European Master in Animal Breeding and Genetics

Runs of Homozygosity patterns in Taurine and Indicine cattle breeds

Student: Zahra Karimi

MAJOR THESIS ANIMAL BREEDING AND GENETICS

July 2013

Supervisor: *Dip.Ing.Prof. Johann Sölkner*

Co-Supervisors*:*

Msc. Maja Ferenčaković, Msc. Ana María Pérez O'Brien

BOKU

Education and Culture

Erasmus Mundus

ABSTRACT

Runs of homozygosity (ROH), which are contiguous homozygous fragments of DNA sequence without heterozygosity in the diploid state, are caused by common ancestors. Long ROH segments are supposed to be autozygous stretches originating from a recent common ancestor, while shorter ROH segments originate from distant ancestors. The main aim of this study was to investigate the existence and comparison of ROH islands across the genomes of Taurine (Angus, Fleckvieh, Brown Swiss) and Indicine (Nelore, Gir, Brahman) breeds. The term ROH islands is used to describe regions of the genome where many individuals of a population share ROH. We also compared ROH estimation using different software tools, e.g. cgaTOH, PLINK and SVS.

We confirmed that small and long segments of ROH are present and abundant in the bovine genome and in some regions they are frequently observable even across unrelated individuals. Their size differed from 1 Megabase (Mb) to long segments spanning multiple Mb. We have also found some specific locations in the genomes in which ROH fragments were significantly conserved (ROH Islands/Hot spots) within and between the Taurine and Indicine cattle breeds. In addition, ROH Islands were observed to be in gene-rich portions of the genome. Even though the density of ROH patterns was always significantly different among the 3 software tools, there was a consensus as to the location of ROH islands discovered. Unique ROH patterns were found for each breed and between breeds, which was consistent with possible signatures of either artificial or natural selection. Since these patterns are an indication of common ancestry, either in the recent or remote past, these results could be beneficial for evolutionary studies as well as for fine mapping gene discovery studies. There is the possibility to use these ROH islands as surrogate markers in case-control experiments to see whether there is any relationship between them and recessive diseases or not. The same can be done for studying economic traits.

Keywords:

Runs of homozygosity; ROH islands, Autozygosity, Taurine, Indicine, cattle

Table of Contents

Acknowledgments

First of all I would like to acknowledge and recognize the EM-ABG program for the financial support without which I could never accomplish this Master of Sciences degree.

I would like to express my gratitude to my supervisor Prof. Johann Sölkner, whose expertise, understanding, and patience added considerably to my experience and for the assistance he provided at all levels of this research project. It was my pleasure to be supervised by him. A very special thanks goes out to my co-supervisors, Maja Ferenčaković and Ana María Pérez O'Brien for their scientific support and especially for their technical direction. I feel extremely grateful I had the opportunity to study at BOKU University and extend my knowledge and experience in both scientific and social aspects in such a good academic environment!

1. Introduction

Inbreeding, the mating of closely genetically related individuals (Wright, 1922), is an important parameter in population genetic studies, being an influencing factor in reducing the genetic variability of populations and causing decrease in production and fitness levels of individuals (Saccheri et al., 1996). These severe negative effects of inbreeding, known as "inbreeding depression", can be seen naturally in small populations were the mating of relatives is unavoidable (Pirchner, 1985).

The recent availability of high-density SNP (Single Nucleotide Polymorphism) data in humans has led to the discovery of continuous regions of autozygosity in individual genomes (Broman & Weber, 1999). Autozygosity occurs when an individual inherits chromosomal fragments that are identical-by-descent (IBD) from both paternal and maternal sides (Wright, 1922). Individual autozygosity can be measured using Runs of Homozygosity (ROH) (Gibson et al., 2006; McQuillan et al., 2008; Keller et al.,2011; Ferenčaković et al., 2013), which are contiguous homozygous fragments of DNA sequence without heterozygosity in the diploid state. Broman & Weber (1999) stated that ROH are caused by common ancestors, and due to recombination events interrupting long chromosome segments over time, extensive ROH segments are supposed to be autozygous stretches originated from recent common ancestors. Shorter ROH segments originate from distant ancestors or may involve some non-IBD stretches (Howrigan et al., 2011). Consequently, an inbred population is expected to display longer homozygous segments as compared to a crossbred population (Gibson et al., 2006).

The first studies on ROH regions were done by Gibson et al. (2006) in humans and by Sölkner et al. (2010) in cattle. Later studies (Bjelland et al., 2013; Ferenčaković et al., 2011; Ferenčaković et al., 2013) have found that level of inbreeding (F) estimated from ROH (F_{ROH}) has several advantages as compared to pedigree estimates (F_{PED}). F_{ROH} can predict the actual percentage of the genome that is autozygous more precisely as compared to F_{PED}; 2) F_{ROH} can capture autozygosity arising from very distant common ancestors (e.g., 50+ generations ago); 3) F_{ROH} can be estimated in any genotyped individual, even though pedigree information is not available; 4) F_{ROH} offer the possibility to examine the distribution of autozygosity across the genome and to find specific locations in the genome with higher levels of autozygosity (e.g., by estimating F_{ROH} separately for different chromosome; Keller et al., 2011; Ferenčakovićet al., 2013).

Nothnagel et al. (2010) has reported regions of "ROH islands" or "ROH hot spots" in humans that are present in more than 50% of the individuals studied, which could theoretically be an indication of strong

past selection in these populations. Additionally, Pemberton et al. (2012) have revealed that ROH patterns are influenced not only by the locations of recessive-disease loci, but also by population history and mating systems. The study of ROH in different populations can provide insight into how population history, genomic properties, and cultural habits could affect the observed islands of ROH in the human genome (Pemberton et al., 2012). In this sense genome-wide autozygosity can be used for recognition of recessive disease variants using homozygosity mapping, as well as for investigating the effects of genome-wide homozygosity on important traits (McQuillan et al., 2008; Keller et al., 2012).

The availability of a bovine high density SNP chip provides the opportunity to investigate the bovine genome for ROH. The current classification of cattle breeds includes three major genetic subdivisions: Taurine, Indicine and Sanga. The Indicine and Sanga groups are both adapted and bred for tropical conditions, while the Taurine group has classically conformed to moderate climate (Chan et al., 2010). Taurine cattle breeds comprise those originated from southwest Asia, with short ears and no hump, while Indicine breeds represent those descendent from South Asian ancestors, typically with long floppy ears and a prominent hump. Indicine animals were introduced to Africa by the Arab traders more than a thousand years ago and mixed with local Taurine cattle to establish the Sanga. Therefore, the geographic influence of this cattle type is referred to East Africa (Chan et al., 2010).

The main aim of this study was to investigate the existence and comparison of ROH islands across the genomes of Taurine and Indicine breeds. The specific objectives included identifying regional patterns with high prevalence of ROH, comparing genome-wide patterns of runs across the different breeds to see whether patterns of ROH are different within and between cattle types, discover genes and QTLs in high prevalence regions and finally compare ROH estimates using different ROH discovery software tools.

2. Materials and methods

Genotypes

Three Taurine breeds, Angus, Brown Swiss and Fleckvieh and three indicine breeds, Brahman, Gir and Nelore were included in the analysis. Illumina BovineHD BeadChip (777K SNPs) (Illumina, 2012) genotypes were used on this study. The genotypes were provided by Zebu Genomic Consortium - Brazil for Nelore, Embrapa - Brazil for Gir, Zuchtdata GmbH - Austria for Fleckvieh and Brown Swiss, and by AGBU (Animal Genetics and Breeding Unit) University of New England - Australia for Angus and Brahman.

Quality Control

Quality control was conducted using SAS software v9.3 (SAS Institute Inc., 2009) for each of the six breeds separately. SNP markers with unknown chromosomes, unknown base pair positions, Y\X chromosomes and mtDNA were taken out. Markers with GenCall score lower or equal to 0.7, as well as those with GenTrain score lower or equal to 0.4 were removed. In addition, removal of SNP markers which were missing at least in 10% of the animals and all animals with more than 5% of missing genotypes was performed. The final number of bulls and the total number of SNPs kept for data analysis is shown in Table 2.1.

Table 2.1: Total number of individuals and SNPs per cattle breed

Breed	Cattle type	Number of individuals	Final Number of SNPs
Fleckvieh	Taurine	97	594632
Brown Swiss	Taurine	46	613990
Angus	Taurine	108	675719
Nelore	Indicine	134	681557
Gir	Indicine	101	677535
Brahman	Indicine	101	675792

Detection of homozygous segments

ROH were detected using three software tools, the Golden Helix SNP & Variation Suite v.7.6.8 (SVS) (Golden Helix Inc., 2013), cgaTOH (Zhang et al., 2013) and PLINK v.1.0.7 (Purcell et al., 2007). The following parameter settings were taken into account and applied for all software:

- Minimum number of SNPs needed to define a segment as a ROH: 30 SNPs
- Number of missing calls allowed inside a ROH segment: 4
- Number of heterozygous calls allowed inside a ROH: 1
- Maximum gap between consecutive homozygous SNPs: 250kb
- Minimum allowed density of SNPs inside a run: 1 SNP for every 50kb
- Minimum length to define a ROH: 1000kb

Afterwards the SNP results on the incidence of ROH patterns for each of the breeds and for grouped results (Taurine, Indicine, and all animals together) were graphed using SVS (Golden Helix Inc., 2013), SAS (SAS Institute Inc., 2009) and R v.2.15.2 (CRAN project, 2013).

Comparison of software tools

Different studies have used different methodologies for prediction of continuous homozygosity in high-density genotyping datasets (Zhang et al., 2013). PLINK (Purcell et al., 2007) and SVS (Golden Helix Inc., 2013) are the most common software tools which have been used recently to identify ROH patterns. The ROH analysis of SVS is programmed to find runs of homozygous SNPs starting at every possible marker and can also recognize and report the position of those runs shared among a user-determined number of samples. PLINK describes the ROH based on the required number of homozygous SNPs extending a specific distance by a sliding-window approach, moving a window of a user-determined number of SNPs along the genome, and deciding if each window complies with the required parameters or not. Zhang et al. (2013) recently presented a method called cgaTOH. This software uses a SNP-wise approach for determining homozygosity runs, which they termed as TOH for Tracts of Homozygosity, and has additional features on the classification of the segments such as Allele matching. In this study ROH analyses were performed using all three software tools. A dependent pair t-test was performed using SAS software v9.3 (SAS Institute Inc., 2009) on the results of the three software to determine whether the frequency of observed ROH patterns reported by each software was significantly different from each other or not.

Comparison of patterns across breeds

Only the results from SVS were used for the breed and group comparison analyses. Patterns of ROH were graphed using SVS for each breed and group, as well as each chromosome separately. Thereafter, patterns of high regional homozygosity common to most animals were found through visual assessment, by comparing the patterns within and between cattle types. A chi-square pair comparison, applying Bonferroni correction for multiple testing, was done using SAS software v9.3 (SAS Institute Inc., 2009) as a homogeneity test to check whether density of ROH islands was significantly different for breeds.

Gene finding for regions of high ROH frequency

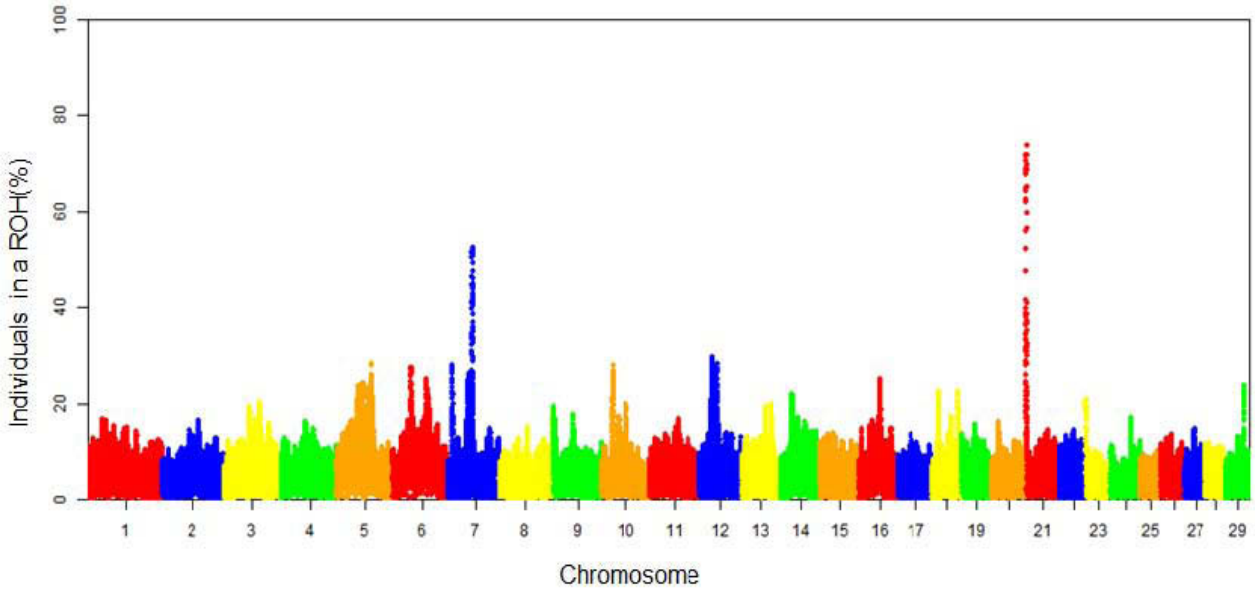
For the regions with high frequency of ROH patterns, the genes located within the pattern boundaries and the orthologs in human were found using the Ensemble Genome Biomart tool (WTSI/EBI, 2013), while, the gene functions were searched using the quickGO (EMBL-EBI, 2013).

3. Results

Evidence of ROH patterns

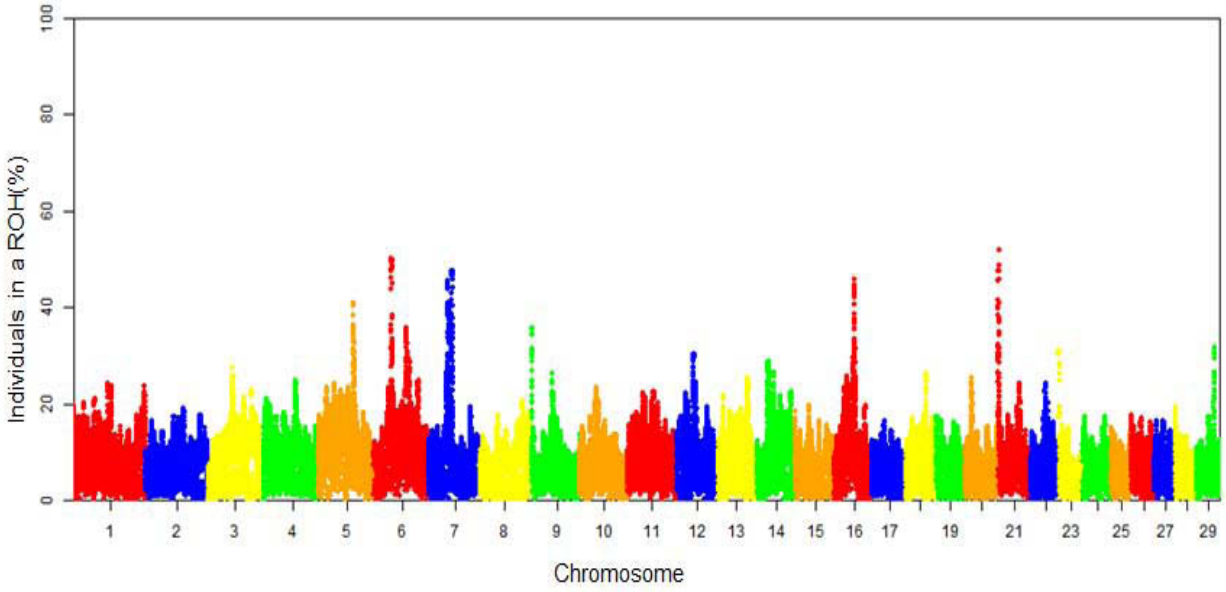
Figure 3.1 shows the incidence of ROH segments among all breeds across the genome. The genomic distribution of ROH segments was clearly non-uniform across chromosomes. ROH patterns were indeed numerous in specific locations across the genome. A total of 7 regions were identified with frequencies of ROH segments exceeding 25% of the whole population. It is clear that these peaks in frequency of ROHs were a sign of common patterns of ROH, i.e. ROH islands, shared by all breeds. Two peaks were particularly strong. For the highest peak, 77% of individuals shared SNP positions in ROH at the start of the chromosome 21. The second strong pattern was observable on chromosome 7 in which 51% of individuals were involved.

Figure 3.1: Distribution of ROH segments in all individuals genomes(Result from Golden Helix SVS)
X axis: The distribution of SNPs in a ROH across the genome
Y axis: The percentage of individuals with SNPs included in a ROH



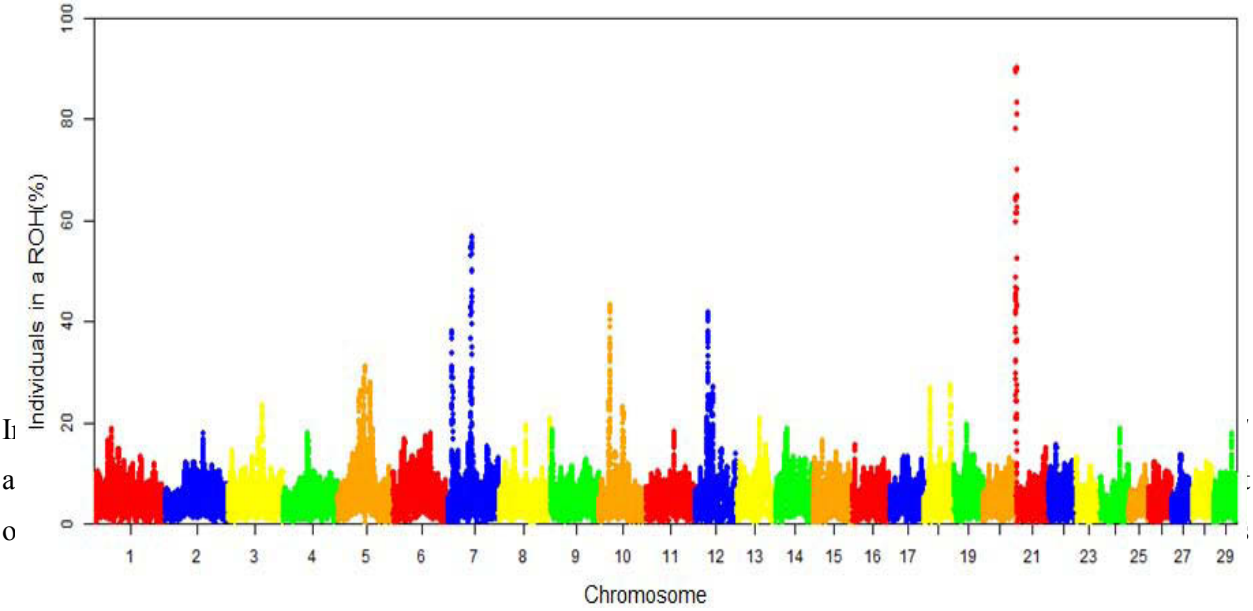
ROH analysis for each of Indicine and Taurine breeds was performed separately in order to differentiate ROH islands between them. Figure 3.2 shows the incidence of ROH patterns within the Taurine cattle breeds, e.g. Fleckvieh, Brown Swiss and Angus, where 5 ROH-rich regions were found across the genome. Notably, 4 ROH islands have been discovered in this subspecies of cattle to be present in more than 45% of individuals (Table 3.1). The first and second most common ROH islands in Taurine breeds were on chromosome 21, being present in 57% of individuals and on chromosome 7, with a frequency of 47%.

Figure 3.2: Distribution of ROH segments in Taurine cattle breeds (Result from Golden Helix SVS)
X axis: The distribution of SNPs in a ROH across the genome
Y axis: The percentage of individuals with SNPs included in a ROH



For the Indicine cattle breeds, i.e. Brahman, Gir and Nelore, 4 ROH islands were evident across the genome, exceeding 40% of individuals (Figure 3.3). The most common pattern was located on chromosome 21 exceeding 92% of individuals and the second strong pattern was on chromosome 7 present in 60 % of individuals, both located in the same regions as in the Taurine breeds, described above.

Figure 3.3: Distribution of ROH segments in Indicine cattle breeds (Result from Golden Helix SVS)
X axis: The distribution of SNPs in a ROH across the genome.
Y axis: The percentage of individuals with SNPs included in a ROH.



I
a
o

Nothnagel et al. (2010) revealed that even though there was a significant relationship between increased regional Linkage Disequilibrium (LD) and incidence of ROH islands, their existence was not completely explicable by high LD alone. In our study we did not perform local linkage disequilibrium analysis, but, by considering the minimum size of ROH segments to be 1MB, we tried to avoid small autozygous segments caused by LD. In ROH studies with the human genome, it is common to consider minimum size of ROH segments of 500kb. We, however, increased the size of segment into 1Mb since it is known that the extent of LD in cattle genomes is usually higher than that of the human genome (Ferenčaković et al., 2013).

Table 3.1: Comparison of "ROH islands" density between Taurine and Indicine cattle breeds.
Regions with high ROH frequency per SNP (ROH islands) in Percentages (%)

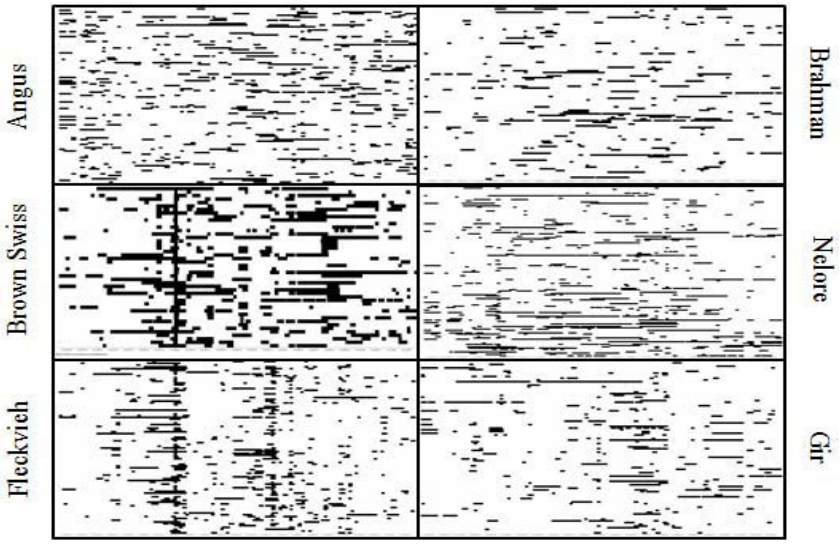
ROH island location	Physical Position(bp)	Taurine	Indicine
BTA6	38,268,200 : 39,451,000	54	-
BTA7	51,502,500 : 52,353,000	47	58
BTA10	24,575,700 : 25,619,800	-	45
BTA12	28,434,000 : 29,628,100	-	42
BTA16	43,802,200 : 44,968,700	44	-
BTA21	1,360,390 : 1,853,150	53	93

Comparison for the distribution of 'ROH islands' across the cattle breeds

We visually studied the results obtained from ROH analyses per chromosome separately to identify the possible ROH islands across the cattle breeds. Graphical representation of individual ROH patterns is given in the Appendix for each chromosome and breed. A total of 6 most conserved common regions were selected across the genome for investigation. These regions were located on chromosomes 6, 7, 10, 12, 16 and 21. The incidence of ROH islands was investigated statistically across different breeds using a chi-squared paired-comparison homogeneity test applying Bonferroni correction to see whether the density of ROH islands is statistically different between different breeds (Table 3.2). Figure 3.4 shows the distribution of ROH segments on chromosome 6 across all breeds. Even though segments of ROH across the chromosome 6 varied among breeds, there was a specific highly conserved region (38,268,200:39,451,000), e.g. ROH island, which was common between Brown Swiss and Fleckvieh. Although the results from chi-square pair comparison applying Bonferroni correction for the incidence of ROH islands across the different breeds indicate that Brown Swiss had significantly greater proportion of

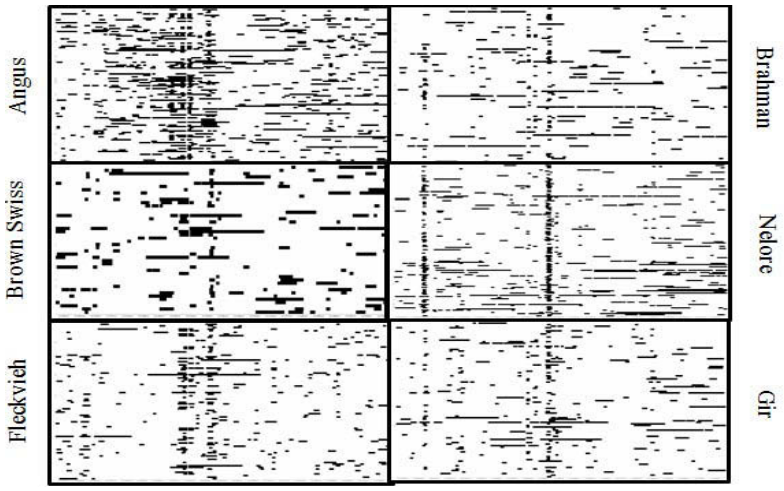
animals in ROH islands than Fleckvieh. The pattern suggests that a large fraction of individuals within Brown Swiss and Fleckvieh breeds may have some distant common ancestor.

Figure 3.4: Distribution of ROH frequency across chromosome 6 for all cattle breeds
Each row represents an individual and each column represents a genotyped SNP position. Black color indicates a SNP included in a ROH, and white color indicates those SNPs not included in ROH. Intensity of black points increases with increasing ROH frequency



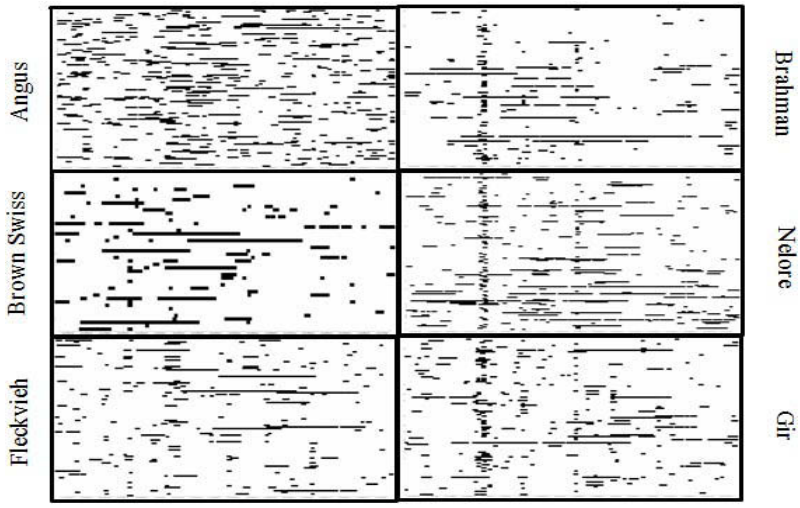
ROH segments were particularly abundant at a specific region (51,502,500:52,353,000) located on chromosome 7 (Figure 3.5). This region was virtually ubiquitous in all cattle breeds studied, however, the results from the chi-square paired-comparison homogeneity test for the incidence of ROH islands across the different breeds, shows that Nelore had the highest proportion of animals for this ROH island. This was followed Brown Swiss, Angus and Gir, having the same density for this region. Brahman and Fleckvieh had significantly lower number of animals sharing this particular ROH islands compared to the above mentioned breeds.

Figure 3.5: Distribution of ROH frequency across chromosome 7 for all cattle breeds
Each row represents an individual and each column represents a genotyped SNP position Black color indicates a SNP included in a ROH, and white color indicates those SNPs not included in ROH. Intensity of black points increases with increasing ROH frequency



We also recognized "ROH islands" on chromosome 10 which was identical only in the Indicine breeds (Figure 3.6).

Figure 3.6: Distribution of ROH frequency across chromosome 10 for all cattle breeds
Each row represents an individual and each column represents a genotyped SNP position. Black color indicates a SNP included in a ROH, and white color indicates those SNPs not included in ROH. Intensity of black points increases with increasing ROH frequency.



There was a ROH island at the beginning of chromosome 21 common across all breeds (Figure 3.7). A careful look at the figure 3.7 shows that the density of these ROH islands was not similar between breeds. Statistical test shows that density of ROH islands was higher in the Indicine cattle type than the Taurine (Table 3.2).

Figure 3.7: Distribution of ROH frequency across chromosome 21 for all cattle breeds
Each row represents an individual and, each column represents a genotyped SNP position. Black color indicates a SNP included in a ROH. Intensity of black points increase with increasing ROH frequency

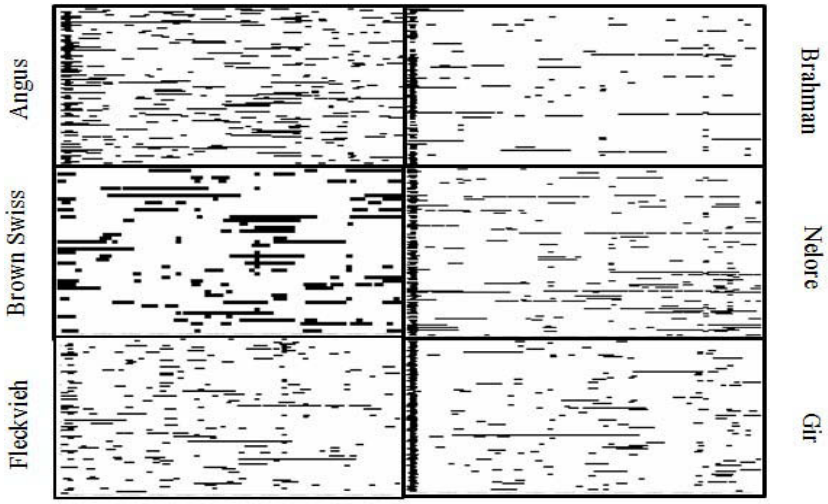


Table 3.2: Chi-square paired-comparison homogeneity test with Bonferroni correction across cattle breeds for ROH islands density (values are in Percentages %)
Each capital letter signifies whether breeds are significantly different from each other in terms of density of observed ROH patterns or not

Breed	Physical Position	Angus	Brown Swiss	Fleckvieh	Brahman	Nelore	Gir
BTA6	38,268,200 : 39,451,000	19^{c}	100^{a}	73^{b}	7^{c}	11^{c}	17^{c}
BTA7	51,502,500 : 52,353,000	51^{b}	56^{b}	38^{c}	39^{c}	78^{a}	46^{b}
BTA10	24,575,700 : 25,619,800	11^{b}	10^{b}	2^{b}	42^{a}	42^{a}	46^{a}
BTA12	28,434,000 : 29,628,100	19^{b}	9^{b}	6^{b}	5^{b}	62^{a}	46^{a}
BTA16	43,802,200 : 44,968,700	71^{a}	39^{b}	18^{bc}	10^{c}	8^{c}	9^{c}
BTA21	1,360,390 : 1,853,150	73^{c}	34^{d}	38^{d}	84^{bc}	87^{b}	100^{a}

Gene discovery inside ROH islands

Table 3.3 includes all genes inside ROH islands selected for this study. In general, these genes had basic functions. For example, the genes within ROH islands on chromosome 16 (43,802,200 : 44,968,700), which was highly conserved in Angus, but, slightly conserved in Brown Swiss, in total 20 genes were found within this region (Table 3.3) and out of these, ten were found to have binding functions. Additionally some of the genes found have been associated to economical traits and recognized as

Quantitative Trait Loci (QTLs) important for beef production (Gutierrez-Gil et al, 2008; McClure MC et al. 2010).

Table 3.3: Conserved genes inside ROH islands

The Top 6 ROH islands on Bovine Autosomes		
CHR	**Location/Size(kb)**	**Genes**
6	38,268.2:39,451	MEPE, IBSP, LAP3, BT.29898, FAM184B, BT.100379, LCORL, BT.94996
7	51,502.5:52,353	HSPA9, BT.63787, LRRTM2, SIL1, GPX4, BT.71626, PAIP2, SLC23A1, PACAP,SPATA24, DNAJC18, ECSCR, C5ORF65, 5S_rRNA, SNORA74, SNORA74
10	24,775.7: 25,007.1	TRAV14DV4, BT.64165, BT.101619
12	28,434:29,628.1	PDS5B, N4BP2L2, N4BP2L1, BRCA2, ZAR1L, FRY, RXFP2
16	43,802.2:44,968.7	BT.104317, BT.103198, DFFA,CORT,APITD1,PGD,KIF1B,UBE4B RBP7, NMNAT1, CTNNBIP1
21	1,360.39:1,853.15	OR5D13, U6, U6, 5S_rRNA, 5S_rRNA

Notably, the ROH island located on chromosome 21 was the only region in which genes did not have any association with QTLs. Intriguingly, this region was the only area which we found some possible orthologs in the human genome at ROH islands regions on human chromosomes 3 and 11 (Table 3.4).

Table 3.4: ortholog genes of human genome with bovine genes in ROH islands on chr21

Ensemble ID	gene name	Function	Ortholog gene name	Ortholog physical position In ROH islands	Description
ENSBTAG00000023079	-	-	PPFIA1	11(70,116,806:70,230,509)	Liprin-alpha-1
ENSBTAG00000013567	OR5D13	Signal transducer activity	OR5D13	11(55,540,914- 55,541,858)	Olfactory receptor 5D13
ENSBTAG00000035268	5S_rRNA	-	RNA5SP132	3(51,728,481-51,728,598)	RNA, 5S ribosomal pseudogene
ENSBTAG00000035268	5S_rRNA	-	RNA5SP131	3: (50,457,212-50,457,330)	RNA, 5S ribosomal pseudogene

In general there were more genes inside the ROH islands, as compared to surrounding area (Table 3.5). Using dependent group t-test, total number of genes inside ROH islands were

significantly more than in regions of identical size chosen for comparison that were 10 or 20 Mb apart (P <0.0001).

Table 3.5: Total number of genes inside ROH islands and neighborhoods

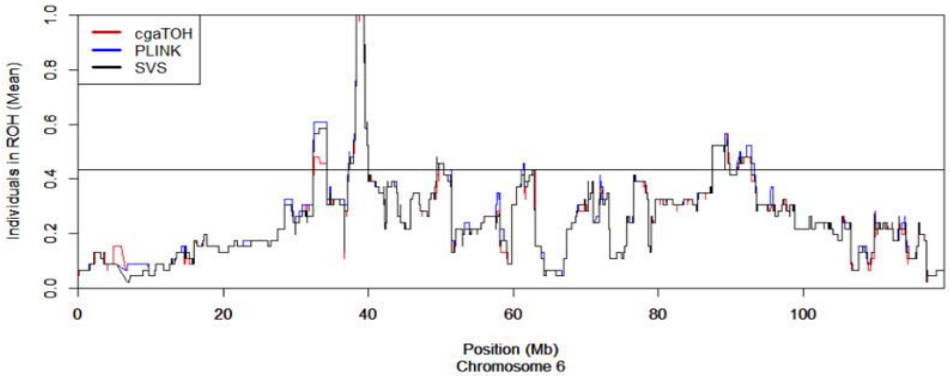
From left to the right: Column 1: ROH islands location; Column 2: Gene number within ROH islands; Column 3: Gene numbers in regions of identical size 10Mb distant from ROH islands; Column 4: Gene numbers in regions of identical size 20Mb distant from ROH islands

CHR	ROH islands	10Mb distance	20Mb distance
16(43,802.2:44,968.7)	20	0	5
6(38,268.2:39,451)	10	2	5
7(51,502.5:52,353)	17	5	1
10(24,775.7: 25,007.1)	4	0	3
12(28,434:29,628.1)	7	0	1
21(1,360.39:1,853.15)	8	0	0

Comparing the result of ROH patterns density across different software tools

The results of ROH incidence from the 3 software tools were compared. The ROH islands discovered by the 3 software tools were always in similar locations (Figure 3.8) with slight difference in the frequency of the signals:

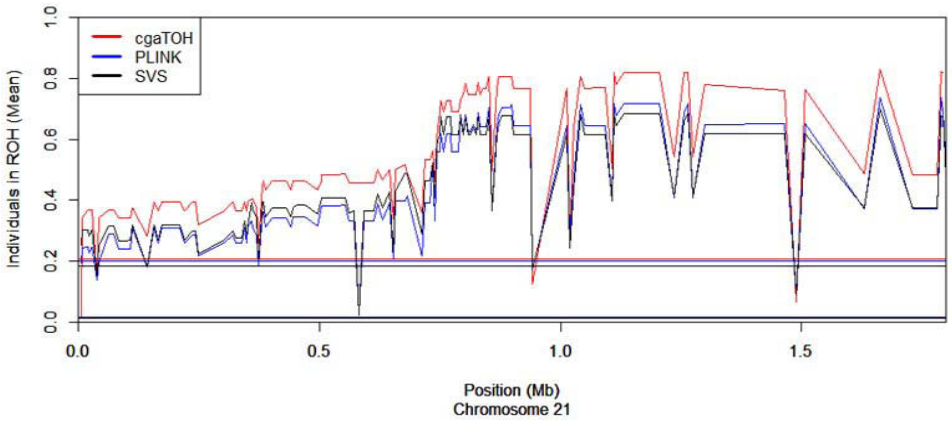
Figure 3.8. Comparing results of 3 software tools across chromosome 6



Apparently, cgaTOH used to capture a higher amount of individuals in ROH patterns than the other software tools. The results from SVS and PLINK were more overlapping (Figur3.9).

a

Figure 3.9: Comparing the results of 3 software tools at conserved area on chr 21



Dependent pair t-test was done across the whole genomes to see whether the density of ROH patterns by different tools were significantly different from each other or not. The results of ROH incidence from different software tools were significantly different from each other ($p < 0.001$) across the genome. In order to make a comparison for smaller segments across the genome 3 conserved and 3. Un-conserved areas were tested for density of ROH patterns. Notably, the density of ROH patterns was always significantly different even for small regions of the genome.

Even though the software tools were found to be significantly different from each other, the regions of ROH islands discovered by the three were always in similar locations with slight differences in the frequency of the signals. This could only make a difference in the grouping of the signals for breed comparisons. We would like to point out that frequency of ROH patterns observed across the genome depends on the program used to calculate them, because different software tools for analyzing ROH patterns do not give identical results.

4. Discussion

We confirmed that small and long segments of ROH are present and abundant in the bovine genome and in some regions they are frequently observable even across unrelated individuals. Their size differs from 1 to multiple Mb, and ROH segments in different individuals have different distributions. Several causes may influence the frequency, size and region of ROH segments. These include natural selection, artificial selection, recombination events, LD patterns, population structure, mutation rate and parental relatedness. We have also found some specific locations in which ROH fragments were significantly ubiquitous and common within and between the Taurine and Indicine cattle breeds, known as ROH islands/hot spots.

The reasons for the existence of ROH fragments in the human genome have been discussed (e.g., McQuillan et al., 2008; Li et al., 2006; Pemberton et al., 2012; Gibson et al., 2006). ROH segments of different size may have different backgrounds. Several studies have stated that small homozygous fragments of ROH patterns have the characteristics of short haplotype block which are abundant in regions of the genome that are highly associated with high linkage disequilibrium regions and low rate of recombination (Gibson et al., 2006). Pemberton et al. (2012) stated that short segments of ROH (<100kb) represent homozygosity originating from distant haplotypes, while, medium segment, e.g. hundreds of kb to several MB, probably result from distant inbreeding causing by small population size; and there is a negative relationship between short/medium segments of ROH and haplotype diversity. Moreover, the similar location of LD and ROH patterns across the populations could suggest a narrow correspondence between historical ROH segments and current recombination patterns through different populations. For long segments of ROH, however, different reasons have been reported.

Long ROH segments exceeding 1 Mb, is not concentrated in regions of high LD as compared to smaller segments (Gibson et al., 2006; Pemberton et al., 2012). One potential reason is that there may be a relation among segments, coming from distant ancestors, which still remained intact because of being located in low recombination regions. The second option is that these segments are a sign of more recent inbreeding and thus they have not experienced many recombination events to be broken down into smaller fragments. Therefore long segments of haplotypes have been speculated to be an evidence of recent positive selection throughout populations (Gibson et al., 2006; Pemberton et al., 2012; Nothnagel et al., 2010). Besides, some studies showed that there is a positive relationship between long segments of ROH and consanguineous unions (Pemberton et al., 2012; McQuillan et al., 2008). Furthermore, it was confirmed that long segments of ROH have significant association with autosomal-dominant diseases than in regions not implicated in recessive disorders (Pemberton et al. 2012). ROH islands, thus are likely found in locations with low recombination and locations of selective sweeps (Pemberton et al., 2012; Nothnagel et al., 2010). Nothnagel et al. (2010) have reported that gene diversity and extent of high LD

explain only 7% and 20% of occurrence of ROH patterns respectively. These reasons, therefore, are neither adequate nor fundamental for presence of ROH islands. Other potential reasons like decreased SNP variability as well as lack of heterozygosity are not either comprehensive enough to explain highly structured of ROH segments across different subpopulations in human genomes. Some studies have revealed that cumulative ROH segments are more often observed in isolated populations (Wang et al 2006, McQuillan et al., 2010; Nothnagel et al., 2010); and there is a significant negative relationship between effective population size and formation of cumulative ROH segments (Nothnagel et al., 2010). ROH islands could also be target locations for positive selection, which has led to increase of homozygosity around target loci (Pemberton et al., 2012). Nothnagel et al. (2010) observed many of ROH islands common throughout European subpopulations, and recognized that systematic distribution of ROH patterns is significantly ubiquitous in geographical widths. Pemberton et al. (2012) discovered many of ROH hot spots across continental regions and found that many of them had before been recognized as goals of resent positive selection. It suggests that positive natural selection is a potential reason for formation of ROH islands. In addition, natural selection will help to shape such strong patterns over a broad span of historical evolution.

We have found some ROH islands where located within the genome that varied across different subpopulations (Table 3.2). The discussion so far concentrated on human genomes because to our knowledge all published studies on ROH patterns are on human data. It is obvious, however, that bovine genomes is different as compared to human genomes in many aspects such as population structure, effective population size, genome wide characteristics, purpose for artificial and natural selection. In this study we have found several common ROH islands which could be a sign of either artificial or natural selection. For example, the common ROH islands on chromosome 6 which were observed only in Brown Swiss and Fleckvieh are supposed to be a sign of strong artificial selection on these breeds. Notably, these breeds have the same breeding purpose, e.g. milk production is important for both breeds, while, Angus has beef production importance. Furthermore, there were ROH islands observable on chromosome 16 only present in Angus, where genes were in association with QTLs important for beef production; e.g. like average daily gain, juiciness, muscle area and etc., which presumably is a sign of strong artificial selection for beef production traits in Angus. For common ROH islands across all breeds, like ROH islands located on chromosome 21, it can be speculated that these ROH islands are just common strands of haplotypes located in the genome at regions with low haplotype diversity. This possibility is more applicable for these patterns, since the genes inside these ROH islands are mostly with unknown function without any association with QTLs. A second possibility could be a historical phenomenon which has been experienced by all subpopulation; e.g. bottleneck event, before the time that these subpopulations

separated from each other. There was also another interesting ROH islands on chromosome 10 only observable in Indicine cattle breeds. We can speculate that this is due to natural selection for favorable traits in the tropical condition, e.g. adaption to heat or parasites infection, which has led to the fixation of best alleles for relevant traits in these breeds. One common characteristic, for all ROH islands was that they were all in gene–rich areas. Number of genes within ROH islands was significantly higher than number of genes in neighborhoods. We can speculate that these ROH islands might be due to selective sweeps which have caused the fixation of physiologically important genes in those regions. We have also found some ortholog genes in ROH islands on chromosome 21 with the ROH islands locating on chromosome 3 and 11 in the human. The analysis of ROH islands that are common across species deserves detailed exploration.

5. Conclusion

We confirmed the existence of ROH islands patterns across breeds, subspecies and production types of cattle. Different patterns were found across the breeds consistent with possible signatures of either artificial or natural selection. Some pattern locations have been reported as highly associated to productive-economical traits (QTLs). We have also found some ortholog genes in ROH islands on chromosome 21 with the ROH islands locating on chromosome 3 and 11 in the human which means that these genes can be highly conserved across many species.

Since these patterns are an indication of common ancestry in past generations, these results could be beneficial for evolutionary studies as well as for GWAS (eg. fine mapping), There is a possibility to use these ROH islands as surrogate markers in case-control experiments to see whether there is any relationship between them and recessive disorders. The same can be done for studying traits of economic importance.

The result of ROH incidence from the 3 software tools was different for the whole genome. Even though the software tools were found to be different, the regions discovered by the three software tools were always in similar locations with slight differences in the frequency of the signals. This difference at incidence of ROH islands could make a difference in the grouping of signals for breed comparisons but not in finding the exact location of ROH signals.

6. References

Broman K.W., Weber J.L. (1999) Long homozygous chromosomal segments in reference families from the Centre d'E ´tude du Polymorphisme Humain. Am. J. Hum. Genet., 65, 1493–1500.

Bjelland D.W., Weigel K.A., Vukasinovic N., Nkrumah J.D. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J. Dairy Science., 96, 4697–4706.

Chan EK., Nagaraj S.H., Reverter A. (2010) The evolution of tropical adaptation: comparing Taurine and Zebu cattle. J. Anim Genet., 41, 67–77

European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI). QuickGO. Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK. 2013.

Ferenčaković M., Hamzic E., Gredler B., Curik I., Sölkner J. (2011) Runs of homozygosity reveal genome-wide autozygosity in the Austrian fleckvieh cattle. Agric. Conspec. Sci., 76, 325–328.

Ferenčaković M., Hamzic, E.Gredler, B. T.R, Solberg, G: Klemetsdal, I. Curik and J. SölknerSölknerSölkner. (2013) Estimates of autozygosity derived from runs of homozygosity:empirical evidence from selected cattle populations. J. Anim. Breed. Genet., 130, 286-293. .

Golden Helix, Inc. website. Available: http://www.goldenhelix.com. Accessed (2013) January 13.

Golden Helix, Inc. website. Available: http://www.goldenhelix.com. Accessed 2013 January 13.Gibson M., Newton E. and Collins, Andrew. (2006) Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics.,15, 789–795.

Gutiérrez Gil B., Wiener P., Nute GR., Burton D., Gill JL., Wood JD., Williams JL. (2008) Detection of quantitative trait loci for meat quality traits in cattle. Anim. Genet., 39, 51-61.

Howrigan D.P., Simonson M.A., Keller M.C. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics., 12, 1471-2164.

Illumina (2011) Bovine SNP50 Genotyping BeadChip (available at: http://www.illumina.com/documents/ products/datasheets/datasheet_bovine_snp50.pdf; last accessed 6 January 2011).

Keller M.C., Visscher P.M., Goddard M.E. (2011) Quantifi- cation of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics., 89, 237–249.

Li L.H., Ho, S.F., Chen, C.H., Wei, C.Y., Wong, W.C., Li, L.Y.,Hung, S.I., Chung, W.H., Pan, W.H., Lee, M.T. (2006) Long contiguous stretches of homozygosity in the humangenome., Hum. Mutat, 27, 1115-1121

McClure MC., Morsci N.S., Schnabel R.D., Kim J.W., Yao P., Rolf M.M., McKay S.D., Gregg S.J., Chapple R.H., Northcutt S.L., Taylor J.F., (2010) A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. Anim. Genet.,41, 597-607.

McQuillan R., Leutenegger A.L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Naran- cic N., Janicijevic B., Polasek O., Tenesa A., Macleod A.K., Farrington S.M., Rudan P., Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Camp- bell H., Wilson J.F. (2008) Runs of homozygosity in European populations. Am. J. Hum. Genet., 83, 359–372.

Nothnagel M., T. Lu, M. Kayser, M. Krawczak. (2009). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Human Molecular Genetics., 19, 2927 -2935.

Pirchner F. (1985). Genetic structure of population. 1. Closed popu- lations or matings among related individuals. In: General and Quantitative Genetics (A.B. Chapman, ed.), World Animal Science A4 Elsevier, Amsterdam. 227-250.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet., 81, 559-575.

R project for Statistical computing. The Comprehensive R Archive Network. Hosted by the Institute for Statistics and Mathematics of Wirtschaftsuniversität Wien, 2013. http://www.r-project.org/

SAS (2009). SAS/STAT® User's Guide Version 9.2. SAS Institute Inc. Cary, NC, USA.

Saccheri I.J., Brakefield P.M., Nichols R.A. (1996) Severe inbreeding depression and rapid fitness rebound in the butterfly Bicyclusanynana (Satyridae) Evolution. 50:2000–2013.

Sölkner. (2011) Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. Agriculturae Conspectus Scientificus., 76, 325-328.

Trevor J., Pemberton, Devin Absher, Marcus W. Feldman, Richard M. Myers, Noah A. Rosenberg and Jun Z. Li. (2012) Genomic Patterns of Homozygosity in Worldwide Human Populations. Am. J. Hum. Genet., 91, 275-292.

Wang, H., Lin, C.H., Service, S., Chen, Y., Freimer, N. and Sabatti, C. (2006) Linkage disequilibrium and haplotype homozygosity in population samples genotyped at a high marker density. Hum. Hered., 62, 175–189.

Wellcome Trust Sanger Institute / European Molecular Biology Laboratory – European Bioinformatics Institute. Ensembl release 72 - June 2013. http://www.ensembl.org

marker density. Hum. Hered., 62, 175–189. Wright S.Coefficients of inbreeding and relationship. Amer Naturalist 1922, 56:330-338.

Ensemble Genome Biomart tool (WTSI/EBI, 2013(Reference: Wellcome Trust Sanger Institute / European Molecular Biology Laboratory – European Bioinformatics Institute. Ensembl release 72 - June 2013. http://www.ensembl.org))

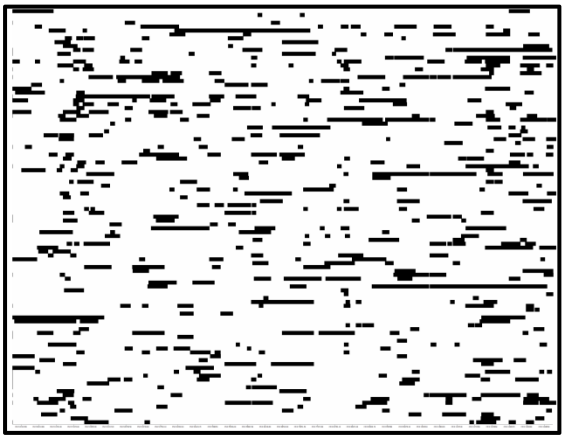
Zhang L, Orloff MS, Reber S, Li S, Zhao Y, et al. (2013) cgaTOH: Extended Approach for Identifying Tracts of Homozygosity. PLoONE8(3):e57772.doi:10.1371/journal.pone.0057772

R project for Statistical computing. The Comprehensive R Archive Network. Hosted by the Institute for Statistics and Mathematics of Wirtschaftsuniversität Wien, 2013. http://www.r-project.org/.

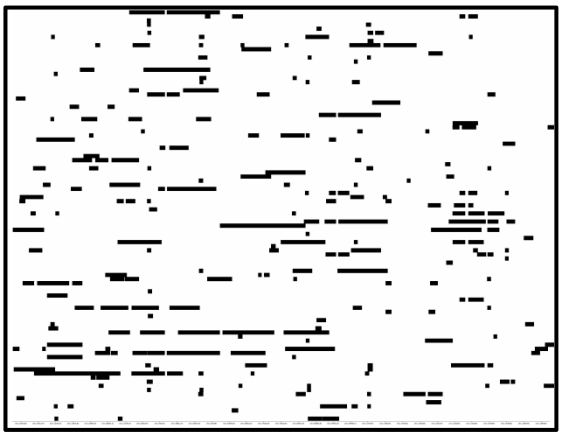
Appendix

CHR 1

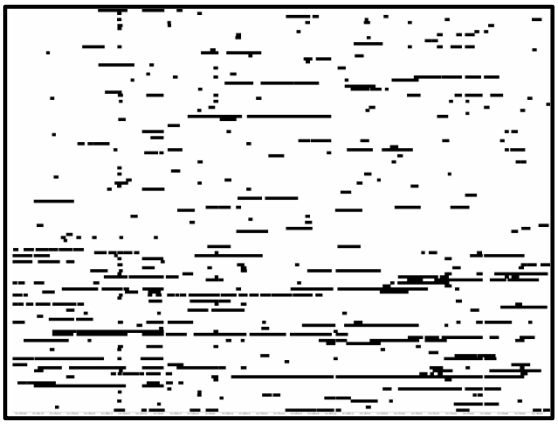
Angus



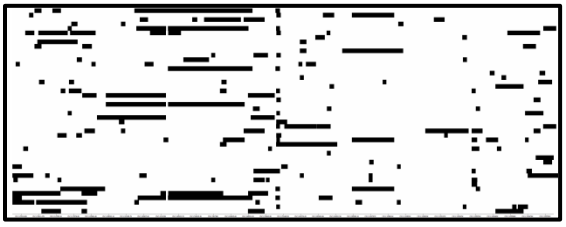
Brahman



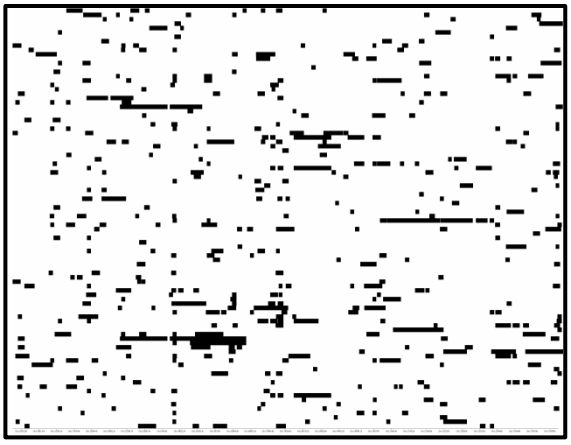
Nelore



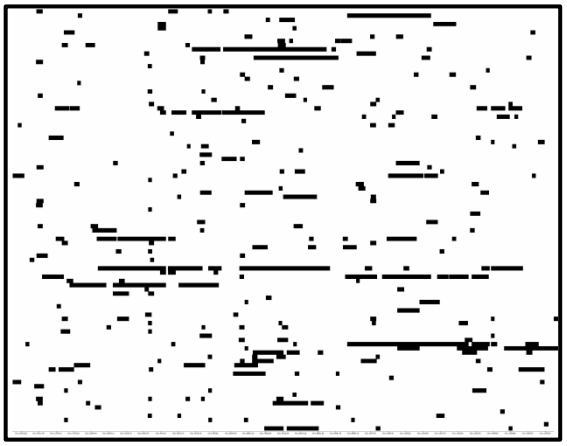
Brown Swiss



Fleckvieh

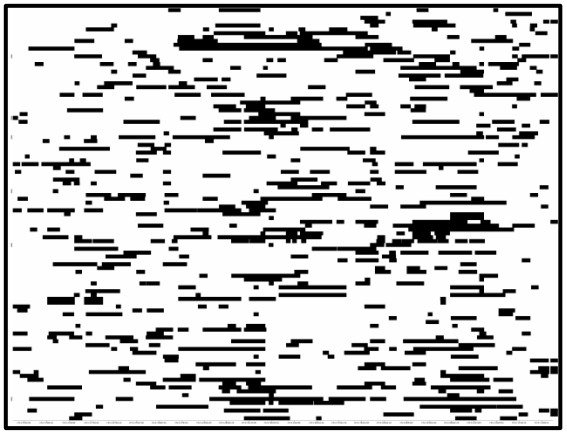


Gir

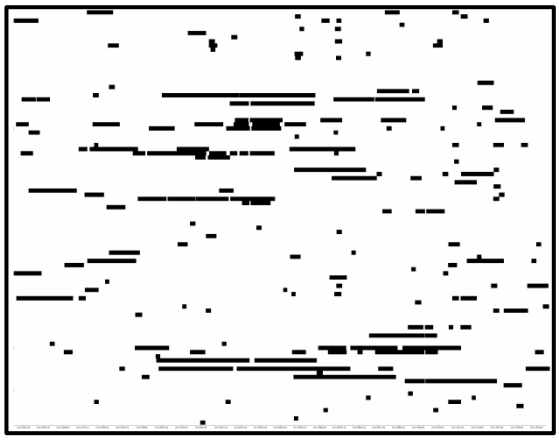


CHR 2

Angus

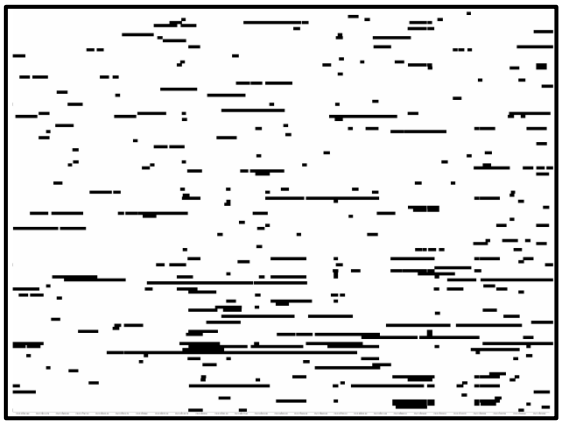


Brahman

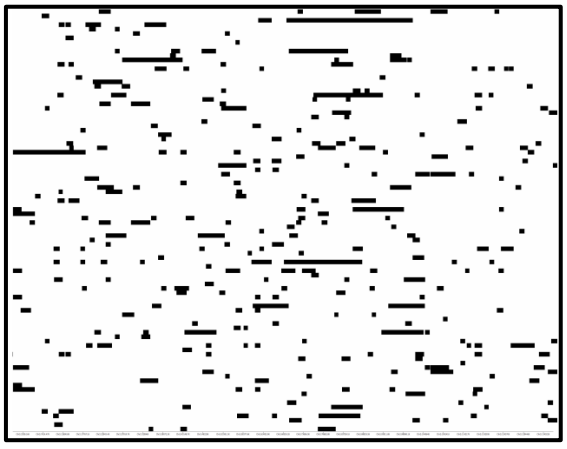


Brown Swiss

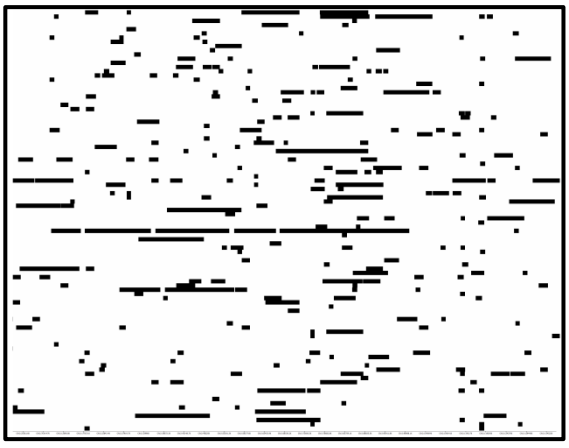
Nelore



Fleckvieh

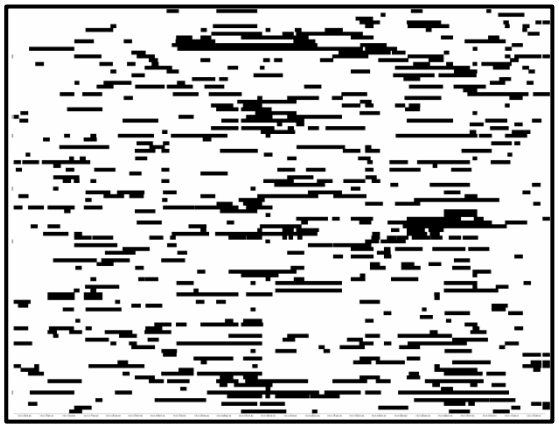


Gir

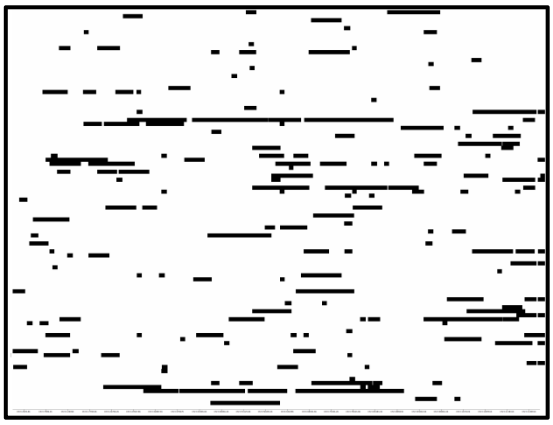


CHR 3

Angus

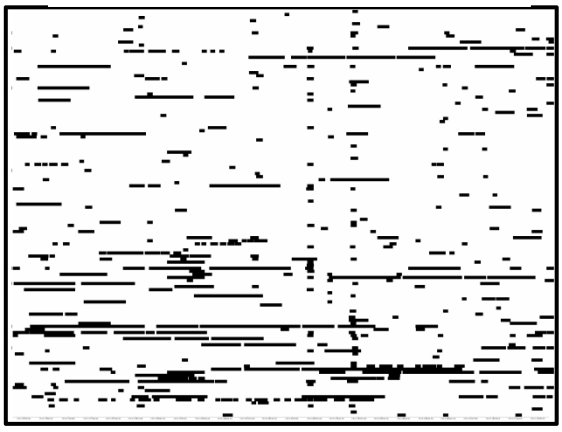


Brahman

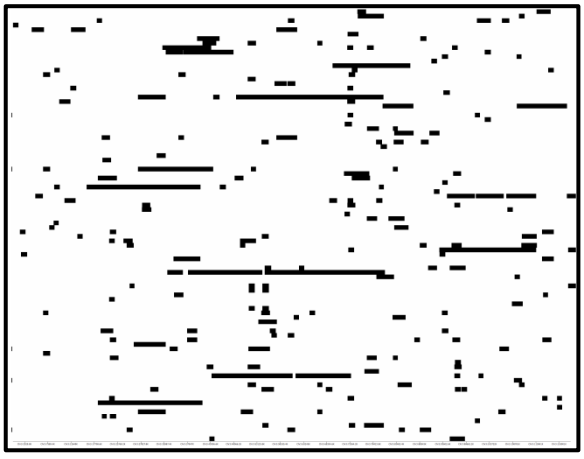


Brown Swiss

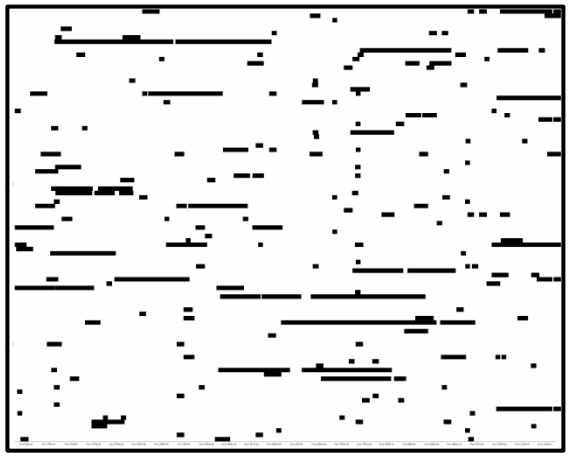
Nelore



Fleckvieh

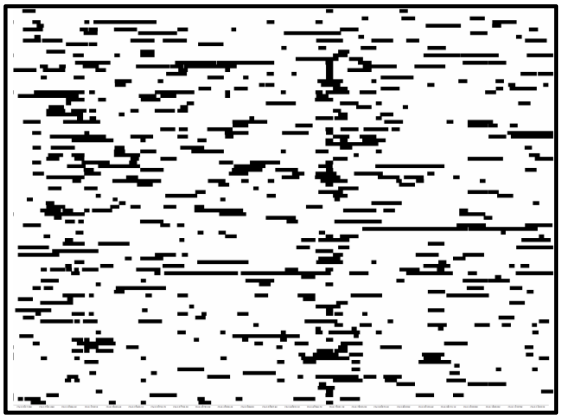


Gir

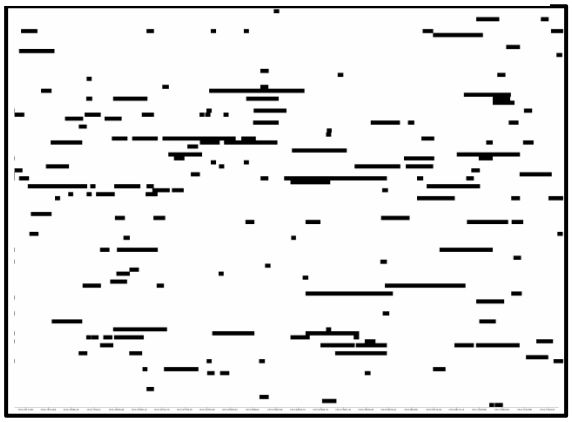


CHR4

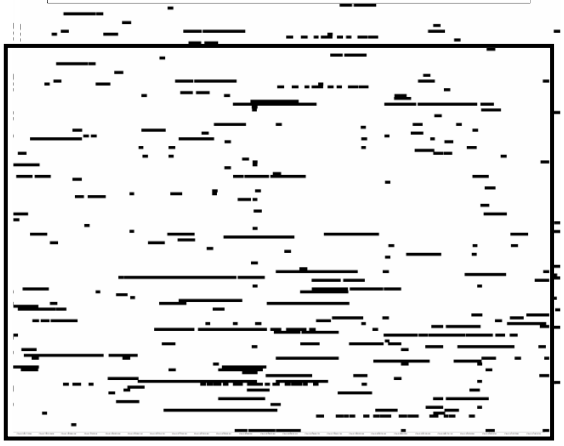
Angus



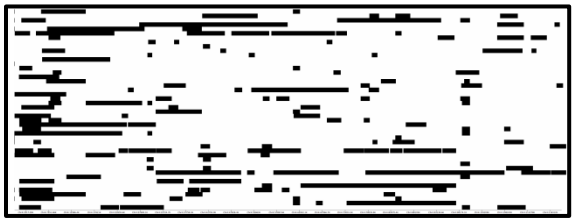
Brahman



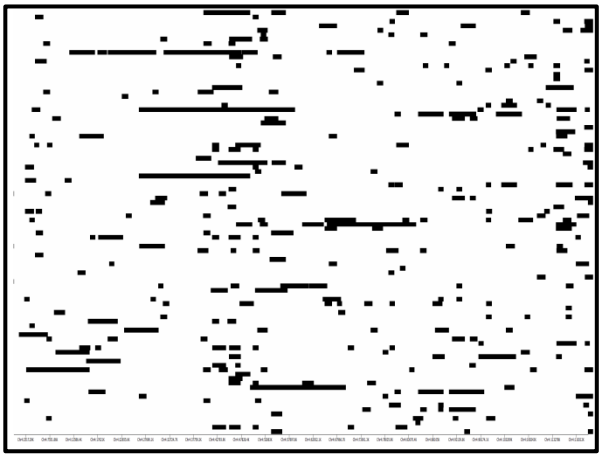
Nelore



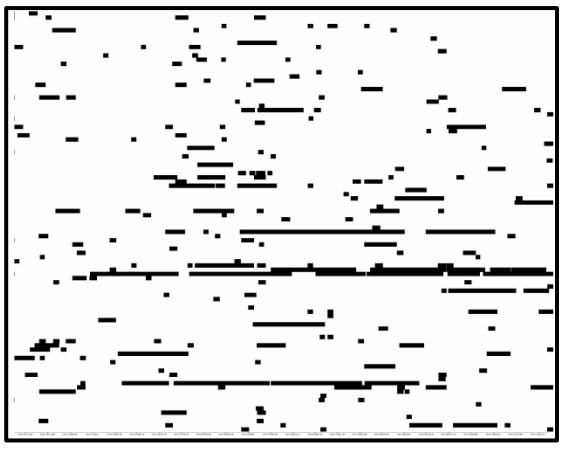
Brown Swiss



Fleckvieh

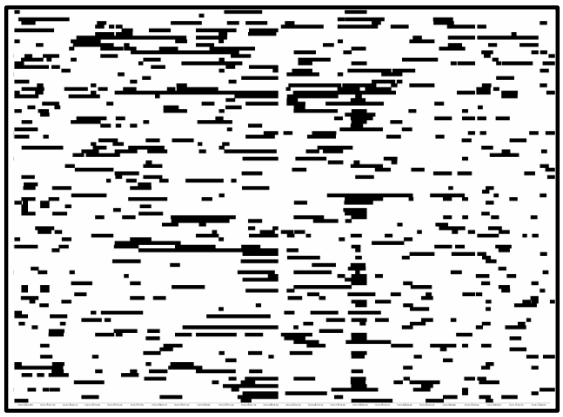


Gir

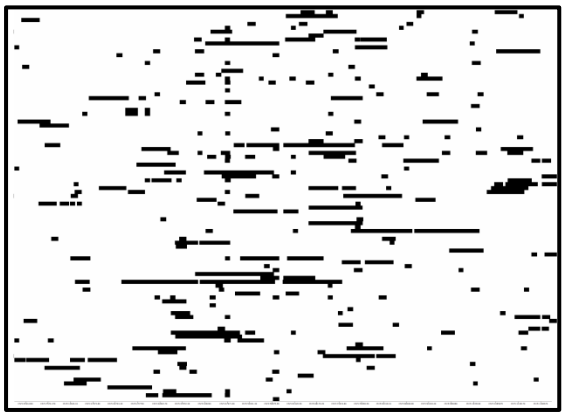


CHR5

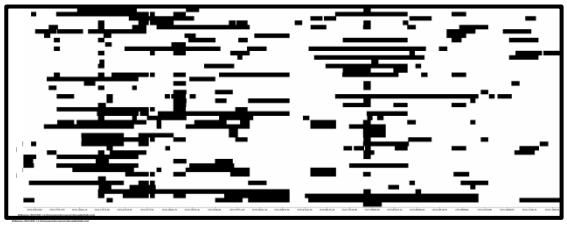
Angus



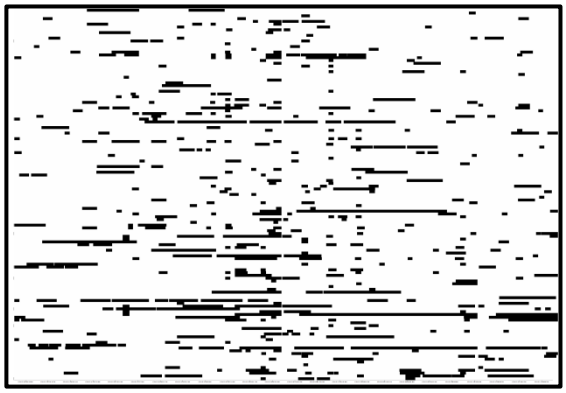
Brahman



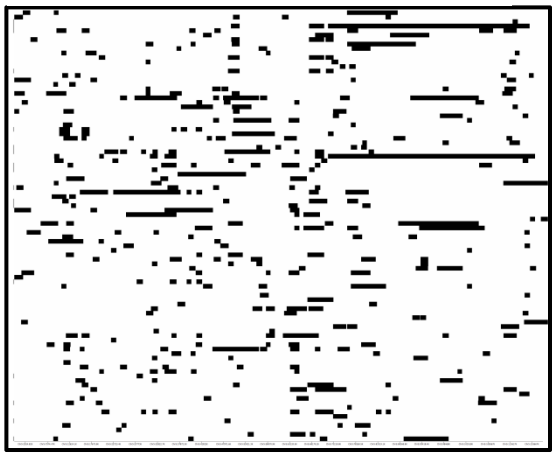
Brown Swiss



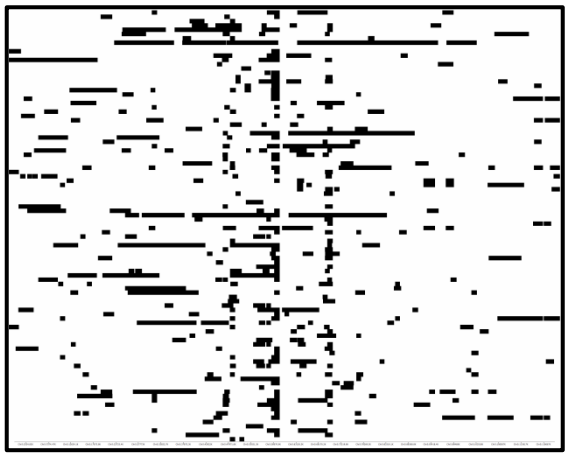
Nelore



Fleckvieh

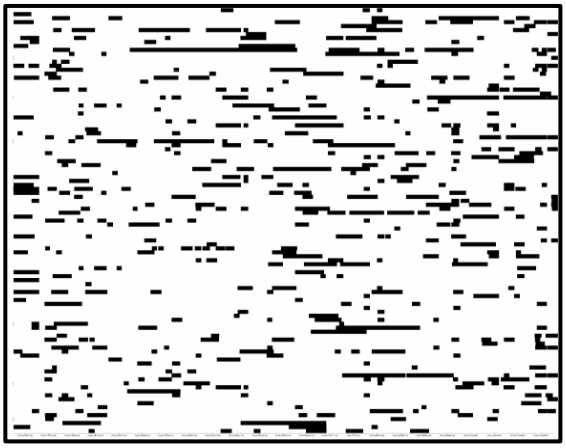


Gir

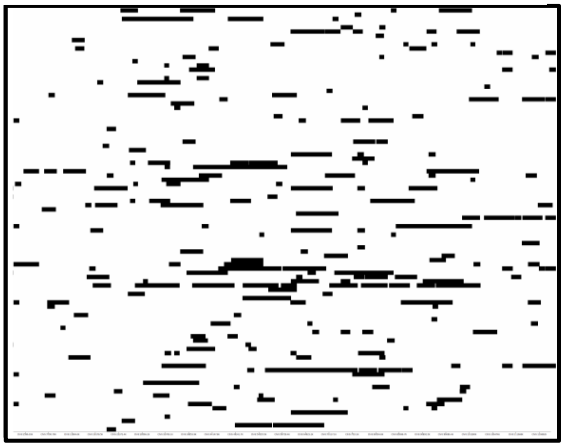


CHR 6

Angus

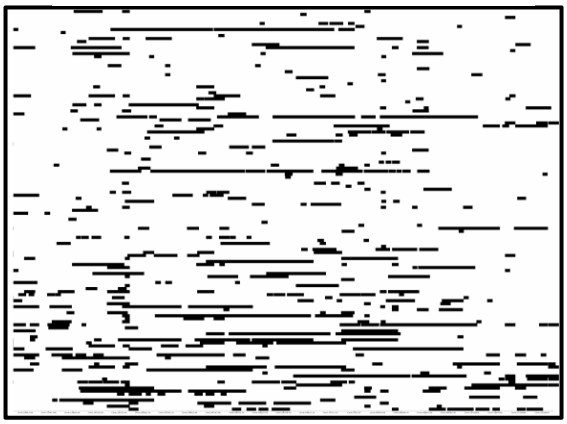


Brahman

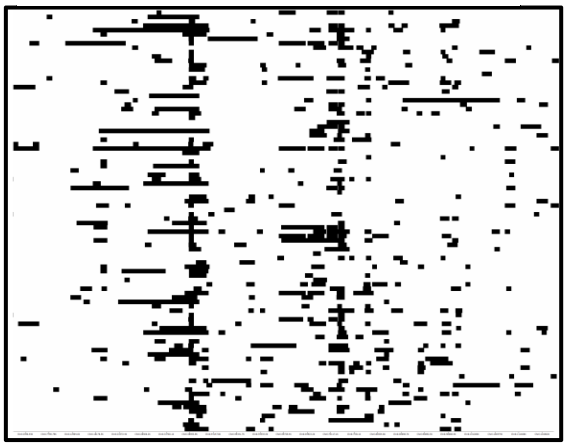


Brown Swiss

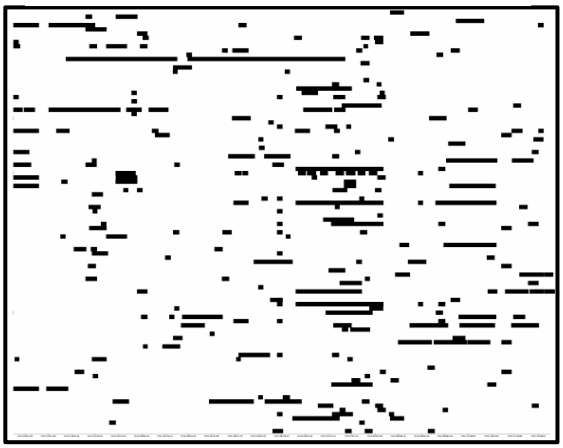
Nelore



Fleckvieh

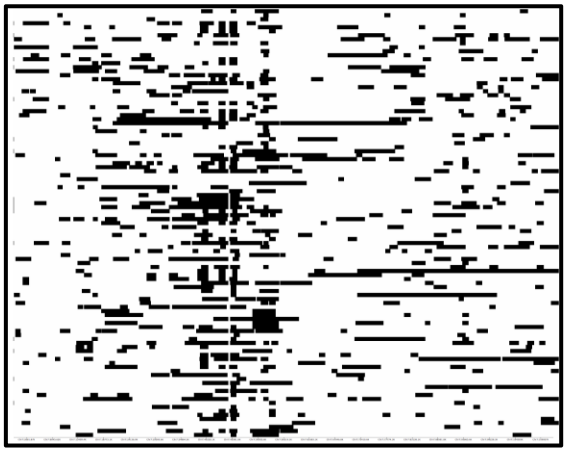


Gir

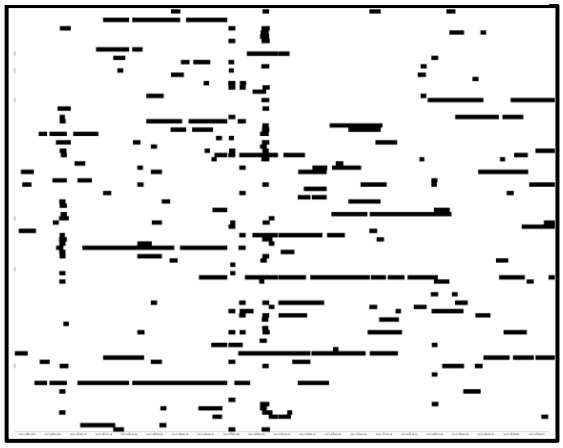


CHR7

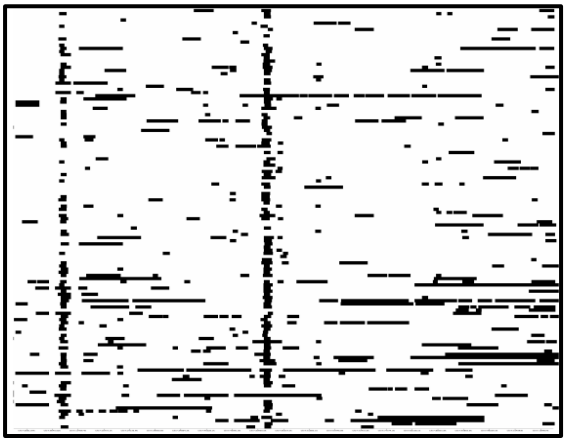
Angus



Brahman

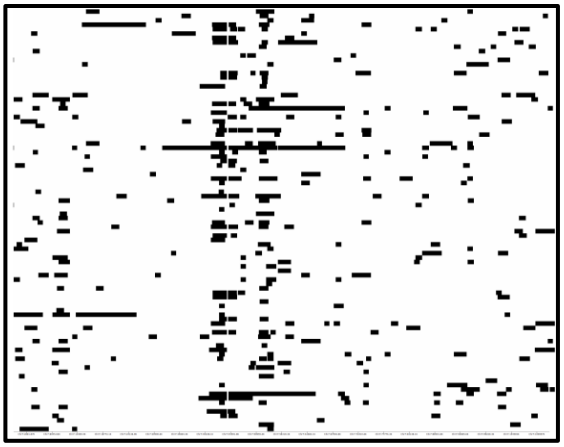


Nelore

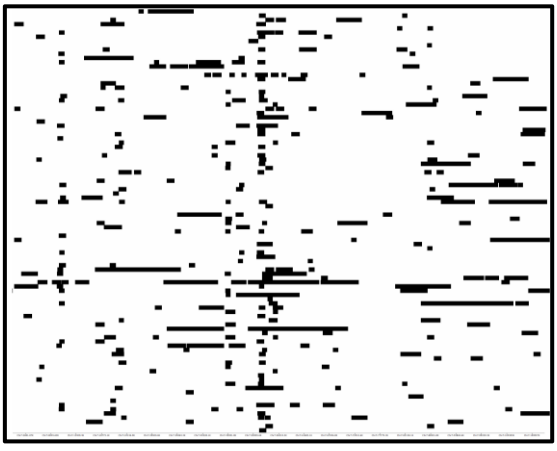


Brown Swiss

Fleckvieh

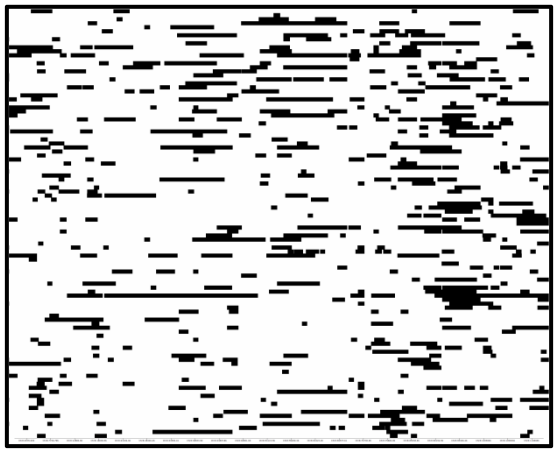


Gir

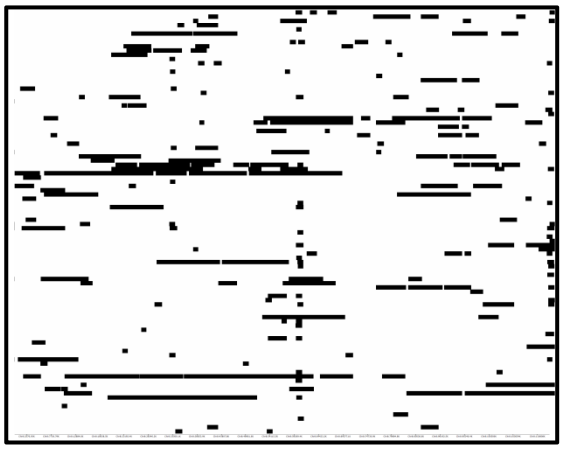


CHR 8

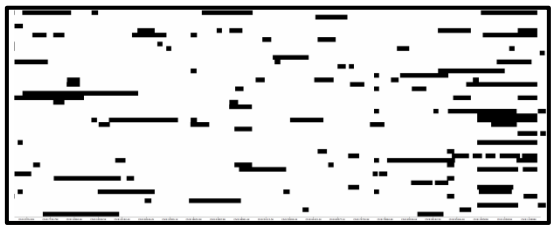
Angus



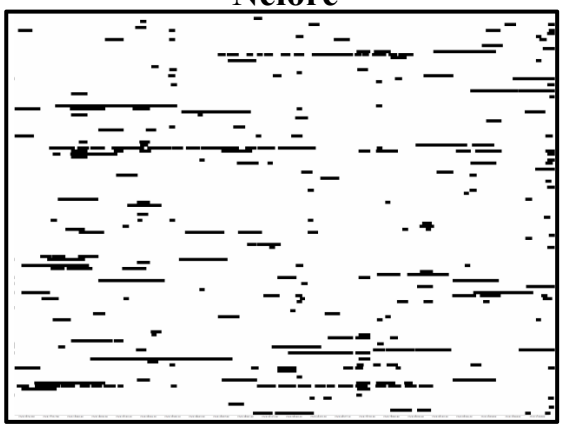
Brahman



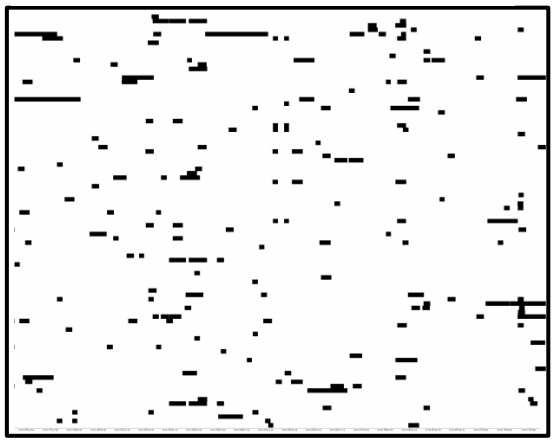
Brown Swiss



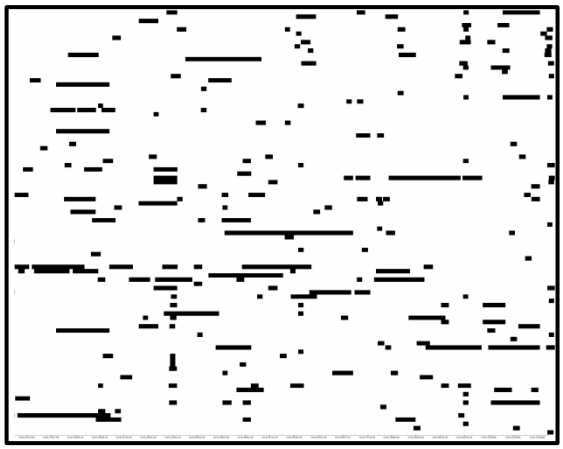
Nelore



Fleckvieh

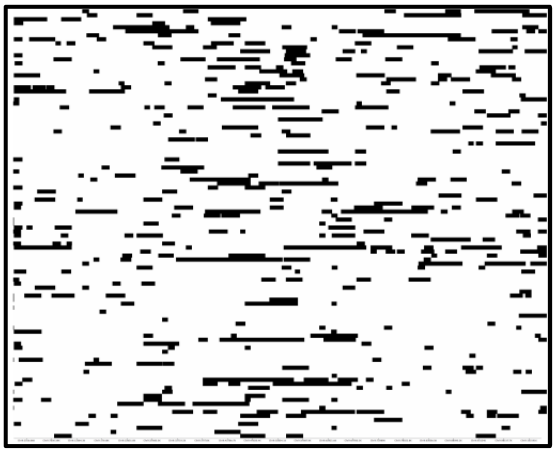


Gir

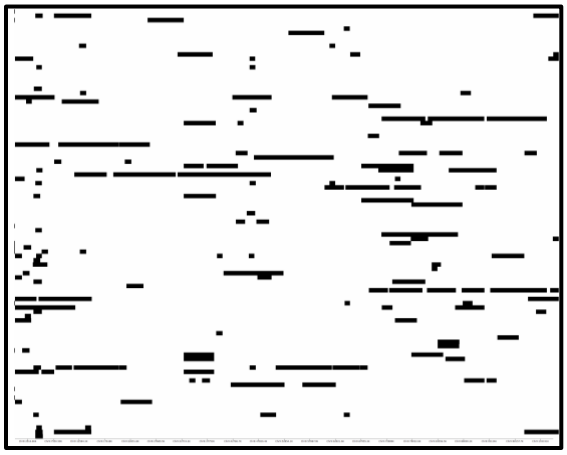


CHR 9

Angus

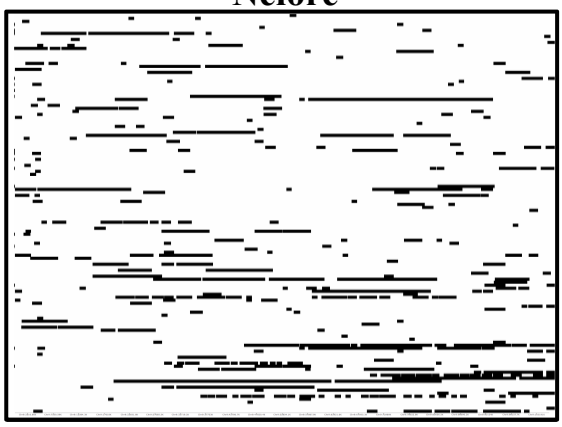


Brahman

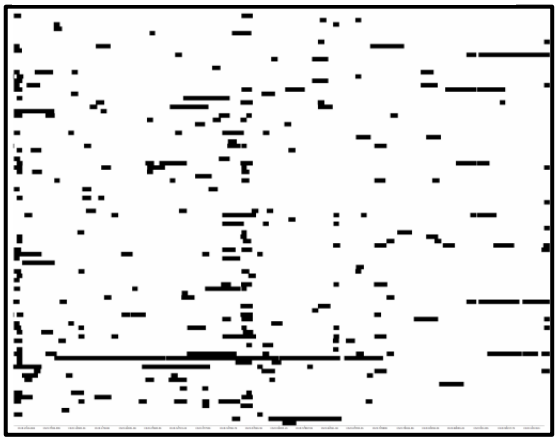


Brown Swiss

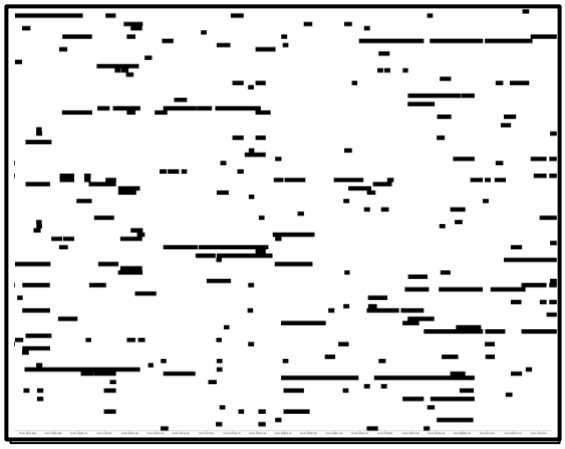
Nelore



Fleckvieh

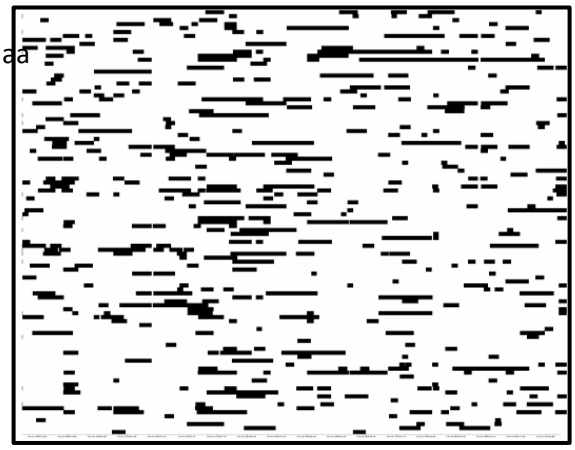


Gir

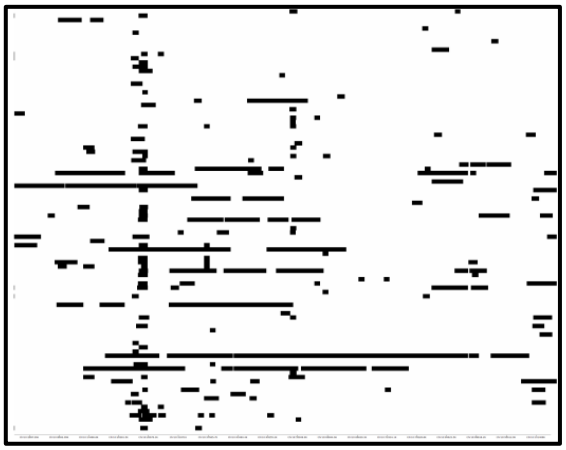


CHR10

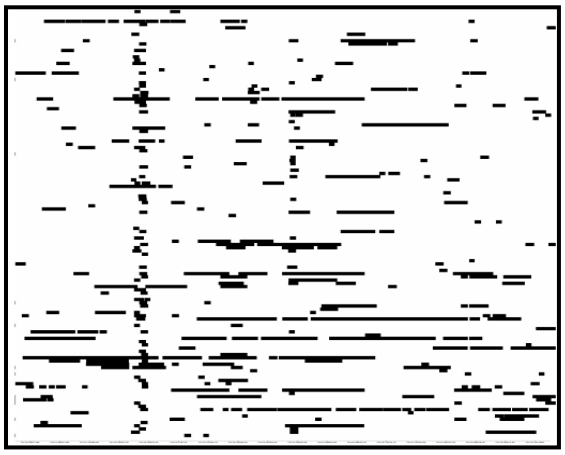
Angus



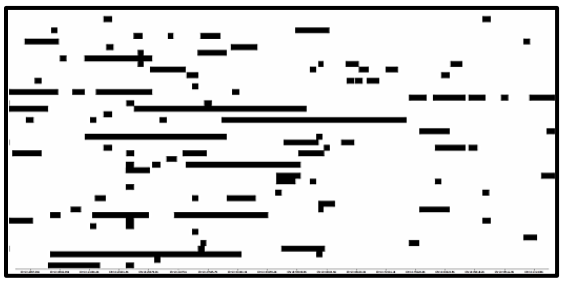
Brahman



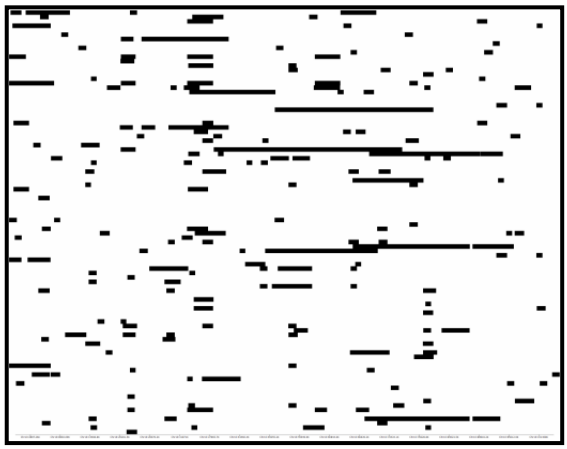
Nelore



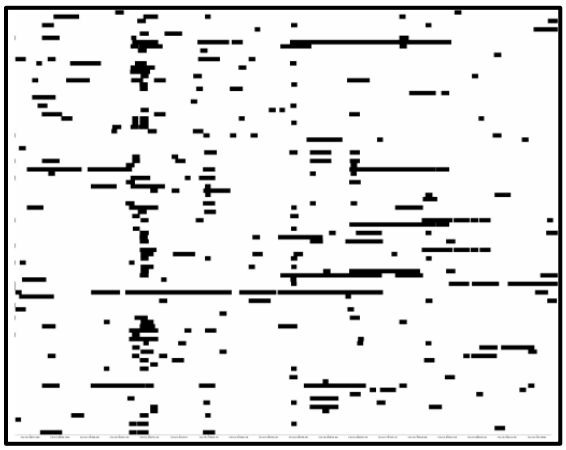
Brown Swiss



Fleckvieh

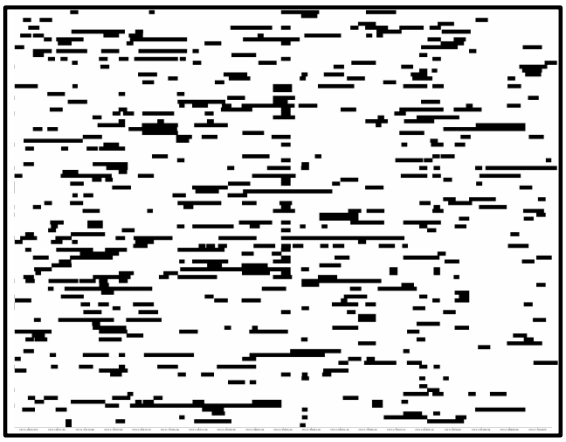


Gir

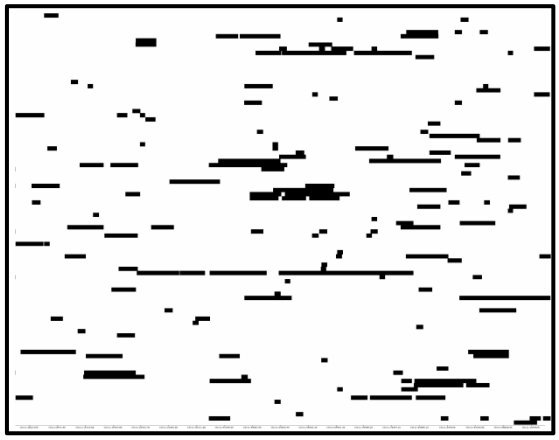


CHR 11

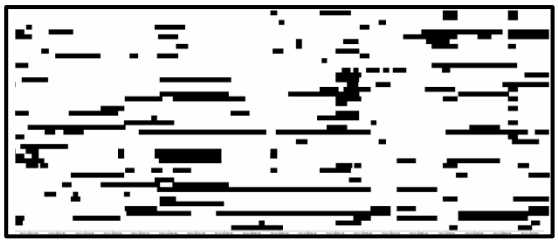
Angus



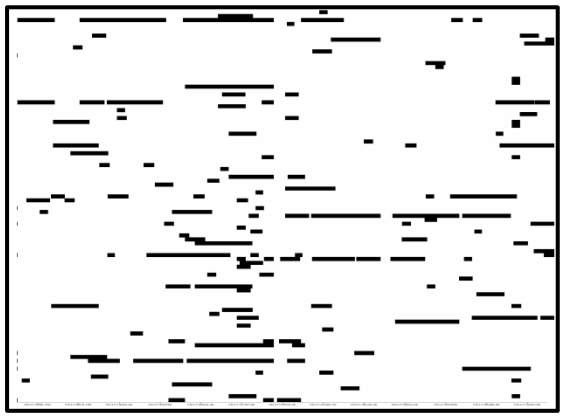
Brahman



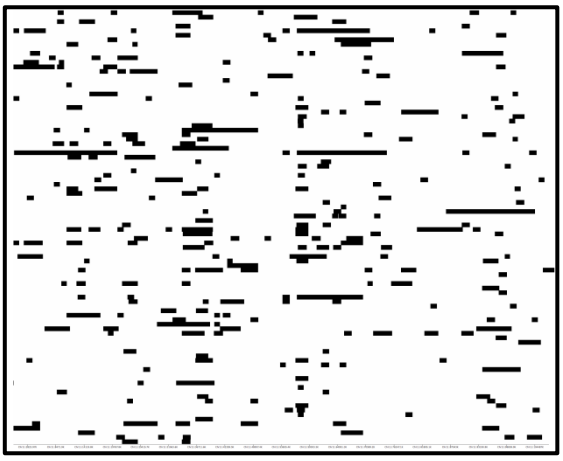
Brown Swiss



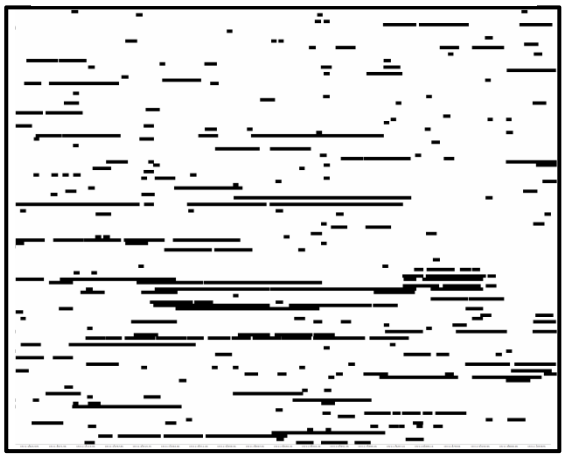
Nelore



Fleckvieh

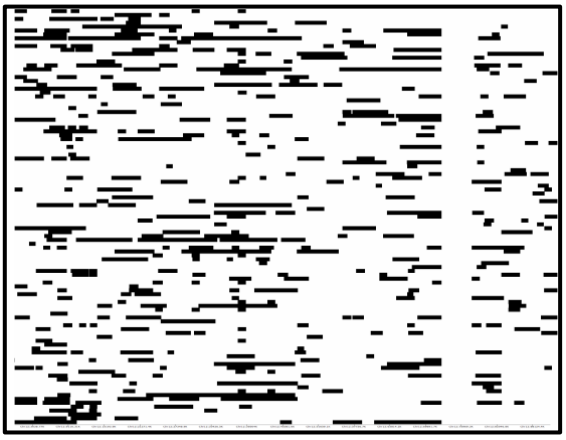


Gir

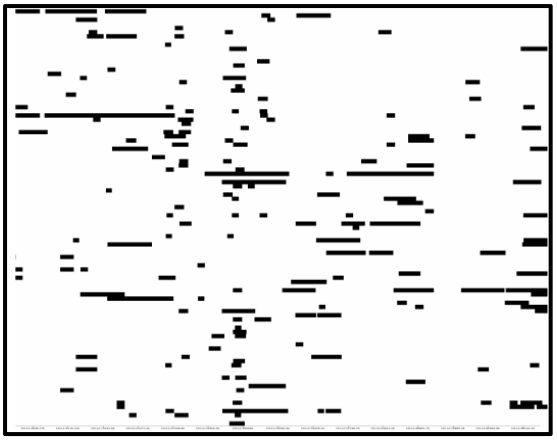


CHR 12

Angus

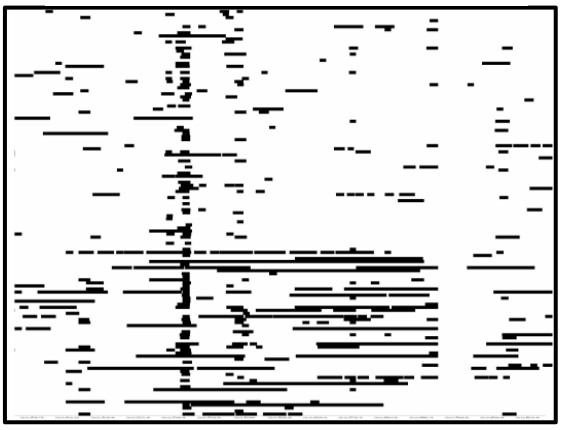


Brahman

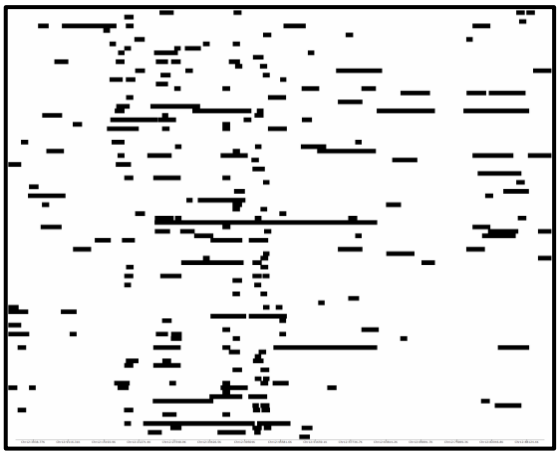


Brown Swiss

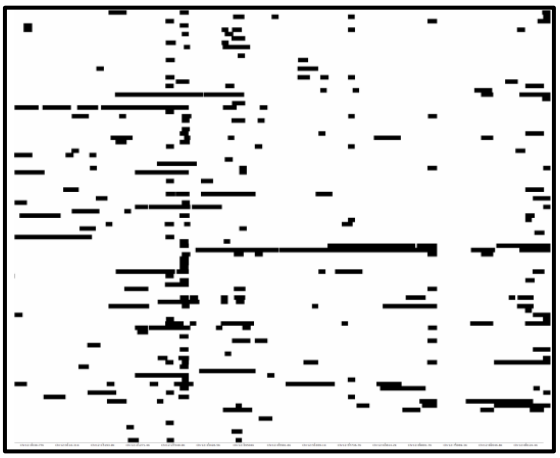
Nelore



Fleckvieh

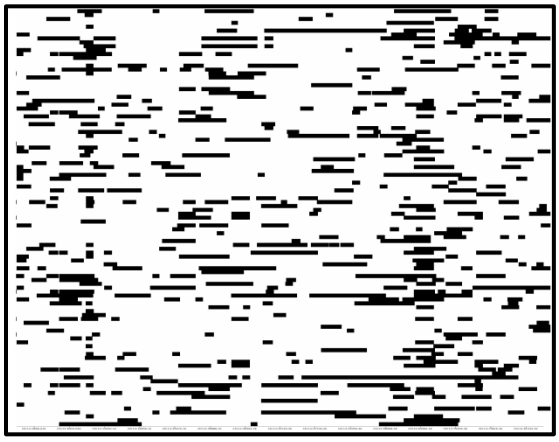


Gir

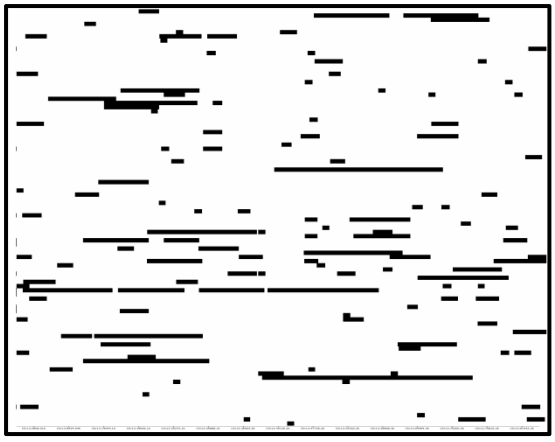


CHR 13

Angus

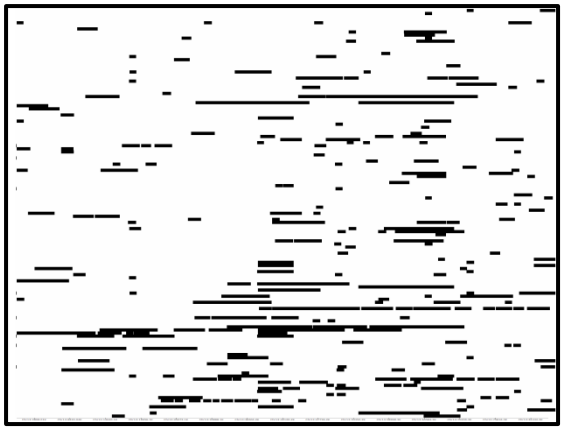


Brahman

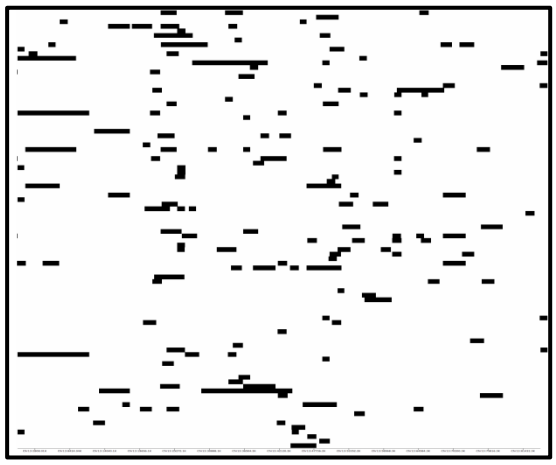


Brown Swiss

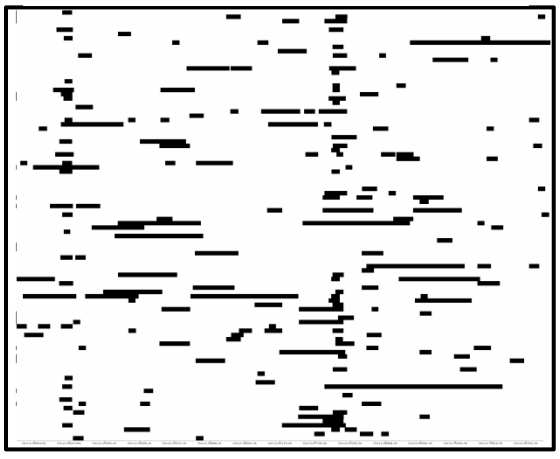
Nelore



Fleckvieh

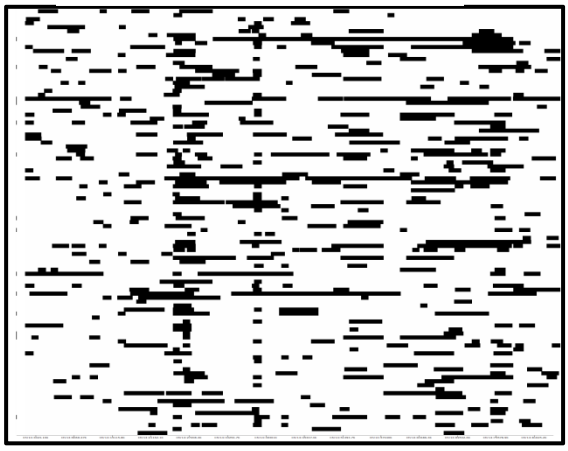


Gir

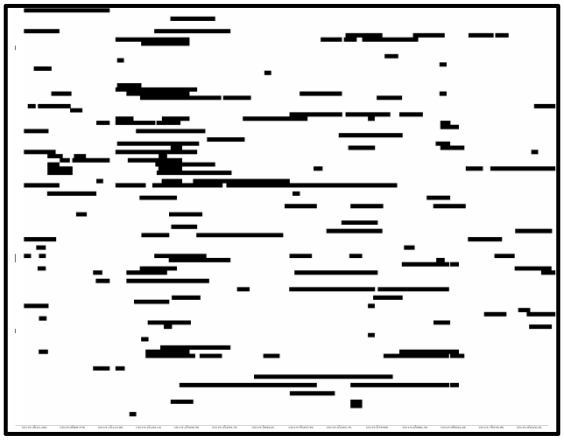


CHR 14

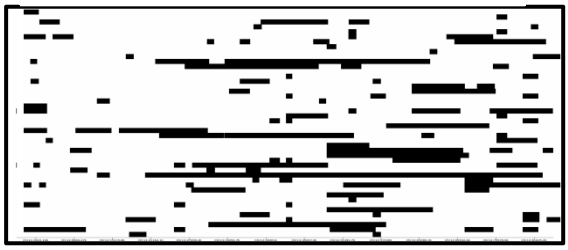
Angus



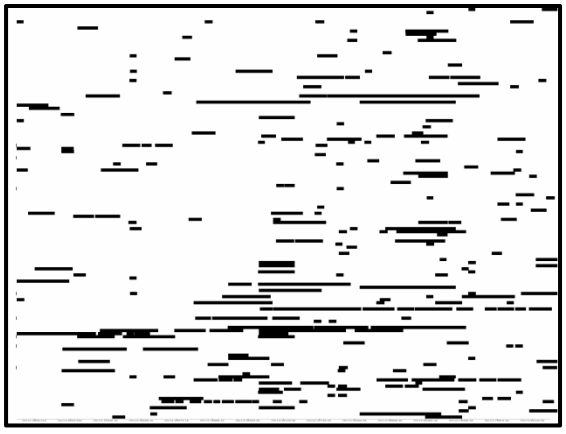
Brahman



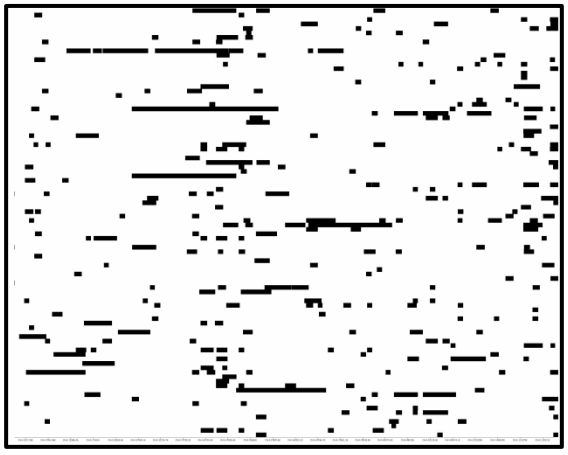
Brown Swiss



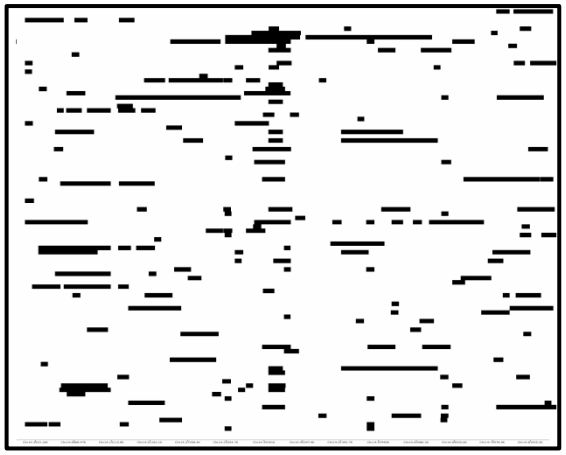
Nelore



Fleckvieh

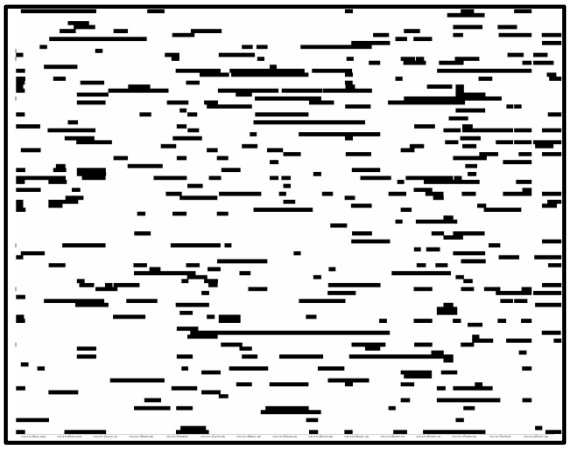


Gir

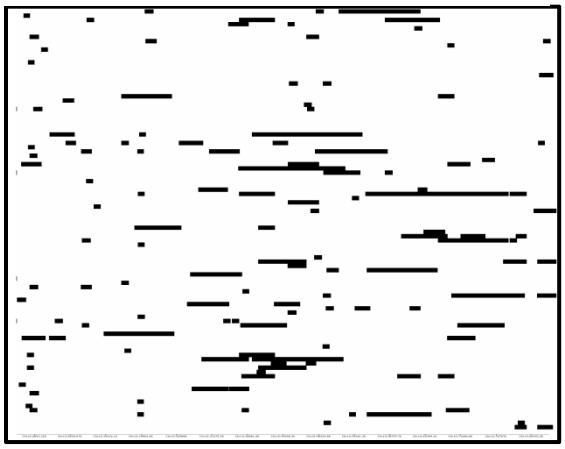


CHR 15

Angus

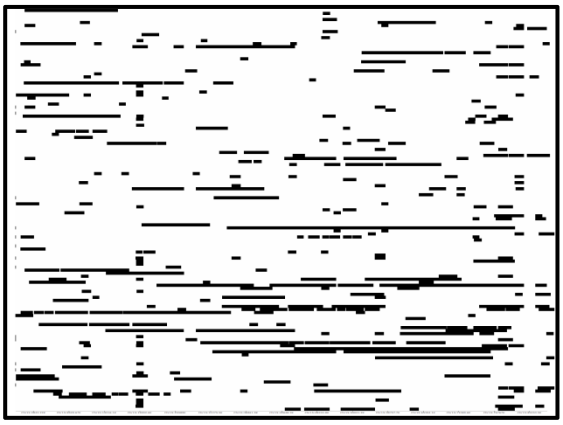


Brahman

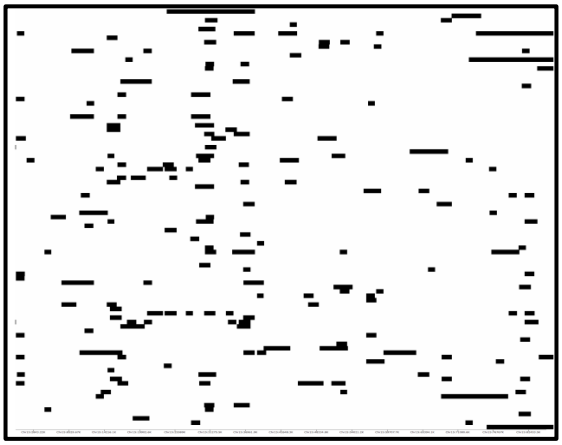


Brown Swiss

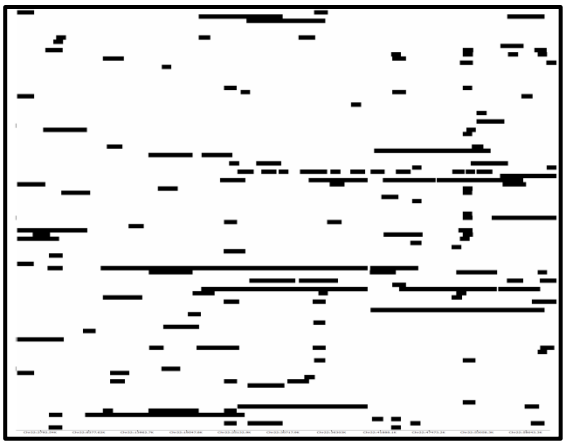
Nelore



Fleckvieh

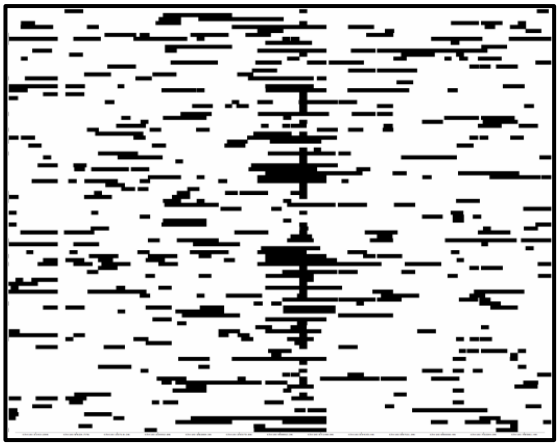


Gir

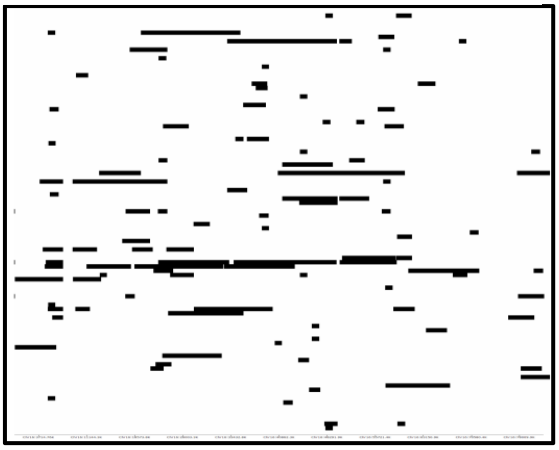


CHR 16

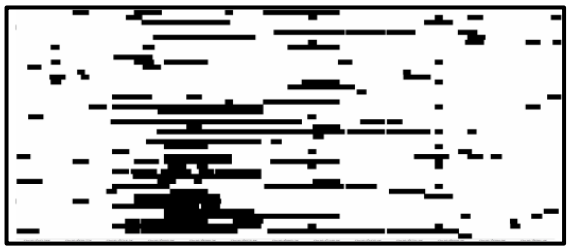
Angus



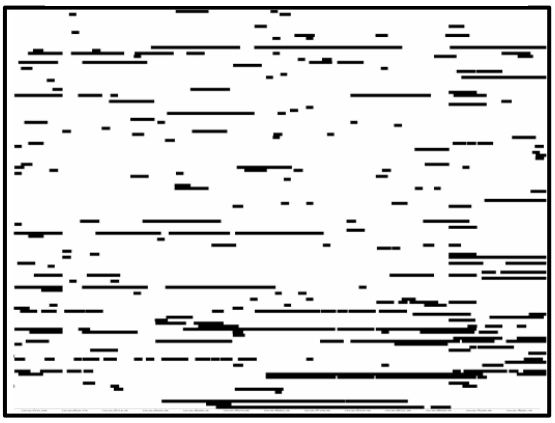
Brahman



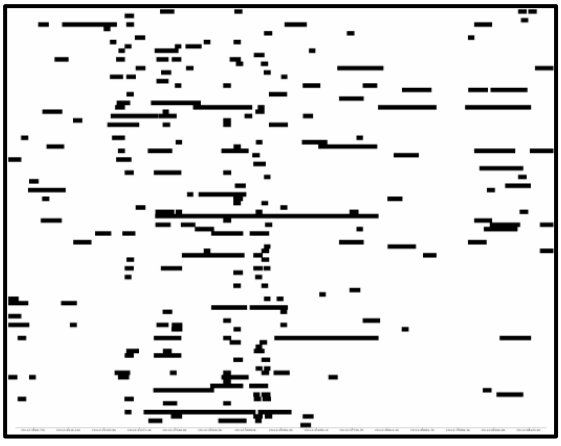
Brown Swiss



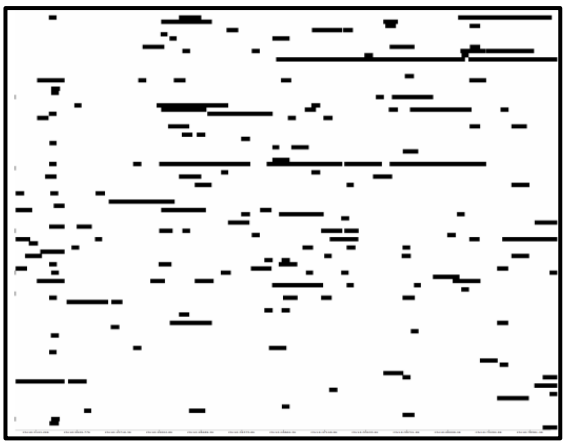
Nelore



Fleckvieh

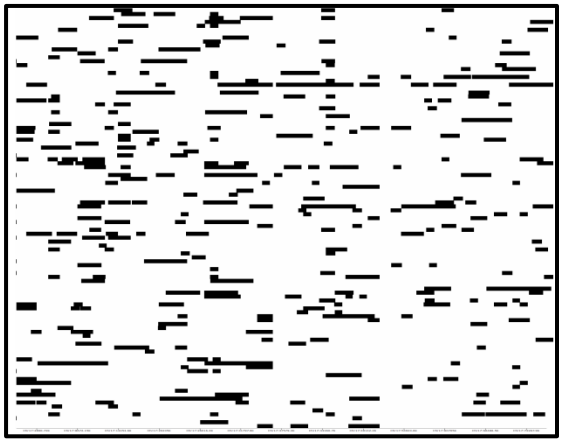


Gir

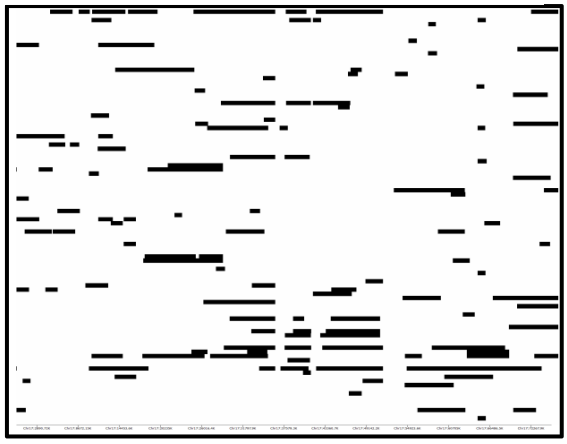


CHR 17

Angus

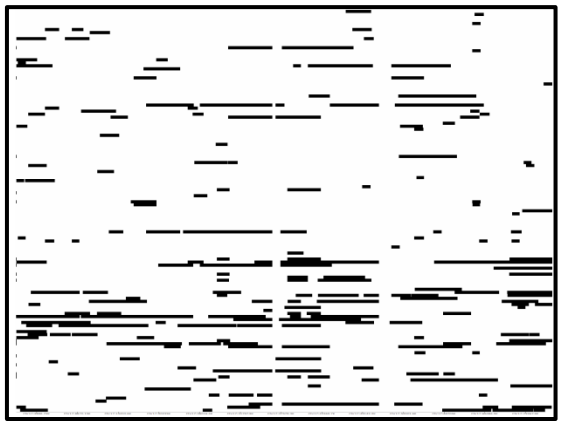


Brahman

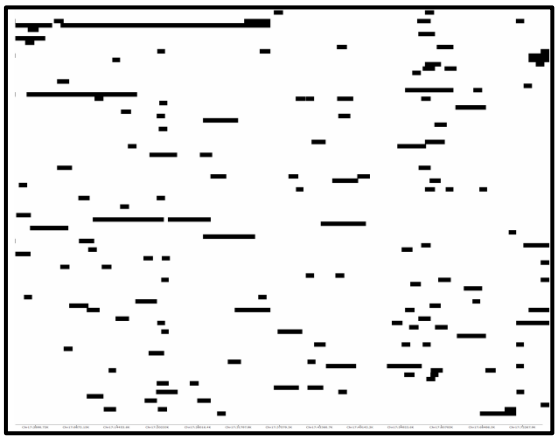


Brown Swiss

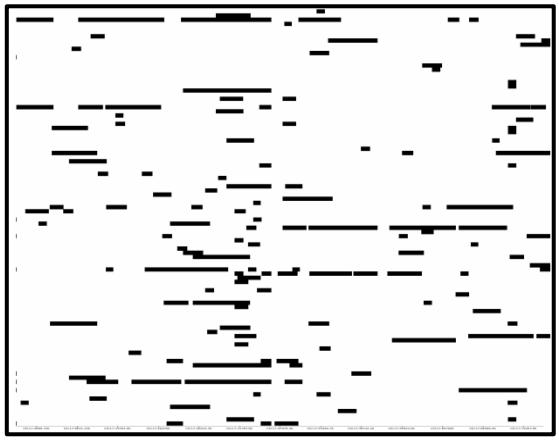
Nelore



Fleckvieh

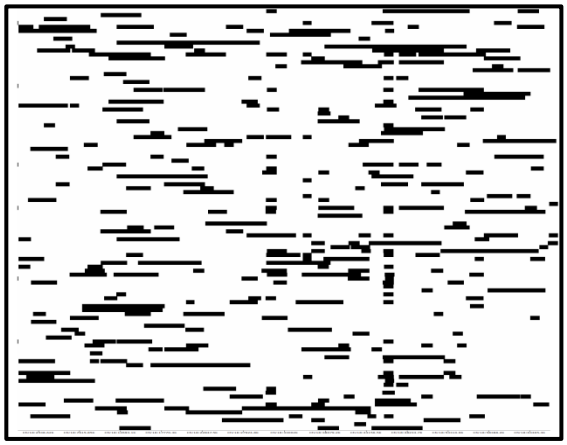


Gir

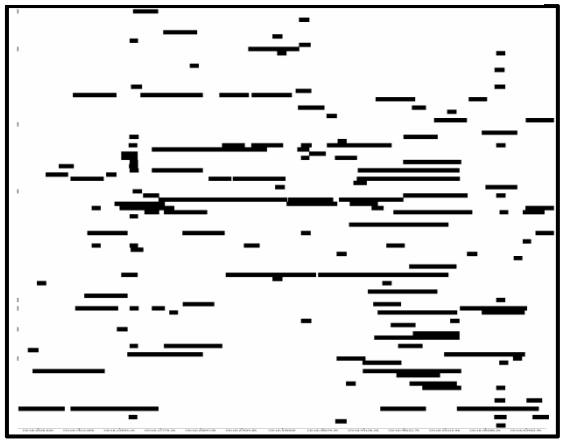


CHR 18

Angus

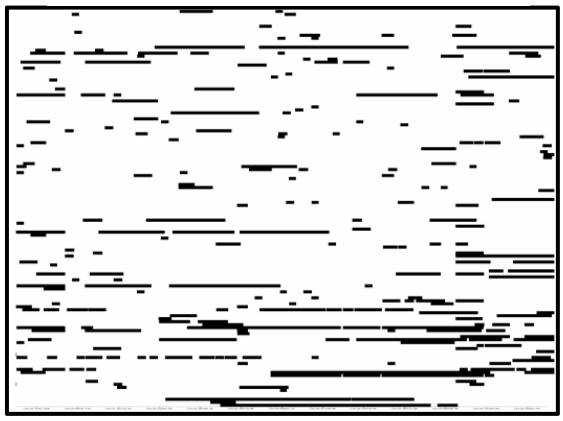


Brahman

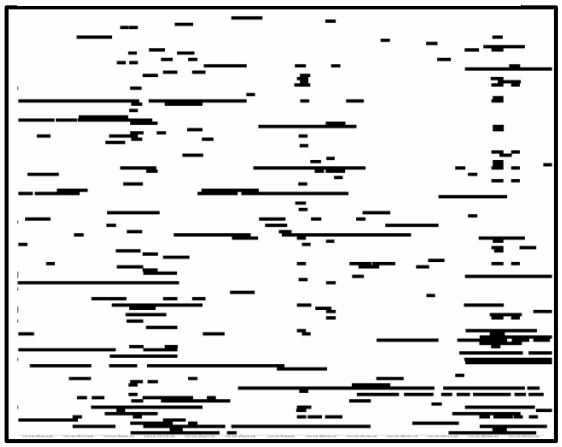


Brown Swiss

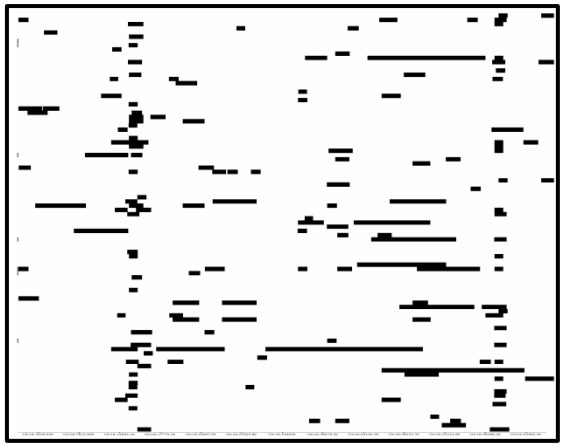
Nelore



Fleckvieh

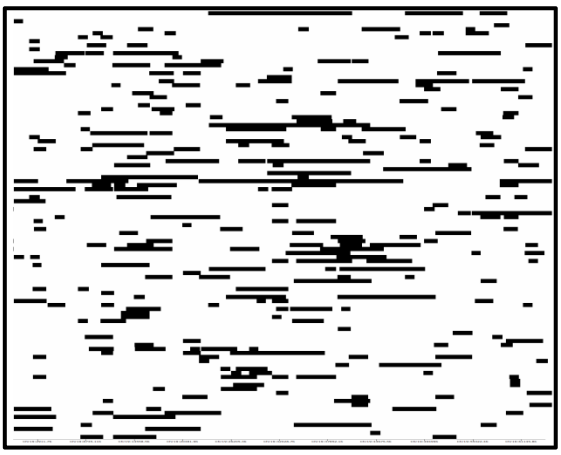


Gir

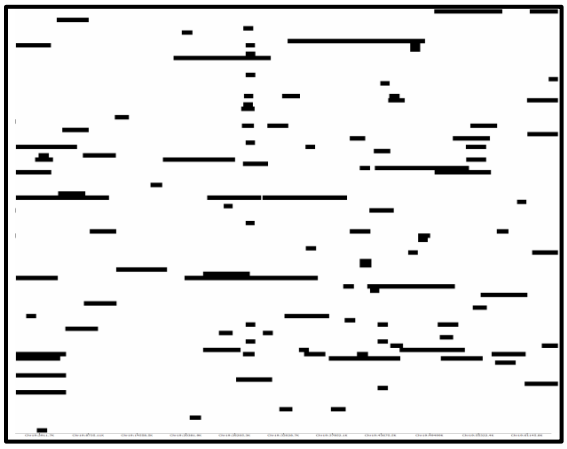


CHR 19

Angus

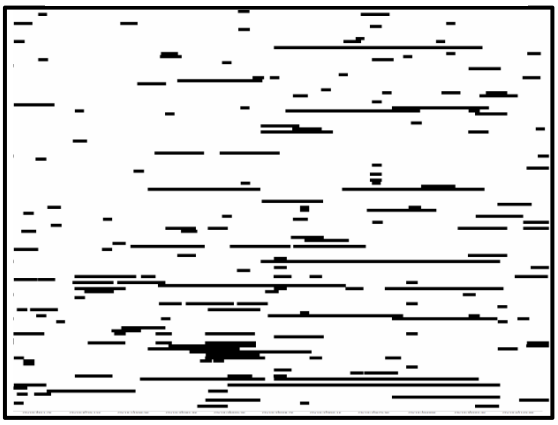


Brahman

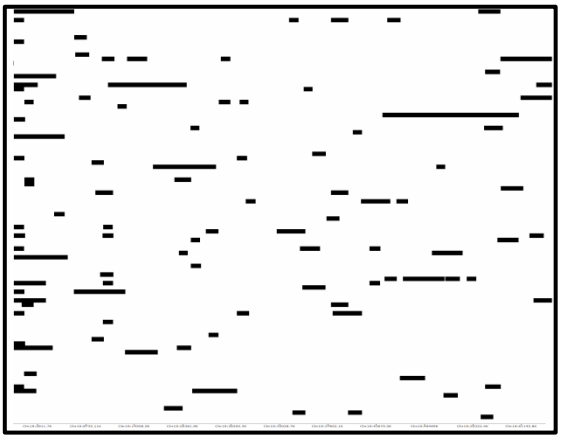


Brown Swiss

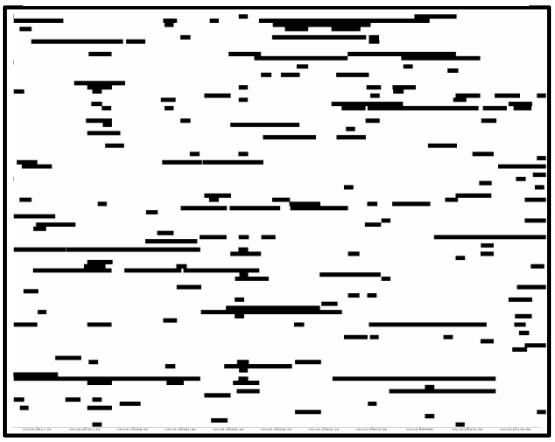
Nelore



Fleckvieh

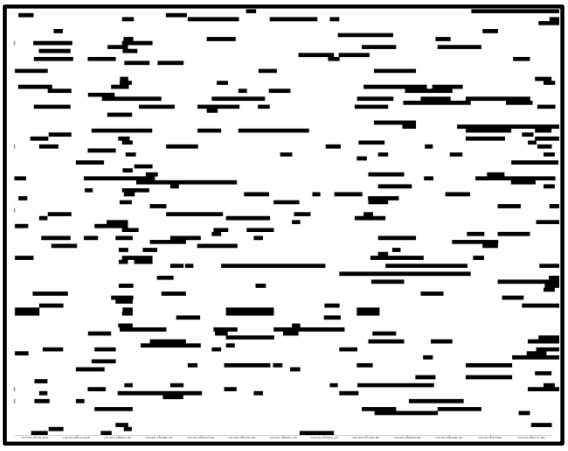


Gir

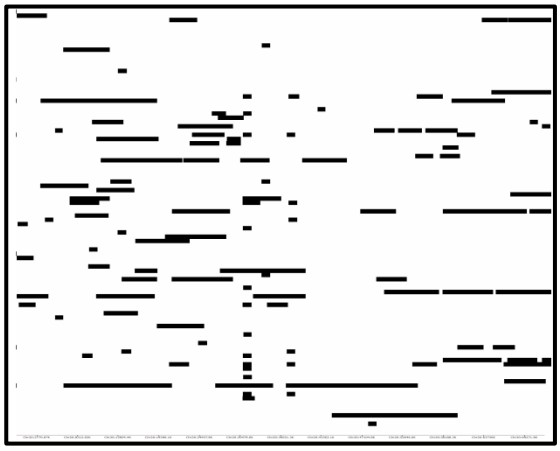


CHR 20

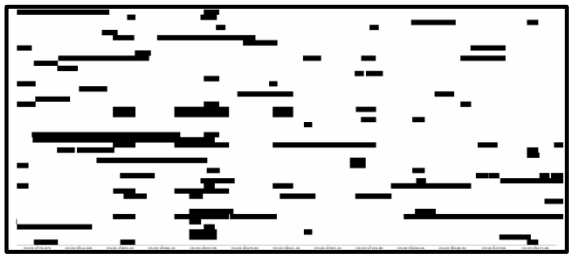
Angus



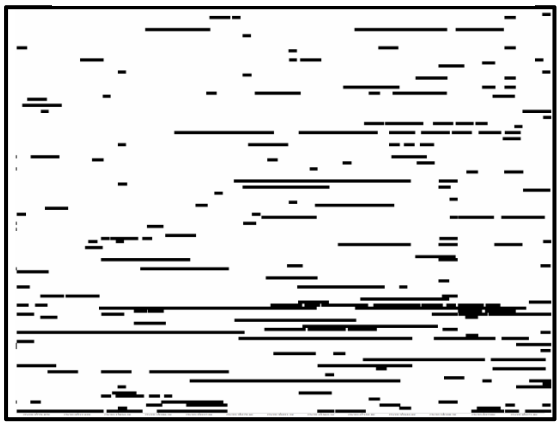
Brahman



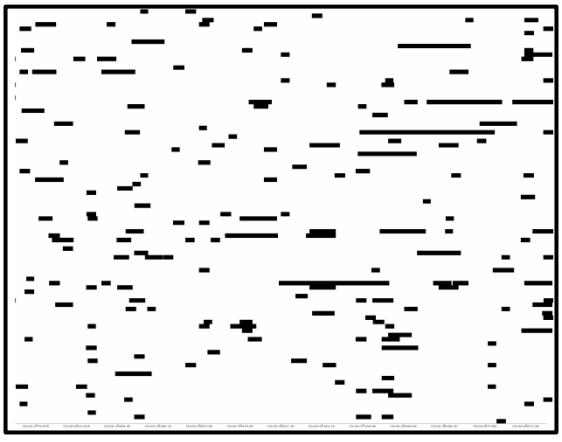
Brown Swiss



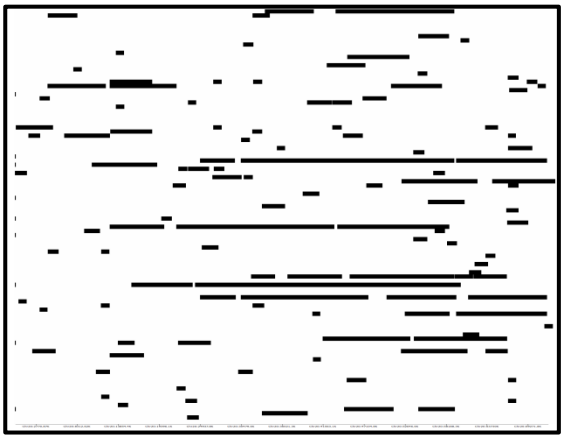
Nelore



Fleckvieh

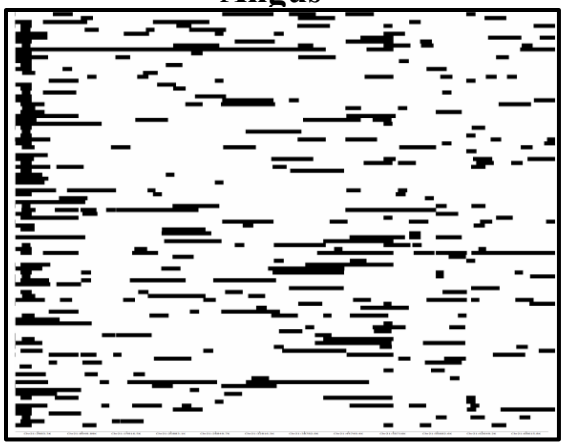


Gir

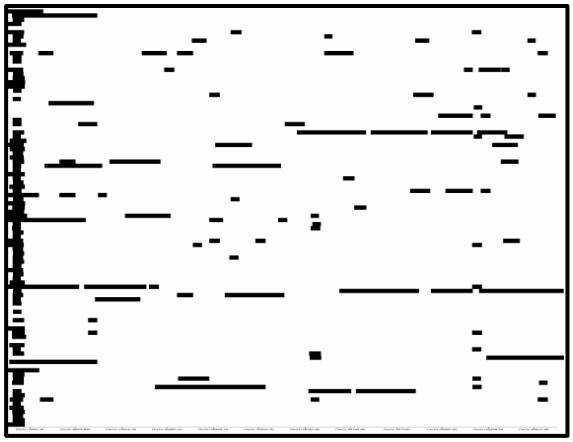


CHR 21

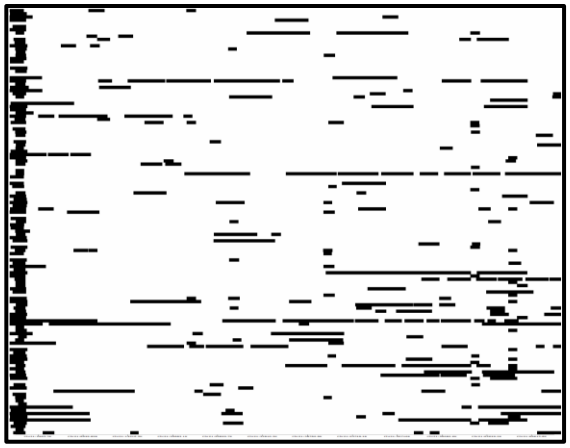
Angus



Brahman

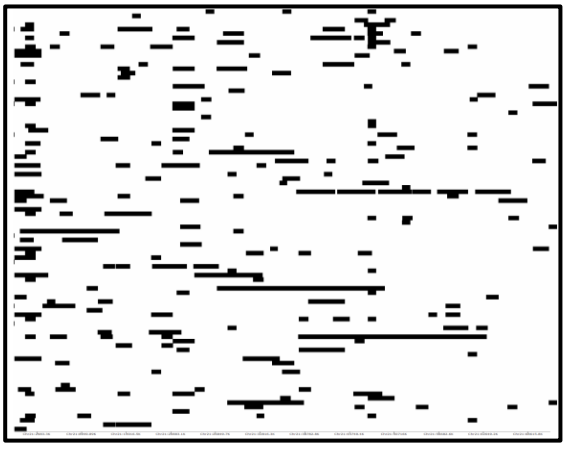


Nelore

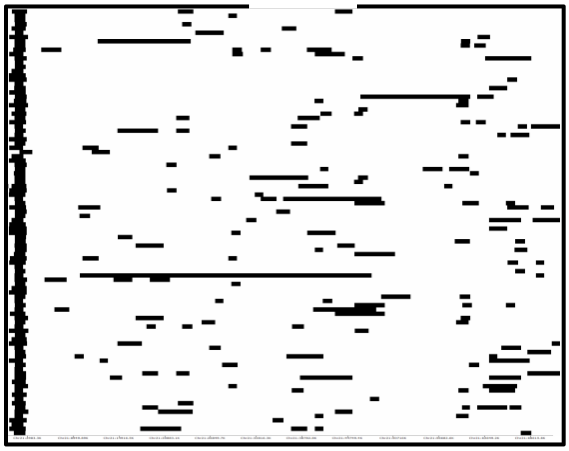


Brown Swiss

Fleckvieh

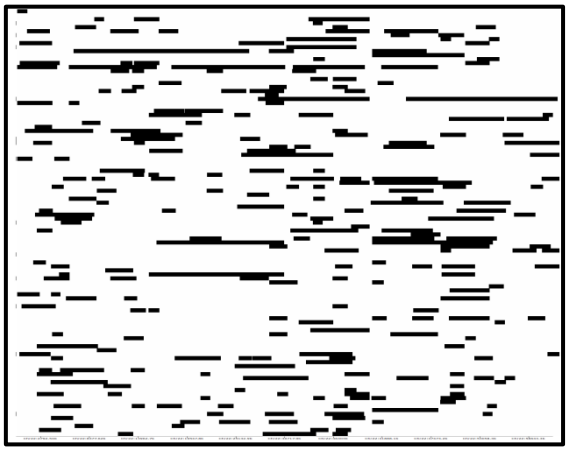


Gir

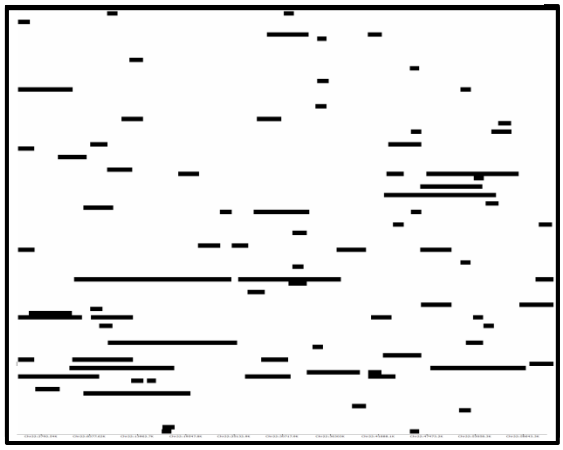


CHR 22

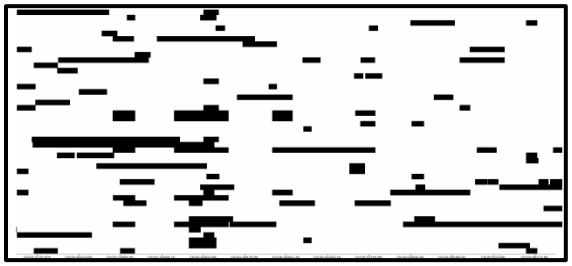
Angus



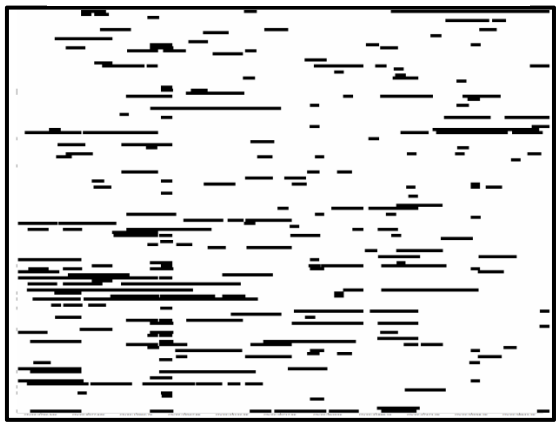
Brahman



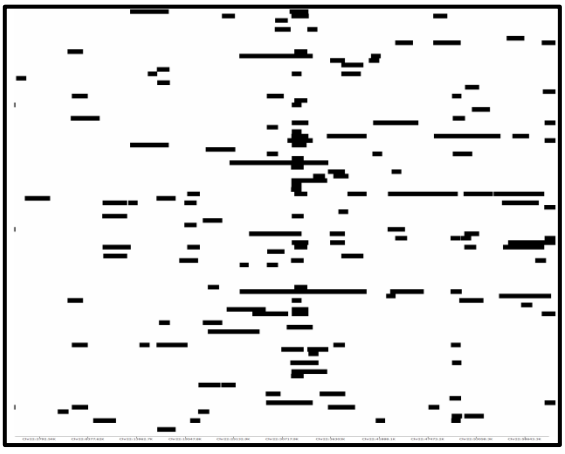
Brown Swiss



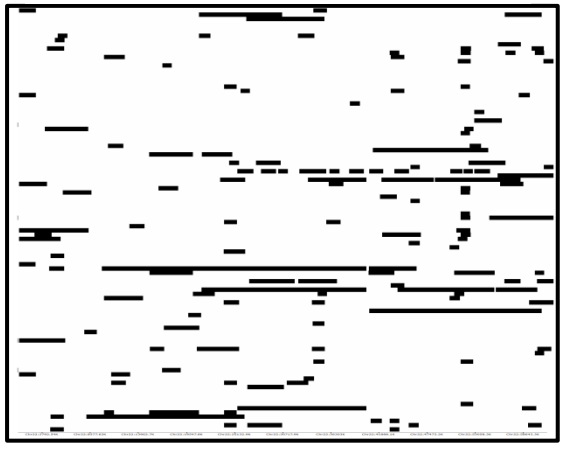
Nelore



Fleckvieh

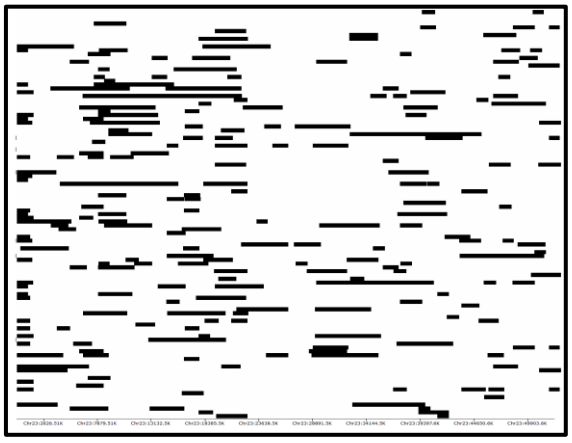


Gir

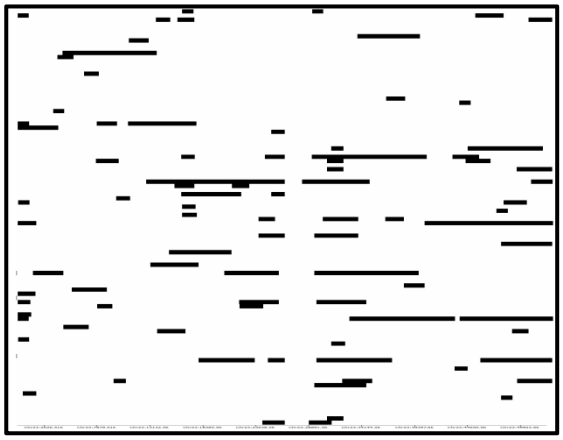


CHR 23

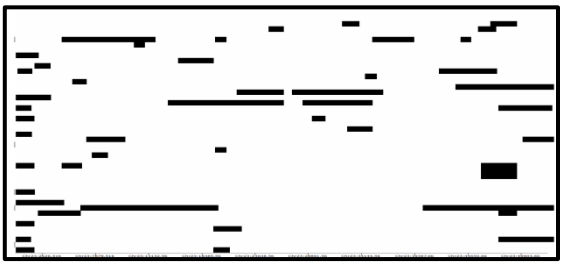
Angus



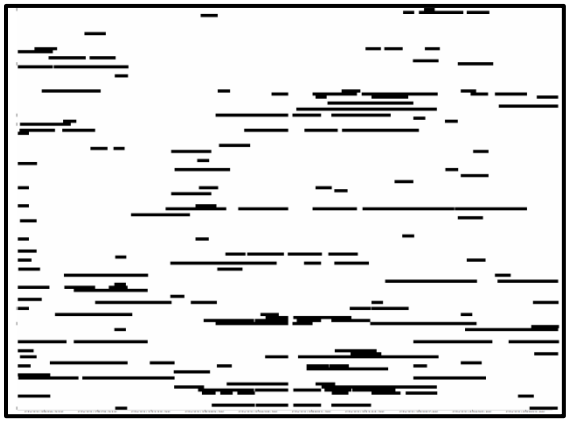
Brahman



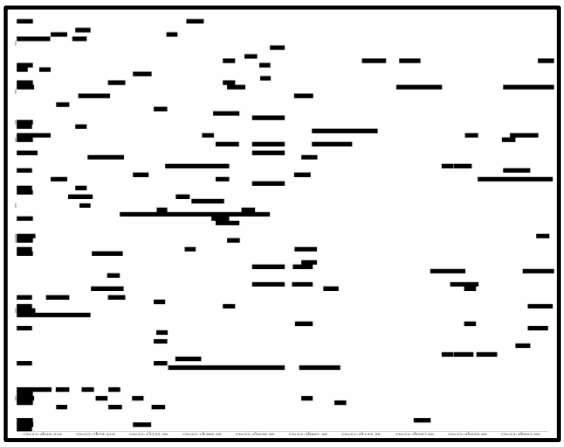
Brown Swiss



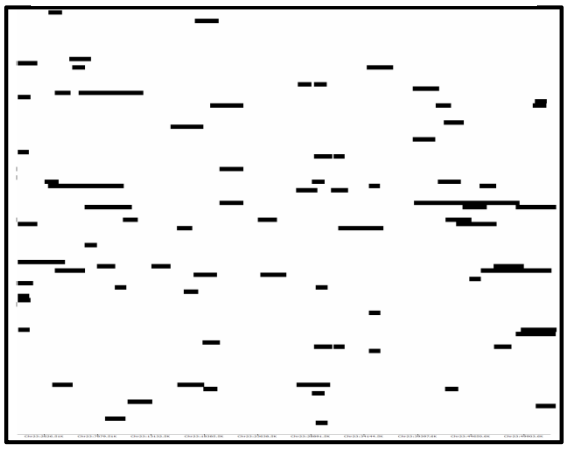
Nelore



Fleckvieh

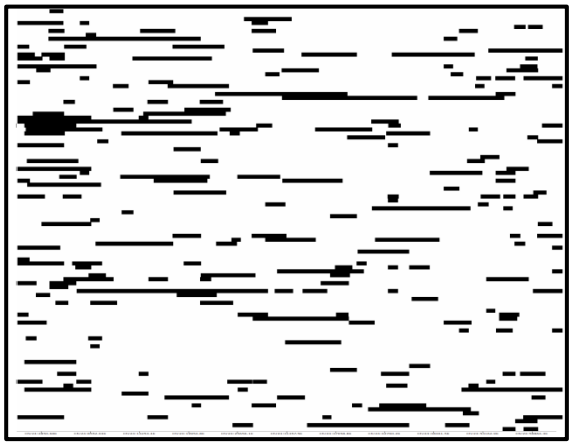


Gir

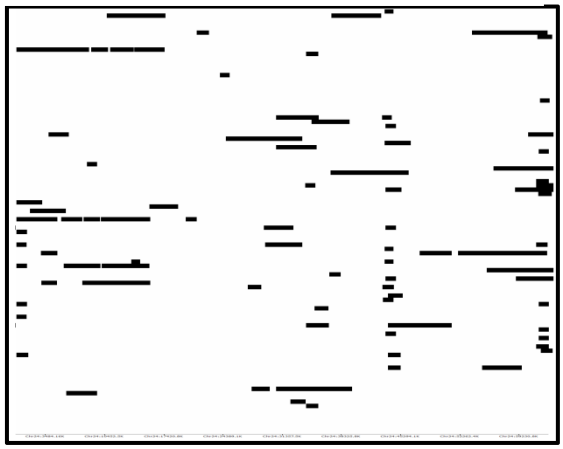


CHR 24

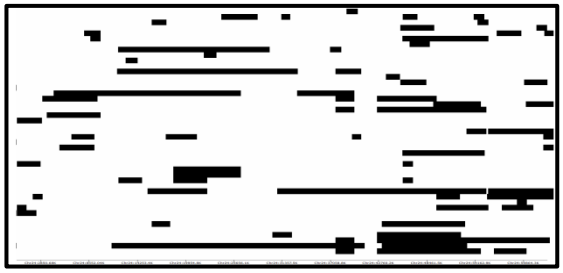
Angus



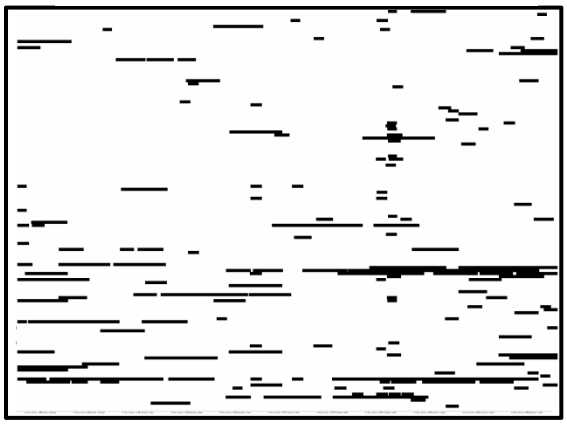
Brahman



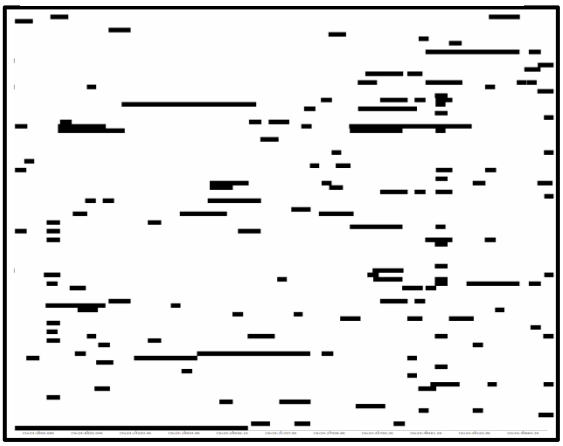
Brown Swiss



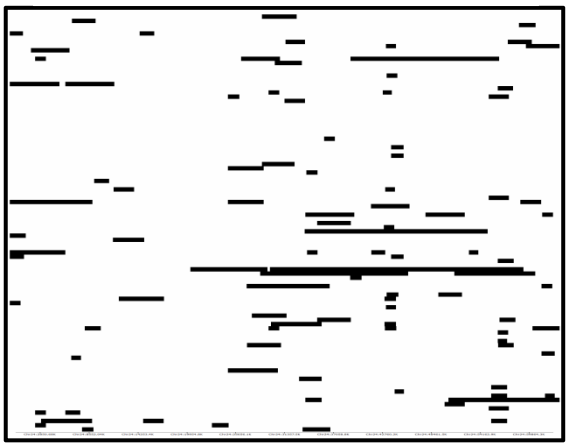
Nelore



Fleckvieh

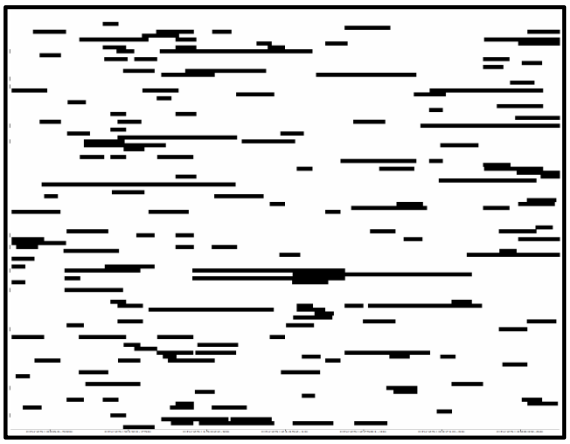


Gir

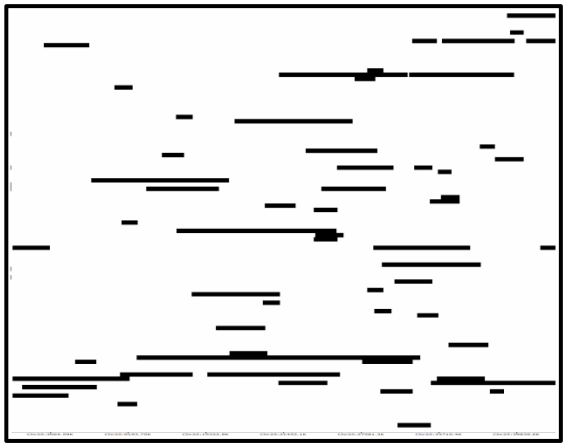


CHR25

Angus

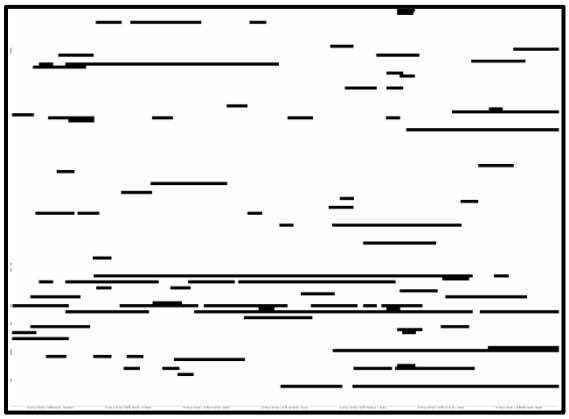


Brahman

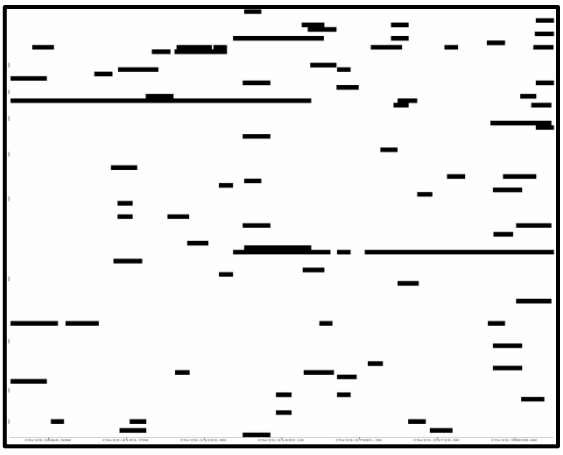


Brown Swiss

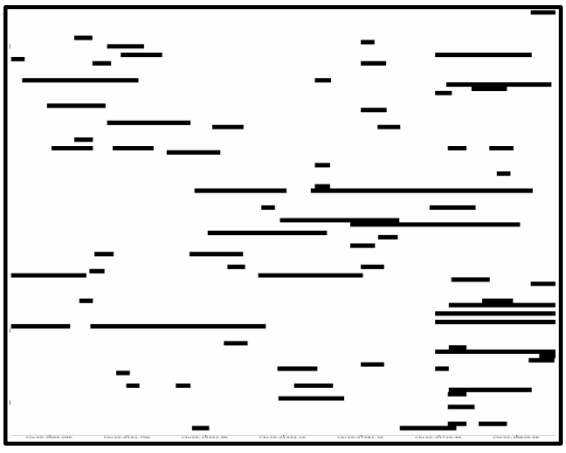
Nelore



Fleckvieh

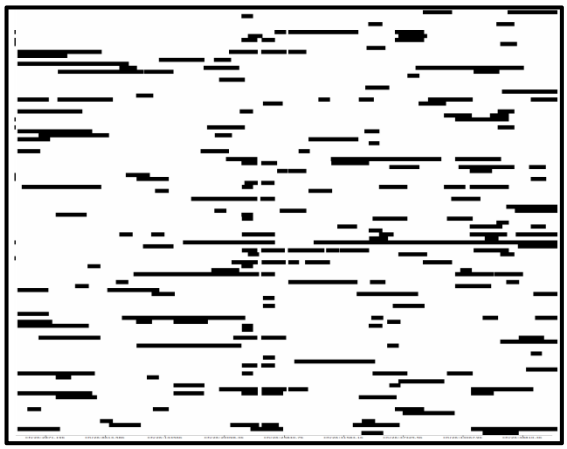


Gir

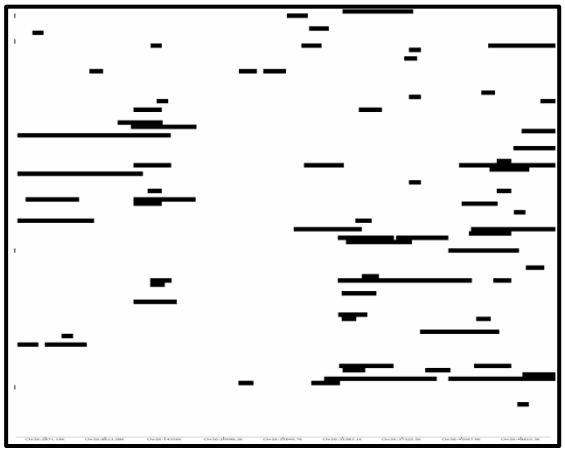


CHR 26

Angus

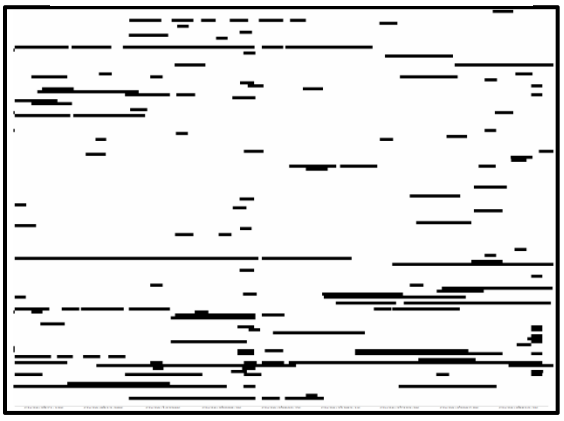


Brahman

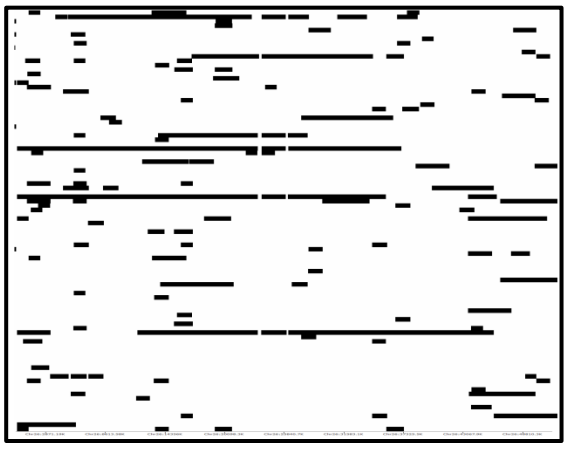


Brown Swiss

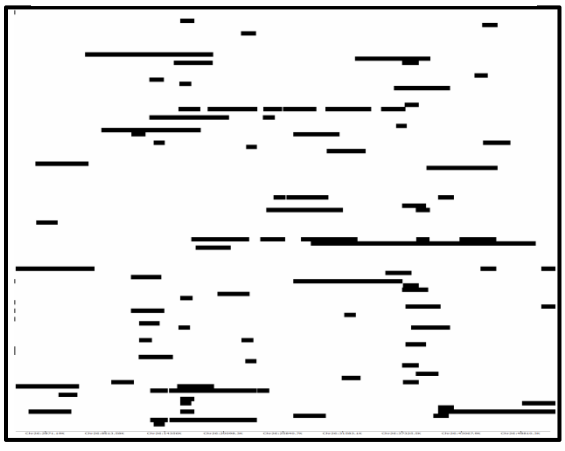
Nelore



Fleckvieh

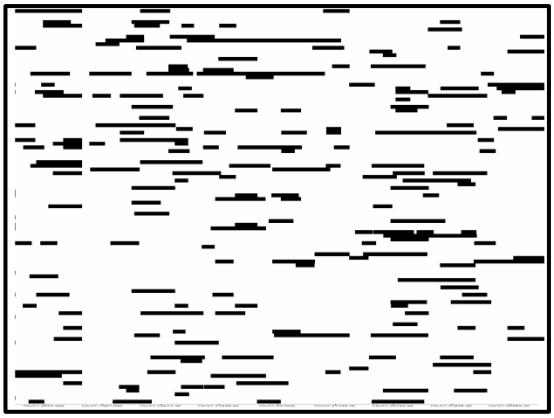


Gir

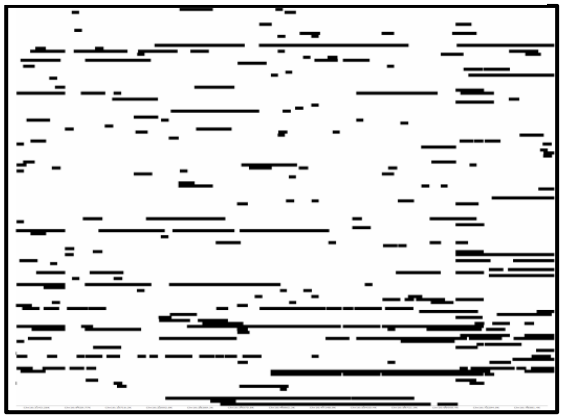


CHR 27

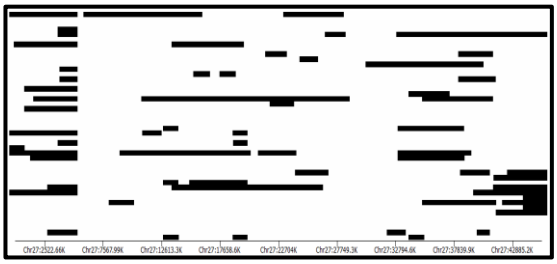
Angus



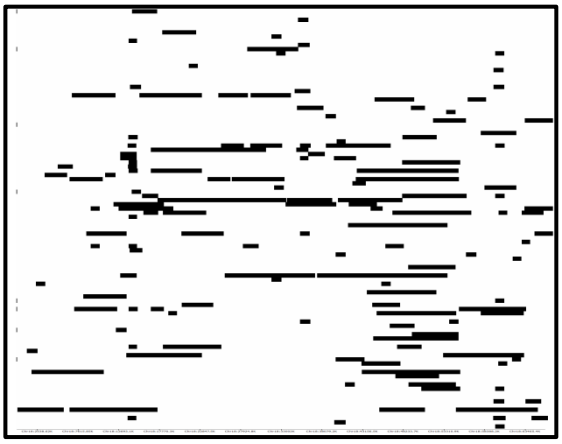
Brahman



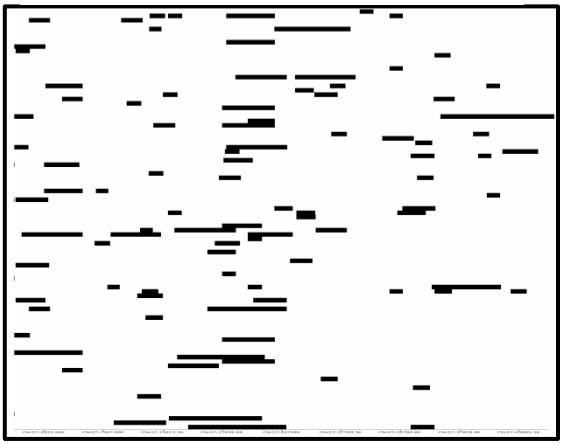
Brown Swiss



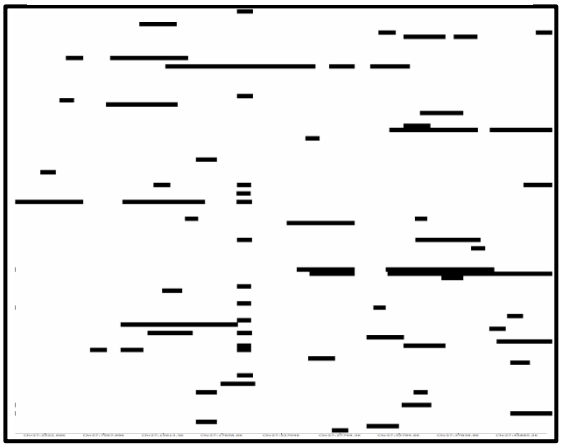
Nelore



Fleckvieh

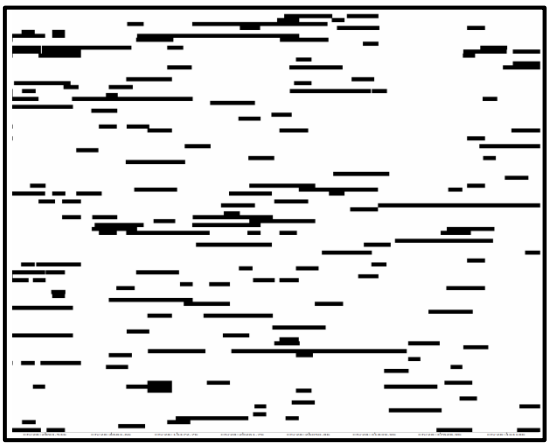


Gir

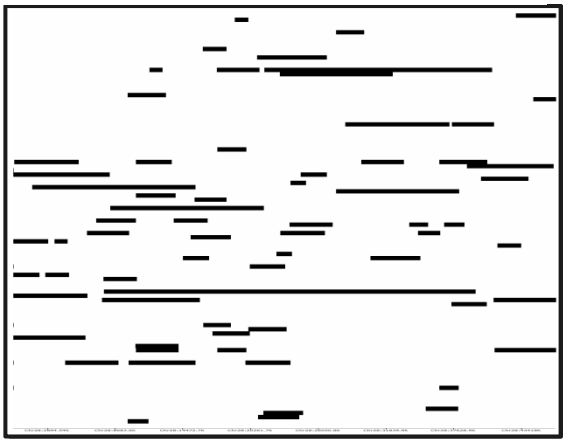


CHR 28

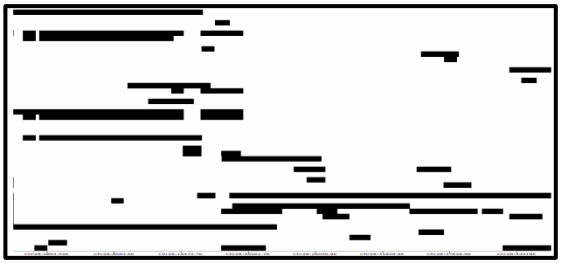
Angus



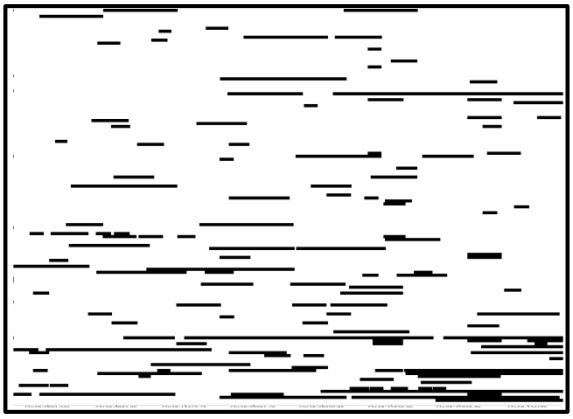
Brahman



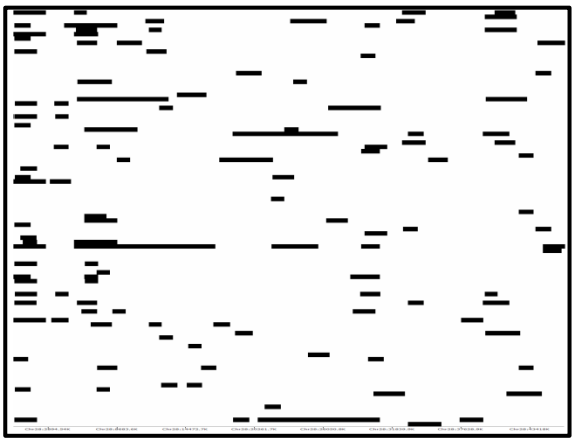
Brown Swiss



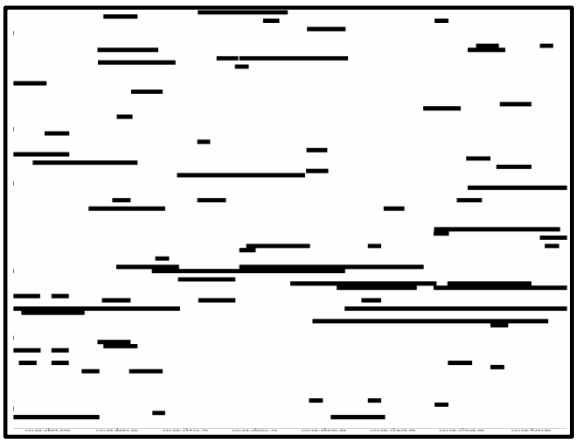
Nelore



Fleckvieh

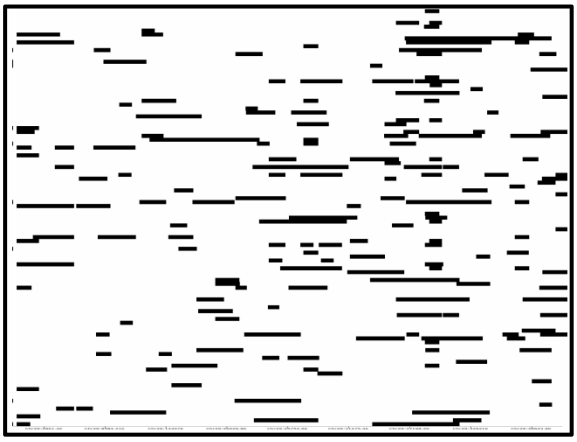


Gir

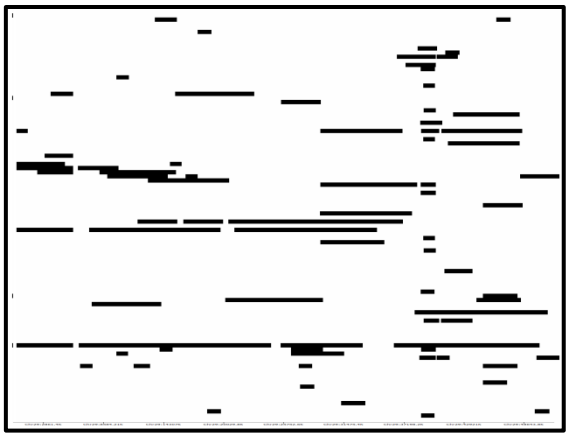


CHR 29

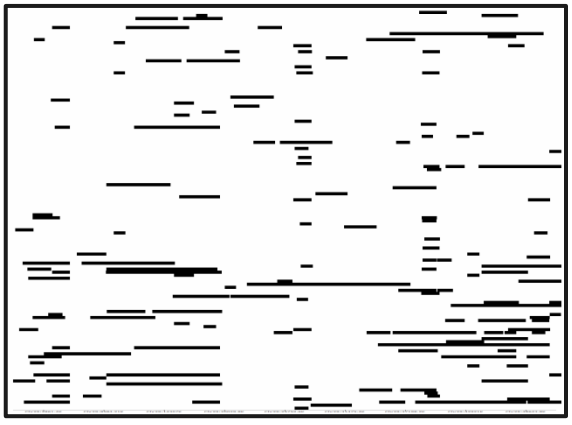
Angus



Brahman

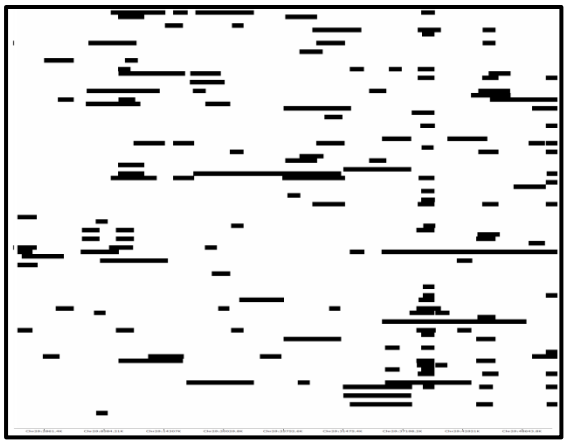


Nelore



Brown Swiss

Fleckvieh



Gir