Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Animal Breeding and Genetics

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

Robert Mukiibi

Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics
449
Uppsala 2014

Master's Thesis, 30 hp

Erasmus Mundus Programme – European Master in Animal Breeding and Genetics

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Animal Breeding and Genetics

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

Robert Mukiibi

Supervisors:

Göran Andersson, SLU, Department of Animal Breeding and Genetics
Anna Johansson, SLU, Department of Animal Breeding and Genetics

Examiner:

Elisabeth Jonas, SLU, Department of Animal Breeding and Genetics

Credits: 30 HEC
Course title: Degree project in Animal Science
Course code: EX0556
Programme: Erasmus Mundus programme
– European Master in Animal Breeding and Genetics
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2014
Name of series: Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics, 449
On-line publication: http://epsilon.slu.se

Key words: indigenous Swedish breeds, enJSRV, primitive, endogenous, polymorphism, proviruses

Education and Culture
Erasmus Mundus

Table of contents

SUMMARY

The present modern sheep is believed to have evolved from its wild ancestors 2.5 million years ago. Furthermore archeological evidence suggest that sheep has been domesticated by man since 8500-9000 years BC for their beneficial products to man including; milk, meat, skin and wool. Through centuries, sheep have been spread from the center of domestication to different parts of the world where they have adapted to environmental factors in such areas giving rise to what is known today as local or indigenous breeds. Indigenous breeds are generally characterized by the presence of horns in both rams and ewes, dark and coarse fleece and moulting coat. Such local breeds in Sweden include; Dala Fur sheep, Gestrike sheep, Helsinge sheep, Klövsjö sheep, Svärdsjö sheep, Åsen sheep, Roslagen sheep, Värmland sheep, Gotland sheep and Gute sheep. Though there have been some studies on some of these breeds, information on genetic distinctiveness between them remains scarce.

The sheep genome contains more than 30 polymorphic endogenous proviruses (enJSRV) that are related to the Jaagsiekte retrovirus (JSRV), a Beta retrovirus known for causing Ovine Pulmonary Adenocarcinoma. These proviruses are remains of exogenous retroviruses that infected germ cells of ancestors or great grandparents and have been passed on in a Mendelian manner to the present day sheep. Some of these endogenous proviruses have been found to have different functions in sheep species including the protection of hosts from infection by related defective exogenous viruses and also aid in conception and implantation during reproduction. Recently these polymorphic loci have been used to study the domestication trend of sheep and genetic distinctiveness between some sheep breeds.

The aim of the current study was to use 6 of the polymorphic proviruses to define genetic difference between five local Swedish breeds. The presence or absence of each of these proviruses in the studied individuals was tested using polymerase chain reaction (PCR) and agarose gel-electrophoresis on genomic DNA of 66 individuals. Results were analyzed based on the frequencies of the individual loci in the tested samples, retrotype or provirus combination frequencies and a principal component analysis was performed based on the provirus frequencies to visualize genetic distances between studied breeds.

The study revealed a close genetic relationship sheep between Gute, Swedish fine wool sheep and Roslag. Furthermore we also found a strong genetic closeness between Värmland and Klövsjö sheep breeds. Although all the local Swedish breeds tested were found to have genomes of ancestral or primitive makeup, Värmland and Klövsjö were more primitive than others. Further comparison of each breed with Texel showed that Swedish fine wool sheep was more close to the Texel breed than any other breed in the study. However, the number of samples used and their distribution among herds was small. Hence, more samples are needed for a larger study to allow for better conclusions on the genetic diversity or relatedness between these breeds and other Swedish breeds not included in our study.

INTRODUCTION

Origin and domestication of modern domestic sheep breeds

It is widely believed that the modern domestic sheep evolved 2.5 million years ago from the three different lineages Urial, Mouflon and Argali which still exist up to today (Dahlberg 2012). However, according to recent studies on mitochondrial DNA, evidence suggests that Urial and Argali are least likely to be ancestors of the modern sheep (Grigaliunaite *et al.* 2004 Bruford and Townsend 2006, Chen *et al.* 2006). According to archeological evidence, the European Mouflon is no longer considered a possible ancestor for the modern sheep but a feral remnant of early domestication (Hiendleder *et al.* 1998, Grigaliunaite *et al.* 2004, Bruford and Townsend 2006). Sheep and goats were the first livestock animals to be domesticated by man about 8500-9000 years BC in the region around the near East. Sheep were primarily kept for meat, and later also for milk, skin, horns, wool and hair as secondary products (Chessa *et al.* 2009, Tapio *et al.* 2010, Dahlberg 2012).

Environmental factors like climate coupled with human-mediated selection of these animals for products such as meat, wool and milk, has through the course of time caused these animals to evolve into breeds that can withstand the local conditions and produce the desired product. In Sweden, according to archeological evidence, sheep production was introduced between 3000-4000 BC (NorthSheD-homepage). Since the introduction of the first sheep population in several Swedish regions, farmers have continued to select for several traits within these sheep, hence creating populations of animals that are well adapted to the agro-ecological systems in those localities; giving rise to what is termed as Swedish local sheep breeds.

Swedish local sheep breeds

Swedish local sheep breeds mainly belong to the North Europe short tailed sheep. These are originally from Russia and they were spread to different North European countries including Sweden by Vikings (Dýrmundsson and Niżnikowski 2010). In their study, Dýrmundsson and Niżnikowski (2010) further noted that through a number of generations of both natural and artificial selection, short tail sheep have adapted to harsh local climates in areas or countries they were reared to evolve into local or indigenous breeds of those regions. Some have been crossed with other non-short tail breeds to improve on the performance for the desired sheep products like wool, mutton and milk both in terms of quality and quantity. In Sweden such breeds include Dala Fur sheep, Gestrike sheep, Helsinge sheep, Klövsjö sheep, Svärdsjö sheep, Åsen sheep, Roslagen sheep, Värmland sheep, Gotland sheep and Gute sheep (Dýrmundsson and Niżnikowski 2010). The current use and historical background of some of these breeds is described further as they were used in the current study.

Gute sheep

Gute sheep is the most primitive among the native Swedish breeds. This sheep breed originates from the Island of Gotland located in the Baltic Sea. In this breed, both rams and ewes have horns that are slightly twisted. Also the sheep has a narrow and slightly wedged head whose color varies from white to black. The wool from this sheep is mixed with white fine wool, black long coarse wool and Kemp fibers, hence giving the sheep a grey color. The normal average weight of ewe lambs at birth is 2.25

kg and 2.5 kg for ram lambs. Fully grown ewes weigh about 45-60 kgs and rams weigh 70-100 kg (FG-homepage).

Figure 1: Ram of Gute sheep with white and black mottled head, and large twisted horns (Photo by Stig Karlsson)

Klövsjö sheep

The present Klövsjö sheep evolved from a couple of flocks surrounding Klövsjö, a small village in the county of Jämtland in Sweden. The sheep are usually white or black in color with some white marks in the face. This breed has got a long life span of about 15 years of age. They are usually polled but rams sometimes have small brittle horns. It is a small breed, with ewes and rams weighing about 40-60 kg and 60-70 kg respectively. The height of ewes and rams is on average 65 cm and 66 cm, respectively. It is a prolific breed common giving birth to twins, triplets and sometimes quadruplets (FSA-homepage , Dahlberg 2012).

Figure 2: Polled Klövsjö sheep ewe with white and black face, and a white neck. (Photo by Jan Gosselman)

Swedish Fine wool sheep

It is a short tailed Swedish breed that originated from Finland. The sheep breed is common in southern and middle regions of Sweden. It is mainly kept for wool and meat production due to its high fertility with an average of 2.7 lambs per lambing. The sheep are mainly white but a small number of black may occur. Both ewes and rams are polled and produce uniform, very soft and glossy elastic wool that is suitable for fine knitting. Rams weigh on average 75 kg while ewes weigh 50 kg at mature age (NorthSheD-homepage).

Roslag sheep

This breed was rediscovered in the 1990s by Nils Dahlbeck. Previously the breed was popular on every farm on the Raggarön Island in Roslagen, in the county of Uppland. They were greatly raised for their long wool that was used to make socks, gloves and other types of clothing. The breed disappeared in the 1900s as farmers adopted more productive breeds, hence leaving only one flock in Raggarön. Roslag sheep are small, with rams weighing about 50 kg and ewes 30-40 kg. The sheep are usually white or black with white markings, or mottled. The rams usually carry large horns while females are polled. Ewes hardly produce more than one lamb per lambing (FSA-homepage , Dahlberg 2012).

Figure 3: Polled Roslag sheep ewe, polled, and all white in color (Photo by Ste phan Noll)

Värmland sheep

The breed originated from the Värmland region of Sweden. The sheep are relatively small with rams weighing about 60-70 kg and 40-65 kg for ewes. The ewes have good mothering ability and high fertility where twins are common, as well as triplets. The sheep can have several coat colors like: black, gray, brown and white. Ewes are usually polled, though buttons and in rare cases small horns occur. Rams are mainly horned but polled rams also occur in the population. The wool varies from rug type to fine wool that can be used for different crafting. Värmland sheep are very easy to handle and good grazers (FSA-homepage)

Figure 4: Värmland sheep ram, with big curved horns, white faces spotted with brown and white wool (By Martin Stoltze, Wikimedia commons)

The above described breeds have been developed through within breed selection and sometimes crossing with other non-native breeds like Texel to improve their productivity as noted by Dýrmundsson and Niżnikowski (2010). This is summarized in the Table 1 below:

Table 1: Summary of the present status of five local Swedish sheep breeds as adapted from Dýrmundsson and Niżnikowski 2010.

Breed Name	Importance as compared to national sheep population	Incidence of crossbreeding (and breeds involved)	Products in the order of their priority ranking	Current breeding population size
Gute sheep	Little	Low (Texel)	Meat, wool, skin , milk	7000
Klövsjö sheep	Little	Low (Gute sheep)	Wool, skin, meat, milk	97
Swedish Fine wool sheep	Considerable	Considerable (Texel, Oxford down and others)	Wool, meat, skin, milk	3669
Roslag sheep	Little	Low	Wool, meat, skin, milk	675
Värmland sheep	Little	Low	Meat, wool, skin, milk	1814

Sheep and Endogenous Retroviruses

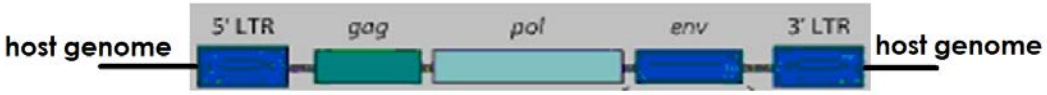

Figure 5: An endogenous retrovirus contains three major genes including; Pol gene which codes for reverse transcriptase, protease, intergrase and ribonuclease, Env-gene codes for envelope protein, transmembrane protein, surface protein membrane and the Gag gene which codes for matrix proteins including major caspid protein, outer matrix membrane.

Eukaryotic genomes contain several retrovirus-like elements (Figure 5). These are called Endogenous retroviruses (ERVs) because they are stably embedded into the hosts' genomes. As an example, 8% of the human genome is composed of ERVs and their remnants (Arnaud *et al.* 2008, Jern and Coffin 2008). ERVs enter into the hosts' genomes through the insertion of the effective exogenous retroviruses into the genome of the germ-line (Chessa *et al.* 2009). After insertion into the germ-line these retroviruses termed as proviruses are passed on to the next generation of that individual vertically in a Mendelian manner, hence circulating into the population with time (Palmarini *et al.* 2004, Arnaud *et al.* 2008, Chessa *et al.* 2009). New integrations could be harmful, neutral or advantageous to their host. The advantageous and neutral proviruses are retained in the population by positive selection and neutral selection respectively. However, those integrations that are harmful to the host are selected against to purge them out of the population. Also through a number of generations, new insertions are gradually rendered inactive or suppressed by mutations or by

epigenetic mechanisms like DNA methylation, Histone modifications and RNA interference (Arnaud *et al.* 2008, Jern and Coffin 2008, Maksakova *et al.* 2008, Sistiaga-Poveda and Jugo 2014).

However, even after going through several modifications, some endogenous retroviruses maintain their intact open reading frames (ORFs) for one or more genes (Arnaud *et al.* 2008). This could imply that these ERVs serve some kind of functions to the host organisms. Such functions have been found to be: involvement in the interference of the replication cycle of related pathogenic retroviruses. Integration sometimes occurs within or near genes and here they regulate transcription as promoters or enhancers of these genes, they are able to do this because of their long terminal repeats (LTR). The LTR has several potential binding sites for transcription factors to regulate and initiate transcription since it is formed at the moment of integration as a promoter for transcription of the retroviral genes (Jern and Coffin 2008). There is also building evidence showing that the development of the placenta in mammals such as humans, mice and sheep needs functional retroviral envelope genes referred to as Syncytins (Arnaud *et al.* 2008). Endogenous proviruses are greatly involved in organization and plasticity of the host genome; here they bring about shuffling of different genomic regions including exons, introns, enhancers and promoter, hence causing new dynamic functions in the host genome (Palmarini *et al.* 2004, Jern and Coffin 2008). This is achieved through homologous recombination between two proviruses on the same chromosome or two chromosomes, partial loss of the provirus as a result of homologous recombination between the two LTRs leaves solo LTRs (Palmarini *et al.* 2004, Jern and Coffin 2008). Some ERVs serve an important function to their hosts of preventing infection of the host by related exogenous retroviruses through early blockage of retroviral receptors, or through late blockage by transdominant interference (Arnaud *et al.* 2008, Jern and Coffin 2008, Sistiaga-Poveda and Jugo 2014).

ERVs could may be referred to as ancient or modern endogenous retroviruses. They are ancient, if they got integrated into the host's genome before speciation and all individuals in the population possess them (i.e., fixed in the population) and they are majorly silenced, while the modern endogenous retroviruses are more similar to the exogenous retroviruses as not much mutations or deletions have accumulated in them. Modern ERVs are in some cases capable of producing infective particles within the host, they also polymorphic as they still undergoing endogenization and are not fixed in the population (Arnaud *et al.* 2008).

It has been estimated that there are over 30 endogenous retroviruses (Sistiaga-Poveda and Jugo 2014) in the sheep genome, that are at the nucleotide level 85-89% related to the Jaagsiekte sheep retrovirus (JSRV) which causes ovine pulmonary adenocarcinoma (OPA) a major viral disease to sheep (Arnaud *et al.* 2008, Sistiaga-Poveda and Jugo 2014). These are referred to as endogenous Jaagsiekte retroviruses (enJSRVs). Most of the enJSRVs are defective due to inactivation resulting from several mutations, deletions and recombinations (Arnaud *et al.* 2008). Howbeit, five of these enJSRVs (enJSR-7, -15, -16, -18 and -26) still possess complete genome arrangements with an undisturbed open reading frames (ORFs) for all the retroviral genes (Arnaud *et al.* 2008, Sistiaga-Poveda and Jugo 2014). Hence, these retroviruses are thought to be recent or modern integrations (Arnaud *et al.* 2008). The integration of these proviruses into the host genomes is random, meaning that integration can be in any part of the genome. Some enJSRVs like enJSRV-6 are already fixed in all the sheep

species while others are only present in some individuals and not in others that is to say they are polymorphic or segregate in the population, and their presence or absence segregate in the population. As described above, enJSRVs have been able to remain stable in the sheep genome because they have been positively selected for since they are thought to serve several functions like helping in conception and development of the placenta, also these proviruses help in interfering with the replication of pathogenic retroviruses with in the host by both early receptor blockade or by a post integration mechanism called late blockade. An example where enJSRVs utilize the late blockade to interfere with replication of exogenous Jaagsiekte sheep retrovirus (JSRV) is exhibited by enJS56A1 which prevents the formation of JSRV viral particles in the post replication stage. Palmarini *et al.* (2004) observed that JSRV particles could not be formed in cells transfected with enJS56A1 plasmid, although Gag was greatly expressed in the lysed cells. This was because enJS5A1 possesses a mutation in the residue position 21 of the Gag in the matrix (MA) protein; in this position enJS5A1 possesses a tryptophan, while JSRV contains a arginine as other *Betaretroviruses*. This difference causes the Gag proteins of JSRV and enJS56A1 that co-localize early in synthesis to co-immunoprecipitate in the cells of expression by the transdorminant mechanism (Palmarini *et al.* 2004, Arnaud *et al.* 2008, Armezzani *et al.* 2011).

enJSRVs as molecular markers in sheep

As mentioned above, the sheep genome contains a great number of enJSRVs (>30), some of which are fixed in all sheep species while others are still polymorphic in the sheep species (Chessa *et al.* 2009). Polymorphic enJSRVs have been used as highly informative markers in several molecular genetics studies; due to the fact the existence of an endogenous retrovirus in the host's (sheep) genome is an outcome of a single and irreversible integration event. This means that any populations sharing the same endogenous locus at a common genomic location are phylogenetically closely related to each other (Chessa *et al.* 2009). Also ENVs are good as genetic markers because of their large size and widely spread within the genomes of their hosts. By comparing the nucleotide mutations in the LTRs of the same ERVs in two species we can infer the age of divergence of these species. To trace sheep domestication in different regions of the world, Chessa *et al.* (2009) used a group of six polymorphic enJSRVs (including enJSRV-7, enJSRV-8, enJSRV-15, enJSRV-16, enJSRV-16 and enJS5F16). Also in a more recent study by Bowels *et al.* (2014), the same six polymorphic enJSRV loci were used to describe the genetic difference between Herdwick sheep breed and two other local breeds (Dalesbred and Rough Fell). Both methods analyzed the distribution of different enJSRV combinations (retrotypes) in the studied breeds from which they were able to draw several conclusions about genetic relationships between the breeds studied. For example, Chessa and colleagues defined for the first time the genetic difference between modern and primitive sheep breeds at genomic level.

Generally there is an extensive call for conservation of Swedish local breeds, but this requires the definition of the genetics of each of the breeds and comparisons between them. Currently only one study by Dahlberg (2012) has been performed to specifically study the genetic variability between local Swedish sheep breeds (Dahlberg 2012). In her study she compared nine of the local breeds using microsatellite markers. Studies carried out by Chessa *et al.* (2009) and Tapio *et al.* (2005)

included also some of these breeds. The current study was motivated by the deficit in molecular genetics information to define genetic relationships between local Swedish sheep breeds. The study aims at using 6 enJSRV polymorphic markers to study genetic relationships between 5 local Swedish breeds including Swedish fine wool sheep, Gute sheep, Roslag sheep, Värmland sheep and Klövsjö sheep.

Materials and Methods

Sample description

Blood samples of 66 sheep, availed by SLU LifeSciLab Biobank at the Department of Animal Breeding and Genetics Swedish University of agricultural sciences, were used, including 8 samples from Swedish fine wool sheep, 12 from Gute sheep, 12 from Roslag sheep, 9 from Värmland sheep, 8 from Klövsjö sheep, 12 from Texel and 1 sample from a Mouflon. For some breeds, samples were from different flocks while for others it was only possible to have samples from one farm (one flock) as indicated in the Appendix 1. Although Texel is not a local Swedish breed, samples were included in the study for a better comparison since there are studies indicating that Texel had been crossed with local Swedish breeds in the past to improve their productivity (Dýrmundsson and Niżnikowski 2010).

DNA extraction and dilution

Genomic DNA was extracted from each of the blood samples using a QIAsymphony® DSP DNA extraction robot and the QIAsymphony DSP DNA Midi Kit (QIAGEN). Concentration of extracted DNA was measured using a NanoDrop 8000 spectrometer and DNA was then diluted to a working concentration of 4ng/µl.

Polymerase chain reaction (PCR)

The genomic DNA of each sample was investigated for the enJSRV loci using both a 5' LTR and 3' LTR PCR for each locus as explained further. A HotStarTaq Plus DNA polymerase kit (QIAGEN) was used to set the PCR reactions. The reaction volume was 20µl composed of 20ng genomic DNA, 0.4Mm dNTPs, 2.8mM $MgCl_2$, and 0.12U HotStarTaq Plus polymerase, 0.4µM of each forward and reverse primer, 2X PCR buffer and 2X Q solution. The thermal cycler was programed for a touchdown PCR as indicated in Table 2 below.

Table 2: Touch down PCR thermo-cycler temperature setting

Cycles	Temperature (^oC)	Time	Rate of decrease
Denature	95	10 min	
Touchdown	95	15 sec	
8 cycles	61	30 sec	Decrease 0.5^oC/cycle
	72	30 sec	
30 cycles	95	15 sec	
	57	30 sec	
	72	30 sec	

(Bowles *et al.* 2014)

We used primer oligonucleotide pairs that were earlier designed by Chessa *et al* (2009), in which the 5' LTR forward (5'FlankF) primers were designed complementary to the specific DNA region flanking the 5' LTR of a specific enJSRV, while the reverse primer (ProvF) was designed to be complementary to the untranslated gag region. 3' LTR PCRs had a forward primer (ProvF) designed to be complementary to the Env gene and the reverse primer FlankR was designed complementary to the specific region flanking the 3'LTR of a specific enJSRV as shown in Figure 6. This meant that all enJSRV amplified loci shared ProvF and ProvR primers as they were located in highly conserved regions within a class of retroviruses. As a control for checking genomic DNA integrity, enJSRV-6 was used to test the quality of genomic DNA before running the PCR for polymorphic loci since it is fixed in the sheep species, hence avoiding false negative results (Chessa *et al.* 2009).

Provirus names	*5'FlankF / ProvR*	*ProvF / 3'FlankR*
enJSRV-6	5'-ccagttccagaagggaagggag/5'-agccctacaacctgggttgccca	5'-agccttcattcactgtggcga/5'-caggggaataactggtgctacct
enJSRV-7	5'-tgtgcacacgtggtgggagtc/5'-gtagtggcgaggaaaactgtcgagc	5'-agccttcattcactgtggcga/5'-aggaactccagggtcgtgcca
enJSRV-8	5'-tcagtggatcaatggtctgcga/5'-gtagtggcgaggaaaactgtcgagc	5'-agccttcattcactgtggcga/5'-gttatgggatttggggaaaagc
enJSRV-15	5'-cttctagtacagaatatagtggg/5'-gtagtggcgaggaaaactgtcgagc	5'-agccttcattcactgtggcga/5'-tcactgttgcttcctgaccagg
enJSRV-16	5'-tgctcagtttcccaggtgcccca/5'-gtagtggcgaggaaaactgtcgagc	5'-agccttcattcactgtggcga/5'-ggccagttgacatctgccagga
enJSRV-18	5'-gggaagattcgtttcttaggcgctc/5'-gtagtggcgaggaaaactgtcgagc	5'-agccttcattcactgtggcga/5'-caagtgccagagcccagagcca
enJS-5F16 (8	5'-ggataagctacactataaaaccaaag/5'-gtagtggcgaggaaaactgtcgagc	5'-agccttcattcactgtggcga/5'-ccatatgtagggattgggggtg

Figure 6: 5' and 3'PCRs performed in the study and primer oligonucleotides (forward and reverse) (Chessa *et al.* 2009)

Electrophoresis of the PCR products

To visualize the presence or absence of the enJSRV loci in the sample, PCR products from each reaction were subjected to Agarose gel electrophoresis. Agarose gels of 2% concentration were prepared using 100ml of 1X TBE buffer solution and stained with 3.5ul SYBR green DNA stain. Each well was loaded with 10ul composed of 9ul of a specific PCR mixed with 1ul of X5 Biorad loading buffer. One of the wells was loaded with 10ul of solution, which included 2ul of 200ug/ml 100bp DNA ladder, 1ul loading buffer and 7ul 1X TBE. To separate different amplified genomic bands the gel was ran on 120 Volts for 90 minutes. Bands were viewed under UV light, a picture was taken using the *AlphaDigiDoc RT* computer program. The expected band size of the amplified enJSRVs was estimated by comparing with the bands of the 100bp DNA ladder as seen in Figure 7.

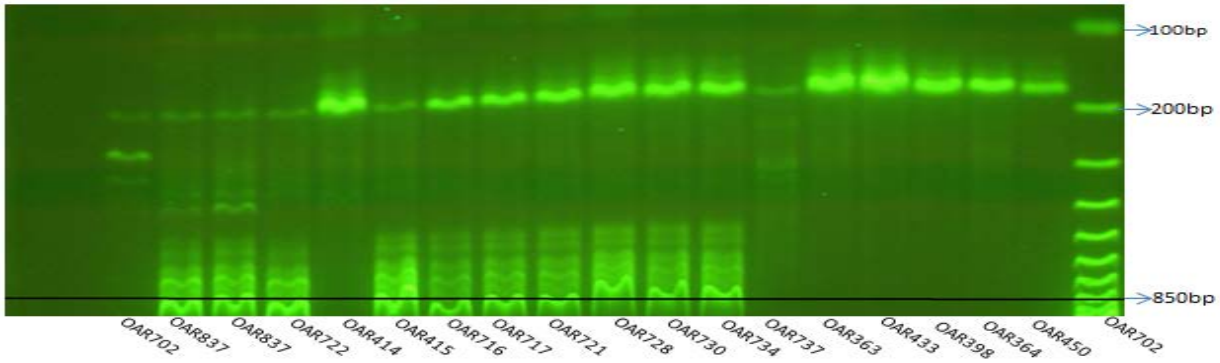

Figure 7: Individual genotyping from electrophoresis gel for enJS5F16, for 18 samples and the expected band size of the amplified locus was 856bp, 10 of these these samples carried the enJS5F16 in their genomes.

Data Analysis

Genetic relatedness between sheep breeds was analyzed based on the distribution of the six enJSRV loci and 15 retrotypes (R0 to R14) in each breed samples. Retrotypes used in the study were defined in a similar way to those used in the studies by Chessa *et al*. (2009) and Bowles *et al*. (2014). Retrotypes and their description are given in Table 3.

Table 3: Combination of retroviruses to categorize retrotypes used for the analysis in the study

Retrotype	Description
R0	no insertionally polymorphic enJSRVs
R1	enJSRV-7 only
R2	enJSRV-18 only
R3	enJS5F16 only
R4	enJSRV-7 + enJSRV-18
R5	enJSRV-7 + enJS5F16
R6	enJSRV-18 + enJS5F16
R7	enJSRV-7 + enJSRV-18+ enJS5F16
R8	enJSRV-8
R9	enJS5F16 + enJSRV-8
R10	enJSRV-7 + enJS5F16 + enJSRV-8
R11	enJSRV-18 + enJSRV-8
R12	enJSRV-18 + enJS5F16 + enJSRV-8
R13	enJSRV-7 + enJSRV-18 + enJSRV-8
R14	enJSRV-7 + enJSRV-18 +enJS5F16+ enJSRV-8

Samples from Texel breed were included in the analysis as an out-group to which breeds in the study could be compared. This is because there have been records of Texel being crossed with the local sheep breeds. PCR for all loci was done for the Mouflon as well but it was excluded from the analysis as it was only one individual.

Results

Provirus frequency analysis

When local breeds were grouped as one population, enJSRV-7 was the most frequent provirus (45%) while enJSRV-16 and enJSRV-15 were the least frequent proviruses in local sheep with frequency of 0 and 4% respectively in the local sheep population. The loci enJSRV- 8, enJSRV-18 and enJS5F16 were moderately present in the samples at frequency of 20, 25 and 10% respectively. However, the most common locus in the Texel breed was enJSRV -18(100%), followed by enJSRV-7 (92%), then enJS5F16 and enJSRV-8 at a frequency of 83 and 67%, respectively. enJSRV-16 had a moderate frequency (25%) and none of the Texel individuals had enJSRV-15 in their genome (Figure 7)

Looking at each local breed independently (Figure 8), it can be seen that Gute sheep, fine wool sheep and Roslag sheep have somewhat a similar distribution pattern of the proviruses; enJSRVs-7 is the most frequent locus in these three breeds (50%, 75% and 75%, respectively). Also enJSRV-18 is moderately distributed with in these three breeds. enJS5F1 is however moderately distributed in the Fine wool breed (42%), low in the Roslag sheep (8%) and none of the Gute sheep had this locus. Also a general overview is that there is a higher presence of enJSRV proviruses in Gute sheep, Fine wool sheep and Roslag sheep than in both Värmland and Klövsjö sheep.

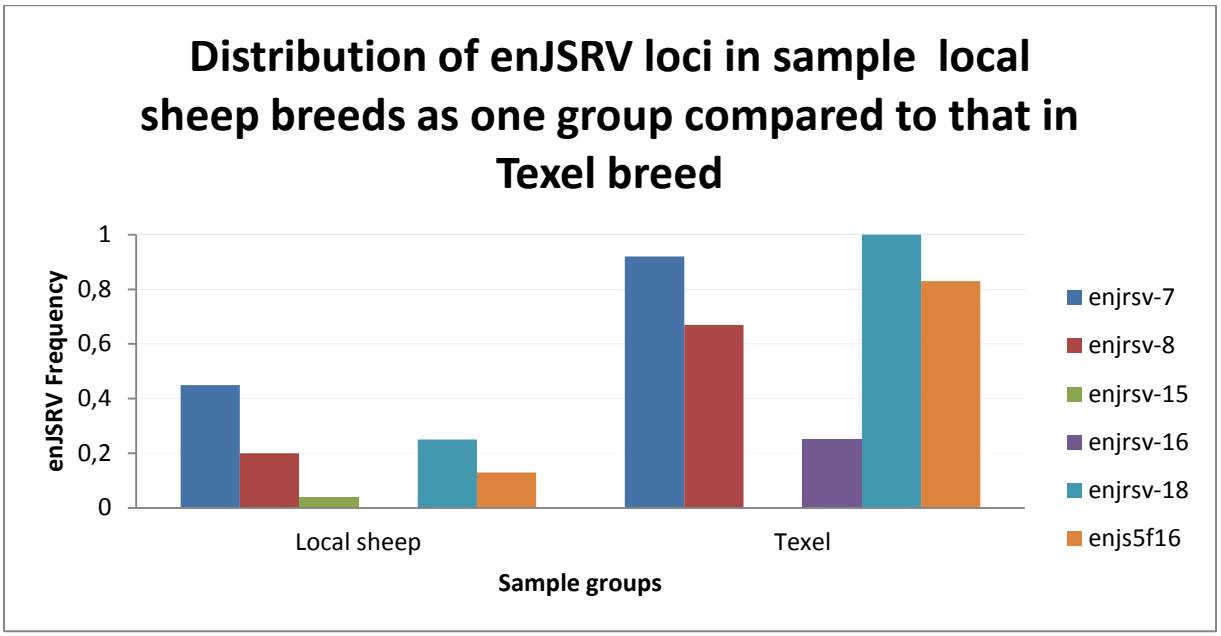

Figure 8: Distribution of the different enJSRV loci investigated in this study in local sheep breeds (collective) in comparison to sheep of the Texel breed

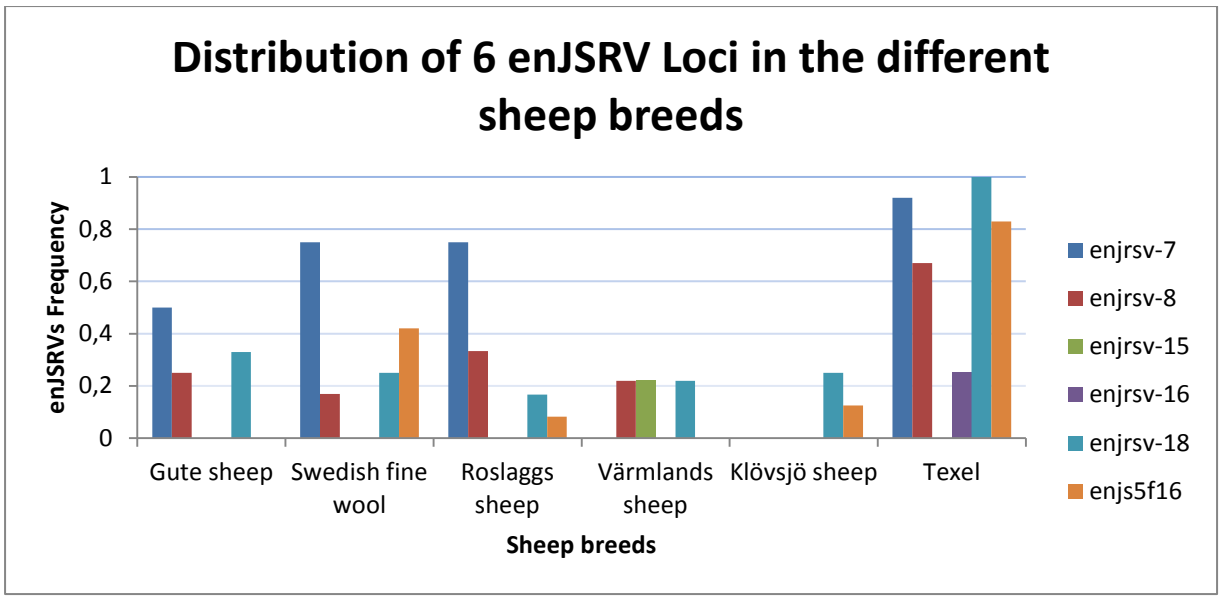

Figure 9: Distribution of the different enJSRV loci in the local sheep breeds in comparison to sheep of the Texel breed

The presence of enJSRVs was low in Värmland and Klövsjö sheep, as shown in Figure 8. The Only enJSRV-18 was present in both breeds, while enJSRV-8 and enJSRV-15 were only present in Värmland and enJS5F16 was only present in Klövsjö sheep. From all individuals genotyped only three individuals had enJSRV-15 insertion and these included 2 Värmland and 1 Mouflon. There were more provirus insertions present in the Texel breed compared to any other breed in the study. Texel also possessed enJSRV-16 at 25% which was absent in all other breeds studied including the Mouflon individual.

Retrotype analysis

Gute, Fine wool and Roslag sheep showed also a great similarity in terms of the retrotype distribution. In all these breeds R1 was the most frequent retrotype at 33% in Gute sheep, 25% fine wool sheep and 50% in the Roslag sheep. There was a moderately lower frequency of R0 within these breeds (17% to 8%). R4 was present in Gute and Fine wool sheep only, though at low frequencies (17% and 8%, respectively), while R11 was only found in Fine wool and Roslag sheep (8% and 17%, respectively). Although, R0 was present in all the five local sheep breeds, it was exceptionally high in Klövsjö and Värmland breeds at 63% and 44%, respectively. These two breed therefore seemed more similar to each other than to any other breed in the study. R8 was only present in Gute, Roslag and Värmland sheep. A new retrotype (enJSRV-7 and enJSRV-8 only) not defined in the retrotype table was found present in on Roslag individual (OAR550). Sheep from the Texel breed had a seemingly different retrotype distribution from any other breed in the study. R14 was the most frequent with 50% of Texel individuals having these loci, followed by R7 at 25%, R12 at 17% and R12 was the least frequent (8%) retrotype. Texel shared also two loci (R14 and R13) with only one of the local Swedish breeds, the Fine wool breed, in the study. As shown in Figure 11, R0 was present in all the indigenous sheep breeds but none of the Texel individuals had this retrotype.

Principle component analysis on provirus frequencies

To visualize the genetic distances between the breeds used in the study we carried out a principal component analysis (PCA) on the frequencies of presence/absence of proviruses in the tested breeds using the R program. From the analysis, the first and second principal components explained 53% and 33% of the variance, respectively. In a plot with the first two Principal Components it can be seen that Värmland sheep and Klövsjö sheep are closest together and furthest away from Texel among the local breeds, and that Swedish Finewool sheep is closest to Texel. Furthermore, Gute, Roslag and Fine wool were relatively close to each other.

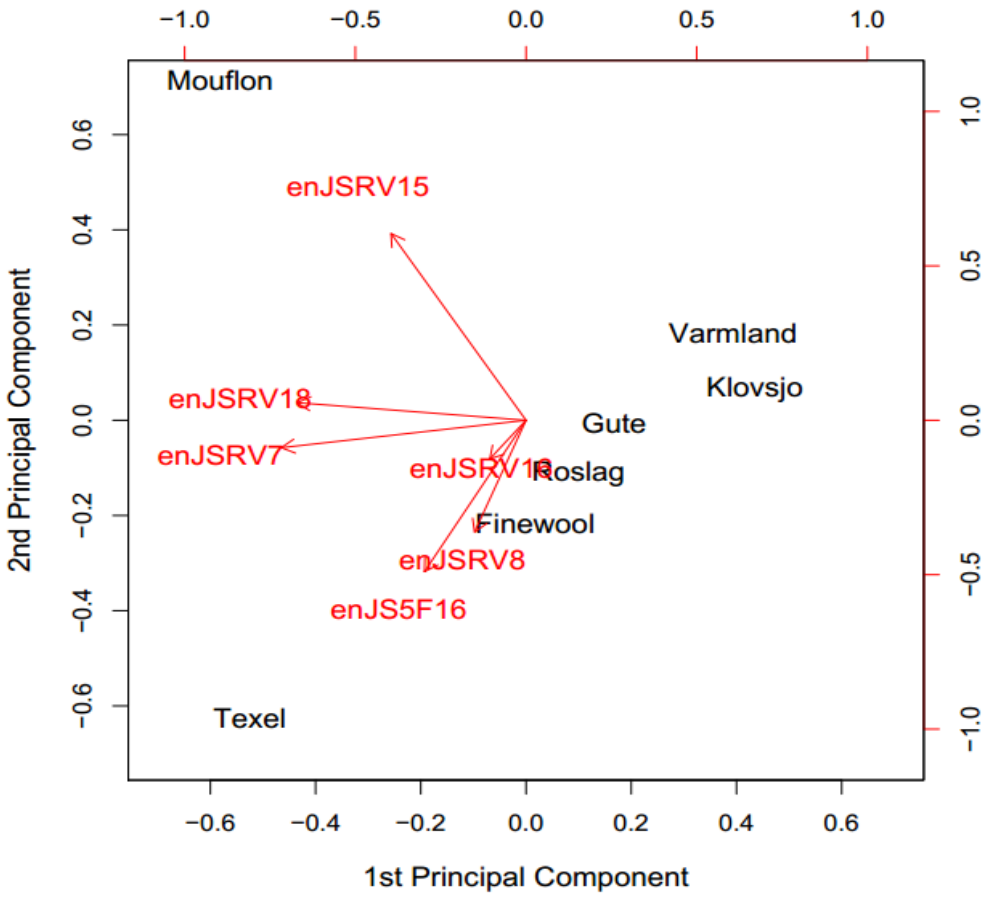

Figure 10: Principal component analysis (PCA) for the distribution of enJSRVs in the breeds investigated in this study, indicated how Mouflon sample was different from other breeds studied and also Texel, is distant from the local Swedish breeds.

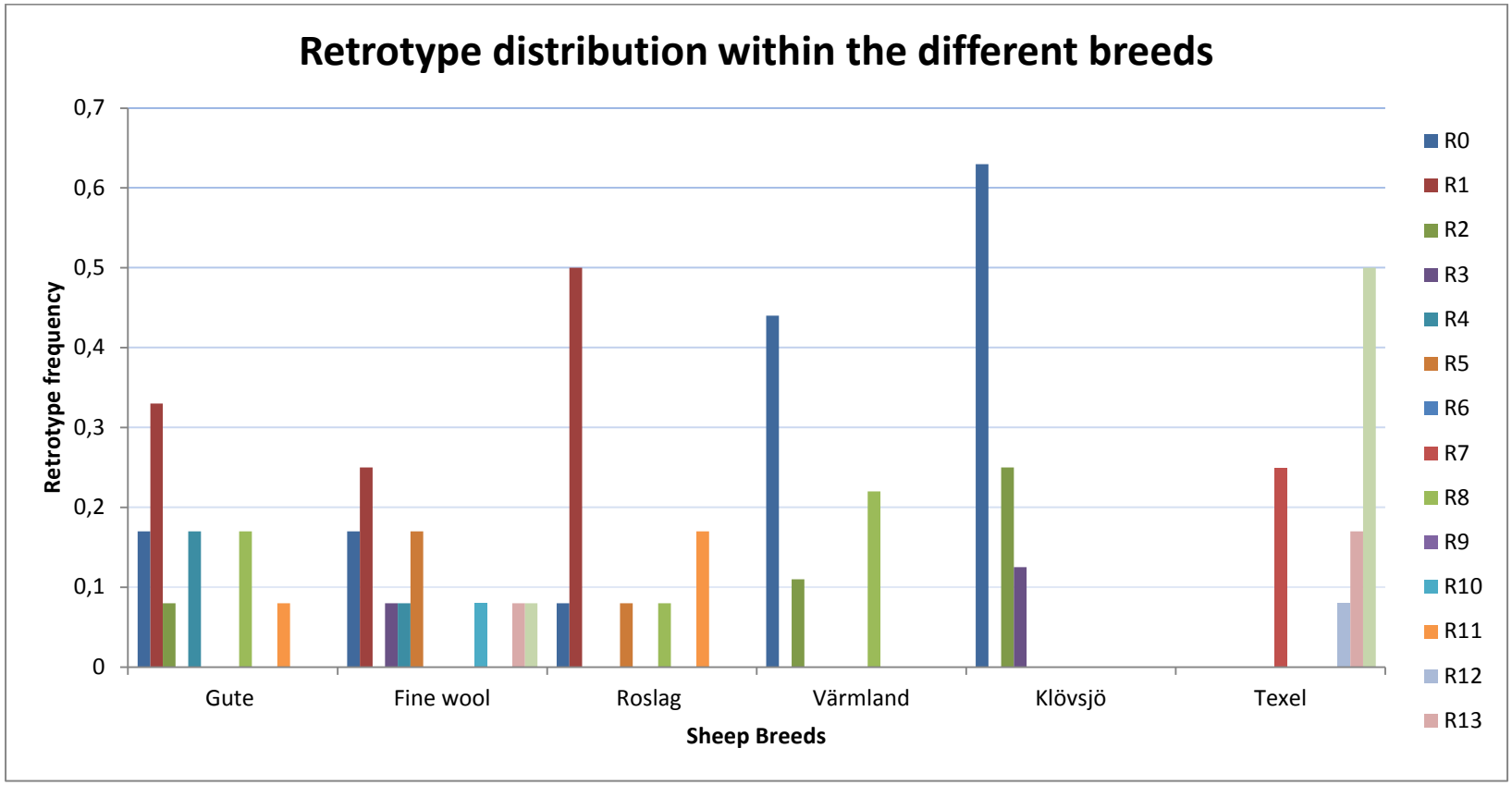

Figure 11: Retrotype frequency distribution (R0 to R14) within Gute, Fine wool, Roslag, Värmland, Klövsjö and Texel sheep breeds

Discussion

According to the obtained results three breeds including Gute, Fine wool and Roslag sheep had a somewhat similar distribution pattern of enJSRVs, which suggests a genetic closeness between these breeds. They also had enJSRV-7 locus as the most frequent locus coupled with moderate frequencies of the enJSRV-18. The existence of enJSRV-7 as the most frequent locus in these breeds suggests that they probably possess ancestral genomes, however, moderate frequencies of enJSRV-18 indicates the presence of modern sheep blood within the genomes of these breeds (Chessa *et al.* 2009, Bowles *et al.* 2014). These breeds had in general a high number of these polymorphic insertions, Fine wool and Roslag had four loci (enJSRV-7, enJSRV-8, enJSRV-18 and enJSF15) while the Gute breed had three loci (enJSRV-7, enJSRV-8 and enJSRV-18). The presence of many provirus loci within the breed has been previously considered as a characteristic of modern sheep breeds like Texel (Arnaud *et al.* 2007, Chessa *et al.* 2009, Bowles *et al.* 2014), this is also supported by results from our study for Texel sheep. This is more proof that Gute, Fine wool and Roslag are not fully primitive breeds but probably they have been crossed with modern breeds hence accounting for the presence of the various enJSRV loci in these breeds. This conclusion agrees with Dýrmundsson and Niżnikowski (2010) who indicated that there have been low and considerable crossing of Gute and Fine wool sheep breeds with Texel respectively. Frequency and distribution of enJSRVs in Värmland and Klövsjö breeds are low, indicating that most of the individuals in these breeds had none of these polymorphic insertions. This according to (Chessa *et al.* 2009, Bowles *et al.* 2014) is a characteristic of primitive sheep genomes.

Results from the retrotype analysis revealed a relatively high frequency of R0 (defined as absence of the polymorphic provirus) in the indigenous breeds except for the Roslag sheep. This to a great extent in agreement with results from a global study by Chessa *et al* (2009) where Nordic breeds including Gute sheep breed were characterized by high frequency of R0. High frequency of R0 or R1 was also a characteristic of all primitive breeds tested by Chessa *et al (*2009*)*. It can be seen from our results that R0 was the most frequent retrotype in both Värmland and Klövsjö breeds indicating close genomic similarity of these breeds and also stronger ancestral genomic composition of these two breeds (Chessa *et al.* 2009) compared to the other tested indigenous breeds. This genetic closeness between Värmland and Klövsjö breeds has also been reported by Dahberg (2012) through pairwise fixation index (F_{ST}) value comparisons in her microsatellite study of genetic variation between local Swedish sheep breeds. A value of 0.21 was obtained from F_{ST} value comparison for these two breeds, which indicated a certain level of closer genetic relationship (Dahlberg 2012).

The provirus combination R1 (enJSRV-7 only) was only present in three breeds, the Gute sheep, Fine wool sheep and Roslag sheep, and it was the predominant retrotype within each of these breeds. The presence of R1 in the three breeds only, further indicated a closer genetic relationship between these breeds. According to Chessa *et al.* (2009) breeds having high frequency of R1 retrotype were relatively primitive due to predominant presence of R1. We obtained interesting results when comparing retrotype distribution between Texel and each of the local Swedish breed tested. Texel shared two retrotypes (R13 and R14) with only one indigenous breed, i.e. Swedish Fine wool sheep

breed. This adds more evidence to the already recorded history of considerable between Texel and Fine wool sheep for improvement (Dýrmundsson and Niżnikowski 2010).

The principle component analysis of the enJSRV frequencies in the different breeds showed additionally that Värmland and Klövsjö were very close to each other as also observed from the retrotype analysis. Fine wool, Gute sheep and Roslag were close to each other further suggesting a stronger genetic relationship between these three breeds compared to other breeds in the study. Even with the principle component analysis Texel was distinctively distant from any of the breeds in the study. This was expected since Texel is a modern breed developed in the 19th century (Arnaud *et al.* 2007, Chessa *et al.* 2009, Bowles *et al.* 2014) that should show significant genetic difference from primitive breeds (including Swedish local sheep breeds). This was also observed in a larger study by Chessa et al. (2009), where Texel was shown to be genetically distant from the primitive sheep breeds that included Scandinavian breeds (including the Gute sheep) basing on their principle analysis results (Chessa *et al.* 2009). Fine wool sheep was the closet breed to Texel, which reaffirms the retrotype analysis that showed a closer genetic relationship between this breed and Texel due to already recorded history of considerable crossing of the two breeds (Dýrmundsson and Niżnikowski (2010).

Conclusions

In conclusion our study reaffirms the usefulness of using polymorphic endogenous retroviruses (in this case enJSRVs) as molecular markers to study genetic distinctiveness between populations of a given species. Furthermore, our study reveals that Gute sheep, Roslag sheep and Swedish fine wool sheep breed are closely related and they are genetically less related to Värmland and Klövsjö. There is also a close genetic relationship between Värmland and Klövsjö sheep breeds. The study further provides objective genetic evidence for the recorded history of considerable crossing between Texel and Fine wool breed.

Looking at the number of samples used and their distribution between flocks, we acknowledge the fact that the sample sizes were small, and collected from a few farms (flocks). Therefore our results could not conclusively describe the genetic differences among the studied breeds, this calls for larger studies using greater sample size; from more flocks and more local sheep breeds that we did not consider. Such extensive studies will give a larger and conclusive picture on the genetic distinctiveness between the different Swedish sheep breeds.

Acknowledgements

I would like to thank my supervisors Prof. Dr. Göran Andersson and Dr. Anna Johansson for their constant guidance and supervision during the course of the project. I would further wish to express my gratitude to all members of the Molecular genetics laboratory (SLU) including Dr. Tomas Bergström, Dr. Sofia Mikko, Dr. Kumar Juneja, Alexander Falk and Charlotte Grundel for their time and technical support during my practical work in the laboratory.

Furthermore, on the behalf of my supervisors and me we would like to show our greatest appreciation to Prof. Dr. Bernardo Chessa and Prof. Dr. Massimo Palmarini for their assistance with their knowledge on practical work related to enJSRVs. They provided useful information on PCR technicality that was very useful for the success of our project.

I cannot forget to thank Agnese Viluma, Merina Shrestha and Christina Rochus for their willingness to share their laboratory practical experience whenever I called on them for help. I am also grateful for the intern student Constance Humblot for practical help during laboratory work in particular genotypic PCRs.

It would have been impossible for me to join the European Masters in Animal Breeding and genetics if it were not for the financial provision by KOEPON foundation. For that reason I am grateful for this opportunity whose impacts on my academic and career value are immeasurable.

Last but not least I would love to thank my family and friends for continuous academic, emotional and spiritual support they have offered to me throughout my Masters Study period.

References

Armezzani, A., *et al.* (2011). "The Signal peptide of a recently integrated endogenous sheep betaretrovirus envelope plays a major role in eluding Gag-mediated late restriction." Journal of virology **85**(14): 7118-7128.

Arnaud, F., *et al.* (2007). "A paradigm for virus–host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses." PLoS pathogens **3**(11): e170.

Arnaud, F., *et al.* (2008). "Coevolution of endogenous betaretroviruses of sheep and their host." Cellular and molecular life sciences **65**(21): 3422-3432.

Bowles, D., *et al.* (2014). "Genetic Distinctiveness of the Herdwick Sheep Breed and Two Other Locally Adapted Hill Breeds of the UK." PLOS ONE **9**(1): e87823.

Bruford, M. W. and S. J. Townsend (2006). "Mitochondrial DNA diversity in modern sheep." Documenting domestication: New genetic and archaeological paradigms: 306-316.

Chen, S.-Y., *et al.* (2006). "Origin, genetic diversity, and population structure of Chinese domestic sheep." Gene **376**(2): 216-223.

Chessa, B., *et al.* (2009). "Revealing the history of sheep domestication using retrovirus integrations." Science **324**(5926): 532-536.

Dahlberg, L. (2012). "Genetic variation in local Swedish sheep breeds."

Dýrmundsson, Ó. R. and R. Niżnikowski (2010). "North European short-tailed breeds of sheep: a review." animal **4**(08): 1275-1282.

FG-homepage. "Föreningen Gutefåret (FG): The association is responsible for maintaining pure breed gutefår as an important heritage of the Swedish animal husbandry." from http://www.gutefar.se/. (accessed: 10/6/2014)

FSA-homepage. from http://www.allmogefar.se/. (accessed: 10/6/2014)

Grigaliunaite, I., *et al.* (2004). "Characterization of genetic diversity in domestic sheep." Consulta en línea: http://www. agronet. fi/maataloustieteellinenseura/julkaisut/poste/kj03grida. pdf Febrero de.

Hiendleder, S., *et al.* (1998). "Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep." Journal of Heredity **89**(2): 113-120.

Jern, P. and J. M. Coffin (2008). "Effects of retroviruses on host genome function." Annual review of genetics **42**: 709-732.

Maksakova, I. A., *et al.* (2008). "Endogenous retroviruses - Keeping active endogenous retroviral-like elements in check: The epigenetic perspective." Cellular and molecular life sciences **65**(21): 3329-3347.

NorthSheD-homepage (13-5-2014). "Coming of sheep and thier breeding to Nordic countries." from http://www.rala.is/beta/History.htm. (accessed: 10/6/2014)

Palmarini, M., *et al.* (2004). "Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation." Journal of general virology **85**(1): 1-13.

QIAGEN. from http://www.qiagen.com/. (accessed: 10/6/2014)

Sistiaga-Poveda, M. and B. M. Jugo (2014). "Evolutionary dynamics of endogenous Jaagsiekte sheep retroviruses proliferation in the domestic sheep, mouflon and Pyrenean chamois." Heredity.

Tapio, M., *et al.* (2010). "Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia." BMC genetics **11**(1): 76.

Appendix 1

sample distribution in different flocks

Swedish Fine wool sheep

Flock	No. of samples
SE3017	4
SE13202	4
SE083438	4
Sub total	12

Gute sheep

Flock	No. of samples
SE26743	1
SE72435	1
SE68899	1
SE079855	1
SE005514	1
SE050451	2
SE76771	2
SE039480	1
SE71937	1
SE083438	1
Sub total	12

Mouflon

Flock	No. of samples
	1
Sub total	1

Klövsjö

Flock	No. of samples
SE069781	8
sub total	8

Roslag

Flock	No. of samples
SE041133	12
Sub total	12

Värmland

Flock	No. of samples
SE47190	6
SE078470	1
SE063891	2
Sub total	9

Texel

Flock	No. of samples
SE38717	4

403900	8
Sub total	12
Total	**66 samples**