



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이 학 석 사 학 위 논 문

Bioinformatics research using next generation sequencing data to compare genomic features between populations and a review of future genetic resources

집단 간의 유전체 비교를 위한 차세대 시퀀싱 데이터를 이용한
생물정보학적 연구 및 미래 유전자원에 대한 리뷰

2014년 8 월

서울대학교 대학원

협동과정 생물정보학전공

권 영 준

**Bioinformatics research using next generation sequencing
data to compare genomic features between populations
and a review of future genetic resources**

By

Young-jun Kwon

Supervisor: Professor Heebal Kim

August, 2014

Interdisciplinary Program in Bioinformatics

Seoul National University

Abstract

Bioinformatics research using next generation sequencing data to compare genomic features between populations and a review of future genetic resources

Young-jun Kwon

Interdisciplinary Program in Bioinformatics

Seoul National University

Genome is a blueprint of organism, so there have been a lot of studies trying to describe phenomenon of life through genome analyses. Given recent development of Next Generation Sequencing (NGS) and reduction of sequencing cost, analyses of whole genome sequencing data between populations became financially possible in individual labs. From gathered genomic variations of individuals, genomic features of each population can be described and compared, and moreover, phenotypic differences between populations can be explained in a view of DNA sequences. Livestock of which characteristics are various have been domesticated for a long time, and signatures of domestication remains on their genome sequences. Thus, those animals have been considered to be good research models and researched by many researchers.

In Chapter 1, I simply summarized and reviewed history of DNA sequencing and NGS technologies. Also, methods for comparative analyses between populations especially which are used in this study such as methods to define positive selection regions were briefly reviewed.

In Chapter 2, I analyzed two of goat populations to compare their genomic contents and understand distinct genetic features of patterns of selection in the populations using NGS whole genome sequence data (Illumina platform) from 15 Korean native goats and 11 crossbred goats. Because goats are the closest animal to cattle, I researched as preceding analyses for the research of African cattle project which I will cover on Chapter 3.

I used GATK software to perform variant calling, and the called SNPs data was used in this study. From the identified variants in two goat populations, I discovered that Korean native goats evidently have lower nucleotide diversity than crossbred goats. Also, I revealed that ancestry of their common ancestor still hold a large majority in both of populations using STRUCTURE although they were clearly distinguished. The results of genetic structure of goat populations also supported the information of population which was already known.

Through analysis using XP-EHH statistics, which is one of the methods to define positive selection regions, 6 candidate genes including predicted genes which show signatures of strong positive selection were

identified.

I also performed gene ontology term analysis, and DAVID tool was used in this study. GO term analyses with genes in highly variable regions showed the potential chances of gaining variations in genes involved with olfactory and neurological systems, but in practice, GO term analyses with non-synonymous SNP revealed that genes involved with olfactory and immune systems show differences between Korean native goat and crossbred goat populations.

In Chapter 3, I reviewed African indigenous cattle as future genetic resources. African indigenous cattle show varieties of characteristics with the result that they have been adapted to various and harsh environmental conditions. I investigated phenotypic and genetic features of African indigenous cattle and described the necessities of research on African indigenous cattle as future genetics resources. I also investigated previous researches on African indigenous cattle and briefly describe expected goals of researches on African cattle in the future.

Key words: Next generation sequencing, genetic comparison, population analysis, genetic resource, African cattle

Student number: 2012-23110

Contents

Abstract	I
Contents.....	IV
List of tables	V
List of figures	VI
Chapter 1. Literature review	1
1.1 Developments in next generation sequencing.....	2
1.2 Re-sequencing analyses using next generation sequencing data	6
Chapter 2. Comparative genome analysis between Korean native goat and crossbred goat populations using next-generation sequencing.....	14
2.1 Abstract	15
2.2 Introduction	16
2.3 Materials and methods.....	19
2.4 Results.....	24
2.5 Discussion	29
Chapter 3. African Indigenous Cattle and their Potential as Genetic Resources	51
3.1 Abstract	52
3.2 Introduction	53
3.3 African cattle.....	55
3.4 Researches on African indigenous cattle	68
3.5 Conclusion and perspectives.....	75
References	90
(요약) 국문초록.....	106

List of tables

Table 2.1. Summary of alignment rate in 26 goat samples	43
Table 2.2. Genome coverage of mapped reads for each sample	45
Table 2.3. Identified variants in 15 Korean native goats and 11 crossbred goats	46
Table 2.4. Variants in genic regions and intergenic regions	46
Table 2.5. List of candidate genes identified from positive selection scan	46
Table 2.6. Breed-specific non-synonymous SNP sites.....	47
Table 2.7. Breed-specific genes in the top 5% of highly variable regions.	49
Table 3.1. Uses and Characteristics of Indigenous African Cattle by African Countries (Excluding Commercial Components)	82

List of figures

Figure 2.1. Male and female Korean native and crossbred goats.....	34
Figure 2.2. Distribution of SNVs in the Korean native goat genome	35
Figure 2.3. Nucleotide diversity plots of Korean native goats (top) and crossbred goats (bottom).	36
Figure 2.4. Comparison nucleotide diversity between Korean native goat and crossbred goat.....	37
Figure 2.5. SNV density of Korean native goat genome and crossbred goat.	38
Figure 2.6. Result of Principal Component Analysis.	39
Figure 2.7. Population structure of goat breeds assessed using STRUCTURE	40
Figure 2.8. Candidate genomic regions of positive selection.....	41
Figure 2.9. Representative biological process GO terms and KEGG pathways associated with genes located in the top 5% of highly variable regions in Korean native goat and crossbred goat populations ($p < 0.05$).	42
Figure 3.1. Approximate migration route and the origin of Africa domestic cattle (adapted and modified from ILRI, 2006. Safeguarding livestock diversity: the time is now).	77
Figure 3.2. Distributions of indigenous cattle in sub-Saharan Africa	78
Figure 3.3. Summary of impacts from the loss of cattle in African agricultural systems.	79
Figure 3.4. Comparison of total area harvested among countries in Sub-Saharan Africa (FAOSTAT)	80

Figure 3.5. Comparison of total cereal production among countries in Sub-Saharan Africa (FAOSTAT)	80
Figure 3.6. African cattle genome consortium	81

Chapter 1. Literature review

1.1 Developments in next generation sequencing

1.1.1 The history of DNA sequencing methods

After DNA double helix structure had been discovered by Watson and Crick in 1953 (Watson and Crick 1953), Decoding of nucleotide sequence of organisms became the primary issue among biologists. During 1970s, there were noticeable advancements in DNA sequencing. Sanger and Coulson introduced a method to detect DNA sequences which was based on primed synthesis in 1975 (Sanger and Coulson 1975). However, this method had a limitation that only ~50 bases could be decoded in a single experiment due to the limited space of gel, so it was unsuitable for getting longer DNA sequences.

To overcome the limitation in previous DNA sequencing method, a new method was introduced. Frederick Sanger developed rapid DNA sequencing methods and published “DNA sequencing with chain-terminating inhibitors” in 1977 (Sanger, Nicklen et al. 1977). This method is based on the incorporation of chain-terminating dideoxynucleotides during DNA replication. There are four separated sequencing reactions, and only one type of labeled dideoxynucleotides (ddATP, ddGTP, ddCTP, ddTTP) is added in each reaction while the other nucleotides are standard deoxynucleotides (dATP, dGTP, dCTP, dTTP). Sequencing gel has 4 lanes for each nucleotide (A, T, G, C), and the band of labeled ddNTPs represent the DNA sequence. In 1977, Gilbert and Maxam also introduced DNA sequencing methods including

chemical degradation methods (Maxam and Gilbert 1977).

In early 80's, a non-radioactive DNA sequencing method was developed by Pohl and co-workers, this comparatively safe sequencing method brought out development of automated DNA sequencing. In 1987, the first fully automated sequencing machine was announced, and in the same year, the sequencing machine Dupont's Genesis 2000 which used a novel fluorescent labeling technique was developed. Because fluorescent labeling technique enabled to identify all the types of dideoxynucleotides in a single lane, large-scale sequencing trials on microorganism began in 1990. From a genome sequence of bacterium (Fleischmann, Adams et al. 1995) to sequence of the human genome (Lander, Linton et al. 2001), whole-random sequencing method which was based on Sanger sequencing method had been used to identify the genome sequence of various organisms.

From the middle of 1990s, new DNA sequencing methods had been developed, and these new methods brought about large changes in genome sequencing. In 1996, pyrosequencing method was published, and in 2000, "Massively parallel signature sequencing (MPSS)" method was published (Brenner, Johnson et al. 2000). This adapter/ligation-mediated and bead-based sequencing method was the first commercial sequencing method of "next generation". In 2004, a parallelized pyrosequencing was commercially served by 454 Life Sciences, and this new next generation sequencing method made sequencing cost 6-fold lower compare to automated Sanger sequencing.

1.1.2 Comparison of DNA sequencing methods

There have been also other developments in commercial next generation sequencing methods, for examples, single-molecule real time sequencing (Pacific Bio), ion semiconductor method (Ion Torrent sequencing), and ligation-based sequencing (Illumina), after demand for low-cost and high-throughput sequencing had increased, and DNA sequencing methods are still under development. Each sequencing method has pros and cons compare to other sequencing methods.

Existing chain termination method (Sanger sequencing) provides the longest read length and shows highest accuracy, but the cost for sequencing is much more expensive than any other methods (Liu, Li et al. 2012). Therefore, chain termination sequencing is not suitable for large-scale sequencing projects. On the other hand, Illumina sequencing provides comparatively short read length and has 1 to 2 percent of sequencing errors per read, but it costs cheaper than other methods (Liu, Li et al. 2012), so it has potential for sequencing projects of large-scale.

Amount of throughput and sequencing cost are typical differences between first generation sequencing (Sanger sequencing method) and next generation sequencing method. Cloning process with plasmid in *E.coli* contributed to these differences. Unlike Sanger method, in next generation sequencing methods, there is no need to construct cloning libraries of DNA fragments using plasmid of *E.coli* (Mardis 2008). For example, in sequencing by synthesis method (Illumina), “Bridge PCR” technique is used. Without

plasmid library, common adapters are covalently ligated to DNA fragments, and the adapters enable DNA fragments are amplified by polymerase and library can be made up in the sequencing machine (Liu, Li et al. 2012). Pyrosequencing (454) and ligation sequencing (SOLiD) use “Emulsion PCR” method which uses in situ bead or glass microfluidic channel to amplify DNA fragments (Liu, Li et al. 2012).

1.2 Re-sequencing analyses using next generation sequencing data

1.2.1 Variant genotyping based on reference genome sequence

One of representative research types using next generation sequencing data is re-sequencing analysis. With previously defined reference sequence, genetic variation including SNP or INDEL between individuals can be identified by whole-genome or targeted re-sequencing (DePristo, Banks et al. 2011). Genotypes of individual can be determined after randomly fragmented DNA reads are aligned to a reference genome sequence, and given the situation that sequencing cost rapidly decreases, high throughput next generation sequencing enabled us to produce genome sequencing data of population level.

Basically, DNA sequences contain the information of protein sequences, and mutation in DNA sequences can bring about different sequence of amino acids. Malformed amino acid sequence usually affect individual negatively, however, there are rare cases that mutated protein sequence gained some fitness so that the DNA mutation remains as a polymorphism of a population (Back 1994). Genetic variants including SNP in each individual can reflect phenotypic variance between entities or different groups, and discovery of the relation between phenotypic differences and genotype variance is one of the main purposes in biological researches.

In case of human, estimated nucleotide diversity is 0.1% to 0.4%, and these slight amounts of genetic variations make individual differences (Goldstein

2009). To uncover detailed genetic factors which are involved in human diseases, “The 1000 Genomes Project” launched as an international research in early 2008. In this project, genetic variation data of about a thousand people who are from different area in the world was gathered through high throughput sequencing platforms (Consortium 2010). Like human research field, animal research communities also have been trying to build database for genetic variants. For example, 1000 bull genomes project is ongoing, and it aims to genotype full sequences in various cattle breeds for genome wide association studies and genomic prediction.

Reference genome sequences for many other species have been produced and ongoing as genome analyses have been developed, and there was a remarkable advance in complete genome sequence field (Pruitt, Tatusova et al. 2012). Based on current UCSC genome browser (Karolchik, Baertsch et al. 2003), a total of 95 reference genome sequences for 68 species of vertebrates including 65 vertebrates which are enable to use complete annotation sets, 3 species of deuterostomes, 15 species of insects, 6 species of nematodes and 3 other category species including yeast are publically able to approach and use for academic researches while the UCSC database in 2009 provided references for only 47 organisms including 24 vertebrates.

1.2.2 Methods for discovery of positive selection

[High proportion of functional mutations]

A new genetic variant that alter amino-acid of protein is normally deleterious and hardly has a proportion in a population. However, when a mutation have beneficial effect to individuals, fixation rate of the variant will increase compared to mutations that bring about no amino-acid substitution (i.e., silent mutations). Signatures can be measured by DNA sequence comparison between species, and these methods are usually able to detect positive selection events about millions years ago, and dN/dS ratio (Yang and Bielawski 2000), and McDonald-Kreitman test (McDonald and Kreitman 1991) are statistical tests that are commonly used.

However, only ongoing or recurrent selection can be detected through this statistical method because to be selected as a strong signature of positive selection against the neutral selection, multiple substitutions will be needed for a gene (Sabeti, Schaffner et al. 2006). To overcome this limited power, sets of multiple genes can be used (Nielsen, Bustamante et al. 2005).

[Reduction in genetic diversity]

A selective sweep is phenomenon of reduction of variation in a certain genomic region due to strong positive natural selection of a mutation. When a new mutation positively contributes to survival of individuals of a population, the mutated variant will increase in frequency. Neighboring DNA sequence of

the mutation will be also affected and become more commonplace. A strong selective sweep drives this phenomenon, which is called *genetic hitchhiking*, and a large reduction of the genetic variation in that region.

Statistical methods which are based on genetic diversity usually can detect signatures within 250,000 years ago, and Tajima's D (Tajima 1989) and Fu and Li's D (Fu and Li 1993) are commonly used.

The strength of positive select affects the size of selective sweep region (Sabeti, Schaffner et al. 2006), and that means the higher speed of positive selection will lead the larger size of sweep regions.

Previous research suggested the general size of sweep region as roughly 600kb when an allele contributes 1% of selective advantage, which is considered strong selection (Sequencing and Consortium 2005). This size is too large to specify the region under positive selection. Moreover, if there were demographic history (e.g., and expanding of population), it will be difficult to distinguish whether the signature indicate a true positive selection or not.

[High-frequency of derived alleles]

A new mutation generates derived allele, and the derived allele is usually less frequent than ancestral alleles which already exist (Watterson and Guess 1977) unless they are dragged by allele which is beneficial (hitchhiking). Thus, in the region under strong positive selection, derived alleles will be frequently presented, and by catching this signature, selective sweep regions can be

measured (<80,000 years).

Fay and Wu's H test (Fay and Wu 2000) is the most common method based on derived allele frequency to detect positive selection, and knowledge of ancestral allele is essential to perform the test. The ancestral alleles can be deduced from the allele type in a species which was derived from most recent common ancestor (MRCA). In this manner, there is the assumption that only one mutation event happened at this nucleotide position after the two related species had been derived (Fay and Wu 2000).

[Differentiation between populations]

After environmental or cultural adaptation happened, the positive selection pressure can bring about the increase in allele frequencies for the locus which was selected compared to other alleles. Then, it will cause higher level of differentiation between populations than expected level (Akey, Zhang et al. 2002).

Among methods which are based on allele frequencies, F_{ST} (Barrett and Schluter 2008) has been commonly used to identify positive selection, and Cross-population Composite Likelihood Ratio Test (XP-CLR; Chen, Patterson et al. 2010) was also introduced recently. These methods are based on allele frequency to detect selective sweep region and able to detect selection events about less than 50,000 to 70,000 years ago (Sabeti, Schaffner et al. 2006).

However, there is a limitation that only when populations are

isolated and have no genetic interchange, population differentiation can exist. Moreover, through this kind of methods, it will be difficult to distinguish between demographic history (e.g., population bottle necks) and true positive selection (Sabeti, Schaffner et al. 2006).

[Long haplotypes]

Haplotype is a combination of alleles in a chromosome that are represented in individuals, and size of haplotype depends on the frequency of recombination events. Measuring linkage disequilibrium through haplotype is one of strategies to detect selective sweep region. Usually, different alleles within a haplotype will be rearranged due to genetic recombination, and a single haplotype will not dominate population under neutral evolution. However, with a strong positive selection force, selected a given haplotype will be predominant and have larger range in the target population rapidly enough that recombination events are hardly able to break the haplotype.

Although haplotype based methods can detect signatures of positive selection within only 30,000 years ago because recombination events keep breaking down haplotypes, it is a useful strategy to identify partial selective sweep (~10% of allele frequencies) and a narrow region that was positively selected.

The long-range haplotype (LRH) test (Sabeti, Reich et al. 2002), integrated haplotype score (iHS; Voight, Kudaravalli et al. 2006) and cross-

population extended haplotype homozygosity (XP-EHH; Sabeti, Varilly et al. 2007) are the methods that are commonly used, and especially, with NGS sequencing data, (XP-EHH) is the most commonly used method for detecting selective sweep regions.

1.2.3 Genetic structure analyses

When individuals are separated from the previous population and interbreeding between populations is restricted (e.g., geographical barrier such as isthmus or river), genetic admixture occurs, and the introduction of new alleles from other population will lead differences in genetic admixture. Therefore, through genetic structure of populations, the history of the populations of ancestors can be deduced.

To clarify genetic structure of populations, which is a degree of similarity of genetic data between different individuals and groups, Principal component analysis (PCA) have been mainly used (Jackson 2005). In this study, GCTA(Yang, Lee et al. 2011) for PCA and STRUCUTRE (Evanno, Regnaut et al. 2005) for estimation of individual ancestries were used with SNP genotype data using NGS sequencing.

1.2.4 GO-term analyses

Gene ontology originated in a consortium studying fruit fly, mouse and yeast in 1998 (Ashburner, Ball et al. 2000). Databases of many organisms have joined together, and it provides defined terms which represents genes that researchers want to know.

In bioinformatics, functional annotation tools are frequently used to extract representative features of genes which researchers are interested in and categorize cellular component, molecular functions biological process, and even KEGG pathways. There are many tools available for gene ontology analyses online and off-line including DAVID tool (Dennis Jr, Sherman et al. 2003). In case of DAVID tool, the results are organized by forms of chart or table and can be downloaded as summarized form.

Chapter 2. Comparative genome analysis between Korean native goat and crossbred goat populations using next-generation sequencing

2.1 Abstract

A study on goat populations using whole-genome sequencing data has not been reported since the goat reference genome was published. This study was performed to examine distinct genetic features and patterns of selection in Korean native goats by comparing the genome between Korean native goat and crossbred goat populations. I got the whole genome data of 15 Korean native goats and 11 crossbred goats using next-generation sequencing (Illumina platform) to perform comparative genome studies between populations; single nucleotide polymorphism (SNP) and insertion and deletion (INDEL) regions were identified in the Korean native goat and crossbred goat populations. I found that nucleotide diversity of Korean native goats decreased compared to crossbred goats. Genetic structural analysis demonstrated that the Korean native goat and crossbred goat populations shared common ancestry, but were clearly distinguished. Gene ontology term analysis with genes involved in nonsynonymous SNPs revealed that Korean native goats and crossbred goats showed differences in olfactory sensors and the immune system. Six selective sweep signals located in genic regions were observed in the Korean native goat genome using cross-population extended haplotype homozygosity (XP-EHH).

2.2 Introduction

Goats (*Capra hircus*) are one of the oldest domesticated animals, and their domestication started in western Iran about 10,000 years ago (Zeder and Hesse 2000). Goats have been used for milk, meat, wool, and hide since the animals were domesticated.

Due to their large differences in body size, weight, muscle mass, milk production, and coat color (GUIDE 2004), goats are good model animals for genetic studies. However, only few studies have analyzed the genetic resource of goat. Moreover, studies on goat populations using the whole-genome sequence have not yet been reported, while some studies have explored breed diversity using SNP chips (Kijas, Ortiz et al. 2013). To explore the goat genome, genetic studies using the whole-genome sequence are required.

Variants [SNPs and insertions and deletions (INDELs)] from whole-genome resequencing data can reveal different genetic features between populations, which could allow us to infer how each population has adapted to environments based on evolutionary evidence such as signatures of selection. I can also explore differences in the amino acid sequences between populations to examine phenotypic variations since amino acid substitutions can affect the phenotypes of individuals, such as Mendelian traits (Yutani, Ogasahara et al. 1977).

Some genes in each goat population may have been naturally selected by fitness or artificially selected through domestication, and the gene regions can be identified using selective sweep region signals. Cross-population extended haplotype homozygosity (XP-EHH) is a representative method for detecting selective sweep regions (Sabeti, Varilly et al. 2007). This method calculates haplotype decay separately for cross-populations using EHH. In addition, when the selective sweep region is fixed, it is sensitive and more powerful than the integrated haplotype score (iHS) when the sample size is small (Pickrell, Coop et al. 2009).

Previous studies on the goat genome were performed based on SNP chip data, and only known SNP sites could be used. However, I could examine known and unknown SNPs using next-generation sequencing (NGS) whole-genome sequencing. In addition, one can examine SNPs located in genic or intergenic regions. Thus, for genome research, generating NGS whole-genome sequencing data is necessary, rather than SNP chip data.

I performed these comparative analyses of whole genomes using goat samples of the Korean native goat and crossbred goat populations (**Figure 2.1**). Korean native goat is known to have been inbred to maintain purebred lines, and the main weak point of this breed is their deficient supply of milk to feed their calves. In order to remedy this, crossbreeds between Korean native goat and imported dairy goat breeds were made. As Saanens are known to be the largest goat breed among dairy types and good milk producers (Boichard, Bouloc et al. 1989), Saanen goats were mostly used to

form the crossbred goat group. In addition, based on the appearance of the crossbred goat breed, the Boer goat, which is developed for meat production, also seemed to be used for crossbreeding. These crossbred black goats in Korea not only have a higher milk yield, but also a higher growth rate than Korean native goats. In addition, when fully grown, the crossbred goats are bigger than the Korean native goat.

In this study, I attempt to discover the different genetic features between Korean native and crossbred goat breeds and elucidate the genetic components that have contributed to these phenotypic differences using whole-genome next generation sequencing data.

2.3 Materials and methods

2.3.1 Sample preparation and whole-genome sequencing

Whole-blood samples were collected from 11 crossbred goats and 15 Korean native goats from Korea. Blood (10 mL) was drawn from the carotid artery and treated with heparin to prevent clotting. DNA was isolated from whole blood using a G-DEXTMIIB Genome DNA Extraction Kit (iNtRoN Biotechnology, Seoul, Korea) according to the manufacturer's protocol. 3 µg of genomic DNA using the Covaris System was randomly sheared to generate inserts of about 300 bp. The fragments of sheared DNA were end-repaired, A-tailed, adaptor-ligated, and amplified using a TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA). Paired-end sequencing of goat genomes to about tenfold coverage were conducted at NICEM (National Instrumentation Center for Environment Management, Seoul, Korea) using the Illumina HiSeq2000 platform with TruSeq SBS Kit vs-HS (Illumina).

2.3.2 Sequencing alignment and variant calling

Approximately 220,000,000–230,000,000 paired-end reads were mapped to the reference goat genome (Dong, Xie et al. 2013) from the Goat Genome Database Web site (<http://goat.kiz.ac.cn/GGD>) using Bowtie 2-2.1.0 with the default settings (Langmead and Salzberg 2012). Open-source packages were used for downstream processing and variant calling for SNPs and INDELs,

including Picard tools 1.94 (<http://picard.sourceforge.net>), SAMTools 0.1.19 (Li, Handsaker et al. 2009), VCFtools 4.0 (Danecek, Auton et al. 2011), and the Genome Analysis Toolkit (GATK);(McKenna, Hanna et al. 2010) 2.6.4 version. Read Groups were added and duplicate reads were removed, and all mate-pair information between each read was confirmed using the module MarkDuplicates, and FixMateInformation of Picard tools. SAMtools was used to index the resulting bam format files and calculate the mapped read length with the flagstat option (Li, Handsaker et al. 2009). The modules RealignerTargetCreator, IndelRealigner, UnifiedGenotyper, SelectVariants, and VariantFiltration in GATK were used for realignment and variant calling. VCFtools was used for handling the VCF format files (Danecek, Auton et al. 2011). Substitution calls were made using the GATK UnifiedGenotyper module (McKenna, Hanna et al. 2010). SNPs and INDELs called with a Phred-scaled quality score of <30 were filtered out. Variants were removed based on MQ0 (mapping quality score 0) >3 , (MQ0/read depth) $>0.1(10\%)$, quality depth <5 , and FS (Phred-scaled p -value using Fisher's exact test to detect strand bias) >200 . After filtering, the SNPs were filtered out again by removing those within 10 bp of INDELs. For the last SNP filtering step, only biallele-type SNPs were considered real SNPs. For haplotype information on each chromosome, I used BEAGLE (Browning and Browning 2007) to infer the haplotype phase and impute missing alleles simultaneously for the entire set of goat populations.

2.3.3 Statistical analysis

To identify and compare highly variable regions, the distribution of variants and nucleotide diversity information was obtained based on the goat reference genome (Dong, Xie et al. 2012) from the Goat Genome Database Web site (<http://goat.kiz.ac.cn/GGD>). VCFtools was used to count the number of variants and nucleotide diversity in each 1-Mb bin region (Danecek, Auton et al. 2011).

2.3.4 Variant annotation

To identify SNPs that contribute to differences in phenotypes between Korean native goats and crossbred goats, VCF format files of candidate SNPs were separately concatenated using VCFtools (Danecek, Auton et al. 2011). Variant regions were then annotated using SNPEff (Cingolani, Platts et al. 2012) with a GFF format file (Dong, Xie et al. 2012) from the Goat Genome Database Web site (<http://goat.kiz.ac.cn/GGD>). I filtered out variants that were not population-specific, and I searched for breed-specific nonsynonymous genes. To identify genes that have distinguished allele types between Korean native goats and crossbred goats, protein IDs of the Ensembl Genome Browser (<http://www.ensembl.org>; Hubbard, Barker et al. 2002) in an annotated file were used.

2.3.5 Principal components analysis (PCA)

I performed PCA (Jackson 2005) to examine population differentiation between Korean native goats and crossbred goats using genotype data from 15 Korean native goats and 11 crossbred goats. I used GCTAtool (Yang, Lee et al. 2011), which implements PCA in EIGENSTRAT (Price, Patterson et al. 2006), to estimate eigenvectors. VCFtools (Danecek, Auton et al. 2011) and PLINK (Purcell, Neale et al. 2007) were used to prepare input data sets for GCTAtool (Yang, Lee et al. 2011).

2.3.6 Admixture analysis

I used STRUCTURE software (Evanno, Regnaut et al. 2005) to identify genetic proportions of each goat individual from ancestral populations. STRUCTURE software used Bayesian algorithms to detect the true number of clusters, K (the number of ancestral populations). Beagle was used to generate input files for running STRUCTURE using SNVs in every 50 SNVs located in only intergenic regions to avoid bias from evolutionary gene acceleration. I used 100,000 iterations with 2,000 burn-in iterations in each analysis from K = 2 to 4.

2.3.7 XP-EHH

I used the XP-EHH method (Sabeti, Varilly et al. 2007) to test for signatures

of selective sweeps, which calculates haplotype decay separately for each group using the EHH (Sabeti, Reich et al. 2002). The test then integrates these values with respect to genetic distance and calculates a log ratio iHH of the Korean native goats and crossbred goats. The calculation for XP-EHH was performed using the software xpehh ((Sabeti, Varilly et al. 2007); <http://hgdp.uchicago.edu/Software/>). I assumed that genetic distance was equal to physical distance. These log ratios (unstandardized XP-EHH) were standardized to have mean 0 and variance 1, and these XP-EHH scores or z-scores were assigned *p*-values assuming a normal distribution, with the exception of tails that diverged from the null expectation. The top 100 loci based on *P*-value were considered strong signals. Since XP-EHH is not sensitive to allele frequencies, no need existed to stratify data into frequency bins before determining significance

2.4 Results

2.4.1 Genetic variations of 15 Korean native goats and 11 crossbred goats

I generated NGS pair-end reads to about ten-fold coverage for 15 Korean native goats and 11 crossbred goats using Illumina HighSeq2000 to obtain goat resequencing data. For each goat individual, over 92% of all reads [excluding possible polymerase chain reaction (PCR) duplicates] were successfully aligned to the reference goat genome (domestic goat, *Capra hircus*, 2n = 60, predicted size = 2.9 Gb; Dong, Xie et al. 2012; **Table 2.1**). The mapped reads covered an average of 94.77% of the reference genome (**Table 2.2**), and read coverage against the reference genome was calculated using BEDTools (Quinlan and Hall 2010).

In total, 22,759,033 single nucleotide variants (SNVs) and 2,450,921 INDELs were identified, and 26.6% of these SNVs and 26.8% of the INDELs were located in genic regions, while 73.4% of the SNPs and 73.2% of the INDELs were located in intergenic regions. Among the genic regions, 845,203 SNVs and 108,653 INDELs were detected in only the Korean native goat population (**Table 2.3**), and 1,661,071 SNVs and 108,653 INDELs were detected in only the crossbred goat population (**Table 2.4**). The average nucleotide change rate was 1 per 110 base pairs. Distribution of SNVs in the goat genome showed that the terminal part of the chromosome was highly variable, and the X chromosome showed less variation than autosomes

(Figure 2.2).

2.4.2 Nucleotide diversity of the Korean native goat population is lower than in crossbred goat populations

I examined the nucleotide diversity of Korean native goats and crossbred goats using VCFtools (Danecek, Auton et al. 2011), and The number of SNVs was also counted and integrated for each 1-Mb bin region of the genome using VCF format files, which contained variant information on 26 goats based on the reference goat genome (Dong, Xie et al. 2012). The overall distributions of nucleotide diversity and the number of SNVs in each 1-Mb bin region of the whole genome were very similar between Korean native goats and crossbred goats. However, Korean native goats generally showed lower nucleotide diversity and SNV density than crossbred goats in the same genomic regions (Figure 2.3, Figure 2.4).

2.4.3 Korean native goats and crossbred goats share common ancestry and also show divergence

According to the result of principal components analysis (PCA; Jolliffe 2005), the Korean native goats seemed to be highly inbred, and they are clearly distinguished from crossbred goats. PC1 explained most variations between Korean native and crossbred goats (Figure 2.5.). I then examined the genetic

structures of Korean native and crossbred goat populations through admixture analysis using STRUCTURE software (Evanno, Regnaut et al. 2005). I observed that Korean native and crossbred goats shared the majority of ancestral state which is expected from Korean native goats. The result supported the assumption that crossbred originated from crossbreeding of Korean native goat and other different breeds (**Figure 2.6**).

2.4.4 Six strong signatures of selective sweep were observed in the Korean native population

I estimated the values of cross-population extended haplotype homozygosity (XP-EHH; Sabeti, Varilly et al. 2007) to detect selective sweep regions and performed a pairwise test of the Korean native goat and crossbred populations (**Figure 2.7**). The top 100 regions were considered significant based on the P-value of XP-EHH test, to identify 106 loci under positive selection. Significant loci were annotated to the closest genes with an up- and downstream span of 50 kb, and I found six genes from twenty-eight loci (**Table 2.5**).

2.4.5 Highly variable goat genomic regions are enriched with genes involved in olfactory sensors or neurological systems

The number of variation sites in each 1 Mb binning region of Korean native

goat and crossbred goat genomes was counted, and I identified the top 5% highly variable regions among all chromosomal regions (empirical p -value <0.05) in each population.

I referenced the Goat Genome Database (Dong, Xie et al. 2012) and Ensembl Genome database (Hubbard, Barker et al. 2002) to identify gene locations and annotation information. I then performed gene ontology (GO)-term analysis of breed-specific gene sets in highly variable regions that I identified to clarify the difference of gene distribution in variable regions (**Table 2.7**) between Korean native and crossbred goats. The DAVID analysis tool (Dennis Jr, Sherman et al. 2003) was used for GO-term analysis in this report.

As a result, the variable regions were significantly enriched with genes involved in sensory perception of smell and neurological systems in both goat populations (**Figure 2.8**).

2.4.6 Nonsynonymous SNP studies showed difference in the immune and olfactory system between Korean goats including crossbred and the Yunnan black goat.

Among 22,759,033 SNV sites, I identified 508 nonsynonymous SNP sites of Korean native and crossbred goats using generated resequencing data and SNPEff software (Cingolani, Platts et al. 2012) to determine genes that show the difference in protein coding sequence between different goat populations. GO-term analysis of these genes was then performed using the DAVID

analysis tool (Dennis Jr, Sherman et al. 2003). DAVID analysis tool identified genes involved in the immune system that were comparatively enriched in GO-term (SOX4, JAG2, PAX1, ZAP70, SHH, SFTPD; p -value <0.05) and olfactory transduction that were enriched in KEGG pathway (OR10J1, OR2A5, OR2F2, OR4L2, OR8A1, OR8J3; p -value <0.05).

I also identified one Korean native-specific and 74 of crossbred-specific nonsynonymous SNPs (**Table 2.6**) to perform GO-term analysis but observed no noticeable result (p -value <0.05).

2.5 Discussion

2.5.1 Goat genome confirmed the history of breeding

It is known that the Korean native goat population is highly inbred due to continuous selection to preserve purebreds, and some outbred goat lines were used to form crossbred goat population based on Korean native goat to improve inferior traits of Korean native goat. From the whole-genome sequencing data, I observed a reduction in nucleotide diversity, which indicates the signal of inbreeding (Nybom and Bartish 2000) in the Korean native goat population (**Figure 2.3**). The result of PCA (Jackson 2005) also supports inbreeding in Korean native goat as I expected (**Figure 2.5**).

The results of admixture analysis using STRUCTURE (Evanno, Regnaut et al. 2005) showed in detail (**Figure 2.6**). The proportion of imported alleles increased during crossbreeding, while a significant component of alleles became an indigenous allele majority in Korean native goat. In crossbred goat population, nucleotide diversity in the genome may have increased due to the consistent introduction of new alleles unlike the Korean native goats.

2.5.2 Gene location may have given the chances of substitutions to genes

The recombination rate is known to affect nucleotide diversity (Nachman 2001), and the terminal region of chromosomes is known to be prone to recombination events. Based on the distribution of variation in the goat genome (**Figure 2.2**), the terminal regions of chromosomes showed more variation than other regions, and the top 5% of highly variable regions were enriched with genes involved in olfactory sensors and neurological systems. From the result of nonsynonymous SNP analysis in this study, I could identify enrichment of amino acid substitution in genes related to olfactory sensors, not in genes involved in neurological systems

Genes involved in both olfactory and neurological systems might have faced more chances to gain variants in nucleotide sequences. However, since goats are livestock, the introduction of new amino acid substitutions in genes involved in smell would not have significant effects as long as livestock are fed by farmers. Therefore, these variations in genes involved in olfactory systems may have been accumulated over time without purifying selection forces, while genes associated with neurological systems are comparatively conserved due to its biological importance.

2.5.3 Selective sweep region between Korean native goat and crossbred goat populations

Generally, when domestic animals are bred, they are selected for specific purposes such as food production, and artificial selection remains a strong signal of selective sweep in the genome. Based on the XP-EHH test, I identified only six significant genic regions in the Korean native goat population.

Korean native goats are reportedly resistant to diseases (especially to lumbar paralysis), and Korean native goat meat is typically used as a medicine and not a food (Son 1999). Nevertheless, differences in the specific purposes of breeding between Korean native goat and crossbred goat populations remain unclear. If no significant difference in selection progress occurs between Korean native goats and crossbred goats, XP-EHH (which is a cross-breed test) would not be able to detect a large number of selective sweep signals. Unfortunately, the number of identified gene regions that counted as selective sweeps was small. Moreover, a considerable number of genes near selective sweep regions were unclear or not well characterized.

Also, when I looked at the differences in nucleotide diversity between Korean native and crossbred, several regions with large difference of nucleotide diversity was found. In addition, performing GO-term analysis, these regions was enriched with genes involved in cell death, and one of them had a non-synonymous SNP (**Figure 2.4**). Thus, Korean native goat and crossbred goat may have different phenotype which is involved in cancer

although XP-EHH didn't identify as a significant signal of selective sweep.

2.5.4 Other possible candidate genes involved in phenotypic variation between Korean native goat and crossbred goat populations

GO term analysis of nonsynonymous SNPs showed differences in the alleles of genes involved in the immune system, including S100A9 (Schnekenburger, Schick et al. 2008), SFTPD (Davé, Childs et al. 2004), SHH (Aifantis, Raetz et al. 2008) although the *p*-value of the result was not significant.

Some differences in allele types were also observed in genes with important roles that GO-term analysis did not reveal. Korean native goat populations are also distinguished from the crossbred goat population based on nonsynonymous SNP patterns of chymosin (CYM) and collagen, type Xi, and alpha 2 (COL11A2). CYM is an enzyme involved in milk ingestion of young ruminant animals (Emmons and Lister 1976), and COL11A2 is involved in osteochondrodysplasias (Vikkula, Madman et al. 1995) and Stickler syndrome (Sirko-Osadsa, Murray et al. 1998). These genes may affect the early period of the goat life cycle and body development, and may contribute to differences in the growth of young calves between Korean native and crossbred goats.

The goat genome project was first published in 2012 (Dong, Xie et al. 2012), and the goat genome has not been fully characterized unlike in humans,

mice, or even other livestock. Therefore, the gene annotation that I used in this study is based on orthologous genes of human or cattle species, and about 20% of annotated genes are unclear or predicted genes. For these reasons, it was limited to identify functions or GO-term of genes which are shown significant from analyses, and further studies will be needed for better understanding of goat genome.



Figure 2.1. Male and female Korean native and crossbred goats

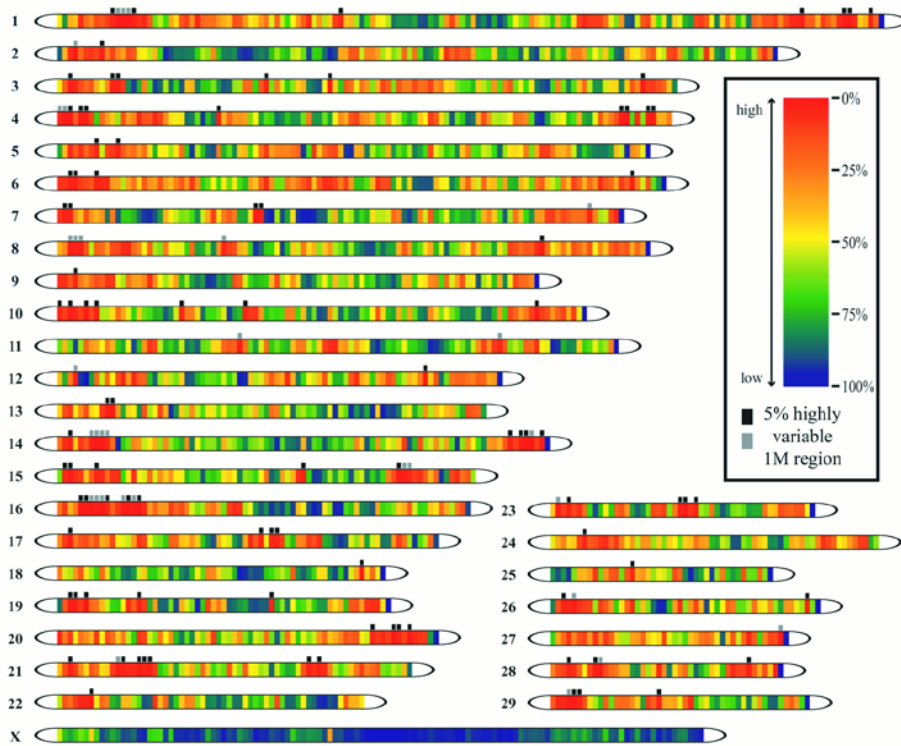


Figure 2.2. Distribution of SNVs in the Korean native goat genome

Each segment represents a 1-Mb binning region of chromosomes, which were colored according to the number of SNVs in the region. The 1-Mb binning region with red color showed more variations, and the binning region with blue color showed less variation. The top 5% of highly variable regions contained a dot on the chromosome bar. A black dot represents the 1-Mb binning region including the genic region, and the gray dot represents the 1-Mb binning region including the intergenic region.

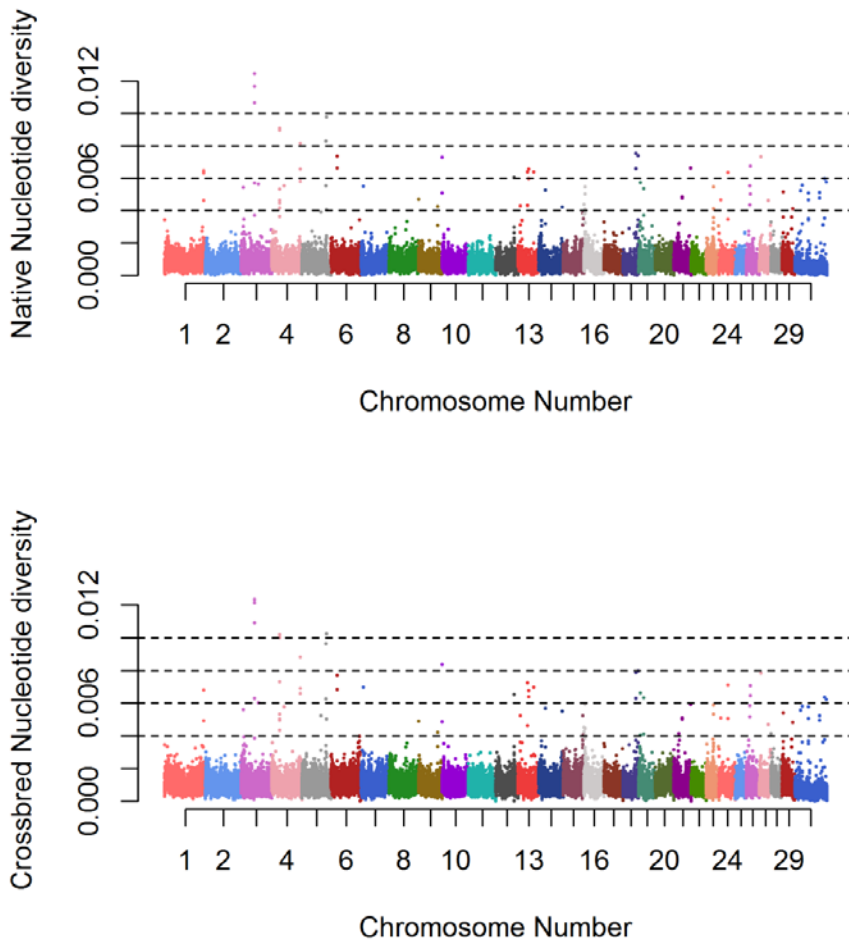


Figure 2.3. Nucleotide diversity plots of Korean native goats (top) and crossbred goats (bottom).

The numbers under horizontal axis of plot are the name of chromosome 1 to X(30). Vertical axis indicates values of nucleotide diversity of genomic regions. Each chromosomal region is colored differently, and one circle means the nucleotide diversity in each 50K bin region

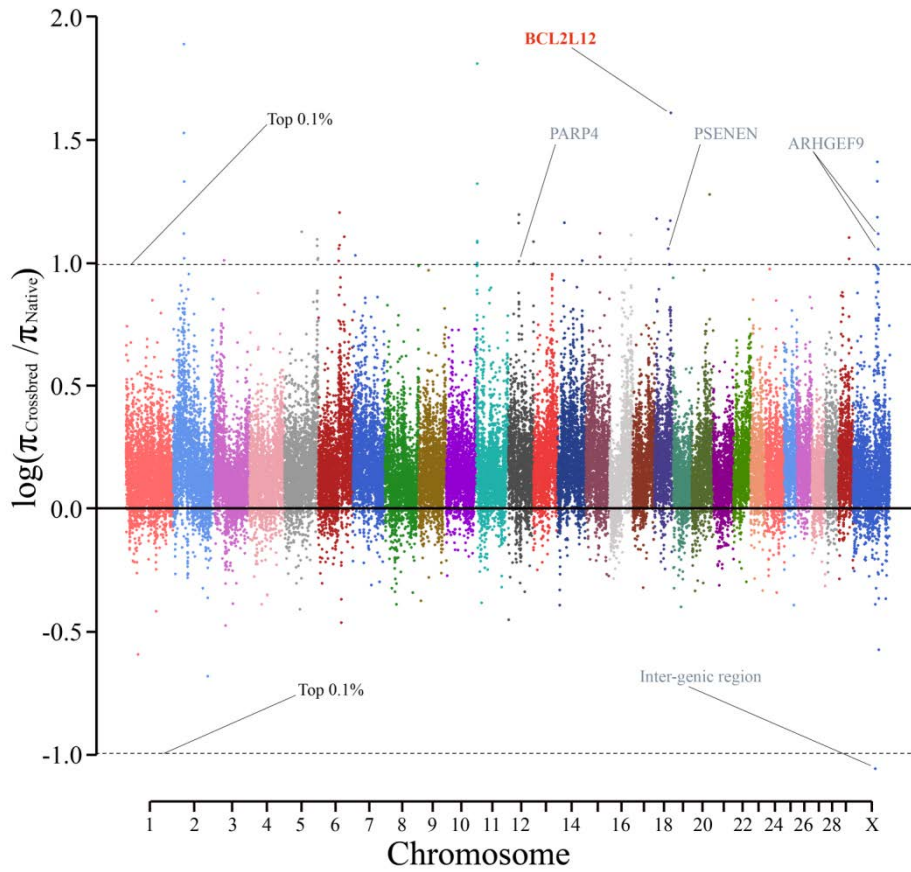


Figure 2.4. Comparison nucleotide diversity between Korean native goat and crossbred goat

Each spot indicates logged ratio of nucleotide diversity of crossbred goat to Native goat's on each 50Kbp binning region. Dotted lines present top 0.1% of the absolute values (0.9937). Among these top 0.1 significantly different binning regions, the regions which contain genes involved in cell death were marked. The gene with red color(BCL2L12) has non-synonymous SNP.

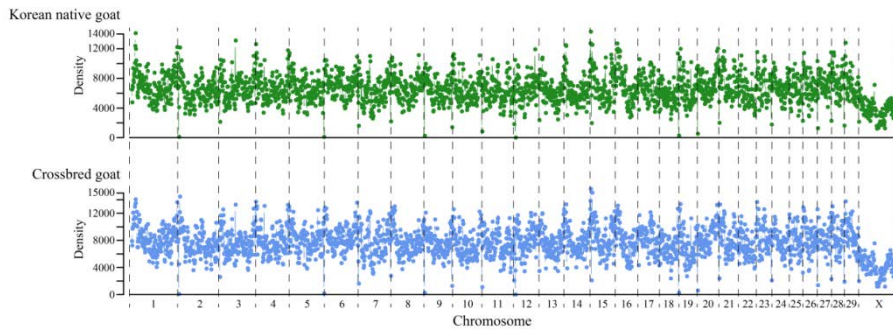


Figure 2.5. SNV density of Korean native goat genome and crossbred goat.

Each point is the density of 1M bin region. The numbers of SNVs were calculated using vcf files. Green dots represent the SNVs density of Korean native goat population, and blue dots represent the SNVs density of crossbred goat population

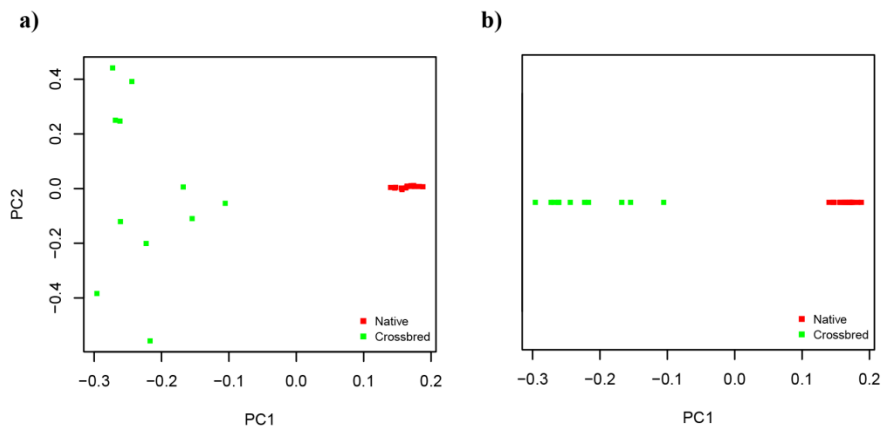


Figure 2.6. Result of Principal Component Analysis.

a) Red squares are individuals in Korean native goats, and light green squares are individuals in crossbred goats. Horizontal axis indicates PC1, and vertical axis indicates PC2. Values of eigenvector were estimated using GCTAtool. **b)** Only PC1 is presented.

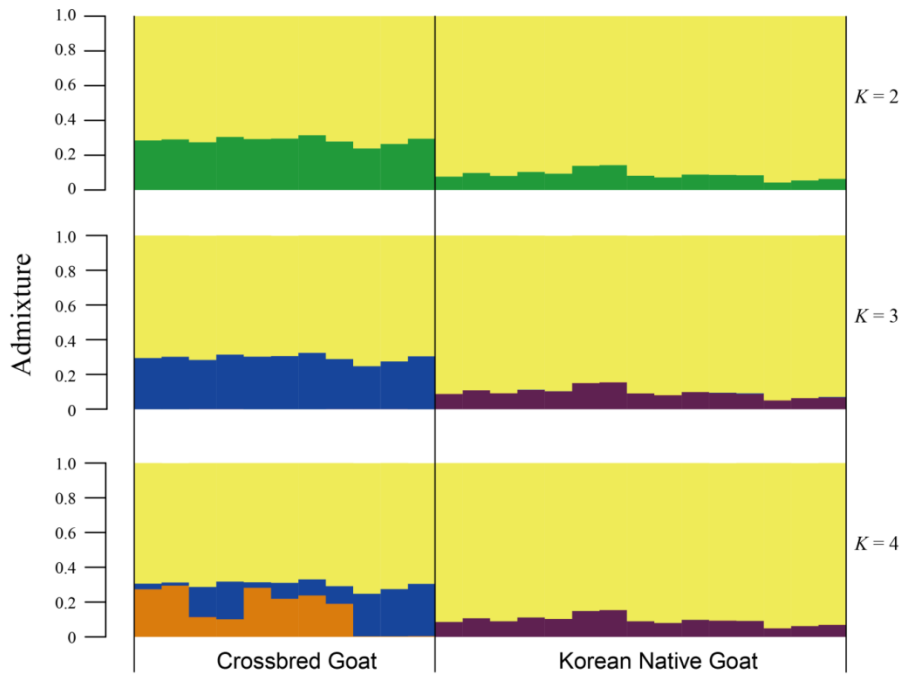


Figure 2.7. Population structure of goat breeds assessed using STRUCTURE

Each segment represents the proportion of a goat individual genome from ancestral populations. Different colored segments in individuals assume that part of the genome originated from different ancestral populations. This figure shows the genetic structure of goat breeds when I assume that the number of ancestral population of goats is 2–4.

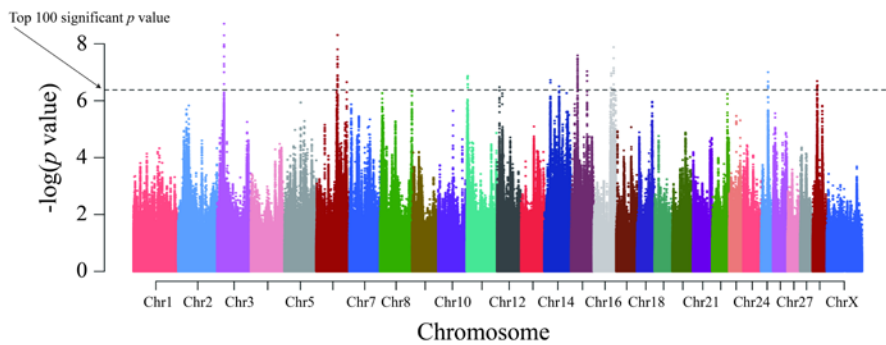


Figure 2.8. Candidate genomic regions of positive selection

The Manhattan plot shows the XP-EHH between Korean native goat and crossbred goat populations. The vertical axis indicates $-\log_{10}(p\text{-value})$ of XP-EHH values.

	Term	p-value	FDR
Biological process GO term	Sensory perception of smell (54) *	1.51E-37	2.36E-34
	Sensory perception of chemical stimulus (54) *	3.29E-35	5.15E-32
	Sensory perception (56) *	1.21E-25	1.90E-22
	Cognition (57)	4.85E-24	7.59E-21
	G-protein coupled receptor protein signaling pathway (57)	1.22E-19	1.90E-16
	Neurological system process (59)	1.44E-19	2.25E-16
	Cell surface receptor linked signal transduction (62)	9.41E-13	1.47E-9
	Collagen metabolic process (4)	0.005496	8.260127
	Multicellular organismal macromolecule metabolic process (4)	0.007331	10.872164
	Multicellular organismal metabolic process (4)	0.011977	17.178426
	Response to metal ion (6)	0.026084	33.862943
KEGG pathway	Sensory perception of chemical stimulus (50) *	1.12E-34	1.73E-31
	Sensory perception of smell (48) *	2.02E-34	3.10E-31
	Cognition (54)	1.85E-25	2.84E-22
	Sensory perception (51) *	5.04E-25	7.74E-22
	G-protein coupled receptor protein signaling pathway (55)	5.48E-22	8.42E-19
	Neurological system process (56)	2.82E-21	4.33E-18
	Cell surface receptor linked signal transduction (60)	2.32E-15	3.59E-12
	Platelet activation (3)	0.047612	52.736316
KEGG pathway	Olfactory transduction (52)	7.99E-29	8.52E-26
	Olfactory transduction (46)	1.25E-27	1.28E-24
	Complement and coagulation cascades (5)	0.034333	30.191410

Korean native goat

Crossbred goat

Figure 2.9. Representative biological process GO terms and KEGG pathways associated with genes located in the top 5% of highly variable regions in Korean native goat and crossbred goat populations ($p < 0.05$).

The red regions indicate the GO terms selected from Korean native goats. The blue regions denote the GO terms selected from crossbred goats. Numbers in parenthesis represent the number of genes associated with the GO term.

Table 2.1. Summary of alignment rate in 26 goat samples

Categories							
Sample	Total number of reads	concordantly 1 time	concordantly >1 time	discordantly 1 time	1 time with single reads	>1 time with single reads	Overall mapping Rate
D11_011 (Native)	223435512 (100.00%)	176095233 (78.81%)	16476421 (7.37%)	3104167 (1.39%)	9641688 (2.16%)	16340099 (3.66%)	208666714.5 (93.39%)
D11_016 (Native)	212755767 (100.00%)	168330691 (79.12%)	16970812 (7.98%)	1647412 (0.77%)	8229049 (1.93%)	14948227 (3.51%)	198537553 (93.32%)
D11_035 (Native)	223098954 (100.00%)	176345793 (79.04%)	16969108 (7.61%)	2129835 (0.95%)	9092243 (2.04%)	15887687 (3.56%)	207934701 (93.20%)
D11_037 (Native)	217591992 (100.00%)	173233485 (79.61%)	17751250 (8.16%)	1853863 (0.85%)	8422585 (1.94%)	14339681 (3.30%)	204219731 (93.85%)
D11_038 (Native)	213139285 (100.00%)	168614089 (79.11%)	17593735 (8.25%)	2410864 (1.13%)	8333268 (1.95%)	14792061 (3.47%)	200181352.5 (93.92%)
J11_007 (Native)	222818884 (100.00%)	175443805 (78.74%)	17820098 (8.00%)	4086504 (1.83%)	9011906 (2.02%)	15049058 (3.38%)	209380889 (93.97%)
J11_008 (Native)	221522250 (100.00%)	174781922 (78.90%)	17489175 (7.89%)	4754867 (2.15%)	8867764 (2.00%)	14847537 (3.35%)	208883614.5 (94.29%)
J11_012 (Native)	224424791 (100.00%)	178697830 (79.62%)	18604701 (8.29%)	2008309 (0.89%)	8373733 (1.87%)	14605918 (3.25%)	210800665.5 (93.93%)
J11_027 (Native)	220517331 (100.00%)	174912241 (79.32%)	18221519 (8.26%)	2493996 (1.13%)	8581381 (1.95%)	14669622 (3.33%)	207253257.5 (93.99%)
J11_035 (Native)	213254835 (100.00%)	168525566 (79.03%)	17323349 (8.12%)	2838915 (1.33%)	8510183 (2.00%)	14767421 (3.46%)	200326632 (93.94%)
T10_004 (Native)	226341463 (100.00%)	178835635 (79.01%)	19671288 (8.69%)	1544266 (0.68%)	8580714 (1.90%)	15889257 (3.51%)	212286174.5 (93.79%)
T10_007 (Native)	222515628 (100.00%)	166989083 (75.05%)	17070018 (7.67%)	10529864 (4.73%)	10442606 (2.35%)	19019647 (4.27%)	209320091.5 (94.07%)
T11_002 (Native)	230447366 (100.00%)	181374576 (78.71%)	20196581 (8.76%)	1579709 (0.69%)	8996228 (1.95%)	16528500 (3.59%)	215913230 (93.69%)

T11_011 (Native)	236159806 (100.00%)	187137806 (79.24%)	21118640 (8.94%)	1383833 (0.59%)	8596604 (1.82%)	16019708 (3.39%)	221948435 (93.98%)
T11_016 (Native)	225251234 (100.00%)	179225057 (79.57%)	19834764 (8.81%)	1219600 (0.54%)	8163977 (1.81%)	15028212 (3.34%)	211875515.5 (94.06%)
Y1 (Crossbred)	211552351 (100.00%)	160048817 (75.65%)	17689501 (8.36%)	3939526 (1.86%)	11060992 (2.61%)	18386920 (4.35%)	196401800 (92.84%)
Y2 (Crossbred)	216320155 (100.00%)	161529410 (74.67%)	18569375 (8.58%)	5951352 (2.75%)	11642051 (2.69%)	18006550 (4.16%)	200874437.5 (92.86%)
Y4 (Crossbred)	209738017 (100.00%)	159206072 (75.91%)	19554792 (9.32%)	2031068 (0.97%)	10408636 (2.48%)	15748167 (3.75%)	193870333.5 (92.43%)
Y6 (Crossbred)	219666517 (100.00%)	165054541 (75.14%)	20620330 (9.39%)	2759404 (1.26%)	11301559 (2.57%)	17044830 (3.88%)	202607469.5 (92.23%)
Y7 (Crossbred)	223375407 (100.00%)	171831766 (76.93%)	22715183 (10.17%)	1300219 (0.58%)	9454336 (2.12%)	13406742 (3.00%)	207277707 (92.79%)
Y8 (Crossbred)	193197973 (100.00%)	152358174 (78.86%)	17073120 (8.84%)	1619981 (0.84%)	8006935 (2.07%)	11812419 (3.06%)	180960952 (93.67%)
Y11 (Crossbred)	218254056 (100.00%)	166273961 (76.18%)	18910312 (8.66%)	4691997 (2.15%)	10606297 (2.43%)	16798020 (3.85%)	203578428.5 (93.28%)
Y12 (Crossbred)	227898122 (100.00%)	172127074 (75.53%)	18281020 (8.02%)	7232480 (3.17%)	12096168 (2.65%)	19675779 (4.32%)	213526547.5 (93.69%)
Y13 (Crossbred)	210573179 (100.00%)	162124619 (76.99%)	21396720 (10.16%)	1453733 (0.69%)	8658341 (2.06%)	13139103 (3.12%)	195873794 (93.02%)
Y14 (Crossbred)	209714819 (100.00%)	161313654 (76.92%)	20975421 (10.00%)	1662312 (0.79%)	8825422 (2.10%)	13867278 (3.31%)	195297737 (93.13%)
Y15 (Crossbred)	215658576 (100.00%)	164695829 (76.37%)	17818607 (8.26%)	2718292 (1.26%)	11400324 (2.64%)	16045353 (3.72%)	198955566.5 (92.25%)

Table 2.2. Genome coverage of mapped reads for each sample

Sample	Genome_Length (bp)	Coverage (%)
Reference	2525732907	100.0
D11_011	2393410083	94.76
D11_016	2392994534	94.74
D11_035	2392527049	94.73
D11_037	2392602405	94.73
D11_038	2393091317	94.75
J11_007	2394104487	94.79
J11_008	2393515439	94.77
J11_012	2393488297	94.76
J11_027	2393766606	94.78
J11_035	2393120415	94.75
T10_004	2392994281	94.74
T10_007	2392723131	94.73
T11_002	2392545585	94.73
T11_011	2391877552	94.70
T11_016	2392324263	94.72
Y1	2394266297	94.79
Y2	2395111611	94.83
Y4	2394967594	94.82
Y6	2395667432	94.85
Y7	2395238590	94.83
Y8	2392410401	94.72
Y11	2394477604	94.80
Y12	2394479625	94.80
Y13	2395084378	94.83
Y14	2395013618	94.82
Y15	2394647156	94.81

Table 2.3. Identified variants in 15 Korean native goats and 11 crossbred goats

Variant	Population	Number of Variants	Number of Total Variants
SNP	Native	16570906	22759033
	Cross-Bred	19577491	
INDEL	Native	1819536	2450921
	Cross-Bred	2037004	

Table 2.4. Variants in genic regions and intergenic regions

Variant Region	Variant Type			
	SNP		INDEL	
Genic region	6050865		657426	
	Native-specific	Crossbred-specific	Native-specific	Crossbred-specific
	845203	1661071	108653	175385
Intergenic region	16708168		1793495	
Total	22759033		2450921	

Table 2.5. List of candidate genes identified from positive selection scan

Gene ID (Goat Genome Database)	Gene Symbol	Note
GOAT_ENSBTAP00000015070	Unknown	no longer in database
GOAT_ENSBTAP00000021251	FLVCR1	
GOAT_ENSBTAP00000048497	TATDN3	
GOAT_ENSP000000358458	FAM46C	
GOAT_ENSP000000396459	Unknown	no longer in database
goat_GLEAN_10017305	Unknown	predicted

Table 2.6. Breed-specific non-synonymous SNP sites.

Position	Gene ID	Codon Change	Substitution
chr1:671485	goat_GLEAN_10006519	Cga/Tga	R6*
chr1:69096942	GOAT_ENSBTAP00000043771	Cct/Act	P850T
chr1:113439307	GOAT_ENSP00000416653	aCc/aAc	T2N
chr1:133755666	goat_GLEAN_10006085	tCt/tTt	S42F
chr1:143513840	GOAT_ENSBTAP00000025663	gCg/gTg	A821V
chr2:4183858	GOAT_ENSBTAP00000046201	Tgt/Cgt	C4R
chr2:132205598	GOAT_ENSBTAP00000010513	Gcg/Tcg	A93S
chr3:9170303	GOAT_ENSBTAP00000001940	cCg/cTg	P122L
chr3:16051776	GOAT_ENSBTAP00000008523	aAt/aGt	N53S
chr3:30838263	GOAT_ENSBTAP00000013970	Atc/Gtc	I306V
chr3:99029021	GOAT_ENSBTAP00000007333	aGg/aTg	R290M
chr3:99029022	GOAT_ENSBTAP00000007333	Agg/Tgg	R290W
chr3:104191661	GOAT_ENSBTAP00000009604	Ccc/Tcc	P66S
chr3:115810924	GOAT_ENSP00000362405	Cgc/Agc	R1399S
chr4:103863252	GOAT_ENSP00000386208-D9	cAa/cGa	Q266R
chr4:113498379	GOAT_ENSBTAP00000034059	cTg/cCg	L92P
chr6:83530113	GOAT_ENSBTAP00000016449	Cgc/Agc	R20S
chr6:112343591	GOAT_ENSP00000339381	agC/agA	S1185R
chr7:5227408	GOAT_ENSBTAP00000022306	Cag/Aag	Q185K
chr7:5753477	GOAT_ENSP00000369131	Gtg/Atg	V951M
chr7:27749220	GOAT_ENSP00000368848	Ccg/Gcg	P156A
chr7:47489623	GOAT_ENSP00000332706	cGg/cTg	R4L
chr8:99053324	GOAT_ENSP00000320650	gCc/gAc	A2D
chr9:84797150	goat_GLEAN_10017139	aAg/aCg	K99T
chr10:22628260	GOAT_ENSBTAP00000027974	cCc/cGc	P4R
chr10:22628266	GOAT_ENSBTAP00000027974	cGg/cTg	R6L
chr10:61473560	GOAT_ENSBTAP00000021603	Aaa/Gaa	K562E
chr10:83857971	GOAT_ENSBTAP00000016290	Gcc/Tcc	A65S
chr11:2825860	GOAT_ENSP00000422513	Gcc/Acc	A1213T
chr11:5380597	goat_GLEAN_10005042	Act/Gct	T51A
chr11:20960241	GOAT_ENSBTAP00000033055	Ata/Gta	I14V
chr11:99532402	GOAT_ENSBTAP00000023920	cCg/cTg	P1117L
chr13:39416277	GOAT_ENSP00000381499	gGg/gTg	G2V
chr14:11190641	GOAT_ENSBTAP00000011004	Cgg/Tgg	R199W
chr14:18401539	goat_GLEAN_10003471	aCa/aTa	T46I
chr14:19034022	goat_GLEAN_10002235	gTg/gCg	V131A
chr14:25053259	goat_GLEAN_10012166	Cgt/Tgt	R89C
chr14:26075485	GOAT_ENSP00000329858	Ggc/Agc	G195S
chr14:37052068	GOAT_ENSP00000307304	cCg/cTg	P1953L
chr14:91342838	GOAT_ENSBTAP00000053499	aGg/aCg	R2T

chr15:13727888	GOAT_ENSP00000397239-D3	tgA/tgG	*132W
chr15:13728238	GOAT_ENSP00000397239-D3	aGc/aCc	S16T
chr15:13761433	goat_GLEAN_10012036	agA/agC	R42S
chr15:20050892	GOAT_ENSBTAP00000045643	Agc/Ggc	S11G
chr15:75584128	GOAT_ENSP00000301529-D47	Cac/Gac	H157D
chr15:76012063	GOAT_ENSP00000330864	cCt/cAt	P57H
chr16:77070267	goat_GLEAN_10001243	Cgc/Tgc	R39C
chr17:65925877	GOAT_ENSP00000399801	Acc/Tcc	T39S
chr17:70559130	GOAT_ENSBTAP00000029404	Gtg/Ctg	V5L
chr17:71345374	GOAT_ENSP00000215882	Atc/Gtc	I162V
chr18:33938740	GOAT_ENSBTAP00000049320	Cgt/Tgt	R58C
chr18:39729482	GOAT_ENSBTAP00000009950	cCc/cAc	P5H
chr18:49839267	GOAT_ENSP00000222339	Cag/Gag	Q722E
chr19:37767394	GOAT_ENSBTAP00000002390	Cgg/Tgg	R223W
chr19:40510267	GOAT_ENSBTAP00000040739-D7	Ggc/Tgc	G48C
chr20:71214910	GOAT_ENSBTAP00000016676	tCc/tTc	S86F
chr21:62110518	GOAT_ENSBTAP00000018163	Gcc/Tcc	A193S
chr22:48613130	GOAT_ENSBTAP00000029212	Cgc/Ggc	R3G
chr23:7051577	GOAT_ENSBTAP00000000793	Ggc/Tgc	G1109C
chr23:27503629	GOAT_ENSP00000415772	cGg/cAg	R182Q
chr23:33733412	GOAT_ENSP00000244745	Tcc/Ccc	S52P
chr24:3232843	goat_GLEAN_10012704	cTa/cCa	L2P
chr24:50278380	GOAT_ENSP00000383234	tCc/tAc	S353Y
chr24:50278419	GOAT_ENSP00000383234	gGc/gAc	G366D
chr24:58059134	GOAT_ENSBTAP00000030954	cGc/cTc	R212L
chr25:1587389	GOAT_ENSP00000191922	gCc/gAc	A8D
chr25:33623642	GOAT_ENSBTAP00000012060	aGc/aAc	S543N
chr25:35832653	GOAT_ENSP00000300181	cCc/cGc	P56R
chr26:50143331	goat_GLEAN_10000730	Acc/Ccc	T171P
chr27:1264445	GOAT_ENSP00000381571	Cga/Gga	R58G
chr28:8768461	GOAT_ENSBTAP00000036829	tTg/tCg	L436S
chr28:33218950	GOAT_ENSP00000361366	cGt/cTt	R122L
chr29:26029439	goat_GLEAN_10008565	gGa/gAa	G16E
chr29:40636735	GOAT_ENSP00000366774	Gcg/Ccg	A1277P
chrX:79388323	GOAT_ENSBTAP00000003318	cCt/cTt	P84L

Table 2.7. Breed-specific genes in the top 5% of highly variable regions.

Variant Type	Korean native goat						Crossbred goat			
SNV	AGPAT1	APC	APEX1	ARF1	ARHGAP42	ATF6B	A1CF	APP	ASAH2B	C8ORF80
	B3GALT2	C12ORF26	CBS	CCDC59	CCL11	CCL7	CCBE1	CCDC25	CDH19	CDH7
	CCL8	CCNB1IP1	CDC73	CHD2	CLIC1	CRYAA	CLDN8	CLEC12A	CLEC1A	CLEC1B
	DCP2	EGFL8	EHMT2	ETFDH	F2RL1	F2RL2	CLEC9A	CLNK	CLU	CPLX4
	FAM198B	FKBPL	GLRX2	GPSM3	GUK1	HBD	CSDA	CYYR1	DDI1	DNAJC15
	HBE2	HBE4	HIST3H2A	HIST3H2BB	HSPA1A	HSPA1L	DOK5	DRD5	ELP3	ENOX1
	IBA57	IQGAP2	JMJD4	KCNH8	KRT12	KRT14	EPSTI1	ESCO2	FAM83B	FBXO16
	KRT15	KRT17	KRT20	KRT222	KRT23	KRT24	FER	FUT4	GABARAPL1	GABPA
	KRT25	KRT26	KRT28	KRT31	KRT32	KRT36	GALC	GPR65	GRP	HS3ST1
	KRT37	KRT38	KRT39	KRTAP29-1	KRTAP3-1	KRTAP3-2	KCNK1	KCNK10	KLRJ1	KRTAP13-1
	KRTAP9-4	LSM2	MMP16	MMTAG2	MRPL55	MSH5-SAPCD1	KRTAP19-5	KRTAP24-1	KRTAP26-1	KRTAP8-1
	NAV3	NCAM2	NEU1	NOTCH4	NOX4	OLFM3	MAGOHB	MAN2A1	MC4R	MLIP
	OR11G2	OR11H1	OR11H4	OR11H6	OR4K1	OR4K15	MLK4	OLR1	OR10D3	OR5AK2
	OR4K2	OR4K5	OR4L1	OR51D1	OR51E1	OR51E2	OR5AP2	OR5J2	OR5R1	OR5T1
	OR51F2	OR51G1	OR51I2	OR51M1	OR51S1	OR51T1	OR5T2	OR8A1	OR8B2	OR8B4
	OR52A1	OR52E2	OR52K2	OR52R1	OR6S1	OSGEP	OR8B8	OR8G5	OR8J1	OR8J3
	PARP2	PBX2	PDE9A	PKNOX1	PNP	PPID	OR8K1	OR8K3	OR8U1	OR9G1
	PPT2	PRSS38	REEP5	RGMA	RNASE1	RNASE10	OR9G4	OTOP1	PBK	PCNXL2
	RNASE11	RNASE4	RNASE9	RNF187	S100Z	S1PR1	PIWIL4	PJA2	PLEKHB2	PNOC
	SAPCD1	SKIV2L	SLC44A4	SLCO3A1	SMC2	SRP19	PPFIA2	PRKG1	RIMS1	SCARA3
	ST8SIA2	STK19	SV2C	TEP1	TFF1	TFF2	SCARA5	SEC11C	SGMS1	SIAE
	TMEM144	TMEM55B	TMEM99	TMPRSS3	TMTC2	TNXB	SLC2A9	SLC35F3	TAS2R7	TAS2R8
	TROVE2	TTC5	TYR	U2AF1	UBASH3A	UCHL5	TBRG1	TINAG	TKTL2	TMA16
	UMODL1	VARS	WDR4	WNT3A	WNT9A	XYLT1	TMEM232	ZNF395	ZNF518B	
	ZBTB12									

Variant Type	Korean native goat				Crossbred goat					
INDEL	AGBL1	B3GALT2	B3GAT2	C12ORF26	BMP5	C4H7orf23	C7ORF62	C8ORF59	CA13	CA3
	C4BPA	C6ORF170	CCDC59	CDC73	CCDC126	CDH9	CENPQ	CLEC12A	CLEC1A	CLEC1B
	CLDN10	CR2	DZIP1	FAIM3	CLEC9A	CLNK	COBL	CRISP1	CRISP2	CROT
	FCAMR	FGA	FGB	FGG	CSDA	DEFB110	DEFB112	DEFB114	DEFB133	DNAJC15
	GABRG1	GLRX2	HBD	HBE2	DRD5	E2F5	ENOX1	ENY2	EPSTI1	FAM174A
	HBE4	IL24	LRAT	LRFN5	FUT4	GABARAPL1	GABRA5	GABRG3	GFRAL	GRIK2
	MAP9	MDGA2	NOX4	NPY2R	HCRTR2	HMGCLL1	KCNK1	KLRJ1	LRRCC1	MAGOHB
	OGFRL1	OR51D1	OR51E1	OR51E2	MLK4	MUT	NUDCD1	OLR1	OR10D3	OR5AK2
	OR51F2	OR51G1	OR51I2	OR51M1	OR5AP2	OR5J2	OR5R1	OR5T1	OR5T2	OR8A1
	OR51S1	OR51T1	OR52A1	OR52E2	OR8B2	OR8B4	OR8B8	OR8G5	OR8J1	OR8J3
	OR52K2	OR52R1	PFKFB2	PLRG1	OR8K1	OR8K3	OR8U1	OR9G1	OR9G4	OTOP1
	RBM46	RPL10L	SARG	SMC2	PCNXL2	PGK2	PIWIL4	PPFIA2	RHAG	SIAE
	TLL1	TMTC2	TROVE2	TYR	SLC2A9	SLC35F3	SYBU	TAS2R7	TAS2R8	TBRG1
	UCHL5	UGGT2	VGLL3	YOD1	TECRL	TKTL2	TMA16	TMEM74	TRHR	ZNF518B
					ZNF804B					

Chapter 3. African Indigenous Cattle and their Potential as Genetic Resources

3.1 Abstract

At least 150 indigenous African cattle breeds have been named, however the majority of African cattle populations remains largely uncharacterized. As cattle breeds and populations in Africa adapted to various local environmental conditions, they acquired unique features. It is known now that the history of African cattle was particularly complex and while several of its episodes remain debated, there is no doubt that African cattle population evolved dramatically over time. Today, many researchers have found a mosaic of genetically diverse population from the purest *B. taurus* to the nearly pure *B. indicus*. African cattle are now found all across the continent, with the exception of the Sahara and the river Congo basin. They are found on the rift valley highlands as well as below sea level in the Afar depression. These unique livestock genetic resources are in danger to disappear rapidly following uncontrolled crossbreeding and breed replacements with exotic breeds. Breeding improvement programs of African indigenous livestock remain too few while paradoxically the demands of livestock products are continually increasing. Many African indigenous breeds are endangered now, poorly characterized their unique adaptive traits may be lost forever. I reviewed the unique known characteristics of indigenous African cattle populations while describing the opportunities, the necessity and urgency to understand and utilize these resources to respond to needs of people.

3.2 Introduction

Once humans started domesticating animals, livestock have continually been selected through the process of selective breeding to respond to the needs of the owners as well as through natural selection to adapt to the local agro-environment. Animal breeding was initially practiced to obtain usable products or services and for socio-cultural (e.g. religions, aesthetic) reasons, (Flint and Woolliams 2008). Although ancient breeding schemes might have existed, knowledge-based selective breeding was first practiced by the Romans (Buffum 1909). Now, with recent technological developments, both phenotypes measurements and genotypes information are now used for breeding improvements (Flint and Woolliams 2008). In particular, methods for obtaining genome information of animals have developed rapidly (e.g. genome wide genotyping and sequencing) opening the new field of genomic selection for breeding improvement (Habier, Fernando et al. 2007, Hayes, Lewin et al. 2013).

However, the commercial livestock sector is facing several new challenges. There is a continuous increased demand for livestock products but long-term sustainability of the intensive livestock sector is being questioned (Otten and Van den Weghe 2011). Climate change is putting new pressures on livestock production while livestock themselves are contributing to the issue (Nardone, Ronchi et al. 2010) (Houghton, Callander et al. 1992). On the other hands indigenous livestock, although adapted to the local environments,

remains poor producers compared to the commercial breeds raised in the extensive system, while the traditional livestock production systems impact on climate changes is higher compared to the commercial sector (Renaudeau, Collin et al. 2012).

These challenges means that more than ever indigenous livestock genetic resources and their diversity are increasingly in danger of disappearing through crossbreeding, changes in the production system and breeds' replacement. It means that we might be losing cattle genetic adaptations to the past and present African environmental challenges, including parasitic infectious diseases heat and drought tolerance.

I summarized the features of indigenous cattle of Africa and explained the reason why we have to take an interest in the genetic resources of Africa.

3.3 African cattle

3.3.1 The origin of African cattle

Currently, around 180 breeds of cattle exist in continental Africa; 150 breeds of indigenous cattle have been locally adapted and the others are commercial composites (Rege 1999, Rege, Ayalew et al. 2006). Zebu cattle *Bos indicus* (humped cattle) constitute the majority of African cattle.

They have inhabited various areas of different conditions and are highly adapted to environmental conditions in which exotic breeds are normally difficult to survive (e.g. high temperatures, long period of drought, tsetse-infection). Indigenous taurine cattle *B. taurus* (humpless cattle) are now found nearly exclusively in West Africa, while commercial taurine breeds and their crossbreed are found across the continent, although their remain relatively low in number compared to indigenous breeds also present mainly. Beside these two main types of cattle the continent is hosting intermediate types of cattle of crossbreeding origin between indigenous taurine and zebu, named as sanga and zenga, which the latter with an expected higher zebu genetic background than the former (Rege 1999). It should be noted, that all African cattle do carry a taurine mitochondrial DNA, in others words they are no pure zebu *B. indicus* on the African continent.

The wild auroch, *Bos primigenius*, is the progenitor of all taurine and zebu cattle (Edwards, Bollongino et al. 2007). They are described by Julius Caesar in his

report of the Gallic Wars as “...a little below the elephant in size, and of the appearance, color, and shape of a bull”. Three subspecies of auroch are currently being recognized. According to fossil record, Northern Africa was the habitat of *B. p. africanus* from Middle Pleistocene onwards (Linseele 2004), and this species was also widely distributed in western Eurasia, *B. p. primigenius*, while the supposedly wild ancestor of zebu *B. p. nomadicus* inhabited South Asia (Stock and Gifford-Gonzalez 2013).

Although depending of the calibration of the molecular clock, all estimates of the *B. taurus* and *B. indicus* divergent time from a common ancestor are pre-Neolithic ranging from 2 million years - to 330,000 years ago (Hiendleder, Lewalski et al. 2008, Ajmone-Marsan, Garcia et al. 2010). It remains unclear whether or not African cattle were domesticated on the African continent. It was originally thought that it might have been the case following molecular evidence indicating the presence of a “unique” African mitochondrial DNA *D*-loop haplogroups (Loftus, MacHugh et al. 1994) and archaeological remains of ancient cattle husbandry in the East of the Sahara desert (Wendorf and Schild 1994). However, more recently, studies using the full cattle mitochondrial DNA sequences have shown that the diversity of African mitochondrial DNA is embedded within the mitochondrial DNA diversity of Near East cattle (Troy, MacHugh et al. 2001), therefore making possible that the first African domestic cattle may have originate from the Fertile Crescent. Moreover, a genome-wide study using single nucleotide polymorphic makers indicate the presence in African cattle of an unique genetic

background which might have originate from the African auroch through a male mediated introgression process (Decker, McKay et al. 2014).

Archaeological and molecular data have allowed to make inferences about the history of African cattle including they possible migration routes within the African cattle (Hanotte, Bradley et al. 2002) (**Figure 3.1**; Hanotte, Bradley et al. 2002, ILRI 2006, Stock and Gifford-Gonzalez 2013). Humpless *Bos taurus* are considered the earliest African cattle as supported by archaeological evidence and pictorial accounts (e.g. Sahara rock painting) (Blench and MacDonald 2000). First, humpless longhorns might have been introduced, followed about 2500 years later by humpless shorthorns cattle (Epstein 1971, Rege 1999). However, it should be pointed out this scenario of two waves of taurine arrival has not been supported yet by any molecular studies. Hanotte et al. (2002) have shown that African *Bos taurus* cattle expanded from the northeast part of the continent to West and East Africa (Hanotte, Bradley et al. 2002). After the initial dispersion taurine cattle, zebu cattle were introduced into the continent, likely in two waves (Hanotte, Bradley et al. 2002). It is possible that the first African cattle reaching the Southern part of the continent were taurine cattle already introgressed with zebu (Epstein 1971, Hanotte, Bradley et al. 2002), Although zebu cattle were probably present as early than 2000BC the first wave of zebu introduction through the Horn of the continent, is believe to be mainly associated with the development of the Swahili-Arabs civilization along coastal East Africa from the 7th century AD. The second wave of *Bos indicus*, was perhaps more recent following the rinderpest epidemic of 19th

century which wiped out most of African cattle population, with zebu cattle believed to be more tolerant to rinderpest than indigenous African taurine and first taurine x zebu crossbreed.

3.3.2 Types of indigenous cattle in Africa

Indigenous African cattle can be broadly classified into four categories: humpless *Bos taurus*, humped *Bos indicus*, sanga (African humpless *Bos taurus* x humped *Bos indicus* hybrid), and zenga, defined as sanga × zebu backcross. In addition to these four categories, other African cattle breeds have been recently derived through recent crossbreeding with exotic (non-African breed) (Rege and Tawah 1999).

African *Bos taurus* include two groups, humpless shorthorns and longhorns. They mostly inhabit West and Central Africa. They have small body size and their productivity is lower compare to most zebu cattle populations in tropical areas, (Rege 1999). However, they have unique evolutionary adaptation to harsh climate (Hansen, 2004) and various endemic diseases (Murray, Trail et al. 1984, Mattioli, Pandey et al. 2000). One of these adaptations is their documented tolerance to trypanosomosis (Roberts and Gray 1973), a parasitic disease following infection with *Trypanosoma* sp. whose vector is the tsetse flight.

Zebu cattle, *Bos indicus*, are the largest cattle type in Africa, and they have a fatty hump on their shoulders and a large dewlap. The zebu is usually susceptible to trypanosomosis (Murray, Trail et al. 1984) (however, tolerance to the disease has documented in the Orma Boran an East African zebu breed ref; Njogu, Dolan et al. 1985). They are adapted to the dry environmental conditions and are known to be more resistant to cattle tick infestation than *Bos taurus* cattle (Mattioli, Pandey et al. 2000). African zebu cattle's inhabits

western and eastern parts of Africa. Their large body size and high production levels in a tsetse-free condition have kept them appealing to African farmers, and may explain their abundance in a large part of the African continent.

The sanga is an intermediate type of cattle, which is a cross between *Bos taurus* and *Bos indicus*. They inhabit eastern and southern Africa and are known to be well adapted to seasonally harsh conditions (Okello and Sabiiti 2006). It is thought that sanga cattle were derived by hybridization between taurine cattle and zebu around 700 AD (Hanotte, Bradley et al. 2002). Crossbreeding between sanga cattle, which were already resident in East Africa, and new introduced zebu led to a new cattle breed called “Zenga” (Rege 1999). Zenga are found in eastern Africa.

Although, cattle diversity remains large on the African continent, indigenous African cattle population or breeds have faced extinction in the recent past. According to Rege (1999), 22% of African cattle breeds have already become extinct in the last century and 32% of indigenous African cattle breeds are in danger of extinction (Rege 1999). Moreover, some breeds that are critically endangered have fewer than 1,000 animals, such as the Ghana Dwarf Muturu (humpless shorthorns), Mkalama Dun (Small East African Zebu), Pare (Small East African Zebu), Chagga (Small East African Zebu), Baria (Small East African Zebu), Nkone (South African sanga), Pedi (South African sanga), and Shangan (South African sanga) (Rege 1999). To compensate the relatively lower productivity of indigenous cattle, African farmers are crossbreeding African indigenous cattle and exotic breeds while

stockbreeding system for African cattle is uncommon. So we may be expecting a continuous loss of cattle diversity in Africa, while the adaptive traits of these genetic resources remain poorly characterized. Therefore, there is an urgent need to establish conservation plan to preserve and utilize these genetic resources before they disappear (Hanotte, Dessie et al. 2010).

3.3.3 Current distribution of African cattle

The humpless *Bos taurus* includes two groups, humpless longhorns and humpless shorthorns, and nowadays, they are mainly distributed in West and Central Africa (Rege, Ayalew et al. 2006). Humped *Bos indicus* contains three groups, Large East African Zebu, Small East African Zebu, and West African Zebu, and they are widely distributed from East Africa to the dry parts of West Africa. Sanga cattle are mainly distributed in eastern and southern Africa, but they are also found in Central Africa. Zenga cattle, which are minorities within the four main African cattle categories, are only found in East Africa (Rege and Tawah 1999). The current distribution of the five African cattle groups including the recent derived breeds is shown at **Figure 3.2**; (Rege 1999, Hanotte, Bradley et al. 2002, Rege, Ayalew et al. 2006).

In general, the density of cattle population in East African highlands has increased compare to other regions because of cattle introduction and human migration (**Figure 3.1**). The two waves of *Bos indicus* migration led to a dominance of zebu type cattle among the African indigenous cattle and even replacement of *Bos taurus* cattle (Payne 1964, Epstein 1971, Hanotte, Bradley

et al. 2002). Y chromosome and autosomal studies have revealed the introgression pattern of zebu cattle. Y chromosome studies have shown dichotomy in African cattle Y chromosome morphology (Bradley, MacHugh et al. 1994, Hanotte, Okomo et al. 1997, Porto-Neto, Sonstegard et al. 2013). These genetic evidences suggest that zebu introgression in the continent of Africa is only through zebu male mediated interbreeding, and in conclusion, there is no true *Bos indicus* in Africa.

Meanwhile, admixture analyses of African cattle through microsatellite data have revealed a level of *Bos indicus* introgression between different areas (Ajmone-Marsan, Garcia et al. 2010), and as a result, the degree of zebu allele domination decreases from eastern Africa to western Africa (Hanotte, Bradley et al. 2002, Freeman, Bradley et al. 2006). Moreover, cattle which inhabit subtropical areas of West and Central Africa barely have zebu ancestry. Given the susceptibility of zebu cattle to trypanosomosis, tsetse-infested area in subtropical areas may have prevented introgression of zebu ancestry (MacHugh, Shriver et al. 1997).

3.3.4 Characteristics of African indigenous cattle

Each African indigenous cattle breed has unique physical features and sometimes the groups can be identified using outward characteristics (Rege, Ayalew et al. 2006). In addition to physical features, non-visible traits such as disease resistance, climatic stress resistance and economic traits also differ among breeds (**Table 3.1**; Rege 1999, Rege and Tawah 1999, Rege, Ayalew et

al. 2006). Some breeds are already known for their unique adaptive attributes or good economic performances, which may have been the result of evolutionary selection or breeding.

One of the well-known outstanding features of African cattle is trypanosomosis resistance. Trypanosomosis is a tsetse-transmitted disease in vertebrates. In cattle the main pathogenic trypanosomes are *Trypanosoma congolense* and *T. vivax*. Non-African cattle breeds are highly susceptible to trypanosomes infection. At the opposite African *Bos taurus* are tolerant while African *Bos indicus* are susceptible although not as susceptible than exotic breeds. For instance, N'Dama cattle (a long humpless longhorns), which are found in central and West Africa, are trypanotolerant (Mattioli, Pandey et al. 2000), which has led to an overall population increase and expansion of their geographical range (Rege and Tawah 1999). Other studies have reported that the Shoko cattle (an East African humpless shorthorns) shows also resistance to trypanosomosis (Lemecha, Mulatu et al. 2006). Unfortunately today, this cattle breed is being intensively crossbred with zebu cattle. Although zebu cattle are usually susceptible to trypanosomosis, in some zebu breeds that inhabit tsetse-infested regions, such as the Nuba Mountain Zebu, Mongolla or the Orma Boran, a degree of trypanotolerance has been reported or is expected, likely a result of local evolutionary adaption (Ruttledge 1928, Rege and Tawah 1999). Also, the Bovine of Tete, which is a zenga type of cattle found in Mozambique and Alur cattle found in Democratic Republic of Congo, are also thought to have some level of trypanotolerance (Lemecha, Mulatu et al. 2006).

Tick infestation represents another major challenge for African cattle. Unlike trypanosomosis, African *Bos indicus* cattle are believed to be more resistant to infestation by cattle ticks compared to taurine animal (Piper, Jonsson et al. 2009). However, among African taurine cattle, the N'Dama (humpless longhorns) is resistant to tick infestation (Mattioli, Pandey et al. 2000). In addition, Tswana cattle (South African sanga), from Botswana, have very high tolerance to heavy tick challenges (Rege, Ayalew et al. 2006). The Tswana is also known to have resistance to the endemic heartwater disease, as do Landim cattle (South African sanga) from Mozambique (Asselbergs, Jongejan et al. 1993).

Distinct evolutionary history between *B. taurus* and *B.indicus* have resulted in different degree thermotolerances at the cellular and physiological levels (Hansen 2004). *Bos indicus* breeds can effectively regulate their body temperature against thermal stress and are better adapted to hot weather than *Bos taurus* breeds (Hansen 2004). In addition, several breeds of zebu and zenga are able to withstand very harsh environmental conditions, and those characteristics have arisen through evolutionary adaptation. For examples, Karamajong Zebu cattle (Large East African Zebu) in Uganda are well adapted to very dry climates and they can survive if they are able to drink only once every two days(Thomas 1943). Turkana cattle (Large East African Zebu) in Kenya which are classified in the same subgroup of cattle, are thought to be able to survive on very poor pasture, scarce water and have good walking abilities (Rege, Ayalew et al. 2006). Angoni cattle (Small East African Zebu) in Zambia and Ugogo Grey cattle (Small East African Zebu) in

Tanzania are adapted to browsing during the long dry seasons (Felix 1995, Rege and Tawah 1999). In the case of Landim cattle (South African Sanga) in Mozambique, they are thought to have resistance to hot, humid weather as well as extended dry periods and foot-and-mouth-disease (FMD; Felix 1995, Rege, Ayalew et al. 2006). In Africa, cattle breeds have adapted to hot and dry weather, but others have adapted to even cold and wet conditions, such as Jem-Jem cattle (Small East African Zebu) in Ethiopia (Rege and Tawah 1999).

Although African indigenous breeds show generally worse economic performance compared to commercial cattle breeds, which have been continuously selected for single productivity trait (e.g. Angus cattle for beef production, Holstein-Friesian cattle for milk), some breeds of indigenous cattle in Africa are known to have both good dairy and beef characteristics (Collier and Gunning 1999). For example, the White Fulani cattle (West African Zebu), which inhabit Cameroon in the Central African Republic and Nigeria, show good performance both as a dairy and beef breed (Pullan and Grindle 1980). Other well-known beef-dairy cattle are Ogaden cattle (Small East African Zebu)(Rege and Tawah 1999), Jiddu cattle (Zenga)(Box 1968), and Fogera cattle (Zenga)(Gebremedhin, Hoekstra et al. 2007). While the Kenana and Butana cattle from Sudan are amongst the best milk producing zebu breeds (Musa, Ahmed et al. 2005).

3.3.5 The roles of cattle in Africa

Sub-Saharan Africa has very low agricultural productivity, even among developing countries, while gross agricultural output is growing (Otsuka 2006). Several factors, such as the degradation of agricultural land resources, declines in soil fertility, and the relatively debased performance of local livestock in comparison with commercial breeds, have been suggested as reasons for the poor agricultural performance in Sub-Saharan Africa.

As a result of shortages in food production, most countries in the region had to import food, and they became net importers of food (Ehui and Pender 2005). To make matters worse, the global food price crisis of 2007–2008 created a global issue and it critically influenced the food and economic situations in African countries (Ehui and Pender 2005, Minot 2011); the demand for livestock production has increased due to economic growth but the countries have problems meeting the growing demands (Steinfeld, Gerber et al. 2006).

Historically, in Africa, wild cattle for food production were domesticated earlier than plants because of the nomadic lifestyle of ancient African people and the risk associated with cultivating farmland under harsh ground conditions (Marshall and Hildebrand 2002). It is likely that some early African societies depended only on livestock given their nomadic lifestyle (Marshall and Hildebrand 2002), and these pastoral societies are still found today on the continent of Africa as exemplified by the Fulani and Masai communities.

African cattle have played a major role in the development agricultural

societies of the continent of Africa until now (**Table 3.1**; Rege 1999, Rege and Tawah 1999, Rege, Ayalew et al. 2006). Across the continent, cattle remain a major socio-cultural item and there are parts of many African ceremonies (e.g marriage, initiation). Besides these major socio-cultural roles, African cattle represent a major source of animal protein (dairy product and beef), workforce for cultivation and manure for fires and cooking. Above all, however, the cultivation of farmland is primarily dependent on cattle for most landed farmers who live in rural areas in Africa (Swallow 2000).

Because of this highly cattle-dependent agricultural system, losses of cattle through incidents such as diseases or maladjustments to climate, or even natural disasters, can be catastrophic for food production and farm household income. Basically, dairy production will decrease and the production of crops will also decrease because of reductions in arable land and the supply of manure as compost. In addition, a decline in manure production can impose a financial burden on farmers because of the need to purchase chemical fertilizers. All of these changes result in reduced farm household income, which negatively influences food and economic conditions in Africa (**Figure 3.3**).

3.4 Researches on African indigenous cattle

3.4.1 Previous researches on African cattle

In 2009, the genome sequence of taurine cattle was published by The Bovine Genome Sequencing and Analysis Consortium (Elsik, Tellam et al. 2009). Before long, bovine genome references became available (Liu, Qin et al. 2009, Zimin, Delcher et al. 2009) and the reference genome has been updated continuously. The cattle breed for the assembled sequence is Hereford cattle (Liu, Qin et al. 2009, Zimin, Delcher et al. 2009), which originally came from the United Kingdom (Heath-Agnew 1983)

Cattle phenotypes may be under the genetic control of a single or a few loci with major effect (e.g. coat color) or multiple loci each explaining a small production of the observed phenotypic variances (e.g. productivity such as milk yield and growth rates) (Ashwell, Heyen et al. 2004). The later are referred as quantitative trait loci or QTLs. Many QTLs have been mapped in cattle (see a cattle QTL database; Hu, Park et al. 2013) including for some of disease tolerance in indigenous African cattle (Reist-Marti, Simianer et al. 2003).

Phenotypes are controlled by genes (Crick 1970), and phenotypic variation among organisms sometimes reveals evolutionary adaptations or artificial selection (Glazier, Nadeau et al. 2002). Quantitative traits, such as the milk yield and growth rate, are controlled by many genes (Goddard and

Hayes 2009), and each gene that affects a complex trait usually has a small genetic effect (Raadsma 2004). QTL have been used to define complex traits (Raadsma 2004), and QTL association analysis is used in modern breeding technology to improve cattle (Kim, Farnir et al. 2003, Raadsma 2004). There were also some monogenic traits, which appear to be controlled by one gene, but only a few traits were truly monogenic (Scriver and Waters 1999, Glazier, Nadeau et al. 2002). Examples of monogenic traits include coat color, resistance to a certain pathogen (Haley, Afanador et al. 1994), or condition (Ammiraju, Veremis et al. 2003). However, in case of trypanosomosis, there is a previous study which reported QTLs contributing to trypanotolerance (Hanotte, Ronin et al. 2003).

3.4.2 Necessities for researches on African indigenous cattle

For commercial cattle breeds artificial selection and management interventions have lead in markedly improved productivity and by extension economic performance but at the cost of reduced genetic diversity and in some case fertility (Pryce, Royal et al. 2004, De Roos, Hayes et al. 2008). At the opposite, productivity of African cattle breeds has not been standardized towards specific breeding objective. Rather the main selection factor is here survival in often unpredictable and changing environmental conditions.

Some studies have already reported and supported the high genetic diversity in African cattle. From an examination of microsatellite DNA variation, African cattle breeds were reported to have higher genetic diversity

than cattle breeds in other regions (MacHugh, Shriver et al. 1997, Loftus, Ertugrul et al. 1999). As an example that explains genetic diversity, linkage disequilibrium is the nonrandom association of alleles at two or more loci (Falconer, Mackay et al. 1996). A study showed that LD values for African cattle breeds (N'Dama) tend to be lower than values for other cattle breeds (Goddard and Hayes 2009). Lower LD values in African cattle breeds also indicate that the inbreeding rate in African cattle is low. In other words, African cattle have larger gene pools as genetic resources than other cattle breeds, such as Holstein or Angus breeds.

Moreover, as we described before, each African indigenous cattle breeds has been well known to have various and unique characteristics, such as resistances to disease or climate stress, and experimental observations in previous studies have proven their known characteristics, for examples, trypanotolerance (Roberts and Gray 1973, Mattioli, Jaitner et al. 1998), cattle tick resistance (Mattioli, Bah et al. 1993, Mattioli, Jaitner et al. 1998, Mattioli, Pandey et al. 2000) and thermal resistance (Hansen 2004).

As a result, future cattle population would be expected to have to maintain high 'functional' genetic diversity, while adapting to their local environments. For this, there have been many trials to improve productivities of cattle through crossbreeding. However, improvement in a breed itself has faced many challenges because of comparatively adverse nutritional and climatic conditions (Scholtz and Theunissen 2010). Moreover, African indigenous cattle are endangered of extinction due to rash crossbreeding with

exotic breeds (Rege 1999). According to Reist-Marti, Simianer (2003), diversity of many of African taurine and zebu already reached marginal diversity of endangerment, and future diversity was expected to be a half of current diversity (Reist-Marti, Simianer et al. 2003).

African countries have tried to accomplish a Green Revolution by following the methods used in other regions such as Asia (Otsuka and Yamano 2005). However, differences in the natural environment and appropriate farming methods resulted in insufficient agricultural growth despite efforts for agricultural development (Otsuka and Yamano 2005). Thus, Africa is still struggling to emerge from poverty although many countries and organizations in the world have donated money and food to Africa.

From the Agricultural Revolution in 18th century England, I could deduce that cattle were the core element for increasing productivity in England. Similarly, in Africa, agriculture is deeply reliant on cattle, as mentioned before, and cattle could be ideal nutrient suppliers for soil and labor sources for plowing ground. Therefore, improving local African cattle through researches could be a practical solution to Africa's economic problem. Also, several cases illustrate the potential for agricultural growth through improving cattle. A new farming system based on the use of manure from improved dairy cows appeared in some regions of Sub-Saharan East Africa, such as Kenya and Uganda (Otsuka and Yamano 2005). This new farming system combines features of previous Agricultural Revolutions in other countries, and it is thought to be a new successful African region-specific

farming system (**Figures 3.4 and 3.5**). In Kenya, increased manure application from improved dairy cattle has had a strong impact on maize yield (Otsuka and Yamano 2005).

Thus, deeper genetic studies on African indigenous cattle would be needed to improve cattle breed against future climate change, preserve current African indigenous cattle as valuable genetic resources and even solve the problem of poverty in the continent of Africa.

3.4.3 The future prospects in African cattle researches.

After the advent of Next-Generation-Sequencing (NGS) technology (Shendure and Ji 2008), the production of genetic information became much easier (Mardis 2008) and it became possible to use not only partial genomic information but entire genomes from individuals (Davey, Hohenlohe et al. 2011). However, not so many African cattle studies based on whole genome sequence have been reported.

As previously mentioned, most African cattle have a *Bos indicus* ancestry, and the common ancestor of *Bos taurus* and *Bos indicus* dates back to over 330,000 or even 2 million years ago (Ajmone-Marsan, Garcia et al. 2010). This distance from the pedigree could represent a challenge for genomic analyses that use the reference sequence. Fortunately, previous study using NGS whole genome sequence showed 92 percent to 99 percent overall alignment rates between *Bos indicus* and the reference genome for taurine

except for the Y chromosome because of incomplete sequence assembly (Canavez, Luche et al. 2012). Thus, major difficulties are not expected when performing fundamental resequencing analyses using NGS methods.

Nevertheless, re-sequencing analyses using reference genome of European *Bos taurus* have clear limitations such as missing information of unaligned genome region or inaccurate reflection of structure variations. As an extreme case, single nucleotide substitutions occur at a mean rate of about 1.2% between human and chimpanzee genome (Sequencing and Consortium 2005), but the structure of the genome between human and chimpanzee differ considerably because of structural variants including genomic insertion, deletion and even inversions (Newman, Tuzun et al. 2005). These limitations could make exact identification of African cattle genome difficult. As a complementary method, de-novo genome assembly would be able to help overcome limitations in genomic research of African indigenous cattle.

Investigating genomic sequences of African indigenous cattle breeds to preserve African cattle breeds which are endangered and identifying target genes which are involved in resistance traits (e.g., heat and disease resistances) of African cattle and their biological process would be the main purpose of African genome researches. For of these purposes, recently, international cattle genome consortium was launched (**Figure 3.6**).

Scientifically, tracing the genes and loci of monogenic traits in African cattle would be relatively simple (Glazier, Nadeau et al. 2002), After conducting a comparative analysis of genetic variation in indigenous African

cattle populations using individual genomic sequences, which will be sequenced using NGS technology, evolutionary selective pressures will be able to be defined by using various and reliable statistical criteria (Back 1994) including methods which were explained in Chapter 1.

However, researching on complex traits in African cattle having high genetic diversity will be more challenging than in inbred cattle having low genetic diversity. In the case of dog species, approximately 30,000 SNPs are needed for between-breed analyses (Sutter, Eberle et al. 2004) because dogs are highly inbred animals, and this will be similar for most livestock (Zajc, Mellersh et al. 1997). In contrast, millions of SNPs would be needed for complex traits in *Bos taurus* and *Bos indicus* subspecies (Goddard and Hayes 2009). Moreover, information from association studies with SNPs have clear limitations for explaining complex traits (Center 1995, Andersson and Georges 2004, Manolio, Collins et al. 2009) even though several phenotypes are determined by only few genes (Kim, Farnir et al. 2003, Andersson and Georges 2004, Schenkel, Miller et al. 2006).

3.5 Conclusion and perspectives

Climatic environment of African continent is more diverse than of any other continents, and African indigenous cattle inhabiting these areas have already adapted to harsh and various environmental conditions. Thus, genetic resource of African indigenous cattle which have various and adapted characteristics is academically valuable. African genetic resource could also make modern commercial cattle breeds, which gained genetic weakness by excessive and sustained inbreeding, have a great breakthrough in the issue of coming climate change.

Given that many indigenous African cattle breeds are at risk of becoming extinct because of crossbreeding or interbreeding or even neglect, analyses of African cattle genome may contribute to preserving the genetic resource found in African cattle by preventing the dilution of African cattle breeds. With the current advances in NGS technology, this is the time to take an interest in African genetic resource that has enormous potential and start research on African cattle populations to keep them from going extinct.

The role of cattle in African agriculture is not only as a manure supplier; small farm households are also almost entirely reliant on them for cultivation (McCown, Haaland et al. 1979). Therefore, in perspective, genomic researches on African indigenous cattle would be able to contribute to building the knowledge infra for the economic development in the continent of Africa considering the circumstance that most impoverished

African countries have an agricultural economic base and rely on cattle.

Although lots of time and effort will be needed for researching on African indigenous cattle breeds, it is time to investigate the genomic contents of them considering its academic, ecological and even economical values before African indigenous cattle become extinct.

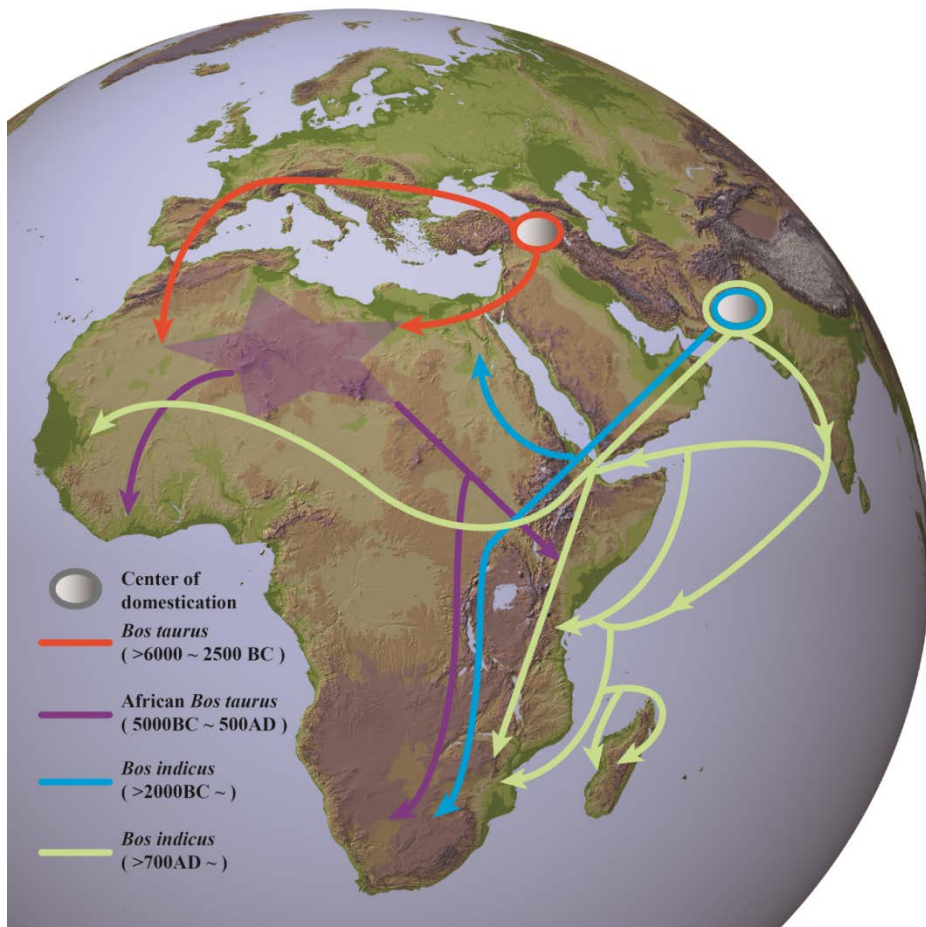


Figure 3.1. Approximate migration route and the origin of Africa domestic cattle (adapted and modified from ILRI, 2006. Safeguarding livestock diversity: the time is now).

Circle regions represent the expected center of cattle domestication. Migration route from the outside Africa and within Africa was shown by color arrows, and the color of arrow represents the migration time and its origin.

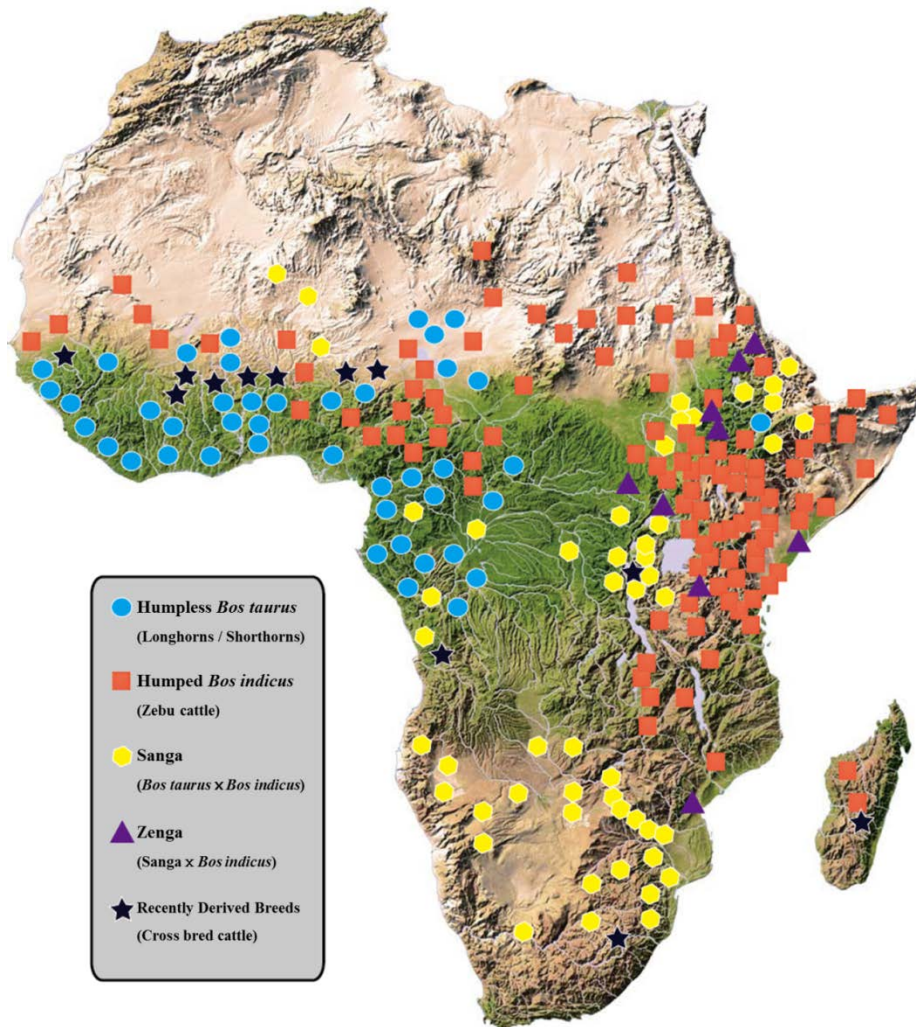


Figure 3.2. Distributions of indigenous cattle in sub-Saharan Africa

Marks on the map represent African cattle distribution, and each mark shows the type of cattle which inhabit the region. Cattle of North Africa and imported commercial cattle breeds are not shown in this figure.

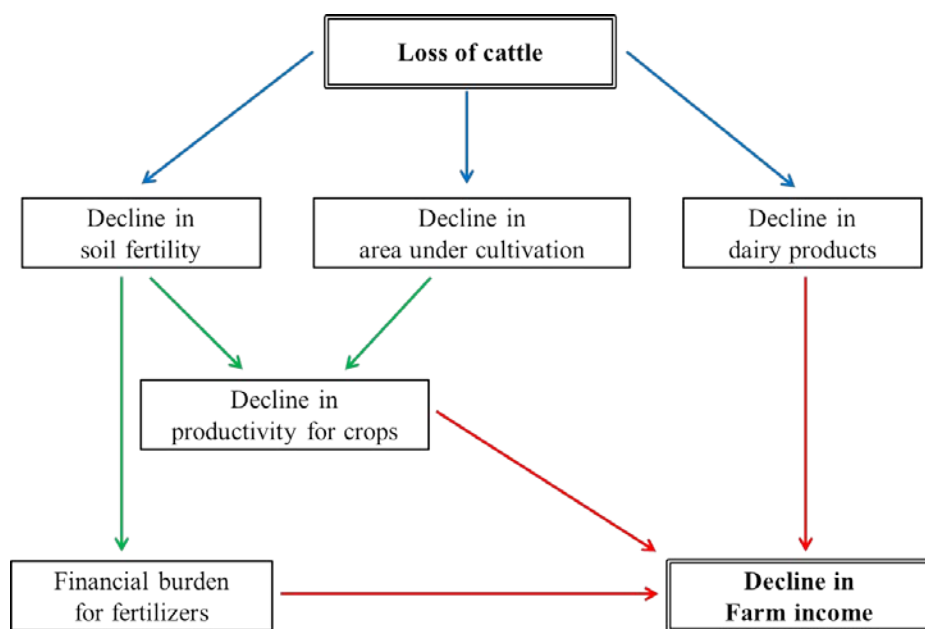


Figure 3.3. Summary of impacts from the loss of cattle in African agricultural systems.

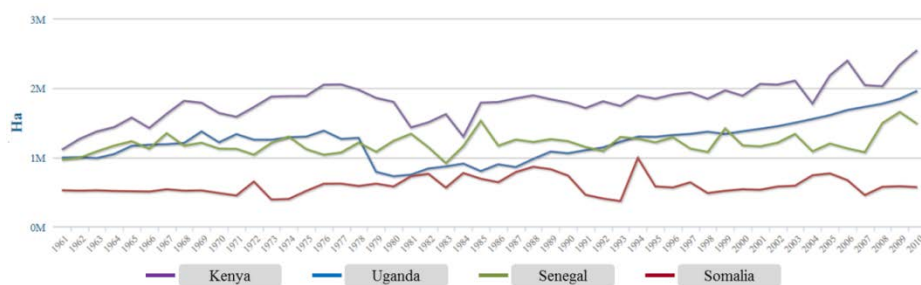


Figure 3.4. Comparison of total area harvested among countries in Sub-Saharan Africa (FAOSTAT)

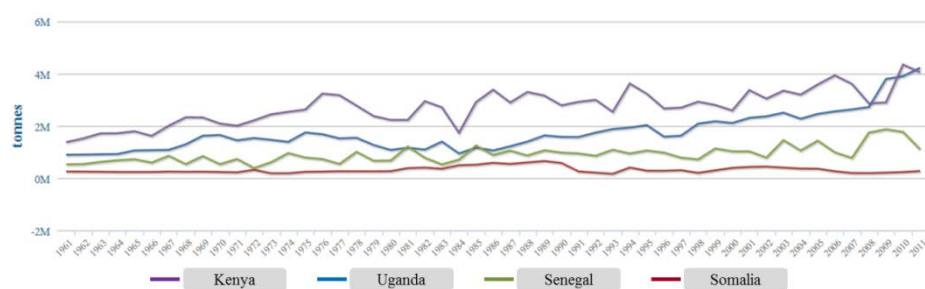


Figure 3.5. Comparison of total cereal production among countries in Sub-Saharan Africa (FAOSTAT)



African Cattle Genome Consortium

the African Cattle Genome Consortium (ACGC) is an international group of academic organizations committed to defining genome structures and useful genes of African indigenous cattle breeds to improve cattle for adaption to future climate change and high agricultural productivity in African countries

Figure 3.6. African cattle genome consortium

Table 3.1. Uses and Characteristics of Indigenous African Cattle by African Countries (Excluding Commercial Components)

Country	Group	Subgroup	Breed	Main Uses (in order of importance)	Note
Angola	South African Sanga	Setswana group	Barotse	Milk, Meat, Work	Docile, good work animal
		Ovambo group	Humbi	Meat, Work, Milk	-
		Zambia/Angola group	Porto-Amboim	Meat	Poor milkers
	Recently Derived Breeds	-	Barra do Cuanzo	Meat, Work, Milk	Good beef characteristics
Benin	Humpless Shorthorns	-	Lagune	Meat, Manure	-
			Somba	Meat, Milk, Rituals	-
	Recently Derived Breeds	Taurine × Zebu cross	Borgou	Meat, Work, Manure	Little sexual dimorphism
Botswana	South African Sanga	Setswana group	Tswana	Meat, Work, Milk	Tolerance to heavy tick challenge, resistance to the endemic heartwater
Burkina Faso	Humpless Longhorns	-	N'Dama	Meat, Work, Manure	Tolerance to trypanosomosis
	Humpless Shorthorns	-	Baoulé	Milk, Meat	-
		A strain of the Baoule	Lobi	Meat, Milk	-
	West African Zebu	Fulani	Sudanese Fulani (Peul Zebu)	Milk, Work, Meat	Walking ability
	Recently Derived Breeds	Taurine × Zebu cross	Méré	Meat, Work, Manure	Variable phenotype from parental breeds
Burundi	East African Sanga	Ankole group	Ruzizi	Milk, Meat, Work	-
			*Watusi	Milk, Meat, Work	-
Central African Republic	Humpless Shorthorns	-	Baoulé	Milk, Meat	-
			N'Dama	Meat, Work, Manure	Tolerance to trypanosomosis
			Gudali	Milk, Meat, Work	-
	West African Zebu	Fulani	Ngaundere	Milk, Meat, Work, Manure	Nervous and intractable temperament, poor milkers
			Red Fulani	Milk, Meat, Work, Manure	
			White Fulani	Milk, Meat, Work, Manure	Good dairy and beef characteristics

*: Vulnerable; **: Endangered; ***: Critically endangered

Country	Group	Subgroup	Breed	Main Uses (in order of importance)	Note
Cameroon	Humpless Longhorns	-	Kuri	Meat, Milk, Work	Excellent swimmers, intolerant to heat and sunlight
	Humpless Shorthorns	-	***Bakosi	Meat, Rituals	-
			***Bakweri	Meat, Rituals	-
			**Doayo (Namichi)	Meat, Rituals	-
			***Kapsiki	Meat, Rituals	-
			Ngaundere	Milk, Meat, Work	-
	West African Zebu	Gudali	Yola	Milk, Meat, Work	Tolerance to trypanosomosis
		Fulani	Red Fulani	Milk, Meat, Work, Manure	Nervous and intractable temperament, poor milkers
			White Fulani	Milk, Meat, Work, Manure	Good dairy and beef characteristics
Chad		Sahel Zebu	Shuwa	Work, Milk, Meat	Good dairy characteristics
	Humpless Longhorns	-	Kuri	Meat, Milk, Work	Excellent swimmers, intolerant to heat and sunlight
	Humpless Shorthorns	-	**Logone	Meat	-
	West African Zebu	Fulani	Fellata (Bororo)	Milk, Meat, Work	-
		Sahel Zebu	Shuwa	Work, Milk, Meat	Good dairy characteristics
Cote d'Ivoire	Recently Derived Breeds	Taurine × Zebu cross	Méré	Meat, Work, Manure	Variable phenotype from parental breeds
D.R. Congo	Humpless Shorthorns	-	Lagune	Meat, Manure	-
	Small East African Zebu	Southern Sudan Zebu	Lugware	Work, Meat, Milk	-
	East African Sanga	Ankole group	*Bahima	Milk, Meat, Work	Susceptible to rinderpest and trypanosomosis
			*Bashi	Milk, Meat, Work	-
			*Ruzizi	Milk, Meat, Work	-
			*Watusi	Milk, Meat, Work	-
	Zenga	-	Alur	Milk, Meat, Work	tolerance to trypanosomosis

*: Vulnerable; **: Endangered; ***: Critically endangered

Country	Group	Subgroup	Breed	Main Uses (in order of importance)	Note
Ethiopia	Humpless Shorthorns		Sheko	Meat, Work	Tolerance to trypanosomosis
	Large East African Zebu	Boran	Ethiopian Boran	Milk, Meat	Good dairy characteristics
		South Sudan Zebu	Murle	Milk, Meat, Work	-
			Adwa	Work, Meat, Milk	Milk yield is low
			Ambo	Work, Meat, Milk	Larger than adjacent Guraghe breed
			Arsi	Work, Meat, Milk	Poor milkers, extremely active and often very aggressive
			Bale	Work, Meat, Milk	-
			Goffa	Work, Milk, Meat	-
	Small East African Zebu	Abyssinian Shorthorned Zebu	Guraghe	Work, Meat, Milk	-
			Hammer	Milk, Work, Meat	-
			Harar	Work, Meat, Milk	-
			Jem-Jem	Work, Meat, Milk	Well adapted to the wet and cold climate
			Jijiga	Work, Milk, Meat	-
			Mursi	Milk, Meat, Work	-
			Ogaden	Milk, Meat, Work	Good dairy and beef characteristics
			Smada	Work, Meat, Milk	Milk yield is quite low
	East African Sanga	Nilotic Sanga	Anuak	Milk, Meat, Work	-
		Abyssinian Sanga	Danakil	Milk, Meat, Work	-
			Raya-Azebo	Work, Milk, Meat	Bigger body size
	Zenga	-	Arado	Work, Milk, Meat	Docile, good work animal, milk yield is low
			Fogera	Work, Milk, Meat	Docile, good dairy and beef characteristics
			Horro	Work, Milk, Meat	Variable milk production
Eritrea	Large East African Zebu	North Sudan Zebu	Barka	Milk, Meat	Active disposition
	Small East African Zebu	Somali Shorthorned Zebu group	Baherie	Milk, Meat	-
	East African Sanga	Abyssinian Sanga	Danakil	Milk, Meat, Work	-
	Zenga	-	Arado	Work, Milk, Meat	docile, good work animal, milk yield is low

*: Vulnerable; **: Endangered; ***: Critically endangered

Country	Group	Subgroup	Breed	Main Uses (in order of importance)	Note
Gabon	Humpless Shorthorns	-	Baoulé	Milk, Meat	-
			Lagune	Meat, Manure	-
Ghana	Humpless Shorthorns	-	***Ghana Dwarf Muturu	Meat	-
			***Ghana Shorthorn	Meat, Work, Milk	Good beef-type conformation
	Recently Derived Breeds	Taurine × Zebu cross	Ghana “Sanga”	Milk, Meat, Work, Manure	-
Kenya	Large East African Zebu	Boran	Improved Boran	Meat, Milk, Work	Excellent tropical beef characteristics
			Orma Boran	Milk, Meat	Tolerance to trypanosomosis
			Kenyan Boran	Milk, Meat	Good dairy characteristics, excellent walking ability, drought resistance, herd instinct, excellent mothering ability, heat tolerance, highly adaptable to a variety of climates, extremely long life, large sex dimorphism, survive on very poor pasture and scarce water, walking ability
		Karamajong Zebu	Turkana	Milk, Meat	
	Small East African Zebu	Kenyan Zebu	Duruma	Meat, Milk, Work	-
			Giriamaa	Meat, Milk, Work	-
			Kamba	Meat, Milk, Work	-
			**Kikuyu	Meat	-
			Kenyan Maasai	Milk, Meat	Heterogeneous phenotypes
			Nandi	Milk, Work, Meat	-
			Samburu	Milk, Meat	Heterogeneous phenotype
			Taita	Meat, Milk, Work	-
			**Watende	Milk, Meat, Work	-
			Winam (Kavirondo)	Milk, Work, Meat	-
		Teso Group	Teso	Milk, Meat, Work	-

*: Vulnerable; **: Endangered; ***: Critically endangered

Country	Group	Subgroup	Breed	Main Uses (in order of importance)	Note
Ivory Coast	Humpless Shorthorns	-	Baoulé	Milk, Meat	-
			Lagune	Meat, Manure	-
Lesotho	Recently Derived Breeds	-	Basuto	Work, Meat, Milk	Low milk production and poor muscular development
Liberia	Humpless Shorthorns	-	**Liberia Dwarf Muturu	Meat	-
Madagascar	Small East African Zebu	Madagascar group	***Baria	Not used (Wild)	Wild, free-roaming humped cattle
			Madagascar zebu	Meat, Milk, Work	-
	Recently Derived Breeds	European taurine x Zebu cross	*Rana (Omby Rana)	Milk, Meat	Fine bodied with dairy conformation
Malawi	Small East African Zebu	Angoni group	Malawi Zebu	Meat, Work, Milk	-
Mali	Humpless Longhorns	-	N'Dama	Meat, Work, Manure	Tolerance to trypanosomosis
		Sahel Zebu	Azaouak (Tuareg)	Meat, Work, Milk	Very well adapted to drought
	West African Zebu	Maure	Maure	Milk, Meat, Work	-
		Fulani	Sudanese Fulani (Peul Zebu)	Milk, Work, Meat	Walking ability
Mauritania	West African Zebu	Sahel Zebu	Maure	Milk, Meat, Work	-
	Small East African Zebu	Angoni group	Angonia (Angone)	Meat, Work, Milk	-
Mozambique	South African Sanga	Nguni group	*Landim	Meat, Work, Milk	Extremely hardy and withstands hot, humid weather as well as extended dry periods, very resistant to FMD
	Zenga	NA	Bovines of Tete	Meat, Milk, Work	Tolerance to trypanosomosis
Namibia	South African Sanga	Ovambo group	Caprivi	Meat, Work, Milk	-
			Kaokoveld	Meat, Work, Milk	-
			Okavango	Meat, Work, Milk	-
			Ovambo	Meat, Work, Milk	-
		Setswana group	**Damara (Herero)	Milk, Work, Meat	-

*: Vulnerable; **: Endangered; ***: Critically endangered

Country	Group	Subgroup	Breed	Main Uses (in order of importance)	Note
Somalia	Large East African Zebu	Boran	Somali Boran	Milk, Meat	-
		Somali Shorthorned zebu group	Garre (Ghera)	Milk, Meat	-
	Small East African Zebu		Gasara	Milk, Meat	-
		North Somali Zebu	Milk, Work	Less variable in morphology	
	Abyssinian Shorthorned Zebu	Ogaden	Milk, Meat, Work	Good dairy and beef characteristics	
	Zenga	NA	Jiddu	Milk, Meat	Fleshy, well proportioned, dairy-type conformation
South Africa	South African Sanga	Africander group	Afrikaner	Meat, Work	Walking and grazing ability, exceptionally good mothering ability, longevity, good quality beef
		Nguni group	Nguni	Meat, Work, Milk	Exceptional fertility under harsh conditions, early sexual maturity of particularly heifers, good foraging and walking ability, good mothering ability
			***Pedi	Meat, Work, Milk	-
			***Shangan	Meat, Work, Milk	-
Sudan	Large East African Zebu	North Sudan Zebu	Baggara	Milk, Meat, Work	-
		South Sudan Zebu	Butana	Milk, Meat	Dairy characteristics
			Kenana	Milk, Meat	Good dairy-type characteristics
		Karamajong Zebu	Murle	Milk, Meat, Work	-
	Small East African Zebu	Southern Sudan Zebu	Toposa	Milk, Meat, Work	Good beef characteristics
			Mongolla	Milk, Meat	Tolerance to trypanosomosis, well fleshed
			Nuba Mountain Zebu	Meat (for ceremonies)	Tolerance to trypanosomosis
	West African Zebu	Fulani	Fellata (Bororo)	Milk, Meat, Work	-
	East African Sanga	Nilotic Sanga	Aliab Dinka	Milk, Meat	-
Aweli Dinka			Milk, Meat	-	
Nuer			Milk, Meat	-	

*: Vulnerable; **: Endangered; ***: Critically endangered

Country	Group	Subgroup	Breed	Main Uses (in order of importance)	Note
Swaziland	South African Sanga	Nguni group	Nguni	Meat, Work, Milk	Exceptional fertility under harsh conditions, early sexual maturity of particularly heifers, good foraging and walking ability, good mothering ability
Tanzania	Large East African Zebu	Boran	Improved Boran	Meat, Milk, Work	Excellent tropical beef characteristics
			***Chagga (Wachagga)	Milk, Work, Meat	Interbreeding and crossbreeding caused declining population size
			Iringa Red	Milk, Meat	-
			Tanzanian Maasai	Milk, Meat	-
			***Mkalama Dun	Milk, Meat, Work	-
	Small East African Zebu	Tanganyika Zebu	***Pare	Milk, Meat, Work	-
			*Singida White	Milk, Meat, Work	-
			*Tarime (Shashi)	Milk, Meat, Work	-
			Ugogo Grey	Milk, Meat	Well adapted to harsh environments, selective grazers with ability to browse during the dry season
			Zanzibar Zebu	Milk, Work, Meat	-
	East African Sanga	Ankole group	*Watusi	Milk, Meat, Work	-
	Zenga	-	Sukuma (Tinde)	Milk, Meat	-
	Humpless Shorthorns	-	Lagune	Meat, Manure	-
			Somba	Meat, Milk, Rituals	-

*: Vulnerable; **: Endangered; ***: Critically endangered

Country	Group	Sub-Group	Breed	Main Uses (in order of importance)	Note
Uganda	Large East African Zebu	Karamajong group	Karamajong zebu	Milk, Meat	Adapted to a very dry climate
	Small East African Zebu	Teso group	*Kyoga	Milk, Meat, Work	-
			**Serere	Milk, Meat, Work	-
			Teso	Milk, Meat, Work	-
			*Usuk	Milk, Meat, Work	-
	Small East African Zebu	Southern Sudan Zebu	Lugware	Work, Meat, Milk	-
			Nkedi	Work, Milk, Meat	-
	East African Sanga	Ankole group	*Bahima	Milk, Meat, Work	Susceptible to rinderpest and trypanosomosis
			Kigezi	Milk, Meat, Work	-
			*Watusi	Milk, Meat, Work	-
Zaire	Zenga	-	Nganda	Milk, Meat, Work	-
	Humpless Shorthorns	-	Lagune	Meat, Manure	-
Zambia	Large East African Zebu	Boran	Improved Boran	Meat, Milk, Work	Excellent tropical beef characteristics
	Small East African Zebu	Angoni group	Angoni	Meat, Work, Milk	Adapted to browsing during long dry season
	South African Sanga	Setswana group	Barotse	Milk, Meat, Work	Docile, good work animal
		Zambia/Angola group	Tonga	Milk, Meat, Work	-
Zimbabwe	South African Sanga	Shona group	Mashona	Meat, Work	High fertility
		Nguni group	***Nkone	Meat, Milk	Good beef characteristics
		Setswana group	Tuli	Meat	high fertility, good mothering ability, low calf mortality

*: Vulnerable; **: Endangered; ***: Critically endangered

References

Aifantis, I., et al. (2008). "Molecular pathogenesis of T-cell leukaemia and lymphoma." Nature Reviews Immunology **8**(5): 380-390.

Ajmone-Marsan, P., et al. (2010). "On the origin of cattle: how aurochs became cattle and colonized the world." Evolutionary Anthropology: Issues, News, and Reviews **19**(4): 148-157.

Akey, J. M., et al. (2002). "Interrogating a high-density SNP map for signatures of natural selection." Genome research **12**(12): 1805-1814.

Ammiraju, J., et al. (2003). "The heat-stable root-knot nematode resistance gene Mi-9 from *Lycopersicon peruvianum* is localized on the short arm of chromosome 6." Theoretical and Applied Genetics **106**(3): 478-484.

Andersson, L. and M. Georges (2004). "Domestic-animal genomics: deciphering the genetics of complex traits." Nature Reviews Genetics **5**(3): 202-212.

Ashburner, M., et al. (2000). "Gene Ontology: tool for the unification of biology." Nature genetics **25**(1): 25-29.

Ashwell, M., et al. (2004). "Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle." Journal of dairy science **87**(2): 468-475.

Asselbergs, M., et al. (1993). "Antibodies to Cowdria ruminantium in Mozambican goats and cattle detected by immunofluorescence using endothelial cell culture antigen." Tropical animal health and production **25**(3): 144-150.

Back, T. (1994). Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. Evolutionary Computation, 1994.

IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on, IEEE.

Barrett, R. D. and D. Schluter (2008). "Adaptation from standing genetic variation." Trends in Ecology & Evolution **23**(1): 38-44.

Blench, R. M. and K. C. MacDonald (2000). The origins and development of African livestock: archaeology, genetics, linguistics and ethnography, Routledge.

Boichard, D., et al. (1989). "Genetic parameters for first lactation dairy traits in the Alpine and Saanen goat breeds." Genet. Sel. Evol **21**(2): 205-215.

Box, T. W. (1968). "Range resources of Somalia." Journal of Range Management: 388-392.

Bradley, D., et al. (1994). "Zebu-aurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations." Animal genetics **25**(S2): 7-12.

Brenner, S., et al. (2000). "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays." Nature biotechnology **18**(6): 630-634.

Browning, S. R. and B. L. Browning (2007). "Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering." The American Journal of Human Genetics **81**(5): 1084-1097.

Buffum, B. C. (1909). Arid Agriculture: A Hand-book for the Western Farmer and Stockman, The author.

Canavez, F. C., et al. (2012). "Genome sequence and assembly of *Bos indicus*." Journal of Heredity **103**(3): 342-348.

Center, C. (1995). "Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results." Nature genetics **11**: 241.

Chen, H., et al. (2010). "Population differentiation as a test for selective sweeps." Genome research **20**(3): 393-402.

Cingolani, P., et al. (2012). "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3." Fly **6**(2): 80-92.

Collier, P. and J. W. Gunning (1999). "Explaining African economic performance." Journal of economic literature: 64-111.

Consortium, G. P. (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Crick, F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-563.

Danecek, P., et al. (2011). "The variant call format and VCFtools." Bioinformatics **27**(15): 2156-2158.

Davé, V., et al. (2004). "Nuclear factor of activated T cells regulates transcription of the surfactant protein D gene (*Sftpd*) via direct interaction with thyroid transcription factor-1 in lung epithelial cells." Journal of Biological Chemistry **279**(33): 34578-34588.

Davey, J. W., et al. (2011). "Genome-wide genetic marker discovery and genotyping using next-generation sequencing." Nature Reviews Genetics **12**(7): 499-510.

De Roos, A., et al. (2008). "Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle." Genetics **179**(3): 1503-1512.

Decker, J. E., et al. (2014). "Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle." PLoS genetics **10**(3): e1004254.

Dennis Jr, G., et al. (2003). "DAVID: database for annotation, visualization, and integrated discovery." Genome Biol **4**(5): P3.

DePristo, M. A., et al. (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nature genetics **43**(5): 491-498.

Dong, Y., et al. (2012). "Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*)."
Nature biotechnology.

Dong, Y., et al. (2013). "Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*)."
Nature biotechnology **31**(2): 135-141.

Edwards, C. J., et al. (2007). "Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs." Proceedings of the Royal Society B: Biological Sciences **274**(1616): 1377-1385.

Ehui, S. and J. Pender (2005). "Resource degradation, low agricultural productivity, and poverty in sub-Saharan Africa: pathways out of the spiral." Agricultural Economics **32**(s1): 225-242.

Elsik, C. G., et al. (2009). "The genome sequence of taurine cattle: a window to ruminant biology and evolution." Science **324**(5926): 522-528.

Emmons, D. and E. Lister (1976). "Quality of protein in milk replacers for young calves. I. Factors affecting in vitro curd formation by rennet (chymosin, rennin) from reconstituted skim milk powder." Canadian Journal of Animal Science **56**(2): 317-325.

Epstein, H. (1971). "Origin of the domestic animals of Africa."

Evanno, G., et al. (2005). "Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study." Molecular ecology **14**(8): 2611-2620.

Falconer, D. S., et al. (1996). "Introduction to Quantitative Genetics (4th edn)." TRENDS in Genetics **12**(7): 280.

Fay, J. C. and C.-I. Wu (2000). "Hitchhiking under positive Darwinian selection." Genetics **155**(3): 1405-1413.

Felius, M. (1995). Cattle breeds: an encyclopedia, C Misset bv.

Fleischmann, R. D., et al. (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." Science **269**(5223): 496-512.

Flint, A. and J. Woolliams (2008). "Precision animal breeding." Philosophical Transactions of the Royal Society B: Biological Sciences **363**(1491): 573-590.

Freeman, A., et al. (2006). "Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle." Animal genetics **37**(1): 1-9.

Fu, Y.-X. and W.-H. Li (1993). "Statistical tests of neutrality of mutations." Genetics **133**(3): 693-709.

Gebremedhin, B., et al. (2007). "Heading towards commercialization? The case of live animal marketing in Ethiopia."

Glazier, A. M., et al. (2002). "Finding genes that underlie complex traits." Science **298**(5602): 2345-2349.

Goddard, M. E. and B. J. Hayes (2009). "Mapping genes for complex traits in domestic animals and their use in breeding programmes." Nature Reviews Genetics **10**(6): 381-391.

Goldstein, D. B. (2009). "Common genetic variation and human traits." New England Journal of Medicine **360**(17): 1696.

GUIDE, L. P. (2004). "GOATS: SUSTAINABLE PRODUCTION OVERVIEW."

Habier, D., et al. (2007). "The impact of genetic relationship information on genome-assisted breeding values." Genetics **177**(4): 2389-2397.

Haley, S. D., et al. (1994). "Selection for monogenic pest resistance traits with coupling-and repulsion-phase RAPD markers." Crop Science **34**(4): 1061-1066.

Hanotte, O., et al. (2002). "African pastoralism: genetic imprints of origins and migrations." Science **296**(5566): 336-339.

Hanotte, O., et al. (2010). "Time to tap Africa's livestock genomes." Science(Washington) **328**(5986): 1640-1641.

Hanotte, O., et al. (1997). "A polymorphic Y chromosome microsatellite locus in cattle."

Hanotte, O., et al. (2003). "Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle." Proceedings of the National Academy of Sciences **100**(13): 7443-7448.

Hansen, P. (2004). "Physiological and cellular adaptations of zebu cattle to thermal stress." Animal reproduction science **82**: 349-360.

Hayes, B. J., et al. (2013). "The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation." TRENDS in Genetics **29**(4): 206-214.

Heath-Agnew, E. (1983). A history of Hereford cattle and their breeders, Duckworth.

Hiendleder, S., et al. (2008). "Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication." Cytogenetic and Genome Research **120**(1-2): 150-156.

Houghton, J. T., et al. (1992). Climate change 1992: the supplementary report to the IPCC scientific assessment, Cambridge University Press.

Hu, Z.-L., et al. (2013). "Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era." Nucleic acids research **41**(D1): D871-D879.

Hubbard, T., et al. (2002). "The Ensembl genome database project." Nucleic acids research **30**(1): 38-41.

ILRI, N. K. (2006). Safeguarding livestock diversity: the time is now, ILRI (aka ILCA and ILRAD).

Jackson, J. E. (2005). A user's guide to principal components, John Wiley & Sons.

Jolliffe, I. (2005). Principal component analysis, Wiley Online Library.

Karolchik, D., et al. (2003). "The UCSC genome browser database." Nucleic acids research **31**(1): 51-54.

Kijas, J. W., et al. (2013). "Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs." Animal genetics **44**(3): 325-335.

Kim, J.-J., et al. (2003). "Detection of quantitative trait loci for growth and beef carcass fatness traits in a cross between Bos taurus (Angus) and Bos indicus (Brahman) cattle." Journal of Animal Science **81**(8): 1933-1942.

Lander, E. S., et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nature methods **9**(4): 357-359.

Lemecha, H., et al. (2006). "Response of four indigenous cattle breeds to natural tsetse and trypanosomosis challenge in the Ghibe valley of Ethiopia." Veterinary parasitology **141**(1): 165-176.

Li, H., et al. (2009). "The sequence alignment/map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Linseele, V. (2004). "Size and size change of the African aurochs." Journal of African Archaeology **2**(2): 165-185.

Liu, L., et al. (2012). "Comparison of next-generation sequencing systems." BioMed Research International **2012**.

Liu, Y., et al. (2009). "Bos taurus genome assembly." BMC genomics **10**(1): 180.

Loftus, R., et al. (1999). "A microsatellite survey of cattle from a centre of origin: the Near East." Molecular ecology **8**(12): 2015-2022.

Loftus, R. T., et al. (1994). "Evidence for two independent domestications of cattle." Proceedings of the National Academy of Sciences **91**(7): 2757-2761.

MacHugh, D. E., et al. (1997). "Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*)." Genetics **146**(3): 1071.

Manolio, T. A., et al. (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Mardis, E. R. (2008). "The impact of next-generation sequencing technology on genetics." TRENDS in Genetics **24**(3): 133-141.

Mardis, E. R. (2008). "Next-generation DNA sequencing methods." Annu. Rev. Genomics Hum. Genet. **9**: 387-402.

Marshall, F. and E. Hildebrand (2002). "Cattle before crops: the beginnings of food production in Africa." Journal of World Prehistory **16**(2): 99-143.

Mattioli, R., et al. (1993). "A comparison of field tick infestation on N'Dama, Zebu and N'Dama× Zebu crossbred cattle." Veterinary parasitology **47**(1): 139-148.

Mattioli, R., et al. (1998). "Trypanosome infections and tick infestations: susceptibility in N'Dama, Gobra zebu and Gobra× N'Dama crossbred cattle exposed to natural challenge and maintained under high and low surveillance of trypanosome infections." Acta Tropica **71**(1): 57-71.

Mattioli, R. C., et al. (2000). "Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (*Bos taurus*) and trypanosusceptible Gobra zebu (*Bos indicus*) cattle." Acta Tropica **75**(3): 263-277.

Maxam, A. M. and W. Gilbert (1977). "A new method for sequencing DNA." Proceedings of the National Academy of Sciences **74**(2): 560-564.

McCown, R., et al. (1979). The interaction between cultivation and livestock production in semi-arid Africa, Springer.

McDonald, J. H. and M. Kreitman (1991). "Adaptive protein evolution at the Adh locus in *Drosophila*." Nature **351**(6328): 652-654.

McKenna, A., et al. (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome research **20**(9): 1297-1303.

Minot, N. (2011). "Transmission of world food price changes to markets in Sub-Saharan Africa." Washington, IFPRI **34**.

Murray, M., et al. (1984). "Genetic resistance to African trypanosomiasis." Journal of infectious diseases **149**(3): 311-319.

Musa, L., et al. (2005). "The reproductive and milk performance merit of Butana cattle in Sudan." ARCHIV FUR TIERZUCHT **48**(5): 445.

Nachman, M. W. (2001). "Single nucleotide polymorphisms and recombination rate in humans." TRENDS in Genetics **17**(9): 481-485.

Nardone, A., et al. (2010). "Effects of climate changes on animal production and sustainability of livestock systems." Livestock Science **130**(1): 57-69.

Newman, T. L., et al. (2005). "A genome-wide survey of structural variation between human and chimpanzee." Genome research **15**(10): 1344-1356.

Nielsen, R., et al. (2005). "A scan for positively selected genes in the genomes of humans and chimpanzees." PLoS biology **3**(6): e170.

Njogu, A., et al. (1985). "Trypanotolerance in East African Orma Boran cattle." Veterinary Record **117**(24): 632-636.

Nybom, H. and I. V. Bartish (2000). "Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants." Perspectives in plant ecology, evolution and systematics **3**(2): 93-114.

Okello, S. and E. Sabiiti (2006). "Milk production of indigenous Ankole cattle on Uganda as influenced by seasonal variations in temperature, rainfall and feed quality, Makerere Univ." Res. J **1**: 73-92.

Otsuka, K. (2006). "Why Can't We Transform Traditional Agriculture in Sub-Saharan Africa?" Applied Economic Perspectives and Policy **28**(3): 332-337.

Otsuka, K. and T. Yamano (2005). "The possibility of a green revolution in sub-Saharan Africa: Evidence from Kenya." Journal of Agricultural and Development Economics **2**(1): 8-19.

Otten, D. and H. F. Van den Weghe (2011). "The Sustainability of Intensive Livestock Areas (ILAS): Network System and Conflict Potential from the Perspective of Animal Farmers." International Journal on Food System Dynamics **2**(1): 36-51.

Payne, W. (1964). "The origin of domestic cattle in Africa." Empire Journal of Experimental Agriculture **32**(126): 97-113.

Pickrell, J. K., et al. (2009). "Signals of recent positive selection in a worldwide sample of human populations." Genome research **19**(5): 826-837.

Piper, E. K., et al. (2009). "Immunological profiles of Bos taurus and Bos indicus cattle infested with the cattle tick, Rhipicephalus (Boophilus) microplus." Clinical and Vaccine Immunology **16**(7): 1074-1086.

Porto-Neto, L. R., et al. (2013). "Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping." BMC genomics **14**(1): 876.

Price, A. L., et al. (2006). "Principal components analysis corrects for stratification in genome-wide association studies." Nature genetics **38**(8): 904-909.

Pruitt, K. D., et al. (2012). "NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy." Nucleic acids research **40**(D1): D130-D135.

Pryce, J., et al. (2004). "Fertility in the high-producing dairy cow." Livestock Production Science **86**(1): 125-135.

Pullan, N. and R. Grindle (1980). "Productivity of white Fulani cattle on the Jos plateau, Nigeria. IV. Economic factors." Tropical animal health and production **12**(3): 161-170.

Purcell, S., et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." The American Journal of Human Genetics **81**(3): 559-575.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Raadsma, H. W. (2004). "Quantitative trait loci mapping in dairy cattle: review and meta-analysis." Genet. Sel. Evol **36**: 163-190.

Rege, J. (1999). "The state of African cattle genetic resources I. Classification framework and identification of threatened and extinct breeds." Animal Genetic Resources Information **25**: 1-25.

Rege, J., et al. (2006). DAGRIS (Domestic Animal Genetic Resources Information System). International Livestock Research Institute, Addis Ababa, Ethiopia.

Rege, J. and C. Tawah (1999). "The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains." Animal Genetic Resources Information **26**(1): 1-25.

Reist-Marti, S., et al. (2003). "Weitzman's approach and conservation of breed diversity: an application to African cattle breeds." Conservation Biology **17**(5): 1299-1311.

Renaudeau, D., et al. (2012). "Adaptation to hot climate and strategies to alleviate heat stress in livestock production." Animal **6**(05): 707-728.

Roberts, C. and A. Gray (1973). "Studies on trypanosome-resistant cattle. II. The effect of trypanosomiasis on N'Dama, Muturu and Zebu cattle." Tropical animal health and production **5**(4): 220-233.

Ruttledge, W. (1928). "Tsetse-fly (*Glossina morsitans*) in the Koalib Hills, Nuba Mountains Province, Sudan." Bulletin of entomological research **19**(03): 309-316.

Sabeti, P., et al. (2006). "Positive natural selection in the human lineage." Science **312**(5780): 1614-1620.

Sabeti, P. C., et al. (2002). "Detecting recent positive selection in the human genome from haplotype structure." Nature **419**(6909): 832-837.

Sabeti, P. C., et al. (2007). "Genome-wide detection and characterization of positive selection in human populations." Nature **449**(7164): 913-918.

Sanger, F. and A. R. Coulson (1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase." Journal of molecular biology **94**(3): 441-448.

Sanger, F., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences **74**(12): 5463-5467.

Schenkel, F., et al. (2006). "Association of a single nucleotide polymorphism in the calpastatin gene with carcass and meat quality traits of beef cattle." Journal of Animal Science **84**(2): 291-299.

Schnekenburger, J., et al. (2008). "The calcium binding protein S100A9 is essential for pancreatic leukocyte infiltration and induces disruption of cell-cell contacts." Journal of cellular physiology **216**(2): 558-567.

Scholtz, M. and A. Theunissen (2010). "The use of indigenous cattle in terminal cross-breeding to improve beef cattle production in Sub-Saharan Africa." Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales **46**: 33-39.

Scriver, C. R. and P. J. Waters (1999). "Monogenic traits are not simple: lessons from phenylketonuria." TRENDS in Genetics **15**(7): 267-272.

Sequencing, T. C. and A. Consortium (2005). "Initial sequence of the chimpanzee genome and comparison with the human genome." Nature **437**(7055): 69-87.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nature biotechnology **26**(10): 1135-1145.

Sirko-Osadsa, D. A., et al. (1998). "Stickler syndrome without eye involvement is caused by mutations in COL11A2, the gene encoding the α 2(XI) chain of type XI collagen." The Journal of pediatrics **132**(2): 368-371.

Son, Y. (1999). "Production and uses of Korean native Black goat." Small Ruminant Research **34**(3): 303-308.

Steinfeld, H., et al. (2006). Livestock's long shadow: environmental issues and options, Food & Agriculture Org.

Stock, F. and D. Gifford-Gonzalez (2013). "Genetics and African cattle domestication." African Archaeological Review **30**(1): 51-72.

Sutter, N. B., et al. (2004). "Extensive and breed-specific linkage disequilibrium in *Canis familiaris*." Genome research **14**(12): 2388-2396.

Swallow, B. M. (2000). Impacts of trypanosomiasis on African agriculture, Food and Agriculture Organization (FAO).

Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-595.

Thomas, A. (1943). "The vegetation of the Karamoja district, Uganda: An illustration of biological factors in tropical ecology." The Journal of Ecology: 149-177.

Troy, C. S., et al. (2001). "Genetic evidence for Near-Eastern origins of European cattle." Nature **410**(6832): 1088-1091.

Vikkula, M., et al. (1995). "Autosomal dominant and recessive osteochondrodysplasias associated with the *COL11A2* locus." Cell **80**(3): 431-437.

Voight, B. F., et al. (2006). "A map of recent positive selection in the human genome." PLoS biology **4**(3): e72.

Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids." Nature **171**(4356): 737-738.

Watterson, G. and H. Guess (1977). "Is the most frequent allele the oldest?" Theoretical population biology **11**(2): 141-160.

Wendorf, F. and R. Schild (1994). "Are the early Holocene cattle in the Eastern Sahara domestic or wild?" Evolutionary Anthropology: Issues, News, and Reviews **3**(4): 118-128.

Yang, J., et al. (2011). "GCTA: a tool for genome-wide complex trait analysis." The American Journal of Human Genetics **88**(1): 76-82.

Yang, Z. and J. P. Bielawski (2000). "Statistical methods for detecting molecular adaptation." Trends in Ecology & Evolution **15**(12): 496-503.

Yutani, K., et al. (1977). "Effect of a single amino acid substitution on stability of conformation of a protein."

Zajc, I., et al. (1997). "Variability of canine microsatellites within and between different dog breeds." Mammalian Genome **8**(3): 182-185.

Zeder, M. A. and B. Hesse (2000). "The initial domestication of goats (*Capra hircus*) in the Zagros Mountains 10,000 years ago." Science **287**(5461): 2254-2257.

Zimin, A. V., et al. (2009). "A whole-genome assembly of the domestic cow, *Bos taurus*." Genome Biol **10**(4): R42.

(요약) 국문초록

집단 간의 유전체 비교를 위한 차세대 시퀀싱 데이터를 이용한
생물정보학적 연구 및 미래 유전자원에 대한 리뷰

권 영준

협동과정 생물정보학 전공

서울대학교 대학원 자연과학대학

유전체는 생물체에서 설계도와 같으며, 따라서 유전체 연구를 통하여 생명현상을 설명하기 위한 많은 연구들이 시도되어 왔다. 최근 차세대 염기서열 기술의 발달과 염기서열 해독 가격의 인하로 인하여 집단수준에서의 전장 유전체 비교 연구가 개인 연구실에서도 가능한 수준에 이르렀다. 개체 별로 수집된 변이정보로부터 각각의 집단의 유전적 특징이 설명되고 비교 될 수 있고, 나아가 집단 사이의 표현형적인 차이 또한 DNA 서열 수준에서 설명 될 수 있다. 다양한 특징들을 가지는 가축들은 오랜 시간 동안 가축화되어 오면서, 그 자취들이 유전체 서열 안에 남아 있게 되었다. 따라서 이러한 가축 동물들은 많은 연구자들로부터 좋은 연구 모델동물로서 여겨져 왔다.

챕터 1에서는 DNA 서열 해독 역사와 차세대 염기서열 기

술에 대하여 간단하게 요약하여 기술하였다. 또한 양성 선택 유전자 지역 발굴 방법과 같이 이 연구에서 사용된 집단간 유전체 비교 분석 기법들에 대하여서도 간략히 요약하였다.

챕터 2에서는 한국 재래 염소 15마리, 교잡종 염소 11마리 각각 개체에 해당하는 차세대 염기서열 해독방법(Illumina)을 이용한 전장 유전체 서열 데이터를 이용하여 두 개의 염소집단에 대해 유전체를 비교하고 집단간 뚜렷이 구분되는 유전적 특징 등을 파악하기 위한 연구를 진행하였다. 염소는 소목 소과의 포유류로서 소와는 가장 가까운 가축이므로, 챕터3에서 다루게 될 아프리카 소 프로젝트에 대한 선행 연구 모델로서 선택하여 연구를 진행하였다.

개체 별로 유전자형을 조사하기 위하여 GATK를 이용한 SNP 분석을 실시하였고, 이 데이터를 바탕으로 연구를 진행하였다. 두 집단에서 확인된 유전 변이들로부터 재래염소 집단에서 유전적 다양성이 확연하게 감소 되어있는 것을 확인하였다. 또한 두 집단이 유전적으로 분명하게 구분되지만, 공통조상으로부터 물려받은 대립형질들이 여전히 양쪽 집단에서 우점 되어 있는 것을 STRUCTURE 분석을 통하여 확인 하였다. 이 결과들은 기존에 알려진 두 집단에 대한 사전정보를 뒷받침 해주는 결과이다.

양성 선택 받은 유전자 지역을 탐지하는 방법 중 하나인 XP-EHH 통계치를 통하여 6개의 후보 유전자(예상되는 유전자 포

함)에서 강한 양성 선택 신호가 발견되었다. 또한 GO term 분석을 수행하였으며, 이 연구에서는 DAVID tool이 GO term 분석을 위하여 이용되었다. 변이가 많이 밀집되어 있는 염색체 구간에 존재하는 유전자들로 GO term 분석을 수행하였을 때, 후각과 신경에 관련된 유전자들이 상대적으로 변이를 획득할 기회를 가지는 것으로 보였으나, 실제로 Non-synonymous SNP을 가지는 유전자들로 GO-term 분석을 실시하였을 때는 후각 및 면역 관련 유전자에서 한국 재래 염소와 교잡종 염소 집단 사이의 차이를 보이는 것으로 확인되었다.

챕터 3에서는 미래 유전자원으로서의 아프리카 재래 소에 대하여 요약 및 기술하였다. 아프리카 재래 소들은 다양하고 혹독한 기후 조건하에 적응 되어 온 결과로 매우 다양한 특징들을 보여준다. 아프리카 재래 소들의 표현형 적인 특징 혹은 유전적인 특징들을 조사하였고, 미래 유전자원으로서의 아프리카 소 연구의 필요성에 관하여 서술하였다. 또한 기존에 행해졌던 아프리카 재래 소 연구에 대하여 간략히 소개하였으며 앞으로 예상되는 아프리카 소 연구에 대하여도 간략히 기술하였다.

주요어: 차세대 염기서열 해독, 유전적 비교, 집단 분석, 유전 자원, 아프리카 소.

학번: 2012-2311