Genetic diversity of domestic sheep examples from Swedish and French populations

Christina Rochus

Genetic diversity of domestic sheep

Examples from Swedish and French populations

Christina Marie Rochus

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala, Sweden

Université Paris-Saclay
AgroParisTech
UFR Génétique, Élevage et Reproduction
Paris, France

Institut National de la Recherche Agronomique
Génétique Physiologie et Systèmes d'Elevage
Castanet Tolosan, France

Graduate School for Veterinary Medicine and Animal Sciences, SLU
École doctorale N^{o} 581, Agriculture, alimentation, biologie, environnement et santé, AgroParisTech, Université Paris-Saclay

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2017

Acta Universitatis agriculturae Sueciae

2017:52

The research presented in this doctoral thesis was conducted under the joint auspices of the Swedish University of Agricultural Sciences, AgroParisTech (Université Paris-Saclay) and Institut National de la Recherche Agronomique and is part of the Erasmus Mundus Joint Doctorate programme "European Graduate School in Animal Breeding and Genetics".

Akademisk avhandling som för vinnande av agronomie doktorsexamen kommer att offentligt försvaras i Ultuna, den 13 juni 2017.

Thèse présentée et soutenue à Ultuna, Suède, le 13 juin 2017. Spécialité de doctorat: Génétique animale.

Opponent:
J.A. Lenstra, Associate Professor, University of Utrecht

Examination committee / Betygsnämnd / Composition du Jury:
F. Pompanon, Professor, Université Grenoble Alpes
E. Ruis-Vilarrasa, Researcher,Växa Sverige
J. Slate, Professor, University of Sheffield

Reserve examiner:
T. Bergström, Associate Professor, Swedish University of Agricultural Sciences

Cover: Designed by the author

ISSN 1652-6880
ISBN (print version) 978-91-576-8877-4
ISBN (electronic version) 978-91-576-8878-1

Print: SLU Service/Repro, Uppsala 2017

Genetic diversity of domestic sheep: examples from Swedish and French populations

Abstract

Domestic sheep are raised for meat, milk and fibre production and are found all around the world in many types of environments. Sheep have been shown to be genetically diverse but this genetic diversity has not been fully described: there are still many sheep populations which have not yet been studied. The purpose of this thesis was to study genetic diversity in Swedish and French sheep breeds using high density marker arrays. Additional methods, including genotyping of microsatellite markers, and endogenous retroviruses and pedigree information were used to study Swedish sheep populations. Inbreeding and heterozygosity estimated in Gute sheep using the pedigree of the entire registered Swedish population and additionally microsatellite genotypes and pedigree from a sample of the population (N=94) indicated a breeding program with the purpose of reducing inbreeding. Studying genetic relationships among breeds by genotyping endogenous retroviruses indicated Klövsjö, Värmland, Finewool, Gute and Roslag sheep breeds had characteristics of primitive breeds (absence of retroviruses or presence of the specific retrovirus event enJSRV-7) although Finewool, Gute and Roslag sheep breeds had moderate frequencies of enJSRV-18 which is indicative of more modern sheep breeds. Studying variants in two coat colour genes, *ASIP* and *MC1R*, and their association with black coat colour revealed different selection histories in five Swedish sheep breeds studied. Studying the population structure of Dalapäls, Fjällnäs, Gotland, Gute and Klövsjö sheep, using high density SNP genotyping revealed that these breeds are genetically distinct breeds. When comparing with other European breeds and south west Asian breeds, they grouped with other north European short-tailed sheep breeds and they had generally accumulated more drift than breeds from other geographical areas. Studying 27 French breeds with high density genotypes revealed that French sheep populations harbour much of European sheep diversity in a small geographic area. Selective sweeps identified: selection hotspots, selection targets in many species; introgression of an adaptive allele; and allelic heterogeneity, which was confirmed with targeted resequencing of a coat colour gene, *MC1R*, in breeds under selection.

Keywords: sheep, genetic diversity, population structure, allelic heterogeneity, signatures of selection, coat colour, *ASIP*, *MC1R*, endogenous Jaagiekte retroviruses

Author's address: Christina Marie Rochus, SLU, Department of Animal Breeding and Genetics,
P.O. Box 7023, 750 07 Uppsala, Sweden
E-mail: Christina.Rochus@slu.se

Genetisk mångfald hos tamfår: exempel från svenska och franska populationer

Abstract

Tamfår föds upp för kött-, mjölk- och ullfiberproduktion och finns över hela världen i många typer av miljöer. Får har visat sig vara genetiskt variabla men denna genetiska mångfald har inte beskrivits fullständigt. Det finns fortfarande många fårpopulationer som ännu inte har studerats. Syftet med denna avhandling var att studera genetisk mångfald i svenska och franska fårraser med hjälp av genetiska markörer med hög densitet. Ytterligare metoder, inklusive genotypning av mikrosatellitmarkörer, och endogena retrovirus och härstamningsinformation användes för att studera svenska fårpopulationer. Inavel och heterozygotigrad som beräknats hos gutefår med hjälp av stamtavlan för hela den registrerade svenska populationen och dessutom mikrosatellitgenotyper och stamtavla från ett stickprov från populationen (N=94) indikerade ett avelsprogram med syfte att minska inavel. Studier av genetiska relationer bland raser genom genotypning av endogena retrovirus indikerade att Klövsjöfår, Värmlandsfår, Finullsfår, Gutefår och Roslagsfår hade egenskaper som är kännetecknande för primitiva raser (frånvaro av retrovirus eller närvaro av den specifika retrovirushändelsen enJSRV-7), även om Finullsfår, Gutefår och Roslagsfår hade moderata frekvenser av enJSRV-18 vilket indikerar mer moderna fårraser. Studier av varianter i två gener för pälsfärg, *ASIP* och *MC1R*, och deras association med svart pälsfärg avslöjade olika urvalshistorier i fem svenska fårraser som studerades. Studier av populationsstrukturen hos Dalapälsfår, Fjällnäsfår, Gotlandsfår, Gutefår och Klövsjöfår, genom användning av SNP-genotyper med hög densitet visade att dessa raser är genetiskt separata raser. När man jämförde med andra europeiska raser och sydvästasiatiska raser grupperade de sig med andra nordeuropeiska kortsvansfår och de hade generellt ackumulerat mer drift än raser från andra geografiska områden. Studier av 27 franska raser med genotyper med hög densitet visade att franska fårpopulationer hyser mycket av mångfalden hos europeiska får i ett litet geografiskt område. Selektiva svep identifierade: hotspots för selektion, urvalsmål i många arter; Introgression av en adaptiv allel; och allelisk heterogenitet, som bekräftades med målinriktad sekvensering av en gen för pälsfärg, *MC1R*, i raser under selektion.

Nyckelord: får, genetisk mångfald, populationsstruktur, allelisk heterogenitet, spår av selektion, pälsfärg, *ASIP*, *MC1R*, endogena Jaagiekte retrovirus

Diversité génétique du mouton domestique: exemple de populations suédoises et françaises

Résumé

Les moutons domestiques sont élevés pour la production de viande, de lait et de laine et sont retrouvés partout dans le monde dans des environnements variés. On a montré que les moutons présentaient une certaine diversité génétique, mais celle-ci n'a pas été complètement caractérisée: il existe encore de nombreuses populations de moutons qui n'ont pas été étudiées. L'objectif de cette thèse était d'étudier la diversité génétique dans des races de moutons suédoises et françaises, en utilisant puces de marqueurs de haute densité. De plus, dans les populations suédoises, d'autres méthodes ont été utilisées: le génotypage de marqueurs microsatellites, de séquences de rétrovirus endogènes et les données de pedigree. Dans la population *Gute*, l'estimation du niveau de consanguinité et d''hétérozygotie, en utilisant le pedigree de toute la population suédoise enregistrée, ainsi que le pedigree et des génotypages de microsatellites complémentaires d'un échantillon de la population (N = 94) ont indiqué un schéma de sélection orienté vers une réduction de la consanguinité. L'étude des relations génétiques entre les races grâce au génotypage de rétrovirus endogènes a montré que les races ovines Klövsjö, Värmland, Finewool, Gute et Roslag avaient des caractéristiques de races primitives (absence de rétrovirus ou présence de la séquence rétrovirale spécifique enJSRV-7) tandis que les races ovines Finewool, Gute et Roslag avaient des fréquences modérées de enJSRV-18, ce qui signe des races de moutons plus modernes. L'étude des variants de deux gènes de coloration de la toison, ASIP et MC1R, et leur association avec la couleur noire a révélé différentes histoires de sélection dans cinq races de moutons suédoises étudiées. L'étude de la structure des populations de moutons Dalapäls, Fjällnäs, Gotland, Gute et Klövsjö, grâce au génotypage de puces SNP à haute densité a révélé que ces races sont génétiquement distinctes. Leur comparaison avec d'autres races européennes et des races du Sud-Ouest asiatique les rapproche d'autres races de moutons à queue courte du nord de l'Europe et montre qu'elles ont accumulé plus de dérive génétique que les races provenant d'autres zones géographiques. L'étude de 27 races françaises avec des génotypages de puces à haute densité a révélé que les populations de moutons français abritent une grande partie de la diversité des moutons européens au sein d'une petite région géographique. Les balayages sélectifs ont montré : des points chauds de sélection, des cibles de sélection partagées par de nombreuses espèces, l'introgression d'un allèle adaptatif et une hétérogénéité allélique, qui a été confirmée par le reséquençage ciblé d'un gène de couleur de la toison, MC1R, dans des races sous sélection.

Mots-clés: mouton, diversité génétique, structure de la population, hétérogénéité allélique, signatures de sélection, couleur de la toison, *ASIP*, *MC1R*, rétrovirus endogènes Jaagiekte,

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Rochus, C.M., Johansson, A.M. (2017). Estimation of genetic diversity in Gute sheep: pedigree and microsatellite analyses of an ancient Swedish breed. *Hereditas* 154:4, doi:10.1186/s41065-017-0026-4.

II Mukiibi, R., Rochus, C.M., Andersson, G., Johansson, A.M. (2015). The use of endogenous retroviruses as markers to describe the genetic relationships among local Swedish sheep breeds. *Animal Genetics* 46, 220-223.

III Rochus, C.M., Westberg Sunesson, K., Jonas, E., Mikko, S., Johansson, A.M. Mutations in *ASIP* and *MC1R*: Dominant black and recessive black alleles segregate in native Swedish sheep populations (manuscript).

IV Rochus, C.M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., Servin, B. (2017). Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep (manuscript in bioRxiv https://doi.org/10.1101/103010).

V Rochus, C.M., Jonas, E., Johansson, A.M. Population structure of five native sheep breeds of Sweden estimated with high density SNP genotypes (manuscript).

Papers I and II are reproduced with the permission of the publishers.

Abbreviations

ABCG2	ATP-binding cassette sub-family G member 2
ADAMTS9	a disintegrin and metalloproteinase with thrombospondin motifs 9
AHCY	adenosylhomocysteinase
ALX4	homeobox protein aristaless-like 4
ASIP	agouti signalling protein
BCDO2	beta-carotene dioxygenase 2
BNC2	basonuclin 2
DNA	deoxyribonucleic acid
EBF2	early B-cell factor 2
EDN3	endothelin 3
EXT2	exostosin glycosyltransferase 2
enJSRV	endogenous Jaasiekte sheep retroviruses
FAANG	Fuctional Annotation of Animal Genomes
FDR	false discovery rate
FLK	extension of Lewontin and Krakauer's test for the heterogeneity of the inbreeding coefficient across loci that takes into account the kinship matrix between populations for biallelic markers
GDF-8	myostation (also known as *MSTN*)
GERP score	genomic evolutionary rate profiling score
hapFLK	haplotypic extension of FLK test
HWE	Hardy Weinberg Equillibrium
ITCH	itchy homolog E3 ubiquitin protein ligase
K	number of ancestral populations
KIT	KIT Proto-oncogene receptor tyrosine kinase
KITLG	KIT ligand
LCORL	ligand dependent nuclear receptor compressor like
MAF	minor allele frequency
MC1R	melanocortin 1 receptor
MITF	microphthalmia-associated transcription factor
MSRB3	methionine sulfoxide reductase B3

MSTN	myostatin (also known as *GDF-8*)
NCAPG	non-SMC condensin I complex subunit g
NPR2	natriuretic peptide receptor 2
OXCT1	3-Oxoacid CoA-Transferase 1
PALLD	palladin
PCA	principal component analysis
PCR	polymerase chain reaction
QTL	quantitative trait locus
RXFP2	relaxin/insulin like family peptide receptor 2
SNP	single nucleotide polymorphism
SOCS2	suppressor of cytokine signalling 2
SOX10	SRY-related HMG-box 10
TYRP1	tyrosinase related protein 1

1 Introduction

Maintaining genetic diversity in livestock is critical for facing future challenges like climate change, emerging diseases, and food security for an increasing human population (Groeneveld et al. 2010; Barker 2001). Sheep are a phenotypically diverse livestock species that are raised around the globe for meat, milk and fibre production. There are a reported 1 542 breeds of sheep: over half of which have an unknown risk status (FAO Commission on Genetic Resources for Food and Agriculture 2015). It is important that genetic diversity in sheep is estimated to identify unique breeds that may be at risk of extinction. Many breeds of sheep are local breeds, breeds which are uniquely and specifically adapted to certain environments (Groeneveld et al. 2010), and therefore sheep is a good model species for adaptation. In addition, livestock have shown to be under selection in similar regions of the genome for example, genes affecting morphology and coat colour (Fariello et al. 2014; Boitard et al. 2016; Petersen et al. 2013; Carneiro et al. 2016; Rubin et al. 2012). Studying selection signatures in sheep can therefore also be beneficial for other species. A number of unique sheep breeds can be found in both Sweden and France. Sweden has native breeds belonging to the north European short-tailed sheep group, breeds that are primitive, have variation in coat colour, and have been shown to be genetically unique and contribute highly to overall genetic diversity in sheep (Dýrmundsson & Niżnikowski 2010; Tapio et al. 2005). France is in a location where many different sheep groups meet in Europe (Kijas et al. 2012), and therefore French sheep populations likely represent much of European sheep diversity (Leroy et al. 2015). Genetic diversity has been previously studied in some of these breeds: some Swedish breeds were included in a study using microsatellite genotypes in North European short-tailed sheep (Tapio et al. 2005) and many French breeds were included in a recent study using microsatellite marker genotypes (Leroy et al. 2015). But further studies are needed to investigate the diversity within and across French and Swedish sheep breeds in more detail. The papers in this thesis estimated

genetic diversity and population structure and history and contributed to the further understanding of genetics of coat colour in domestic sheep. New genetic data sets (microsatellite genotypes, retrovirus genotypes, sequencing of genes under selection and high density SNP genotypes) were used to study Swedish and French sheep genetic diversity in more detail and higher resolution than previously and the data applied here also allowed for an in-depth comparison with other breeds from around the world.

2 Background

2.1 Sheep domestication

Domesticated approximately 10 000 years ago in the Anatolia region, sheep were first hunted by humans and over time became managed wild populations, then herded, and finally humans began to manage sheep breeding (Zeder 2008; Larson & Burger 2013). Today domestic sheep (*Ovis aries*) are a separate species from their progenitor, the Asian Mouflon (*Ovis orientalis*), which still exists to this day in south west Asia (Hiendleder et al. 2002). Sheep and goats were the first domesticated livestock species and since that time have spread around the globe: in Europe, there were two main routes of domestication that followed the Neolithic expansion of plant domestication (Zeder 2008). Animals adapted to being kept by humans show changes in behaviour and morphology (Mignon-Grasteau et al. 2005) as seen in sheep. Around 200 years ago sheep breeds began to be formed which led to fragmented populations and reduced genetic diversity (Taberlet et al. 2008). A few decades ago modern artificial selection began to further reduce genetic diversity (Taberlet et al. 2008). Today, sheep are raised and bred for: meat, milk, or fibre production; conservation purposes; or a combination of purposes. Sheep are found in extensive and intensive production systems and in hot to cold environments. Sheep also face challenges with coping with different diseases in different environments (FAO Commission on Genetic Resources for Food and Agriculture 2015).

2.2 Genetic diversity

Genetic diversity is defined as the variety of alleles and genotypes present in a population (Frankham et al. 2002). Conserving genetic diversity in livestock is important for improving production and quickly meeting future challenges

including food security, increasing demand for animal products, changing environments and emerging diseases (Groeneveld et al. 2010; Barker 2001). Every breed has a unique set of gene combinations which is due to: mutation and drift from geographic isolation and bottlenecks; artificial selection and adaption to climate; available nutrition; and diseases and parasites (Barker 2001). More homogeneous breeds with improved production and fixation of traits (such as coat colour) are a result of: pedigree recording, the formation of breeding organizations, advances in technology making the transportation of animals and the dissemination of genetic material easier, and more controlled production environments (FAO Commission on Genetic Resources for Food and Agriculture 2015; Groeneveld et al. 2010). Highly productive breeds have been replacing local breeds and unique gene combinations from some local breeds are being lost (Groeneveld et al. 2010). Maintained genetic diversity is important for adapting quickly to challenges (Andersson 2012). Natural and artificial selection pressures on sheep leading to their adaptation to different demands and environments makes domesticated sheep an interesting model species to study the evolution of populations (Andersson 2012). In addition, although not a focus of this thesis, local breeds have adapted to specific locations for thousands of years and are closely tied to culture and history and maintaining and studying genetic diversity of these breeds will help with understanding human history (Gandini & Villa 2003).

In order to maintain genetic diversity, it is important to measure and monitor it. One challenge with studying genetic diversity is the variety of methods, each with its advantages and disadvantages (which will be further reviewed in the discussion). In this thesis, genetic diversity was studied using methods to estimate inbreeding using pedigree or microsatellite genotypes, endogenous retrovirus retrotypes to study population structure, sequencing of genes under selection (coat colour genes *ASIP* and *MC1R*) and high density SNP genotypes to study population structure and signatures of selection.

2.3 ASIP and *MC1R*

Selection on coat colour happens early in domestication of animals because: coat colour and pattern variation appears rapidly; natural selection against variation in coat colour and pattern relaxes; and humans select animals with unique coat colours and patterns (Andersson 2001; Linderholm & Larson 2013). Over 300 genetic loci at over 150 genes affecting coat colour and pattern have been identified in mammals (Hubbard et al. 2010; Cieslak et al. 2011). In sheep, coat colour and pattern are a result of pigmentation of wool fibres with melanin. Cells called melanocytes produce two different types of

melanin which are transported in granules called melanosomes (Birbeck et al., 1956). Eumelanin or pheomelanin is produced after the binding of a melanocortin agonist or antagonist agouti signalling protein (*ASIP*) respectively to the melanocortin 1 receptor (*MC1R*) (Ducrest et al., 2008). Melanosomes transfer melanin to keratinocytes, a predominant cell type in the outer layer of skin (Birbeck et al., 1956) and keratinocytes can then incorporate melanin into fibre growth (Forrest et al., 1985). In a study of Asiatic sheep, white animals did not have any pigment type (eumelanin or pheomelanin), tan animals had small amounts of pheomelanin, light red, brown, red and dark brown animals had some combination of eumelanin and pheomelanin, and black individuals only had eumelanin present in their wool (Aliev et al., 1990). Even though sheep show a wide variety of coat colours and patterns, they are a good model for studying coat colour and pattern because many breeds are fixed at coat colour alleles. Genome wide linkage, association and selective sweep studies in sheep have identified many candidate genes associated with coat colour and pattern in sheep including *ASIP*, *EDN3*, *KIT*, *KITLG*, *MC1R*, *MITF*, *SOX10*, and *TYRP1* (Kijas et al. 2012; Fariello et al. 2014; M.-H. Li et al. 2014). In addition, causative mutations have been identified in *ASIP*, *MC1R* and *TYRP1* (Norris & Whan 2008; Smit et al. 2002; Fontanesi et al. 2011; Våge et al. 1999; Våge et al. 2003; Gratten et al. 2007). In *ASIP*, the white allele is an approximately 190 kbp tandem duplication which includes *ASIP*, *AHCY* and *ITCH* (Norris & Whan 2008), the recessive black allele is caused by either of two mutations, g.100-105del and g.5172T>A (Smit et al. 2002; Norris & Whan 2008), and the grey allele is caused by a different form of the duplication causing white coat colour (Fontanesi et al. 2011). In *MC1R* there is the dominant black allele which is caused by either of two mutation c.218T>A and c.361G>A (Våge et al. 1999; Våge et al. 2003). Not only does studying coat colour help determine the function of genes, but mutations associated with coat colour and pattern in modern animals can inform on population history: *MC1R* variation in Tibetan and Landrace pigs showed different levels of variation reflective of the different selection pressures (Liu et al. 2016); domestic chickens are descended from red jungle fowl however show some hybridization with grey jungle fowl because of variation in *BCDO2* (which is responsible for yellow legs in domestic chickens and grey jungle fowl) (Eriksson et al. 2008); and the Franches-Montages horse breed has a unique mutation for white coat colour traced to a horse born in 1957 (Haase et al. 2007).

2.4 Swedish sheep

Four of the five papers in this thesis focus on Swedish sheep breeds. These breeds belong to a group of sheep called the north European short-tailed sheep. These breeds are characterized by their short fluke shaped tails, variation in coat colour and pattern, their primitive phenotypes, dual-coated wool, robustness and prolificacy (Dýrmundsson & Niżnikowski 2010). North European short-tailed sheep were spread by Norse Vikings from the eighth to the middle of the eleventh century and now cover an area from Russia to Iceland (Dýrmundsson & Niżnikowski 2010). In studies estimating genetic diversity using microsatellite genotypes, north European short-tailed sheep, including some Swedish breeds, have been shown to be unique, with some breeds contributing more than expected to overall domestic sheep diversity, although many of these populations have small effective population size and have low within breed diversity (Tapio et al. 2005). Some of these (but not any Swedish) breeds were also included in the Sheep HapMap project where breeds from around the world were genotyped using a medium density SNP array with the data (available from http://www.sheephapmap.org). Many of the north European short-tailed sheep breeds today are local breeds having adapted to specific areas. In Sweden these small local populations are recorded and maintained as separate heritage breeds (Dýrmundsson & Niżnikowski 2010). Genetic diversity has been studied in some of these breeds including using microsatellite markers comparing north European short-tailed sheep (Tapio et al. 2005) and northern Eurasian breeds (Tapio et al. 2010). Two breeds (Gotland and Rya) were included in a study of retroviruses in some European, Asian and African breeds (Chessa et al. 2009). North European short-tailed sheep were also included in studies of inheritance of coat colour alleles including Gotland and Gute sheep that were believed to have two grey alleles different from the grey allele segregating in Icelandic sheep (Adalsteinsson et al. 1978).

2.5 French sheep

Sheep raised in France are a heterogeneous group: France borders many of the geographic areas associated with western European sheep population groups of the Sheep HapMap dataset (Kijas et al. 2012). The two French breeds included in this study were Lacaune and Rambouillet sheep, both of which belong to the Mediterranean sheep group (Kijas et al. 2012). In a recent study of 49 sheep breeds raised in France, population structure was studied using microsatellite genotypes and researchers confirmed that French sheep breeds show a structure consistent with influences from both Northern Europe, specifically British

breeds, and southern European breeds (Leroy et al. 2015). The study of French sheep diversity should therefore provide valuable insights on the establishment of genetic structure in European sheep, and in particular inform us on the history of sheep husbandry in Europe as they are likely to harbour a large part of European sheep diversity in a relatively small geographical region.

2.6 Gaps in knowledge

Sheep are a good model for studying adaptation and there are many unique breeds in Sweden and France that could offer more insight into genetic diversity and adaptation. More detailed information on population structure and genetic diversity of many of these breeds would assist to fill current gaps in knowledge. Especially in concern with the need to maintain genetic diversity to be able to meet future demands quickly and with breeds that are being kept for conservation purposes. Therefore the study of Swedish and French breeds is especially interesting as they cover a wide range of environments. Finally, high density data allows for comparisons with other world-wide breeds which allows for even further studies and comparisons of population structure.

3 Aims of the thesis

The aim of this thesis was to study population history, structure and genetic diversity of domesticated sheep with a particular focus on Swedish and French sheep populations.

Although not well represented in peer reviewed literature, many local Swedish sheep breeds are genetically unique. In addition, these breeds have small population sizes and many of these breeds are characterized in part by their variation within breed for coat colour and pattern. Further research is important for informing future discussions on the preservation of these breeds. The aim was to characterize and describe the population structure and history of these native Swedish sheep breeds.

Based on historical information of sheep domestication routes in Europe, studies on population structure of European sheep, and a study of population structure in French sheep using low density DNA information, there is evidence that there is a high level of genetic diversity in French sheep populations. This is also seen phenotypically as French sheep are raised for a variety of purposes and thrive in different farming systems. French sheep populations offer a powerful model for the study of adaptation. The aim was to study population structure, history and genetic diversity in these sheep using high density DNA marker information.

Specific aims:

- Study genetic diversity and population structure in Swedish breeds of sheep using low density DNA marker information and pedigree information.
  - Study Gute sheep (an ancient Swedish breed of sheep) inbreeding and genetic diversity using whole population pedigree information and seven microsatellite markers.
  - Study population structure of five Swedish sheep breeds using endogenous retroviruses as DNA markers.

- Study genotypic variation in two coat colour genes (*ASIP* and *MC1R*) and phenotypic variation in Swedish sheep.
- Study genetic diversity and population structure in Swedish and French breeds of sheep using high density DNA marker information (500 000 SNPs).

4 Summary of studies

This thesis is comprised of five papers in which genetic information was employed to study genetic diversity, and population structure and history of Swedish and French sheep breeds. In paper I, inbreeding and genetic diversity were estimated in a local Swedish breed using pedigree data and microsatellite genotypes. In paper II, the population structure and history of five native Swedish breeds was studied using retrovirus genotypes. In paper III, diversity and population history was studied in seven native Swedish breeds by studying variants in the exons of two coat colour genes, *ASIP* and *MC1R*. In paper IV, population structure, signatures of selection and allelic heterogeneity were studied in 27 French sheep breeds using high density SNP genotypes. Finally, in paper V, population structure and history and diversity were studied in five local Swedish breeds.

4.1 Materials and Methods

4.1.1 Estimation of genetic diversity in Gute sheep: pedigree and microsatellite analyses of an ancient Swedish breed

The pedigree for all registered Gute sheep in Sweden, with anonymous animal identification, was provided by Elitlamm, the recording program for sheep in Sweden. Blood samples were collected by a trained technician and DNA was extracted using DNeasy mini kit for QIAsymphony robot (Qiagen®, Hilden, Germany) for 100 Gute sheep from a total of 13 flocks. Animals were genotyped using eight DNA markers (AME, INRA005, INRA023, INRA063, INRA172, MAF214, MAF65 and McM527) at the Department of Animal Breeding and Genetics laboratory (SLU, Uppsala, Sweden). Pedigree data for the 100 genotyped individuals was also provided by Elitlamm.

The whole population pedigree was used to estimate population parameters including: generation intervals, effective number of founders, effective number of ancestors, effective number of founder genomes and marginal contributions of ancestors using Pedig (Boichard 2002); generation intervals, coefficient of inbreeding and average coancestry by birth year, pedigree completeness indices and number of complete generation equivalents were estimated with EVA (Meuwissen & Luo 1992; Berg et al. 2006); and effective population size was calculated as one divided by double the annual rate of inbreeding multiplied by generation interval (Falconer & Mackay 1996). Pedigree information from the 100 genotyped sheep was combined into one pedigree and the coefficient of inbreeding by birth year and pedigree completeness were calculated for comparison with the whole population pedigree.

Diversity was estimated using microsatellite genotypes including: diversity within populations and between populations, inbreeding, and Hardy Weinberg tests for heterozygosity deficiency and excess in microsatellites with Genepop v.4.3 (Rousset 2008); Ritland inbreeding and relatedness among individuals using Coancestry v.1.0.1.5 (Ritland 1996; Wang 2011); multilocus heterozygosity (Coltman et al. 1999); past changes in effective population size (average harmonic mean effective population size) using VarEff v.1.2 (Nikolic & Chevalet 2014) with an assumed mutation rate of 0.00013 (Crawford & Cuthbertson 1996) and generation interval of three years; and population structure using principal component analysis (PCA) using the R package adegenet (Jombart 2008).

4.1.2 The use of endogenous retroviruses as markers to describe the genetic relationships among local Swedish sheep breeds

Blood samples were collected by a trained technician and DNA was extracted using DNeasy mini kit for QIAsymphony robot (Qiagen®, Hilden, Germany) for adult sheep of various ages. A total of 66 animals from seven breeds (Swedish Finewool, Gute, Roslag, Värmland, Klövsjö, Texel and Mouflon sheep) were included in this paper. Breeds for this study represented populations from different geographical areas and different population sizes. Swedish Finewool sheep originate from the Swedish Landrace and has been crossed with other breeds including Finnsheep, while Gute, Roslag, Värmland and Klövsjö sheep are local breeds. In addition to Swedish breeds, Texel sheep, which originated in the Netherlands and are commonly raised in Sweden, as well as Mouflon sheep, which originates from sheep that became feral shortly after being brought to Europe for agricultural purposes, were included. The presence of six enJSRV loci (enJSRV-7, enJSRV-8, enJSRV-

16, enJSRV-18, enJSRV-15 and enJS5F16) were tested with touchdown PCR (Taq Plus polymerase QIAGEN Kit) using primers from Chessa et al. (2009) and products were then separated by agarose gel electrophoresis. The Fisher exact test was used to test the differences in frequencies of the presence/absence of enJSRV between the breeds, principal component analysis was performed using the presence of the enJSRVs in Swedish sheep and three British breeds from Bowles et al. (2014), and multidimensional scaling was performed using the presence/absence of enJSRV in individuals. Retrotypes, combinations of the presence/absence of four enJSRV (7, 8, 18 and 16), defined Chessa et al. (2009) were also studied in the breeds studied here.

4.1.3 Mutations in ASIP and MC1R: Dominant black and recessive black alleles segregate in native Swedish sheep populations

Blood samples were collected by a trained technician and DNA was extracted using DNeasy mini kit for QIAsymphony robot (Qiagen®, Hilden, Germany) for Swedish Finewool, Gotland, Gute, Helsinge, Klövsjö, Roslag, Värmlands, Texel and Mouflon sheep. Included in this study were sheep from as many flocks as possible and sheep that had "white", "light grey", "grey", "dark grey", "black grey" or "black" fleece colour. Exons of *ASIP* and *MC1R* were sequenced using BigDye© Direct Cycle Sequencing Kit and protocol (Applied Biosystems, Foster City, CA, USA). *ASIP* primers from Gratten et al. (2010) and *MC1R* primers from Fontanesi et al. (2010) were used in this study and forward and reverse primers had M13-21 and M13-29 5' tails respectively. Sequenced products were purified using a BigDye XTerminator® Purification Kit (Applied Biosystems, Foster City, CA, USA) and capillary electrophoresis with Applied Biosystems® 3500xL Dx Genetic Analyzer (Life Technologies, Foster City, CA, USA). Sequences were visually inspected and aligned using CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA).

4.1.4 Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep

Blood samples were collected by veterinarians or under veterinarian supervision from sheep on commerical French farms for routine veterinary care. A total of 691 French sheep were genotyped for 653 305 autosomal SNPs (Ovine Infinium® HD SNP BeadChip) from 27 sheep breeds. Animals were chosen to be as unrelated as possible based on pedigree records. Hardy Weinberg Equilibrium (HWE) was calculated for each SNP within each breed using PLINK (Purcell & Chang 2014; Chang et al. 2015) and SNPs not is HWE (False discovery rate (FDR) 5% calculated with R package qvalue

(Storey 2015)) in one or more breeds were removed. For each breed, a genomic relationship matrix was computed (Yang et al. 2011). The distribution of kinship coefficients was then modelled as a mixture of normal distribution, with the major component of the mixture representing pairs of unrelated individuals. Pairs of related individuals were identified as those that did not belong to this component (FDR $<$ 5%). Finally, one individual for each pair was removed until no related individuals remained. All further analyses were performed on the set of unrelated individuals. SNPs with a minor allele frequency (MAF) = 0, SNPs with a missing call rate $>$ 0.01 and sheep with individual missing call rate $>$ 0.05 were removed.

Three methods were used to study population structure in French sheep from SNP genotypes: principal components analysis using the software PLINK (Purcell & Chang 2014; Chang et al. 2015); a model based approach to estimate individual ancestry coefficients, using the software sNMF (Frichot et al. 2014); A model based approach to infer populations splits and mixtures using the software treemix (Pickrell & Pritchard 2012). For sNMF, the optimal number of ancestral populations (K) was the one that had the lowest cross entropy criterion (Frichot et al. 2014). For treemix, Asian Mouflon sheep were included in the analysis to root the population tree when estimating population splits and mixtures. Allele frequencies from 30 Iranian Asian Mouflon sheep from the NEXTGEN project were included to root the population tree. Only the 498 651 SNPs genotyped in both Asian Mouflon and French sheep were included. The number of migration events included in the population tree was chosen based on the comparison of the fraction of the variance in relatedness between populations that is accounted for by the model (Frichot et al. 2014).

The model based analyses to study population structure were performed a second time to include a subset of breeds from the Sheep HapMap data set, the same European sheep used by Fariello et al. (2014) including breeds from; Central Europe; South Western Europe; North Europe; and Italy. Shared markers between the French sheep breeds and the Sheep HapMap subset of breeds, 39 976 SNPs, were used to estimate individual ancestry coefficients using sNMF (Frichot et al. 2014). Asian Mouflon sheep were again included in the analysis to root the population tree. A total of 38 596 SNPs common to the French sheep, Sheep HapMap subset and Asian Mouflon sheep were used in the analyses.

FLK (Bonhomme et al. 2010) and hapFLK (Fariello et al. 2013) tests were used to detect signatures of selection. The evolutionary model underlying FLK and hapFLK assumes that SNPs were polymorphic in an ancestral population therefore SNPs with estimated ancestral minor allele frequency $<$ 5% were removed, leaving 465 2351 SNPs for use in the analyses. For the FLK analysis,

significant SNPs within 1 Mbp from each other were grouped into a common selection signature (FDR 1%) . Candidate genes were identified in regions containing 10 or more significant SNPs. For hapFLK, a number of haplotype clusters has to be specified in order to fit the Scheet and Stephens model (Scheet & Stephens 2006). This number was set to 30, from results obtained using the cross validation procedure included in the fastPHASE software. P-values for hapFLK were obtained by fitting a chi-squared distribution to the empirical distribution as explained by Boitard et al. (2016) using the script available on the hapFLK software web page. For the hapFLK analysis, regions were constructed from significant SNPs (FDR 5%) and grouped together if they were 1 Mbp or less apart. The FLK analysis was run with all breeds and northern and southern breeds separately and these results were compared to determine the regions missed when northern and southern breeds are run separately. The hapFLK statistics were computed for northern and southern groups separately. Candidate genes were searched for in regions with five or more significant SNPs. Regions from both FLK and hapFLK results had all protein coding genes present extracted. The SNP with the lowest p-value in each region is referred to as the best SNP. The distance from each gene to the best SNP was determined by the distance of the midpoint of the gene to the best SNP and then genes were than ranked with the closest gene labelled as the best gene. Of all results, nine regions were selected for further analysis based on the candidate genes located within the regions. Local trees were constructed by recomputing Reynold's distances between populations in a region and re-estimating branch length of the whole genome tree from local Reynold's distances as in the study by Fariello et al. (2013). Local trees from single SNPs and allele frequencies in FLK results were evaluated to determine breeds selected on and for hapFLK results local trees from single SNPs and haplotypes, allele frequencies and haplotype cluster frequencies.

The *MC1R* locus (OAR14:14 228 283-14 235 506 on OARv3.1 assembly) was amplified as overlapping PCR fragments using appropriate PCR conditions regarding the expected length of the product: either conventional amplification using goTaq polymerase (Promega) or Long-range PCR amplifications using the Long PCR Enzyme Mix provided by Fermentas (http://www.fermentas.de) was performed, using 50 ng of genomic DNA as a template and the manufacturer's protocol. After treatment with 0.5 U of Tsap (Promega) and 10 U of exonucleaseI (Biolabs), 10 to 90 ng of PCR product were used for sequencing with either internal primers or PCR primers used for the amplification. Sequencing reaction was carried out via the BigDye® Terminator v3.1 Cycle Sequencing Kit (http://www.appliedbiosystems.com). The primers are listed in Supplementary Table 2. Sequences from three

animals from each breed exhibiting a selection signature were aligned against scaffold 00839 of the *Ovis aries musimon* assembly (GenBank accession HG925721.1), which we found of better quality than the OAR3.1 reference genome. The CLC software was used for the alignment as it allowed the detection of polymorphisms. The discovered polymorphisms were genotyped for all the animals of the relevant breeds by sequencing purified PCR products with internal primers using the same protocol. The sheep genome browser from EBI (www.ensembl.org) was used to determine conserved regions through 39 eutherian mammals and calculate GERP scores.

4.1.5 Population structure of five native sheep breeds of Sweden estimated with high density SNP genotypes

Blood samples were collected by a trained technician and DNA was extracted using DNeasy mini kit for QIAsymphony robot (Qiagen®, Hilden, Germany) from five breeds (Dalapäls, Fjällnäs, Gotland, Gute and Klövsjö) that represented different geographic origins (within Sweden) and population sizes. Sheep were genotyped for 653 305 autosomal SNPs (Ovine Infinium® HD SNP BeadChip) at Labogena, INRA, Paris, France. SNPs with a minor allele frequency (MAF) of zero, SNPs with a missing rate greater than 0.01 and sheep with a missing genotype rate greater than 0.15 were excluded from further analysis. A total of 93 sheep and 502 144 SNPs passed quality control. Three methods were used to study population structure in the five Swedish breeds: principal component analysis with PLINK (Purcell et al. 2007; Purcell & Chang 2014), a mixture model to estimate individual ancestry coefficients estimated with sNMF (Frichot et al. 2014) and a maximum likelihood population tree model with estimated mixture events with treemix (Pickrell & Pritchard 2012). In treemix analyses, Asian Mouflon frequencies of SNPs were included from the NEXTGEN project.

After analysis of only Swedish breeds, a French sheep dataset (also genotyped with Ovine Infinium® HD SNP BeadChip, first studied in paper IV, and European and south west Asian sheep (genotyped for 49 034 SNP for the Sheep HapMap project) were added. The common SNPs were used to study population structure using the same three methods previously used. A total of 38 589 SNPs and 2938 sheep were included in this study. In treemix analyses, Asian Mouflon frequencies of SNPs were included from the NEXTGEN project.

4.2 Main findings

4.2.1 Estimation of genetic diversity in Gute sheep: pedigree and microsatellite analyses of an ancient Swedish breed

This study was the first analyses of genetic diversity of the Gute sheep population in Sweden using pedigree information and microsatellite markers. Inbreeding is being successfully managed in this small local population and this was determined by studying both the whole population pedigree and a sample of the population using microsatellite genotypes and pedigree information.

There were a total of 70 474 births recorded in our data set since 1960. There were 30 616 animals born from 2007 to 2012 and they had an average pedigree completion index for three generations greater than 0.8 and therefore this cohort were used to calculate population parameters. A total of 94 animals from 13 flocks were successfully genotyped.

The coefficient of inbreeding estimated from the whole population pedigree was low (0.038) and the coefficient of inbreeding estimate from the sample population pedigree was also low (0.018). The average inbreeding estimated from sheep born between 2003 to 2012 has remained relatively constant. Ritland inbreeding (-0.03932) and average multilocus heterozygosity (1.01845) estimated from microsatellite genotypes of the sample population indicated that sheep were more heterozygous than expected. Effective population size estimates from the population pedigree and the microsatellite genotypes were 155.4 and 88.3.

These results indicated that the current breeding programs for reducing inbreeding in Swedish Gute sheep are successful.

4.2.2 The use of endogenous retroviruses as markers to describe the genetic relationships among local Swedish sheep breeds

Genetic relationships between breeds were estimated using enJSRV and retrotype frequencies. All pairwise differences in frequency of enJSRV-18 for Swedish breeds with Texel were significant. All but Swedish Finewool and Texel pairwise differences in frequencies for enJS5f16 were significant. For enJSRV-7, pairwise differences between Värmlands and Texel and Klövsjö and Texel were significant and for enJSRV-8 pairwise differences between Klövsjö and Texel and Swedish Finewool and Texel were significant. The principal component analysis plot showed Värmlands and Klövsjö to be the closest together breeds and furthest breeds from Texel. The multidimensional scaling plot (Figure 1A in paper II) showed large variation within breeds and some differentiation between breeds (all Texel were on the right of the plot, and all Klövsjö and Värmlands were on the left). The study of retrotypes

showed all Swedish breeds had a primitive retrotype at the highest frequency. In Klövsjö and Värmlands this was R0 and in Swedish Finewool, Gute and Roslags this was R1 although R0 was also present in all three breeds. In addition, Swedish Finewool, Gute, Roslag sheep had moderate frequencies of enJSRV-18 indicating the presence of more recently derived sheep breeds. The Texel shared only one retrotype with Swedish sheep, and that was only shared with the Swedish Finewool. Roslag sheep had a unique and new retrotype comprising of enJSRV-7 and enJSRV-8. These results show that all the Swedish breeds in this study have some primitive origins especially the Värmlands and Klövsjö breeds.

4.2.3 Mutations in *ASIP* and *MC1R*: Dominant black and recessive black alleles segregate in native Swedish sheep populations

Recessive and dominant black alleles were found to be segregating in some Swedish populations and *ASIP* and *MC1R* genotypes reflected breed selection histories.

A total of 17 combinations of genotypes were found in 151 animals in *ASIP* coding regions. Black coat colour can be explained when individuals are homozygous for the deletion g.100-105del (D_5) or homozygous for the mutation at g.5172 or heterozygous for both mutations (as it is possible that they only had non-functional copies of *ASIP* (one non-functional because of D_5 and the other because of g.5172A)). Black Roslag sheep, and the one black Texel included in this study were explained by two copies of the deletion D_5. Black Swedish Finewool, black Klövsjö and the one black Gotland sheep were explained by D_5 and/or g.5172A. Only 11 of 16 black sheep from Gute, Helsinge and Värmland breeds could be explained by D_5 and/or g.5172A mutations. The five black individuals that could not be explained were heterozygous for either D_5 or g.5172A, but not both at the same time.

Seven other mutations in *ASIP* were found: six in introns and one synonymous mutation found in the fourth exon.

There was evidence of a duplication of *ASIP* in Gotland sheep because all sheep had one of two *ASIP*-genotypes: 17 sheep were heterozygous at all mutations and 28 sheep were heterozygous at all mutations but at g.100-105, where they were homozygous for N_5 (no deletion). There were a total of 82 animals with grey in their fleece colour description that were N_9D_9. The four remaining animals with grey in their phenotypic description that were not N_9D_9 included: a grey spotted Helsinge sheep, a dark grey Värmland sheep, a dark grey Helsinge sheep, and a grey Värmland sheep. Explanations for these four individuals not having a copy of D_9 could include: grey colour or spotting is

controlled by alleles in other genes and phenotyping or genotyping errors. There were five black and eight white sheep that also had one copy of D_9 so further study of this duplication and the mutations involved is needed.

There were a total of seven mutations sequenced in *MC1R* (and 10 genotype combinations) in a total of 128 sheep from eight breeds. All seven mutations were located in coding regions, three were synonymous and four were missense mutations. The four missense mutations included two previously known mutations, c.218T>A and c.361G>A (Våge et al. 1999; Våge et al. 2003), and two previously unknown mutations, c.452G>A and c.785C>T. The novel missense mutation, c.785C>T, was found in one black and one grey Gute sheep. The other novel mutation at c.452 was found segregating in three populations: Gotland, Gute and Värmland sheep. A total of 60 animals were heterozygous at c.452 including all 40 Gotland sheep sequenced. All Gotland animals had the same combination of genotypes in *MC1R* and therefore there is evidence that there is a duplication of *MC1R* in Gotland sheep. This genotype could be associated with the black heads and legs seen in all non-white Gotland sheep. The mutation at c.452 was found to be segregating in Gute and Värmland sheep.

The two previously known missense mutations, c.218T>A and c.361G>A, segregated only in Swedish Finewool in this study and could explain black coat colour in five of six black sheep (these are the same black sheep sequenced at *ASIP* where all six animals could have their coat colour explained by D_5 and/or g.5172A). The two previously known mutations in *MC1R*, c.218T>A and c.361G>A, were not present in local Swedish breeds and therefore the dominant black allele may not have been present in early Swedish sheep populations.

4.2.4 Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep

This paper studied population structure and detected signatures of selection, many of which demonstrated allelic heterogeneity. After quality control, there were 527 823 SNP markers used for the analyses. The 27 populations chosen represented the majority of commercial breeds present in France but also included some breeds maintained for conservation purposes. The three approaches (PCA, estimated ancestry coefficients and estimated population tree) showed a clear structure with two main groups of breeds plus two highly differentiated breeds, the Ouessant and the Mérinos de Rambouillet sheep. Both populations are small and isolated with the Ouessant breed possibly an ancient European breed of sheep such as the Soay sheep (Chessa et al. 2009)

and the Mérinos de Rambouillet breed has been a single flock of animals without the introduction of additional animals since 1786. Although it has a very long branch in the population tree, it clusters with the other Merino population in the dataset, the Mérinos d'Arles. Two outgroup populations were included in the analyses: allele frequencies in the Asian Mouflon (the ancestor of domesticated sheep (Nadler et al. 1973; Bunch et al. 1976)) were obtained from the NEXTGEN project and were used to root the population tree in the maximum likelihood tree analyses and two genotyped samples of European Mouflon from Corsica were included in all three analyses. In the maximum likelihood tree analyses, European Mouflon sheep were between Asian Mouflon and domesticated sheep, consistent with their known origins (Poplin 1979) therefore European Mouflon were used as an outgroup for FLK and hapFLK analyses. The two groups of breeds found in the remaining sheep corresponded to northern and southern origins and had contrasting structure: northern breeds were clearly separated in PCA, had clear cluster assignment when estimating individual ancestry coefficients and generally had longer branch length in the population tree; while southern breeds were less differentiated, PCA clusters were closer together, individual clustering was sublter and shorter branch length in the population tree. The division between northern and southern French breeds in population structure analysis is possible evidence of the effects of routes of domestication on modern sheep breeds and because of longer established formal breeding programs in northern Europe.

The three analyses were redone to include other European breeds from the Sheep HapMap project. The French sheep complement the Sheep HapMap dataset in the European sheep population tree. Most of the breeds in our dataset added to the global diversity, as they tend to root in internal branches of the population tree. The other French breeds branched in population groups that were already present in the Sheep HapMap. As when considering French breeds only, the tree obtained from the inclusion of the Sheep HapMap breeds separated into southern breeds and northern breeds. Here also, the northern breeds tended to have longer branch lengths and more drift than southern breeds.

While hapFLK methods are robust to bottlenecks and moderate admixture, both of these can affect power for detecting signatures of selection (Fariello et al. 2014). Therefore four breeds were not included in the selection scan: three breeds, the Ouessant, Mérinos de Rambouillet and Berrichon du Cher, that have experienced severe bottlenecks, corresponding to very long branch lengths in the population tree; and the Romane breed, as it is a recent composite of two breeds. When considering all breeds together, 50 selection

signatures were detected using FLK methods. Three of these signatures were only found when analyzing all the breeds together, while the other 47 were also found in the within group analyses. As only three signatures were specific to analyzing all breeds together, and there was more power to detect signatures of selection when separating breeds into the two groups, signatures of selection were detected in the two groups separately using FLK and hapFLK: 61 and 26 regions were detected respectively for northern sheep populations and 65 and 42 regions for southern sheep populations.

Selective sweep regions in paper IV using high density SNP information were on average smaller than those detected by Fariello et al. (2014) using medium density (50 000) SNP information. The higher resolution obtained could come from the fact that paper IV had a better haplotype diversity description with higher SNP density and the size of the candidate region was reduced by analyzing more breeds showing more recombinant haplotypes.

The candidate genes identified included genes for coat colour (*ASIP*, *MC1R*, *TYRP1*, *MITF*, *EDN3*, and *BNC2*), stature and morphology (*NPR2*, *MSTN* (*GDF-8*), *LCORL* and *NCAPG*, *ALX4* and *EXT2*, *PALLD*), milk production (*ABCG2*), horns (*RXFP2*) and a region where a wool quality QTL has been found (Kijas et al. 2012; Fariello et al. 2014). Other candidate genes that we identified included *SOCS2*, associated with growth and mammary gland development (Rupp et al. 2015), *OXCT1*, associated with milk fatty acid traits in dairy cattle (C. Li et al. 2014), EBF2, involved in brown fat fate and function in mice (Rajakumari et al. 2013), *ADAMTS9*, under selection in Tibetan pigs and boars living at high and moderate altitudes and *MSRB3*, associated with floppy ear position in dogs (Boyko et al. 2010).

A known causal SNP (rs408469734) for increased muscularity in *MSTN* previously found in Texel sheep was identified as the best SNP in a region under selection. Local population trees showed the Texel and Rouge de l'Ouest under selection with the Rouge de l'Ouest carrying a similar but smaller haplotype than in the Texel. A comparison of FLK tests with and without Texel sheep on chromosome 2 showed that it is likely that both breeds share a common ancestral population where this SNP segregated and it was not introduced into the Rouge de l'Ouest via introgression from the Texel.

Allelic heterogeneity is the presence of multiple selected alleles, or haplotypes, at the same genomic location. Allelic heterogeneity was identified in selection signatures when there was evidence for more than one breed affected by selection in the same region and different haplotypes having arisen to high frequency in the selected breeds. The six examples presented in more detail in paper IV are: *ADAMST9*, *MSRB3*, *SOCS2*, *RXFP2*, *LCORL* and *MC1R*.

The selection signature on chromosome 14 was investigated further. Selection in this region affected the only three black faced breeds in this study: the Romanov, the Suffolk and the Noire du Velay. Haplotype diversity plots showed they were clearly selected on different haplotypes. These three breeds plus Texel, a breed not found to be selected in this region, were resequenced in this region. A total of six SNPs and one 11 bp insertion were detected and genotyped (dbSNP accession numbers ss# 1996900605 to 1996900611). The three breeds were almost completely fixed around *MC1R*, and the three main breed haplotypes were different. The Noire du Velay was the only breed that carried the two known mutations for the dominant black allele in sheep (named c.218T > A and c.361G > A) (Våge et al. 1999; Våge et al. 2003). These results confirmed the presence of three different haplotypes around the *MC1R* gene.

4.2.5 Population structure of five native sheep breeds of Sweden estimated with high density SNP genotypes

The principal component analysis (PCA) showed the five Swedish breeds, Dalapäls, Fjällnäs, Gotland, Gute and Klövsjö sheep, were distinguishable and unique from one another and the individuals within each breed clustered close together (with the exception of two Dalapäls individuals). When estimating individual ancestry coefficients, Klövsjö was the most different breed followed by Dalapäls. Gotland sheep and Gute sheep were more related and there appeared to be some relationship between Fjällnäs and Gotland sheep. The same two Dalapäls sheep were different from other Dalapäls sheep and in addition there were three Fjällnäs sheep that were different from other Fjällnäs sheep. The animals that were a bit different from their breed mates were likely because of the way these breeds were formed.

In the population tree rooted by the Asian mouflon sheep, the five breeds were very distinct and had long branch lengths indicating high amounts of drift. Gute and Gotlands sheep had a closer relationship than the other breeds: they both were on the same branch although they both had long branch lengths.

When the other European and south west Asian breeds were included in the PCA, breeds generally grouped with others from the same geographical area of origin. There were three exceptions to this: the Mérinos de Rambouillet; and the Boreray and Soay sheep breeds. These breeds are small, isolated populations that have accumulated drift. Swedish sheep breeds and north European short-tailed sheep from other countries grouped together in the principal component analysis, the estimates of individual ancestry coefficients and in the population tree. Sheep breeds in the north European short-tailed

sheep group (including Swedish sheep) generally had long branch lengths in comparison with other geographic groups of sheep breeds.

5 General discussion

5.1 Genetic diversity of sheep

Sheep are genetically diverse in part because they are adapted to the different environments and production systems they have been bred for. When compared with cattle, there are more sheep breeds and fewer extinct breeds (Taberlet et al. 2011). Many sheep breeds are based in specific local areas which they are adapted to (Taberlet et al. 2008). This diversity is older than diversity from breed formation which started around 200 years ago (Taberlet et al. 2008). Unlike cattle, where a small number of breeds dominate the world-wide industry, sheep have not been disseminated in the same way: artificial insemination is not used as extensively in sheep as it is in cattle and has been limited to a small number of highly productive breeds (Taberlet et al. 2008). Additionally, in sheep there has been extensive crossbreeding, favouring some breeds such as the Merino, British breeds and Texel (Groeneveld et al. 2010). One of the future challenges for maintaining genetic diversity in sheep will be its fragmentation in small populations. These small, unique breeds add to overall diversity of the species but the within breed diversity is low. For example, North European short-tailed sheep breeds are generally small in population size (Dýrmundsson & Niżnikowski 2010) but using microsatellite genotypes they were shown to be genetically unique, add to overall diversity in sheep but have low within breed diversity (Tapio et al. 2005). Small populations can accumulate a lot of drift as seen in papers IV and V where breeds with small populations like Ouessant, Mérinos de Rambouillet, Fjällnäs, Klövsjö and Dalapäls, in addition to other north European short-tailed sheep breeds, have accumulated more drift than other groups of breeds). Even in comparison to most of the southern breeds in paper IV, the northern breeds had more drift. One possible explanation for more drift in northern breeds was that there have been longer established formal breeding programs and breed

organizations in northern France (and in the British breeds historically used for crossbreeding in these populations) in comparison with southern breeds. Drift will be a challenge in the future for Sweden's local sheep populations because these breeds have low population sizes. To safeguard the future of genetic diversity in Swedish breeds, future conservation decisions will have to balance the importance of cultural and historical identity with future needs for genetic diversity.

5.2 Allelic heterogeneity

Paper III and paper IV have examples that highlight the complexity of coat colour and demonstrate that when there is parallel selection in breeds for a specific phenotype or allele, it is not necessarily the same mutations or genes being selected on. In paper III where there are two different black alleles being studied there are two different genes and in each gene there are two mutations (dominant black allele in *MC1R*: c.218T>A or c.361G>A (Våge et al. 1999; Våge et al. 2003), and recessive black allele in *ASIP*: g.100-105del or g.5172T>A (Smit et al. 2002; Norris & Whan 2008)). The recessive black and dominant black allele are impossible to phenotypically distinguish from one another (Adalsteinsson 1983). In Swedish sheep, local breeds were black because of mutations in *ASIP* while a native Swedish breed with a history of crossbreeding to improve fleece quality had mutations in both *ASIP* and *MC1R* that could cause black coat colour. In paper IV, *MC1R* was re-sequenced in three breeds under selection in this region. These three breeds had black faces, necks and legs (but not necessarily black fleece) and only one breed had the two known mutations, c.218T>A or c.361G>A, while the other two breeds are black because of mutations outside of the coding region. In the literature there are also differences in the mutations present in breeds causing black coat colour (Våge et al. 1999; Smit et al. 2002; Våge et al. 2003; Norris & Whan 2008; Royo et al. 2008; Fontanesi et al. 2010; Gratten et al. 2010; Fontanesi et al. 2012). This highlights that mutations associated with a phenotypic trait in one breed doesn't necessarily translate to those causative mutations being present in other breeds with the same phenotype and the causative mutations could even be at a different gene locus.

5.3 Methods

This thesis employed a number of different methods to estimate genetic diversity in domestic sheep. When choosing methods, a careful consideration

of the available resources, the stakeholders, the information out and the possible applications, is important. The use of pedigree information is powerful for conservation programs (Fernández et al. 2005) and this was true in paper I too. However, in small populations the exact level of inbreeding is not known because pedigree records aren't extensive enough (Li et al. 2011). Pedigree information is often limited in quality and quantity: there are errors in recorded pedigrees and accurate inbreeding and effective population size can be estimated only for the recent years where data is available and complete (Li et al. 2011). In more intensive production systems, accurate pedigree is easier to record however in some extensive production systems matings are not as controlled or producers are not present at lambings and there may be lamb adoption by other ewes (Li et al. 2011). It is a challenge to accurately record parentage unless there is parentage testing using molecular markers. Microsatellite markers have been used extensively in the past for diversity studies in many species but they are being replaced by other markers like SNPs. Both microsatellites and SNPs can have ascertainment biases: for example, in wild species, the most polymorphic microsatellites are used and SNP discovery is done by resequencing a small sample of individuals to find polymorphisms (Taberlet et al. 2011; Väli et al. 2008; Clark et al. 2005). Another challenge with microsatellite markers is that there is not a lot of consistency with the markers used between studies because of inconsistent reproducibility (FAO Commission on Genetic Resources for Food and Agriculture 2015) and this makes it difficult to compare results between studies or to build on studies. For example, of the seven microsatellite genotyped in paper I only four were in a previous study of north European short-tailed sheep breeds genotyped for 25 microsatellites (Tapio et al. 2005). Microsatellites are fewer and the genetic distance between them is larger, and cost more per data point than other genetic markers (FAO Commission on Genetic Resources for Food and Agriculture 2015). In contrast, medium and high density SNP arrays are available for genotyping sheep. Many studies are using the same SNP markers because they can cover the genome and are much closer together than microsatellite markers. Examples can be seen when comparing the seven microsatellite markers in paper I to the over 500 000 SNPs in paper IV and V. The use of SNP arrays also allowed for comparison with breeds genotyped for 50 000 SNPs from the Sheep HapMap project. Endogenous retroviruses are used for studying population history and structure. They don't give detailed information about breeds but can indicate the presence of ancient or more recent breeds (Chessa et al. 2009). This data is easy to build on as seen in paper II and the comparison of Swedish breeds with British breeds from Bowles et al. (2014). Sequencing parts of specific genes shows the population

selection history at that exact region and can help with finding causative mutations although a study where researchers detected signatures of selection in cattle with whole genome sequences found that most causal mutations are found outside of coding regions (Boitard et al. 2016).

5.4 Future studies

The main theme of the papers was to study genetic diversity however some of these studies lend themselves to learning about the function of genes. There are reoccurring signatures of selection in different species with candidate genes associated with coat colour and morphology in sheep, cattle, horses, rabbits and pigs (Fariello et al. 2014; Boitard et al. 2016; Petersen et al. 2013; Carneiro et al. 2016; Rubin et al. 2012). The study of domestic animals will not only improve animal production and welfare but the better understanding of genotype to phenotype will also help with other species including in human medicine (Andersson et al. 2015). When detecting signatures of selection (as done in paper IV) candidate genes in regions under selection were found that are important for agriculture and adaptation. While paper IV did not use phenotypes, it does direct future research to possible candidate genes and the breeds to study them in. In the future, these additional data can help these types of studies not only in sheep. Using high density SNP genotyping narrowed down the selection signatures. On average the identified regions were smaller than regions detected using a medium density SNP array by Fariello et al. (2014). To improve even further, whole genome sequences would allow the detection of actual causal mutations under selection (Boitard et al. 2016). With high density SNP genotyping there is a chance of including causal mutations. An example of this was seen in paper IV where the known causal mutation in *MSTN*, rs408469734, for increased muscling was the SNP with the most significant FLK signal of the selection signature detected in that region. However, the high density SNP array will not include all causal mutations. Detecting selection signatures can allow for targeted sequencing of those candidate genes to determine causal mutations. However, in a recent study where researchers detected signatures of selection using whole genome sequences of cattle, they had more detection power and could detect causal mutations or at least narrow down candidates mutations and they found that the majority of these mutations regions are not in coding regions but regulatory regions (Boitard et al. 2016). Including additional information like environment and phenotypes could aid in determining the purpose of selection signatures. Finally, function of genes can be better studied in the future by

consortiums like Functional Annotation of Animal Genomes (FAANG), which is committed to collecting phenotypes and tissues at different stages of development of domesticated animals for the purpose of better understanding genotype to phenotype and study other non-additive genetic effects (Andersson et al. 2015).

While papers IV and V include many breeds from around the world, there are still many breeds not included and future studies could benefit from an even greater number of sheep breeds being genotyped. For example, although in paper IV, breeds represented many of the commercially important French breeds and some of the conserved French breeds, this study was not as extensive in studying French sheep populations as a previous study including 49 sheep breeds in a study of French sheep population structure and genetic diversity using microsatellite genotypes (Leroy et al. 2015). In paper V, only five Swedish breeds were included, three of which belong to the Föreningen Svenska Allmogefår, an organization for Swedish local breeds which includes a total of ten breeds (http://allmogefar.se/). By including some of the rare, local breeds from both France and Sweden knowledge about their genetic diversity could be used for future conservation decisions but would also allow the study of adaptation to specific environments. In paper IV some selective sweeps detected are likely associated with adaptation. For example, the region with the candidate gene *a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif 9* (*ADAMTS9*), which was also a candidate gene in a study where wild boars and pigs at different altitudes (Li et al. 2013), was under selection in Romanov and Causse du Lot sheep. Because this candidate gene was previously seen and Romanov and Causse du Lot sheep are hardy breeds living in colder regions, this could be a gene involved in temperature regulation. However, to discover further regions important for adaptation, sheep breeds from other climates should be included: French and Swedish sheep are in temperate to cold climates while sheep adapted to warmer climates are also important to include in studies on adaptation. Breeds from the tropics are not well studied at the moment but these populations could be critical for responding to future challenges including those which are related to climate change (Thornton et al. 2009). Finally, local breeds from all climates are important to include in future studies because local breeds are low input breeds (Hoffmann 2011) and this could also prove important in the future where livestock will need to produce well on lower quality feed (Hayes et al. 2013). Livestock will be strategic in adapting to higher temperatures, converting low quality forage to high quality nutrients for humans (meat and milk) and efficiently meeting increased demands (Hoffmann 2010).

Historical events were inferred from modern DNA genotypes in all five papers included in this thesis. To validate these conclusions and further study population history of sheep, future studies of modern Swedish and French populations could be compared with ancient sheep DNA. One example of this was a MSc thesis defended in 2015: the coding regions in *ASIP* and *MC1R* were genotyped in Swedish breeds (some of these results were included in paper III) and coding mutations in *ASIP* and *MC1R* were genotyped from ancient DNA from old sheep bones from the Iron Age (from Stora Förvar), fourth to seventh century (from Runsaborg) and Middle Ages (from Stora Förvar) (Westberg Sunesson 2015). In paper III, the Swedish Finewool had dominant black alleles (mutations c.218T>A and c.361G>A) while the local breeds did not. Because Swedish Finewool have had a recent history of crossbreeding to improve wool quality, one conclusion was that these mutations were never present in early populations of sheep in Sweden. Ancient sheep DNA genotypes for *MC1R* dominant black allele mutations (c.218T>A and c.361G>A) revealed one individual from the fourth to seventh century was heterozygous at c.361 and one individual from the Middle Ages was heterozygous at c.361 however these genotypes have not yet been replicated (Westberg Sunesson 2015). If these results can be verified by additional genotyping, this suggests that at least one of the *MC1R* mutations was present in Swedish sheep populations much earlier than hypothesized in paper III. Another example of the use of ancient sheep DNA in resolving population history was in a paper where researchers studied ancient DNA from Estonia and surrounding area from the Bronze Age to modern times by genotyping mitochondrial DNA: these researchers found continuity in haplotypes over this 3000 year time period (Rannamäe et al. 2016). Another study of ancient Finnish sheep DNA showed that there has been no major population replacement in Finland since the Iron Age (Niemi et al. 2013). Studies of continuity and genetic diversity over time in French sheep would be a way to confirm that northern and southern French sheep genetically divide into two groups because they originate from two different routes of domestication as hypothesized in paper IV.

6 Conclusions

Population history, structure and genetic diversity of domestic sheep in Swedish and French sheep populations were the focus for this thesis.

Genetic diversity and population structure in Swedish breeds of sheep was studied using low density DNA marker information and pedigree information. Gute sheep, an ancient Swedish breed, was studied using whole population pedigree information and seven microsatellite markers. Estimates of inbreeding and diversity for the Swedish Gute sheep population were consistent with a population selected to reduce inbreeding.

Population structure of five Swedish sheep breeds was studied using endogenous retroviruses as DNA markers. Swedish Finewool, Gute, Klövsjö, Roslag and Värmland sheep had retrotypes characteristic of ancient breeds however Swedish Finewool, Gute and Roslag sheep also had retrotypes indicative of influence from more modern sheep breeds.

Phenotypic variation in Swedish sheep and genotypic variation of two coat colour genes (*ASIP* and *MC1R*) was studied by sequencing of coding regions. Black coat colour was explained in Klövsjö and Roslag sheep by mutations in *ASIP*. Black coat colour was explained in Swedish Finewool by mutations in *ASIP* or mutations in *MC1R*. Black coat colour in Gute, Helsinge and Värmland sheep was not always explained by coding region mutations in *ASIP*. A novel mutation was found in *MC1R* in Gotlands, Gute and Värmland sheep and there was evidence of a novel duplication of *MC1R* in Gotland sheep.

Genetic diversity and population structure in Swedish and French breeds of sheep was studied using high density DNA marker information. French sheep were divided into two groups, northern and southern breeds. Signatures of selection were found within northern and southern groups, in many of the identified regions, important candidate genes for agriculturally traits were located. The results highlighted allelic heterogeneity found in selection signatures and this was validated by resequencing *MC1R* in three breeds that

were under selection in this region. The selection signature results indicated that each breed had a different haplotype that had been selected on, a result which was confirmed by resequencing. The five Swedish breeds genotyped with the 600K SNP array were unique and had accumulated drift and when compared with other European and south west Asian breeds, they grouped with north European short-tailed sheep.

7 Thesis summary

Domestic sheep (*Ovies aries*) are raised all over the world. They are a heterogeneous group raised for meat, milk, or fibre production, or for conservation purposes. Some sheep breeds have low population sizes and are found in very specific environments while other breeds are used around the world. Sheep are raised in a wide range of production systems from extensive to intensive. The differing selection pressures, both natural and artificial, on domestic sheep make this species a good model for studying adaptation. In addition, it is important that the genetic diversity in domestic sheep is studied, especially in local breeds that are at risk of extinction to maintain them for the future. By maintaining genetic diversity, future challenges including changing environment and emerging diseases can be met more quickly. This thesis includes five studies of genetic diversity and population structure with a focus on Swedish and French sheep breeds.

Paper I presented estimates of Gute sheep, a Swedish breed that is primitive, horned, short-tailed and native to Gotland, an island in the Baltic Sea. Inbreeding and effective population size was calculated using pedigree information from the registered population of Gute sheep in Sweden until 2012. A sample of 100 Gute sheep was used to estimate inbreeding and effective population size using pedigree data and microsatellite genotypes. The inbreeding coefficient estimates from the whole population pedigree and the sample population pedigree were low (0.038, 0.018). Estimates from microsatellite genotypes showed that the average multilocus heterozygosity was high (1.01845) and Ritland inbreeding was low (-0.03931) and five of seven markers were not in Hardy-Weinberg equilibrium because of excess in heterozygosity. The effective population size was 155.4 and average harmonic mean effective population size was 88.3. In total, there were clear indications that this breed is currently part of a successful conservation program.

Paper II presented estimates of genetic relationships among five Swedish breeds. Individuals from Swedish Finewool, Gute, Klövsjö, Roslag and Värmland sheep breeds were genotyped for six endogenous Jaasiekte retroviruses of sheep (enJSRVs). Klövsjö and Värmland sheep had low frequencies of enJSRVs and for Swedish Finewool, Gute and Roslag sheep, enJSRV-7 was the most frequent. Both the absence of enJSRVs and the presence of enJSRV-7 are characteristics of primitive breeds. Swedish Finewool, Gute and Roslag sheep had moderate frequencies of enJSRV-18 which is indicative of the presence of more recently derived sheep breeds.

In paper III, agouti signalling protein (*ASIP*) and melanocortin 1 receptor (*MC1R*) coding regions were sequenced to study mutations associated with black coat colour in seven Swedish breeds. Not only does studying coat colour help determine the function of genes, but mutations associated with coat colour and pattern in modern animals can inform on population history. Swedish breeds are good models for studying coat colour because they display a variety of coat colours and patterns. Klövsjö and Roslag black sheep could be explained by two previously known variants in *ASIP* (recessive black allele: g.100-105del and g.5172T>A) and black Swedish Finewool by either mutations in *MC1R* (dominant black allele: c.218T>A or c.361G>A) or *ASIP*. In contrast, only one third of individuals with black fleece in Gute, Gotland and Värmland sheep had genotypes that could explain black coat colour. This could be because these breeds have grey individuals and the grey allele is a duplication, which could make also the black coat colour inheritance more complicated. Finally, a novel missense mutation (c.452G>A) was identified in *MC1R* in Gute, Gotland and Värmland sheep and there was evidence of a novel duplication of *MC1R* in Gotland sheep.

Paper IV presented population structure and signatures of selection in French sheep. A total of 27 breeds were genotyped for over 500 000 SNPs. Sheep were divided into two groups, northern and southern populations and admixture events between the two groups were identified when studying population structure using principal component analysis, individual ancestry coefficients and estimated population trees. There was evidence for parallel selection events in regions that had candidate genes potentially involved in morphological traits, coat colour and adaptation. There was confirmation of different mutations responsible for the same phenotypic trait in *MC1R* (coat colour). This was confirmed by resequencing *MC1R* in the three breeds that had hard selective sweeps in the region that included *MC1R*. The paper concluded that dense genetic data in many populations can decipher evolutionary history of populations and adaptive mutations.

In paper V, population structure in Swedish breeds was studied by genotyping five Swedish sheep breeds for over 500 000 SNPs. The five breeds, Dalapäls, Fjällnäs, Gotland, Gute and Klövsjö sheep, included in this paper were chosen because of their differing geographic origins and population. It was found that these are five distinct breeds with Gute and Gotland more closely related. The analyses of population structure were then repeated with the inclusion of the French sheep dataset from paper IV and additional medium density genotypes from world-wide sheep breeds from the Sheep HapMap project. All five Swedish breeds grouped with other north European short-tailed sheep in the study. Breeds were all distinct and Swedish breeds in general have accumulated more drift in comparison with other breeds.

8 Sammanfattning

Tamfår (*Ovies aries*) föds upp över hela världen. De är en heterogen grupp som föds upp för produktion av kött, mjölk eller ullfibrer eller för bevarande. Vissa fårraser har små populationsstorlekar och finns i mycket specifika miljöer medan andra raser används över hela världen. Får föds upp i ett brett spektra av produktionssystem, från omfattande till intensiv. De olika selektionstrycken, både naturliga och artificiella, på tamfår gör denna art till en bra modell för att studera anpassning. Dessutom är det viktigt att den genetiska mångfalden hos tamfår studeras, särskilt i lokala raser som riskerar att utrotas, för att bevara dem för framtiden. Genom att upprätthålla genetisk mångfald kan framtida utmaningar, inklusive förändrad miljö och nya sjukdomar, mötas snabbare. Avhandlingen innehåller fem studier av genetisk mångfald och populationsstruktur med fokus på svenska och franska fårraser.

Artikel I presenterade beräkningar hos Gutefår, en svensk ras som är primitiv, behornad, kortsvansad och ursprungligen från Gotland, en ö i Östersjön. Inavel och effektiv populationsstorlek beräknades med hjälp av stamtavleinformation från den registrerade populationen av Gutefår i Sverige fram till 2012. Ett stickprov på 100 Gutefår användes för att uppskatta inavel och effektiv populationsstorlek med data från stamtavla och mikrosatellitgenotyper. Inavelskoefficienten beräknad från hela stamtavlan och från stickprovets stamtavla var låg (0,038 respektive 0,018). Beräkningar från mikrosatellitgenotyper visade att den genomsnittliga hetererozygotin från flera loci var hög (1,01845) och Ritland-inavel var låg (-0,03931) och fem av sju markörer var inte i Hardy-Weinberg-jämvikt på grund av överskott i heterozygoti. Den effektiva populationsstorleken var 155,4 och effektiv populationssstorlek beräknad som genomsnittligt harmoniskt medelvärde var 88,3. Totalt var det tydliga tecken på att denna ras för närvarande har ett framgångsrikt bevarandeprogram.

I artikel II presenterades beräkningar av genetiska relationer bland fem svenska raser. Individer från Finullsfår, Gutefår, Klövsjöfår, Roslagsfår och Värmlandsfår genotypades för sex endogena Jaasiekte retrovirus hos får (enJSRVs). Klövsjöfår och Värmlandsfår hade låga frekvenser av enJSRV och för Finullsfår, Gutefår och Roslagsfår var enJSRV-7 den vanligaste. Både frånvaron av enJSRV och närvaron av enJSRV-7 är egenskaper hos primitiva raser. Finullsfår, Gutefår och Roslagsfår hade måttliga frekvenser av enJSRV-18 vilket är en indikation på förekomsten av mer moderna fårraser.

I artikel III sekvenserades de kodande regioner i generna ”agouti signalling protein” (*ASIP*) och “melanocortin 1 receptor” (*MC1R*) för att studera mutationer associerade med svart pälsfärg i sju svenska raser. Att studera pälsfärg hjälper inte bara till att bestämma genernas funktion, utan mutationer i samband med pälsfärg och mönster i moderna djur kan även ge information om populationernas historia. Svenska raser av får är bra modeller för att studera pälsfärg eftersom de har olika pälsfärger och mönster. Svart färg hos Klövsjöfår och Roslagsfår kan förklaras av två tidigare kända varianter i *ASIP* (recessiv svart allel: g.100-105del och g.5172T>A) och svarta Finullsfår av antingen mutationer i *MC1R* (dominant svart allel: c.218T>A eller c.361G>A) eller *ASIP*. Däremot hade endast en tredjedel av individerna med svart färg hos Gutefår, Gotlandsfår och Värmlandsfår genotyper som kunde förklara svart färg. Detta kan bero på att dessa raser har gråa individer och den grå allelen är en duplikation, vilket också kan göra den svarta pälsfärgens nedärvning mer komplicerad. Slutligen identifierades en ny missensmutation (c.452G>A) i *MC1R* hos Gutefår, Gotlandsfår och Värmlandsfår och det fanns bevis för en ny duplikation av *MC1R* i Gotlandsfår.

Artikel IV presenterade populationsstruktur och spår av selektion i franska får. Totalt 27 raser genotypades för över 500 000 SNP. Får delades in i två grupper, nordliga och sydliga populationer, och blandning mellan de två grupperna identifierades när man studerade populationsstrukturen med hjälp av principalkomponentanalys, individuella släktskapskoefficienter och beräknade populationsträd. Det fanns bevis för parallella urvalshändelser i regioner som hade kandidatgener som potentiellt var inblandade i morfologiska drag, färg och anpassning. Det fanns olika mutationer som ansvarar för samma fenotypiska egenskaper i *MC1R* (pälsfärg). Detta bekräftades av sekvensering av *MC1R* i de tre raserna som hade hårda selektiva svep i regionen som inkluderade *MC1R*. I artikeln drogs slutsatsen att täta genetiska markörer i många populationer kan dechiffrera evolutionär historia hos populationer och adaptiva mutationer.

I artikel V studerades populationsstrukturen i svenska raser genom att genotypa fem svenska fårraser för över 500 000 SNP-markörer. De fem raserna

Dalapälsfår, Fjällnäsfår, Gotlandsfår, Gutefår och Klövsjöfår, som ingår i denna artikel, valdes på grund av deras olika geografiska ursprung och populationer. Det visade sig att dessa är fem olika raser med Gutefår och Gotlandsfår närmare släkt med varandra. Analyserna av populationsstrukturen upprepades sedan tillsammans med det franska datasetet från artikel IV och ytterligare genotyper med medelhög densitet från fårraser från hela världen från Sheep HapMap-projektet. Alla de fem svenska raserna grupperade sig med andra nordeuropeiska kortsvansfår i studien. Raserna var alla distinkta och svenska raser har i allmänhet ackumulerat mer drift i jämförelse med andra raser.

9 Synopsis de la thèse

Les moutons domestiques (*Ovies aries*) sont élevés dans le monde entier. Ils forment un groupe hétérogène et sont élevés pour la production de viande, de lait ou de laine, ou à des fins de conservation. Certaines races ovines ont une faible taille de population et sont présentes dans des environnements très spécifiques tandis que d'autres races sont utilisées partout dans le monde. Les moutons sont élevés dans un large éventail de systèmes de production allant du système extensif au système intensif. Les différentes pressions de sélection, naturelles et artificielles, opérées sur les moutons domestiques, font de cette espèce un bon modèle d'étude de l'adaptation. En outre, l'étude de la diversité génétique des moutons domestiques est importante, en particulier pour la conservation des races locales qui présentent un risque d'extinction. En maintenant la diversité génétique, les défis futurs, comme un environnement changeant ou les maladies émergentes, peuvent être relevés plus rapidement. Cette thèse comprend cinq études sur la diversité génétique et la structure de la population en mettant l'accent sur des races ovines suédoises et françaises.

L'article I présente l'étude de la race ovine Gute, une race suédoise primitive, à cornes, à queue courte et originaire de Gotland, une île de la mer Baltique. La consanguinité et la taille effective de la population ont tout d'abord été calculées en utilisant les informations généalogiques provenant de la population Gute enregistrée en Suède jusqu'en 2012. De plus ces mêmes paramètres ont été calculés sur un échantillon de 100 moutons Gute en utilisant des données de pedigree et des génotypages de microsatellites. Dans les deux cas, les estimations des coefficients de consanguinité de ces deux populations étaient faibles (0,038 ; 0,018). Les estimations à partir de génotypages de microsatellites ont montré que l'hétérozygotie multilocus moyenne était élevée (1,01845), que la consanguinité calculée selon la méthode de Ritland était faible (-0,03931) et que cinq des sept marqueurs n'étaient pas en équilibre de Hardy-Weinberg en raison d'un excès d'hétérozygotie. La taille effective de la population était de 155,4 et la moyenne harmonique de 88,3. Au final, il y avait

des indications claires que cette race fait actuellement partie d'un programme de conservation efficace.

L'article II présente des estimations des relations génétiques entre cinq races suédoises. Des individus des races ovines suédoises de Finewool, Gute, Klövsjö, Roslag et Värmland ont été génotypés pour six rétrovirus endogènes de moutons (enJSRV). Les moutons des races Klövsjö et Värmland avaient des fréquences faibles pour les enJSRV et pour les moutons suédois Finewool, Gute et Roslag, la fréquence la plus élevée a été trouvée pour enJSRV-7. L'absence de enJSRVs et la présence de enJSRV-7 signent des races primitives. Les races Finewool, Gute et Roslag avaient des fréquences modérées pour enJRSV-18, ce qui signe des races de moutons plus modernes.

Dans l'article III, les séquences codantes des gènes codant pour la protéine *ASIP* (agouti signalling protein) et *MC1R* (melanocortin 1 receptor) ont été séquencées pour étudier les mutations associées à la couleur noire de la toison de sept races suédoises. L'étude de la couleur de la toison permet non seulement de déterminer la fonction des gènes et l'étude des mutations associées à la couleur et à la qualité de la toison mais aussi de recueillir des informations sur l'histoire des races. Les races suédoises sont de bons modèles pour étudier la couleur de la toison car elles présentent des couleurs et des motifs variés. La coloration noire des moutons Klövsjö et Roslag pourrait s'expliquer par deux variants connus dans *ASIP* (allèle noir récessif: g.100-105del et g.5172T>A) et celle des Finewool soit par une des deux mutations dans *MC1R* (allèle noir dominant: c.218T>A ou c.361G>A) ou dans *ASIP*. En revanche, seulement un tiers des animaux à toison noire dans les races Gute, Gotland et Värmland ont des génotypes expliquant leur couleur. Cela pourrait être dû à la présence d'individus gris, et que l'allèle gris, qui est une duplication, pourrait rendre plus complexe la transmission génétique de la couleur noire de la toison. Enfin, une nouvelle mutation non synonyme (c.452G>A) a été identifiée dans *MC1R* chez les races Gute, Gotland et Värmland, et une nouvelle duplication de *MC1R* a été identifiée dans la race Gotland.

L'article IV présente l'étude de la structure des populations et les signatures de sélection chez les moutons français. Au total, 27 races ont été génotypées pour plus de 500 000 marqueurs SNP. L'étude de la structure de ces populations, par une analyse en composantes principales, le calcul des coefficients de coancestralité et l'estimation d'arbres phylogénétiques, a révélé que ces races forment deux groupes, les populations du Nord et celles du Sud et a permis d'identifier des événements de croisements anciens entre les deux groupes. Des situations de sélection parallèles dans des régions porteuses de gènes candidats potentiellement impliqués dans des caractères

morphologiques, la couleur de la robe ou l'adaptation ont été identifiés. La présence de différentes mutations dans *MC1R*, responsables du même caractère phénotypique (couleur de la toison) a été démontrée. Cela a été confirmé par le reséquençage de *MC1R* dans trois races qui présentaient un balayage sélectif dans la région du gène *MC1R*. Ce travail conclut que les données génétiques denses dans de nombreuses populations peuvent élucider l'histoire des populations et des mutations adaptatives.

Dans l'article V, la structure de population des races suédoises a été étudiée en génotypant cinq races ovines suédoises pour plus de 500 000 marqueurs SNP. Les cinq races, Dalapäls, Fjällnäs, Gotland, Gute et Klövsjö, incluses dans cet article ont été choisies en raison de leurs différentes origines géographique et de population. Nous avons démontré que cinq races sont distinctes, les Gute et les Gotland étant les plus proches. Les analyses de la structure de la population ont ensuite été répétées en incluant les d'une part les données des races ovines français de la quatrième publication et d'autre part les génotypes supplémentaires de moyenne densité des races mondiales du projet Sheep Hap Map. Les cinq races suédoises sont groupées avec d'autres races de moutons à queue courte nord-européennes. Les races étaient toutes distinctes les unes des autres et les races suédoises en général ont accumulé plus de dérive génétique que les autres races.

References

Adalsteinsson, S. et al., 1978. A possible genetic interpretation of the colour variation in the fleece of the Gotland and Goth sheep. *Annales de Génétique et de Sélection Animale*, 10(3), pp.329–342.

Adalsteinsson, S., 1983. Inheritance of colours, fur characteristics and skin quality traits in North European sheep breeds: A review. *Livestock Production Science*, 10(6), pp.555–567.

Andersson, L. et al., 2015. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome biology*, 16(1), p.57. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84935833067&partnerID=tZOtx3y1.

Andersson, L., 2001. Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet*, 2(2), pp.130–138. Available at: http://dx.doi.org/10.1038/35052563.

Andersson, L., 2012. How selective sweeps in domestic animals provide new insight into biological mechanisms. *Journal of Internal Medicine*, 271(1), pp.1–14.

Barker, J.F.S., 2001. Conservation and management of genetic diversity: a domestic animal perspective. *Canadian Journal of Forest Research*, 31(4), pp.588–595. Available at: http://www.nrcresearchpress.com/doi/abs/10.1139/x00-180.

Berg, P., Nielson, J. & Sørensen, M., 2006. EVA: Realized and predicted optimal genetic contributions. In *8th World Congress on Gentics Applied to Livestock Production*.

Boichard, D., 2002. Pedig : a Fortran Package for Pedigree Analysis Suited for Large Populations. In *7th World Congress on Genetics Applied to Livestock Production*.

Boitard, S. et al., 2016. Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics*, 203(1), pp.433–450. Available at: http://www.genetics.org/cgi/doi/10.1534/genetics.115.181594.

Bonhomme, M. et al., 2010. Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics*, 186(1), pp.241–262.

Bowles, D., Carson, A. & Isaac, P., 2014. Genetic distinctiveness of the Herdwick sheep breed and two other locally adapted hill breeds of the UK. *PLoS ONE*, 9(1), pp.0–6.

Boyko, A.R. et al., 2010. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biology*, 8(8), pp.49–50.

Bunch, T.D., Foote, W.C. & Spillett, J.J., 1976. Translocations of acrocentric chromosomes and their implications in the evolution of sheep (Ovis). *Cytogenetic and Genome Research*, 17(3), pp.122–136.

Carneiro, M. et al., 2016. Dwarfism and Altered Craniofacial Development in Rabbits Is Caused by a

12.1 kb Deletion at the HMGA2 Locus. *Genetics*, 205(February), pp.955–965.

Chang, C.C. et al., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), p.7. Available at: http://gigascience.biomedcentral.com/articles/10.1186/s13742-015-0047-8.

Chessa, B. et al., 2009. Revealing the History of Sheep Domestication Using Retrovirus Integrations. , 324(April), pp.532–536.

Cieslak, M. et al., 2011. Colours of domestication. *Biological Reviews*, 86(4), pp.885–899.

Clark, A.G. et al., 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, 15(11), pp.1496–1502.

Coltman, D.W. et al., 1999. Parasite-Mediated Selection against Inbred Soay Sheep in a Free-Living , Island Population Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2640828 Accessed : 06-04-2016 09 : 24 UTC Your use of the JSTOR archive i. *Evolution*, 53(4), pp.1259–1267. Available at: http://www.jstor.org/stable/2640828.

Crawford, A.M. & Cuthbertson, R.P., 1996. Mutations in sheep microsatellites. *Genome research*, 6(643), pp.876–879.

Dýrmundsson, Ó.R. & Niżnikowski, R., 2010. North European short-tailed breeds of sheep: a review. *Animal*, 4(8), pp.1275–1282.

Eriksson, J. et al., 2008. Identification of the Yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics*, 4(2).

Falconer, D.S. & Mackay, T.F.C., 1996. *Introduction to Quantitative Genetics* fourth ed., Harlow, England: Pearson Education Limited.

FAO Commission on Genetic Resources for Food and Agriculture, 2015. The second report on the state of the world's animal genetic resorces for food and agriculture.

Fariello, M.-I. et al., 2014. Selection signatures in worldwide sheep populations. *PLoS ONE*, 9(8), p.e103813.

Fariello, M.I. et al., 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193(3), pp.929–941.

Fernández, J. et al., 2005. Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics*, 170(3), pp.1313–1321.

Fontanesi, L. et al., 2012. Analysis of polymorphisms in the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) genes and association with coat colours in two Pramenka sheep types. *Small Ruminant Research*, 105(1–3), pp.89–96.

Fontanesi, L. et al., 2011. Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) genes. *Animal*, 5(1), pp.8–17.

Fontanesi, L. et al., 2010. Sequence characterization of the melanocortin 1 receptor (MC1R) gene in sheep with different coat colours and identification of the putative e allele at the ovine Extension locus. *Small Ruminant Research*, 91(2–3), pp.200–207.

Frankham, R., Briscoe, D.A. & Ballou, J.D., 2002. *Introduction to Conservation Genetics*, Cambridge University Press, 2002.

Frichot, E. et al., 2014. Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), pp.973–983.

Gandini, G.C. & Villa, E., 2003. Analysis of the cultural value of local livestock breeds: A methodology. *Journal of Animal Breeding and Genetics*, 120(1), pp.1–11.

Gratten, J. et al., 2007. Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proceedings. Biological sciences / The Royal Society*, 274(1610), pp.619–626.

Gratten, J. et al., 2010. The genetic basis of recessive self-colour pattern in a wild sheep population. *Heredity*, 104(2), pp.206–214. Available at: http://dx.doi.org/10.1038/hdy.2009.105.

Groeneveld, L.F. et al., 2010. Genetic diversity in farm animals - A review. *Animal Genetics*, 41(SUPPL. 1), pp.6–31.

Haase, B. et al., 2007. Allelic heterogeneity at the equine KIT locus in dominant white (W) horses. *PLoS Genetics*, 3(11), pp.2101–2108.

Hayes, B.J., Lewin, H.A. & Goddard, M.E., 2013. The future of livestock breeding: Genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics*, 29(4), pp.206–214. Available at: http://dx.doi.org/10.1016/j.tig.2012.11.009.

Hiendleder, S. et al., 2002. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. , (April), pp.893–904.

Hoffmann, I., 2010. Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, 41(SUPPL. 1), pp.32–46.

Hoffmann, I., 2011. Livestock biodiversity and sustainability. *Livestock Science*, 139(1–2), pp.69–79. Available at: http://dx.doi.org/10.1016/j.livsci.2011.03.016.

Hubbard, J.K. et al., 2010. Vertebrate pigmentation: from underlying genes to adaptive function. *Trends in Genetics*, 26(5), pp.231–239. Available at: http://dx.doi.org/10.1016/j.tig.2010.02.002.

Jombart, T., 2008. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), pp.1403–1405.

Kijas, J.W. et al., 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, 10(2).

Larson, G. & Burger, J., 2013. A population genetics view of animal domestication. *Trends in Genetics*, 29(4), pp.197–205. Available at: http://dx.doi.org/10.1016/j.tig.2013.01.003.

Leroy, G. et al., 2015. How do introgression events shape the partitioning of diversity among breeds: a case study in sheep. *Genetics Selection Evolution*, 47(1), p.48. Available at: http://www.gsejournal.org/content/47/1/48.

Li, C. et al., 2014. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. *PLoS ONE*, 9(5).

Li, M. et al., 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet*, 45(12), pp.1431–1438. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24162736.

Li, M.-H., Tiirikka, T. & Kantanen, J., 2014. A genome-wide scan study identifies a single nucleotide substitution in ASIP associated with white versus non-white coat-colour variation in sheep (Ovis aries). *Heredity*, 112(2), pp.122–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24022497.

Li, M.H. et al., 2011. A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS ONE*, 6(11).

Linderholm, A. & Larson, G., 2013. The role of humans in facilitating and sustaining coat colour variation in domestic animals. *Seminars in Cell and Developmental Biology*, 24(6–7), pp.587–593. Available at: http://dx.doi.org/10.1016/j.semcdb.2013.03.015.

Liu, R. et al., 2016. Detection of genetic diversity and selection at the coding region of the melanocortin

receptor 1 (MC1R) gene in Tibetan pigs and Landrace pigs. *Gene*, 575(2), pp.537–542.

Meuwissen, T.H.E. & Luo, Z., 1992. Computing inbreeding coefficients in large populations. *Genetics, selection, evolution*, 24, pp.305–313.

Mignon-Grasteau, S. et al., 2005. Genetics of adaptation and domestication in livestock. *Livestock Production Science*, 93(1), pp.3–14.

Nadler, C.F. et al., 1973. Cytogenetic differentiation, geographic distribution, and domestication in Palaearctic sheep (Ovis). *Zeitschrift für Säugetierkunde*, 38, pp.109–125.

Niemi, M. et al., 2013. Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (Ovis aries) in Finland: comparison with contemporary sheep breeds. *Genetics Selection Evolution*, 45(2), pp.1–14. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3558444&tool=pmcentrez&rendertype=abstract.

Nikolic, N. & Chevalet, C., 2014. Detecting past changes of effective population size. *Evolutionary Applications*, 7(6), pp.663–681.

Norris, B.J. & Whan, V. a., 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Research*, 18(8), pp.1282–1293.

Petersen, J.L. et al., 2013. Genome-Wide Analysis Reveals Selection for Important Traits in Domestic Horse Breeds. *PLoS Genetics*, 9(1).

Pickrell, J.K. & Pritchard, J.K., 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, 8(11).

Poplin, F., 1979. Origine nouvelle perspective paléontologique par marronnage. *Annales de genetique et de selection animale*, 11(2), pp.133–143.

Purcell, S. et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), pp.559–575.

Purcell, S. & Chang, C., 2014. PLINK. Available at: https://www.cog-genomics.org/plink2.

Rajakumari, S. et al., 2013. EBF2 determines and maintains brown adipocyte identity. *Cell Metabolism*, 17(4), pp.562–574. Available at: http://dx.doi.org/10.1016/j.cmet.2013.01.015.

Rannamäe, E. et al., 2016. Three thousand years of continuity in the maternal lineages of ancient sheep (Ovis aries) in Estonia. *PLoS ONE*, 11(10), pp.1–19.

Ritland, K., 1996. Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetic Research*, 67, pp.175–185.

Rousset, F., 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8(1), pp.103–106.

Royo, L.J. et al., 2008. Differences in the expression of the ASIP gene are involved in the recessive black coat colour pattern in sheep: Evidence from the rare Xalda sheep breed. *Animal Genetics*, 39(3), pp.290–293.

Rubin, C.-J.J. et al., 2012. Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences*, 109(48), pp.19529–19536. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23151514.

Rupp, R. et al., 2015. A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *PLoS Genetics*, 11(12), pp.1–19. Available at: http://dx.doi.org/10.1371/journal.pgen.1005629.

Scheet, P. & Stephens, M., 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American*

journal of human genetics, 78(4), pp.629–44. Available at: http://www.sciencedirect.com/science/article/pii/S000292970763701X.

Smit, M.A. et al., 2002. Identification of an agouti-like locus in sheep. *Animal genetics*, 33(5), pp.383–385.

Storey, J., 2015. qvalue: Q-value estimation for false discovery rate control.

Taberlet, P. et al., 2008. Are cattle, sheep, and goats endangered species? *Molecular Ecology*, 17(1), pp.275–284.

Taberlet, P. et al., 2011. Conservation genetics of cattle, sheep, and goats. *Comptes Rendus - Biologies*, 334(3), pp.247–254.

Tapio, M. et al., 2010. Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. *BMC genetics*, 11, p.76.

Tapio, M. et al., 2005. Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Molecular Ecology*, 14(13), pp.3951–3963.

Thornton, P.K. et al., 2009. The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. *Agricultural Systems*, 101(3), pp.113–127. Available at: http://dx.doi.org/10.1016/j.agsy.2009.05.002.

Våge, D.I. et al., 2003. Mapping and Characterization of the Dominant Black Colour Locus in Sheep. *Pigment Cell Research*, 16(6), pp.693–697.

Våge, D.I. et al., 1999. Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome*, 10(1), pp.39–43.

Väli, Ü. et al., 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, 17(17), pp.3808–3817.

Wang, J., 2011. Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, 11(1), pp.141–145.

Westberg Sunesson, K., 2015. *Genetic variation of ASIP and MC1R in past and present sheep of Gotland*. Swedish University of Agricultural Sciences. Available at: http://stud.epsilon.slu.se/7613/11/westberg_sunesson_k_150309.pdf.

Yang, J. et al., 2011. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1), pp.76–82. Available at: http://dx.doi.org/10.1016/j.ajhg.2010.11.011.

Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), pp.11597–11604.

Acknowledgements

I would like to thank all the people and organizations that have supported me and my research over the past four years.

Thanks to the European Graduate School in Animal Breeding and Genetics and SLU. Having this opportunity to come to Europe and study in Sweden and France has been very special: I have met many people and learned many things.

Thank-you to my supervisors Anna, Lisa, Gwenola and Bertrand: I've really appreciated your guidance, knowledge and patience. Thanks Anna and Lisa for stepping up and filling supervisor and assistant supervisor positions. Thanks Gwenola and Bertrand for welcoming me in France for a year and all the time since that we've spent working together. Finally, I need to thank Anna Näsholm, who unfortunately passed away: I really appreciated the time I had with her and everything she contributed.

I would like to thank all of my collaborators as they have made it possible for me to explore many research topics. I want to specifically thank Sofia Mikko, for her collaboration and enthusiasm for our coat colour project.

I would like to thank the breed organizations and producers that have helped with this thesis. I appreciate the interest and support for my research. I'd also like to thank technicians and veterinarians who collected samples, and phenotypes.

Thanks to all my colleagues who have made me feel welcome at both SLU and INRA. Thanks for all the fika and interesting conversations that we've had at SLU. Thanks to my colleagues in the student unions at SLU for introducing me to the workings of the university and new and diverse perspectives.

Thanks to all the new friends that I've made in the past four years. Thanks to my Uppsala friends (Agnese, Alberto, Ahmed, André, Amabel, Berihu, Bingjie, Daniela, Emelie, Hadrien, Jovana, Juan, Kim, Margot, Merina, Nancy, Sandrine, Shizhi, Sofia, Suvi, Thu, Tomas and Xiaowei). We've had a lot of fun together having effective fikas, exploring new places and spending

holidays together. Thanks to my Toulouse friends (Clémentine, Jason, Marjorie, Marjorie, Morgane and Pauline) who introduced me to French cuisine and countryside. Thanks also to my fellow EGS-ABG students who I've gone to conferences, taken courses and had holidays with.

Finally, thanks to my family, and friends in Canada. Thanks for your support and thanks for understanding that I couldn't always be there like I wish I could've been.