Genomic variation across European cattle: contribution of gene flow

Maulik Upadhyay

Thesis committee

Promotor:
Prof. Dr M.A.M. Groenen
Professor of Animal Breeding and Genomics
Wageningen University & Research

Co-promotors:
Dr R.P.M.A. Crooijmans
Assistant Professor, Animal Breeding and Genomics
Wageningen University & Research

Prof. Dr G. Andersson
Professor, Department of Animal Breeding and Genetics
Swedish University of Agricultural Sciences,Sweden

Dr S. Mikko
Assistant Professor, Department of Animal Breeding and Genetics
Swedish University of Agricultural Sciences,Sweden

Other members (assessment commitee)
Prof. Dr C.H.J. van Oers, Wageningen University & Research
Dr E. Jonas, Swedish University of Agricultural Sciences, Sweden
Prof. Dr M. Boichard, National Institute of Agricultural Research (INRA), France
Dr D.K. Aanen, Wageningen University & Research

The research presented in this doctoral thesis was conducted under the joint auspices of the Swedish University of Agricultural Sciences and the Graduate School Wageningen Institute of Animal Sciences of Wageningen University and is part of the Erasmus Mundus Joint Doctorate program “EGS-ABG”.

Genomic variation across European cattle: contribution of gene flow

Maulik Upadhyay

ACTA UNIVERSITATIS AGRICULTURAE SUECIAE
DOCTORAL THESIS No 2019:16

Thesis
submitted in fulfilment of the requirements for the joint degree of doctor between
Swedish University of Agricultural Sciences
by the authority of the Board of the Faculty of Veterinary Medicine and
Animal Science and
Wageningen University
by the authority of the Rector Magnificus, Prof. Dr. A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board of Wageningen University and
the Board of the Faculty of Veterinary Medicine and Animal Science at
the Swedish University of Agricultural Sciences
to be defended in public
on Tuesday 12 March, 2019
at 4.00 p.m.in the Aula of Wageningen University.

Maulik Upadhyay
Genomic variation across European cattle: contribution of gene flow
69 pages.

Joint PhD thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden
and Wageningen University, the Netherlands (2019)
With references, with summaries in English,Dutch and Swedish.

ISBN (print version) 978-91-7760-350-4
ISBN (electronic version) 978-91-7760-351-1
ISSN 1652-6880
ISBN 978-94-6343-420-1
DOI https://doi.org/10.18174/469250

Abstract

Upadhyay M.R. (2019). Genomic variation across European cattle: contribution of gene-flow. Joint Ph.D. thesis, between Swedish University of Agricultural Sciences, Sweden and Wageningen University and Research, the Netherlands.

European cattle display vast phenotypic diversity which can be attributed to genomic variation such as single nucleotide polymorphisms (SNPs) and structural variations (SVs).The distribution of these genomic variations in a population is heavily influenced by different population genomic forces. In this thesis,I used genome-wide SNPs to characterize genomic variation and admixture across different European cattle populations. Broadly,I show the difference in the domestication histories for north-western and southern European cattle.I argue that this difference can be attributed to a differential pattern of genomic admixture involving wild local aurochs and zebu cattle. Genomic admixture analysis revealed share ancestry between Balkan and Italian cattle (BAI) breeds, and zebu cattle. Moreover,I also show that southern European cattle breeds displayed shared ancestry with African taurine cattle.Using linked SNP based approaches, I inferred a common origin of the African taurine and zebu cattle ancestry in BAI cattle breeds. Furthermore,I also characterized the genomic diversity and structure in European cattle populations. I show that, on average, nucleotide diversity is higher in southern European cattle than western European (British and commercial) cattle. However, some of these southern European cattle breeds such as Romagnola and Maltese appeared to have undergone a recent bottleneck. On the other hand, Swedish native cattle breeds like Swedish Mountain cattle, despite recorded bottleneck in the past, still display significant genomic diversity. However, southern Swedish cattle breeds like Väneko and Ringamålako requires attention for conservation management as these breeds display lowest genetic diversity among all the Swedish cattle breeds. To understand the patterns of genomic variations comprehensively, I also characterized the structural variations (SVs) in the genome of European cattle. I inferred the influence of demographic changes in the distribution of SVs in the cattle genome. In addition, I also identified an SV CNV overlapping the *KIT* gene in English Longhorn cattle which has previously been associated with color-sidedness. Finally, using whole genome sequencing data, I identified various protein-coding genes and regulatory elements encompassing SVs which represents valuable resources for future studies aimed at finding the association between physiological processes and SVs in cattle.

Contents

List of publications

This thesis is based on the work contained in the following publications:

M. R. Upadhyay, W. Chen, J.A. Lenstra, C.R. Goderie, D.E. MacHugh, S.D. Park, D.A. Magee, D. Matassino, F. Ciani, H.J. Megens, J.A.M. van Arendonk, P. Ajmone-Marsan, V.A. Bâlteanu, S. Dunner, J.F. Garcia, C. Ginja, J. Kantanen, M.A.M. Groenen and R.P.M.A. Crooijmans, Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle (2017), Heredity, 118(2), 169–176.

M. R. Upadhyay,C. Bortoluzzi, M. Barbato , P. Ajmone-Marsan, L. Colli,J.A. Lenstra, C. Ginja, T. Sonstegard, M. Bosse , M.A.M. Groenen and R.P.M.A. Crooijmans, Deciphering the pattern of genetic diversity and admixture using Genome-wide SNPs in Southern European cattle (2019), Evolutionary Applications. doi:10.1111/eva.12770.

M. R. Upadhyay, S. Eriksson, S. Mikko, E. Strandberg, M.A.M. Groenen and R.P.M.A. Crooijmans, G. Andersson and A.M. Johansson, Genomic relatedness and diversity of Swedish native cattle breeds. Under review.

M. R. Upadhyay, V.H. da Silva, H.J. Megens, M.H.P.W. Visker, P. Ajmone-Marsan, V.A. Bâlteanu, S. Dunner, J.F. Garcia, C. Ginja, J. Kantanen, M.A.M. Groenen and R.P.M.A. Crooijmans, Distribution and Functionality of Copy Number Variation across European Cattle Populations (2017), Frontiers in Genetics, 8. doi: 10.3389/fgene.2017.00108.

M. R. Upadhyay, M.F.L. Derks, G. Andersson, M.A.M. Groenen and R.P.M.A. Crooijmans, Comparative evaluation of structural variations in taurine and indicine cattle using individual whole genome sequences.Under review.

Chapter 1

General introduction

1.1 Evolution of Bovinae sub-family

The mammalian sub-family Bovinae comprises several diverse species (Figure 1.1), some of which are culturally and economically very important throughout the world. The sub-family is further classified into the three major tribes: Tragelaphini, Boselaphini, and Bovini. While the first two tribes comprise of spiral, four-horned, large ox-like antelope, the Bovini tribe comprises almost all domestic and wild bovine species. The first split within the Bovini tribe occurred somewhere between 5-10 million years ago (MYA) when the subtribe Bubalina (*Bubalus* and *Syncerus spp.*) diverged from the subtribe Bovina (*Bos* and *Bison spp.*) (Hartl et al., 1988;L. Janecek et al., 1996;Ritz et al., 2000). These two subtribes have consistently shown to be forming dichotomous groups and no evidence of viable hybrid offspring has been reported from the mating involving these two subtribes (Hartl et al., 1988;L. Janecek et al., 1996;Ritz et al., 2000;Hassanin and Ropiquet, 2004;MacEachern et al., 2009;Dorian J. Garrick and Ruvinsky, 2014). Within the subtribe Bovina, divergence events involving the remaining species appeared to have occurred recently, in the last 2 MYA. As a result, the species within this sub-tribe can still produce viable offspring indicating incomplete speciation. In fact, the introgression events involving domestic cattle in the yak (*Bos Grunniens*) and wisent (*Bison Bonasus*) lineage have already been inferred using whole genome sequencing data (Soubrier et al., 2016;Medugorac et al., 2017).

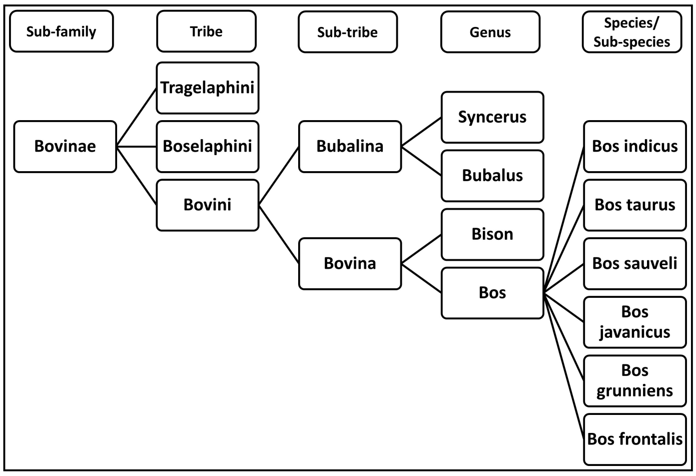

Figure 1.1: Taxonomic classification of sub-family Bovinae

Mitochondrial DNA (mtDNA) and genome-wide SNP based analyses have estimated the divergence date between the two most economically important Bos sub-species, *Bos indicus*, and *Bos taurus*, somewhere between 0.117 to 0.275 MYA (Loftus et al., 1994;Bradley et al., 1996;Gautier et al., 2016). The majority of the world cattle populations can be categorized under these two Bos sub-species with cross-breeding practices between these sub-species being widely prevalent in many parts of the world such as North America and Africa. The major morphological differences between these two sub-species are the presence of a thoracic hump, floppy rather than upright ears, and a large dewlap in *Bos indicus.* Both sub-species also display identical karyotypes with 29 autosomal pairs and a pair of sex chromosomes (X/Y). The Y-chromosome, however, is submetacentric in taurine and acrocentric in zebu, respectively (Kieffer and Cartwright, 1968;Jorge, 1974). Both these sub-species also display differences in physiological adaptation; while indicine cattle are very well adapted to harsh environmental conditions, most taurine cattle have been intensively selected for production related traits.

1.2 Initiation of domestication and early dispersion of cattle in Europe and Africa

The geographic origin and number of domestication events of cattle are arguably one of the most debated questions among bovine geneticists. Evidence based on archaeological and molecular data, points towards at least two centres of cattle domestication: domestication of *Bos primigenius namadicus* (Indian aurochs) in the Indus valley and domestication of *Bos primigenius primigenius* (European aurochs) in the Near East (Loftus et al., 1994;Bradley et al., 1996). The independent domestication of African aurochs has also been proposed (Grigson, 1991). However, a recent study has refuted this hypothesis (Decker et al., 2014).

The taurine lineage might have been domesticated first ∼10,000 years before present (YBP) in the Near East, most likely near the regions of the upper Euphrates basin and adjacent to the uppermost Tigrin basin(Helmer et al., 2005). Based on approximate Bayesian computation approach on mtDNA of ancient and modern cattle samples, it has been estimated that only about 80 female aurochs were initially domesticated (Bollongino et al., 2012;Scheu et al., 2015). Like other successful innovations, agriculture and animal husbandry also dispersed to other human populations, which can partly be attributed to migrations of early Neolithic farmers. Based on the archaeological and molecular evidence, it is possible to reconstruct the demographic events leading up to the dispersion of domestic cattle throughout Europe. Following domestication, it has been suggested (Martins et al., 2015;Hofmanová et al., 2016) that Neolithic farmers along with their livestock took at least two distinct routes (Figure 1.2) to reach mainland Europe: the Mediterranean Sea route and the Danube river route. Following these migrations, the earliest evidence of domestic cattle in Europe are reported in the form of cattle bones found at a Neolithic site in Greece which are dated ∼8,500 YBP (Conolly et al., 2012). Evidence also suggests that, via the Mediterranean Sea route, farming was introduced in Corsica, the southwest of France and in eastern Spain between ∼7,700–7,600 and ∼7,400–7,300 YBP, respectively (De Lagrán,2014). Via the Danube river route, domestic cattle reached central Europe and Northern Europe ∼ 7,500 YBP and ∼ 6,500 YBP respectively (Tresset, 2003). Indeed, studies involving Isotope analyses of organic residues of the major milk fatty acids preserved in archaeological pottery have indicated the use of milk products by European farmers from as early as ∼8,000 YBP (Salque et al., 2013).

During the early dispersion of domestic taurine, the wild population of ancestral European aurochs was still prevalent across mainland Europe. In fact, the last aurochs died at the beginning of the 17th Century (Kędzierska 1959; 1965 cited by van Vuure 2005). At its peak, the aurochs were distributed all over Eurasia; the distribution ranged from the Atlantic coast of Europe to the Pacific coast of China (Wright, 2013). Aurochs remains, however, have not yet been found

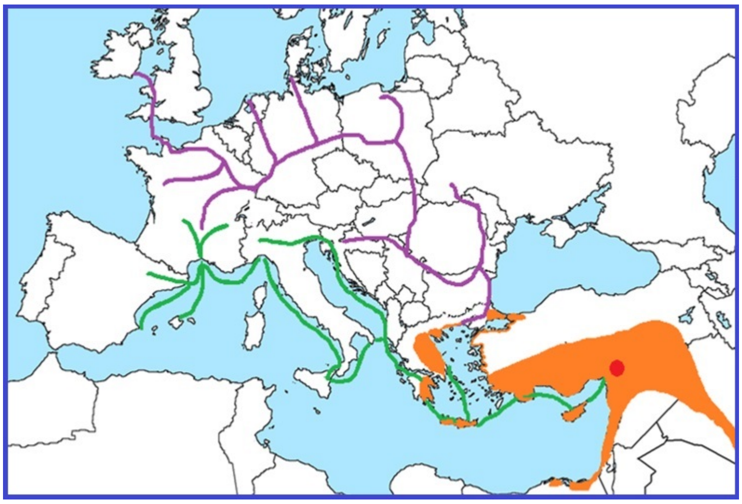

Figure 1.2: Representation of migration routes of Neolithic farmers. Red colour represents the center of domestication; the green line represents the Mediterranean Sea route, while the violet line represents the Danube river route. The figure is adapted from Felius et al.,(2014).(map outline from D-maps.com-https://d-maps.com/carte.php?num_car=2232&lang=en)

in Ireland, making West Iberia as the westernmost range of its distribution (Wright, 2013). Due to the long history of shared geography between aurochs and domestic cattle, the possibility of inter-crossing between them cannot be ruled out. In fact, several studies have investigated this hypothesis of post-domestication contact between domestic cattle and aurochs, some of which are discussed elsewhere in the thesis.

The theory of independent cattle domestication of the now-extinct African aurochs *(Bos primigenius africanus)* is highly disputed. The supporters of the theory often point out the prevalence of mitochondrial T1 haplotypes (Bradley et al., 1996;Edwards et al., 2004) in African cattle and osteological evidence found in the western Egyptian desert dating from ∼10,000 YBP (Wendorf et al., 1989) as proofs for backing their claim. Zooarchaeologists have cast their doubt on the origin of osteological evidence, and analysis of complete mtDNA sequences has shown that the T1 mtDNA haplotype is also found among Southwest Asian cattle, albeit at low frequency (Troy et al., 2001). Moreover, it has also been shown that the T1 haplogroup node is only one mutation (np 16113) away from the common mtDNA T3 haplotypes of European taurine, and hence, a near Eastern origin is very likely (Achilli et al., 2009). Uncontroversial dates for the arrival of domestic taurine has been estimated from ∼7500 YBP; based on archaeological evidence, it has been suggested that it appeared first in the region around the eastern Sahara (Gifford-Gonzalez and Hanotte, 2011). Archaeological and pictorial evidence also suggest that humpless *Bos taurus* were among the first cattle to appear on the African continent, which later got replaced or admixed by the arrival of zebu cattle. Two waves of zebu arrival in Africa have been proposed: the first wave of zebu arrival is associated with the development of Swahili-Arab civilization that started taking its root from the 7th century AD, while the second wave of zebu cattle expansion is associated with the rinderpest epidemics of the 19th century. Therefore, modern African cattle are mosaics of European taurine and zebu ancestry, though breeds like N'Dama and Mutarin have a unique genetic component which has been hypothesized as a legacy of African aurochs

(Decker et al., 2014).

1.3 European cattle diversity: Domestication to Modern times

1.3.1 Cattle genetic diversity from Neolithic to Roman era

The process of domestication initiated a symbiotic human-animal relationship in which humans started providing food and shelter to livestock in exchange for animal products and services such as fur, food, and protection. The process also allowed the transition of human society from being hunter-gatherers to settled farmers. Gradually, humans started the process of selective breeding of livestock to fulfill their specific needs. Many generations of this human-controlled livestock breeding and adaptation in their respective habitat greatly influenced behavioral and physiological traits of livestock.

The shift in livestock traits, from their ancestral forms to the more derived forms as we see today, occurred gradually as some of the traits that were desirable in wild cattle became a hindrance in domestic habitats. For instance, long horns in the ancestral bovids protected potential predators to some extent, while in the domestic setting long horns are redundant and undesirable as it made the task of handling livestock difficult. Therefore, short-horned cattle emerged somewhere in Mesopotamia in the early Bronze Age and they gradually replaced long-horned cattle in Europe from 5000 BP onward (Epstein, 1971). In the late Bronze age, short-horned cattle were widely distributed across central and Northern Europe while long-horned cattle were more common in many parts of the Mediterranean area as well in the region of today's Hungary (Bőkőnyi et al., 1974;Mason, 1984).

Though it is likely that breeding schemes might have existed in ancient times, the first detailed contemporary account of animal husbandry and knowledge-based selective breeding comes from ancient Roman literature. In his classic book "History of Animals" the Greek philosopher and scientist Aristotle gave accounts of large size cattle roaming about in rich pasturelands of Epirus (Balme, 1965). Skeletal remains recovered in Epirus also indicated that between 7^{th} and 8^{th} century BC, the region was inhabited by large size cattle with wither heights ranging from 115 to 135 cm (Kron, 2002). Large Roman cattle that had large horns and wither heights ranging from 120 to 140 cm, also inhabited the ancient Etruria region. However, soon after the fall of the Roman empire, large cattle also disappeared (Felius et al., 2014).

1.3.2 Cattle genetic diversity in Middle ages to the present times

During the Middle Ages, the small-sized cattle became prevalent in most parts of Europe. This has been attributed to various factors such as ease in management, poor availability of nutritious diet and castration of the large size bulls. Further, the number of livestock was greatly affected during the 14th century due to the Great Famine as well as Great Cattle Plague (Bőkőnyi et al., 1974;Kron, 2002;Campbell, 2009). Following the disastrous 14th century, cattle population gradually recovered owing to cultural and technological development (Felius et al., 2014). This was also the time, when a grey coat colored long-horned cattle of Podolian origin began replacing the local breeds in several parts of Eastern Europe (Bodo et al., 2004). Two hypotheses (Bőkőnyi et al., 1974;Ferdinando and Donato, 2001) have been put forward to explain the origin of Podolian cattle: 1) they arrived from the Podolian steppe of Ukraine where they were kept and bred until the 12th century, 2) they are descendants of large cattle which were kept during the Roman era.

During the 17th and 18th century, knowledge-based animal breeding started taking its root across north-western Europe. Literature related to animal husbandry and breeding became commonly available, partly due to improvement in literacy. Cattle migrations were also important aspects of animal husbandry practices during this time. Dutch cattle, due to their superiority in milk production, were exported to Germany, France, and Britain (Felius et al., 2014). However, still until the first industrial revolution of the late 18th century, the majority of the cattle diversity that existed among European cattle was due to adaptation and selection of local cattle breeds to the local circumstances rather than selection for certain traits which were desired by a broad range of consumers (Felius et al., 2014).

In Britain, the industrial revolution that began in the late 18th century provided impetus to the innovation in the field of agriculture and animal husbandry (Thomas, 2005). To meet the demands for animal-related products such as milk and beef in a growing urban population, the farmers began selecting animals based on their performance in desired production traits. For this selection process to work, the record-keeping of a herd as well as pedigree had to be of prime importance. Therefore, the concept of herdbook was introduced in animal husbandry practices. During the 1760's, the Englishman Robert Bakewell—one of the pioneers in Animal breeding—started improving cattle by selecting cows and bulls based on long horns, early growth, docility and other phenotypes (Stanley, 1995). Many English beef cattle breeds, such as Hereford and Aberdeen-Angus, were developed following the breeding success of English Longhorn cattle (Hall and Clutton-Brock, 1988). In fact, to keep the bloodline pure, dairy breeds such as Jersey and Guernsey were forbidden for cross-breeding and kept isolated from as early as 1789 (Hall and Clutton-Brock, 1988). Following these suits of success in record keeping and breeding objectives to develop systematic breeds, many western European countries adapted these techniques. In the Netherlands, the first herdbooks were established in the late 19th Century. Also, by performing cross-breeding between local cattle populations, breeds like Holstein Friesian (HF) and Meuse-Rhine-Yssel (MRY) were developed in the last two decades of the 19th Century (Felius et al., 2014).

1.3.3 Primitive and traditional cattle breeds of Europe

Although the process of domestication led to a transformation of traits that were seen frequently in the wild ancestors, the modern animal breeding practices (such as selection, herd isolation) accelerated this process of transformation of wild traits and led to the emergence of derived traits, such as early maturity, polledness, and docility in modern cattle that might had rarely been present in its ancestral form. However, many cattle breeds of Europe still display many ancestral features such as horn shape and size, sexual dimorphism, and aggressive behaviour (van Vuure, 2005). I refer to such cattle breeds as primitive in this thesis throughout, and in the following paragraphs, give an overview of the primitive cattle breeds of Europe (Table 1.1).

Primitive Cattle breeds of Balkan and Italian regions largely fall under the category of Podolic cattle breeds. As I mentioned in the previous section, the origin of a Podolic group of cattle and their diffusion to southern Europe is highly debated among bovine geneticists. This group of cattle along with Busha represents some of the most underdeveloped taurine populations. Apart from displaying common characteristics such as long horns and grey coat colour, these Podolic cattle breeds are some of the hardiest European taurines that can be raised under extensive management (Ferdinando and Donato, 2001;Felius et al., 2014;Di Lorenzo et al., 2018). These cattle are also adapted to a wide range of environments and display high disease resistance (Bartosiewicz, 2011). Some of the sampled Podolian cattle breeds include the following: Romanian grey, Boskarin, Chianina, Maremmana, Podolica, Romagnola, and Marchigiana (Table 1.1). Busha and Maltese are two other cattle breeds that we included in the Balkan and Italian group. Busha is distributed throughout the Balkan peninsula including Bulgaria and Greece. This group of cattle is characterized by small height, red to grey coat colour and small horns (Broxham et al., 2015). Busha cattle are hypothesized to have originated from small cattle of Medieval Europe. At present, several strains of Busha exist throughout the Balkan peninsula (Broxham et al., 2015). Maltese cattle are an ancient cattle breed of Malta; it is characterized by large body size and red coat colour. Although not much is known about its origin, it is hypothesized that the origin of Maltese traces back to the prehistoric era.

Iberian cattle breeds are the group of cattle displaying a large variety of coat colours and horn morphology. Many of the Iberian cattle display ancestral characteristics such as sexual dimorphism, coat colour, and horn morphology. It has been suggested that, mostly, Iberian cattle breeds have been developed in many different types with relatively little contribution from the outside (Felius et al., 2014). However, during the 1950's, sires from exotic cattle breeds, which displayed the same coat colour as some Iberian breeds, were used in "upgrading" some of the local Iberian cattle breeds such as Alentejana and Pajuna (Felius et al., 2014).

As described in the previous section, many modern British cattle breeds such as Hereford, Longhorn, Shorthorn, are were developed using modern animal breeding principles. However, several British, Scottish, and Irish cattle breeds, such as White Park, British White cattle, and Highland cattle, have been developed with minimal human interventions. Further, many British and Irish cattle breeds have individuals that display various ancestral traits (van Vuure, 2005). In this thesis, we also used genotyping data of commercial European cattle breeds such as Dutch cattle breeds and Jersey for comparative purposes. Some of the Dutch cattle breeds investigated in the study have undergone a drastic reduction in effective population size, for example, Dutch Friesian. HF and Jersey are among the most widespread cattle breeds in the world. HF originates from the Dutch provinces of North Holland and Friesland, while Jersey originates from Jersey Island. These cattle breeds are suitable for intensive farming which aims at maximizing the overall production and economic profit.

Apart from primitive and commercial cattle breeds, we also studied various Swedish and Dutch traditional cattle breeds. Note that the term “primitive” used in this context, refers to the selection of breeds based on their ancestral phenotypes. However, no such distinction is made while using the term “traditional”, the breed defined as traditional should be native to a particular region and maintained using traditional ways. Swedish traditional cattle used in the study includes various mountain breeds from the northern and western part of Sweden and some commercial cattle breeds of southern Sweden. A large phenotypic diversity exists among these Swedish cattle breeds. For example, white coat colour and polledness are predominant traits in Swedish mountain cattle breeds, while a large number of southern Swedish cattle breeds display red coat colour and a relatively high frequency of horned individuals. These breeds also display large temporal variation regarding the foundation of breed standards and herd books. For instance, Swedish mountain cattle was recognized as a breed way back in the 19th Century, while Vaneko was recognized as a breed in the late 20th Century.

Table 1.1: Table showing information of samples genotyped in this thesis. Sampling information: First column is Breed information where, in bracket, “C” stands for commercial breed, second column is Breed code, third and fourth columns displays information about country and region of origin for the breed respectively, fifth column displays number of samples collected per breed, sixth column display present conservation status which is obtained from Domestic Animal Diversity Information System (DAD-IS) on 06th November 2018. The last column displays the types of markers used in the present thesis; note that it does not necessarily indicates the number of individuals genotyped using each type of markers. Abbreviations: ALP-Alpine, BRI-British and Irish, NLD- Dutch, JE- Jersey, IBR- Iberian, BAI-Balkan and Italy, SAN-Scandinavian, WGS-whole genome sequencing data, 777K SNP-array-bovine 777K SNP High density array- (Illumina Inc.), 150K SNP array- bovine 150K Genomic Profiler High-Density SNP array (Illumina Inc. through GeneSeek©).Note that generally the conservation status—at risk—is allotted to the population with an effective population size less than 10,000.

Breed	**Code**	**Country of origin**	**Region/ species if not taurine**	**Sample size**	**Conservation status**	**Genetic markers used**
Brown Swiss (C)	BS	Switzerland	ALP	4	Not at risk	777K SNP array
Fleckvieh (C)	FL	Switzerland	ALP	4	Not at risk	777K SNP array
Chianina	CH	Italy	BAI	3	Not at risk	777K SNP array and WGS
Maremmana	MA	Italy	BAI	5	Not at risk	777K SNP array and WGS
Podolica	PO	Italy	BAI	1	Not at risk	777K SNP array and WGS

Maremmana x Pajuna	MP	NLD	BAI X IBR	1		777K SNP array
Busha	BU	Balkan region	BAI	6	At risk	777K SNP array and WGS
Romanian grey	RO	Romania	BAI	4	Not known	777K SNP array
Maltese	MT	Malta	BAI	4	At risk	777K SNP array and WGS
Boskarin	BK	Croatia	BAI	4	At risk	777K SNP array and WGS
Nellore	NE	Brazil	Bos indicus	4	Not at risk	777K SNP array
Aurochs	AU	Britain	Bos primigenius	1	Extinct	WGS
Angler (C)	AN	Germany	NLD	1	Not at risk	777K SNP array
Dutch Belted (C)	DB	The Netherlands	NLD	2	At risk	777K SNP array
Dutch Friesian (C)	DF	The Netherlands	NLD	4	At risk	777K SNP array
Groningen Whiteheaded (C)	GW	The Netherlands	NLD	5	Not at risk	777K SNP array
Holstein Friesian (C)	HF	The Netherlands	NLD	5	Not at risk	777K SNP array
MRY (C)	MR	The Netherlands	NLD	4	Not at risk	777K SNP array
English Longhorn	EL	England	BRI	4	At risk	777K SNP array
Galloway	GA	Scotland	BRI	5	At risk	777K SNP array
White Park	WP	England	BRI	3	At risk	777K SNP array
Highland	HL	Scotland	BRI	5	At risk	777K SNP array
Kerry Cattle	KC	Ireland	BRI	4	At risk	777K SNP array
Heck	HE	Germany	NLD	5		777K SNP array
Alentejana	AL	Portugal	IBR	2	Not at risk	777K SNP array
Arouquesa	AR	Portugal	IBR	3	At risk	777K SNP array
Cachena	CC	Portugal	IBR	3	Not at risk	777K SNP array
Caldela	CL	Portugal	IBR	1	At risk	777K SNP array
Mirandesa	MI	Portugal	IBR	2	At risk	777K SNP array

Berrenda en colorado	BC	Spain	IBR	3	At risk	777K SNP array
Berrenda en negro	BN	Spain	IBR	3	At risk	777K SNP array
Cardena	CA	Spain	IBR	5	At risk	777K SNP array
Lidia	LI	Spain	IBR	3	Not at risk	777K SNP array
Limia	LM	Spain	IBR	4	At risk	777K SNP array
Maronesa	ME	Spain	IBR	6	At risk	777K SNP array and WGS
Pajuna	PA	Spain	IBR	6	At risk	777K SNP array and WGS
Sayaguesa	SA	Spain	IBR	5	At risk	777K SNP array and WGS
Tudanca	TU	Spain	IBR	2	Not at risk	777K SNP array and WGS
Jersey (C)	JE	Jersey Island	Jersey	4	Not at risk	777K SNP array
Swedish Mountain Cattle	SMC	Sweden	SCAN	23	At risk	150K SNP array
Fjallnara cattle	FNC	Sweden	SCAN	16	At risk	150K SNP array
Swedish Polled cattle	SPC	Sweden	SCAN	3	At risk	150K SNP array
Bohus Polled	BPC	Sweden	SCAN	6	At risk	150K SNP array

1.4 Present genetic diversity status of primitive and traditional cattle

Advancement in quantitative genetics theory and techniques related to biotechnology, after the end of the second world war, led to a rapid increase in beef and dairy production in Europe. However, this rapid increase in production was brought about by using only a handful of north-western European (NWE) cattle breeds. Moreover, the effective population size for some of these NWE cattle breeds reduced to less than fifty (Gautier et al., 2007) because of intensive selection and repetitive usage of germplasm from proven sires. At the same time, industrial demand in some of the countries, where the development in animal husbandry was still in its nascent stage, led to the import of germplasm from these productive NWE breeds. As a result, the number of local cattle breeds with a long history of adaptation in their respective environments reduced drastically (Medugorac et al., 2009). Moreover, in some European regions, where livestock was mainly used for draft purposes, the mechanization of agriculture led to a decline in effective population size in those cattle breeds. For instance, the effective population size for Andalusian black cattle breeds reduced steeply in the last decade of the twentieth century as a result of agriculture mechanization (Felius et al., 2014). Similarly,the effective population size of Romanian grey cattle dropped from about ∼0.2 million at the end of 19^{th} century to just ∼500 animals in the beginning of 21^{st} century.

According to the FAO report (FAO, 2015), cattle are among the mammalian species with the highest number of breeds at risk. In fact, the report also provides some other worrisome statistics. For instance, of the total 1,408 global cattle breeds, the diversity status of more than 750 breeds remains unknown. Further, of the total 640 global cattle breeds with known "risk status", 171 breeds have been classified under "at risk" category while 184 breeds are already extinct (FAO, 2015). Therefore, using genetic markers to estimate the status of genetic diversity of traditional cattle breeds is an import step towards breed conservation. One of the questions that might arise from this chapter is: what is the need of conserving primitive cattle breeds? Based on the literature that I surveyed, I give the following three broad arguments to underscore the importance of primitive cattle breeds:

1). Long adaptation history in their respective environments: primitive cattle breeds represent cattle populations that have a long history of adaptation in their respective indigenous environment. For instance, Italian Podolic cattle breeds such as Chianina and Maremmana are well adapted to the harsh environment, and they also display a good growth ability and resistance against parasitic diseases (Sargentini et al., 2010).

2). The abundance of rare alleles: It has been postulated that, because some Balkan cattle breeds such as Busha have large effective population sizes for a very long time, they might have conserved an abundance of rare alleles, some of which are lost alleles in production cattle breeds (Medugorac et al., 2009). Therefore, diversity in traditional cattle breeds represents gene pool which may play an important role to fulfil the needs of future generations.

3). Heritage values and unique products: In many instances, primitive cattle breeds are linked to socio-cultural values of local tradition. Moreover, the products obtained from local breeds might have some additional value that could distinguish them from commercial breeds.

1.5 Measures of genetic variation/diversity

Genetic variation can be measured as the differences in two DNA sequences sampled randomly from a panmictic population or any other well-defined population. Therefore, it can refer to variation within a population or a genome. Further, variation within an individual genome can also capture variation in a population as the haplotypes of an individual are a sample of the haplotypes segregating in a population. Two important sources of variation are de novo mutations and recombination. Genetic variation arises depending on the consequences of mutations in a genome. For instance, sometimes mutations can lead to single base pair substitution which is called single nucleotide polymorphism (SNP) when the frequency in the population has reached a minor threshold typically more than 1%. Recombination generally does not create any de novo mutation, but rather it creates new combinations of alleles by reshuffling the genetic materials between homologues chromosomes during meiosis.

Heterozygosity is among the first parameter that often has been used by researchers to represent genetic variation in a natural population (Beja-Pereira et al., 2003;Cymbron et al., 2005). The term heterozygosity refers to the state of having two distinct alleles at a locus. The overall heterozygosity in a genome gives insight on genetic structure and demographic history of a population. For instance, reduced heterozygosity can indicate low genetic variability which can be the result of selection or a demographic process that severely reduced the population size (i.e., Bottleneck). As selection only acts on specific genetic segments, which depends on its contribution to the overall fitness of the individuals, its effect on heterozygosity would be local compared to genetic drift which would affect the entire genome.

Another parameter, which not only measures heterozygosity in a population but also provides additional information about the factors that generated it, is called runs of homozygosity (ROH). ROH are segments of identical haplotypes in an individual that are identical by descent (IBD). Inbreeding and selection are the most common causes that result in ROH within a genome. Another cause being non-random association between alleles, the phenomenon also known as linkage disequilibrium (LD). More often, ROH due to ancestral LD are much smaller in size compared to recent inbreeding as in the latter case, haplotypes have not had enough time to break-down due to recombination. Therefore, varying length of ROH provide insight into the level of inbreeding and demographic history of a population (Bosse et al., 2012).

1.6 Gene flow and genetic variation/diversity

Typically, the term migration in genetics refers to "gene flow" which is defined as the movement of alleles from one population to another.It is also an important factor affecting genetic variation.It reduces the genetic variation between previously isolated populations. This reduction in variability, however, depends on the rate and duration of gene flow. At the genomic level, gene flow followed by recombination makes the chromosomes of admixed populations mosaics of chromosomal blocks from different admixing populations (Lawson et al., 2012). Further, other population genetic forces such as selection or/and drift would determine the dispersal of introgressed segments in a population (Bosse et al., 2014).

As the events involving gene flow usually reduce allele frequency differences between the populations, several statistical approaches have been developed to classify individuals into "K" different clusters based on genetic similarity. The maximum-likelihood based approaches as implemented in the software-ADMIXTURE (Alexander et al., 2009) and STRUCTURE (Pritchard et al., 2000)-estimates underlying global admixture coefficients for each of the user-defined ancestral populations. These methods assume independence of markers. Therefore, it is important to filter SNPs based on a threshold of squared Pearson coefficient of correlation (r^2) estimate of

LD before performing the analysis.

Another way of estimating admixture events using independent SNP markers is by measuring the shared drift between populations (Patterson et al., 2012). These measures are the extension of Wright's F statistics which measures the population differentiation based on allele frequencies. Shared-drift based measures assume the null hypothesis that a tree-like fashion relates populations under investigation, i.e., they evolved independently after divergence (Figure 1.3A). Therefore, the branch lengths in the population phylogeny correspond to the amount of drift that has occurred after the divergence. The alternative model, in addition to branches, extends the phylogeny by allowing edges that represent migration events (Figure 1.3B and 1.3C). In other words, in the case of gene flow events, there will be an allele frequency correlation between source and admixed populations. However, the significant drift in either/or both admixing and source population after admixture can distort the correlation in allele frequencies. The shared-drift based measures calculated based on allele frequencies of three and four population are known as f3 and f4 tests, respectively. The algorithm implemented in the tool Treemix (Pickrell and Pritchard, 2012) is another interesting approach which assumes independence in allele frequencies between populations, and by modelling their relationships as bifurcating tree, it infers the migration events among sets of populations.

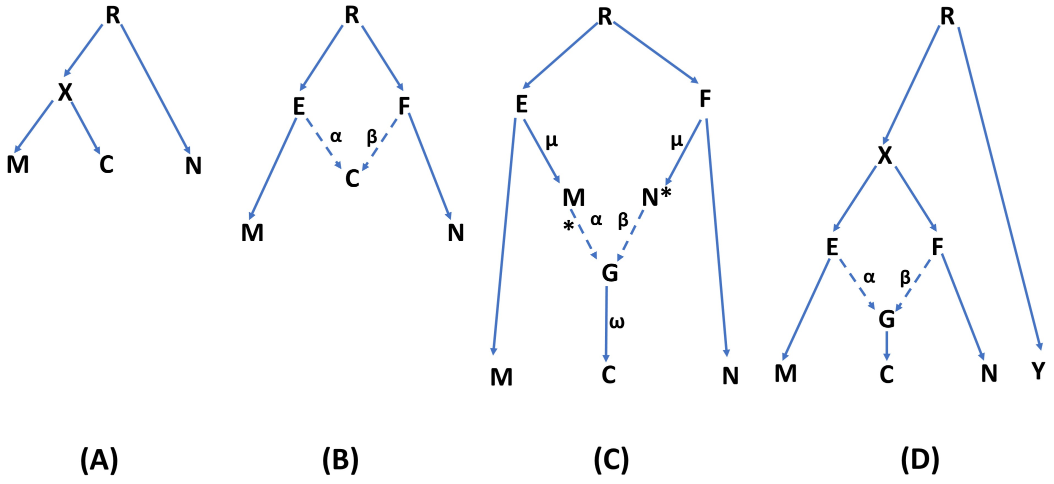

Figure 1.3: .Different demographic models: (A). Present day population M, N, and C evolved independently without any significant admixture. (B). Population C formed as a result of interbreeding between the population E and F that are ancestors of modern populations M and N respectively. (C). After receiving gene flow from population M and population N, population G undergoes a significant genetic drift. Note that "α" and "β" represents a proportion of gene flow, while "ω" represents the amount of genetic drift. The figure is adapted from Patterson et al., (2012).

Almost all the previously defined algorithms use SNPs individually and assume independence between successive markers. However, the advent of cost-effective high throughput technologies has resulted in array-based approaches which can genotype thousands or hundreds of thousands of closely positioned markers, and the analysis of such data can, sometimes, violate this assumption of independence. Haplotype-based analyses can harness the information from such closely linked data, leading to improvement in the inference of population structure. One such algorithm is implemented in a suite of a program called fineStructure (Lawson et al., 2012). The algorithm reconstructs each “recipient” haplotype as a mosaic of haplotypic blocks of all the other “donor” haplotypes in the dataset using a Hidden Markov Model method as introduced by Li and Stephen (Li and Stephens, 2003). Essentially, this reconstruction results in the co-ancestry matrix wherein each value corresponds to the shared ancestry between any two haplotypes in the dataset. Later, the co-ancestry matrix is used by fineStructure to assign individuals into population using a Markov chain Monte Carlo (MCMC) algorithm.

1.7 Brief review on studies of genetic admixture in primitive European cattle

Because primitive cattle breeds show many ancestral phenotypes, the hypothesis of post-domestication contact between aurochs and the ancestors of these cattle has been proposed (Achilli et al., 2008;Bonfiglio et al., 2010). Many studies, using uniparental markers such as mtDNA and Y-chromosome haplotypes, investigated the hypothesis of post-domestication gene flow in European cattle (Götherström et al., 2005;Achilli et al., 2008;Bollongino et al., 2008;Bollongino et al., 2012). For instance, gene flow between Italian domestic cattle and Italian aurochs has been proposed based on the observation that Italian aurochs also carried mitochondrial T3 haplogroups which is the most common haplogroup among European taurine (Beja-Pereira et al., 2006). Additionally, Italian cattle breeds such as Romagnola and Chianina also displayed low frequency of several novel mtDNA haplogroups such as Q and R which also has been proposed as a legacy of local Italian aurochs (Achilli et al., 2008;Bonfiglio et al., 2010). On the other hand, Götherström et al., (2005) observed a high frequency of Y1 haplogroups in (Y-chromosomal markers) aurochs samples retrieved from north-western Europe and, since Y1 haplogroup is also the most common haplogroup among north-western European domestic cattle, they proposed that gene flow between aurochs and domestic cattle of north-western Europe might have occurred after domestication. Later, (Bollongino et al., 2008) refuted this hypothesis by showing that there was no difference in the frequencies of Y1 and Y2 haplogroups among the aurochs samples retrieved from north-western Europe. It should be noted that until now mitochondrial haplogroup of the British aurochs, i.e., P-haplogroup, has only been found in one to two individuals of modern European taurine (Achilli et al., 2008).

Previous studies have also hypothesized that gene flow has occurred between south-eastern European cattle and non-European cattle (zebu and African taurine) (Beja-Pereira et al., 2003;Cymbron et al., 2005;Ginja et al., 2010;Decker et al., 2014). For instance, studies using genome-wide SNPs and microsatellite markers have shown that African taurine ancestry is a common feature of Iberian cattle (Decker et al., 2014). On the other hand, a gradient of zebu ancestry from southern Europe to Northern Europe cattle also had been proposed (McTavish et al., 2013). However, a recent study refuted this hypothesis and instead proposed that only a handful of cattle breeds, especially from Italy, carry zebu ancestry in their genome (Decker et al., 2014). The majority of these studies focused only on major breeds from Iberia, Italy, and North-western Europe and lacked in genotypes from other Eastern European regions that are close to the center of domestication.

1.8 Structural variation and its contribution to cattle diversity

Structural variation (SVs) is a term that includes various genomic alterations (Layer et al., 2014) such as insertions, deletions, duplications, inversions, translocations, or other complex rearrangements of large genomic segments (Figure 1.4). Though SVs are not very common, they may have a great impact on gene structure and function (Bickhart and Liu, 2014). Therefore, SVs are an important source of genetic and phenotypic variation between individuals.

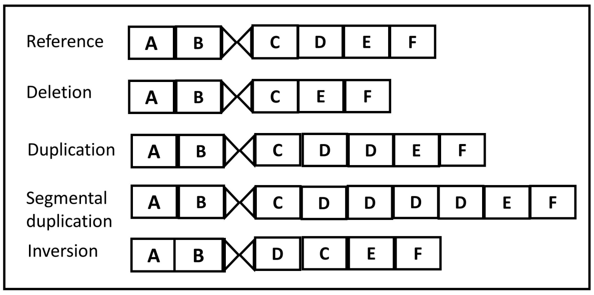

Figure 1.4: Some examples of structural variation in genome sequences

The advancement in the methodologies of genome sequencing has not only accelerated the discovery and genotyping of SVs but have also increased our understanding of its type and formation. Based on the effect of overall genome size, SVs can be categorized in two types: Balanced (translocation and inversion) and unbalanced (insertion, deletion, and duplication) (Bickhart and Liu, 2014). The unbalanced class of SVs can also be called copy number variation (CNV). CNV encompasses a large proportion of cattle and human genomes. For example, in the human genome CNV is estimated to have covered between 4.8-9.5%, while in cattle, it covered approx. 3% of the genome (Zarrei et al., 2015;Bickhart et al., 2016). The methods to identify CNV can either make use of whole genome sequencing data or array probe signal intensities. These methods are also more refined compared to the methods aimed at identifying balanced SVs as the sequence breakpoints in balanced SVs are difficult to pinpoint (Bickhart and Liu, 2014).

CNV play an important role depending on where in the genome it is present. Genic CNVs, for instance, can influence phenotypes of an organism through at least three different mechanisms: change in gene dosage, exposure to recessive alleles and expression regulation changes.

Further, it has been shown that CNVs, if present in the regulatory elements of developmental genes, can also change the phenotypic expression (Spielmann and Klopocki, 2013). Moreover, duplication of the genic region may lead to another gene copy acquiring a novel functional role *(neofunctionalization)*, or the gene's functional role may get divided between these paralogs *(sub-functionalization)*, thereby contributing to the genome evolution. However, genes that are conserved across species or genes that are essential for multiple biological pathways are predicted to be sensitive to CNVs affecting gene expression (Schuster-Bockler et al., 2010). Moreover, different types of repeat regions in a genome contribute to the formation of CNVs. For instance, CNVs are reported to be associated with segmental duplications in mammalian genomes (Sharp et al., 2006).These repetitive regions in the genome facilitate the formation of CNV through mechanisms such as non-allelic homologous recombination (Warburton et al., 2008) . In fact, a recent study has shown that such repetitive regions in the genome are five times more likely to harbour CNVs when compared to germline CNVs (Monlong et al., 2018).

The availability of the Bovine50K and BovineHD 777K SNP arrays has revolutionized the field of bovine genomics. Extensive use of such arrays has led to the identification of many CNVs in the bovine genome (Fadista et al., 2010;Bickhart et al., 2012;Bickhart et al., 2016;Sasaki et al., 2016;Wang et al., 2016). As a result, a complex landscape of CNVs in the bovine genome has emerged. It has been shown that some gene families such as an Olfactory receptor (OR) and genes that play a role in the immune system harbour an abundance of CNVs. Because both these complex gene families serve important functions associated with a sense of smell and ability to resist pathogens, respectively, the evolutionary selection pressure might have played an important role in generating and maintaining variable copy numbers. In cattle, like sheep and pig (Moller et al., 1996;Han et al., 2015), SVs affecting coat colour have also been reported (Durkin et al., 2012;Brenig et al., 2013).

1.9 Identification of structural variations

1.9.1 SNP-array based identification

SNP array platforms typically target biallelic SNPs by including two types of probes, usually coded as A and B, for every single SNP. The resulting hybridization between targeted DNA fragments and probes generates hybridization intensity, which can be used to determine SNP genotypes (Wang et al., 2007). For instance, SVs involving deletions and duplications decrease or increase the total signal intensity, respectively. Apart from the signal intensity, other genomic factors such as GC content around the targeted sites or population allele frequencies can also be included in models to increase the accuracy of identification of SVs. In principle, these methods can only identify SVs involving deletions or duplications. Further, these methods cannot reliably identify break-points around SVs. Some examples of computation programs that can identify SVs based on signal intensity data of SNP array include PennCNV (Wang et al., 2007) andQuantiSNP (Colella et al., 2007) .

1.9.2 WGS-based identification of SVs

The methods used to identify SVs from whole genome sequence (WGS) data can be categorized in four classes: Read-pair (RP), Split-read (SR), Read depth (RD) and assembly-based methods.

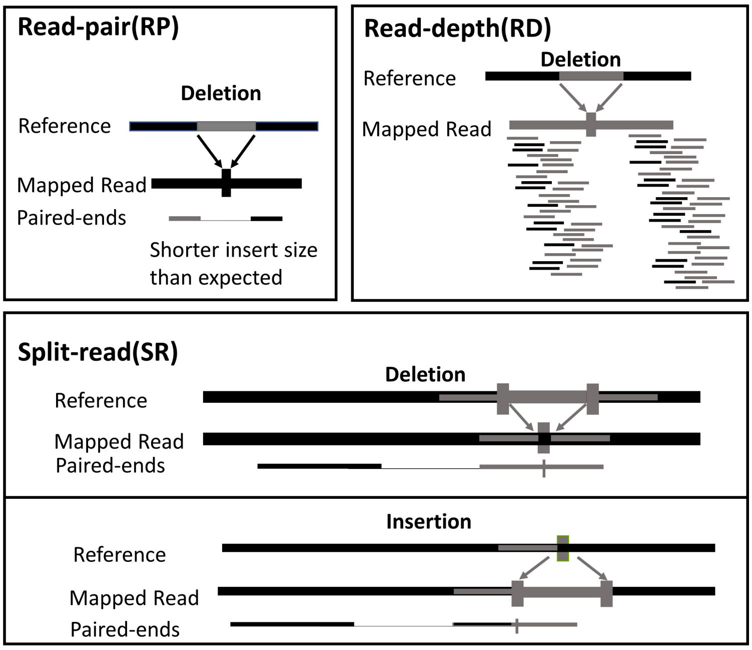

Figure 1.5: Some examples of identification of structural variation events using different whole genome re-sequencing approaches. The figure is adapted from Pirooznia et al.,(2015)

In the paired-end sequencing, DNA fragments are likely to display a specific distribution around the insert size (Korbel et al., 2007). Therefore, read spanning SVs may display a different insert size compared with the genomic average and read pair-based methods use these discordant paired-end reads to identify SVs. However, small sized SVs are difficult to detect using these methods as small disruptions in insert size are difficult to separate from the normal background dispersion in insert size distribution (Medvedev et al., 2009). Further, read pair-based methods are not preferred for detection of SVs in low complexity regions of the genome (Pirooznia et al., 2015). Some methods, in addition to read pair, also consider the split read information to locate precise break-points of SV events (Zhang et al., 2011). Split read methods use reads that remain completely or partially unmapped to the reference genome. Read depth methods exploit the depth of coverage information of genomic alignments to identify deletions or insertions, as there is a direct correlation between the copy number events and depth of coverage (Pirooznia et al., 2015). As opposed to read pair and split read, read depth methods can identify the exact copy number of an event, while the former methods only report the position and the type of event. Moreover, compared to read pair and split read, read depth methods have a higher sensitivity to large CNVs. However, read depth has low efficiency when identifying small CNVs (<1 kbp) (Pirooznia et al., 2015).

There are methods implemented in various tools (Sindi et al., 2012;Layer et al., 2014) that consider a combination of one or more methods described in the previous paragraph. This combined approach often results in a better accuracy of SV identification compared to any single method. In principle, the combination approach-based method combines information from multiple methods, taking advantage of their strength. In doing so, they also overcome the limitation of one method with the unique feature of the another. For instance, combining Read pair, Split read with Read depth has resulted in a high accuracy of identification of small as well as large-sized CNVs (Pirooznia et al., 2015).

1.10 Thesis outline

The overall goal of my research is to investigate the pattern of genetic variation, gene flow and demography in primitive cattle breeds of Europe. By analyzing genotyping data of a large number of cattle breeds, I disentangle the complex relationships between European, African and zebu cattle. Additionally, I also give a broad overview of genetic diversity in some of the least studied cattle breeds of Europe. The practical implications and future direction of the research associated with the results of this thesis are also discussed.Finally, I conclude the thesis by discussing my findings, their importance and applicability in a broader context.

Chapter 2

General discussion

2.1 Introduction

European cattle display vast phenotypic diversity which can be attributed to genomic variations such as single nucleotide polymorphisms (SNPs) and structural variations (SVs). The distribution of these genomic variations in a population is heavily influenced by different population genomic forces such as migration, drift, and selection. In this thesis,genomic variations were characterized in traditional and primitive European cattle breeds using genome-wide SNPs. Specifically, hypotheses concerning gene flow from zebu,African taurine and wild local aurochs ancestry were investigated in detail. To understand the patterns of genomic variations comprehensively, I also characterized the structural variations in the genome of European cattle. In this final chapter, I will discuss the main findings of all the previous chapters in the context of existing literature and knowledge about the genetic structure, admixture, and variations in European cattle.

2.2 Patterns of genomic admixture

2.2.1 On geneflow between European and non-European cattle

The divergence between populations is directly proportional to the time since they shared a most recent common ancestor and differential selection pressure they experienced in their respective environments unless gene flow occurred in these populations. Indeed, the dynamics of population divergence is heavily influenced by gene exchange between isolated populations. In general, gene exchanges between previously isolated populations counter the divergence due to population scaled mutation rate and genetic drift. This demographic model, which is also known as Isolation with Migration (IM), has been investigated widely to explain the genomic divergence observed in a various population of livestock and wild species. For instance, studies have reported the presence of a high frequency of mtDNA haplotypes of Asian origin in various European pig breeds due to admixture. In fact, Bosse et al., (2014) also identified introgressed Asian pig haplotype in European domestic pigs which most probably contributed to increased fertility. These results are in good concordance with the historical record of the early nineteenth century which mention the import of Chinese pigs in Europe because of the renowned fertility of Chinese pigs. Another example is the introgression from Chinese pigs into European pigs of the regulatory gene variant at the porcine IGF2 gene that explains increased muscle growth (Van Laere et al., 2003). However, even though historical records associated with import/migration of zebu cattle are scant, the gene flow from indicine cattle in many European cattle breeds has been hypothesized. For example, based on the similarity of a β-globin variant, Pieragostini et al. (2000) proposed a contribution of zebu cattle in the gene pool of Podolica cattle. Furthermore, analyzing microsatellite markers in different Eurasian cattle breeds, Cymbron et al. (2005) reported that among all mainland European cattle breeds which they studied, Italian cattle breeds—particularly Maremmana and Modicana—followed by Greek cattle breed— Sykia— displayed the highest frequency of indicine population-associated alleles (PAA). They proposed a

Near Eastern origin for this indicine ancestry in Italian and Greek cattle breeds. This hypothesis was further supported by the identification of indicine mtDNA haplotypes in individuals of the Ukrainian Whitehead cattle breed (Kantanen et al., 2009). Further, analyzing genome-wide SNP data, McTavish et al. (2013) reported indicine ancestry in multiple southern European cattle breeds, and they also proposed a north-south gradient of indicine ancestry in Europe. Decker et al., (2014), however, refuted this hypothesis as they reported indicine ancestry only in three Italian cattle breeds—Chianina, Romagnola and Marchigiana. Nevertheless, all these studies lack in the genetic information of cattle breeds from the Balkan region which lies between Anatolia and Italy and therefore, may provide a more comprehensive understanding of indicine ancestry gradient in European cattle breeds.

In this thesis, I used genome-wide SNPs genotyped in different cattle breeds of Balkan and Italian regions (BAI) to characterize indicine ancestry in detail. Using unlinked SNPs and a haplotype-based approach, I show that indicine ancestry is a common feature of several BAI breeds. In chapter 2, I carried out standard population genomics analyses (such as ADMIXTURE and D-statistics) based on high-density SNP array data and proposed that high divergence of BAI breeds can be attributed to indicine ancestry. Interestingly, the signals of indicine ancestry were not observed in any of the Iberian cattle breeds that were investigated, confirming the previous hypothesis that indicine ancestry is uncommon in southern European breeds (Decker et al., 2014). Further, in chapter 3, I carried out a detailed characterization of indicine ancestry in European cattle and showed that different Italian cattle breeds as well as the breed called Busa, of Balkan origin—display a similar proportion of indicine ancestry in their genomes. These results could indicate that BAI breeds received this indicine ancestry from a common ancestor and subsequently, differentiated relatively recently. However, ADMIXTURE analysis is known to be affected by sample size, and moreover, several demographic scenarios often lead to same ADMIXTURE patterns as noted by Lawson et al. (2018). Similarly, the result of D-statistics does not necessarily imply gene flow between the lineages as a subdivision of ancestral populations, if this remains persistent for a long time, also leads to signals similar to recent gene flow (Theunert and Slatkin, 2017). However, sub-structure is unlikely to affect these results as Indian and European cattle have been domesticated independently. Nevertheless, based on the results of chapter 2 and chapter 3, I propose several models as shown in Figure 7.1 that can be tested on whole genome sequencing data (WGS) using a Bayesian approach for thorough investigation of demographic events in BAI breeds.

The fact that BAI breeds still display indicine ancestry in their genomes indicates the possibility that indicine genomic segments might be under selection because of some adaptive advantages they confer to BAI breeds. Indeed, this phenomenon—also known as 'adaptive introgression'—whereby introgressed segments from distantly related populations provide increased fitness to the donor population, has been reported in many animal species (Hedrick, 2013). For instance, Song et al. (2011) identified a large genomic segment in a new world mouse population *(Mus musculus)* which has been introgressed from old world mice and contained the warfarin resistance gene *vkorc1* encoding the vitamin K epoxide reductase subcomponent 1. The BAI cattle breeds display many zebu-like traits such as adaptation to relatively hot climates and better general disease immunity. In fact, Modicana, which is an Italian cattle breed, displays

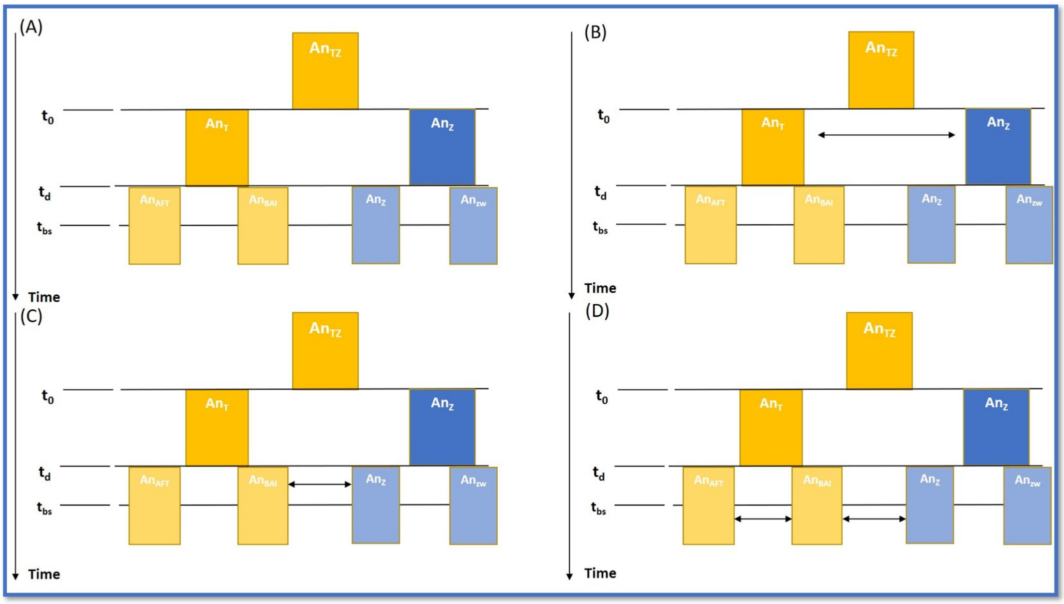

Figure 2.1: Schematic of the proposed demographic models to be tested on whole genome sequencing data. Double-headed arrows represent migration events that should be modelled as two continuous parameters. Barring the model (A), which represents null model without migration events, all other demographic models include migration events. The term “An” refers to the ancestral effective population size or simply, effective population size. Other abbreviations used as subscripts: TZ- term for the ancestors of taurine and zebu before they split, T-Taurine, Z-Zebu, AFT-African taurine, BAI- Balkan and Italian taurine, ZW: ancestral wild Zebu. The “t_0” refers to the number of generations (back in time) in the past at which the ancestral taurine and zebu population separated. The “t_d“ refers to the number of generations (back in time) at which the domestic cattle separated from their wild ancestors. The “t_{bs}“ refers to the number of generations (back in time) at which the African cattle separated from the European domestic cattle.

bifid processes in the last thoracic vertebrae— traditionally considered as a zebu-specific characteristic (Grigson 2000). Therefore, an intensive sampling of various BAI breeds is needed to investigate this hypothesis of adaptive introgression.

Many studies analyzing uniparental markers such as mitochondrial DNA and Y chromosomal haplogroups as well as analyzing microsatellite and genome-wide SNP markers have identified African cattle ancestry in various southern European cattle breeds (Beja-Pereira et al., 2003; Cymbron et al., 2005, Ginja et al., 2010a, 2010b; Decker et al., 2014). In fact, Decker et al. (2014) reported indicine as well as African taurine cattle ancestry in central Italian cattle breeds of Chianiana, Romagnola, and Marchigiana. In Chapter 3, using a haplotype-based approach with genome-wide SNP markers, we proposed that other BAI cattle breeds like Busa and Maremmana also display shared ancestry with non-European cattle which is quantitatively similar to central Italian cattle breeds. Moreover, ADMIXTURE analysis of high-density SNP array data also identified signals of shared ancestry between Iberian and African taurine cattle. These results could be interpreted as a legacy of the Moors who inhabited the Iberian Peninsula between the 8th to 15th centuries and certainly would have brought livestock with them during their more than 600 years presence on the Peninsula. However, because European and African taurine cattle originated from the same domestication center, i.e., the Near East, the possibility of shared ancestry (without migration) cannot be ruled out. Shared genetic variation between relatively closely related populations but without migration has been observed for many species. For instance, by applying a Bayesian approach on microsatellite data, Sousa et al. (2012) showed that the observed genetic patterns in fish populations, which was attributed to admixture using a clustering-based approach, could be better explained by the demographic model with a population split but without admixture. However, introgression in BAI cattle breeds from a ghost population carrying both—African and indicine cattle—ancestry cannot be ruled out.

In this thesis,relationship between southern European cattle and East African zebu was also explored. Although, as described in chapter 2, no shared indicine ancestry between Iberian and Nellore (derived from Indian zebu) was observed,the inclusion of genotyping data of East African zebu in chapter 3 indicated shared ancestry between East African zebu and southern European cattle.However,as East African zebu itself is a cross-breed between African taurine and zebu, this signal of shared ancestry is difficult to interpret.

2.2.2 On geneflow between domestic European cattle and wild local aurochs

Backcrossing between the wild ancestor and its domesticated form is not uncommon in livestock species (Barbato et al., 2017; Frantz et al., 2015; Vilà et al., 2005; Frantz et al., 2013). For instance, Frantz et al., (2013) reported that wild local pigs and domesticated pigs in Eurasia interbred quite often, contrary to the general assumption of reproductive isolation between these two species. However, such events of interbreeding between domesticated cattle and local wild aurochs are highly debated. Before Park et al., (2015) first published the WGS data of British aurochs, researchers had used only uniparental markers (mtDNA and Y-chromosome SNPs) to investigate this research question (Achilli et al., 2008, 2009; Bollongino et al., 2008; Bonfiglio et al., 2010; Götherström et al., 2005; Svensson and Götherström, 2008). While mtDNA studies identified novel haplogroups in Italian cattle breeds such as Chianina and Romagnola, supporting some level of aurochs introgression in Italian cattle breeds (Bonfiglio et al., 2010), the results based on Y-chromosome analysis has remained inconclusive (Bollongino et al., 2008; Götherström et al., 2005). Park and colleagues (2015) analyzed WGS data of wild local aurochs in relation to worldwide cattle breeds, and they concluded, perhaps not surprisingly, that cattle breeds of Britain and Ireland share the highest level of genetic variants with the British aurochs sample among all the cattle breeds that they studied. Although they incorporated genetic data of more than 1200 animals, their dataset lacked in the genetic information of Iberian and some important primitive cattle breeds of BAI regions.

In this thesis,comparative analysis of genomic variants were performed between the British aurochs sample,which was used in the study of Park et al.,(2015), and various primitive cattle breeds of Europe.The results as described in chapter 2 indicated instances of interbreeding between wild local aurochs and ancestors of domestic cattle. However, this gradient of derived alleles across European cattle should be interpreted with caution as this analysis (D-statistics) can provide similar results even in case of a genetic structure in a population. Moreover, significant diversity existed among wild local aurochs as inferred by diverse mitochondrial haplogroups identified in ancient samples of wild aurochs (Bonfiglio et al., 2010). Therefore, the possibility of secondary geneflow between Italian domestic cattle and other distinct sub-population of local aurochs (not related to British aurochs) cannot be ruled out. Overall, our results not only reinforce the earlier findings of Park et al., (2015) but also provide an overview of the distribution of aurochs specific variants in major primitive cattle breeds of Europe.

In the future, availability of WGS data of ancient aurochs bones sampled from different parts in Europe and representing different time periods after the event of cattle domestication may provide more detailed insight into the level of introgression and possible adaptive advantage of these introgressed segments in extant European cattle breeds. Moreover, such studies also have the potential to provide insight into how livestock farming evolved over time, since the beginning of cattle domestication. For instance, a recent study (Bro-Jørgensen et al., 2018) analyzing mtDNA extracted from the horn of the last aurochs bull, identified the T3 haplogroup which is the most common haplogroup in domestic taurine cattle. Based on this result, it can be speculated that the last individuals of the surviving aurochs population might already have exchanged gene-flow with domesticated cattle before they went extinct.

2.3 Patterns of Genetic relatedness/structure and demographic history

Knowledge of genetic relatedness, demographic history and genetic status of a population play a decisive role in conservation management. This information as sometimes recorded in historical literature is often biased or not available. Genetic markers such as SNPs or microsatellites serve as powerful tools to retrieve unambiguous breed information that can reliably be used to design a conservation program.The results as described in chapter 2-4 provided detailed insights into the relationship among European cattle populations and demographic history using genome-wide SNP markers.

2.3.1 Genetic relatedness/structure

The information about genetic structure enables the assignment of individuals to their genetic origin and to identify admixed individuals in a population (Herrero-Medrano et al., 2013; Negrini et al., 2009). In this thesis, unlinked SNP marker-based analyses such as PCA, ADMIXTURE, estimating genetic distance, and haplotype-based approaches as implemented in CHROMOPAINTER and the fineStructure pipeline,were performed to assess genetic structure of European cattle. Generally, high-density SNP arrays suffer from ascertainment bias which can, sometimes, distort inferences about population structure (Albrechtsen et al., 2010). Diversity-related statistics for Swedish traditional cattle breeds as reported in Chapter 4 of this thesis indicates that the Genomic Profiler High-Density Bovine150K (GGP HD150K) array can reliably be used to assess population structure and genetic diversity in European traditional cattle breeds. A similar assessment was not made for the Illumina Bovine777K array as a relatively low number of samples per breed was genotyped using this array, and therefore, it may have provided biased estimations of the diversity statistics. However, as a large number of markers from this array has been validated with MAF > 0.05 across various Eurasian cattle breeds, it can be safely assumed that estimation of genetic structure was unbiased.

The results presented in chapter 2 clearly differentiate cattle populations based on their origin. In fact, genetic structure/population split as recovered in PCA can be interpreted as a legacy of ancient migration events involving Neolithic farmers and their livestock through the Danube River route and the Mediterranean Sea route. In chapter 3 and 4, population clustering using a haplotype-based approach not only successfully retrieved the clustering pattern of unlinked SNP based analyses (PCA and ADMIXTURE) but also provided additional information about sub-structures within a population. For instance, in Chapter 4, the fineStructure-inferred tree clearly identified sub-structure within a population of Swedish mountain cattle breeds that corresponds to the farms from which the samplings were carried out. These results suggest that a haplotype-based approach as implemented in fineStructure is a powerful tool which can even identify sub-structures in a population.

2.3.2 Demographic history

Selection and demography both play significant roles in how genomic variants are distributed among populations. While selection changes the allele frequency of local variants, demographic changes affect variants across the entire genome (Bosse et al., 2015). Moreover, current effective population sizes are of major interest in conservation management. However, due to the lack of pedigree information of local cattle breeds, this information is often difficult to retrieve. Therefore, in this thesis, inferences of demographic history and recent changes in effective population size were carried out using the analysis based on runs of homozygosity (ROH) and linkage disequilibrium (LD). The presence of ROH in a genome indicates relatedness between the parents of an individual. However, the length of ROH decreases over time as recombination events break down ROH into smaller pieces. Therefore, usually, the presence of small ROH in a genome indicates ancestral relatedness, while long ROH indicates inbreeding, e.g., consanguineous mating as the haplotypes have not had enough time to break-down (Bosse et al., 2012). Similarly, LD-based methods also infer past and recent effective population size where small values of r2 at small genetic distances and rapid decay of r2 values indicate vast haplotype diversity (large Ne), and large values of r2 across large interval of genetic distances indicates small Ne (Hayes et al., 2003; Tenesa et al., 2007).

In chapter 2-4,the inferences of demographic events were drawn using the ROH profile of different European cattle populations. I show that the abundance of ROH counts in British cattle breeds indicates the combined effect of genetic isolation and long selection history. Further, low nucleotide diversity outside ROH as estimated using WGS data indicated a relatively small number of founding individuals. Conversely, the presence of only a relatively few ROH segments and high nucleotide diversity outside ROH in several BAI individuals indicated large ancestral founding population or highly diverse ancestral population. Similarly, the abundance of long ROH in several individuals of Iberian and BAI cattle indicated frequent mating between related individuals which can be attributed to a recent reduction in effective population size. Indeed, the recent reduction in the effective population size of Mirandesa and Maltese as reported in Chapter 2-3 is in good agreement with historical and scientific literature concerning these breeds. For example, it was reported that last pure-bred Maltese bull was culled in 1990 and since then, semen of Chianina bulls, which showed Maltese-cattle like feature, were used to recover the population size of Maltese cattle (Lancioni et al., 2016).

In Chapter 4, we investigated demographic history of Swedish traditional cattle populations most of which had never been studied before. The results showed that despite a reduction in population size recorded until the end of the 20^{th} Century, Swedish mountain cattle still has a large effective population size. Additionally,the analyses also suggested that despite genetic isolation the breeders have managed to avoid frequent mating between related individuals in the southern Swedish cattle breeds—Väne cattle and Ringamåla cattle.

2.3.3 Status of genetic diversity

In this thesis, I characterized the genetic diversity of European cattle breeds using data sets obtained by both WGS and high-density SNP arrays. Both these data suggested that southern European cattle breeds—especially BAI cattle—have relatively higher genetic diversity compared to Iberian and North-Western European cattle. This observation can be attributed partly to the complex admixture history of BAI cattle. Moreover, its proximity to the center of domestication means that a founder effect might not have been as strong as it may have been in Western Europe and Iberia. Moreover, estimates of genetic diversity were in concordance with a documented breed history for each of the breeds that were analyzed. For example, breeds with genetic isolation and the recent reduction in Ne like the Maltese and Mirandesa have low genetic diversity. Furthermore, in chapter 4, diversity statistics suggested that Swedish mountain cattle display high genetic diversity, while another subpopulation of SMC—Fjällnära—display high variation in genetic diversity.

2.4 Patterns of Structural variations

SVs such as insertion, deletions, inversions, translocations, and duplications are important classes of genetic variations. These variations can drive mammalian adaptive evolution. In livestock, studies have associated SVs with coat colour and other morphological traits. Moreover, like other genomic markers such as SNPs and microsatellites, the distribution of SVs in a genome is affected by demography and selection. However, compared to SNPs, SVs affect large portions of a genome. Additionally, they may contribute to individual fitness by influencing mRNA and protein expression levels, and therefore, subjected to selection. For instance, genes such as *CATHL4* and *ULBP17* which have been associated with parasitic infection, display a difference in copy number between indicine and taurine cattle (Bickhart et al., 2012). Therefore, defining the distribution and nature of SVs in cattle genomes is crucial to understand the underlying genetic factors responsible for the observed phenotypic diversity between different cattle breeds. In the following paragraphs, I discuss the distribution of SVs in the cattle genome and some important genes encompassed by SVs that we identified in this study. Additionally, I also discuss the strengths and limitations of the tools that were for SVs identification in this thesis.

2.4.1 Structural variations and demography

In chapter 5, copy number variations (CNVs) in cattle genomes were identified using signal intensity data of bovine high-density SNP arrays. We show that, on average, BAI and British cattle display a significantly higher number of CNVs and non-redundant CNV regions (CNVRs) compared to Dutch and Alpine cattle. We also suggest that differential selection pressure and drift effects between cattle breeds can lead to differential CNV counts. However, to validate this hypothesis, additional samples representing different cattle populations need to be genotyped. Nevertheless, this observation is in agreement with a recent study by Mielczarek et al. (2018), in which they reported a significant inter-as well as intra-population variability in copy number loci between different European cattle populations. Similarly, in chapter 6, using WGS data, we also reported higher SV counts in African and Indicine cattle compared to European cattle. A study (Paudel et al., 2013) analyzing CNVs in Eurasian pig populations also reported higher CNV counts in Asian pigs compared to European pigs which they attributed to higher effective population size. On the contrary, Bickhart et al. (2016) observed comparable SV counts across different European and Asian cattle breeds. Therefore, part of the differences between SV counts across cattle populations can be attributed to the fact that the UMD3.1 reference genome, which was used for sequence alignment in this thesis, is assembled from sequences of a Hereford (European taurine) cow.

It has been shown that population inferences based on the pattern of SVs and SNPs produce identical results in geographically distinct populations (Jakobsson et al., 2008). In chapter 5, we show that CNVRs data successfully clustered individuals belonging to the breeds that displayed low genetic diversity using SNP data (such as English longhorn and Maltese). However, hierarchical clustering failed to cluster the individuals based on the geographical similarities, indicating the effect of small sample size and sharing of high frequent CNVRs. Moreover, the possibility of false positive CNVs distorting the sharing of CNVRs cannot be excluded.

2.4.2 Structural variation and functional annotation

Many CNVs affecting phenotypic traits related to coat colour and morphology have been identified in livestock as well as in companion animals (Durkin et al., 2012; Jakobsson et al., 2008; Salmon Hillbertz et al., 2007). For instance, Salmon Hillbertz et al., (2007) identified a 133-kb duplication in the genome of Ridgeback dogs which encompass three fibroblast growth factor (FGF) genes and causes hair ridge and predisposition to dermoid sinus. In chapter 5 and 6, we identified SVs encompassing various genes related to important livestock traits. Additionally, in both these studies, we also reported over-representation of genes related to immunity and olfaction processes. In chapter 5, we identified and validated the structural variant (Cs29) encompassing the *KIT* gene in English Longhorn cattle. This variant was first identified in Belgian blue cattle and was shown to be associated with coat-colour sidedness (Durkin et al., 2012). Later, Brenig et al. (2013) identified the same variant in White Park and Galloway cattle. In addition, they also suggested a dose-dependent effect of Cs29 in these breeds. Interestingly, English Longhorn cattle also display considerable variation in coat colour, i.e., such as red, brown, grey or white. Therefore, it is likely that such a dose-dependent effect of the Cs29 variant might be responsible for coat colour variation in English Longhorn cattle.

In this thesis, SVs were identified in various genes related to metabolism, meat quality, and immunity-related traits. For instance, in chapter 6, we described SVs encompassing genes such as *CAST* and *CAPN13* that are associated with meat quality and tenderness (Barendse et al., 2007; Casas et al., 2006; Tizioto et al., 2013). However, to verify their effect on gene expression requires that transcriptome data be generated from relevant tissues.

In human, studies have suggested that a large number of SVs are shared across different populations (Sjödin and Jakobsson, 2012). SV can arise independently in a population and, if selected upon, can spread in a population. Studies have identified population-specific SVs in many cattle populations (Bickhart et al., 2016; Xu et al., 2016). In chapter 6, we identified several population-specific SV in African and Zebu cattle populations. Moreover, several novel SVs were also identified in primitive cattle breeds. For instance, a novel SV was identified in the gene *HERC2*. This gene has been associated with pigmentation in human (Visser et al., 2012). However, as I mentioned earlier in the section, verification of such novel SVs as identified in this thesis requires that many samples with recorded phenotypes related to coat colour and body conformation traits should be investigated using gene expression data.

2.4.3 Structural variation in ancient aurochs sample

In recent times, the advancement in experimental and bioinformatics approaches has led to sequencing and analysis of hundreds of ancient genomes (Orlando et al., 2015), which in turn, have transformed our understanding of population genomics forces leading to speciation and adaptation. Studies involving ancient genomes in livestock and humans have shown how the genetic make-up of populations has changed substantially over a short period of time owing to the selection pressure (Lazsaridis et al., 2016; Orlando et al., 2013, 2015; Somel et al., 2016). Moreover, studies of ancient genomes also allow researchers to trace back the age of functional alleles across time. However, the fragmentation of ancient DNA due to post-mortem changes, limits the read length of DNA molecules between 60 to 150 bp length, which is shorter than the read length generated by Illumina sequencing technology (Miller et al. 2008; Briggs et al. 2009). Moreover, this fragmented nature of ancient DNA (aDNA) also prevents sequencing using paired-end approaches. Therefore, often single-end sequencing has been preferred to sequence aDNA and subsequently, read-depth approaches have been used to identify SVs (Lin et al., 2015; Sudmant et al., 2015). In this thesis, SVs were identified in aDNA prepared from an aurochs sample using a read-depth approach as implemented in CNVnator. We reported that about 80% of the total duplications identified in the aurochs sample are still segregating among modern cattle. In fact, we also identified one shared deletion between ancient aurochs and the studied cattle breeds which likely have the same break-points. Therefore, it can be hypothesized that many of the SV between aurochs and modern cattle are identical by descent. Moreover, It is likely that such SVs might be under selection because of the adaptive advantage they confer. In human, a recent study has identified a shared deletion event between ancient Neanderthals and modern non-African human populations which the authors attributed to introgression from the Neanderthals (Sudmant et al., 2015). Similarly, we hypothesized that secondary introgression from aurochs in European cattle might have led to frequency differences of "introgressed SVs" between European and non-European taurine. However, such study awaits sequence data from ancient aurochs sample with substantially much higher coverage than those currently available.

2.4.4 Challenges of SVs identification in livestock

SVs can be identified from various types of data generated by WGS, comparative genomic hybridization (CGH) and SNP arrays (Alkan et al., 2011). However, studies have reported a low agreement in the SV identified from different data sources in the same individual (Pinto et al., 2011; Zhan et al., 2011). For instance, Zhan et al. (2011) identified SVs using three different platforms (WGS, SNP array, and CGH) in the same individual and observed only a maximum of 23% overlap among these platforms. Moreover, studies have also reported low agreement between different CNV callers used on the same platform (Legault et al., 2015; Pinto et al., 2011). Indeed, different algorithms used for SVs identification have their own strengths and limitations. For example, *Lumpy*, which uses split-read and discordant reads to identify SVs, can identify small SV events reliably compared to the large events in repetitive regions of a genome. Moreover, in other platforms such as SNP array, a large fluctuation in signal intensity data due to relatively bad DNA quality can lead to the identification of false positive SVs. Furthermore, sometimes SV identification algorithms are optimized using data only generated in humans, which makes the comprehensive SVs identification in non-model organisms difficult. Therefore, selection of proper tools and optimizing post-filtering strategies to generate reliable and reproducible SV set in livestock is a major challenge for researchers.

Across all the SVs identification platforms, the quality and quantity of SVs heavily rely on a good reference genome assembly. For example, in SNP array, the hybridization probes to capture the variants of interest are designed from the reference genome, while alignment against the reference genome is often the first step in re-sequenced data produced using WGS approaches. Therefore, it is essential that the reference genome is as complete, correctly assembled and error-free as possible so that SV can be reliably identified. Unfortunately, often the reference assembly in livestock genomes are incomplete with relatively high errors, which may cause misinterpretation of the underlying sequences involved in SV. For example, Zimin et al. (2012) identified 39 Mb of sequences which were incorrectly assembled as segmental duplications in the *Btau4.1* cattle reference assembly. Using an SNP array platform, Zhou et al. (2016) identified 9 frequent false-positive copy number variable regions which were attributed to assembly errors. In fact, in chapter 5, we also reported that underlying probes covering the same regions show different signal intensities between cows and bulls, supporting the findings of Zhou et al., (2016). Also, in the same chapter, we identified an abundance of CNVs between 72 and 74 Mb region of chromosome 12, which partly can be attributed to assembly errors as the size of chromosome 12 in the new cattle genome build (ARS-UCD 1.2) is about 2 Mb shorter compared to that of UMD3.1. Therefore, working with incomplete genome-builds for SVs identification requires that results should be interpreted with caution.

Balanced SVs such as translocations and inversions, which do not change the overall copy number of the sequences, are difficult to identify with the current sequencing technologies. Therefore, this likely is one of the reasons for the underrepresentation of such events in genomic data obtained from livestock studies. Nevertheless, studies have shown that events such as translocations can have a significant impact on phenotypic diversity like, e.g., the *KIT* gene translocation affecting coat colour in Belgian blue and Brown-swiss (Durkin et al., 2012). In chapter 5, we also confirmed that this translocation is quite frequently present in British cattle breeds. However, the lack of phenotypic information meant that the effect of this translocation in these breeds could not be investigated. In fact, linking phenotypes with underlying genotypes is the most crucial goal of livestock genomics. Therefore, detailed and extensive phenotypic information from large numbers of individuals is essential to allow the proper understanding of underlying genotypes. To this end, it is also required that proper genome annotations are available for the genome assemblies for livestock.

2.5 Genomic characterization of European cattle: combining information from SNP arrays and whole genome sequences

Conservation management of a population requires a thorough understanding of the pattern of admixture, demography and genetic diversity. Moreover, scanning the genomic sequences in a population may unravel the genomic basis of adaptation. Native local cattle breeds have inhabited their respective environments for many centuries. However, limited availability of literature related to breed history for some populations may act as a hurdle in their conservation management. The distribution of genomic variations in a population provides reliable information about breed history. However, a question may arise regarding the type of genomic variations that should be used for genetic characterization. Although genotyping microsatellite markers in animals is more economical than SNP arrays, it provides only partial information concerning demographic history assessed through LD and ROH (Herrero-Medrano, 2013). Inferences based on SNP arrays and WGS indicated similar demographic history for European cattle populations. These results show that despite the small ascertainment bias, the currently available BovineHD genotyping arrays are very useful in deriving statistics related to genetic diversity and demography.

The results obtained from the studies described in this thesis demonstrate demographic history and admixture as two of the most important forces driving the distribution of genomic variations in cattle populations. Moreover, I also demonstrated that substantial genetic diversity exists among European cattle population which can be attributed to the founder effects involving migration of Neolithic farmers as well as gene-flow from non-European taurine cattle populations. For instance, BAI cattle breeds displayed high heterozygosity as well as an abundance of common and unique SV. These results partly can be attributed to the fact that among all the cattle breeds studied in this thesis, geographically, BAI cattle breeds are the closest to the centre of domestication.

Most of the native cattle breeds are still reared in small farms using conventional management. However, differences in breeding strategy between farms can lead to heterogeneous population structure. For instance, in chapter 4, sub-structure was identified in the two Swedish cattle breeds-Fjällnära and Ringamålako. Moreover, It was also demonstrated that cross-breeding between local cattle breeds is also a prominent factor contributing to genetic diversity. Conversely, low genetic diversity due to genetic isolation in Iberian cattle breeds—Mirandesa and Cachena—and Swedish cattle breeds—Väneko—requires conservation efforts. I propose cross-breeding with individuals from phenotypically similar breeds might be a sustainable approach to conserve the breeds at risk, for instance, Ringamålako in case of Väneko. This might enhance the genetic diversity in such genetically isolated breeds. In fact, such conservation steps have already been carried out in the Maltese cattle breed, where Chianina bull has been used to increase the genetic diversity in this breed.

In this thesis, using SNP array as well as a whole genome sequencing approach, many common as well as novel SVs were identified. These results could indicate that native cattle breeds harbor unique genomic variants which might play an important role in adaptation. Moreover, in chapter 6, novel SVs have been exclusively identified in African taurine and Indian zebu. These results could indicate that SVs plays a vital role in population differentiation. However, determining the break-points of SV events was a major challenge in the studies performed in this thesis. Perhaps, in the future, the availability of sequence data produced by long read sequencing approaches may help resolve this issue.

2.6 Concluding remarks

This thesis provided detailed insights into how demographic changes and admixture patterns have contributed to genomic variation among European cattle breeds. The results in this thesis suggest a contribution of non-European taurine and ancestral wild aurochs populations, which warrants further investigation concerning adaptive introgression. Moreover, the results related to genetic diversity and population structure are valuable for conservation management of native cattle breeds. In this thesis, I also identified novel and lineage-specific structural variations which can be targeted by future association studies.

References

Achilli, A. et al. (2009). The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS ONE*, *4*, e5753.

Achilli, A. et al. (2008). Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology*, *18*, R157–R158.

Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment biases in snp chips affect measures of population divergence. *Molecular Biology and Evolution*, *27*, 2534–2547.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–64.

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, *12*, 363–376.

Bahbahani, H. et al. (2017). Signatures of selection for environmental adaptation and zebu × taurine hybrid fitness in east african shorthorn zebu. *Frontiers in Genetics*, *8*.

Balme, D. B. (1965). *History of animals translated from Historia animalium by Aristotle.*. London : Heinemann ; Cambridge, Mass. : Harvard University Press, 1965-1991.

Barbato, M., Hailer, F., Orozco-terWengel, P., Kijas, J., Mereu, P., Cabras, P., Mazza, R., Pirastru, M., & Bruford, M. W. (2017). Genomic signatures of adaptive introgression from european mouflon into domestic sheep. *Scientific Reports*, *7*.

Barendse, W., Harrison, B. E., Hawken, R. J., Ferguson, D. M., Thompson, J. M., Thomas, M. B., & Bunch, R. J. (2007). Epistasis between calpain 1 and its inhibitor calpastatin within breeds of cattle. *Genetics*, *176*, 2601–2610.

Bartosiewicz, L. (2011). The hungarian grey cattle: a traditional european breed. *Animal Genetic Resources Information*, *21*, 49–60.

Beja-Pereira, A., Alexandrino, P., Bessa, I., Carretero, Y., Dunner, S., Ferrand, N., Jordana, J., Laloe, D., Moazami-Goudarzi, K., Sanchez, A., & Cañon, J. (2003). Genetic characterization of southwestern european bovine breeds: A historical and biogeographical reassessment with a set of 16 microsatellites. *Journal of Heredity*, *94*, 243–250.

Beja-Pereira, A. et al. (2006). The origin of european cattle: evidence from modern and ancient dna. *Proceedings of the National Academy of Sciences*, *103*, 8113–8.

Bickhart, D. M. et al. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Research*, (pp. 778–790).

Bickhart, D. M., & Liu, G. E. (2014). The challenges and importance of structural variation detection in livestock. *Frontiers in Genetics*, *5*, 37.

Bickhart, D. M. et al. (2016). Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. *DNA Research*, *23*, 253–262.

Bodo, I., Gera, I., & Koppany, G. (2004). *The hungarian grey cattle breed book*. The hungarian grey cattle breed book. Budapest: ASSOCIATION OF THE HUNGARIAN GREY CATTLE BREEDERS.

Bollongino, R., Burger, J., Powell, A., Mashkour, M., Vigne, J. D., & Thomas, M. G. (2012). Modern taurine cattle descended from small number of near-eastern founders. *Molecular Biology and Evolution*, *29*, 2101–4.

Bollongino, R., Elsner, J., Vigne, J.-D., & Burger, J. (2008). Y-snps do not indicate hybridisation between european aurochs and domestic cattle. *PLoS ONE*, *3*, e3418.

Bonfiglio, S., Achilli, A., Olivieri, A., Negrini, R., Colli, L., Liotta, L., Ajmone-Marsan, P., Torroni, A., & Ferretti, L. (2010). The enigmatic origin of bovine mtdna haplogroup r: Sporadic interbreeding or an independent event of bos primigenius domestication in italy? *PLoS ONE*, *5*, e15760.

Bosse, M., Megens, H.-J., Frantz, L. A. F., Madsen, O., Larson, G., Paudel, Y., Duijvesteijn, N., Harlizius, B., Hagemeijer, Y., Crooijmans, R. P. M. A., & Groenen, M. A. M. (2014). Genomic analysis reveals selection for asian genes in european pigs following human-mediated introgression. *Nature Communications*, *5*, 4392.

Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L. A. F., Schook, L. B., Crooijmans, R. P. M. A., & Groenen, M. A. M. (2012). Regions of homozygosity in the porcine genome: Consequence of demography and the recombination landscape. *PLoS Genetics*, *8*, e1003100–e1003100.

Bo¨ko¨nyi, S., Hala´py, L., & Tringham, R. (1974). *History of domestic mammals in central and eastern Europe*. Akade´miai Kiado´.

Bradley, D. G., MacHugh, D. E., Cunningham, P., & Loftus, R. T. (1996). Mitochondrial diversity and the origins of african and european cattle. *Proceedings of the National Academy of Sciences*, *93*, 5131.

Brenig, B., Beck, J., Floren, C., Bornemann-Kolatzki, K., Wiedemann, I., Hennecke, S., Swalve, H., & Schütz, E. (2013). Molecular genetics of coat colour variations in white galloway and white park cattle. *Animal Genetics*, *44*, 450–453.

Bro-Jørgensen, M. H., Carøe, C., Vieira, F. G., Nestor, S., Hallström, A., Gregersen, K. M., Etting, V., Gilbert, M. T. P., & Sinding, M.-H. S. (2018). Ancient dna analysis of scandinavian medieval drinking horns and the horn of the last aurochs bull. *Journal of Archaeological Science*, *99*, 47–54.

Broxham, E. T., Kugler, W., & Medugorac, I. (2015). A case study on strains of busa cattle structured into a metapopulation to show the potential for use of single-nucleotide polymorphism genotyping in the management of small, cross-border populations of livestock breeds and varieties. *Frontiers in Genetics*, *6*, 73.

Campbell, B. M. S. (2009). The great famine: Northern europe in the early fourteenth century. by william chester jordan. princeton: Princeton university press, 1996. pp. 318. $29.95. *The Journal of Economic History*, *57*, 725–726.

Casas, E., White, S. N., Wheeler, T. L., Shackelford, S. D., Koohmaraie, M., Riley, D. G., Chase, C. C., Johnson, D. D., & Smith, T. P. L. (2006). Effects of calpastatin and µ-calpain markers in beef cattle on tenderness traits1,2. *Journal of Animal Science*, *84*, 520–525.

Colella, S., Yau, C., Taylor, J. M., Mirza, G., Butler, H., Clouston, P., Bassett, A. S., Seller, A., Holmes, C. C., & Ragoussis, J. (2007). Quantisnp: an objective bayes hidden-markov model to detect and accurately map copy number variation using snp genotyping data. *Nucleic Acids Research*, *35*, 2013–25.

Conolly, J., Manning, K., Colledge, S., Dobney, K., & Shennan, S. (2012). Species distribution modelling of ancient cattle from early neolithic sites in sw asia and europe. *The Holocene*, *22*, 997–1010.

Cymbron, T., Freeman, A. R., Isabel Malheiro, M., Vigne, J. D., & Bradley, D. G. (2005). Microsatellite diversity suggests different histories for mediterranean and northern european cattle populations. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 1837–1843.

Cymbron, T., Loftus, R. T., Malheiro, M. I., & Bradley, D. G. (1999). Mitochondrial sequence variation suggests an african influence in portuguese cattle. *Proceedings of the Royal Society B: Biological Sciences*, *266*, 597–603.

Daniel, H., Lionel, G., Hervé, M., Peters, J., & Segui, M. S. (2005). Identifying early domestic cattle from pre-pottery neolithic sites on the middle euphrates using sexual dimorphism. In J. D. Vigne, J. Peters, & D. Helmer (Eds.), *First Steps of Animal Domestication New archaeozoological approaches* book section 9. (pp. 86–96). Oxbow Books.

de, L., & Íñigo, G.-M. (2014). Recent data and approaches on the neolithization of the iberian peninsula. *European Journal of Archaeology*, *18*, 429–453.

Decker, J. E. et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics*, *10*, e1004254.

Di Lorenzo, P. et al. (2018). Mitochondrial dna variants of podolian cattle breeds testify for a dual maternal origin. *PLOS ONE*, *13*, e0192567.

Dorian, J. G., & Ruvinsky, A. (2014). *The genetics of cattle*. Wallingford, Oxfordshire: CAB International.

Durkin, K. et al. (2012). Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature*, *482*, 81–4.

Edwards, C. J., MacHugh, D. E., Dobney, K. M., Martin, L., Russell, N., Horwitz, L. K., McIntosh, S. K., MacDonald, K. C., Helmer, D., Tresset, A., Vigne, J.-D., & Bradley, D. G. (2004). Ancient dna analysis of 101 cattle remains: limits and prospects. *Journal of Archaeological Science*, *31*, 695–710.

Epstein, H. (1971). *The origin of the domestic animals of Africa*. Africana Publishing Corporation.

Fadista, J., Thomsen, B., Holm, L.-E., & Bendixen, C. (2010). Copy number variation in the bovine genome. *BMC Genomics*, *11*, 284–284.

Fao (2015). The second report on the state of the world's fao commission on genetic resources for food and agriculture assessments • 2015, . (p. 36).

Felius, M., Beerling, M.-L., Buchanan, D., Theunissen, B., Koolmees, P., & Lenstra, J. (2014). On the history of cattle genetic resources. *Diversity*, *6*, 705–750.

Ferdinando, C., & Donato, M. (2001). Bovino grigio allevato in italia: origine. nota 1: il bovino macrocero. *Taurus speciale*, *13*, 89–99.

Frantz, A. C., Zachos, F. E., Kirschning, J., Cellina, S., Bertouille, S., Mamuris, Z., Koutsogiannouli, E. A., & Burke, T. (2013). Genetic evidence for introgression between domestic pigs and wild boars (sus scrofa) in belgium and luxembourg: a comparative approach with multiple marker systems. *Biological Journal of the Linnean Society*, *110*, 104–115.

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., Paudel, Y., Crooijmans, R. P. M. A., Larson, G., & Groenen, M. A. M. (2015). Evidence of long-term gene flow and selection during domestication from analyses of eurasian wild and domestic pig genomes. *Nature Genetics*, *47*, 1141–1148.

Gargani, M., Pariset, L., Lenstra, J. A., De Minicis, E., Holm, L. E., Moazami-Goudarzi, K., Martinez, A., & Valentini, A. (2015). Microsatellite genotyping of medieval cattle from central italy suggests an old origin of chianina and romagnola cattle. *Frontiers in Genetics*, *6*.

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J. G., Boichard, D., Lathrop, G. M., Gut, I. G., & Eggen, A. (2007). Genetic and haplotypic structure in 14 european and african cattle breeds. *Genetics*, *177*, 1059–70.

Gautier, M., Moazami-Goudarzi, K., Leveziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., & Flori, L. (2016). Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences. *Molecular Biology and Evolution*, *33*, 2801–2814.

Gifford-Gonzalez, D., & Hanotte, O. (2011). Domesticating animals in africa: Implications of genetic and archaeological findings. *Journal of World Prehistory*, *24*, 1–23.

Ginja, C., Penedo, M. C. T., Melucci, L., Quiroz, J., Martínez López, O. R., Revidatti, M. A., Martínez-Martínez, A., Delgado, J. V., & Gama, L. T. (2010a). Origins and genetic diversity of new world creole cattle: inferences from mitochondrial and y chromosome polymorphisms. *Animal Genetics*, *41*, 128–141.

Ginja, C., Telo Da Gama, L., & Penedo, M. C. T. (2010b). Analysis of str markers reveals high genetic structure in portuguese native cattle. *Journal of Heredity*, *101*, 201–210.

Gotherstrom, A., Anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D. G., & Ellegren, H. (2005). Cattle domestication in the near east was followed by hybridization with aurochs bulls in europe. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 2345–2351.

Grigson, C. (1991). An african origin for african cattle? some archaeological evidence. *The African Archaeological Review*, *9*, 119–144.

Götherström, A., Anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D. G., & Ellegren,

H. (2005). Cattle domestication in the near east was followed by hybridization with aurochs bulls in europe. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 2345–50.

Hall, S., & Clutton-Brock, J. (1988). *Two Hundred Years of British Farm Livestock*. British Museum, London, UK.

Han, J. L., Yang, M., Yue, Y. J., Guo, T. T., Liu, J. B., Niu, C. E., & Yang, B. H. (2015). Analysis of agouti signaling protein (asip) gene polymorphisms and association with coat color in tibetan sheep (ovis aries). *Genetics and Molecular Research*, *14*, 1200–9.

Hartl, G. B., Göltenboth, R., Grilltsch, M., & Willing, R. (1988). On the biochemical systematics of the bovini. *Biochemical Systematics and Ecology*, *16*, 575–579.

Hassanin, A., & Ropiquet, A. (2004). Molecular phylogeny of the tribe bovini (bovidae, bovinae) and the taxonomic status of the kouprey, bos sauveli urbain 1937. *Mol Phylogenet Evol*, *33*, 896–907.

Hayes, B. J. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, *13*, 635–643.

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*, 4606–4618.

Herrero-Medrano, J. M., Megens, H. J., Crooijmans, R. P., Abellaneda, J. M., & Ramis, G. (2012). Farm-by-farm analysis of microsatellite, mtdna and snp genotype data reveals inbreeding and crossbreeding as threats to the survival of a native spanish pig breed. *Animal Genetics*, *44*, 259–266.

Hofmanová, Z. et al. (2016). Early farmers from across europe directly descended from neolithic aegeans. *Proceedings of the National Academy of Sciences*, *113*, 6886–6891.

Hu, Q., Ma, T., Wang, K., Xu, T., Liu, J., & Qiu, Q. (2012). The yak genome database: an integrative database for studying yak biology and high-altitude adaption. *BMC Genomics*, *13*, 600.

Jakobsson, M. et al. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, *451*, 998–1003.

Jorge, W. (1974). Chromosome study of some breeds of cattle. *Caryologia*, *27*, 325–329.

Kantanen, J., Edwards, C. J., Bradley, D. G., Viinalass, H., Thessler, S., Ivanova, Z., Kiselyova, T., Ćinkulov, M., Popov, R., Stojanović, S., Ammosov, I., & Vilkki, J. (2009). Maternal and paternal genealogy of eurasian taurine cattle (bos taurus). *Heredity*, *103*, 404–415.

Kieffer, N. M., & Cartwright, T. C. (1968). Sex chromosome polymorphism in domestic cattle. *Journal of Heredity*, *59*, 35–36.

Kim, J. et al. (2017). The genome landscape of indigenous african cattle. *Genome Biology*, *18*, 34.

Korbel, J. O. et al. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science*, *318*, 420.

Kron, G. (2002). *Archaeozoology and the Productivity of Roman Livestock Farming* volume 21.

L. Janecek, L., Honeycutt, R., M. Adkins, R., & K. Davis, S. (1996). Mitochondrial gene sequences and the molecular systematics of the artiodactyl subfamily bovinae. *Molecular Phylogenetics and Evolution*, *6*, 107–119.

Lancioni, H. et al. (2016). Survey of uniparental genetic markers in the maltese cattle breed reveals a significant founder effect but does not indicate local domestication. *Animal Genetics*, *47*, 267–269.

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret structure and admixture bar plots. *Nature Communications*, *9*.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, *8*, e1002453–e1002453.

Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). Lumpy: a probabilistic framework for structural variant discovery. *Genome Biology*, *15*, R84.

Lazaridis, I. et al. (2016). Genomic insights into the origin of farming in the ancient near east. *Nature*, *536*, 419–424.

Legault, M.-A., Girard, S., Lemieux Perreault, L.-P., Rouleau, G. A., & Dubé, M.-P. (2015). Comparison of sequencing based cnv discovery methods using monozygotic twin quartets. *PLOS ONE*, *10*, e0122287.

Li, N., & Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, *165*, 2213.

Lin, Y.-L., Pavlidis, P., Karakoc, E., Ajay, J., & Gokcumen, O. (2015). The evolution and functional impact of human deletion variants shared with archaic hominin genomes. *Molecular Biology and Evolution*, *32*, 1008–1019.

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., & Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences*, *91*, 2757.

MacEachern, S., McEwan, J., & Goddard, M. (2009). Phylogenetic reconstruction and the identification of ancient polymorphism in the bovini tribe (bovidae, bovinae). *BMC Genomics*, *10*, 177.

Martins, H., Oms, F. X., Pereira, L., Pike, A. W. G., Rowsell, K., & Zilhão, J. (2015). Radiocarbon dating the beginning of the neolithic in iberia: New results, new problems. *Journal of Mediterranean Archaeology*, *28*, 105–131.

Mason, I. L. (1984). *Evolution of domesticated animals*. Prentice Hall Press.

McTavish, E. J., Decker, J. E., Schnabel, R. D., Taylor, J. F., & Hillis, D. M. (2013). New world cattle show ancestry from multiple independent domestication events. *Proceedings of the National Academy of Sciences*, *110*, E1398–E1406.

Medugorac, I. et al. (2017). Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of mongolian yaks. *Nature Genetics*, *49*, 470.

Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C. E., Taberlet, P., Luntz, B., Mix, H. M., & Forster, M. (2009). Genetic diversity of european cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Molecular Ecology*, *18*, 3394–410.

Medvedev, P., Stanciu, M., & Brudno, M. (2009). Computational methods for discovering structural variation with next-generation sequencing. *Nature Methods*, *6*, S13.

Mielczarek, M., Fraszczak, M., Nicolazzi, E., Williams, J. L., & Szyda, J. (2018). Landscape of copy number variations in bos taurus: individual – and inter-breed variability. *BMC Genomics*, *19*.

Moller, M. J., Chaudhary, R., Hellmén, E., Höyheim, B., Chowdhary, B., & Andersson, L. (1996). Pigs with the dominant white coat color phenotype carry a duplication of the kit gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome*, *7*, 822–830.

Monlong, J., Cossette, P., Meloche, C., Rouleau, G., Girard, S. L., & Bourque, G. (2018). Human copy number variants are enriched in regions of low mappability. *Nucleic Acids Research*, *46*, 7236–7249.

Negrini, R., Nicoloso, L., Crepaldi, P., Milanesi, E., Colli, L., Chegdani, F., Pariset, L., Dunner, S., Leveziel, H., Williams, J. L., & Ajmone Marsan, P. (2009). Assessing snp markers for assigning individuals to cattle populations. *Animal Genetics*, *40*, 18–26.

Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics*, *16*, 395–408.

Orlando, L. et al. (2013). Recalibrating equus evolution using the genome sequence of an early middle pleistocene horse. *Nature*, *499*, 74–78.

Park, S. D. E. et al. (2015). Genome sequencing of the extinct eurasian wild aurochs, bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biology*, *16*, 234.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*, 1065–93.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*.

Pieragostini, E., Scaloni, A., Rullo, R., & Di Luccia, A. (2000). Identical marker alleles in podolic cattle (bos taurus) and indian zebu (bos indicus). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *127*, 1–9.

Pinto, D. et al. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology*, *29*, 512–520.

Pirooznia, M., Goes, F. S., & Zandi, P. P. (2015). Whole-genome cnv analysis: advances in computational approaches. *Frontiers in Genetics*, *6*, 138.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945.

Ritz, L. R., Glowatzki-Mullis, M. L., MacHugh, D. E., & Gaillard, C. (2000). Phylogenetic analysis of the tribe bovini using microsatellites. *Animal Genetics*, *31*, 178–185.

Salmon Hillbertz, N. H. C. et al. (2007). Duplication of fgf3, fgf4, fgf19 and oraov1 causes hair ridge and predisposition to dermoid sinus in ridgeback dogs. *Nature Genetics*, *39*, 1318–1320.

Salque, M., Bogucki, P. I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M., & Evershed, R. P. (2013). Earliest evidence for cheese making in the sixth millennium bc in northern europe. *Nature*, *493*, 522–5.

Sargentini, C., Riccardo, B., Pablo, D.-R., Alessandro, G., Andrea, M., Paola, L., Cazzola, P., Simona, B., & Tiziana, C. (2007). Onset of puberty in maremmana heifers. *Italian Journal of Animal Science*, *6*, 385–394.

Sasaki, S., Watanabe, T., Nishimura, S., & Sugimoto, Y. (2016). Genome-wide identification of copy number variation using high-density single-nucleotide polymorphism array in japanese black cattle. *BMC Genetics*, *17*, 26–26.

Scheu, A., Powell, A., Bollongino, R., Vigne, J. D., Tresset, A., Cakirlar, C., Benecke, N., & Burger, J. (2015a). The genetic prehistory of domesticated cattle from their origin to the spread across europe. *BMC Genetics*, *16*, 54.

Scheu, A., Powell, A., Bollongino, R., Vigne, J.-D., Tresset, A., Çakırlar, C., Benecke, N., & Burger, J. (2015b). The genetic prehistory of domesticated cattle from their origin to the spread across europe. *BMC Genetics*, *16*, 54.

Schuster-Bockler, B., Conrad, D., & Bateman, A. (2010). Dosage sensitivity shapes the evolution of copy-number varied regions. *PLoS One*, *5*, e9474.

Sharp, A. J., Cheng, Z., & Eichler, E. E. (2006). Structural variation of the human genome. *Annual Review of Genomics and Human Genetics*, *7*, 407–442.

Sindi, S. S., Önal, S., Peng, L. C., Wu, H.-T., & Raphael, B. J. (2012). An integrative probabilistic model for identification of structural variation in sequencing data. *Genome Biology*, *13*, R22.

Sjödin, P., & Jakobsson, M. (). Population genetic nature of copy number variation. In *Methods in Molecular Biology* (pp. 209–223). Springer New York.

Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F.-R., Shih, C.-H., Nachman, M., & Kohn, M. (2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, *21*, 1296–1301.

Soubrier, J. et al. (2016). Early cave art and ancient dna record the origin of european bison. *Nature Communications*, *7*, 13158.

Sousa, V. C., Beaumont, M. A., Fernandes, P., Coelho, M. M., & Chikhi, L. (2011). Population divergence with or without admixture: selecting models using an abc approach. *Heredity*, *108*, 521–530.

Spielmann, M., & Klopocki, E. (2013). Cnvs of noncoding cis-regulatory elements in human disease. *Curr Opin Genet Dev*, *23*, 249–56.

Stanley, P. (1995). *Robert Bakewell and the Longhorn breed of cattle*. Ipswich :: Farming Press.

Sudmant, P. H. et al. (2015). Global diversity, population stratification, and selection of human copy-number variation. *Science*, *349*, aab3761.

Svensson, E., & Gotherstrom, A. (2008). Temporal fluctuations of y-chromosomal variation in bos taurus. *Biology Letters*, *4*, 752–754.

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., & Visscher, P. M. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Research*, *17*, 520–526.

Theunert, C., & Slatkin, M. (2017). Distinguishing recent admixture from ancestral population structure. *Genome Biology and Evolution*, *9*, 427–437.

Thomas, R. (2005). Zooarchaeology, improvement and the british agricultural revolution. *International Journal of Historical Archaeology*, *9*, 71–88.

Tizioto, P. C. et al. (2013). Genome scan for meat quality traits in nelore beef cattle. *Physiological Genomics*, *45*, 1012–1020.

Tresset, A. (2003). French connections ii: of cows and men. In E. M. E. N. Ian Armit, & S. Derek (Eds.), *Neolithic Settlement in Ireland and Western Britain* book section 3. (pp. 18–30). Oxford, UK: Oxbow Books.

Troy, C. S., MacHugh, D. E., Bailey, J. F., Magee, D. A., Loftus, R. T., Cunningham, P., Chamberlain, A. T., Sykes, B. C., & Bradley, D. G. (2001). Genetic evidence for near-eastern origins of european cattle. *Nature*, *410*, 1088.

Upadhyay, M. R. et al. (2016). Genetic origin, admixture and population history of aurochs (bos primigenius) and primitive european cattle. *Heredity*, *118*, 169.

Van Laere, A.-S. et al. (2003). A regulatory mutation in igf2 causes a major qtl effect on muscle growth in the pig. *Nature*, *425*, 832–836.

Vilà, C., Seddon, J., & Ellegren, H. (2005). Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends in Genetics*, *21*, 214–218.

Visser, M., Kayser, M., & Palstra, R. J. (2012). Herc2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the oca2 promoter. *Genome Research*, *22*, 446–455.

van Vuure, C. (2005). *Retracing the aurochs history, morphology and ecology of an extinct wild ox*. Pensoft Pub.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., Hakonarson, H., & Bucan, M. (2007). Penncnv: an integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome snp genotyping data. *Genome Research*, *17*, 1665–74.

Wang, M. D., Dzama, K., Rees, D. J. G., & Muchadeyi, F. C. (2016). Tropically adapted cattle of africa: Perspectives on potential role of copy number variations. *Animal Genetics*, *47*, 154–164.

Warburton, P. E., Hasson, D., Guillem, F., Lescale, C., Jin, X., & Abrusan, G. (2008). Analysis of the largest tandemly repeated dna families in the human genome. *BMC Genomics*, *9*, 533.

Wendorf, F., Close, A., & Schild, R. (). Early domestic cattle and scientific methodology. In L. Krzyz˙aniak, & M. Kobusiewicz (Eds.), *Proceedings of the International Symposinm organized by the Archaeological Commission of the Polish Academy of Sciences* (pp. 61–67). volume 2 of *Studies in African archaeology (Poznan´, Poland)*.

Wright, E. (2013). *The history of the European aurochs (Bos primigenius) from the Middle Pleistocene to its extinction: an archaeological investigation of its evolution, morphological variability and response to human exploitation*. Doctoral.

Xu, L., Hou, Y., Bickhart, D. M., Zhou, Y., Hay, E. H. a., Song, J., Sonstegard, T. S., Van Tassell, C. P., & Liu, G. E. (2016). Population-genetic properties of differentiated copy number variations in cattle. *Scientific Reports*, *6*.

Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nature Reviews Genetics*, *16*, 172.

Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, F., & Bendixen, C. (2011). Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics*, *12*.

Zhang, Z. D., Du, J., Lam, H., Abyzov, A., Urban, A. E., Snyder, M., & Gerstein, M. (2011). Identification of genomic indels and structural variations using split reads. *BMC Genomics*, *12*, 375.

Zhou, Y., Utsunomiya, Y. T., Xu, L., Hay, E. H. a., Bickhart, D. M., Sonstegard, T. S., Van Tassell, C. P., Garcia, J. F., & Liu, G. E. (2016). Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations. *Scientific Reports*, *6*.

Zimin, A. V., Kelley, D. R., Roberts, M., Marçais, G., Salzberg, S. L., & Yorke, J. A. (2012). Mis-assembled “segmental duplications” in two versions of the bos taurus genome. *PLoS ONE*, *7*, e42680.

Summary

Summary

A large diversity exists among European cattle breeds which can be attributed to population genomics forces such as migration, demography, and selection. These forces determine the distribution of variations such as single nucleotide polymorphisms (SNP) and structural variations (SV) in the genome. While the effect of these population genomics forces, as well as the distribution of genomic variations, have been studied extensively in commercial European cattle breeds, very few studies have focused on traditional and primitive cattle breeds of Europe. Therefore, this thesis aimed at providing a comprehensive overview of genomic admixture, demography, and variations in European cattle breeds using genome-wide SNP markers.

In chapter 2, the genome-wide SNP data indicated that many European cattle breeds display zebu and wild local aurochs ancestry in their genome. Therefore, the high divergence of Balkan and Italian cattle breeds can be attributed to zebu ancestry present in their genome. In this chapter, comprehensive overview of shared genomic variants was provided between wild local aurochs and different European cattle populations. These results suggested the possibility of several instances of intermating between wild local aurochs and ancestors of modern European cattle. Additionally, runs of homozygous genotypes also indicated that several Iberian and Italian cattle breeds have undergone a recent reduction in effective population size.

In chapter 3, we collated the genotyping data of various African and zebu cattle breeds available in public database to characterize African taurine and zebu ancestry in more detail. Additionally, we also made use of individual whole genome sequencing information to perform a comparative evaluation of genetic diversity in European cattle. The results showed that Italian cattle display high genetic diversity which can be attributed to the diverse founder population. In line with the previous studies, the influence of African taurine ancestry in Iberian cattle was also reported in this chapter. Further, we show that like central Italian cattle breeds, Busha also displays complex non-European ancestry—African taurine and zebu—, probably indicating the common origin of this ancestry in various Balkan and Italian cattle populations.

In chapter 4, genetic relatedness, diversity and demographic history of native Swedish cattle breeds were studied using genome-wide SNP markers. The results indicated that these cattle breeds could be classified into two genetic clusters: Swedish mountain cattle breeds (including Bohus Polled) and horned cattle breeds from southern Sweden. Interestingly, we also identified sub-substructure within a Fjallanra population which corresponded to their farm of origin. Moreover, the results also indicated a relatively high genetic diversity in Swedish mountain cattle. Additionally, low genetic diversity was observed in Vaneko and Ringamalako which can be attributed to their genetic isolation. Comparative evaluation of genotyping data of various northern-western European cattle breeds indicated little contribution in gene pool of native Swedish cattle.

In chapter 5, identification and distribution of copy number variations (CNVs) were assessed using signal intensity from SNP genotyping array. The results indicated differences in CNV profile between different European cattle populations, indicating the effect of demography and selection. Moreover, enrichment analysis showed that CNVs are enriched in the genomic regions

related to olfactory processes and immunity. Additionally, structural variation involving *KIT* genes, which has been associated with coat-color sidedness in Belgian blue cattle in the previous study, was also identified in English longhorn samples, indicating its role in coat color diversity in the population.

In chapter 6, three different approaches—paired-end read, split reads and read-depth—were employed to identify SVs in individual whole genome sequences of cattle. Similar to chapter 5, differences in SVs profile was observed between African taurine, zebu and European taurine, indicating the effect of demography. However, these results in parts can also be attributed to reference genome which is assembled from European taurine individual. The analysis also identified lineage-specific SVs in different cattle populations, indicating their role in population differentiation. Additionally, CNVs were also identified from ancient aurochs whole genome sequences using the read-depth approach. The results identified a high sharing of duplication event between aurochs sample and modern European cattle, which can be attributed to recent evolutionary split between these two populations. Finally, the approaches identified many SV associated with traits related to meat quality, coat color and metabolism, which can be targeted by future association studies.

Finally, in chapter 7, I discuss the findings of all the previous chapters in relation to what is known so far with respect to genomic admixture, diversity, and demography of European cattle. I also discuss the strength and limitations of the approaches employed in the thesis. I also put forward hypotheses of a demographic scenario involving complex admixture pattern in Balkan and Italian cattle breeds. I also discuss the possibility that zebu ancestry could be playing an important role by providing increased fitness to some European cattle. I also discuss the need of additional ancient aurochs samples representing different time periods and geographical ranges to answers the questions related to cattle domestication in Europe. I conclude by highlighting the practical implication of the information related to genetic diversity and structure in conservation management of native cattle breeds.

Samenvatting

Samenvatting

Er bestaat een grote diversiteit tussen Europese runderrassen welke toegeschreven kan worden aan diverse factoren die van invloed zijn op het genoom, zoals migratie, demografie en selectie. Deze factoren bepalen de verdeling van de waargenomen variaties in een genoom zoals puntmutaties ("single nucleotide polymorphism", SNP) en structurele variaties (SV). Terwijl het effect van deze factoren op het genoom alsmede de verdeling van de genoomvariatie uitvoerig zijn bestudeerd in commerciële Europese runderrassen, hebben slechts enkele studies zich gefocust op traditionele en primitieve runderrassen in Europa. Daarom is er in dit proefschrift getracht om met behulp van genoom-wijde SNP merkers een uitvoerig overzicht te verkrijgen van de demografie en genomische vermenging en variatie binnen Europese runderrassen.

In hoofdstuk 2, laat ik zien dat het genoom van veel Europese runderrassen, sporen vertoont van het Europese oeros en zebu. Ik veronderstel dat de hoge mate van divergentie van Balkan en Italiaanse runderrassen toegewezen kan worden aan het inkruisen van zebu in het verleden. In dit hoofdstuk, geef ik ook een uitgebreid overzicht van overeenkomstige genetische varianten tussen het wilde lokale oeros en diverse Europese runderpopulaties. Deze resultaten suggereren de mogelijkheid dat in meerdere gevallen een kruising tussen het wilde oeros en voorouders van de moderne Europese rundveerassen heeft moeten hebben plaatsgevonden. Verder toon ik ook aan dat verschillende Iberische en Italiaanse rundveerassen een recente reductie in effectieve populatiegrootte hebben ondergaan.

In hoofdstuk 3, combineer ik de genotypeerdata van diverse Afrikaanse en zebu runderrassen, waarvoor data voorhanden is in publieke databases en mijn data van de Europese runderrassen, en kom zo op een meer gedetailleerde karakterisatie van de mate van de genetische inbreng van zebu en Afrikaanse runderen in deze Europese rundverrassen. Daarnaast, heb ik gebruik gemaakt van complete genoomsequenties van individuele dieren voor een vergelijkende evaluatie van de genetische diversiteit van Europese runderen. Hiermee toon ik aan dat Italiaanse runderen de hoogste genetische diversiteit laten zien, wat toegeschreven kan worden aan een diversiteit voorouder populatie. De resultaten beschreven in dit hoofdstuk tonen ook een Afrikaanse taurine oorsprong aan in Iberische runderen en dit is in lijn met bevindingen uit eerdere studies. Verder laat ik zien dat, vergelijkbaar met Centraal Italiaanse runderrassen, ook Busa een complexe niet-Europese, Afrikaans taurine en zebu, oorsprong laat zien, mogelijk als gevolg van een gezamenlijke oorsprong van de diverse Balkan en Italiaanse runderen.

Hoofdstuk 4 beschrijft een studie naar de verwantschap, diversiteit en demografische geschiedenis van lokale Zweedse runderrassen op basis van genoom-wijde SNP merkers. De resultaten wijzen erop dat deze rassen geclassificeerd kunnen worden in twee genetische clusters: Zweedse bergrunderassen (inclusief Bohus Polled) en gehoornde runderrassen van Zuid Zweden. Interessant is ook dat ik in staat was om een substructuur aan te tonen binnen de Fjallanra populatie, welke correspondeert met de boerderij waarvan de dieren afkomstig zijn. De resultaten tonen verder een relatief hoge genetische diversiteit aan van de Zweedse bergrunderen. Dit in tegenstelling tot de lage genetisch diversiteit van de rassen Vaneko en Ringamalako, die verklaard kan worden uit de genetisch isolatie van deze rassen. Een vergelijking van de genotypeerdata van diverse noordwest Europese runderrassen toont een geringe bijdrage aan de genetisch diversiteit van lokale Zweedse runderen.

In hoofdstuk 5, beschrijf ik de identificatie en distributie van CNVs (afkorting voor de Engelse term “copy number variation”). De resultaten, gebaseerd op de signaal intensiteit van SNP genotypeerdata, tonen verschillen aan in de CNV profielen van verschillende Europese runderrassen waarschijnlijk het gevolg van verschillen in demografie en selectie. Een verrijkingsanalyse toont aan dat er meer CNVs gevonden worden in genoomregio’s gerelateerd aan reukzin en afweermechanismen. In het Engelse ras “English Longhorn” vond ik een structurele variatie rond het *KIT* gen, die eerder gevonden was in het runderras Belgische Blauwe. Deze structurele variatie speelt waarschijnlijk een rol bij de specifieke kleurvariaties binnen dit ras.

Hoofdstuk 6 beschrijft de resultaten van drie verschillende methoden voor het opsporen van structurele variaties in het genoom van runderen gebaseerd op complete genoom sequenties. Vergelijkbaar met de resultaten beschreven in hoofdstuk 5, worden verschillen in SV profielen gevonden tussen zebu, Afrikaanse en Europese runderen die wijzen op het effect van verschillen in demografie. Echter, deze verschillen kunnen voor een deel ook het resultaat zijn van het gebruik van het referentiegenoom, dat afkomstig is van een Europees individu. In de analyse, worden ook een aantal ras specifieke SV gevonden binnen diverse runderrassen, wat een rol hiervan suggereert in populatie differentiatie. Ook in het oeros sample konden CNVs aangetoond worden en deze resultaten tonen een hoge mate van overeenkomst met CNVs van moderne Europese runderen. Dit kan verklaard worden uit de vrij recente splitsing tussen deze twee populaties. De verschillende methoden resulteerden uiteindelijk in een groot aantal SV waarvan velen geassocieerd blijken te zijn met kenmerken als vleeskwaliteit, kleur van de vacht en metabolisme, allen goede targets voor toekomstige associatiestudies.

In hoofdstuk 7, tenslotte, bediscussieer ik alle resultaten die in de voorgaande hoofdstukken zijn beschreven in relatie tot de huidige kennis m.b.t. genomische vermenging, diversiteit en demografie van Europese runderen. Ik bediscussieer ook de sterkte en beperkingen van de diverse methoden zoals beschreven in dit proefschrift. Ik postuleer een aantal hypothesen voor demografische scenario’s gebaseerd op complexe patronen van genomische vermenging van Balkan en Italiaanse runderrassen. Daarnaast bediscussieer ik de mogelijkheid dat genetisch variatie afkomstig van zebu, een belangrijke rol speelt in een verhoogde fitness van sommige Europese runderen. Verder benadruk ik de noodzaak voor additionele DNA samples van het oeros die verschillend tijdsfasen en geografische gebieden vertegenwoordigen. Deze zijn nodig om een beter inzicht te verkrijgen van het domesticatieproces van het rund in Europa. Ik sluit het hoofdstuk af met het benadrukken van de praktische toepassing van de informatie m.b.t. genetische diversiteit en populatiestructuur voor het behoud en beheer van lokale runderrassen.

Sammanfattning

Sammanfattning

En stor fenotypisk variation förekommer hos Europeiska nötkreatursraser, som populationsgenetiskt kan hänföras till effekterna av migration, demografi och selektion. Dessa faktorer påverkar fördelningen av genetiska varianter, såsom SNP-markörer och strukturella variationer (SV), mellan olika populationer. Effekten av dessa faktorer samt distributionen av genetiska varianter är väl dokumenterade i kommersiella europeiska nötkreatursraser. Däremot är kunskapen begränsad om hur situationen är hos traditionella och primitiva europeiska nötkreatursraser. Denna avhandling syftar därför till att ge en ökad kunskap och en överblick om genomisk och genetisk variation hos de traditionella europeiska nötboskapsraserna med användning av SNP-markörer som täcker hela genomet.

I kapitel 2 visar jag att många europeiska nötboskapsraser har genvarianter som härstammar från både zebu och lokala uroxar. Jag föreslår att den höga divergens som finns hos boskapsraser från Balkan och Italien kan hänföras till zebu-inslag. I detta kapitel ger jag också en övergripande beskrivning av genomisk variation mellan vilda lokala uroxar och olika europeiska nötkreaturspopulationer. Dessa resultat indikerar ett signifikant genflöde mellan vilda lokala uroxar och de tidigaste populationerna av domesticerade europeiska nötkreatur. Dessutom visar jag att flera iberiska och italienska nötboskapsrasers effektiva populationsstorlekar nyligen har minskat.

I kapitel 3 samlade jag genotypdata för olika afrikanska nötkreatursraser samt raser av zebutyp som finns tillgängliga i publika databaser för att mer i detalj karakterisera graden av taurint (bos taurus)- respektive zebu (bos indicus)-ursprung i dessa boskapsraser. Dessutom använde jag genomsekvenser från specifika individer för att göra en jämförande utvärdering av genetisk mångfald i europeiska nötkreatursraser. Jag visar att italienska nötkreatur har hög genetisk variation som kan hänföras till att ett stort antal individer domesticerades. Inflytandet av afrikanska taurina nötkreatur som tidigare rapporterats hos iberiska nötkreatur bekräftades i detta kapitel. Även busa-djur från Balkan uppvisar ett komplext ursprung med genetiskt bidrag från både afrikanska taurin och zebu, vilket indikerar att vissa nötkreatursraser från Balkan och Italien har ett gemensamt ursprung.

I kapitel 4 studerades genetiskt släktskap, variation och demografisk historia hos alla kända svenska lantraser. Vi genotypade hela arvsmassan med en uppsättning av SNP markörer. Resultaten visade att dessa nötkreatursraser kan klassificeras i två genetiska grupper: svenska fjällkoraser (inklusive bohuskulla) och nötkreatursraser från södra Sverige (väneko och ringamålako) vilka båda har horn. Intressant nog kunde jag också påvisa understrukturer inom rasen fjällnära ko som motsvarade olika ursprungsgårdar. Dessutom indikerade resultaten också relativt hög grad av genetisk variation hos svenska fjällkor. Däremot observerades endast låg grad av genetisk variation hos både väneko och ringamålako vilket kan hänföras till deras genetiska isolering. Analyserna visade dessutom att de svenska lantraserna saknar eller har mycket lågt släktskap med andra analyserade nordvästeuropeiska nötkreatursraser och därmed har en unik uppsättning genetiska varianter.

I kapitel 5 identierades kopietalsvariation (copy number variation; CNV) och hur dessa är dis-

tribuerade hos olika europeiska nötkreatursraser. Det gjordes genom att analysera signalintensitet från SNP-genotypning av arvsmassan. Resultaten påvisade skillnader i CNV-profil mellan olika europeiska nötkreaturspopulationer, vilket indikerar effekter av demografi och selektion. CNV visades vara vanliga i genomiska regioner som innehåller gener med funktioner relaterade till olfaktoriska processer och immunsystemet. Dessutom identifierades hos rasen engelsk långhornad boskap strukturella variationer som involverar *KIT*-gener, som tidigare har förknippats med en specifik pälsfärg hos rasen belgisk blå.

I kapitel 6 användes tre olika analysmetoder av helgenomsekvensdata (paired end read, split read, samt läsdjup), för att identifiera strukturell variation (SV) i enskilda helgenomsekvenser av nötkreatur. I likhet med vad som observerades i kapitel 5 upptäcktes även skillnader i SV-profil mellan afrikansk taurin, zebu och europeisk taurin, vilket indikerar effekten av olika demografi. Resultaten kan delvis ha påverkats av att referensgenomet är från en europeisk taurin individ. Analysen identifierade också linjespecifika SV i olika boskapspopulationer, vilket indikerar att SV har en betydelse vid differentiering av populationer. Dessutom identifierades även CNVs i sekvenser som kommer från arkeologiska uroxe-prover. Resultaten visade mycket stora likheter mellan moderna europeiska nötkreatur och uroxen vilket bäst kan förklaras av den relativt korta tid som har förflutit sedan domesticeringen samt ett visst genflöde mellan populationerna innan uroxen dog ut. Sammanfattningsvis identifierades många SV relaterade till köttkvalitet, pälsfärg och metabolism. Resultat som kan vara värdefulla att beakta vid framtida genetiska associeringsstudier.

Slutligen diskuterar och sammanfattar jag i kapitel 7 resultaten från alla tidigare kapitel i förhållande till vad som hittills är känt rörande genomiska varianter och deras demografi hos europeiska nötboskap, d.v.s. hur populationer har korsats och därigenom påverkat varandra genom effekterna av s.k. admixture. Jag diskuterar också styrkan och begränsningarna i de metoder som används i avhandlingen. Jag lägger fram hypoteser om ett demografiskt komplext scenario där nötkreatur från Balkan och Italien har korsats och vilka effekter detta har haft. Jag diskuterar också möjligheten att inkorsning av zebu kan ha spelat en viktig roll genom att ge ökad fitness i vissa europeiska nötkreatursraser. Jag diskuterar också behovet av att sekvensera ytterligare uroxe-prover som representerar olika tidsperioder och geografiska områden för att svara på frågor som rör domesticering i Europa. Jag konkluderar med att belysa vilken praktisk implikation kunskapen om genetisk variation och struktur har i bevarandet av nötkreaturslantraser i olika länder i Europa.

Acknowledgements

Acknowledgements

I would like to take this opportunity to express my gratitude to all those people who have contributed to this thesis as well as supported me during this wonderful journey of my PhD. John Donne aptly remarked that “no man is an island”, indeed, human being can only thrive when he is a part of community. I feel very lucky that for the period of four years I was a part of this ABGC community at Wageningen University and Research (WUR). Moreover, as this PhD project was a collaborative effort of WUR and Swedish University of Agricultural Sciences (SLU), Uppsala, I also got opportunity to carry out part of my research in Department of Animal Breeding and Genetics at SLU which has enriched my professional expertise.

I am grateful to my WUR supervisors Prof. Dr. Martien Groenen and Dr. Richard Crooijmans for their guidance and constructive criticism on my scientific work which have shaped my scientific attitude. In fact, I also appreciate their patience which they have shown at a time when I was unable to meet deadlines. Without their faith in me, this thesis might not have seen the light of day. Overall, they have played a major role in nurturing the researcher out of me. I am also grateful to my SLU supervisors Prof. Dr. Göran Andersson and Dr. Sofia Mikko for their supervision and advices on analysis and wet laboratory procedures that I carried out during my PhD.

I am thankful to all the present and past members of weekly genomics meeting group—Ole, Hendrik-Jan, Kyle, Juanma, Mirte, Vinicius, Martijn, Chiara, Langqing, Zhou, Henri, Lim—for bearing my presentation and their scientific suggestions. Although I have learnt one thing or another from everyone in the group, I would like to single out Martijn from whom I learnt a lot in bioinformatics. He is the person I would give credit to for improving my Python skills. I also appreciate Vinicius for his help in R programming. My special thanks to Mirte for her scientific suggestions and advices in carrying out population genomics analysis. I also would like to thank other member of ABGC community—Gwen, Bert, Lissette, Ada,Maya—for their continuous support during my PhD. Thank you to the researchers of ABG dept at SLU—Susanne E., Anna, Susanne G., Naveed, Gabriela—-for their support during my PhD duration in Uppsala. Special thanks to Sandrine and Jovana for their guidance and suggestions regarding mobility and other matters related to European Graduate School in Animal Breeding and Genetics.

Thank you to all my friends in Wagenigen, especially, Priyanka Di,Sidhuji and Tanvi, for the delicious dinner that we often prepared and had together. Also, I would cherish the time that we spent together playing “dumb-charades” and board games. I would also like to thank the members of “chess club” Wageningen with whom I spent every Thursday evening of the first six months after I arrived in Wageningen. Thanks to ”freepik.com” for providing graphics for the thesis cover.

I am grateful to my ma and papa for their unconditional love and support throughout my career. I am also grateful to my grandma for her love and care. Special thanks to my sister Namrata and my brother-in-law Kunal for their love and support. In the end, I would like to thank my love Nehal for her constant support, care and company through think and thin of my life so far.

Curriculum vitae

Curriculum vitae

About the author

Maulik Upadhyay was born on 23'rd May 1988 in Junagadh district of India. After completing his bachelor's degree in veterinary sciences and animal husbandry (B.V.Sc. & A.H.) from Anand Agricultural University (AAU), he got enrolled for master's programme in animal breeding and genetics in the same university. During his masters, he also worked as a teaching assistant in the practical courses of Biostatistics and Animal Genetics which were offered to the undergraduate students. In his master's thesis, he analyzed whole genome sequencing data generated by Roche GS-FLX Titanium and Ion Torrent PGM sequencer to identify single nucleotide polymorphisms (SNPs) from buffalo genome. After completing his masters, he worked as a bioinformatician in the project of "Single Nucleotide Polymorphisms detection (SNP) in the coding region of the genome and its association with feed conversion ratio in broilers". In September 2014, he started his joint PhD under the aegis of the European Graduate School in Animal Breeding and Genetics (EGS-ABG). While most of his PhD research works were carried out at Wageningen (Netherlands), he had the opportunity to spend eight months at Uppsala (Sweden). During his PhD, he worked on characterizing genomic admixture and variation in native and primitive cattle breeds of Europe. The results of his PhD are presented in this thesis entitled "Genomic variation across European cattle: contribution of gene flow".

Peer-reviewed journal publications

M. R. Upadhyay,C. Bortoluzzi, M. Barbato , P. Ajmone-Marsan, L. Colli,J.A. Lenstra, C. Ginja, T. Sonstegard, M. Bosse , M.A.M. Groenen and R.P.M.A. Crooijmans, Deciphering the pattern of genetic diversity and admixture using Genome-wide SNPs in Southern European cattle (2019), Evolutionary Applications. doi:10.1111/eva.12770.

M. R. Upadhyay, V.H. da Silva, H.J. Megens, M.H.P.W. Visker, P. Ajmone-Marsan, V.A. Bâlteanu, S. Dunner, J.F. Garcia, C. Ginja, J. Kantanen, M.A.M. Groenen and R.P.M.A. Crooijmans, Distribution and Functionality of Copy Number Variation across European Cattle Populations (2017), Frontiers in Genetics, 8. doi: 10.3389/fgene.2017.00108.

M. R. Upadhyay, W. Chen, J.A. Lenstra, C.R. Goderie, D.E. MacHugh, S.D. Park, D.A. Magee, D. Matassino, F. Ciani, H.J. Megens, J.A.M. van Arendonk, P. Ajmone-Marsan, V.A. Bâlteanu, S. Dunner, J.F. Garcia, C. Ginja, J. Kantanen, M.A.M. Groenen and R.P.M.A. Crooijmans, Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle (2017), Heredity, 118(2), 169–176.

R. B. Onzima, **M. R. Upadhyay**, H.P. Doekes, L.F. Brito, M. Bosse, E. Kanis, M.A.M. Groenen and R.P.M.A. Crooijmans, Genome-wide Characterization of Selection Signatures and Runs of Homozygosity in Ugandan Goat Breeds (2018), Frontiers in Genetics.doi:10.3389/gene.2018.00318

R. B. Onzima,**M. R. Upadhyay**, R. Mukiibi, E. Kanis, M.A.M. Groenen and R.P.M.A. Crooijmans, Genome-wide population structure and admixture analysis reveals weak differentiation among Ugandan goat breeds (2018), Animal Genetics, 49, 59-70. doi: 10.1111/age.12631.

T.M. Shah, N.V. Patel, A.B. Patel, **M.R. Upadhyay**, A. Mohapatra, K.M. Singh, S.D. Deshpande, 0.G. Joshi, A genome-wide approach to screen for genetic variants in broilers (Gallus gallus) with divergent feed conversion ratio (2016), Molecular Genetics and Genomics, 291,1715–1725.

A.C.Patel, T.K.Jishaa, D. Upadhyay, R. Parikh, **M.R. Upadhyay**, R. Thaker, S. Das, J.V. Solanki, D.N. Rank, Molecular characterization of camel breeds of Gujarat using microsatellite markers (2015), Livestock Science, 181, 85-88.

M. R. Upadhyay, A.B. Patel, R.B. Subramanian, T.M. Shah, S.J. Jakhesara, V.D. Bhatt, P.G. Koringa, D.N. Rank & C.G. Joshi, Single nucleotide variant detection in Jaffrabadi buffalo (Bubalus bubalis) using high-throughput targeted sequencing (2015), Frontier in Life Sciences, 8(2), 192-199.

Manuscripts under review

M. R. Upadhyay, S. Eriksson, S. Mikko, E. Strandberg, M.A.M. Groenen and R.P.M.A. Crooijmans, G. Andersson and A.M. Johansson, Genomic relatedness and diversity of Swedish native cattle breeds. Genetics Selection Evolution (2019).

M. R. Upadhyay, M.F.L. Derks, G. Andersson, M.A.M. Groenen and R.P.M.A. Crooijmans, Comparative evaluation of structural variations in taurine and indicine cattle using individual whole genome sequences. BMC Genomics (2019).

M. Barbato, M.D. Corvo,**M. R. Upadhyay**, E. Kim, F. Hailer, L. Colli, R. Negrini, R.P.M.A. Crooijmans, T.Sonstegard, P. Ajmone-Marsan, Signals of adaptive introgression in white cattle breeds from Central Italy. Evolutionary Applications (2019).

Training and Supervision Plan

Training and Supervision Plan (TSP)	**Graduate School WIAS**

Section 3. EDUCATION AND TRAINING (minimum 30 credits)

A. The Basic Package	year	credits *
WIAS Introduction Day **(mandatory)**	2014	0.3
Course on philosophy of science and/or ethics **(mandatory)**	2014	1.5
European Graduate school of Animal Breeding and genetics (Introduction week)	2014	2.0
Subtotal Basic Package		**4**

B. Disciplinary Competences	year	credits
Introduction to Phylogenetics analysis using R	2014	2.0
The sustainabiity concept in Animal Breeding	2015	2.0
Data management and Planning	2015	0.4
Emerging technologies in Animal Breeding	2017	1.5
Population Genomics: background and tools	2017	3.0
Code club meeting of Wagenignen University and Research	2017-2018	1.0
HPC (High performance computing) Advanced Couse	2018	0.2
Summer school France	2018	0.8
Linear Models in Animal Breeding	2018	3.0
HPC Basic course	2018	0.2
WIAS Course Statistics for the Life Sciences	2018	2.0
M.Sc. Level course		
Programming in Python	2014	6.0
Subtotal Disciplinary Competences		**22**

C. Professional Competences	year	credits
Course on essential skills (Frank Little) (recommended)	2014	1.2
Techniques for Writing and Presenting a Scientific Paper (TWP)	2015	1.2
Effective behavior in your professional surroundings (EB)	2016	1.3
project and Time management	2016	1.0
Writing a grant proposal	2017	2.0
Searching and Organizing literature	2018	0.6
How to beat Procastination	2018	0.3
Subtotal Professional Competences		**8**

D. Presentation Skills *(maximum 4 credits)*				year	credits
< title of presentation, name of conference/seminar, place, date, oral / poster >					
1). Genetic structure of primitive European cattle breeds and aurochs	Wageningen PhD symposium	Wageningen, 26th April, 2016	Oral	2016	1.0
2). Genetic structure of primitive European cattle breeds and aurochs	Mini-symposium on the search for genotypic, phenotypic & ecotypic equivalents of an extinct species	Uppsala, Oct.6, 2016	Oral	2016	1.0
3). Genetic differentiation of primitive cattle breeds illustrate aurochs admixture	50th Population Genetics Group Meeting	Cambridge, 4th-7th January, 2017	Oral	2017	1.0
4). Characterization of copy number variation in European cattle	International Conference on Animal Genetics, 2017	Dublin, 16th-21st July, 2017	Poster	2017	1.0
5). Inferring pattern of genomics admixture in Southern Europen cattle breeds	Joint Congress on Evolutionary Biology, 2018 DrosEU Summer School on Adaptation Genomics	Montpellier, France, 23'rd-24th August, 2018	Poster	2018	0.0
Subtotal presentations					**4**

E. Teaching competences *(max 6 credits)*	year	credits
Supervising thesis (B.Sc.)	2015	1.0
Tutorship	2015	1.8
Subtotal Teaching competences		**3**

Education and Training Total (minimum 30 credits)*		**40**

Data availibility and supplementary material

Data availibility and supplementary material

All the genotyping and signal intensity data have been submitted in DRYAD with links provided in the manuscripts published based on the chapter 2 and chapter 5 of this thesis.

The supplementary material for the chapter 2 and chapter 5 are available through the journal websites:

chapter 2: (https://www.nature.com/articles/hdy201679#supplementary-information)
chapter 5: (https://www.frontiersin.org/articles/10.3389/fgene.2017.00108/full#h11)

The supplementary materials for the unpublished articles are uploaded on google drive, please visit this link to download the supplementary materials:

https://drive.google.com/file/d/1JxUqg25NKEG1ZRXnLGvot8Bb3n4_QMfV/view

Colophon

Colophon

The work performed in Chapter 4 was financed by the funds of Swedish Research Foundation (FORMAS).

The author was supported by the European Commission (within the framework of the Erasmus-Mundus joint doctorate "EGS-ABG") and Taurus foundation (Stichting Taurus), Nijmegen, The Netherlands.

The cover of this thesis was designed by Nehal Ranabhatt and Razia Sultana.

The thesis was printed by Digiforce — Proefschriftmaken.nl, De Limiet 26, 4131NC, Vianen, the Netherlands.