D. 1993. Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* 72, 827-834.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* 242, 84-89.
- Ronaghi, M., Uhlen, M. & Nyren, P. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281, 363-365.
- Spillman, W.J. 1906. Inheritance of coat colour in swine. *Science* 24, 441-443.
- Sutton, R.H. & Coleman, G.T. 1997. Melanoma and the Greying Horse. *RIRDC Research Paper Series* 55, 1-27.
- Swinburne, J.E., Hopkins, A. & Binns, M.M. 2002. Assignment of the horse grey coat colour gene to ECA25 using whole genome scanning. *Animal Genetics* 33, 338-342.

- Syvanen, A.C., Aalto-Setälä, K., Harju, L., Kontula, K. & Soderlund, H. 1990. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* 8, 684-692.
- Syvanen, A.C., Sajantila, A. & Lukka, M. 1993. Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *American Journal of Human Genetics* 52, 46-59.
- Takeuchi, S., Suzuki, H., Yabuuchi, M. & Takahashi, S. 1996. A possible involvement of melanocortin 1-receptor in regulating feather color pigmentation in the chicken. *Biochimica et biophysica acta* 1308, 164-168.
- Terry, R.B., Archer, S., Brooks, S., Bernoco, D. & Bailey, E. 2004. Assignment of the appaloosa coat colour gene (LP) to equine chromosome 1. *Animal Genetics* 35, 134-137.
- Townson, J.R., Barcellos, L.F. & Nibbs, R.J. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *European Journal of Immunology* 32, 3016-3026.
- UCSC Genome Browser. <http://genome.ucsc.edu/>; (accessed 20-Oct-2004).
- Vage, D.I., Klungland, H., Lu, D. & Cone, R.D. 1999. Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome* 10, 39-43.
- Walter, J. & Paulsen, M. 2003. The potential role of gene duplications in the evolution of imprinting mechanisms. *Human Molecular Genetics* 15, 215-220.
- Waskow, C., Paul, S., Haller, C., Gassmann, M. & Rodewald, H. 2002. Viable c-Kit(W/W) mutants reveal pivotal role for c-kit in the maintenance of lymphopoiesis. *Immunity* 17, 277-288.
- Xu, C., Goodz, S., Sellers, E.M. & Tyndale, R.F. 2002. CYP2A6 genetic variation and potential consequences. *Advanced drug delivery reviews* 54, 1245-1256.

Acknowledgements

I would like to express my gratitude to the following people:

Leif Andersson, my main supervisor. Your passion for science is contagious and that makes you one of the best teachers I've ever had. All these hours in your office have been for my own good and I'm grateful to you taking all that time for me. I will be looking forward to our future lively discussions!

Göran Andersson and Stefan Marklund. You have been the help that is always there when I needed it.

Gudrun Wieslander, our guardian angel. Thank you for all the help with paperwork – I could have gone crazy without you!

Ulla Gustafson and Erik Bongcam-Rudloff – your excellent help during the projects has been priceless.

Carolyn Fitzsimmons and Martin Braunschweig, my officemates. You two have given the most constructive criticism on my experiments and writing. I appreciate it.

All the present members of our group – Ana Fernandez, Meiyang Fang, Frida Berg, Ulrika Gunnarsson, Carl-Gustaf Thulin, Ellen Markljung, Per Wahlberg, Gabriella Lindgren and Emmanuelle Bourneuf.

Hee-Bok Park – baking cakes for you is a pleasure!

Lina Jacobsson – you are the most helpful person I know!

Susanne Kerje – you saved me from living in a tent...twice. I'll never forget that!

Nicolette Salmon Hillbertz – all these hugs have given me the strength that I desperately needed. Thank you for letting me pour my stress on your dogs!

Anne-Sophie van Laere – you have been a true friend all through this PhD. It's memories like wicked-hungry ducks at Fyrisån, weekends at the Swedish army base and Geishas at the Nations that make all these times special!

All previous members of our group – Siw Johansson, James Kijas, Gabriella Nyman, Jin-Tae Jeon, Örjan Carlborg, Elisabetta Giuffra, Maria Möller, Anna Törnsten. You were the ones that made me stay in this group!

Anna-Karin Fridolfsson, one of my previous supervisors. You were the reason I started my PhD!

Valerie Amarger, another one of my supervisors. You gave me a good base for future development and made me see 'the whole'. Thank you for that and all the discussions about life. I will never forget your ability to draw parallels between computers and men!

Ann-Christine Syvänen, Lotta Olsson, Graham Plastow, Andy Day, Kaj Sandberg and Sofia Mikko, my collaborators. Thank you for your expertise, for your time and effort.

All my students – Ulrica Svensson, Jessica Pettersson, Kajsa Himmelstrand, Malin Andersson, Karolina Holm, Maria Bergsland, Jimmy Larsson, Anna Lord, Magnus Lundberg, Amelie Johansson and Marta Marilli. I hope my guidance made at least a little bit of difference on your scientific journeys.

To everybody at the Evolutionary Biology group back in Estonia, especially Jüri Parik – you showed me the way to the PCR-world.

To the UGSBR-people and Catharina Svensson – this school was one of the best chances I've ever been given.

Pilvi – on hetki mis iial ei unune ja sõprusi mis iial ei purune. Hea on et oled olemas – kilomeetrid ja aastad meid juba ei lahuta!

Ema ja isa – ma tean et palju on neid asju ja tegusid, mis Teile mõistmatutena tunduvad. Samas võin ma alati kindel olla, et on olemas koht siinilmas, kus inimlikku mõistmist, rahu ja lihtsalt olemist jagub kuhjaga. See teadmine annab jõudu!

Kärt – ei teagi kohe kustkohast alustada – hea on et Sa oma rebasest õe eest hoolt kandsid. Veelgi parem on et Sa alati olemas oled olnud, just seal kus Su abi vaja on läinud. Peamine on elu huumoriga võtta ja kas on siinilmas midagi veel naljakamat kui meie ise?

Lasse – You're the best I have in my life! The way we make things right, the way we laugh at ourselves, the way we stand there for each other – that's our way, the best I know...