University of Natural Resources and Life Sciences, Vienna
Department of Sustainable Agricultural Systems
Division of Livestock Sciences
Working group Animal Breeding and Genetics

Assessing taurine introgression in the current Brazilian Nelore cattle population

Master Thesis

Daniela Höller, BSc

Supervisor:
Univ.-Prof. Dipl.-Ing. Dr. rer. nat. Johann Sölkner

Co-Supervisor:
Ana María Pérez O'Brien, MSc

Vienna
December 2013

Statutory declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources or resources and that I have explicitly marked all material, which has been quoted either literally or by content from the used sources.

Vienna, 12th December 2013

Date

Signature (Daniela Höller, BSc)

Abstract

The history of Nelore cattle in Brazil is that imported Indian animals were interbred with available taurine Creole populations to spread the zebuine germplasm, creating an admixture event by crossbreeding two formerly distant populations. With the use of currently available high-density genotypic information it is possible to investigate the genetic origins of livestock populations. The main aim of this thesis was to assess the taurine introgression in the current Brazilian Nelore cattle population by comparison to a reference group consisting of taurine (Angus, Fleckvieh, Hereford, Holstein, Limousin, Piedmontese) and indicine (ancestral Nelore, Brahman, Gir) cattle breeds. This was performed by estimating individual admixture levels, constructing haplotypes and calculating their frequencies, and building phylogenetic trees to show the relationships of the breeds graphically. The taurine introgression found in the young Nelore animals is lower than initially expected, with an average genome-wide taurine admixture level of just 0.4 % and a high proportion of the indicine admixture seemingly arising from the ancestral Nelore. Analysis of the Y-chromosome shows similar results to the genome-wide, with an average taurine introgression of 0.7 %. Moreover, differences between young Nelore subgroups were found with animals from pedigree type herds showing higher ancestral Nelore and lower taurine origin than the production type, and this result was additionally confirmed by the phylogenetic trees. In contrast, mitochondrial data shows an average of 95.1 % of taurine maternal ancestry in the young Nelore population. These results support the initial hypothesis of the historical use of taurine females, yet the analyses show that very little taurine ancestry has remained in the genome of the current population with variations at the individual and group level.

Table of contents

Acknowledgements

I would like to express my gratitude essentially to my master thesis supervisor Prof. Johann Sölkner, who sparked my interest deeply in this topic. He supported me in all levels of this research work. Special thanks are also given to my co-supervisor Ana María Pérez O'Brien. Without her scientific and technical assistance in all the times of operating the analyses it would have never been possible for me to finish my thesis. Thanks also for being such a good friend and the mental support all the time.

I also would like to acknowledge José Fernando Garcia from the ZGC (Zebu Genomic Consortium) from the “Universidade Estadual Paulista” in Brazil, as well as Tad Sonstegard and Curt Van Tassell from the Bovine HapMap Project from the USDA (United States Department of Agriculture), who provided all the genotypes of the animals.

Last but not least important are my friends Kerstin and Sylvia, but I owe more than thanks to my family. With the encouragement from my parents and my older brother all the time throughout my whole study I got prospects to graduate. Gratefully thanks to Christian for being my tower of strength and giving me all the time a start and the pressure I needed for finishing my thesis.

1 Introduction

The introduction of this thesis will give an overview and will show the importance of the topic "admixture" in general. Moreover the objectives will be stated. This will respond to the structure of the work and show the chosen mode of operation, to build up towards the interpretation of the results and the response to the research questions, which can be found in the conclusions on chapter 6.

1.1 Importance and topicality

Estimation of admixture levels was, like in many other different fields of research, initially used exclusively in human genetics. Today, because of low cost availability of genetic marker data and the fast progress in genotyping techniques, admixture assessment is also common practice in animal, plant and other species genetics. Few years ago genetic analyses were mostly performed using microsatellite markers, which even though useful at the time, are known to have limitations on the genome coverage and are a less frequent variation than single nucleotide polymorphism (SNP) markers. Since genotyping is continuously getting cheaper, it is also possible now to genotype the whole bovine genome with a high SNP density of 777 thousand, and the technology is extending as well to other livestock species such as chicken, pigs, goats and sheep.

Admixture mapping studies are often performed to gain understanding in the migration patterns of different populations, for finding genomic regions and genes derived from specific populations in mixed origin individuals, which can be associated to a disease status (BASU et al., 2008), and in livestock species to increase our understanding of domestication events. In livestock the migration and mixing of different populations is commonly known as "crossbreeding". Trustable and good pedigree information has been satisfactorily used for the assessment of crossbreeding levels, but the use of SNP chip data makes the estimations more accurate and precise taking into account the stochastic nature of recombination (FRKONJA et al., 2012). Additionally, it is important to remember that a large amount of livestock populations lack good pedigree records, therefore making these methodologies more and more important for estimating breeding values and designing breeding programs in crossbred populations.

Many methods and programs have been developed for these types of analyses. In the study "Prediction of breed composition in an admixed cattle population" FRKONJA et al. (2012) implemented different freely available genetic admixture software and algorithms and performed an extensive comparison of the results. Based on their results, the decision for the software used in these analyses was reached.

This study will provide an insight into the formation and the evolution of the Brazilian Bos primigenius indicus (indicine) cattle breed Nelore. Genome-wide autosomal admixture studies as well as focus studies on maternal and paternal inheritance, using mitochondrial DNA (MT) and Y-chromosome (Y) SNPs, have been performed to elucidate the genetic descent of Nelore, a zebuine beef breed of high importance in Brazil.

1.2 Objectives

The objectives of this master thesis are structured into the main and the specific ones. The main target is the investigation of the taurine introgression in the current Brazilian Nelore cattle population through the comparison to different reference populations. Moreover a separation of the current population – which are later referred as young Nelore – into pedigree and production purpose animals will be operated. For a more exact definition of this main objective the specific objectives are defined and explained as follows:

1) Evaluation of the admixture levels of the autosomal, Y-chromosomal and mitochondrial genome. The estimation of the admixture levels of all three data sets will give an idea of the origin and the ancestry of the individuals.
2) Assessment of the connection and the grade of relationship via the genetic distances of the breeds illustrated graphically through the construction of phylogenetic or neighbour joining trees.
3) Finally, building haplotypes (haps), based on Y-chromosome and mitochondrion, will en quire if there are any breed specific haplotypes adopted from paternal and maternal inheritance.

1.3 Structure of the thesis

This manuscript is arranged in six chapters. The introductory chapter is giving a preface to the general subject and guidelines to the transcript itself (chapter 1), followed by the background and the literature review (chapter 2), which shows the connection of the topic to the literature, the history and importance of admixture studies and describes the Nelore breed and its relevance for the Brazilian context. The main component of the analyses consists of the materials and methods (chapter 3), were all used methods, software and data are delineated at first, and the results obtained (chapter 4), where all relevant graphs and plots are exhibited and described. In the end, a thorough discussion of the results (chapter 5) and the conclusions (chapter 6) are stated.

2 Literature review and background

2.1 Genetic admixture

To clarify the term "genetic admixture" a general definition will be given: Admixture occurs when two previously isolated populations begin interbreeding (BALDING et al., 2007). In the case of the target population in this study, Nelore, this concept would denote that the Ongole, the most influential Indian population which gave origin to the Nelore, was very likely interbred with taurine breeds and the Creole cattle available at the moment, highly derived from Portuguese and Spanish breeds, due to the limited number of "pure" zebuine females available (AJMONE-MARSAN et al., 2010). Out of these breed crossings the current Nelore population, composed of more than 100 million individuals (ACNB, 2006), came to be. Every single individual is then likely to have different levels of admixture depending on the intercrossings in its ancestral history, and this can be evaluated with the use of different genetic admixture analyses models and tools.

As mentioned in the introduction, genetic admixture analysis was first developed in human genetics. It can be differentiated between regional/local and continental genetic admixture, which is of big interest, due to recent migration patterns (FRKONJA et al., 2012). Regional admixture refers to the formation of a new breed or population in a certain region, e.g. the Latinos in South America. PRICE et al. (2007) describe in their study four different Latino populations (Brazilians, Mexicans, Columbians and Latinos) and estimated their ancestry, which is coming from Europe and Africa for the most part (continental admixture), but also to a certain percentage from the regional ancestors native to North and South America.

Continental admixture arises when a new population develops comprehending a large continental area. In the study of NASSIR et al. (2009) the continental origins of humans all over the world were estimated. They used ancestry informative markers (AIMs) to identify the descent of subjects from Europe, Sub-Saharan Africa, the Americas, and East Asia. AIMs are genetic loci involving the alleles with large frequency differences between the populations (SHRIVER et al., 2003). For example in case-control studies they can be used as tools for minimization of bias derived from population stratification. In genome-wide association studies it is not necessary to use them, because the high quantity of SNPs is able to control these stratifications (NASSIR et al., 2009). Similar to that study, this thesis was designed under the same methodology: Estimating the ancestry of different cattle populations using supposed diverging continental origin breeds as a reference.

Admixture studies have become more common in animal genetics as a consequence of technological progress in genotyping platforms and therefore the recent availability of cheaper SNP genotyping techniques. Some years ago microsatellites were used as the main markers for genetic studies; nowadays SNP chip data are predominantly utilized. For cattle populations the most widely used are LD (low-density), 54k and HD (high-density) SNP chips which contain a few thousand (~ 6000), 54609 and 777962 SNPs respectively. Most studies in admixture have utilized 50k or HD chips, in this master thesis a high-density chip was used (for more details see chapter 3, materials and methods).

Herein, the genetic admixture level was estimated individual per individual based on genome-wide autosomal, Y-chromosomal and mitochondrial SNP data sets. MT and Y analyses were performed to retrace the maternal and paternal inheritance and taurine source introgression. We hypothesized a higher percentage of Indicine genome should be in the male Y-chromosome due to the wider use of imported "pure" zebuine males and larger number of progeny in bulls which explain the main reason. The mitochondrial genome was hypothesized to be predominantly taurine, as mitochondria are transmitted almost exclusively along the maternal line and the most of the first mating partners of Ongole bulls in Brazil were taurine Creole cows.

2.2 Reference to literature

Nowadays many studies concerning admixture in cattle exist and many more are published every year. For establishing the connection from this study to others, a few of the most relevant will be described in the following passages.

FLORI et al. (2012) analysis of the Senepol cattle is very similar to this thesis in reference to the estimations, analyses and used software. The aim of their work was to clarify the breed origin and its current genomic admixture, and to detect footprints of selection characterizing the breed origin of selected genomic regions. The results showed that the breed is mainly of European taurine ancestry with around 10.4 % of indicine descent and no significant ancestry of the West African N'Dama was found, contradicting the historical hypothesis of a high African taurine composition which was supposed to provide trypanotolerance in the breed. This information is not only important for historical interest, but also for the management and diffusion of a cattle breed into new environments. Another pertinent example is that of the Italian Piedmontese breed, which was thought to be admixed cattle out of the extinct Aurochs and Indo-Pakistani zebu, but admixture studies showed that this breed is just a mixture of European taurine breeds (FLORI et al., 2012).

There was a similar topic using dairy cattle in Kenya faced by GORBACH et al. (2010). Knowing the pedigree structure of a herd is very important to avoid inbreeding and the loss of species diversity, but this is a big challenge in developing countries, where very often pedigree information is not documented, or very incomplete. For this study they sampled animals stemming from small and large herds, insemination stations and presumably of three different breeds: Holstein, Guernsey and Jersey, all having different levels of pedigree records. The genomic inbreeding coefficients were estimated, maternity and paternity were checked and also the breed composition or the admixture level was appraised. The results of these analyses show that the recorded pedigree information was often wrong, incorrect and inaccurately recorded and it did not matter if the animals are coming from big or small farms. This insight is important: Conservation of the genetic resources from Kenyan cattle like adaption to harsher conditions is fundamental, because the current crossbreeding schemes with European breeds will lead to more losses of these special adaptive characteristics. These findings can help to antagonise the loss of species diversity, find better and more suitable crossbreeding strategies and control inbreeding levels in the population.

Analyses of the geographic distribution of taurine and indicine Y-chromosome haplotypes of Argentinian and Bolivian Creole cattle using microsatellites were performed by GIOVAMBATTISTA et al. (2000). They found out that there is an east/west and a north/south gradient of the zebuine Y-chromosome, which is paternally inherited: The highest frequencies were found in Brazilian cattle populations. In Uruguay and Argentina there were no indicine Y-chromosomal haplotypes detected, though in Bolivia, which is in the centre of the continent, the values were at intermediate level. This decline of the gradient could be explained by historical events like importations and environmental factors like temperate weather, which could explain a reduced interest for zebuine breeding.
Investigation of X- and Y-specific SNPs were performed as well in African cattle. Domesticated African cattle are of taurine ancestry, but as former studies show there are also indicine and crossbred types. Like in human genetics the Y- and X-chromosome markers were combined to simulate eventually (like in the study of the South American cattle) geographic events. Taurine and indicine subspecies could be separated. Likewise in Africa it was detected that there is a higher indicine introgression in the East than in the South part of the continent. Reasons for higher levels in the East there were hypothesized to be recent admixture and crossbreeding events (ANDERUNG et al., 2007).
Referring to the Nelore breed there was a study performed by MEIRELLES et al. (1999) analysing mtDNA (mitochondrial DNA) for tracing the maternal inheritance of the registered American Bos indicus breeds Nelore, Brahman and Gir, which include both the target and the indicine reference group used in this master thesis. These three breeds were subdivided into POI (purebred of imported origin) and PO (purebred origin). On average Nelore and Gir showed 58 % taurine mitochondrial ancestry (higher in PO than in POI), and Braham was completely of taurine descent. This descent traces back to the history of the Brahman breed, which originated from four Indian breeds, imported from available South American herds, with some infusion of taurine British-bred cattle (AUSTRALIAN BRAHMAN BREEDERS' ASSOCIATION LIMITED, s. a.).
Comparison of different types of software used in genetics to estimate admixture levels was the main aim of the study of FRKONJA et al. (2012). They assessed the performance of hidden Markov models like STRUCTURE v2.3 (PRITCHARD et al., 2010) for admixture estimation by using 50k SNP chip data from Swiss Fleckvieh animals, a cross of Simmental and Red Holstein Friesian. Four methods were applied in this study for estimating admixture levels: STRUCTURE, PLSR (partial least squares regression), BAYESB (a Bayesian approach) and LASSO (least absolute shrinkage and selection operator). The results show very high correlations (~ 0.97) of the estimations of these methods with pedigree admixture, with LASSO having a lower power (0.93). While pedigree admixture is not a good reference, SÖLKNER et al. (2010) showed for a crossbred population of Merino and Awassi sheep, where identity-by-descent (IBD) states could be traced, that STRUCTURE and PLSR performed equally.

2.3 Nelore cattle description and history

The Nelore breed belongs to the species Bos primigenius, subspecies indicine, and is used for beef production purposes. It is widely used in Southern America, especially in Brazil, which is the largest Nelore cattle breeder and exporter of their meat in the world with an estimated

current population of 100 million animals (ACNB, 2006). Brazilian Nelore has descended from the Indian Ongole breed (Bos indicus), and is named after the Nellore district of the Indian federal state Andhra Pradesh, where the first exporters gathered and shipped the animals to Brazil. Around 7000 individuals were imported in total (VOZZI et al., 2007) and were bred for increasing the population up to current numbers, very likely interbreeding with Creole cattle (Bos taurus). In 1868 there was the first acknowledgement of Nelore in Brazil and afterwards, in the year 1938, the herd book was created and the breed standards were defined (FLECHA, 1997; ACNB, 2006; DANI et al., 2008).

Like any typical zebuine breed the Nelore animals have a big hump on top of the shoulders. They have long legs and for indicine standards very short ears. Their skin is very dark, mostly black, and covered by a light (white or light grey) coat colour, which protects the skin from sunlight. They are well adapted to harsh conditions. Because of many adequate characteristics of the breed like hardiness, insect and heat resistance, metabolic and reproductive efficiency and maternal instinct, this breed is still expanding its population (FLECHA, 1997).

2.4 Reference population description

To estimate the levels of taurine ancestry in the young Nelore production and pedigree type animals, reference populations were built out of eight different indicine and taurine breeds. The taurine breeds include: Angus, a beef breed originated in Scotland and imported to the United States (AMERICAN ANGUS ASSOCIATION, 2013); Fleckvieh, which is a dual purpose breed originated in Austria from the Swiss Simmental (BMLFUW, 2011); Hereford, a beef breed from the County of Herefordshire in the United Kingdom (THE HEREFORD CATTLE SOCIETY, 2013); Holstein, the most popular and widely used dairy breed in the world, with the first breeding association in Germany (DEUTSCHER HOLSTEIN VERBAND E. V., 2013); Limousin, an old French beef breed (BRITISH LIMOUSIN CATTLE SOCIETY, 2010); and Piedmontese, a white coloured beef breed from the Italian region Piedmont (ANABORAPI, 2013). The indicine breeds used were composed of: Nelore ancestral animals imported to Brazil from India, and some of their direct descendants (FLECHA, 1997; ACNB, 2006; DANI et al., 2008); Brahman, which is a highly indicine composite breed, developed in the United States and widely used in the tropics (AUSTRALIAN BRAHMAN BREEDERS' ASSOCIATION LIMITED, s. a.; SANDERS, 1980); and Gir, another originally Indian breed, imported to Brazil and used in dairy production (SANDERS, 1980).

3 Materials and methods

3.1 SNP genotypes

The extracted DNA of all animals was genotyped with the Illumina BovineHD Genotyping BeadChip (ILLUMINA, 2012) with a total number of 777962 (777k) SNPs. The ZGC (Zebu Genomic Consortium) from Brazil and the Bovine HapMap Project from the USDA (United States Department of Agriculture) provided all the genotypes.

3.2 Animals

A total number of 706 animals were included in this study, divided into eleven groups with nine different breeds. The Nelore breed was split into the ancestral (ANL) and young Nelore (YNL). Moreover, to differ between the types, young Nelore were again separated into production (YNLpro) and pedigree (YNLped) animals according to the herd management and purpose. The production animals are derived from herds with breeding programs focused on commercial production of beef, and with an intense selection for productive characteristics such as muscling, weight gain and carcass traits. The pedigree type animals are coming from herds focused on type traits and breed specific characteristics such as body shape, leg length and amount of skin folds, among others (DANI et al., 2008). The reference group (all animals except YNL) consists of 171 B. p. taurus and 81 B. p. indicus individuals. In the following table the different groups and their subspecies are listed with the number of individuals.

Table 1: List of the breeds (abbreviations), subspecies and number of individuals per breed

Breed (breed abbreviation)	Subspecies	Number of individuals
Angus (ANG)	Taurine	30
Fleckvieh (FLV)	Taurine	30
Hereford (HFD)	Taurine	27
Holstein (HOL)	Taurine	30
Limousin (LMS)	Taurine	30
Piedmontese (PMT)	Taurine	24
Nelore (NEL)	Indicine	475
Young Nelore (YNL)	Indicine	454
Production type (YNLpro)	Indicine	306
Pedigree type (YNLped)	Indicine	148
Ancestral Nelore (ANL)	Indicine	21
Brahman (BRA)	Indicine	30
Gir (GIR)	Indicine	30

3.3 Quality control (QC)

Quality control was performed using the software PLINK v1.07 (PURCELL et al., 2007). The parameters were individually chosen for the autosomal (AS), the Y-chromosome (Y) and mito-

chondrial (MT) data. First, to get the different types of SNPs, the autosomal, Y and MT SNPs were extracted. Afterwards the quality control was arranged with the parameters “geno” and “mind”, which give the maximum rate of missingness per-SNP and per-individual. This quality control step was done breed per breed in order to keep the maximum number of SNPs for analysis.

For the autosomal SNPs a rate of 0.1 or 10 %, for Y and MT SNPs a rate of 0.2 or 20 % was selected for the maximum missingness rate and all SNPs or individuals with higher missingness rate were excluded. Additionally, just for AS-SNPs, the command “hwe 0.000001” was added. This command refers to a p-value higher or equal to 0.00001 on the exact test statistic for Hardy-Weinberg-equilibrium (HWE) described by WIGGINGTON et al. (2005). SNPs with strong deviations from HWE are observed in the case of inbreeding, population stratification or, more frequently, problems and errors in genotyping. For MT and Y it was set to 0.

The next step of the quality control was finding the common SNPs between the different breeds, which was done with the software R v3.0 (R DEVELOPMENT CORE TEAM, 2008). For the AS data there were 710608, for Y 1194 and for MT 314 common SNPs. A text file with all common SNPs was created and afterwards with the help of PLINK this SNPs were extracted breed per breed. Subsequently, all the files with the single breeds were merged using PLINK and afterwards the minor allele frequency analysis was done using the command “maf 0.001” for all types of SNPs. With this command all SNPs with a minor allele frequency rate lower than 0.1 % were excluded. The last QC step (MAF) was the only one operated with the merged file consisting of all breeds.

After frequency and genotyping pruning of the SNP data there were 704 animals with 706017 SNPs for the autosomal, 693 animals with 98 SNPs for the Y and 703 animals with 27 SNPs for the mitochondrial remaining for analyses.

For a better understanding of the quality control steps see tables 2 to 4, which give the remaining and lost numbers of SNPs and animals after every single QC step for the autosomal, Y and mitochondrial data sets.

Table 2: Autosomal data set quality control steps with remaining and lost SNPs and animals per breed

Breed	Extracted AS-SNPs	SNPs left (lost) "geno" + "hwe"	Indiv. left (lost) "mind"	Common SNPs	SNPs left (lost) "maf"
ANG	734176	732514 (2725)	30 (0)	710608	706017 (4591)
FLV		733030 (2209)	30 (0)		
HFD		733328 (1911)	27 (0)		
HOL		733507 (1732)	30 (0)		
LMS		733564 (1675)	30 (0)		
PMT		733160 (2079)	24 (0)		
YNLpro		719343 (15896)	304 (2)		
YNLped		723147 (12092)	148 (0)		
ANL		722555 (12684)	21 (0)		
BRA		728468 (6771)	30 (0)		
GIR		727723 (7516)	30 (0)		

Table 3: Y-chromosome data set quality control steps with remaining and lost SNPs and animals per breed

Breed	Extracted Y-SNPs	SNPs left (lost) "geno"	Indiv. left (lost) "mind"	Common SNPs	SNPs left (lost) "maf"
ANG	1222	1212 (10)	28 (2)	1194	98 (1096)
FLV		1203 (19)	30 (0)		
HFD		1222 (0)	27 (0)		
HOL		1214 (8)	28 (2)		
LMS		1222 (0)	30 (0)		
PMT		1222 (0)	24 (0)		
YNLpro		1222 (0)	306 (0)		
YNLped		1222 (0)	148 (0)		
ANL		1222 (0)	21 (0)		
BRA		1208 (14)	22 (8)		
GIR		1210 (12)	29 (1)		

Table 4: Mitochondrial data set quality control steps with remaining and lost SNPs and animals per breed

Breed	Extracted MT-SNPs	SNPs left (lost) "geno"	Indiv. left (lost) "mind"	Common SNPs	SNPs left (lost) "maf"
ANG	343	335 (8)	30 (0)	314	27 (287)
FLV		340 (3)	30 (0)		
HFD		335 (8)	27 (0)		
HOL		337 (6)	30 (0)		
LMS		341 (2)	30 (0)		
PMT		343 (0)	24 (0)		
YNLpro		334 (9)	305 (1)		
YNLped		329 (14)	147 (1)		
ANL		321 (22)	21 (0)		
BRA		340 (3)	30 (0)		
GIR		325 (18)	29 (1)		

3.4 Admixture analyses

With the remaining SNPs and individuals of all data sets (autosomal, Y and mitochondrial) admixture analyses were done with the help of the softwares ADMIXTURE v1.2 (ALEXANDER et al., 2009) and R. ADMIXTURE software was chosen first because of its efficiency, and second because of the high similarity of its results with STRUCTURE and other analytical software for this purpose. FRKONJA et al. (2012) compared various publicly available software, most of them based on the same estimation model, hidden Markov model (HMM) clustering algorithms, a stochastic statistical model which attempts to detect an infinite number of hidden states (BEAL et al., 2001), and found out a very high correlation among all of their results.

ADMIXTURE is a freely available software which estimates the ancestry of unrelated individuals with the help of SNP marker data sets. All three data set analyses were performed unsupervised (without population information) and supervised (with population information). Unsupervised estimation of admixture levels means the computation without giving population membership as additional information for the analyses. In contrast to this, the supervised estimation of admixture levels clearly states the reference populations which should be used to estimate the ancestry of a subset population.

For supervised analyses it was necessary to create new files, called pop files, giving the population information. That means that every single individual was assigned to one breed. The young Nelore cattle individuals (production and pedigree type) were given as unknown to estimate their ancestry. Moreover, also K, the number of assumed populations, should be given for the calculation. All unsupervised estimations were calculated with K=2 to K=10 (nine different breeds with a subdivision of Nelore into young and ancestral) and for all supervised with K=2 to K=9 (two subspecies and nine reference breeds respectively).

Running ADMIXTURE creates a Q and a P file. The P file gives the allele frequencies of the ancestral population, and Q gives the fractions of ancestry for each chosen K. Therefore, for

plotting the results, the Q file was used and the R function “barplot” was applied to graph them.

3.5 Phylogenetic analyses

To show the relationship between all the breeds graphically, phylogenetic or neighbour joining trees were constructed using the genotype data. Before the trees were built, the data was prepared in a few steps. The quality controlled data sets of the autosomal, Y and MT were taken and run again in PLINK with the command “distance-matrix”, which estimates IBS (identity-by-state) distance matrices. The IBS distance matrix expresses the genetic relationship between individuals. As the target was the clustering of animals into the different breeds, the tree should show the genetic distances just between the breeds and not between the individuals to get a clear arrangement. Therefore it was necessary to group the animals. This was operated with the help of R. Out of the distance matrices new ones were built; all animals were grouped breed-wise and the distances were calculated again by estimating the arithmetic means of the individuals belonging to each breed. The phylogenetic trees were ultimately built with R using the package “hclust” (KAUFMAN et al., 1990) out of these new and combined distance matrices.

3.6 Haplotype analyses

Haplotype analyses were conducted with the software FASTPHASE v1.2 (SCHEET et al., 2006) and R. Only for Y and MT data the haplotypes were built, due to the high number of haplotypes that would result from the use of autosomal data, which was out of the scope of this thesis. On the other hand mitochondrial data shows the maternal and Y the paternal inheritance, allowing to trace the possible parental origin of the ancestry. Therefore it would be possible to estimate first if there are haplotypes, which are only inherited from the sires or the dams, and second if there are any haplotypes, which are indicine or taurine specific.

Initially for the Y-chromosome all the heterozygote SNPs (78 of 98) were removed to get the homozygous ones, which were in sum 22 SNPs. The heterozygous SNPs in Y were assumed to be part of the pseudoautosomal region that suffers recombination with the X-chromosome, which consequently does not represent strictly the paternal inheritance. With this number of SNPs the correlations of the reference groups were calculated with R. All SNPs in perfect correlation, meaning that they have a correlation of one and are therefore giving the same information, were removed (13 SNPs). It ended up with a total number of 9 SNPs, with which the analysis was started. For mitochondrial data these steps were not necessary, all 27 SNPs were used.

Before running FASTPHASE with the remaining SNPs the binary files of Y and MT were again recoded by PLINK using the command “recode-fastphase” to get the right input files. These files consist of the number of diploid individuals, the number of SNP sites and the genotypes for each single individual. For the purpose of this thesis haplotypes were constructed by merging all the SNPs left in their right order.

The software was run for both Y and MT with the input file, a text file, which names the population information and three different K values: KL=2, KU=9 and KI=1. These values fix

the lower, the upper and the intervals between the numbers of populations or – as FASTPHASE uses the name – clusters. Additionally the command "F" was used to estimate the frequencies of the haplotypes.

To predict the haplotypes of the young Nelore cattle, the estimated haplotypes from the other animals were consulted. A text file was made with the number of the already known haplotypes, their hap ID and the haplotypes themselves. This file (command "b", which utilizes the already known haplotypes) was run with the input file consisting of the young Nelore animals and the command "F". The frequencies of the haplotypes were disclosed in a separate output file with the ending "freqs", where the table with the final haplotype frequency results was built from. Using hidden Markov models (HMM) to model cluster membership of the genotypes, together with expectation-maximization (EM) algorithms and Monte Carlo simulations, FASTPHASE imputes the missing SNPs through a "best guess" for each genotype approach and estimates haplotypes with their expected frequency in the population (SCHEET et al., 2006).

To verify the obtained results the haplotype frequencies were also calculated manually.

4 Results

In this chapter all relevant results of unsupervised and supervised estimation of admixture levels, the phylogenetic trees and the haplotype assessment of the genome-wide (except haplotype analysis), the Y-chromosomal and the mitochondrial analyses are shown and described.

4.1 Genome-wide (autosomal) analyses

Genome-wide analyses were performed by using the autosomal SNPs. All chromosomes except sex chromosomes (gonosomes) are defined as autosomes. The analyses are operated therefore nearly over the whole genome.

4.1.1 Genome-wide admixture

Genome-wide admixture plots show the different levels of admixture as the fraction of the genome belonging to a certain population expressed by the different colours. The admixture level of every single individual ranging from 0 to 1 is shown on the y-axis, and individuals are grouped according to breed, which is presented as the breed abbreviations on the x-axis.
For a better understanding and clarity the breeds on the x-axis are always presented in the same order in all plots of this thesis. First there are the six taurine breeds Angus (ANG), Fleckvieh (FLV), Hereford (HFD), Holstein (HOL), Limousin (LMS) and Piedmontese (PMT). The middle of the graphs is formed by the young Nelore animals of the production (YNLpro) and pedigree (YNLped) type. At the end there are the three indicine breeds ancestral Nelore (ANL), Brahman (BRA) and Gir (GIR).

4.1.1.1 Unsupervised admixture level estimation

Unsupervised admixture level estimation for the autosomal data was performed for K=2 to K=10. For the interest of this thesis only the results from K=2 and K=9 will be shown here. The admixture plots corresponding from K=3 to K=8, as well as K=10, can be found in the appendix of this thesis (see appendix 1 to 7).
In the first plot (figure 1) the number of populations or clusters was set to 2 (K=2). By setting K to 2 it was expected that there will be just a differentiation between the taurine and indicine subspecies. Clearly it can be seen that all taurine breeds have the same colour (red) throughout the whole y-axis which stands for an admixture level of 1.0 (100 % taurine descent). Indicine ancestry is marked with the colour blue. There can also be asserted that ANL, BRA and GIR are not totally (100 %) indicine in their autosomal genome origin. Brahman has the highest ratio of taurine ancestry with an average taurine admixture level of 0.133, Gir has the second highest with an average of 0.048 and the ancestral Nelore shows merely 0.001 of taurine admixture.
Young Nelore production and pedigree type animals have been sorted according to their taurine admixture level in a descending manner and in total the young Nelore population shows less than 1 % average taurine admixture. About half of the production type individuals

illustrate a taurine admixture, whereupon the highest taurine admixture level is at 0.041 for the production and at 0.025 for the pedigree type. For more details see figure 1.

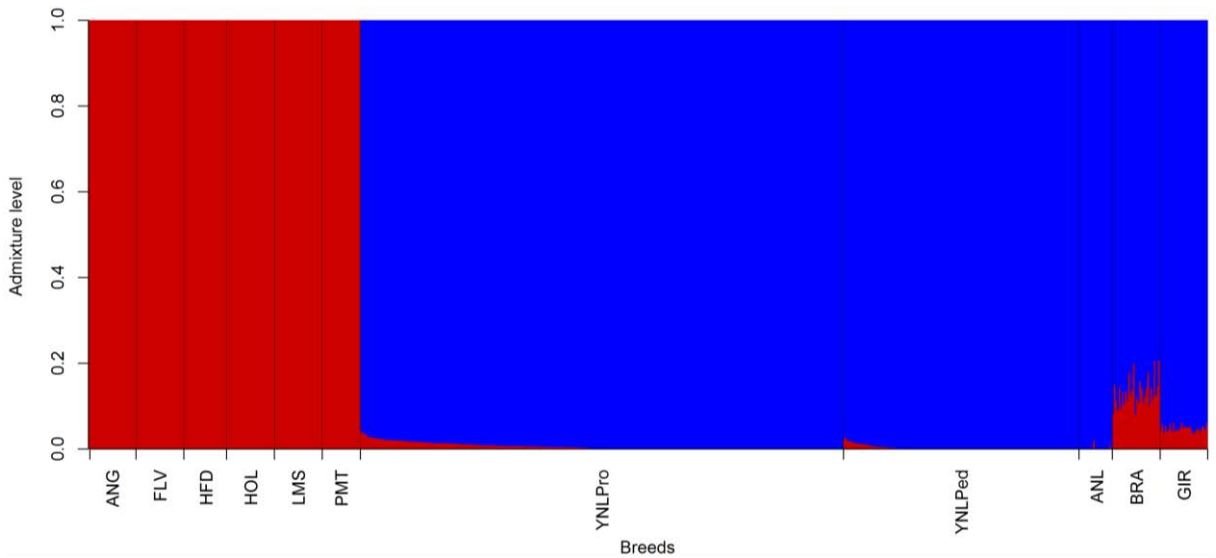


Figure 1: Unsupervised genome-wide admixture plot with two ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red colour represents the taurine and the blue the indicine ancestry.

Figure 2 shows the unsupervised genome-wide admixture level estimation, but with a different chosen number of populations. Here nine clusters (K=9) were selected, which stand for all breeds except the YNLs. It is evident that the colours purple, yellow, magenta and pink represent the taurine descent, while indicine colours are orange, dark and light blue, red and green. Noticeable is that yellow and purple colours are again represented in the indicine breed Brahman. Moreover we can see a colour pattern differentiation between the YNL groups, where the production and pedigree types are distinguishable in the varying heights of the admixture level fractions. For example the pedigree shows a higher level of dark blue while the production group shows more green and red fractions.

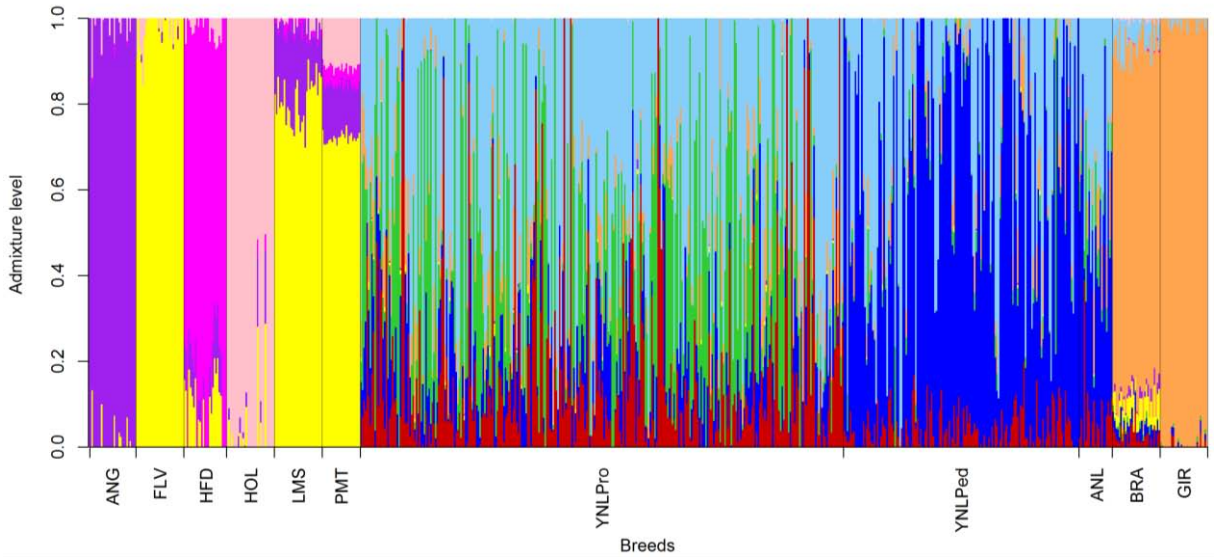


Figure 2: Unsupervised genome-wide admixture plot with nine ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The purple, yellow, pink and magenta colours represent the ancestry of taurine and the dark and light blue, red, orange and green colour of indicine breeds.

4.1.1.2 Supervised admixture level estimation

The next graph (figure 3) shows a supervised analysis with two set reference populations (K=2): taurine, including all taurine breeds, and indicine, including the indicine reference populations ANL, BRA and GIR. The software was then asked to estimate the admixture only for the YNLpro and YNLped. It was calculated that both production and pedigree group have an average taurine admixture level of 0.002, but the production group shows the highest peak of an individual at 0.028, while the pedigree group highest level is at 0.015.

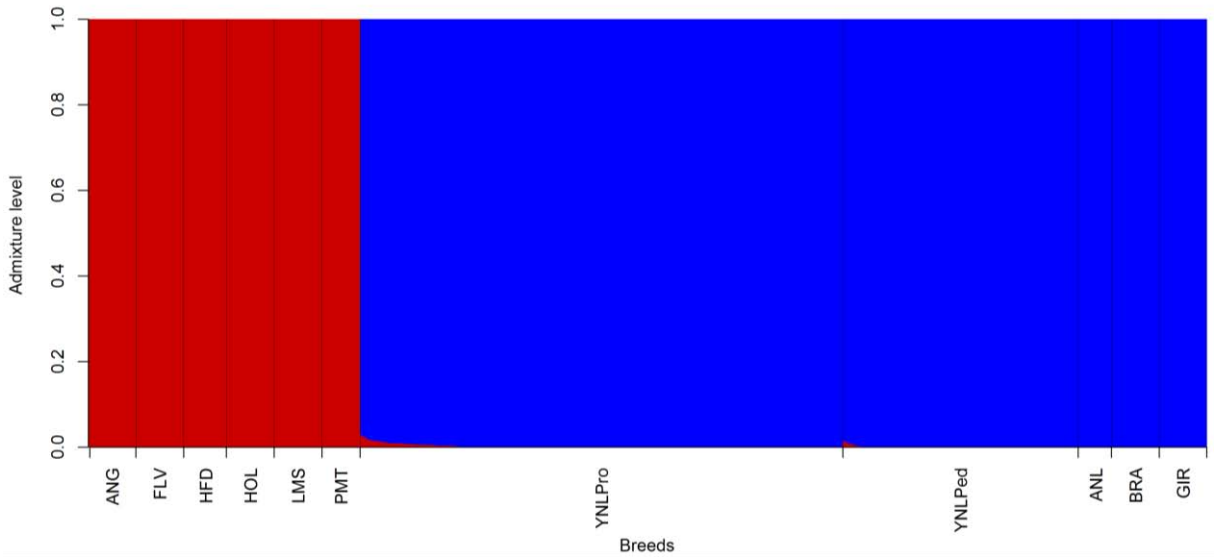


Figure 3: Supervised genome-wide admixture plot with two ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red colour represents the taurine and the blue the indicine ancestry.

The next plot (figure 4) shows the supervised estimation of admixture levels for the YNL groups using nine reference populations (K=9) corresponding to the nine breeds. A separation of the young Nelore mainly into ancestral Nelore, Brahman and Gir can be seen in the graph, with only a few spots of ancestry coming from different taurine breeds. It is also observable that the YNLped shows a higher ANL ancestry than the YNLpro.

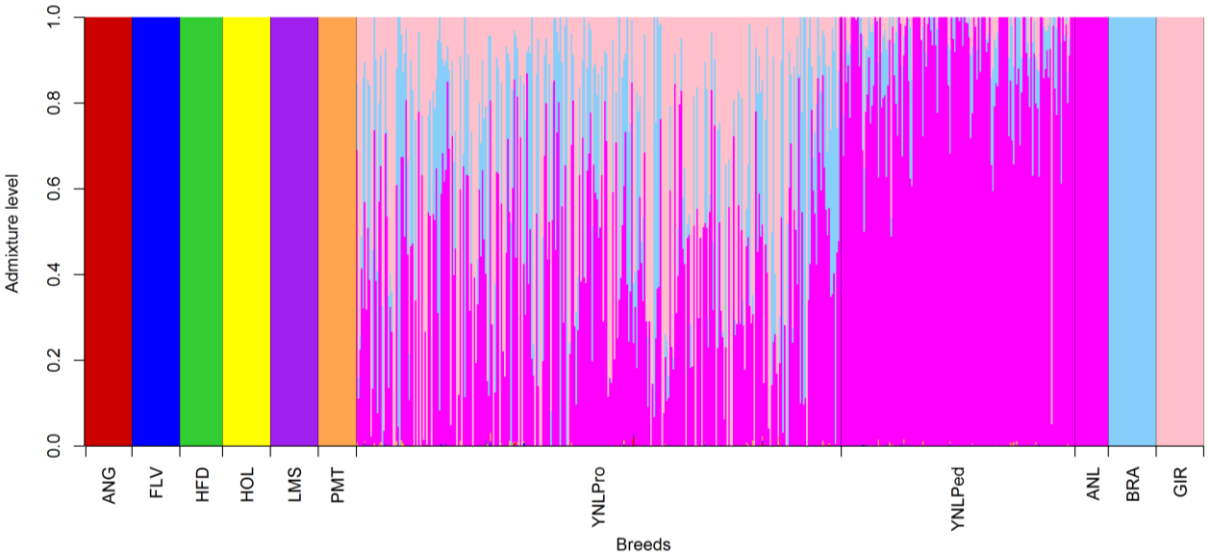


Figure 4: Supervised genome-wide admixture plot with nine ancestral populations
The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red, dark blue, green, yellow, purple and orange represent the taurine breeds ANG, FLV, HFD, HOL, LMS and PMT and magenta, light blue and pink the indicine breeds ANL, BRA and GIR.

4.1.2 Genome-wide phylogenetic tree

The created neighbour joining tree, based on autosomal data, shows a clear differentiation between taurine (right-hand side with HFD, PMT, FLV, LMS, ANG and HOL) and indicine breeds (left-hand side with BRA, Gir, YNLpro, YNLped and ANL). It is noticeable that YNLped and ANL are very close to each other standing at the same branch.

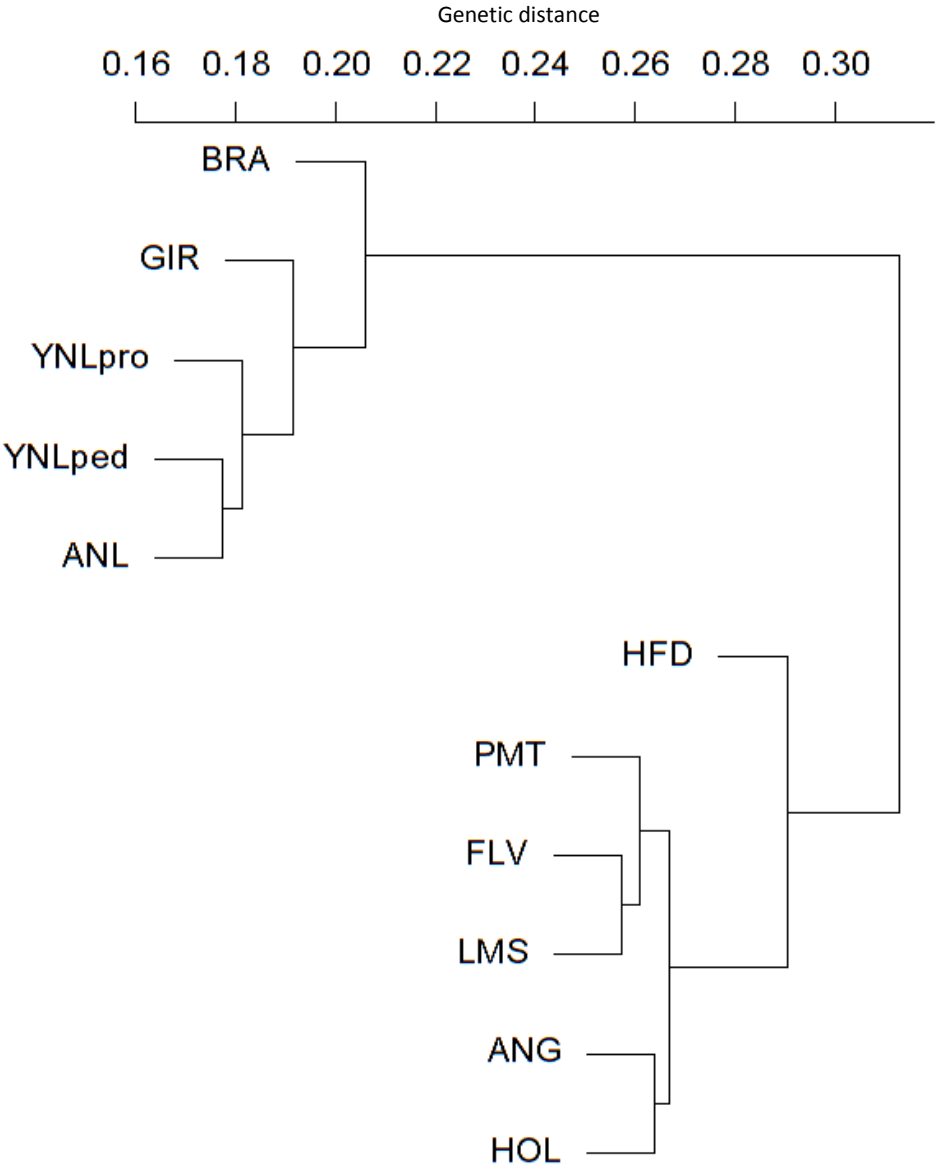


Figure 5: Genome-wide phylogenetic tree

In the top-axis the estimated genetic distance is shown. The length of the horizontal branches represents the distance between breeds. A clear differentiation of the taurine breeds, bottom right corner, and the indicine breeds, top left corner, can be seen.

4.2 Y-chromosome analyses

Y-chromosome analyses were performed to investigate the influence of the paternal inheritance on the taurine introgression. Additionally also an analysis of haplotypes and their frequencies was operated.

4.2.1 Y-chromosomal admixture

Also with Y-chromosomal data admixture plots were generated. There are unsupervised and supervised analyses from K=2 and K=9.

4.2.1.1 Unsupervised admixture level estimation

Unsupervised admixture level estimation for the Y-chromosome data was performed for K=2 to K=10. For the interest of this thesis only the results from K=2 and K=9 will be shown here. The admixture plots corresponding from K=3 to K=8, as well as K=10, can be found in the appendix of this thesis (see appendix 8 to 14).

The next plot shows the unsupervised admixture level estimation with two assumed populations (K=2). In Fleckvieh (FLV) and Limousin (LMS) one outlier was found: In FLV there is one individual with an indicine introgression of 0.186 and in LMS one with 0.240. Among the YNLpro animals 28 show taurine ancestry, the highest rate is at 0.312 with an average of 0.009 of the whole group. In YNLped twelve animals out of 148 show some level of taurine admixture. The highest level is at about 14 % (0.139) with an average of the whole group of 0.007. In the ancestral Nelore two out of 21 animals have taurine introgression according to this analysis. This is different to Brahman (BRA) and Gir (GIR) where more than half and about a quarter of the animals respectively, exhibit taurine descent; the highest peaks are 0.684 and 0.651, and the arithmetic means of the breeds are 0.146 and 0.077 correspondingly.

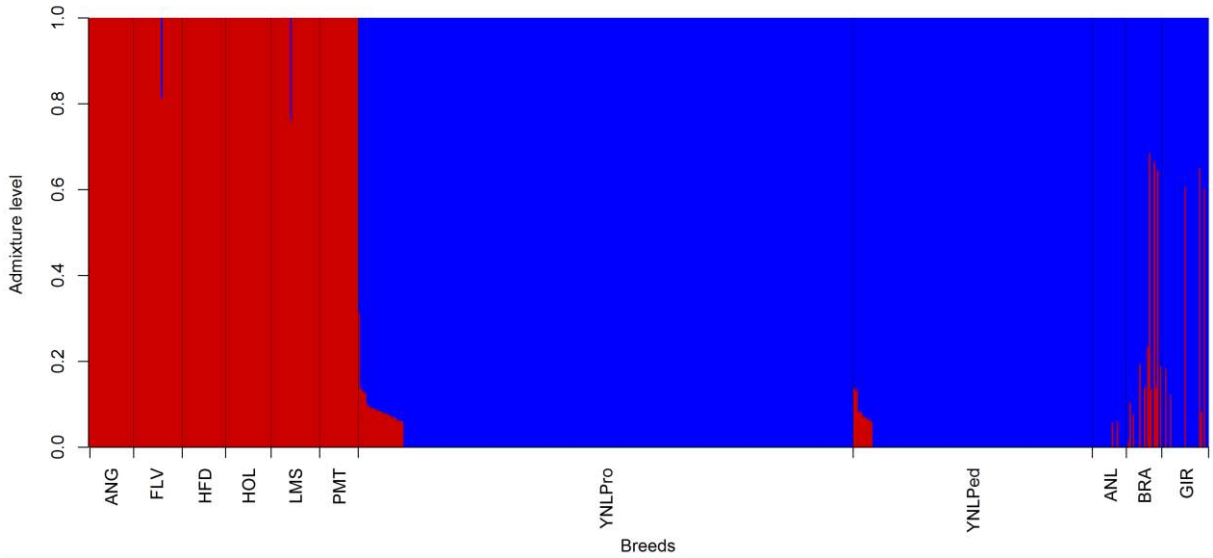


Figure 6: Unsupervised Y-chromosome admixture plot with two ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red colour represents the taurine and the blue the indicine ancestry.

Figure 7 shows unsupervised Y-chromosomal estimation of admixture levels with a chosen population of nine (K=9). A colour separation of taurine (light blue, magenta, green and purple) and indicine (dark blue, red, yellow, orange and pink) breeds can be noticed showing a breakup between the subspecies, but there is a fractioning of the admixture particular for every single individual, not allowing a differentiation between breeds.

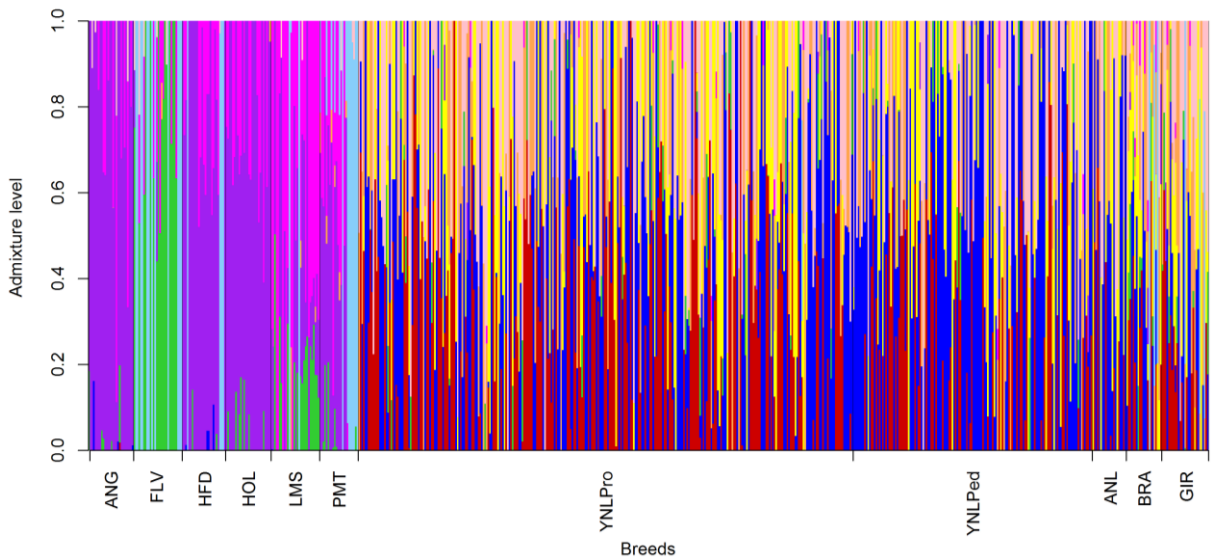


Figure 7: Unsupervised Y-chromosomal admixture plot with nine ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The purple, green, light blue and magenta colours represent the ancestry of taurine and the dark blue, red, orange, yellow and pink colour of indicine breeds.

4.2.1.2 Supervised admixture level estimation

Supervised, or in other words giving the population information, analyses were also operated for the Y-chromosome with two and nine populations (K=2 and K=9) respectively.

The following graph (figure 8) shows the admixture levels of taurine and indicine (K=2). The taurine breeds (left side, red colour) were fixed as 100 % taurine and the indicine breeds (right side, blue colour) as 100 % indicine. In the middle there are again the young Nelore animals, where the level of admixture was estimated. Production type individuals (28 animals) show taurine introgression at an average of only 0.005 (highest rate of a single individual at 0.131). The YNLped has twelve animals out of 148 with taurine ancestry; the average for the whole cluster is 0.004 and the highest peak of one individual is at about 11 % (0.107).

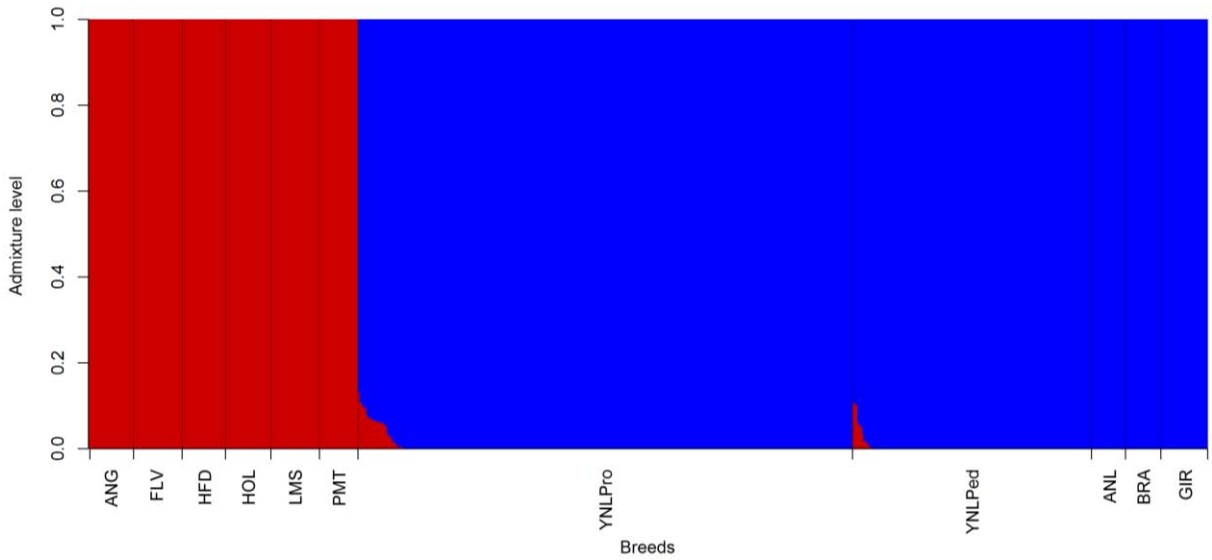


Figure 8: Supervised Y-chromosomal admixture plot with two ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red colour represents the taurine and the blue the indicine ancestry.

In the following graph (figure 9) nine clusters were chosen for the assessment of the admixture levels of the YNL groups. The population information (breeds) was again stated. YNLpro as well as YNLped show a predominance of the colours magenta, light blue and pink, which represent the indicine breeds ANL, BRA and GIR. There is a relatively high percentage (in comparison to the other taurine breeds) of Angus ancestry, although small derivation of the other taurine breeds can also be seen in the young Nelore animals.

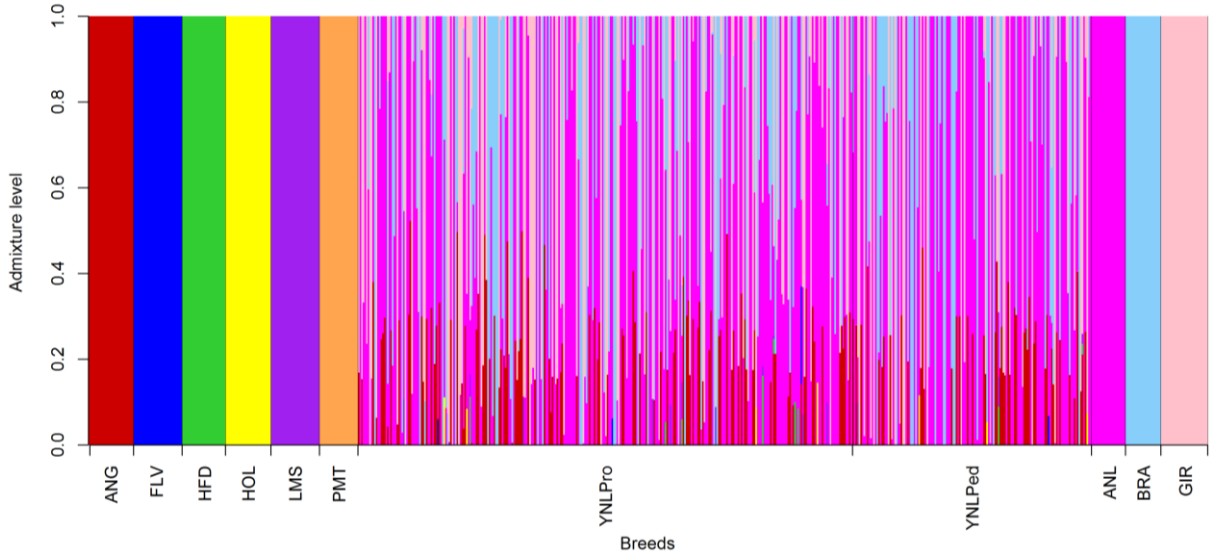


Figure 9: Supervised Y-chromosomal admixture plot with nine ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red, dark blue, green, yellow, purple and orange represent the taurine breeds ANG, FLV, HFD, HOL, LMS and PMT and magenta, light blue and pink the indicine breeds ANL, BRA and GIR. The young Nelore groups were estimated.

4.2.2 Y-chromosomal phylogenetic tree

The following phylogenetic tree (figure 10) shows the calculated genetic distances between the eleven groups estimated from the Y-chromosome SNPs. It is visible that the indicine and taurine breeds are separated from each other. The young Nelore pedigree individuals are closer to the ancestral Nelore animals than the production type animals. The order of the genetic distance from the indicine subspecies is the same as in the genome-wide analysis. In the taurine breeds ANG and HOL stand on the same level and the next closest breed is HFD, followed by LMS, PMT and FLV. Noticeably, FLV is the furthest away from the rest of the taurine breeds. Here we can clearly see the difference from autosomal or genome-wide to the paternal inheritance as shown by the Y-chromosome SNPs.

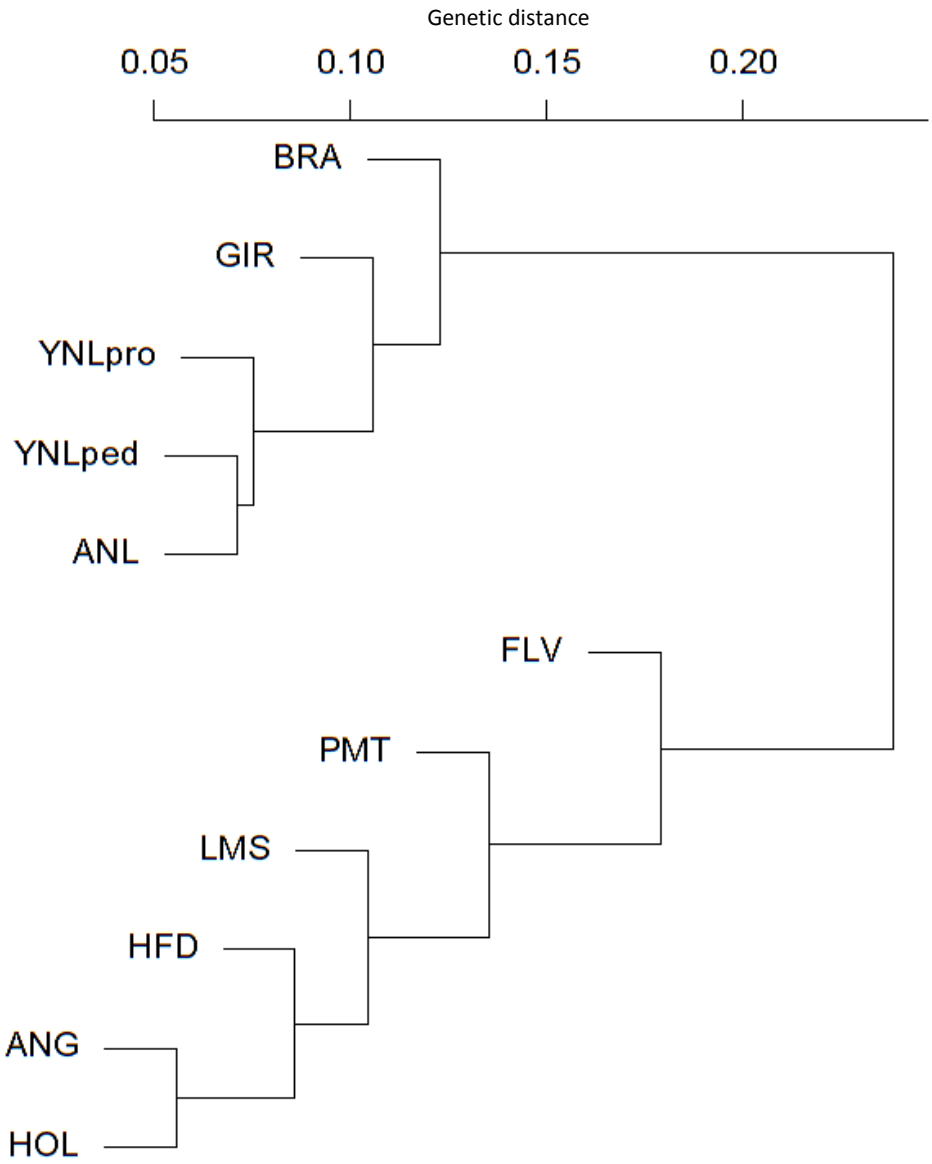


Figure 10: Y-chromosomal phylogenetic tree

In the top-axis the estimated genetic distance is shown. The length of the horizontal branches represents the distance between breeds. A clear differentiation of the taurine breeds, bottom, and the indicine breeds, top left corner, can be seen.

4.2.3 Y-chromosomal haplotype analysis

Breed or subspecies specificity can be explored by estimating haplotypes. In total there were five different haplotypes found. One haplotype appears in every one of the eleven groups in both calculation methods (Monte Carlo and manual with identical results) and at the same time it has the highest frequency in every breed. In the both computations one hap was found in five breeds (HFD, LMS, PMT, BRA and GIR), and the other three haps were in each case in one breed (FLV, PMT and ANG). Fleckvieh is represented by two haplotypes. The first hap, which is in all other breeds relatively high (0.85 to 1), shows a frequency of 0.6 (only in PMT it is lower with 0.417) and is therefore worth mentioning. The second one shows a frequency of 0.4 and is at the same time a breed unique haplotype. In the following table (table 5) there are the breed per breed estimated Y-chromosome haplotypes and their frequencies, calculated by Monte Carlo method (FASTPHASE) and manually, rounded to three decimals.

For better understanding table 5 only shows the haplotype identification number instead of the whole haplotype. The estimated haplotypes with the appropriate hap-ID can be found in the appendix of this thesis (see appendix 15).

Table 5: Y-chromosome haplotype frequencies, calculated by FASTPHASE and manually

The colour background shows the highest and the second highest (more than 0.1) frequency in a breed in yellow and blue respectively, as well as the breed unique characterized haplotypes in orange.

	Breed										
Hap-ID	**ANG**	**FLV**	**HFD**	**HOL**	**LMS**	**PMT**	**ANL**	**BRA**	**GIR**	**YNLpro**	**YNLped**
1	0.964	0.600	0.852	1.000	0.933	0.417	1.000	0.864	0.897	1.000	1.000
2			0.148		0.067	0.333		0.136	0.103		
3		0.400									
4						0.250					
5	0.036										

4.3 Mitochondrial analyses

Equally to the before described Y-chromosomal results, mitochondrial analyses were done: Estimating admixture levels, calculating and creating a neighbour joining tree and the haplotype analysis.

4.3.1 Mitochondrial admixture

On the next pages the mitochondrial unsupervised and supervised admixture plots will be described. Again the levels were estimated with K=2 and K=9.

4.3.1.1 Unsupervised admixture level estimation

Unsupervised admixture level estimation for the mitochondrial data was performed for K=2 to K=10. For the interest of this thesis only the results from K=2 and K=9 will be shown here.

The admixture plots corresponding from K=3 to K=8, as well as K=10, can be found in the appendix of this thesis (see appendix 16 to 22).

In the next plot (figure 11), unsupervised with two chosen populations, there is a substantial difference to all precedent graphs: Much more admixture in every breed, except Piedmontese (PMT), can be asserted on both taurine and indicine breeds. In the taurine ANG, FLV and LMS the highest indicine fraction of an animal is found at 0.085 with an average from 0.003 to 0.008. Differently to that are the breeds HFD and HOL: The average of them lies in each case at 4 % (0.040), but both have an animal with 100 % indicine mitochondrial or, in other words, indicine maternal origin. A similar picture shows the right side, the indicine subspecies, of the graph: In the ancestral Nelore (ANL) there are seven of 21 animals with complete taurine ancestry, and only eight have a level of 100 % indicine. In Brahman (BRA) and Gir (GIR) there are 25 and 23 animals accordingly with 100 % taurine origin and the rest of the individuals of both populations show a complete indicine descent. YNLpro and YNLped behave in a similar way: 299 out of 305 production and 134 out of 147 pedigree type animals show complete taurine ancestry, while only four and twelve animals respectively exhibit 100 % indicine mitochondrial DNA. Although there are some individuals with a total maternal ancestry from indicine breeds, the arithmetic means of these groups are 0.013 and 0.085.

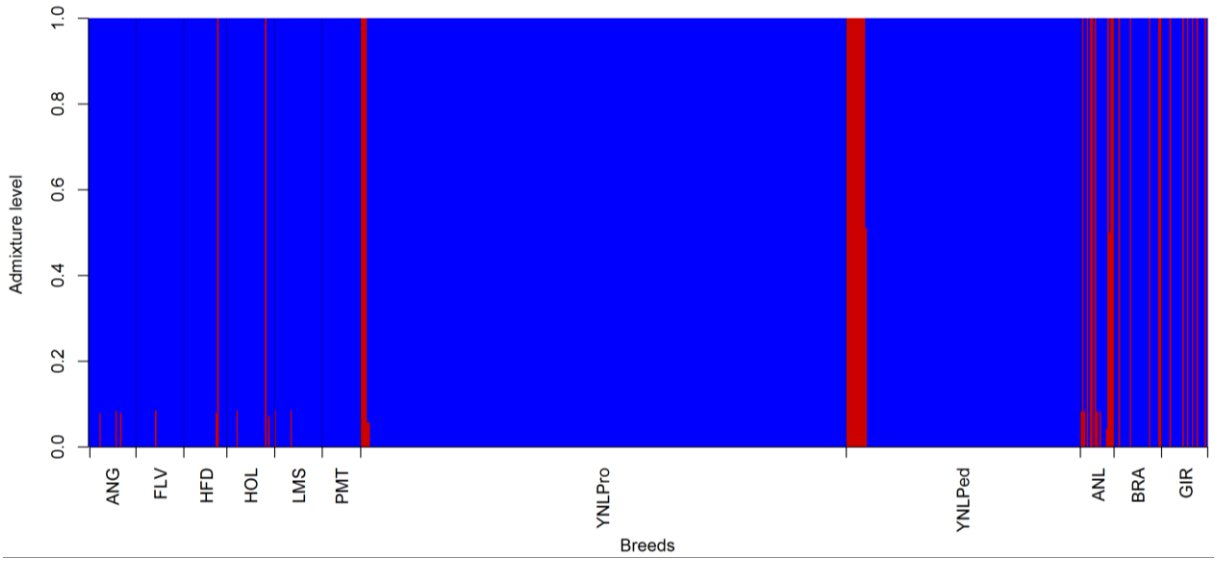


Figure 11: Unsupervised mitochondrial admixture plot with two ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The blue colour represents the taurine and the red the indicine ancestry.

The next graph (figure 12) shows the unsupervised estimation with nine populations (K=9). As it is visible pink, followed by purple, is the most dominant colour and is not totally relatable to one specific breed. This graph shows that maternal inheritance is hard to reconstruct with the available data, because several ancestries are seen in every single breed, and there is also no clear taurine or indicine differentiation.

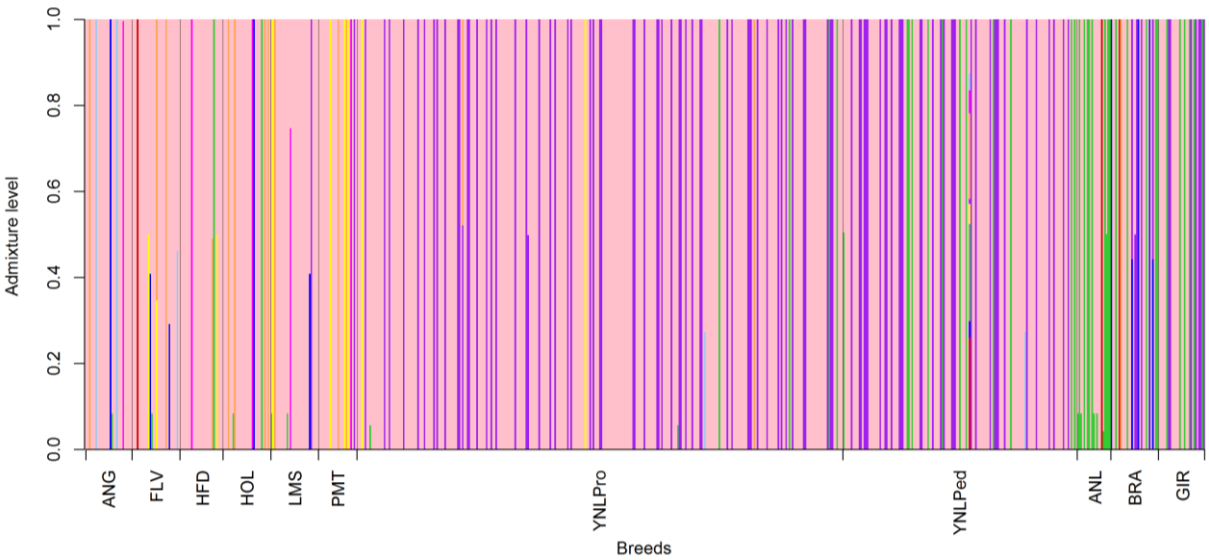


Figure 12: Unsupervised mitochondrial admixture plot with nine ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and indicine breeds.

4.3.1.2 Supervised admixture level estimation

The supervised admixture level estimation with two given populations (figure 13) gives a clearer picture: The left side in red colour representing the taurine, and the right side in blue colour again showing the indicine breeds. The YNL animals display more taurine than indicine ancestry. From 305 production individuals just six show indicine admixture and the taurine introgression level is on average 95.4 %. The pedigree group has 13 individuals with a total indicine and 134 with a total taurine origin. The average taurine introgression is 91.2 %.

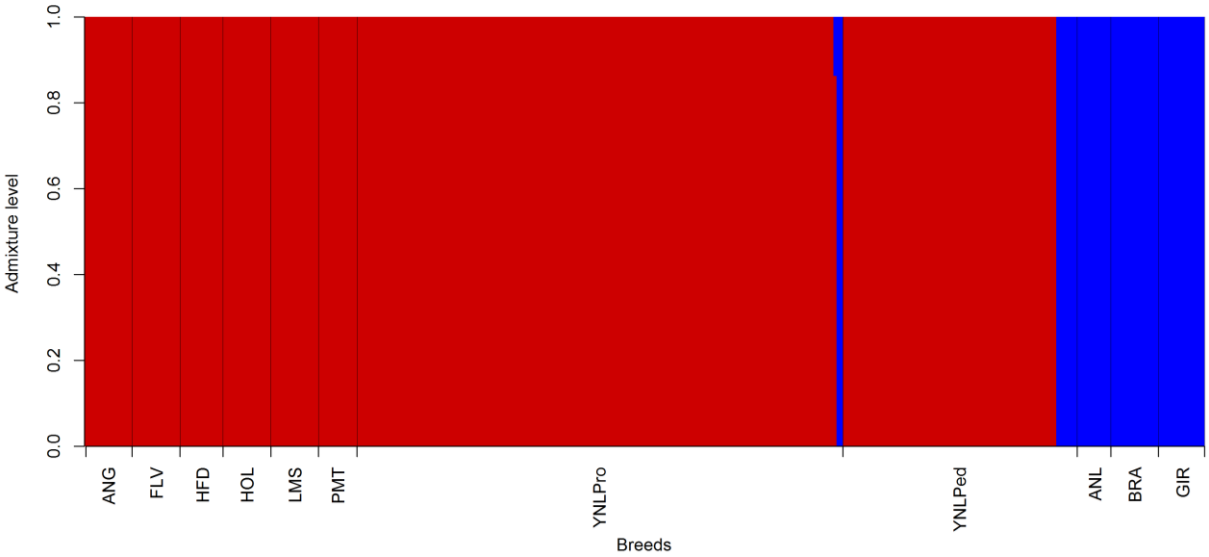


Figure 13: Supervised mitochondrial admixture plot with two ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red colour represents the taurine and the blue the indicine ancestry.

The mitochondrial supervised admixture graph with nine chosen populations (figure 14) confirms again the before shown graph from maternal inheritance. The most dominant colour in the two young Nelore groups is dark blue, which corresponds to the taurine breed Fleckvieh (FLV). Second most frequent colour is pink, which stands for GIR. Moreover the colours magenta (ANL) and orange (PMT) are also in YNL, the remaining breeds do not appear in YNL.

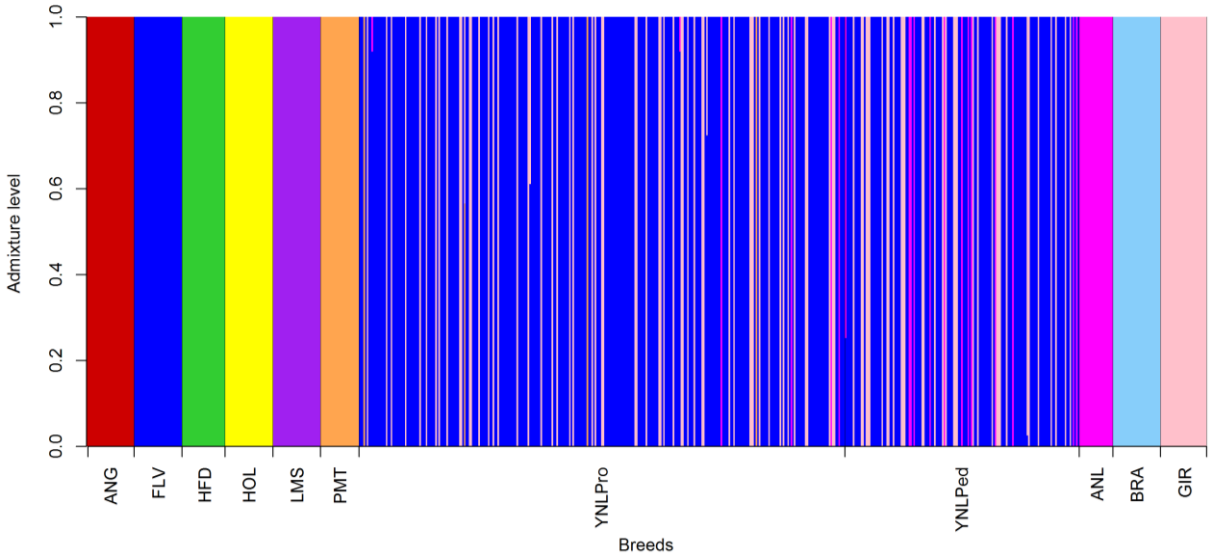


Figure 14: Supervised mitochondrial admixture plot with nine ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The red, dark blue, green, yellow, purple and orange represent the taurine breeds ANG, FLV, HFD, HOL, LMS and PMT and magenta, light blue and pink the indicine breeds ANL, BRA and GIR.

4.3.2 Mitochondrial phylogenetic tree

In the same way like the admixture plots the mitochondrial phylogenetic tree shows the same situation of maternal ancestry (figure 15). Unlike before in autosomal and Y-chromosomal analyses, the breeds here are not correctly grouped in taurine and indicine subspecies. PMT and YNLpro are the closest, in terms of genetic distance, to each other, followed by all other taurine breeds and then the YNLped. After that the indicine Brahman (BRA), Gir (GIR) and ANL (ancestral Nelore) are coming with a higher genetic distance to the taurine.

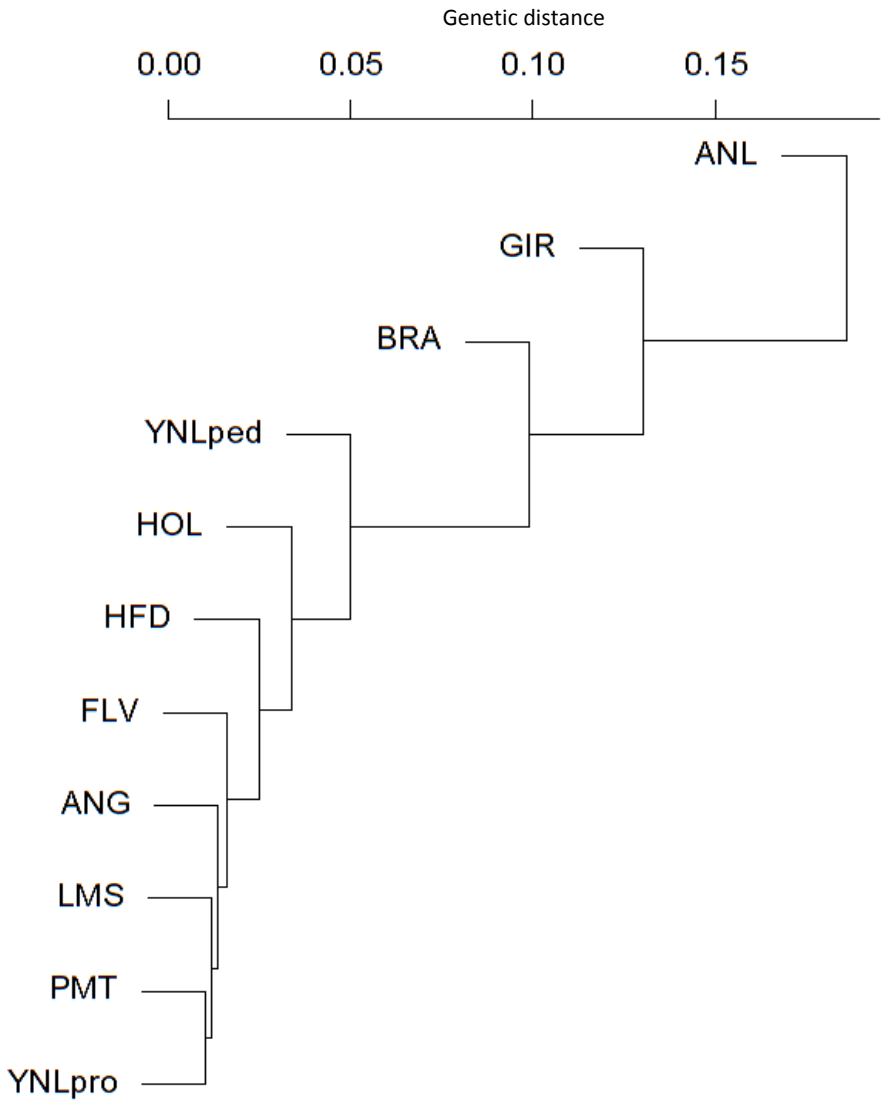


Figure 15: Mitochondrial phylogenetic tree

In the top-axis the estimated genetic distance is shown. The length of the horizontal branches represents the distance between breeds. There is no clear differentiation of the taurine and the indicine breeds.

4.3.3 Mitochondrial haplotype analysis

The calculations of the mitochondrial haplotypes (consisting of 19 SNPs) show in both Monte Carlo method and manual calculation very similar results with 15 haplotypes over all. One haplotype was found in every one of the eleven groups again with the highest frequency in all. Another hap was detected in seven (HFD, HOL, ANL, BRA, GIR, YNLpro and YNLped), one in five (ANG, FLV, HOL, LMS and BRA) and one in two (FLV and HOL) breeds. In the manual calculation all other eleven haplotypes are breed unique, while the Monte Carlo method finds just ten breed unique haps, and the one which is different appears is shared by HFD and LMS. In the following tables (tables 7 and 8) the breed per breed estimated mitochondrial haplotypes and their frequencies, calculated by Monte Carlo method (FASTPHASE) and manually, rounded to three decimals can be seen.

For better understanding the tables 7 and 8 only show the haplotype identification instead of the whole haplotype. The estimated haplotypes with the appropriate hap-ID for both calculation methods can be found in the appendix of this thesis (see appendix 23).

Table 6: Mitochondrial haplotype frequencies, calculated by FASTPHASE

The colour background shows the highest and the second highest (more than 0.1) frequency in a breed in yellow and blue respectively, as well as the breed unique characterized haplotypes in orange.

	Breed										
Hap-ID	ANG	FLV	HFD	HOL	LMS	PMT	ANL	BRA	GIR	YNLpro	YNLped
1	0.833	0.833	0.962	0.899	0.900	1.000	0.545	0.632	0.793	0.982	0.911
2			0.037	0.033			0.403	0.167	0.207	0.013	0.085
3	0.033	0.033		0.033	0.033			0.101			
4		0.033		0.033							
5	0.067										
6								0.067			
7			0.001		0.033						
8					0.033						
9	0.033										
10	0.033										
11								0.033			
12							0.047				
13		0.033									
14		0.033									
15		0.033									

Table 7: Mitochondrial haplotype frequencies, calculated manually

The colour background shows the highest and the second highest (more than 0.1) frequency in a breed in yellow and blue respectively, as well as the breed unique characterized haplotypes in orange.

	Breed										
Hap-ID	ANG	FLV	HFD	HOL	LMS	PMT	ANL	BRA	GIR	YNLpro	YNLped
1	0.833	0.833	0.963	0.900	0.900	1.000	0.524	0.633	0.793	0.987	0.918
2			0.037	0.033			0.429	0.167	0.207	0.013	0.082
3	0.033	0.033		0.033	0.033			0.100			
4		0.033		0.033							
5	0.067										
6								0.067			
7					0.033						
8					0.033						
9	0.033										
10	0.033										
11								0.033			
12							0.048				
13		0.033									
14		0.033									
15		0.033									

5 Discussion

In this chapter the described results (chapter 4) will be discussed, commented and interpreted. Inasmuch there are not many studies of Nelore cattle ancestry research available at the moment, most results are discussed without any comparisons to already existing literature.

5.1 Materials and methods

Before the results are going to be discussed, first there is a focus on the materials and methods of this master thesis. The used SNP genotypes had a total number of 777962 (777k) SNPs. This high number of SNPs can make the conclusions of the estimated results more powerful than with lower numbers, for example like investigations with microsatellites (GIOVAMBATTISTA et al., 2000), LD or 54k (FRKONJA et al., 2012) SNP chips. Most studies, which were presented in the literature background, used smaller SNP chip data. Yet, FRKONJA et al. (2012) found that for accurate estimation of autosomal admixture, the 54k SNP chip was more than sufficient, admixture results were virtually identical down to smaller sets of 8000 evenly spaced SNPs. While 777k information might not be particularly beneficial for evaluation of levels of crossbreeding analysis, FERENČAKOVIĆ et al. (2013) showed that it is substantially better for estimation of levels of inbreeding.

For the analyses 706 animals were used, which were most of the time powerful enough to conclude. Quality control parameters were chosen individually for the three data sets. In Y-chromosome there were many losses of SNPs, but most of them occurred in the mitochondrial data set. There were just 314 common SNPs and after minor allele frequency pruning only 27 left. These SNPs were not powerful enough to distinguish between the breeds or subspecies in the haplotypic analysis or in the phylogenetic trees. The same was found for Y, although the number of remaining SNPs was higher (98) and the phylogenetic tree seemed to be able to still divide the subspecies correctly.

Chosen software for the analyses of admixture levels, phylogenetic trees and building haplotypes were optimal for the purposes of this study and allow the completion of all objectives.

5.2 Autosomal (genome-wide) results

The autosomal analyses show the genome-wide results. The differentiation between the subspecies taurine and indicine with a chosen number of two populations and unsupervised admixture estimation (figure 1) illustrate that the supposed taurine breeds ANG, FLV, HFD, HOL, LMS and PMT are all sharing a common ancestry. In contrary to this the expected indicine breeds ANL, BRA and GIR are not totally purebred indicine; they have in some cases parts of taurine admixture. In Brahman the average of 13.3 % could be explained by their crossbreeding history (AUSTRALIAN BRAHMAN BREEDERS' ASSOCIATION LIMITED, s. a.). YNL animals both production and pedigree are showing lower taurine percentages. As these animals have historically been interbred with the taurine Creole cattle (VOZZI et al., 2007), it was expected that there will be higher levels of taurine admixture in these two groups than what the results show: The whole population of young Nelore is nearly up to 100 % on average indicine purebred, with a taurine admixture of 0.405 %.

The same analysis with a different number of clusters, K=9 (figure 2), shows disparity of the production and the pedigree type. YNLped correlate more with the ancestral Nelore animals whereas YNLpro show more indicine admixture. This might be explained by the differences in the breeding strategies of these two groups: YNLped are coming from registered herds, where the breeding was kept to only registered animals with high emphasis on specific lineages (DANI et al., 2008) and it focused on type traits and breed specific characteristic. In a different way from that, YNLpro descended out of breeding programs where the main target was on beef production in a commercial way and little importance is given to the registration status and "pureness" of the fathers (DANI et al., 2008). Again it can be seen that there are just small traces of taurine ancestry in the young Nelore animals, thus they can be categorized as zebus with this result as well. Also BRA shows taurine introgression on the same level as with two chosen populations.

When the population information is announced with K=2 (figure 3) the share in taurine in the YNL groups is decreasing in comparison to the unsupervised analysis. With 0.2 % in the pedigree and 1.5 % in the production group on average they wouldn't be considered as admixed animals by most breed association standards. However, the production group shows higher taurine introgression than the pedigree, which can be traced back again to the differences in the breeding strategies of these two groups as explained by DANI et al. (2008).

Estimation of the admixture level with nine populations (figure 4) highlights a partitioning of the YNL animals into ANL, BRA and GIR. YNLped is more related to the ANL, whereas YNLpro is relatively fragmented in equal parts to ANL, BRA and GIR. That acknowledges also the assumption that the pedigree type should be more equal to its ancestry than the production type. Only a few spots of ancestry coming from different taurine breeds can be recognized. Comparing it to the unsupervised with nine chosen populations (figure 2), the supervised analysis again acknowledges the assumptions, which were made with the previous results.

These assumptions from estimating the admixture level are also reflected in the phylogenetic tree (figure 5): Examination of the genetic distances of the breeds to each other demonstrates that YNLped and ANL are very close together standing on the same branch. However, the arrangements in two groups show a clear differentiation in taurine and indicine subspecies. Brahman is the closest neighbour to the taurine breeds in reference to the estimated genetic distances. Also with the neighbour joining tree it is acknowledged that the young Nelore individuals are genome-wide of zebuine ancestry, mostly coming from the ancestral Nelore progenitors.

An analysis chromosome by chromosome of the whole genome could indicate if there is admixture happening preferentially on certain genomic regions. This was not part of the thesis and should therefore be taken into consideration for later studies on autosomal data of Nelore cattle.

5.3 Y-chromosomal results

The Y-chromosomal analyses show the paternal way of inheritance: Unsupervised admixture level estimation with a chosen number of two populations (figure 6) shows more, but not significant, taurine origin in the indicine breeds than the genome-wide. Although the highest

taurine peak in YNLped is at about 14 %, the whole group shows on average only 0.7 %. YNLpro group represent 0.9 % taurine introgression with the highest rate of an individual at about 31 %. It can be interpreted that the indicine ancestry, although there are some outliers, is coming from the sire side. In BRA the crossings with taurine breeds are acknowledged with an average of nearly 15 % again, with more than the half of the animals apparently descending from taurine sires.

The graph with K=9 (figure 7) displays a trend towards the genetic information from the paternal side of YNL belonging to the zebuine subspecies showing similar colours as the breeds ANL, BRA and GIR. The colour patterns suggest a differentiation between taurine and indicine subspecies but a clear separation between the single breeds cannot be distinguished.

Supervised estimation with K=2 (figure 8) and K=9 (figure 9) show similar results to the unsupervised ones, namely a small percentage of taurine introgression in the current Nelore animals. The admixture levels of the animals of both YNL groups are higher than in genome-wide estimation, but again lower than 1 %. The result for investigation with stated population information (figure 9) tells, as well as for genome-wide, in most cases paternal zebuine ancestry. Compared to the autosomal analysis the separation of the YNLped and YNLpro groups, now is not so definite anymore. This might be attributed to the low number of SNPs and therefore the lower accuracy of the data used in the Y-chromosomal analyses.

The neighbour joining tree, drawn with the help of the paternal data set (figure 10), illustrates again the clear separation of the two subspecies. The order of the genetic distance from the indicine subspecies is the same as in the genome-wide analysis, while the distance is here shorter. Comparing it to the previous neighbour joining tree (figure 4), this one confirms the structure again: The YNLped group has the shortest genetic distance to the ANL and is therefore more closely related to them than the YNLpro, as nearly all before shown results already exhibited. In the group of the taurine breeds there is a short genetic distance of ANG and HOL, which are standing on the same branch. They are thus the closest related breeds from the taurines based on Y-chromosome. FLV is the furthest away from the rest of the taurine breeds and at the same time the nearest related breed to the indicine group. With the different ordering a difference with the autosomal phylogenetic tree can be seen.

With the estimation of the Y-chromosome haplotypes it can be possible to distinguish breed or subspecies specificity, which were, if detectable, inherited from the paternal side. Surprisingly, in table 5, the first assessed haplotype (111111111) is occurring with the highest percentage, from about 42 % to 100 %, in all breeds regardless of which subspecies, so specificity cannot be alluded. The same is the case for the second hap (111110111), because it is traceable in three taurine and two indicine breeds. Different to that are the haplotypes numbers three (000011000), four (011111011) and five (111101111). The third only exists in the breed FLV, the fourth in PMT, and the fifth in ANG. These haplotypes show breed, or in other words Fleckvieh, Piedmontese and Angus, specificity. Therefore they can be termed as breed unique haplotypes. In FLV the hap has the second highest frequency at the same time. This estimation of the Y-chromosomal haplotypes shows a different result to the phylogenetic tree. With the tree it is possible to distinguish between the breeds and the subspecies, the haps do not give this information. The reason is the different used data set for these two

analyses, the results from the table and the tree stand therefore in a conflict. For building the haplotypes, all heterozygous SNPs (in theory coming from the pseudoautosomal Y region) and SNPs found in perfect correlation were deleted. Apparently these SNPs expressed variation that allowed breed and species differentiation, but is also important to take into account that they were also much more in number (nine SNPs for the haplotypes and 98 SNPs used for the tree estimation), which might explain the increased information obtained from them.

To summarize, it is not possible to explain the differences of the breeds or the subspecies with these SNPs, because they are not powerful enough. For future analyses it would be better to use sequence data for example, to get more clear conclusions and haplotypes, with which it is possible to distinguish the subspecies and maybe the breeds or at least some breed groupings.

5.4 Mitochondrial results

The calculated mitochondrial results show the maternal inheritance and have a big disparity to the genome-wide and Y-chromosomal estimations. The separation of taurine and indicine with unsupervised K=2 admixture analyses (figure 11) leads to the statement that all young Nelore animals are to a high part of taurine descent along the maternal line, which they have inherited historically from use of taurine dams: 299 of production and 134 of pedigree type animals show up to 100 % taurine ancestry. In comparison to this, the graph with nine given clusters (figure 12) does not show a typical differentiation between the breeds. The pink colour can be seen as taurine characteristic, and it can be found in every breed. Unsupervised estimation of breed belonging does not show breed or subspecies specificity, although it can still be interpreted that the young Nelore groups have their maternal ancestry more from all the taurine breeds.

In the assessment of the young Nelore animals with K=2 (figure 13), YNLpro and YNLped have much more taurine than indicine maternity origin and the whole groups have on average 95.4 % and 91.2 % taurine ancestry respectively. Splitting this conclusion into breeds, the following picture appears: Unsupervised estimation with K=9 (figure 14) shows that the taurine maternal ancestry in both YNL is more closely related to the breed FLV, followed by ANL and GIR. Also some mixtures of PMT are visible. ANG, HFD, HOL, LMS and BRA do not appear in the YNL groups; this could argue that YNL have no mitochondrial ancestry from these breed groups.

Unlike before in autosomal and Y-chromosomal analyses, the breeds in the created mitochondrial phylogenetic tree (figure 15) are not correctly separated into taurine and indicine subspecies. PMT and YNLpro are the closest related. Mitochondrial genetic distances of young Nelore production and pedigree type individuals are less to the taurine breeds and this shows a more likely taurine descent from the maternal side than from the indicine breeds.

In mitochondrial haplotype construction more breed unique haps were found than in the Y-chromosomal analysis (table 7 is going to be discussed here, because both tables have nearly the same results): The first haplotype found appears in all breeds, regardless of their sub-

species. Simultaneously it is again the hap with the highest frequency (from 52.4 % to 100 %) in all breeds, even though the frequencies are higher in taurine than in indicine breeds. The second hap indicates a little bit more subspecies specificity: Although it can be found in HFD, HOL and in all the indicine breeds, in ANL, BRA and GIR the frequencies are from about 17 % up to about 43 %, while in HFD and HOL we can only denote an incidence of approximately 4 %. The third and the fourth could be interpreted as taurine haplotypes, although their frequencies are not very high (0.033). The rest of the built haplotypes (eleven haps) can be described as breed specific haplotypes, because of their occurrences in just one breed. Eight of them belong to taurine (three to Angus, three to Fleckvieh and two to Holstein) and three of them to indicine breeds (one to ancestral Nelore and two to Brahman). Different from the Y-chromosomal analyses, a little conflict of the results from the phylogenetic tree and the haplotype tables was found. The data set used for these two analyses was only slightly different as only nine SNPs with a correlation of one were removed.

Comparing these mitochondrial results to the study of MEIRELLES et al. (1999) there are many similarities: In their investigation Nelore and Gir showed 58 % taurine mitochondrial ancestry on average, and Braham was completely of taurine descent. Overall, a similarity of genome-wide and Y-chromosomal results was observed, while results of the mitochondrial SNPs were substantially different.

6 Conclusions

All these discussed results lead to the following conclusions: Autosomal admixture analyses show on average a minor taurine introgression (YNLpro with 0.525 % and YNLped with 0.285 %) of the YNL, which are therefore almost of complete indicine ancestry, whereupon in both cases (supervised and unsupervised) the taurine admixture is higher in the production group. Paternal inheritance, studied from Y-chromosome, shows a high ancestral Nelore descent, supporting the hypothesis of mainly use of ANL sires on the transmission of indicine genetics. Opposite to this, maternal inheritance, as evaluated from mitochondrial analyses, shows a taurine origin in both Nelore groups, with slightly more indicine in pedigree (8 %) than in production (5 %) type.

Phylogenetic analyses confirm with all three datasets, autosomal, Y-chromosomal and mitochondrial, the statements about the descent of the current Nelore animals: A clear separation of taurine and indicine breeds is recognized, where YNLped is related slightly more closely to the ancestral Nelore than YNLpro, while in mtDNA analyses both YNLs are closer in their genetic distances to taurine than to zebuine breeds. Mitochondrial inheritance and maternal descent as such is therefore almost completely coming from taurine ancestry.

Haplotype analyses show only a small differentiation of taurine and indicine haplotypes, adapted from paternal and maternal inheritance, and only very few breed specific haplotypes. The haplotype analyses therefore failed to distinguish breed and subspecies, with only one dominating haplotype found in all populations. To get a clearer picture of haplotypes it would be more efficient and meaningful to use next generation sequence data, because the used SNPs in this thesis were not powerful enough to make clear conclusions. This was not part of the thesis, though it should be contemplated for next and deeper analyses of the current population of Nelore cattle in Brazil.

Finally, the main conclusion is that taurine introgression in the current Brazilian Nelore cattle population is not as high as expected from the historical records, but there is still minimal taurine admixture observed in parts of the individuals.

References

ACNB (2006): A Raça: Histórico. Associação dos Criadores de Nelore do Brasil.
http://www.nelore.org.br/Raca/Historico
visited on 2013-11-24

AJMONE-MARSAN P., GARCIA J. F., LENSTRA, J. A. (2010): On the origin of cattle: How aurochs became domestic and colonized the world. Evol. Anthropol., Volume 19, 148-157

ALEXANDER D. H., NOVEMBRE J., LANGE K. (2009): Fast model-based estimation of ancestry in unrelated individuals. Genome Research, Volume 19, 1655-1664

AMERICAN ANGUS ASSOCIATION (2013): Angus History.
http://www.angus.org/pub/Anghist.aspx
visited on 2013-11-27

ANABORAPI – ASSOCIAZIONE NAZIONALE ALLEVATORI BOVINI DI RAZZA PIEMONTESE (2013): History of the Breed.
http://www.piedmontese.org/History%20of%20Breed.html
visited on 2013-11-27

ANDERUNG C., HELLBORG L., SEDDON J., HANOTTE O., GÖTHERSTRÖM A. (2007): Investigation of X- and Y-specific single nucleotide polymorphisms in taurine (Bos taurus) and indicine (Bos indicus) cattle. Animal Genetics, Volume 38, 595-600

AUSTRALIAN BRAHMAN BREEDERS' ASSOCIATION (s. a.): About Brahmans – A brief history.
http://www.brahman.com.au/history.html
visited on 2013-11-27

BALDING D. J., BISHOP M., CANNINGS C. (2007): Handbook of Statistical Genetics – Third Edition. John Wiley & Sons Ltd., Volume 1, XXXVII

BASU A., TANG H., XIAOFENG Z., GU C., HANIS C., BOERWINKLE E., RISCH N. (2008): Genome-wide distribution of ancestry in Mexican Americans. American Journal of Human Genetics, Volume 124, 207-214

BEAL M. J., GHAHRAMANI Z., RASMUSSEN C. E. (2001): The Infinite Hidden Markov Model. Advances in Neural Information Processing Systems 14, Cambridge, MIT Press, 1-3

BMLFUW – Österreichisches Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft (2011): Fleckvieh.
http://wisa.lebensministerium.at/article/articleview/89129/1/25578/
visited on 2013-11-27

British Limousin Cattle Society (2010): Breed History.
http://limousin.co.uk/the-breed/breed-history/
visited on 2013-11-27

Dani M. A. C., Heinneman M. B., Dani S. U. (2008): Brazilian Nelore Cattle: a melting pot unfolded by molecular genetics. Genetics and Molecular Research, Volume 7, 1127-1137

Deutscher Holstein Verband e. V. (2013): Über 130 Jahre Deutsche Holsteinzucht
http://www.holstein-dhv.de/geschichte.html
visited on 2013-11-26

Ferenčaković M., Sölkner J., Curik I. (2013): Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genetics Selection Evolution, 45:42

Flecha J. P. (1997): Nelore. Oklahoma State University, Department of Animal Science, Division of Agricultural Sciences and Natural Resources.
http://www.ansi.okstate.edu/breeds/cattle/nelore/
visited on 2013-09-29

Flori L., Gonzatti M. I., Thevenon S., Chantal I., Pinto J., Berthier D., Aso P. M., Gautier M. (2012): A Quasi-Exclusive European Ancestry in the Senepol Tropical Cattle Breed Highlights the Importance of the slick Locus in Tropical Adaptation. PLoS ONE, Volume 7, Issue 5, 1-10

Frkonja A., Gredler B., Schnyder U., Curik I., Sölkner J. (2012): Prediction of breed composition in an admixed cattle population. Animal Genetics, Volume 43, 696-703

Giovambattista G., Ripoli M. V., De Luca J. C., Mirol P. M., Liro´n J. P., Dulou F. N. (2000): Male-mediated introgression of Bos indicus genes into Argentine and Bolivian Creole cattle breeds. Animal Genetics, Volume 31, 302-305

Gorbach D. M., Makgahlela M. L., Reecy J. M., Kemp S. J., Baltenweck I., Ouma R., Mwai O., Marshall K., Murdoch B., Moore S., Rothschild M. F. (2010): Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. Journal of Animal Breeding and Genetics, Volume 127, 348-351

KAUFMAN L., ROUSSEEUW P. (1990): Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons, Inc.

ILLUMINA (2012): BovineHD Genotyping BeadChip.
http://res.illumina.com/documents/products/datasheets/datasheet_bovinehd.pdf
visited on 2013-09-05

MEIRELLES F. V., ROSA A. J. M., LÔBO R. B., GARCIA J. M., SMITH L. C., DUARTE F. A. M. (1999): Is the American Zebu really Bos indicus? Genetics and Molecular Biology, Volume 22, Issue 4, 543-546

NASSIR R., KOSOY R., TIAN C., WHIT P. A., BUTLER L. M., SILVA G., KITTLES R., ALARCON-RIQUELME M. E., GREGERSEN P. K., BELMONT J. W., DE LA VEGA F. M., SELDIN M. F. (2009): An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. BMC Genetics 2009, 10:39

PRICE A. L., PATTERSON N., YU F., COX D. R., WALISZEWSKA A., MCDONALD G. J., TANDON A., SCHIRMER C., NEUBAUER J., BEDOYA G., DUQUE C., VILLEGAS A., BORTOLINI M. C., SALZANO F. M., GALLO C., MAZZOTTI G., TELLO-RUIZ M., RIBA L., AGUILAR-SALINAS C. A., CANIZALES-QUINTEROS S., MENJIVAR M., KLITZ W., HENDERSON B., HAIMAN C. A., WINKLER C., TUSIE-LUNA T., RUIZ-LINARES A., REICH D. (2007): A Genomewide Admixture Map for Latino Populations. American Journal of Human Genetics, Volume 80, 1024-1036

PRITCHARD J. K., WEN X., FALUSH D. (2010): Documentation for STRUCTURE software, Version 2.3.
http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/structure_doc.pdf
visited on 2013-12-07

PURCELL S., NEALE B., TODD-BROWN K., THOMAS L., FERREIRA M. A. R., BENDER D., MALLER J., SKLAR P., DE BAKKER P. I. W., DALY M. J., SHAM P. C. (2007): PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, Volume 81, 559-575

R DEVELOPMENT CORE TEAM (2008): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0
http://www.R-project.org
visited on 2013-09-05

SANDERS O. C. (1980): History and Development of Zebu Cattle in the United States. Journal of Animal Science, Volume 50, 1188-1200

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, Volume 78, 629-644

SHRIVER M. D., PARRA E. J., DIOS S., BONILLA C., NORTON H., JOVEL C., PFAFF C., JONES C., MASSAC A., CAMERON N., BARON A., JACKSON T., ARGYROPOULOS G., JIN L., HOGGART C. J., MCKEIGUE P. M., KITTLES R. A. (2003): Skin pigmentation, biographical ancestry and admixture mapping. American Journal of Human Genetics, Volume 112, 387-399

SÖLKNER J., FRKONJA A., RAADSMA H. W., JONAS E., THALLER G., GOOTWINE E., SEROUSSI E., FUERST C., EGGER-DANNER C., GREDLER B. (2010): Estimation of Individual Levels of Admixture in Crossbred Populations from SNP Chip Data: Examples with Sheep and Cattle Populations. Interbull Bulletin No. 42. Riga, Latvia, May 31-June 4, 2010

THE HEREFORD CATTLE SOCIETY (2013): History.
http://www.herefordcattle.org/about-us/history/
visited on 2013-11-27

WIGGINTON J. E., CUTLER D. J., ABECASIS G. R. (2005): A Note on Exact Tests of Hardy-Weinberg Equilibrium. American Journal of Human Genetics, Volume 76, 883-887

VOZZI P. A., MARCONDES C. R., BEZERRA L. A. F., LÔBO R. B. (2007): Pedigree analyses in the Breeding Program for Nellore Cattle. Genetics and Molecular Research, Volume 29, 482-485

List of tables

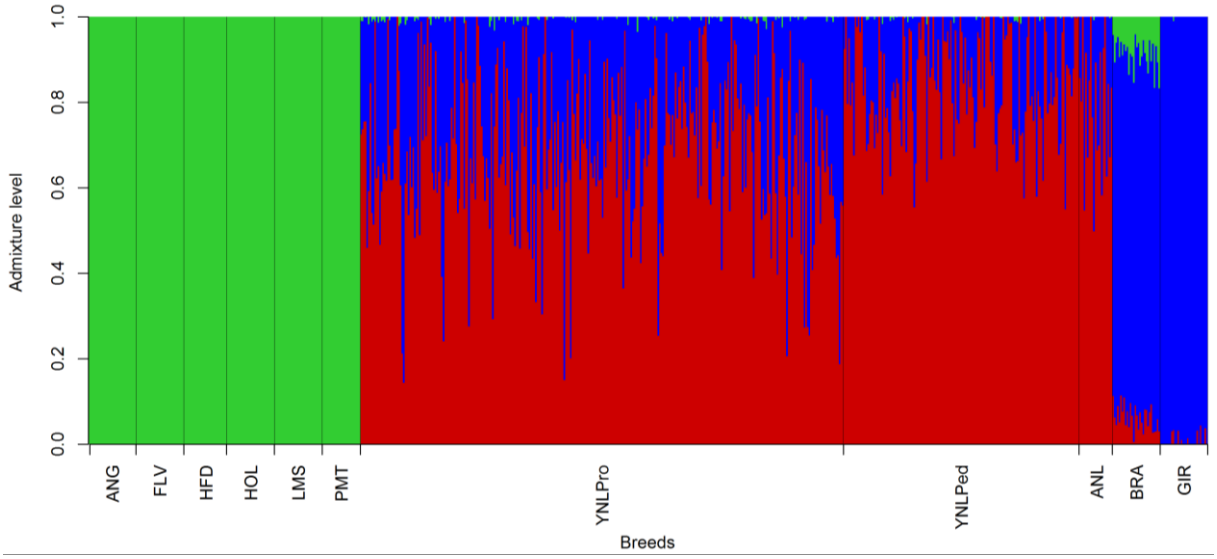
List of figures

List of additional files

Abbreviations

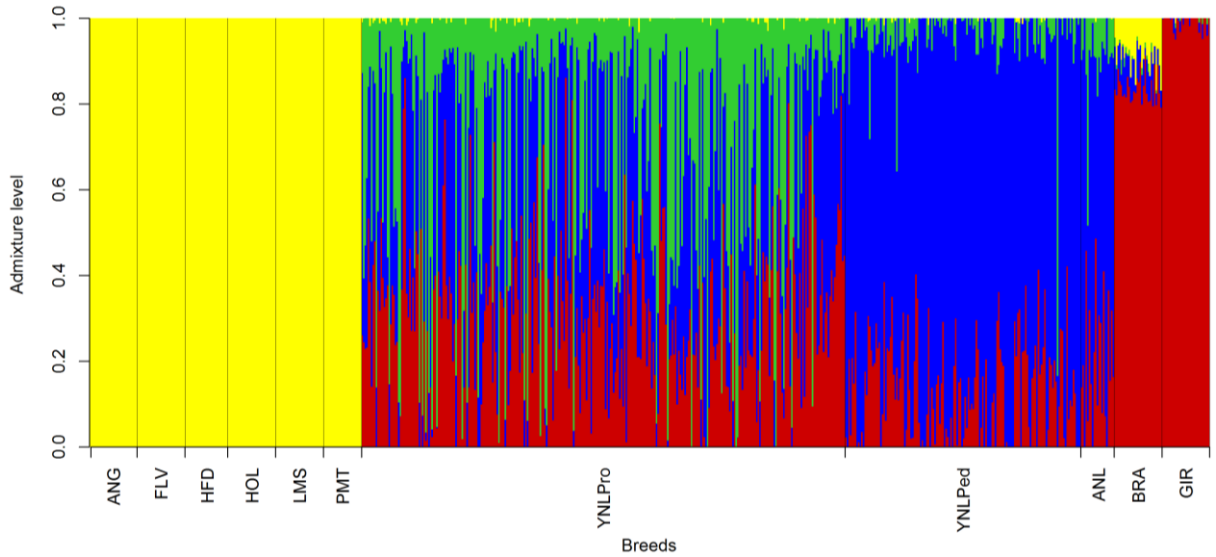
AIM(s) ..... ancestry informative marker(s)
ANG ..... Angus
ANL ..... ancestral Nelore
AS ..... autosomal
BRA ..... Brahman
B. p. ..... Bos primigenius
DNA ..... deoxyribonucleic acid
EM ..... expectation-maximization
FLV ..... Fleckvieh
GIR ..... Gir
Hap(s) ..... Haplotype(s)
HD ..... high-density
HFD ..... Hereford
HMM ..... hidden Markov model
HOL ..... Holstein
HWE ..... Hardy-Weinberg-equilibrium
IBD ..... identity-by-state
IBS ..... identity-by-state
ID ..... identification
Indiv. ..... individuals
K ..... number of populations/clusters
k ..... thousand
LD ..... low-density
LMS ..... Limousin
mt ..... mitochondrial
mtDNA ..... mitochondrial deoxyribonucleic acid
MAF ..... minor allele frequency
NEL ..... Nelore
PMT ..... Piedmontese
PO ..... purebred origin
POI ..... purebred of imported origin
s. a. ..... sine anno (undated)
SNP(s) ..... single nucleotide polymorphism(s)
QC ..... quality control
YNL ..... young Nelore
YNLped ..... young Nelore pedigree type
YNLpro ..... young Nelore production type

Appendix



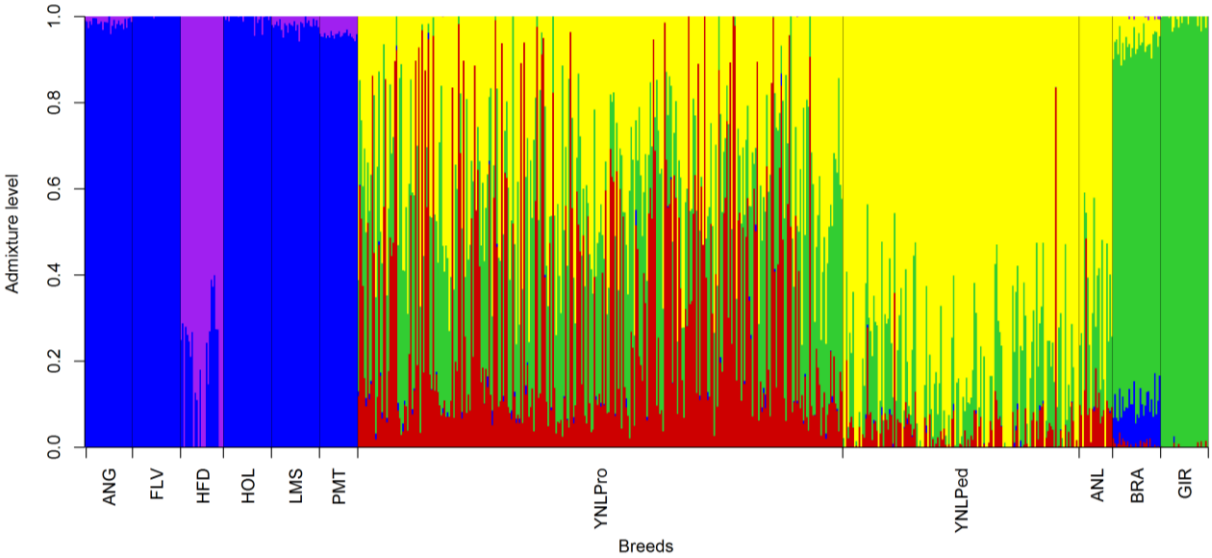
Appendix 1: Unsupervised genome-wide admixture plot with three ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The green colour represents the taurine and the red and dark blue the indicine ancestry.



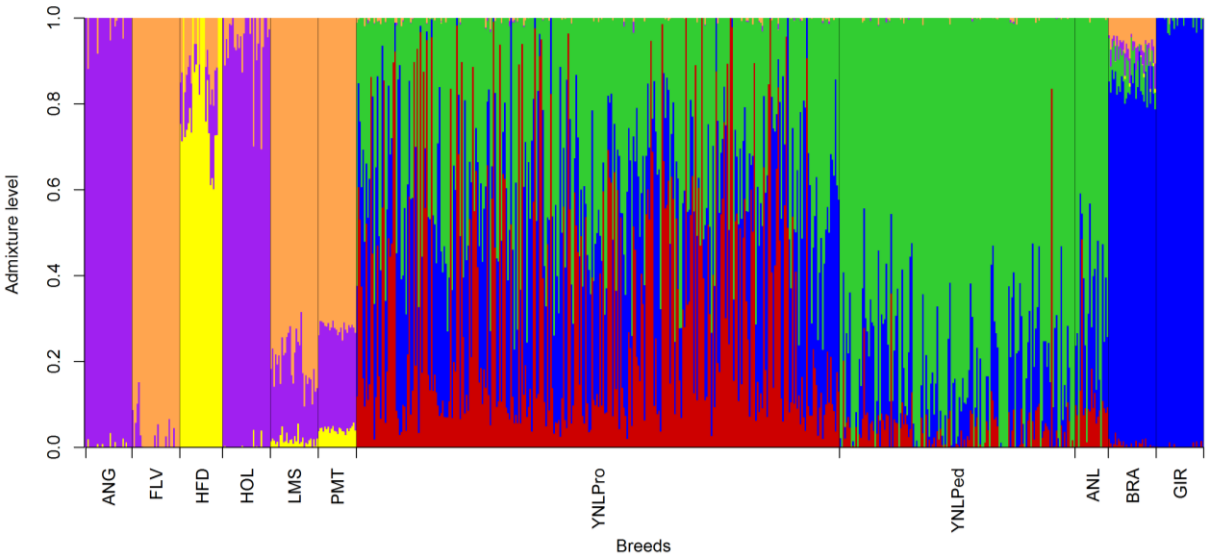
Appendix 2: Unsupervised genome-wide admixture plot with four ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The yellow colour represents the taurine and the green, red and dark blue the indicine ancestry.



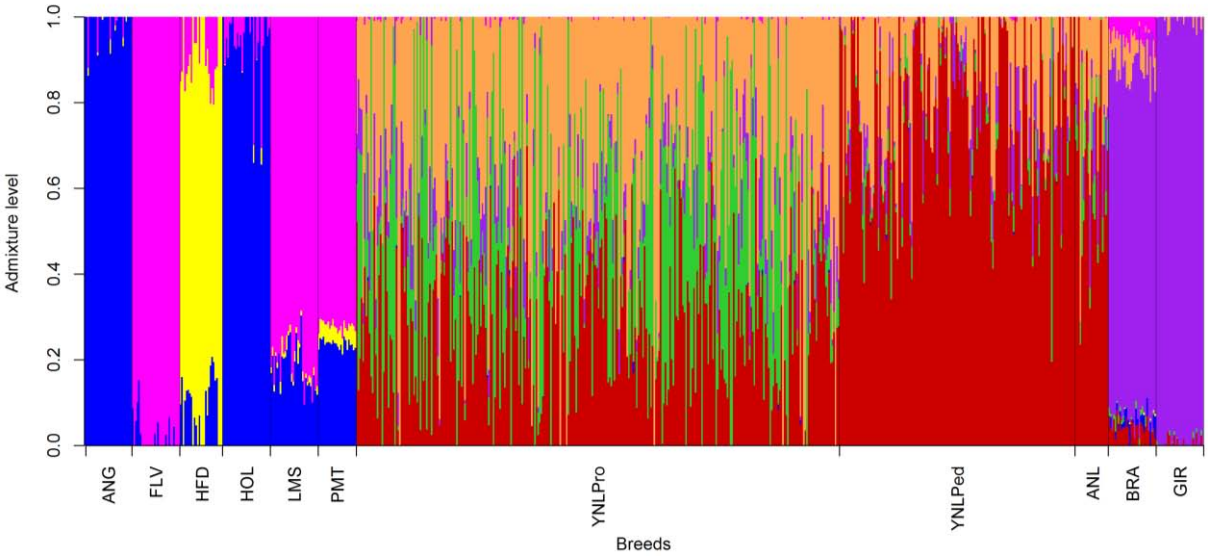
Appendix 3: Unsupervised genome-wide admixture plot with five ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The dark blue and purple colours represent the taurine and the green, yellow and red the indicine ancestry.



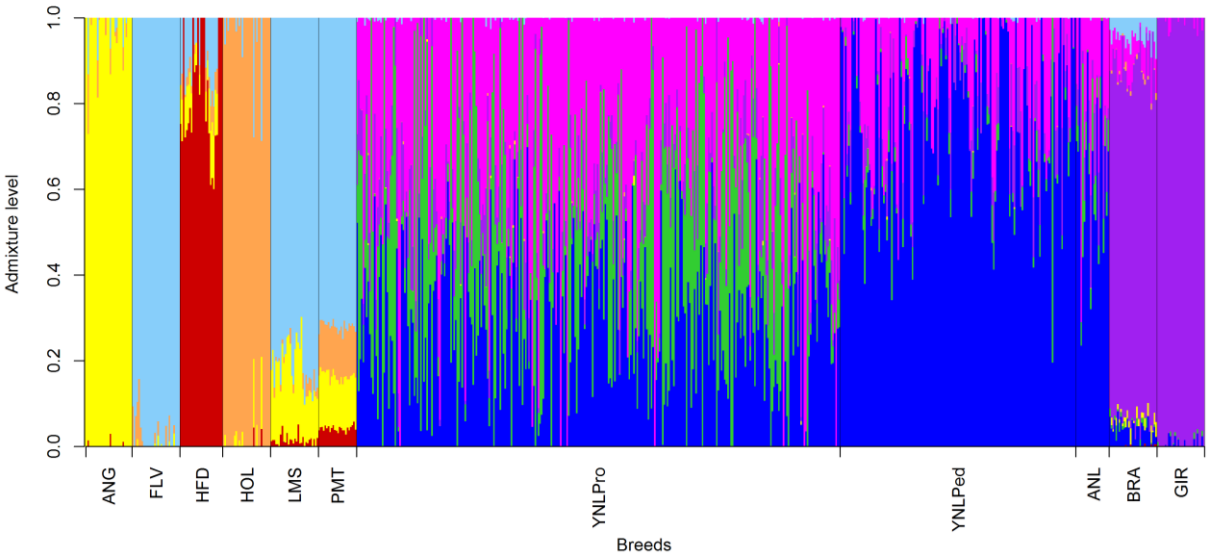
Appendix 4: Unsupervised genome-wide admixture plot with six ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The purple, orange and yellow colours represent the taurine and the green, red and dark blue the indicine ancestry.



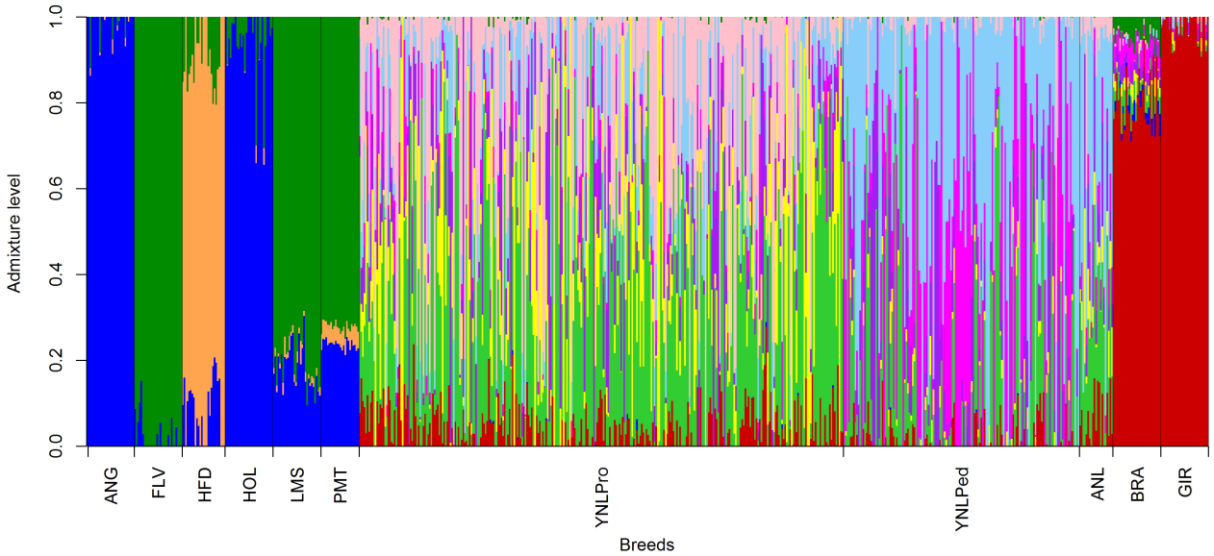
Appendix 5: Unsupervised genome-wide admixture plot with seven ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The dark blue, magenta and yellow colours represent the taurine and the green, red, orange and purple the indicine ancestry.



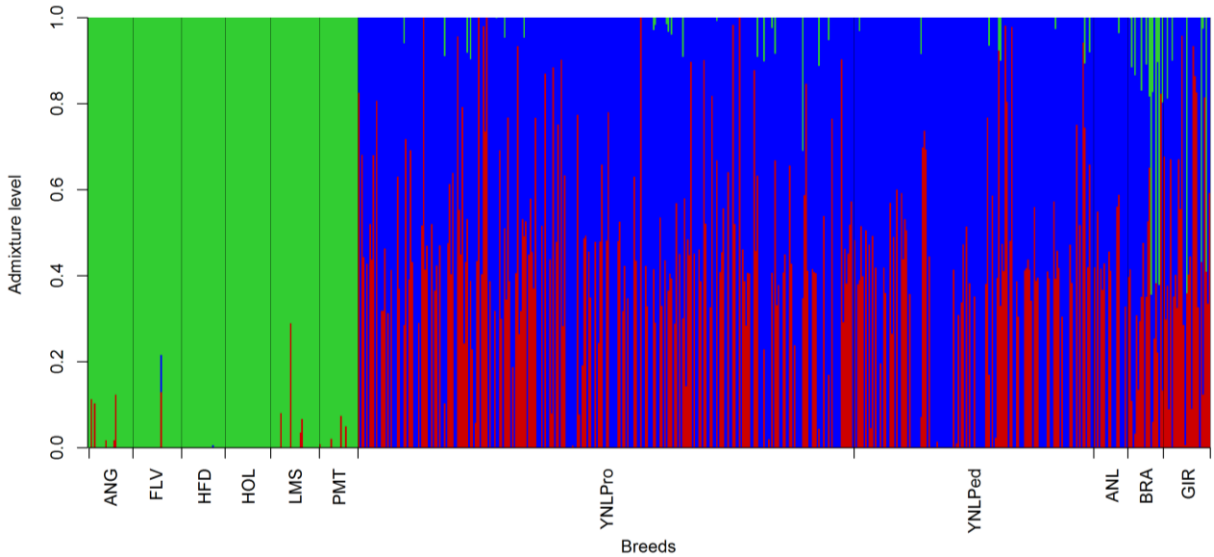
Appendix 6: Unsupervised genome-wide admixture plot with eight ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The yellow, light blue, red and orange colours represent the taurine and the green, magenta, dark blue and purple the indicine ancestry.



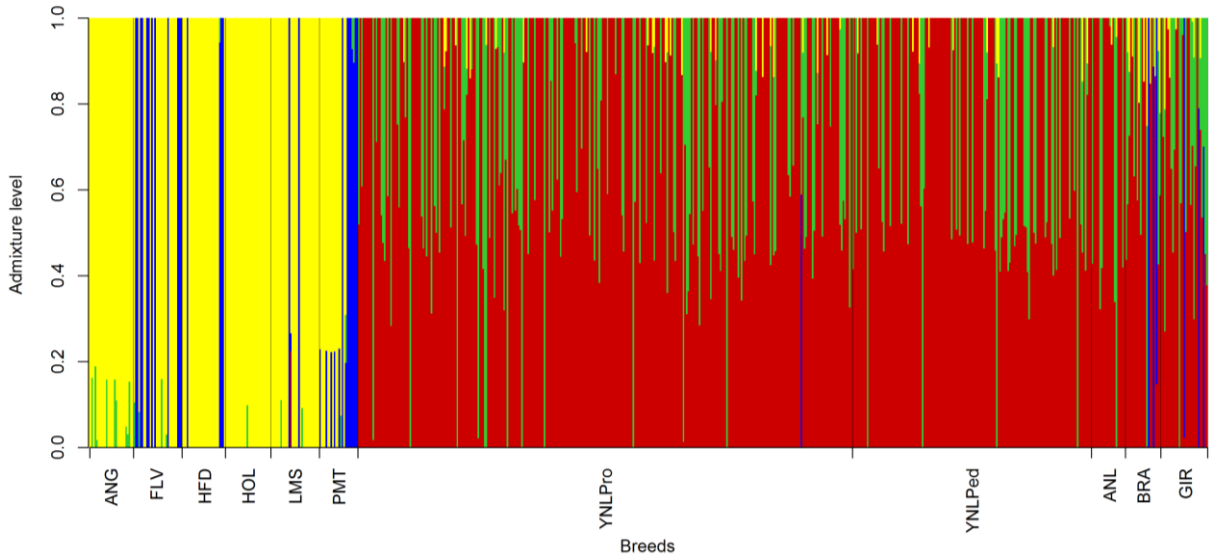
Appendix 7: Unsupervised genome-wide admixture plot with ten ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The dark blue, orange and dark green colours represent the taurine and the green, magenta, yellow, red, pink, light blue and purple the indicine ancestry.



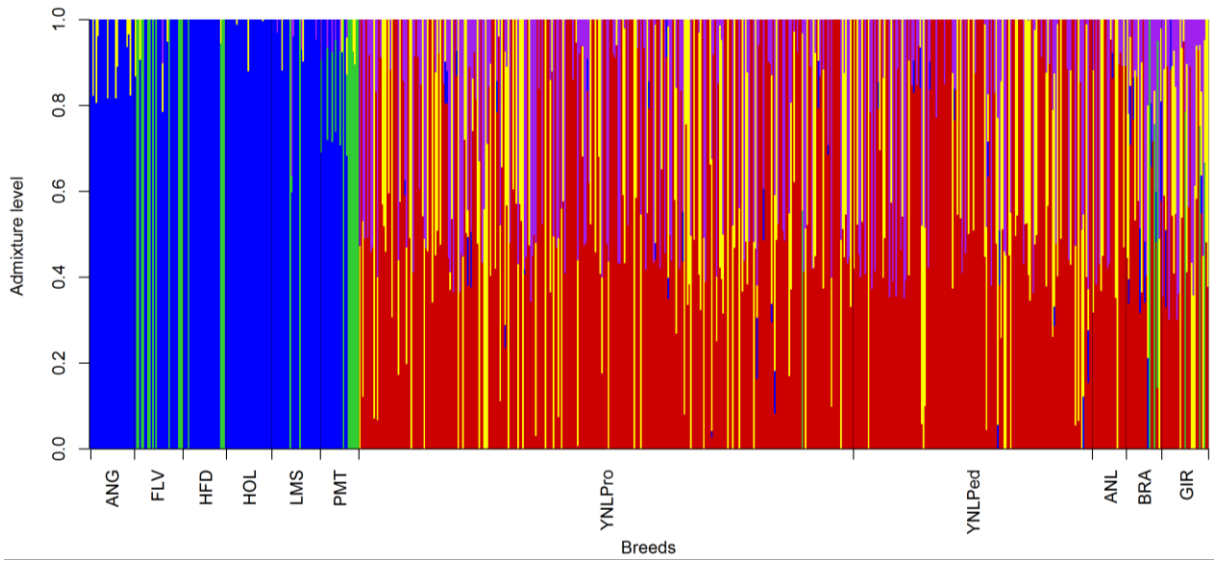
Appendix 8: Unsupervised Y-chromosomal admixture plot with three ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The green colour represents the taurine and red and dark blue the indicine ancestry.



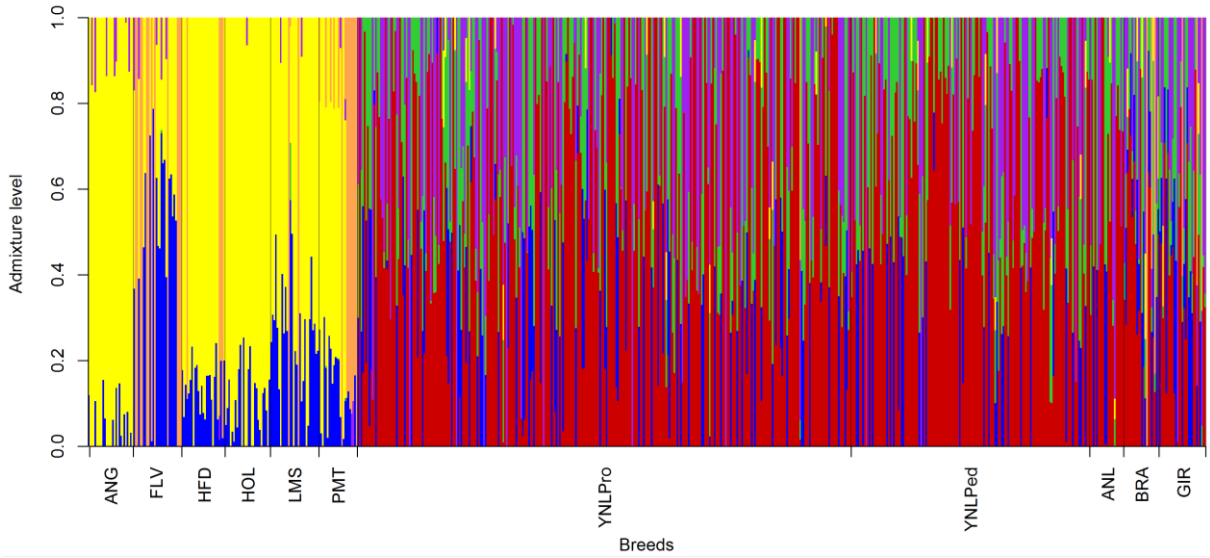
Appendix 9: Unsupervised Y-chromosomal admixture plot with four ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The blue and yellow colours represent the taurine and red and green the indicine ancestry.



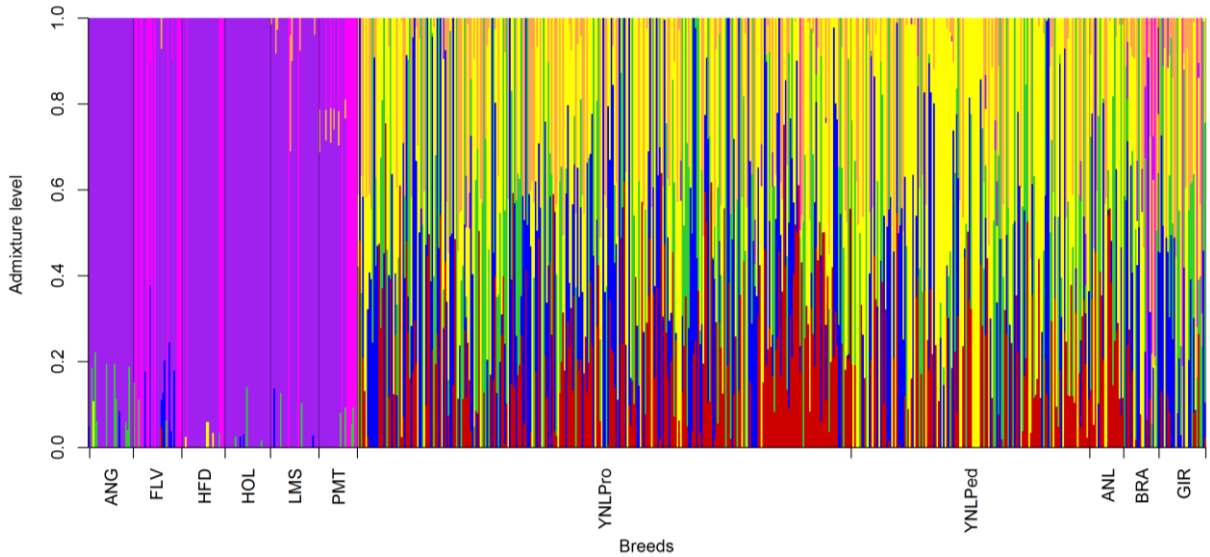
Appendix 10: Unsupervised Y-chromosomal admixture plot with five ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The dark blue and green colours represent the taurine and red, purple and yellow the indicine ancestry.



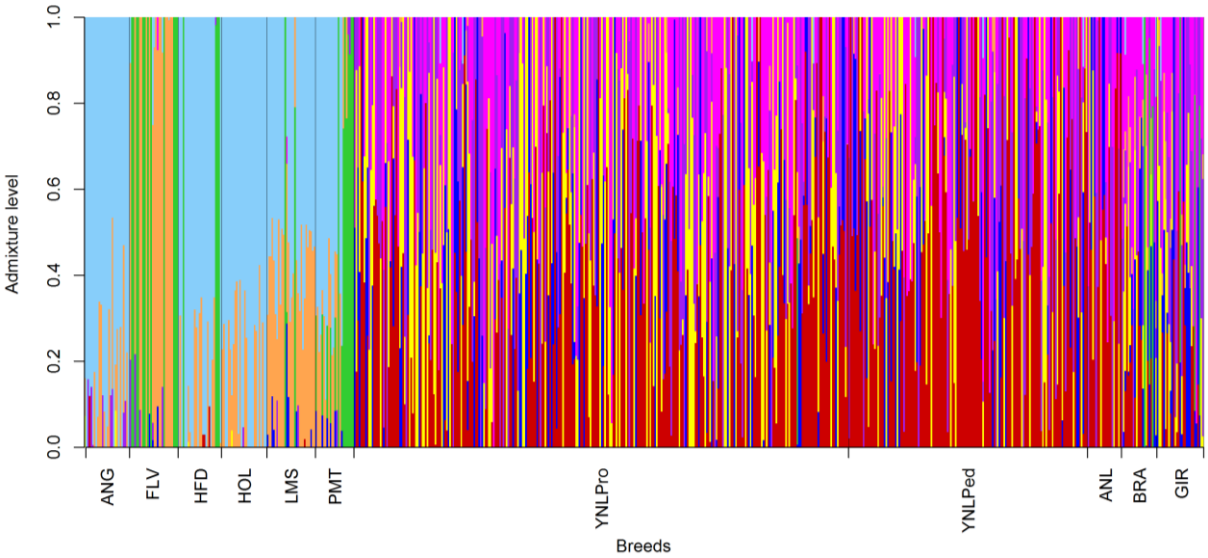
Appendix 11: Unsupervised Y-chromosomal admixture plot with six ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The yellow, dark blue and orange colours represent the taurine and red, green and purple the indicine ancestry.



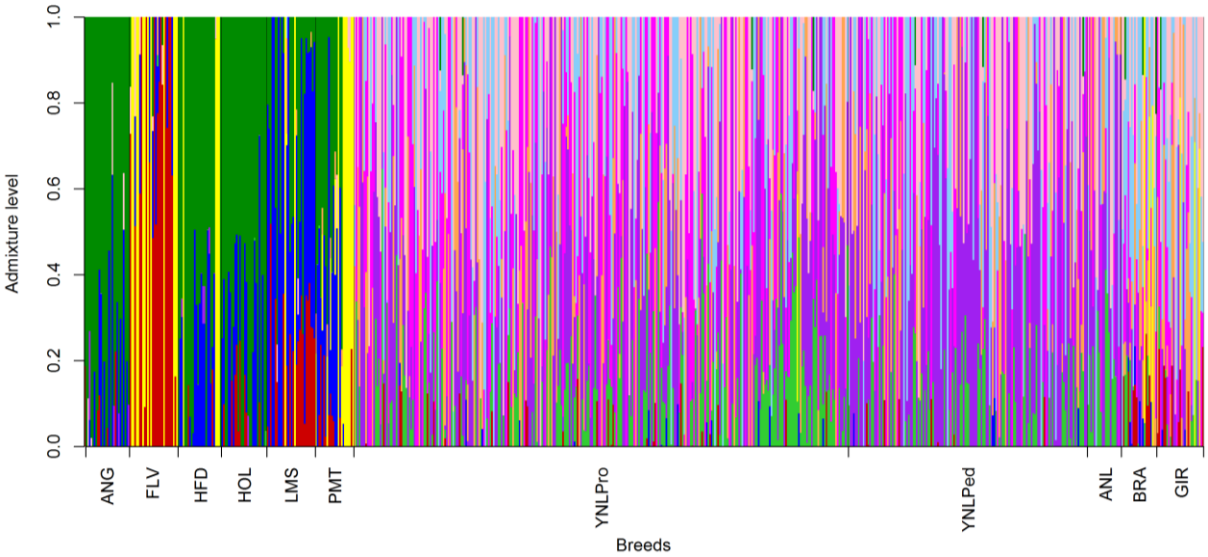
Appendix 12: Unsupervised Y-chromosomal admixture plot with seven ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The purple and magenta colours represent the taurine and red, green dark blue, yellow and orange the indicine ancestry.



Appendix 13: Unsupervised Y-chromosomal admixture plot with eight ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The orange, light blue and green colours represent the taurine and red, yellow, dark blue, magenta and purple the indicine ancestry.



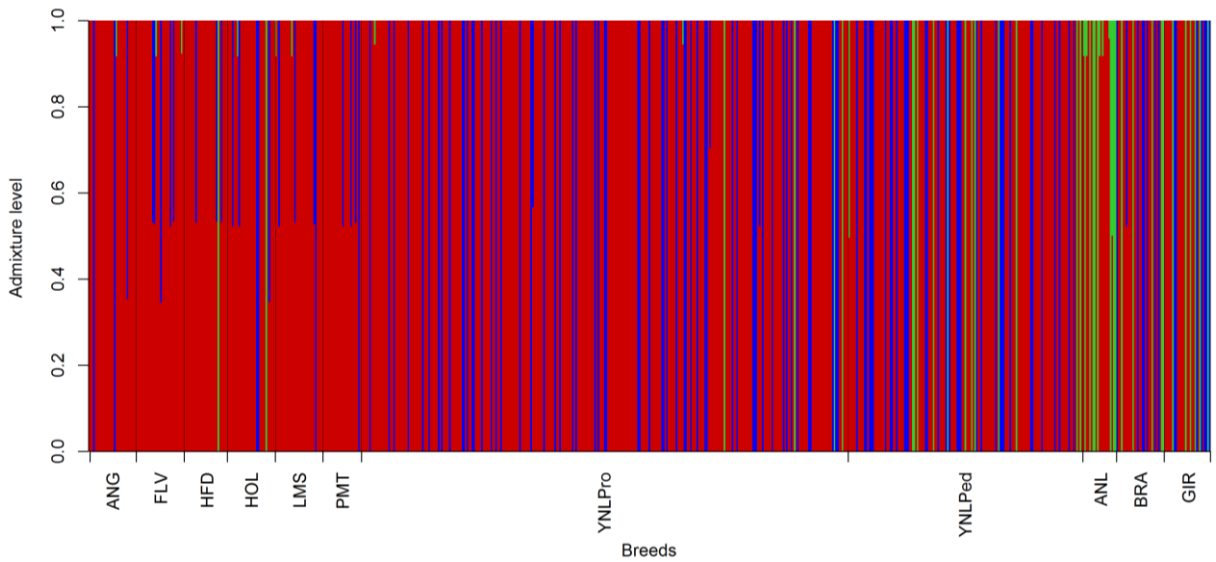
Appendix 14: Unsupervised Y-chromosomal admixture plot with ten ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. The orange, dark blue, yellow, red and dark green colours represent the taurine and green, light blue, pink, magenta and purple the indicine ancestry.

Haplotype identification (hap ID)	Haplotype (SNP progression)	Appearance	
		Total count	Percentage (%)
1	111111111	11	100,00
2	111110111	5	45,45
3	000011000	1	9,09
4	011111011	1	9,09
5	111101111	1	9,09

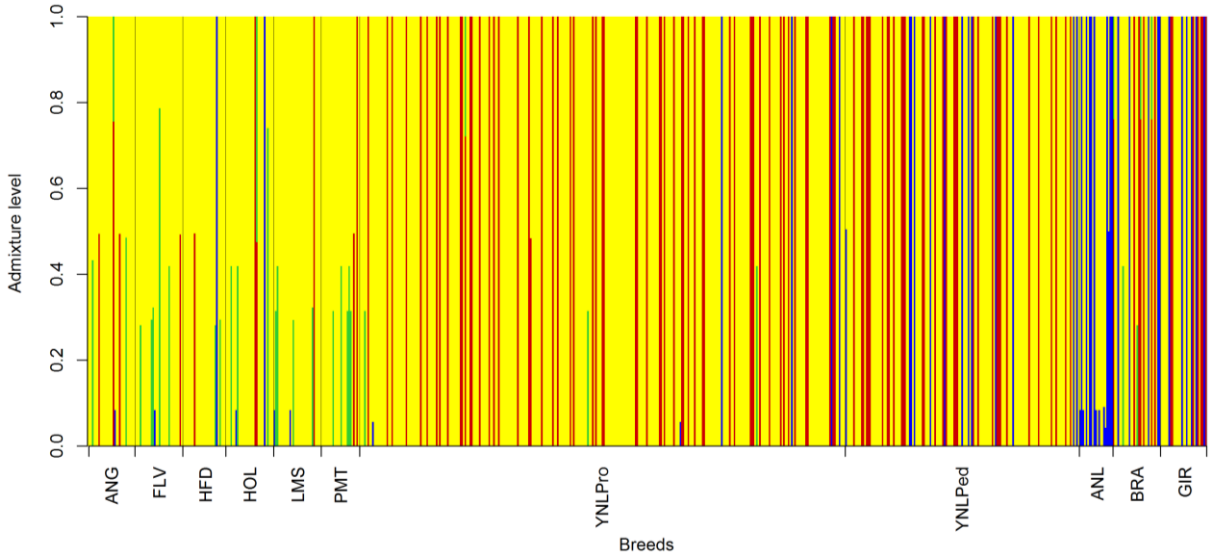
Appendix 15: Y-chromosomal haplotype identification numbers, valid for both calculation methods (FASTPHASE and manual)

This table is an extension to table 5. It shows the haplotype identification numbers with the appropriate estimated Y-chromosomal haplotypes and their appearance as total number and percentage of the breed groups showing the haplotype.



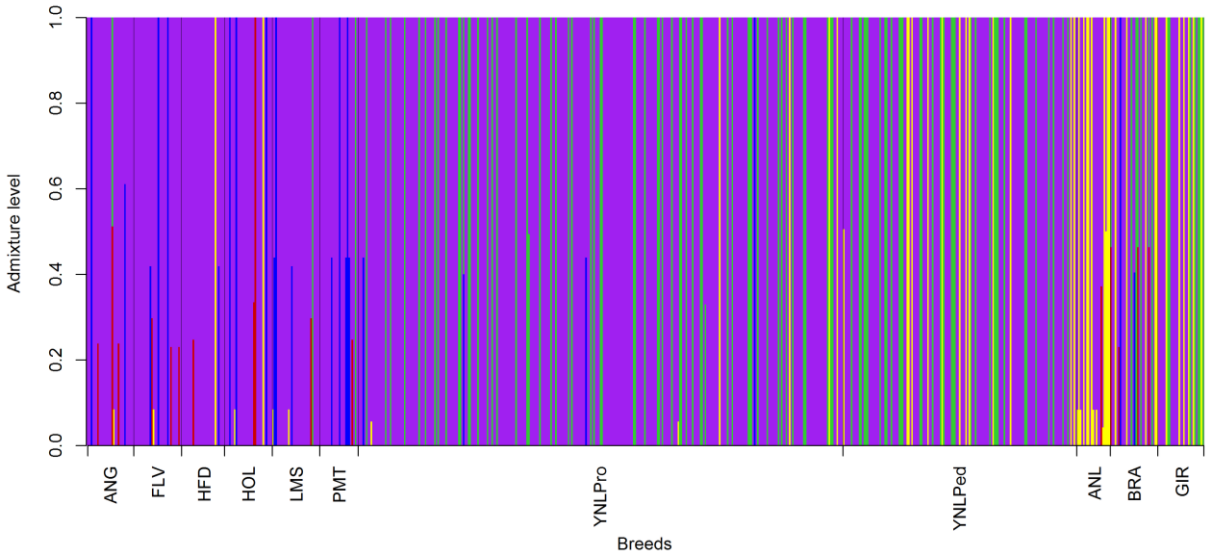
Appendix 16: Unsupervised mitochondrial admixture plot with three ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and the indicine breeds.



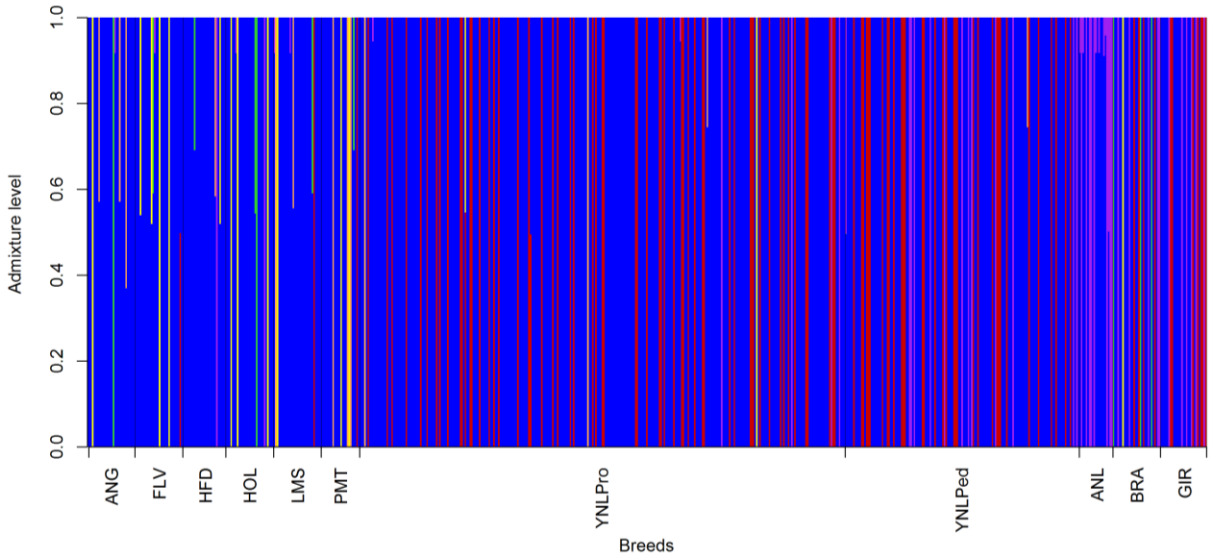
Appendix 17: Unsupervised mitochondrial admixture plot with four ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and the indicine breeds.



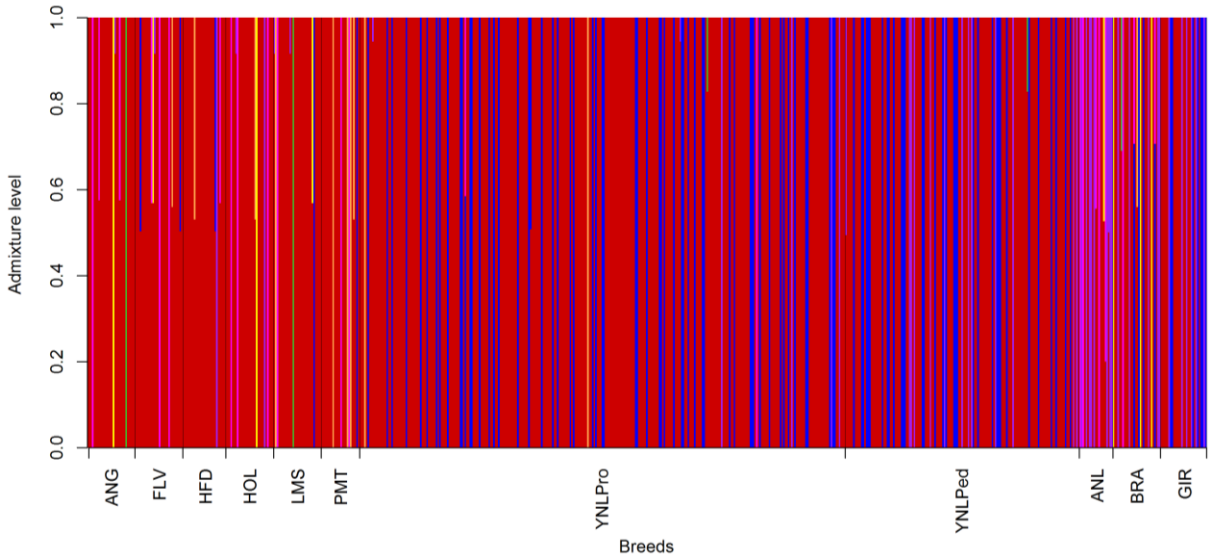
Appendix 18: Unsupervised mitochondrial admixture plot with five ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and the indicine breeds.



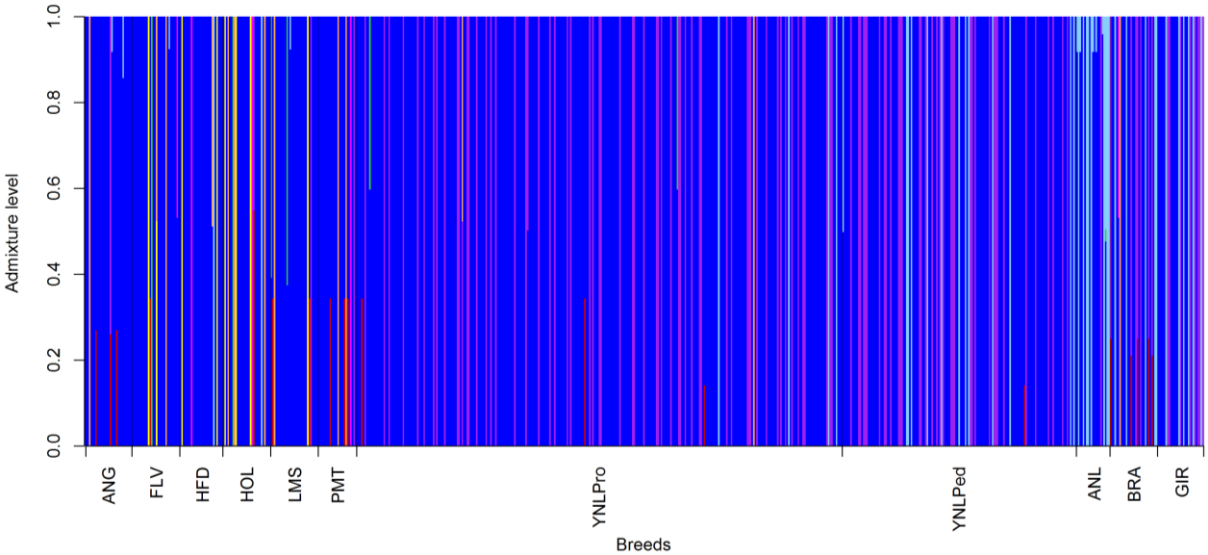
Appendix 19: Unsupervised mitochondrial admixture plot with six ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and the indicine breeds.



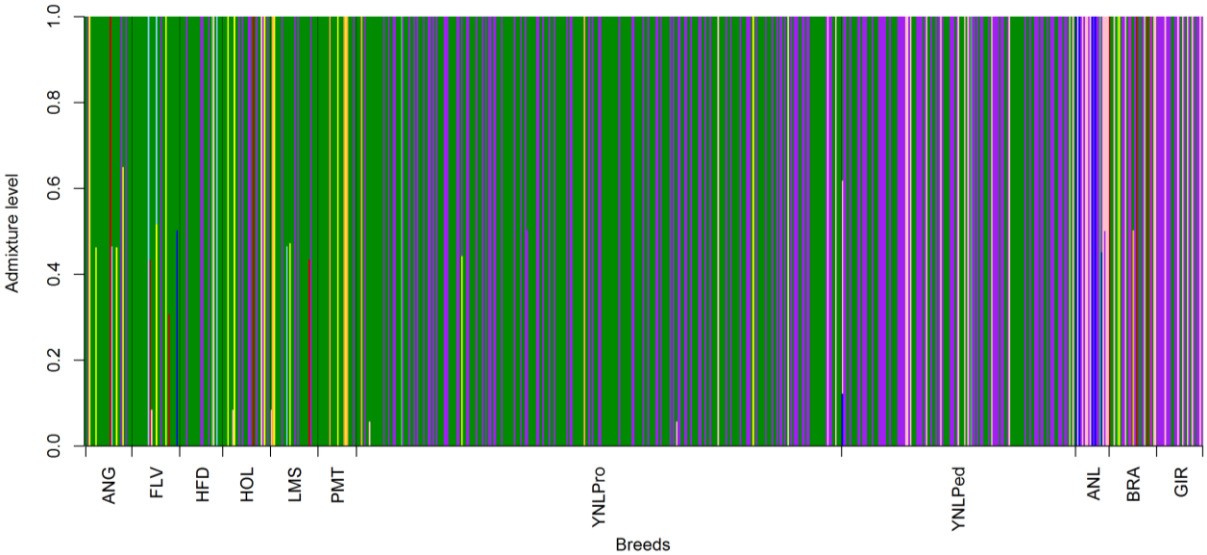
Appendix 20: Unsupervised mitochondrial admixture plot with seven ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and the indicine breeds.



Appendix 21: Unsupervised mitochondrial admixture plot with eight ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and the indicine breeds.



Appendix 22: Unsupervised mitochondrial admixture plot with ten ancestral populations

The x-axis shows the different breeds and groupings, and the y-axis shows the fraction of estimated admixture. Colours cannot clearly be assigned to the taurine and the indicine breeds.

Haplotype identification (hap ID)	Haplotype (SNP progression)	Appearance	
		Total count	Percentage (%)
1	1111111111111111111	11	100,00
2	0101111001101000110	7	63,64
3	1110111111111111111	5	45,45
4	1111111011111111111	2	18,18
5	1111101111111111111	1	9,09
6	1111111111110111111	1	9,09
7	0111111111111111111	2 (1)	18,18 (9,09)
8	1111111111111101111	1	9,09
9	1111111111111111101	1	9,09
10	1111111111111011111	1	9,09
11	1111111111111111011	1	9,09
12	1111110111111111111	1	9,09
13	1011111111111111111	1	9,09
14	1111011111111111111	1	9,09
15	1111111110111111111	1	9,09

Appendix 23: Mitochondrial haplotype identification numbers, valid for both calculation methods (FASTPHASE and manual)

This table is an extension to tables 6 and 7. It shows the haplotype identification numbers with the appropriate estimated mitochondrial haplotypes and their appearance as total number and percentage of the breed groups showing the haplotype (haplotype number 7: In brackets is the appearance in the manual calculation).