UNIVERSITÀ DEGLI STUDI DI SASSARI

SCUOLA DI DOTTORATO DI RICERCA
Scienze e Biotecnologie
dei Sistemi Agrari e Forestali
e delle Produzioni Alimentari

Indirizzo Scienze e Tecnologie Zootecniche

Ciclo XXVI

APPROACHES FOR INVESTIGATING GENOME VARIABILITY IN CATTLE

dr. Gabriele Marras

Direttore della Scuola	prof. Alba Pusino
Referente di Indirizzo	prof. Nicolò P.P. Macciotta
Docente Guida	prof. Nicolò P.P. Macciotta
Tutor	dott.ssa Silvia Sorbolini
	dott. Giustino Gaspa

Anno accademico 2012- 2013

UNIVERSITÀ DEGLI STUDI DI SASSARI

SCUOLA DI DOTTORATO DI RICERCA
Scienze e Biotecnologie
dei Sistemi Agrari e Forestali
e delle Produzioni Alimentari

Indirizzo Scienze e Tecnologie Zootecniche

Ciclo XXVI

La presente tesi è stata prodotta durante la frequenza del corso di dottorato in "Scienze e Biotecnologie dei Sistemi Agrari e Forestali e delle Produzioni Alimentari" dell'Università degli Studi di Sassari, a.a. 2012/2013 - XXVI ciclo, con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività l.3.1 "Finanziamento di corsi di dottorato finalizzati alla formazione di capitale umano altamente specializzato, in particolare per i settori dell'ICT, delle nanotecnologie e delle biotecnologie, dell'energia e dello sviluppo sostenibile, dell'agroalimentare e dei materiali tradizionali".

Gabriele Marras gratefully acknowledges Sardinia Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.)

Anno accademico 2012- 2013

Anno accademico 2012- 2013

La mente è come un paracadute.
Funziona solo se si apre.
Albert Einstein

Anno accademico 2012- 2013

Index

CHAPTER 1

General introduction
INTRODUCTION 2
From phenotype to genotype 5
The molecular (r)evolution 6
Encoding DNA 8
Intergenic DNA 9
Genomic Tools For Investigating Differences Between Individuals 10
OBJECTIVE OF THE THESIS 16
REFERENCES 17

CHAPTER 2

Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds
SUMMARY 26
INTRODUCTION 27
MATERIALS AND METHODS 28
RESULTS AND DISCUSSION 31
CONCLUSIONS 44
REFERENCES 46

CHAPTER 3

Detection of selection signatures in five Italian cattle breeds with different productive specialization
ABSTRACT 53
INTRODUCTION 53
MATERIAL AND METHODS 55
Experimental population and genotypic data 55
Detection of signatures of selection 56
RESULTS AND DISCUSSION 59
CONCLUSIONS 69
REFERENCES 70

CHAPTER 4

Analysis Of Distribution Of Runs Of Homozygosity And Of Their Relationship With Inbreeding In Five Cattle Breeds Farmed In Italy
ABSTRACT 76
INTRODUCTION 77
Inbreeding definition, its measure and consequences 77
A Measure of Autozigosity derived from SNP data: Run of Homozigosity 78
Usefulness of ROH in animal breeding and formulas to calculate molecular inbreeding 79
OBJECTIVE OF THE WORK 80
MATERIALS AND METHODS 81
Data 81
Criteria used for ROH detection 81
Basic statistics and genomic inbreeding calculation 82

RESULTS AND DISCUSSION 83
Number of ROH 83
Average Length of ROH and Sum of ROH 87
Length of ROH vs Number of ROH 91
Inbreeding measure based on ROH 93
CONCLUSION 100
REFERENCES 102

CHAPTER 5

General Conclusions
GENERAL CONCLUSIONS 106

SUPPLEMENTAL MATERIAL

"Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds" 110

"Detection of selection signatures in five Italian cattle breeds with different productive specialization" 140

CHAPTER 1

General introduction

INTRODUCTION

Evolution is a continuous and progressive process of small mutations whose accumulation results in a huge variability among individuals. These mutations can be both favorable and unfavorable and represents the basis of the biological ability to adapt to different environments. In the human genome it has been estimated that approximately 70-75% of the mutations affecting a change in amino acid undergo a negative selection. This is because strongly deleterious mutations are eliminated immediately, or within a few generations. In contrast, a mutation can bring a positive benefit to the organism and through a selection artificial or natural, is also maintained in successive generations (Nielsen et al., 2007).

Domestication is considered one of the first event responsible for the phenotypic changes. In this process, which lasted thousands of years, man has selected plants and animals to be used as a support for his own growth. The first domesticated animal was the dog around 15,000 years ago. Only later, around 10,000 years ago, were selected cattle, sheep, goats and pigs as a source of fresh food (FAO, 2007; Toro and Màki-Tanila, 2007; Ajmone-Marsan et al., 2010). The main purposes of animal domestication have been the reduction of aggressiveness, the increase of the degree of adaptation, and the enhancement of some production characteristics (Toro and Màki-Tanila, 2007).

The combination of artificial selection and evolutionary processes has led the appearance of new breeds (Ross-Ibarra et al., 2007). In the world there are more than 4,000 breeds of farm animal species (Table 1), including local and international breeds. More of 50% of local breeds are present in Europe (about 1,500 breeds) and Asia (about 1,200 breeds). Sheep and cattle are among the most represented livestock species in the world.

Cattle is an interesting model for studying the molecular basis of genetic variability between individuals, due to the intense action of genetic improvement operated by man. Through the combination of natural and artificial selection, there has been a gradual and continuous change in the phenotype that led to the about 1,000 local breeds currently distributed worldwide (FAO, 2007). The modern species of the genus *Bos* are the *Bos Taurus* and the *Bos indicus*. They are both descendant of the *Bos primigenius* (Bradley et al., 1996). The divergence between the two species has been estimated between 250,000 and 850,000 years ago and their domestication occurred separately in near East and India, respectively (Bolormaa et al., 2013).

Table 1. Number of breeds of farm animals in the world (FAO, 2007)

	Country								
Species	Africa	Asia	Europe & the Caucasus	Latin America & the Caribbean	Near & Midlle East	North America	Southwest Pacific	International Breeds	World
Buffalo	2	88	11	11	8	0	2	5	127
Cattle	154	239	277	129	43	29	26	112	1009
Goat	86	182	170	26	34	3	11	40	552
Sheep	109	265	458	47	50	31	35	100	1095
Pig	49	229	165	67	1	18	12	33	574
Ass	17	39	40	21	16	4	3	6	146
Horse	36	141	269	65	14	23	22	66	636
Rabbit	11	16	125	14	5	0	0	23	194
Total	464	1199	1515	380	171	108	111	385	4333

The genetic variability of a population results from the action of demographic (drift, migrations, mating), genetic forces (mutations, recombination and selection) (Cavalli-Sforza and Feldman, 2003) or of random processes (Windig and Engelsma, 2010). The modification of phenotypes caused by selection involved not only morphological features but, as expected, also traits of economic interest. In particular, in the last decades artificial selection has resulted in the development of specialized cattle breeds (dairy, beef) that represent the largest part of the cattle stock in many developed countries. The other main category is represented by local breeds, i.e. those that had been lees subjected to selection and whose main feature is represented by a great adaptability to specific environment.

If it is widely accepted that selection has been one of the major driving forces for breed differentiation, on the other hand it has to be remembered that the high selection pressure often resulted in the elimination of less favorable alleles thus reducing the genetic variability within breeds. Generally, the effects of artificial selection on the genome tend to drastically reduce the levels of heterozygosity (Andersson, 2012). For positive mutations, selected alleles tend to become more frequent in the population. (Nielsen, 2005; Scheinfeldt and Tishkoff, 2013). This selection does not only affect a particular allele, but can also involve neighbouring neutral sites. This phenomenon is called hitch-hiking (Smith and Haigh, 1974). In a structured population, the fixation of a beneficial allele distorts patterns of variation also at linked sites leaving distinct signatures from that expected under the standard sweep model (Przeworski et al., 2005). To better understand the genetic basis of phenotypic variation induced by directional selection the pattern of heterozygosity can be studied in domestic species (Wiener and Wilkinson, 2011).

In cattle, the effective population size of highly selected breeds experienced a further huge reduction due to the widespread use of artificial insemination (Hayes et al., 2009). On the contrary, the poor or absence of selection for local breeds has guaranteed the maintenance of a large allelic richness. The recent development of studies on biodiversity of cattle breeds, due to the action of several international research projects (Elsik et al., 2009), has highlighted the role of these populations as reservoirs of allelic diversity. The maintenance of genetic diversity of local breeds represents a key point for their contribution to current or future traits of interest and for their potential role in future breeding options (Groeneveld et al., 2010).

As a consequence of the different evolution history of the various breeds, two main sources of genetic variability can be considered: between and within breeds, respectively. These two have different meaning and different practical relevance. The study of the genetic basis of differences between two breeds with different production aptitude (for example beef and dairy) may be of great help for identifying genomic regions involved in the genetic determinism of the different categories of traits. On the other hand, genetic differences between animals of the same breed represent the source of variation that is exploited by purebred selection programmers.

The variability between individuals has been investigated for long time. In the specific case of livestock, the understanding of the genetic architecture of complex traits remains one of the main goals for animal breeders and evolutionary geneticists (Stranger et al., 2011). Describing the genetic architecture of a phenotype means to estimate the heritability, the number and distribution of loci underlying phenotypic variation, as well as to highlight possible interaction phenomena such as dominance, epistasis or pleiotropy (Stranger et al., 2011). The unravelling of the genomic architecture of a polygenic trait is fundamental to study variability within and between population.

The tools changed according to achievements in molecular and quantitative genetics, nanotechnology, computer science, statistics and bioinformatics. Such a dynamic situation has resulted in a progressive increase of the power of investigation and in a subsequent continuous improvement of the knowledge on the molecular basis of genetic variability. Changes as the switch from phenotypic to molecular markers or the outbreak of molecular genetics had a great impact not only on the knowledge about the organism investigated but also on the upgrade of theoretical models used to explain the genetics of the different traits.

From phenotype to genotype

First characteristics of animals used to measure diversity were the external phenotypes as, for example size, morphology, coat colour, polledness, double muscling and some production traits (milk yield, growth, wool). These traits are characterized by different level of genetic complexity: single gene for "double muscle" (Kambadour et al., 1996); few genes with large effect but complicated relationships between loci for coat colour (Capitan et al., 2009) and for polledness (Seichter et al., 2012); many loci with small effects as in the case of type traits (Hayes et al., 2010); several loci with both large and small effects for many yield traits. In case of single genes, the identification of underlying genetic structure from the phenotype variability was complicated by the existence of dominance (as in the case of polledness) or exceptions of Mendelian inheritance such as epistasis (as the case of coat colour). Identification of genotypes was also based on the use of pedigree information. These external phenotypes have been fundamental for the definition of the standard characteristics of the various breeds.

A particular case of measure of genetic difference is the evaluation of the genetic merit of candidates to selection. This measure is essential in breeding programs, where it is used to rank individuals in order to plan matings aimed at obtaining a genetic improvement in the offspring. The genetic merit is estimated through the breeding value (EBV), which is obtained by combining phenotypes and pedigree information using a suitable statistical methodology as for example BLUP (Henderson, 1975). EBV have been conceived according to the infinitesimal model theory of Fisher, that assumes that observed phenotypes are determined by an infinite number of loci, each with an infinitesimal additive effect plus the deviation due to environmental effect (Fisher, 1918; Falconer and Mackay, 1996). The use of EBV to select best animals has resulted in a great enhancement of performances in specialised breeds: for example, the average 305d milk production per cow for the Italian Holstein shifted from less than kg/year 7000 in 1985 up to kg/year 9000 of milk in 2012 (www.anafi.it). For beef traits, the average daily gain of young bulls during performance test increased from kg/d 1.51 and kg/d 1.61 in 1990 to kg/d 1.82 and kg/d 1.92 in 2011 for Marchigiana and Chianina breeds, respectively (ANABIC, 2012).

Since the middle of the last century, however, the availability of new laboratory techniques made feasible the use of biochemical markers. The widespread use of electrophoresis allowed for detecting the genotype of individuals for many loci coding for proteins expressed in the blood or milk, or in many organs and tissues. Biochemical markers have represented a powerful tool for studying genetic diversity. In livestock they have been used

for many purposes as the study of genetic distances between populations, the identification of individuals, the parentage testing (as the blood groups), the identification of monozygotic twins (Masina et al., 1989; Katanen et al., 1999). Many researches have been carried out on relationships between polymorphisms and traits of economic interest: examples are studies on milk protein genes in cow and goat (Aschaffenburg, 1966; Aleandri et al., 1990; Ng-Kway-Hang, 1997; Kantanen et al., 2000), the *phosphohexose isomerase* (PHI) for meat quality in pigs (Davies et al., 1988).

Main limitations of biochemical markers were the relative low degree of polymorphism (i.e. limited number of alleles) and the need for phenotypic expression, as the case of milk protein alleles (expressed only in lactating females). A main consequence of the dependence from phenotypes was that this approach was able to detect genetic variation only in the expressed portion of the genome. But, also when the use of biochemical markers was widespread, the disproportion between the estimated number of genes (around a hundred thousand) and the size of the whole genome ($3*10^9$ bp in mammals) was evident. These issues were addressed by the advent of molecular genetics, which allowed the study of the basis of genetic variability at DNA level.

The molecular (r)evolution

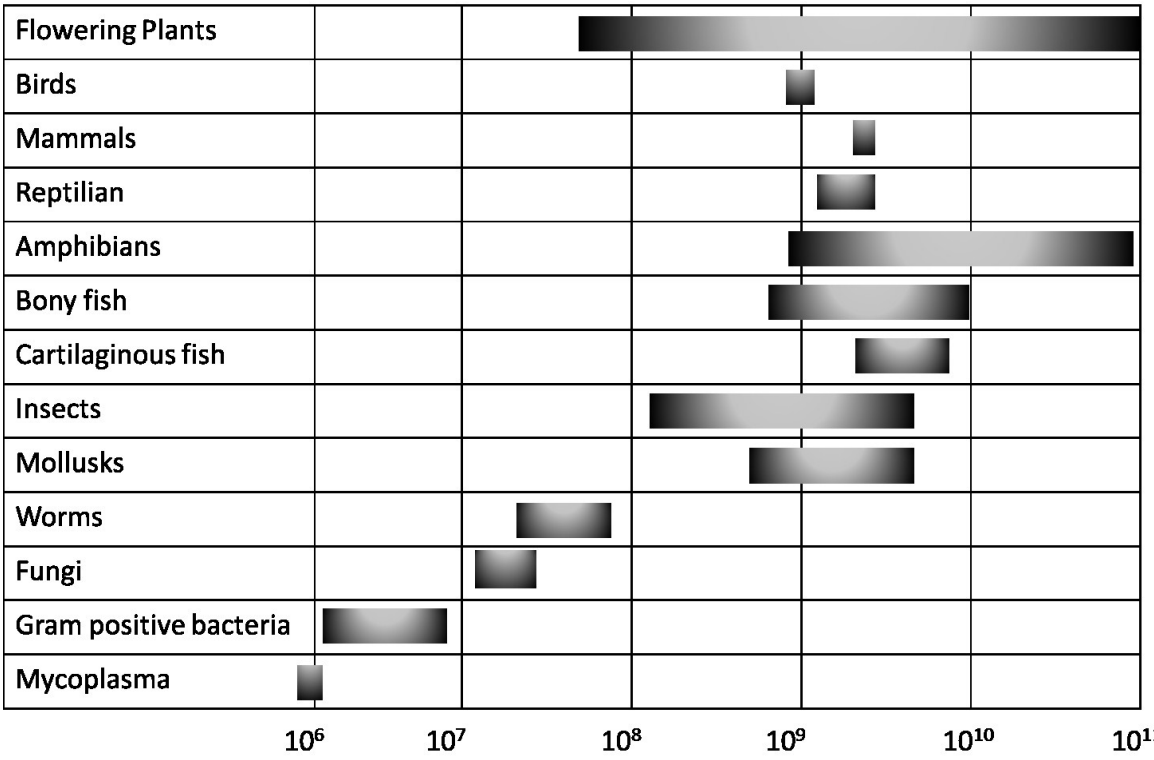
The development of molecular genetics technologies that has occurred during the last five decades and the integration with tools of quantitative genetics has had a tremendous impact on the study of genetic variability of farm animals (Przeworski et al., 2005). Whole genome sequencing and mapping allowed the quantification of differences between genomes of various species, the development of new categories of powerful genetic markers, the deciphering of the genetic architecture of traits of importance, and the study of structural element that contribute to the genetic differences between individuals (Windig and Engelsma, 2010).

First sequencing analyses of living organisms were performed on bacteria (*Haemophilus influenzae*), yeasts *(Saccharomyces cerevisiae*) and insects (*Drosophila melanogaster*). Only in 2001 it was possible to sequence the entire human genome (Lander et al., 2001). Afterwards, other organisms such as plants (i.e. *Arabidopsis thaliana*) and animals (i.e. cattle, dog, mice) have been sequenced (Griffiths, 2013).

Table 2 reports the comparison between the size of the genome (in number of base pairs) for different living organisms. A huge variation could be clearly observed.

Sequencing has allowed not only the comparison between sizes of the genomes of different species, but also the detection of differences in their structural organization. By using molecular genetics techniques, scientists have realised that phenotypic variability was not only due to differences in the expressed sequences (genes), but could also reside in portions of the genome that were not directly expressed.

Table 2 Genome size in different living organisms (Lewin, 2004).



The organization of the genomes of eukaryotic cells is characterized by two main types of DNA: the encoding DNA, which comprises all the genes; and the intergenic DNA, that is mainly represented by transposons and retrotransposons (Lander et al., 2001).

The encoding DNA, represents a small portion of the genomes (about 3% in the human genome) (Kass and Batzer, 2001), whereas most part is represented by intergenic DNA, also called junk DNA, which is divided into unique DNA and repeated sequences, respectively.

Thanks to the analysis of genomic sequences, it is now possible to compare different organisms. For example, it was possible to determine that 5% of the human genome has been strongly conserved during evolution (Strachan and Read, 2011). Moreover, among human populations, the individual genetic diversity was quantified in ~0.1% base pairs.

This apparent little number represents however differences among 3.2 billions base pairs. The variability among genomes resides mainly in the intergenic DNA, which is characterized by repeated sequences, and to a lesser degree, in regions of the encoding DNA.

Encoding DNA

The part of DNA which consists of genes is highly conserved across species, i.e. sequences are similar in structure and position. The first studies to identify genes were made on mutant organisms of *Drosophila melanogaster*, by exploiting the difference between phenotypes. Advanced technologies have made also possible to identify genes in which mutations do not generate phenotypic changes (Lewin, 2004; Griffiths, 2013). This has speeded up the process of identifying and cataloging genes. To date, in humans have been identified about 21,000 genes that represent 1.5% of the total genome (http://www.ensembl.org/index.html).

The studies of genetic divergence has given the opportunity to identify genes that have the same biological function in different species. These genes, called homologues can be divided into two categories: the orthologous, which are derived from a common ancestor; and the paralogs, which are represented by genes that have undergone a process of duplication and independent evolution in the same genome (Griffiths, 2013). In mouse, about 80% of the 25,000 annotated genes have been identified as orthologous with the genes present in humans, although a phylogenetic separation occurred 75 million years ago (Guénet, 2005).

It has been estimated that there are about 26,000 genes in the cattle genome. A comparison with genes of other six different species (dog, human, mouse, rat, opossum, platypus), showed that about 14,000 genes are orthologous. (Elsik et al., 2009). Another study conducted on a set of genes of cattle, human, dog and mouse, showed that about 3,800 bovine genes are not orthologs between the species being compared. (Mazza et al., 2009).

A very interesting aspect that emerged from bioinformatics studies, is the presence of non-coding genes defined pseudogenes. These sequences are numerous in the genome, and can arise for accumulation of nonsense mutations or by the destruction of the gene function (Griffiths, 2013).

Intergenic DNA

The intergenic DNA is the represented by all those sequences which do not apparently exhibit a biological function. These regions have been previously defined "junk DNA" or "non-coding sequences". They are characterized by a high genetic variability between individuals 60-70%) (Kass and Batzer, 2001).

The intergenic DNA can be divided into two categories: unique DNA, which consists of regulatory sequences, which modulate gene expression (promoters and enhancers); and the repeated DNA, which consist of repeated sequences. These sequences in turn are divided into moderately and highly repeated sequences, respectively (Lewin, 2004).

The unique DNA contain promoters that are located upstream of the gene and are divided into three highly conserved sequences (TATA box, CAAT box, GC box). The enhancers are located distant from the gene in position upstream or downstream and play the role of promoting the frequency of transcription (Lewin, 2004).

The repeated sequences are mainly represented by transposons, genomic elements that move within the genome with a mechanism to cut-and-past. Within the genome could be found also retrotransposons, which are able to move within the genome using RNA as intermediate and represent the main mechanism by which copies of a sequence in the genome are generated. The amount of transposons in the genome can vary in different organisms. In fact, they may represent approximately 46% of the genome in bovine, 37% in mouse and 45% in humans. (Lander et al., 2001; Waterston et al., 2002; Lewin, 2004).

In addition to transposable DNA, there are other repeated sequences that occur several times in the genome. They are represented by the satellite DNA. There are three types of satellite DNA:

- Satellite DNA, consisting of sequences larger than 100 bp that may be repeated hundreds of times;
- Minisatellites, which are tandem repeats of sequences that have between 10 and 100 base pairs;
- Microsatellite repeats of short sequences (<6 bp) within the genome, also called Simple Sequence Repeats (SSRs) or Short Tandem Repeats (STRs).

The most widespread are the simple sequence repeats (SSR). These evenly dispersed markers are characterized by tandem repeats of 1-6 bp. and it is thought that their origin is due to slippage of DNA polymerase during DNA replication. Although it is not clear its role within the genome, the microsatellite DNA has been most studied especially for the

analysis of the genetic variability or for phylogenetic reconstructions. The amount and the kind of SSR present in the genome of different species is variable. For example, the density of these markers in the human genome can reach 3%, and the most frequent repeated sequence is the triplet AAT.

In cattle the percentage of SSR present in the genome amounts to 2.27%, and the most represented triplet is the AGC (55%), as in sheep (40%) (Elsik et al., 2009).

Microsatellites have been frequently identified in the intergenic DNA. However, studies in different organisms, report that microsatellites can localize also in the coding regions. (Moran, 1993); VanLith and VanZutphen, 1996; Edwards et al., 1998). However the amount of microsatellites present in these regions is very low and it can represent between 9-15% of the total of the SSR in the genome of vertebrates. (Chistiakov et al., 2006).

Genomic Tools For Investigating Differences Between Individuals

The study of genetic variability nowadays essentially relies on the use molecular markers. Such a definition comprises DNA sequences that are abundant throughout the genome, express a large variability between individuals, could be unambiguously detected, and are transmitted to offspring with simple Mendelian inheritance. (Woolliams and Toro, 2007).

The development of PCR (Polymerase Chain Reaction) technology made feasible the set up of several classes of powerful molecular markers such as RAPD (Random Amplification Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism), RFLP (Restriction Fragment Length Polymorphism) and SSR (Simple Sequence Repeat) (Abdel-Mawgood, 2012).

The most widely used class of molecular markers in the last decades have been the microsatellites, that belong to the SSR category. There are several reasons for their success. One is represented by their high reproducibility compared to the other markers: for example, RAPD are randomly generated and are therefore not reproducible. Another important feature of microsatellites is their high degree of allele polymorphism, whereas RFLP generally have only two alleles. Moreover they are abundant throughout the genome. Last, but not least, the analysis is easy to perform: small amounts of DNA are sufficient to obtain rapid results (Vignal et al., 2002; Abdel-Mawgood, 2012) and the process can be automatized by the use of sequencers.

Microsatellites have been used in studies of animal biodiversity, mainly to study genetic distance between breeds. Researches have been carried out on different livestock species: Cattle (MacHugh et al., 1997, MacHugh et al., 1998), Sheep (Arranz et al., 2001), Goat

(Okpeku et al., 2011), Swine (SanCristobal et al., 2006), Chicken (Hillel et al., 2003). The presence of high variability of microsatellites in the genome has prompted FAO to propose a panel of 30 microsatellites for the study of animal genetic diversity for the following species: cattle, buffalo, sheep, goat, horse, donkey, camelid, pig and chicken (FAO, 2011).

Microsatellites have been also extensively used in experiments carried out on several livestock species for QTL detection. They have been powerful tools to identify genomic regions involved in the genetic determinism of quantitative traits in both in studies where ad hoc populations were generated (i.e. F2 or back-cross) and in investigations carried out outbred populations, as dairy cattle breeds, using specific statistical approaches as the Daugther and the Granddaughter designs (Weller et al., 1990; Georges et al., 1995).

In spite of their great power, the use of microsatellites has experienced a relevant reduction in the last decade. Starting from the 2000', Single Nucleotide Polymorphisms become the genetic marker of reference in animal genetics studies on biodiversity, on association phenotype-genotype and on marker assisted selection. A main reason of the success of SNP is represented by their great abundance of SNP throughout the genome, in comparison with microsatellites.

Figure 1 shows a sequence of about 10,000 bp in the human genome, along which densities of SNP and microsatellites are compared. It can be clearly seen that SSR represent a small fraction of this sequence, whereas the SNP exhibit a higher coverage and an homogeneus distribution.

In the human genome SNPs represent about 90% of sequence variants. It has been estimated that there are about 1.42 million SNPs with an average distribution of about 1.9 per kilobase (Sachidanandam et al., 2001, Abdel-Mawgood, 2012). About 2.2 million SNPs have been discovered in the cattle genome at a density of 1 kb (Fan et al., 2010). Only a small part of these SNP is used for association studies and genetic diversity (Table 3).

A further reason for the great success of SNP as genetic markers is represented by the development of high throughput platforms (SNP arrays) able to simultaneously genotype hundred thousand SNP for each individual. Currently, there are chips with different densities for the various species (Table 3) (http://www.illumina.com/) (http://www.affymetrix.com/estore/).

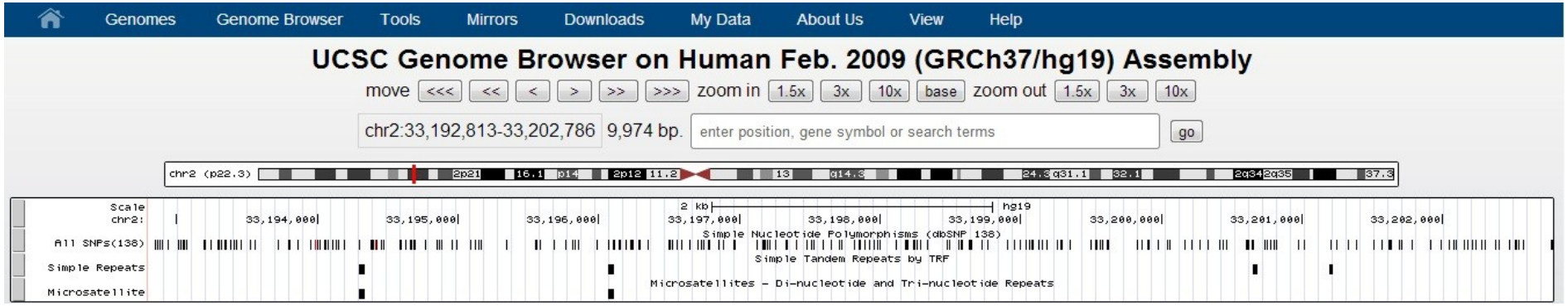


Figure 1. Genetic variability along a stretch of 10,000 bp in Human genome **(**http://genome.ucsc.edu/**)**

Table 3. Number of SNPs contained in the different chips produced by the Illumina and Affymetrix companies.

	Illumina		Affymetrix	
Species	Medium density	High Density	Medium Density	High Density
Human	730.000	1.14 million	500.000	906.000
Dog	50.000	170.000	50.000	127.000
Bovine	54.000	777.000		640.000
Porcine	60.000	510.000		
Sheep	54.000	600.000		
Horse		54.000		
Chicken		60.000		600.000

The SNP have proved to be excellent molecular markers for genetic studies. The high repeatability of the analysis and the easy interpretation of the results makes them very reliable and comparable among different laboratories (FAO, 2011). However some problems still exist, as the ascertainment bias of alleles that may occur when are genotyped individuals from different breeds of those that have been used to select SNP included in the bead-chip (Nielsen and Signorovitch, 2003).

The combined use of sequencing techniques and high density SNP genotyping has led to the discovery of structural elements of DNA sequence that may be of great importance in the identification of sources of variation between and within breeds. Examples of these structural features are copy number variations and runs of homozygosity.

The CNVs are the results of DNA deletions or duplications that lead to a change in the number of copies of a specific chromosome region. The CNVs may have a size that varies from one kilobase (kb) to several megabases (Mb). These structural variations may involve segments of DNA containing whole genes (Redon et al., 2006). The mechanisms of formation of CNV can be different, resulting in different modifications of the genome structure. The decrease or increase in the number of copies of a gene can be the results of micro-homologous recombination (Hastings et al., 2009). Currently it is estimated that the CNV cover about 12% of the human genome (Hastings et al., 2009). CNV regions have a

different lengths: 35% range between 100 bp and 1 Kb and 65% have a size between 1-10 Kb, respectively (Scherer et al., 2007) (http://projects.tcag.ca/variation/). CNV have been investigated in cattle. Kijas et al. (2011) identified 51 CNV in 10 cattle (3 Angus, 6 Brahman and one composite animal one third *Bos Taurus Indicus* e two thirds *Bos Taurus*) using a microarray-based comparative genomic hybridization (array-GCH) technique. Fadista et al. (2010), using the same technique, has identified 304 CNV summing to a total of 23 Mb on a sample of 14 Holsteins, 3 Simmental, 2 Red Danish and 1 Hereford animals. Hou et al. (2011) reports a set of about 680 regions of CNV with a total size of 139 Mb identified in a population of 539 animals using a SNP chip.

As a general definition, a run of homozygosity (ROH) is intended as a chromosomal segment where an interrupted succession of homozygotes markers occurs. There are some specific features in the definition of ROH that may vary between studies as the minimum length of a ROH, the minimum number of SNP included, the average SNP density, presence of just one heterozygote SNP (Ferencakovic et al., 2013).

The length of the ROH is connected to the degree of relationship of animals and to the inbreeding coefficient of the parents of the individual under consideration. Short ROH can be found in abundance on outbred individuals whereas long ROH are generated mainly by a recent inbreeding (Ferencakovic et al., 2013). However the length of the ROH can also be generated by a low recombination of a particular region of the genome (Kirin et al., 2010).

Runs of homozygosity have been investigated mainly in human populations. In particular, relationships between ROH and diseases (Lencz et al., 2007) and inbreeding (Woods et al., 2006) have been investigated. Studies on ROH have been also carried in livestock, mainly for estimation of inbreeding levels (Ferencakovic et al., 2012). Measures of inbreeding based on ROH could then further be used for estimation of inbreeding depression.

In cattle the estimate of ROH is a valid method for the identification of the level of inbreeding of individuals. Studies have shown that a greater length of ROH is attributable to a recent inbreeding (about 3 generation ago for ROH larger than 16 Mb) and it is a good indication of the levels of inbreeding (Ferencakovic et al., 2011, 2012). It is much more complicate to detect shorter ROH (under 4 Mb in length using 54K chipset) since most of the time these homozygous fragment are not truly Identical by descend, hence they do not represent the true level of autozigosity. Ferencakovic (2013) and Purfield (2012) demonstrate how only autozigous segment >5 Mb can be effectively detected using 54K SNP chip, for shorter true autozygous segment it is necessary to increase the SNP density.

Analysis on segments of homozygosity can be used for the study of the history of domesticated populations and to determine inbreeding coefficient when pedigree is not available or is not sufficiently depth (Purfield et. Al 2012). Purfield analyzed cattle ROH data of 42 different breeds from different continents, using a core of nearly 900 bulls of multiple breed genotyped at 800K and a group of 1,162 animals gentyped at 54K.

Furthermore, the association between phenotypes and ROH has been used to measure the inbreeding depression in US dairy Cattle. It has been shown that an increase in the values of genomic based inbreeding coefficient (FROH) lead to a lowering of production values for most of productive, reproductive traits (Bjelland et al., 2013).

OBJECTIVE OF THE THESIS

Aim of the present thesis is to study the genetic variability of cattle breeds farmed in Italy. These were selected in order to represent different origin, selection history and production aptitudes. The variation at genome level was investigated by using three approaches that allow for disentangling different aspects of the genetic difference between animal populations. One approach checked the difference between two breeds by using a modification of the F_{st} metric methodology, developed in order to make the signal pattern more readable through a local regression smoothing and to identify outlier signals based on statistical evidence. This study was focused a identifying genomic regions that harbor genes involved in the differences between dairy and beef cattle. Then a multi-breed investigation was performed based on the comparison between allele frequencies under the hypothesis of the binomial distributions. The research was mainly aimed at looking for genes shared between breeds grouped according to production aptitudes. Finally a third study was carried out to analyse the structural differences in the genome by investigating the variation of the distribution of ROH between breeds. Relationships between runs of homozygosity features and inbreeding coefficients have been investigated within and between breeds.

REFERENCES

http://jul2012.archive.ensembl.org/index.html

http://www.anafi.it/

Abdel-Mawgood, A. L. 2012. DNA Based Techniques for Studying Genetic Diversity, Genetic Diversity in Microorganisms. Genetic Diversity in Microorganisms.

Ajmone-Marsan, P., Garcia, J. F., Lenstra, J. A. & Globaldiv, C. 2010. On the Origin of Cattle: How Aurochs Became Cattle and Colonized the World. Evolutionary Anthropology, 19, 148-157.

Aleandri, R., Buttazzoni, L. G., Schneider, J. C., Caroli, A. & Davoli, R. 1990. The Effects Of Milk Protein Polymorphisms On Milk Components And Cheese-Producing Ability. Journal of Dairy Science, 73, 241-255.

Aschaffenburg, R. 1966. Modified procedure of starch gel electrophoresis for β-casein phenotyping. J. Dairy Sci. 49: 1284.

Associazione Italiana Allevatori Bovini Italiani Carne (ANABIC). 2012. Annuario.

Andersson, L. 2012. How selective sweeps in domestic animals provide new insight into biological mechanisms. Journal of Internal Medicine, 271, 1-14.

Arranz, J. J., Bayón, Y. & San Primitivo, F. 2001. Differentiation among Spanish sheep breeds using microsatellites. *Genetics Selection Evolution,* 33**,** 529-542.

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. *Journal of Dairy Science,* 96**,** 4697-4706.

Bolormaa, S., Pryce, J. E., Kemper, K. E., Hayes, B. J., Zhang, Y., Tier, B., Barendse, W., Reverter, A. & Goddard, M. E. 2013. Detection of quantitative trait loci in Bos indicus and Bos taurus cattle using genome-wide association studies. Genetics Selection Evolution, 43.

Bradley, D. G., Machugh, D. E., Cunningham, P. & Loftus, R. T. 1996. Mitochondrial diversity and the origins of African and European cattle. Proceedings of the National Academy of Sciences of the United States of America, 93, 5131-5135.

Capitan, A., Grohs, C., Gautier, M. & Eggen, A. 2009. The scurs inheritance: new insights from the French Charolais breed. Bmc Genetics, 10, 33.

Cavalli-Sforza, L. L. & Feldman, M. W. 2003. The application of molecular genetic approaches to the study of human evolution. Nature Genetics, 33, 266-275.

Cavalli-Sforza, L. L., Menozzi, A., and Piazza, A. (1994). The History and Geography of Human Genes. Princeton: Princeton University Press

Chistiakov, D. A., Hellemans, B. & Volckaert, F. A. M. 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture, 255, 1-29.

Davies, W. , Harbitz, I., Fries, R., Strazinger, G., Hauge, J. G. 1988. Porcine malignant hyperthermia carrier detection and chromosomal assignment using a linked probe. Animal Genetics 19: 203-212.

Edwards, Y. J. K., Elgar, G., Clark, M. S. & Bishop, M. J. 1998. The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish, Fugu rubripes: Perspectives in functional and comparative genomic analyses. Journal of Molecular Biology, 278, 843-854.

Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A., Abatepaulo, A. R. R., Abbey, C. A., Adelson, D. L., Aerts, J., Ahola, V., Alexander, L., Alioto, T., Almeida, I. G., Amadio, A. F., Anatriello, E., Antonarakis, S. E., Anzola, J. M., Astashyn, A., Bahadue, S. M., Baldwin, C. L., Barris, W., Baxter, R., Bell, S. N., Bennett, A. K., Bennett, G. L., Biase, F. H., Boldt, C. R., Bradley, D. G., Brinkman, F. S. L., Brinkmeyer-Langford, C. L., Brown, W. C., Brownstein, M. J., Buhay, C., Caetano, A. R., Camara, F., Carroll, J. A., Carvalho, W. A., Casey, T., Cervelatti, E. P., Chack, J., Chacko, E., Chandrabose, M. M., Chapin, J. E., Chapple, C. E., Chen, H.-C., Chen, L., Cheng, Y., Cheng, Z., Childers, C. P., Chitko-McKown, C. G., Chiu, R., Choi, J., Chrast, J., Colley, A. J., Connelley, T., Cree, A., Curry, S., Dalrymple, B., Diep Dao, M., Davis, C., de Oliveira, C. J. F., de Miranda Santos, I. K. F., de Campos, T. A., Deobald, H., Devinoy, E., Dickins, C. M., Yan, D., Dinh, H. H., De Donato, M., Donohue, K. E., Donthu, R., Dovc, P., Dugan-Rocha, S., Durbin, K. J., Eberlein, A., Edgar, R. C., Egan, A., Eggen, A., Eichler, E. E., Elhaik, E., Ellis, S. A., Elnitski, L., Ermolaeva, O., Eyras, E., Fitzsimmons, C. J., Fowler, G. R., Franzin, A. M., Fritz, K., Gabisi, R. A., Garcia, G. R., Garcia, J. F., Genini, S., Gerlach, D., German, J. B., Gilbert, J. G. R., Gill, C. A., Gladney, C. J., Glass, E. J., Goodell, J., Grant, J. R., Graur, D., et al. 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science, 324, 522-528.

Fadista, J., Thomsen, B., Holm, L.-E. & Bendixen, C. 2010. Copy number variation in the bovine genome. Bmc Genomics, 11.

Fan, B., Du, Z. Q., Gorbach, D. M. & Rothschild, M. F. 2010. Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals. Asian-Australasian Journal of Animal Sciences, 23, 833-847.

Falconer D.S., and Mackay T.F.C. Introduction to quantitative genetics. Fourth Edition. Longman, UK, 1996.

FAO. 2007. The State of the World's Animal Genetic Resources for Food and Agriculture, edited by Barbara Rischkowsky & Dafydd Pilling. Rome.

FAO. 2011. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines, N. 9.

Ferencakovic, M., Hamzic, E., Gredler, B., Curik, I. & Sölkner, J. 2011. Runs of homozygosity reveal genomewide autozygosity in the Austrian Fleckvieh cattle. Agriculturae Conspectus Scientificus, 76, 325-328.

Ferencakovic, M., Hamzic, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I. & Soelkner, J. 2012. Estimates of autozygosity derived from runs of homozygosity:

empirical evidence from selected cattle populations. Journal of Animal Breeding and Genetics, 130, 286-293.

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

Fisher R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edimburgh, 52:399-433.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X. Y., Womack, J. E. & Hoeschele, I. 1995. Mapping Quantitative Trait Loci Controlling Milk-Production In Dairy-Cattle By Exploiting Progeny Testing. Genetics, 139, 907-920.

Griffiths A.J., W. S. R., Carroll S.B. 2013. Genetica. In: Zanichelli (ed.).

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., Negrini, R., Finlay, E. K., Jianlin, H., Groeneveld, E. & Weigend, S. 2010. Genetic diversity in farm animals - A review. Animal Genetics, 41, 6-31.

Guénet, J. L. 2005. The mouse genome. Genome Research, 15, 1729-1740.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. 2009. Mechanisms of change in gene copy number. Nature Reviews Genetics, 10, 551-564.

Hayes, B. J., Chamberlain, A. J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., Sethuraman, L. & Goddard, M. E. 2009. A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. Animal Genetics, 40, 176-184.

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J. & Goddard, M. E. 2010. Genetic architecture of complex traits and accuracy of genomic Prediction: Coat colour, Milk-fat percentage, and type in holstein cattle as contrasting model traits. Plos Genetics, 6.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics, 31, 423-447.

Hillel, J., Groenen, M. A. M., Tixier-Boichard, M., Korol, A. B., David, L., Kirzhner, V. M., Burke, T., Barre-Dirie, A., Crooijmans, R. P. M. A., Elo, K., Feldman, M. W., Freidlin, P. J., Mäki-Tanila, A., Oortwijn, M., Thomson, P., Vignal, A., Wimmers, K. & Weigend, S. 2003. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genetics Selection Evolution,* 35**,** 533-557

Hou, Y., Liu, G. E., Bickhart, D. M., Cardone, M. F., Wang, K., Kim, E. S., Matukumalli, L. K., Ventura, M., Song, J., VanRaden, P. M., Sonstegard, T. S. & Van Tassell, C. P. 2011. Genomic characteristics of cattle copy number variations. Bmc Genomics, 12, 127.

Kambadur, R., Sharma, M., Smith, T. P. L. & Bass, J. J. 1997. Mutations in myostatin (GDF8) in double-muscled Belgian blue and Piedmontese cattle. Genome Research, 7, 910-916.

Kantanen, J., Olsaker, I., Adalsteinsson, S., Sandberg, K., Eythorsdottir, E., Pirhonen, K. & Holm, L. E. 1999. Temporal changes in genetic variation of North European cattle breeds. Animal Genetics, 30, 16-27.

Kantanen, J., Olsaker, I., Brusgaard, K., Eythorsdottir, E., Holm, L. E., Lien, S., Danell, B. & Adalsteinsson, S. 2000. Frequencies of genes for coat colour and horns in Nordic cattle breeds. Genetics Selection Evolution, 32, 561-576.

Kass, D. H. & Batzer, M. A. 2001. Genome Organization: Human. eLS. John Wiley & Sons, Ltd.

Kijas, J. W., Barendse, W., Barris, W., Harrison, B., McCulloch, R., McWilliam, S. & Whan, V. 2011. Analysis of copy number variants in the cattle genome. Gene, 482, 73-77.

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M. & Wilson, J. F. 2010. Genomic Runs of Homozygosity Record Population History and Consanguinity. Plos One, 5.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al., 2001. Initial sequencing and analysis of the human genome. Nature, 409, 860-921.

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., Kucherlapati, R. & Malhotra, A. K. 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proceedings of the National Academy of Sciences of the United States of America, 104, 19942-19947.

Lewin, B. 2004. Il Gene VIII, Bologna.

MacHugh, D. E., Loftus, R. T., Cunningham, P. & Bradley, D. G. 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. Animal Genetics, 29, 333-340.

MacHugh, D. E., Shriver, M. D., Loftus, R. T., Cunningham, P. & Bradley, D. G. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and Zebu cattle (Bos taurus and Bos indicus). Genetics, 146, 1071-1086.

Masina P., Rando A. & P., D. G. 1989. Polimorfismo e geni a effetto maggiore. In: Reda (ed.) Miglioramento genetico degli animali domestici. Roma.

Mazza, R., Strozzi, F., Caprera, A., Ajmone-Marsan, P. & Williams, J. L. 2009. The other side of comparative genomics: Genes with no orthologs between the cow and other mammalian species. Bmc Genomics, 10.

Moran, C. 1993. Microsatellite Repeats In Pig (*Sus-Domestica*) And Chicken (*Gallus-Domesticus*) Genomes. Journal of Heredity, 84, 274-280.

Ng-Kway-Hang, K.F. 1997. A review of the relationship between milk protein polymorphism and milk composition/milk production. Proceedings of the IDF seminar on Milk Protein Polymorphism, Palmerston North, New Zealand: 22-37.

Nielsen, R. 2005. Molecular signatures of natural selection. Annual Review of Genetics.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. & Clark, A. G. 2007. Recent and ongoing selection in the human genome. Nature Reviews Genetics, 8, 857-868.

Nielsen, R. & Signorovitch, J. 2003. Correcting for ascertainment biases when analyzing SNP data: Applications to the estimation of linkage disequilibrium. Theoretical Population Biology, 63, 245-255.

Przeworski, M., Coop, G. & Wall, J. D. 2005. The signature of positive selection on standing genetic variation. Evolution, 59, 2312-2323.

Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. 2012. Runs of homozygosity and population history in cattle. *Bmc Genetics,* 13.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G. & Simianer, H. 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. Bmc Genomics, 12.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W. & Hurles, M. E. 2006. Global variation in copy number in the human genome. Nature, 444, 444-54.

Ross-Ibarra, J., Morrell, P. L. & Gaut, B. S. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proceedings of the National Academy of Sciences of the United States of America, 104, 8641-8648.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z. M., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-

Thomann, N. S., Zody, M. C., Linton, L., Lander, E. S., Altshuler, D. & Int, S. N. P. M. W. G. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 409, 928-933.

SanCristobal, M., Chevalet, C., Haley, C. S., Joosten, R., Rattink, A. P., Harlizius, B., Groenen, M. A. M., Amigues, Y., Boscher, M. Y., Russell, G., Law, A., Davoli, R., Russo, V., Désautés, C., Alderson, L., Fimland, E., Bagga, M., Delgado, J. V., Vega-Pla, J. L., Martinez, A. M., Ramos, M., Glodek, P., Meyer, J. N., Gandini, G. C., Matassino, D., Plastow, G. S., Siggens, K. W., Laval, G., Archibald, A. L., Milan, D., Hammond, K. & Cardellino, R. 2006. Genetic diversity within and between European pig breeds using microsatellite markers. *Animal Genetics,* 37**,** 189-198.

Scheinfeldt, L. B. & Tishkoff, S. A. 2013. Recent human adaptation: Genomic approaches, interpretation and insights. Nature Reviews Genetics, 14, 692-702.

Scherer, S. W., Lee, C., Birney, E., Altshuler, D. M., Eichler, E. E., Carter, N. P., Hurles, M. E. & Feuk, L. 2007. Challenges and standards in integrating surveys of structural variation. Nature Genetics, 39, S7-S15.

Seichter, D. , Russ, I., Rothammer, S., Eder, J., Forster, M., and Medugorac, I. 2012. SNP-based association mapping of the polled gene in divergent cattle breeds. Animal Genetics, 43, 595–598

Smith, J. M. & Haigh, J. 1974. The hitch-hiking effect of a favourable gene. Genetical research, 23, 23-35.

Strachan, T. & Read, A. P. 2011. Organization of the Human Genome. Human Molecular Genetics 4. Garland Science/Taylor & Francis Group.

Stranger, B. E., Stahl, E. A. & Raj, T. 2011. Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. Genetics, 187, 367-383.

Toro M.A., Màki-Tanila, A. Genomics reveals domestication history and facilitates breed development. In: Oldenbroek, K. (ed.) Utilisation and conservation of farm animal genetic resources. Wageningen: Wageningen Academic Publishers; 2007 p 75-102

VanLith, H. A. & VanZutphen, L. F. M. 1996. Characterization of rabbit DNA microsatellites extracted from the EMBL nucleotide sequence database. Animal Genetics, 27, 387-395.

Vignal, A., Milan, D., SanCristobal, M. & Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution, 34, 275-305.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis,

E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature, 420, 520-62.

Weller, J. I., Kashi, Y. & Soller, M. 1990. Power Of Daughter And Granddaughter Designs For Determining Linkage Between Marker Loci And Quantitative Trait Loci In Dairy-Cattle. Journal of Dairy Science, 73, 2525-2537.

Wiener, P. & Wilkinson, S. 2011. Deciphering the genetic basis of animal domestication. Proceedings of the Royal Society B-Biological Sciences, 278, 3161-3170.

Windig, J. J. & Engelsma, K. A. 2010. Perspectives of genomics for genetic conservation of livestock. Conservation Genetics, 11, 635-641.

Woods, C. G., Cox, J., Springell, K., Hampshire, D. J., Mohamed, M. D., McKibbin, M., Stern, R., Raymond, F. L., Sandford, R., Sharif, S. M., Karbani, G., Ahmed, M., Bond, J., Clayton, D. & Inglehearn, C. F. 2006. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. American Journal of Human Genetics, 78, 889-896.

Woolliams J. and Toro M. (2007). What is genetic diversity. In: Utilisation and conservation of farm animal genetic resources, Oldenbroek, K. (Ed.), Wageningen Academic Publishers, The Netherlands, pp. 29-54.

CHAPTER 2

Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds

Elia Pintus[1], Silvia Sorbolini[1], Andrea Albera[2], Giustino Gaspa[1], Corrado Dimauro[1], Roberto Steri[1], Gabriele Marras[1], Nicolò Pietro Paolo Macciotta[*1]

Article published on Animal Genetics, 2013, doi: 10.1111/age.12076

Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds

SUMMARY

Selection is the major force affecting local levels of genetic variation in species. The availability of dense marker maps offers new opportunities for a detailed understanding of genetic diversity distribution across the animal genome. Over the last fifty years, cattle breeds have been subjected to intense artificial selection. Consequently, regions controlling traits of economic importance are expected to exhibit selection signatures. The fixation index (F_{st}) is an estimate of population differentiation, based on genetic polymorphism data and it is calculated using the relationship between inbreeding and heterozygosity. In the present study, the locally weighted scatterplot smoothing regression (LOWESS) and a Control Chart approach were used to investigate selection signatures in two cattle breeds with different production aptitude (dairy and beef).

F_{st} was calculated for 42,514 SNPs marker loci distributed across the genome in 749 Italian Brown and 364 Piedmontese bulls respectively. The statistical significance of F_{st} values was assessed using a Control Chart. The LOWESS technique was efficient in removing noise from the raw data and was able to highlight selection signatures in chromosomes known to harbour genes affecting dairy and beef traits. Examples are, the peaks detected for BTA2 in the region where the *myostatin* gene is located and for BTA6 in the region harbouring the *ABCG2* locus. Moreover, several loci not previously reported in cattle studies were detected.

Key words: SNPs, F_{st}, LOWESS, cattle breeds

INTRODUCTION

The study of the genetic basis of differences among animal populations is a hot topic of animal genetics. The quantification of allelic richness and the evaluation of their association with phenotypes represent tools for the safeguard and the management of local populations. Moreover, identification of genomic regions involved in phenotypic differences between individuals provide useful knowledge for gene assisted selection programmes.

Specialized breeds have been subjected to intense selection. A main consequence has been the progressive erosion of local levels of genetic variation that may have compromised the ability to challenge environmental factors (Mäki-Tanila et al., 2010). Thus a genetic comparison between selected and autochthonous populations may allow for the identification of genomic regions involved in the control of fitness traits. On the other hand, studies involving highly selected breeds with different production aptitudes, as the case of dairy and beef cattle, can provide an exciting opportunity for studying signatures of selective breeding (Hayes et al., 2008a; Qanbari et al., 2010). Actually, little is known about the effects of intensive, directional and prolonged selection on genome sub-structure of domestics species.

In population genetics, the identification of a locus target of selection is based on the existence of a reduction in nucleotidic diversity, or on an increase linkage disequilibrium (LD) and/or a changed allele frequency (Doebley et al., 2006). Currently, different statistical methods are used for the detection of selection signatures. Many of them are based on the comparison of allele frequencies or haplotype structure (for a review see Biswas and Akey 2006). The most commonly used metrics are the r^2 for measuring LD and the fixation index F_{st} (Weir and Cockerham, 1984). A quite recent approach, extensively studied in human populations, is based on the detection of runs of homozygosity (ROH), defined as uninterrupted stretch of homozygous genotypes (Gibson et al., 2006). The extent and frequency of ROHs can be used as an indication of past or recent inbreeding (Khatkar et al., 2010; Purfield et al., 2012; Ferencakovic et al., 2012). However, for many of these methods it is difficult to develop a proper statistical test. This is particularly true when searching for selective signatures within a single population.

High throughput platforms able to simultaneously genotype for many thousands of SNP offer a powerful tool for the assessment of the genetic diversity across the genome (Andersson and Georges 2004). Genome Wide Analysis (GWA) have been performed to

clarify the role of selection and drift in the evolutionary processes (Biswas and Akey 2006). Several recent studies have proposed the hitch-hiking mapping approach for identification of target of positive selection. The basic assumption is that the substitution of favourable allele at one site results in a reduction of variability at closely linked sites and lead to the allele fixation in a population (Przeworski et al., 2005). Actually, the abundance of SNP throughout the genome makes them particularly suitable in the detection of such selective sweeps (Andersson and Georges 2004).

However, such a huge amount of information has become rather problematic to interpret. A major issue is represented by the great variability of the signal pattern (for example heterozygosity or other related statistics as F_{st}) along the chromosome. An usual empirical practice to smooth data is to work on average values of sliding windows of predetermined size (Weir et al., 2005; Barendse et al., 2009). A common problem is represented by the development of a suitable statistical test able to assess whether an index of the genetic difference between two populations can be considered significant. Different approaches have been proposed to assess F_{st} statistical significance, as the calculation of q-values for Kernel-smoothed values (Flori et al., 2009), permutation test based on binomial distribution of the SNP allelic frequencies (Stella et al., 2010) or the setting of a threshold of one standard deviation from the mean for smoothed F_{st} values (Kijas et al., 2012).

In the present work, an approach for studying selection signatures in two Italian cattle breeds with different production aptitude, Italian Brown and Piedmontese (dairy and beef, respectively), is proposed. In particular, a local regression is used to smooth raw F_{st} data and a Control Chart is applied to predicted data for identifying significant values. The method is challenged to identify genes that have been reported to be involved in the genetic determinism of dairy and beef traits in cattle.

MATERIALS AND METHODS

A sample of 749 Italian Brown and 364 Piedmontese bulls was considered in the study. Animals were genotyped with the Illumina Bovine bead-chip containing 54,001 SNP (http://www.illumina.com). Only SNP located on the 29 autosomes in the Btau4.0 build of the Bovine Genome assembly were considered. Quality control was performed on the genotypes. SNP were removed if: monomorphic in both breeds; had a percentage of missing data higher than 2.5% had a minor allele frequency lower than 1%. After data

editing, 42,514 markers were retained for the study. Missing data were replaced with the most frequent allele at that specific locus for each breed.

Allele frequencies, observed and expected heterozygosity were calculated for each breed. Total allelic frequencies for each locus, f_p and f_q, considering all animals as a single population were calculated as:

$$f_p = [f_B (2 \cdot n_B) + f_P \cdot (2 \cdot n_P)] / (n_B + n_P);$$

Where f_B and n_B are frequencies of alleles and number of individuals in Brown, and f_P and n_P are frequencies of alleles and number of individuals in Piedmontese.

$$f_q = 1 - f_p$$

Then, expected heterozygosity in populations (Hs) and overall (Ht) were calculated. Finally, F_{st} was calculated according to Weir and Cockerham (1984) as:

$$F_{st} = Ht - Hs/Ht$$

In order to smooth F_{st} pattern, data were fitted with a Locally Weighted Scatterplot Smoothing (LOWESS) regression using the PROC LOWESS of SAS/STAT software version 9.2 (SAS Institute, Inc; Cary NC) (Cleveland 1979). The LOWESS has been used in genetics for smoothing model R^2 in the statistical analysis of molecular marker data (Questa-Marcos et al., 2010). In the LOWESS, the space of the independent variable is fragmented into different intervals for which separate regressions are fitted. The aim of the method is to remove noise from raw data and to clarify graphical presentations. A critical point in fitting LOWESS is the identification of a suitable dimension of the data interval to be included in the analysis. In other words, if x is the number of adjacent points to be used in the estimation procedure of a set of n data, each region contains a fraction of points given by x/n. This ratio is defined as the smoothing parameter S of the LOWESS regression. As S increases, the fitted line will be smoother until $S = 1$ that corresponds to a single line (i.e. the standard linear regression). Consequently, the goodness of fit depends strongly on the smoothing parameter used (Cohen 1999).

In general, the number of markers considered in the local regression was different across chromosomes, being directly related to their length. Therefore the use of the same S parameter in all chromosomes could not be feasible. In the present work, a smoothing parameter corresponding to an interval of 20 SNPs for each separated regression gave the best results. The different smoothing parameters used for each chromosome are reported in Table 1.

Table 1 LOWESS smoothing parameter (S) and interval length in base pair (bp) for each chromosome (BTA)

BTA	Smoothing parameter (S)	Length in bp
1	0.007057163	161,021,444
2	0.008594757	140,672,838
3	0.009017133	127,908,629
4	0.009363296	124,125,394
5	0.011049724	125,804,605
6	0.009149131	122,543,360
7	0.01048218	112,064,213
8	0.009813543	116,938,581
9	0.011554015	107,962,209
10	0.010775862	119,596,824
11	0.010487677	110,120,689
12	0.014224751	85,277,438
13	0.013236267	84,344,187
14	0.013689254	81,323,942
15	0.013831259	84,598,267
16	0.014869888	77,895,388
17	0.014534884	76,454,249
18	0.017421603	66,116,595
19	0.017035775	65,213,966
20	0.014673514	75,705,448
21	0.017021277	69,171,298
22	0.018298262	61,825,382
23	0.021574973	53,329,482
24	0.018281536	64,945,342
25	0.024009604	44,021,516
26	0.021881838	51,726,098
27	0.023781213	48,726,297
28	0.024691358	46,020,951
29	0.022271715	51,979,343

In order to identify F_{st} values different from the average pattern that could be evidence of selection signatures, LOWESS smoothed F_{st} were analysed with a Control Chart approach. This methodology aims at checking a process and its variability and it can be used to identify sources of variation. In the specific case of the present study, the goal

was to partition F_{st} variation into a component due to selection, that causes a drop of heterozygosity, and a remaining random variation along the chromosome. Control Charts are graphically displayed as stream of data falling within control limits. Data exceeding these limits are flagged as outlier signals. A similar approach has been used by Kijas et al., (2012) for identifying selection signatures in sheep breeds.

A Control Chart approach has been recently used to identify sites of preferential location of genetic variation in *Mycobacterium tuberculosis* (Das et al., 2012). In the present study, smoothed F_{st} values were plotted against their position along the chromosome. Limits of the Control Chart were set at 3 standard deviations from the mean.

In order to compare the results with an assessed methodology for studying selection signatures, F_{st} values were also smoothed with a sliding windows approach. The genome was divided into windows and average F_{st} values for each interval were calculated. Sliding windows are a graphical method widely used for detect genomic regions under positive or balancing selection (Hayes et al., 2008a; Stella et al., 2010). In the present study, the size of the window was fixed at 20 SNP each (i.e. of the same size of those used in the LOWESS smoothing). The two methods were compared by examining patterns of smoothed F_{st} signals.

Annotated genes in genomic regions corresponding to peaks exceeding Control Chart limits were derived from the UCSC Genome Broswer Gateway **(**http://genome.ucsc.edu/**)**. Intervals of 500Kbp (0,25 Mbp upstream and downstream the significant region) were considered.

RESULTS AND DISCUSSION

The comparison of chromosome average heterozygosity (Hobs) between the two breeds highlights lower values for the Italian Brown (average difference of 0.04) (Figure 1). The largest difference was found for BTA6 (0.07) the smallest for BTA2 (0.02). Differences in heterozygosity between cattle breeds have been reported by other authors (Ciampolini et al., 1995; Cañón et al., 2001).

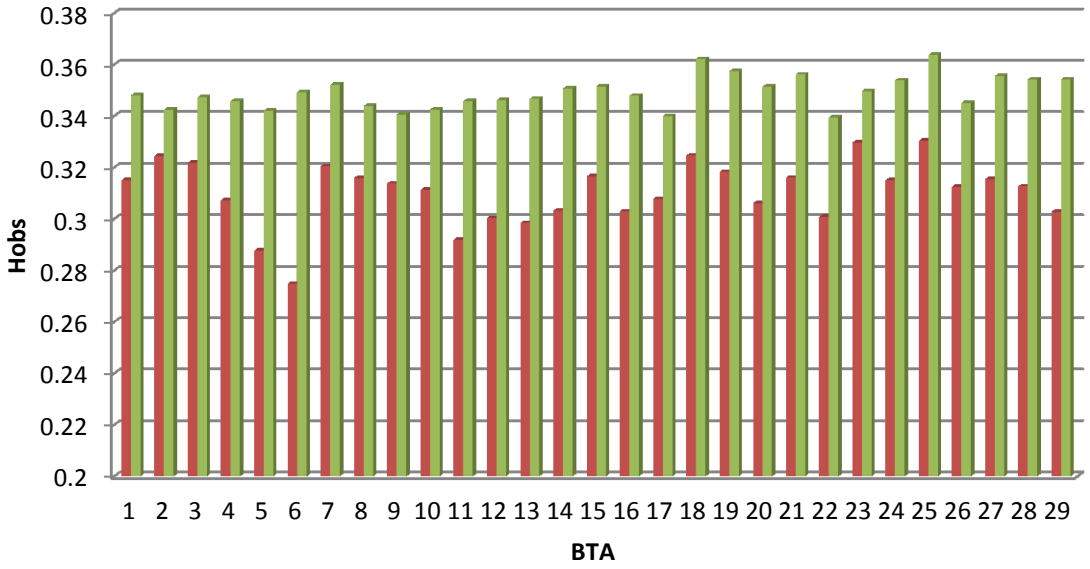
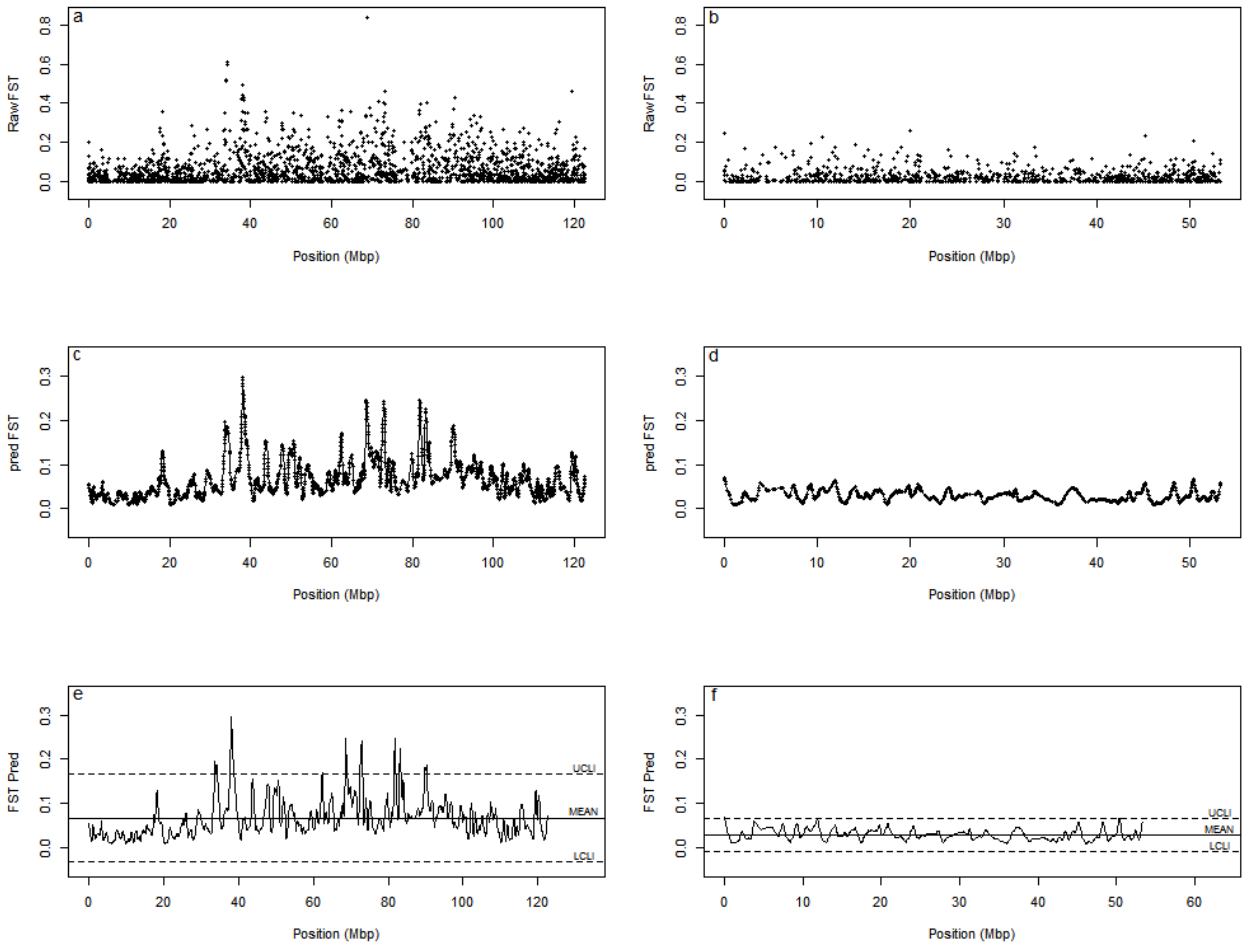


Figure 1 Comparison of average heterozygosity (Hobs) per chromosome between the two breeds (green = Piedmontese, red = Italian Brown).

In general, non smoothed SNP F_{st} values were characterized by some well defined peaks that could be evidence of divergent selection, and by a large background of low to moderate values indicating random noise. The largest number of high raw F_{st} values (n= 17, between 0.4 and 0.9) was detected on BTA6, the smallest (n= 1 with F_{st} value= 0.26) on BTA23 (Figures 2a and b, respectively). Few raw F_{st} signals (both in frequency and magnitude) were also detected on BTA28 and 29 (supplemental material). The pattern of raw F_{st} data for BTA6 was more regular compared to BTA23. This result may be interpreted as a consequence of the hitchhiking effect, because a reduction of heterozygosity (selective sweep) affect polymorphism of both individual and associated loci (Stephan et al., 2006).



Figures 2. Pattern of raw F_{st} data calculated for SNP located along the BTA 6 (a) and 23 (b); predicted F_{st} values for the SNP located along BTA6 (c) and 23 (d) using the LOWESS regression with a smoothing parameter of 0.009 and 0,021 respectively; Control Chart of predicted F_{st} values for BTA6 (e) and 23 (f). Solid line: Mean, dotted lines are: upper control limit (UCLI) and lower control limit (LCLI).

The LOWESS correction resulted in a better definition of highest peaks, even if with an expected reduction in scale due to the regression (supplemental material). Moreover, other peaks of moderate height have been disentangled from the background noise of raw F_{st} data. As an example, LOWESS corrected F_{st} values for BTA6 and BTA23 are reported in Figures 2c and d, respectively. It can be clearly seen that the smoothing procedure enhanced clustered peaks whereas isolated signals were regressed towards lower values. This behaviour was observed for the whole genome (supplemental material).

Compared to other methods currently used for studying selection signatures, such a relative simplicity could be interpreted as a sign of weakness. Actually, the LOWESS

regression is a robust non parametric method, does not relies on strong assumptions on data distribution and it could be considered as a sort "of vertical sliding windows" (Jacoby, 2000). Such a property was evident also in the comparison with the sliding windows approach performed in this study (Figure 3): the LOWESS was actually able to yield more defined and clear signals. The enhancement of cluster of peaks and the lowering of isolated signals are evidence of robustness of the method that is not affected by the variation of a single marker. This feature is particularly useful for fitting the hitch-hiking effect that occurs in the surroundings of a selectively favourable mutation (Maynard Smith and Haigh, 1974).

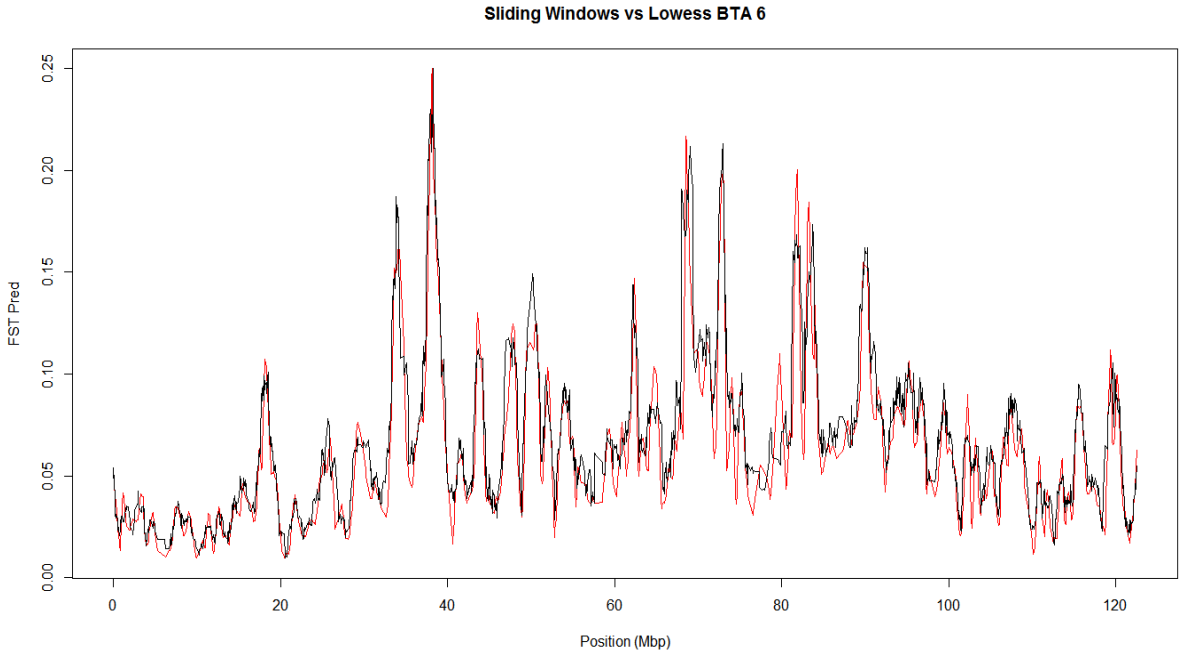


Figure 3 Plot of comparison between Sliding Windows versus LOWESS on BTA6. black line: Sliding Windows method, red line: LOWESS methodology.

The control chart analysis detected a total of 98 outliers on the whole genome. BTA6, 8 and 15 showed the largest number of signals (8, 8 and 7 respectively). On BTA11, 12, 17 and 22 a single signal was detected, whereas no peaks were found on BTA23, 25 and 29. These Figures are lower than those reported by Stella et al. (2010) that, on a large number of breeds, found 699 different putative selection signatures on the whole genome. However, Flori et al. (2009) using smoothed F_{st} across three different dairy or dual purpose breeds identified a total of 13 significative regions under selection distributed on

seven distinct chromosomes. Some of these regions correspond to those detected in the present work. An example are peaks of smoothed F_{st} that have been found on BTA6 regions where *LAP3* (*leucine aminopeptidase 3* at 37,871,423-37,896,860 bp) and *LCORL* (*ligand dependent nuclear receptor corepressor-like* at 38,137,617-38,288,047 bp) loci map.

Five peaks distributed throughout the chromosome were detected BTA19. A total of 66 different annotated loci for the corresponding genomic regions were retrieved from UCSC Genome Browser Gateway data base. This is the highest number of genes per chromosome found in the present analysis. This result is in agreement with the study of Band et al. (2000), that reported a significantly larger number of mapped genes for BTA19 compared to the other autosomes.

As far as the use of the Control Chart for testing outliers is concerned, the way confidence limits are set implies an assumption of normality for data distribution. Actually F_{st} often shows a heavily skewed distribution (Deng et al., 2007). However, a way to deal with this problem is to divide data into subgroups and then use their averages which could be considered approximately normally distributed (Morrison, 2008). Such an approach is similar to what has been done in the present work, where actually intervals of F_{st} data were considered for the LOWESS smoothing. Recently, to investigate levels of genetic diversity and to characterise the role of domestication and selection on the sheep genome, Kijas and co-workers (2012) performed a genome-wide analysis using smoothed SNP-specific F_{st} plotted for values on excess of one standard deviation from the mean. In any case, being straightforward to interpret without specific statistical background and simple to update, the Control Chart approach has been widely used in genetics, medicine and other fields of applied biology (Westgard et al., 1981; Coskun et al., 2008; Das et al., 2012).

The reliability of the proposed method was confirmed by smoothed F_{st} values that exceeded Control Chart limits in regions of the genome where genes known to affect productive traits are located. An evident example was the highest peak detected at about 37 Mbp on BTA6 (Figure 2e). It was the largest smoothed F_{st} predicted value (0.30) observed across the whole genome in the present study. Some genes known to affect milk production traits have been mapped in this region. Examples are *Family with sequence similarity 13 member A* (*FAM13A1*) (36,740,247-36,843,133 bp) (Cohen et al., 2004), *ATP-binding Cassette, sub-family G (WHITE), member 2* (*ABCG2*) (37,342,201-37,433,870 bp), *secreted phosphoprotein 1* (*SPP1*) (37,511,672-37,511,830 bp) and *peroxisome proliferator-activated receptor gamma, coactivator 1 alpha* (*PPARGC1A)*

(44,797,216-44,935,623 bp) (Cohen-Zinder et al., 2005; Ron and Weller 2007; Sheehy et al., 2009). On the other hand, no LOWESS predicted F_{st} peaks were detected on BTA23 (Figure 2f).

A further example is represented by a peak exceeding the chart limits that was detected between 6,5-7,5 Mbp on BTA2. It is well known that *myostatin* (*MSTN*) locus that controls double muscling phenotype in cattle is located in position 2q14-q15 between 6,532,697 and 6,539,265 bp. Actually this gene is reported to be fixed for the p.Cys313Tyr variant in the Piedmontese breed (Casas et al., 1999). Even though this causative mutation is not present in the SNP chip, the signal has been detected in the adjacent markers.

A rather unexpected result has been obtained on BTA14, where no relevant signals in the region where the *DGAT1* locus (*dyacylglycerol O-acyltransferase 1*) maps were found. However, it should be remembered that some studies have reported the fixation of the p.Lys232Ala variant both for the Italian Brown and Piedmontese breeds (Kaupe et al., 2004).

Other detected genomic regions, in agreement with previous researches on selection signatures in cattle, were those harbouring genes affecting coat colour. These loci have been under strong selection considering the importance of this trait in defining cattle breeds (Flori et al., 2009; Wiener and Wilkinson 2011). In the present study, two selection signatures were observed on BTA18 (12-13Mbp) and between 72-73 Mbp still on BTA6. In these chromosomic regions are located the *Melanocortin 1 receptor (MC1R) and* the *Kit* (*V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog),* loci respectively. The Extension locus controls melanine synthesis. The presence of three different alleles (E, E^1 and e) in both cattle breeds considered in this study has been reported (Russo et al., 2007). The *Kit* locus is responsible for the "Piebald" spotted coat-colour pattern in cattle and other species. This is interesting because Brown Swiss and Piedmontese breeds did not show Piebald phenotype (Stella et al., 2010), confirming the complex genetic architecture of coat colour in mammals.

In the present study, strong selection signals have been identified also in genomic regions not previously associated to traits of economic importance.

Several genes related to calcium homeostasis and metabolism were found. *Osteocrin* (*OSTN*) on BTA1, the *calcitonin receptor (CALCR)* and *calmodulin 2 (CAM2)* on BTA4 encode for bone specific proteins that appears to act as soluble osteoblast/osteoclast regulators (Thomas et al., 2003). The analysis of BTA10 outliers has revealed the presence of one interesting gene, the *GREM1* that encode for the *gremlin 1*, a

protein required for the osteoblastic activity and mineral apposition (Canalis et al., 2012). Moreover, on BTA15 *STIM1* (*Stromal interaction molecule 1*, was highlighted. It is expressed in mammary gland and it is essential for the cellular storage of calcium and the activation of the calcium influx pathway (Li et al., 2012). Actually, milk production is a complex biological process involving different tissues and governed by many genes (Finucane et al., 2007; Lemay et al., 2009). Bone and mammary tissues are related via the same signalling pathways (Cohen at al., 2004). Bone is a dynamic tissue continually modelled through the coordinate actions of the bone forming osteoblast and resorbing osteoclast (Budayr et al., 1989). Lactation is considered one of the most important events that determine bone remodelling due to the relevant calcium mobilization (Qing et al., 2012). During the production of milk, plasmatic Ca^{+2} entries into the mammary epithelial cells through a phenomenon called calcium influx pathway (McAndrew et al., 2011).

A further set of highlighted genomic regions are those were genes related to epithelial cell proliferation, skeletal muscle and bone morphogenesis map. As reported in the previous section, a selection signature in a large region spanning between 7 and 10 Mbp was found on BTA2. Several genes involved in the biology of muscular apparatus have been mapped in this region. Examples are the *Bridging integrator 1* (*BIN1*), that plays an important role in muscle cell biology (Sedwick 2010), and the *Solute Carrier family 40 (iron regulated transporter), member1* (*SLC40A1*) locus, that codes for the ferroportin 1 (FPN1) a protein with an essential role in the regulation of iron levels on the body.

On BTA11 the *Bone morphogenetic protein 10 (BMP10)* a growth factor belonging to the TGF-ß superfamily known for its ability to induce bone and cartilage development (Groenveld and Burger, 2000) was found. Moreover, *SNAI3 (Snail homolog 3*) and *CDH15 (cadeherin 15, type1, M-cadherin (myotubule)* were highlighted on BTA18 (12,908,122-13,260,964 bp). They are involved in the skeletal morphogenesis and myoblast differentiation (Moran et al., 2002; Zhuge et al., 2005). An evident peak around 26-27 Mbp was observed on BTA20. In this genomic region is annotated the *Follistatin* (*FST*) locus. This protein acts blocking the binding of *Myostatin* to its receptor and causing an abnormal muscle development (McPherron and Lee, 2001). Table 2 reports other regions identified by peaks exceeding Control Chart limits and the annotated genes involved in skeletal muscle development and metabolism. Strong selection signatures observed in regions of genes related to muscle development, differentiation and metabolism could be interpreted as signs of selection within the Piedmontese. However, it should be

remembered that the Brown Swiss was originally a dual purpose breed. Thus some of these genes might have also contributed to determine the Brown's phenotype. A deeper knowledge of the role of these genes in muscular cells could be of help for selecting markers useful for beef cattle breeding.

Table 2 List of putative candidate genes obtained on the basis of Control Chart outliers

Biological Function	**BTA**	**Position Mbp**	**Gene name**
Immune Response	5	81,763,516-81,779,866	*USP18 ubiquitin specific peptidase 18*
	16	23,622,572-23,625,154	*TLR5 toll-like receptor 5*
	17	57,084,217-57,115,368	*HVCN1 hydrogen voltage-gated channel 1*
	18	1,880,236-12,887,173	*CYBA cytochrome b-245, alpha polypeptide*
	19	21,395,686-21,409,196	*TMIGD1 transmembrane and immunoglobulin domain containing 1*
	26	23,471,864-23,478,382	*NFKB2 nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)*
Reproduction	1	155,943,716-155,956,150	*EAF1 ELL-associated factor 1*
	3	86,007,282-86,200,728	*AK4 Adenylate kinase 4*
	5	20,587,724-20,612,963	*KITLG Kit ligand*
	6	37,961,724-37,987,164; 38,153,046-38,199,153; 38,227,954-38,378,385	*LAP3 leucine aminopeptidase 3; NCAPG non-SMC condensing I complex, subunit G; LCORL ligand dependent nuclear receptor corepressor-like*
	8	104,876,401-104,908,801	*TXNDC8 Thioredoxin domain containing 8 (spermatozoa)*
	9	41,225,543-41,246,855	*AMD1 adenosylmehtionine decarboxylase 1*
	10	36,873,000-36,890,219	*TYRO3 TYRO3 protein tyrosine kinase*

	11	70,119,086-70,174,863	*GMCL1 germ cell-less, spermatogenesis associated 1*
	14	60,023,782-60,033,403	*ODF1 outer dense fiber of sperm tails 1*
	15	18,520,179-18,520,292; 24,046,636-24,397,152	*FDX1 ferredoxin 1;* *CADM1 cell adhesion molecule 1*
	18	13,648,996-13,652,641	*SPATA2L spermatogenesis associated 2-like*
	19	24,498,808-24,501,792; 24,628,862-24,646,107; 50,216,969-50,223,538	*GSG2 germ cell associated 2 (haspin);* *P2RX1 purinergic receptor P2X, ligand-gated ion channel,1 ;* *DDX5 DEAD (Asp-Glu-Ala-Asp) box helicase 5*
Cell growth, proliferation and differentiation	1	76,216,039-76,832,685	*FGF12 fibroblast growth factor 12*
	12	70,974,850-71,682,818	*HS6ST3 heparan-sulfate 6-O-sulfotransferase 3*
	13	47,627,052-47,683,993 48,488,115-48,536,904	*CDS2 CDP-diacylglycerol synthase (phosphatidate cytydyltransferase) 2 ;* *FERMT1 fermitin family member1*
	14	60,169,396-60,307,900	*UBR5 ubiquitin protein ligase E3 component n-recognin 5*
	16	20,839,081-20,931,656 26,878,683-26,905,046	*TGFB2 transforming growth factor, beta2;* *PSEN2 presenilin 2 (Alzheimer disease 4)*
	17	57,146,787-57,165,849	*PPP1CC protein phosphatase 1, catalytic subunit, gamma isozyme*
	19	*35,124,710-35,129,750* *35,535,495-35,544,295* *35,953,771-35,969,817* *45,567,703-45,574,688*	*MAPK7 mitogen- activated protein kinase 7;* *DRG2 developmentally regulated GTP binding protein 2;* *FLCN folliculin;* *GRN granulin*

Ions metabolism	1	78,466,667-78,488,928	*CLDN16 claudin 16*
	2	9,451,265-9,580,452	*CALCRL calcitonin receptor- like*
	4	11,016,143-11,126,171	*CALCR calcitonin receptor*
	5	81,136,111-81,146,812	*KCTD17 potassium channel tetramerisation domain containing 17*
	8	*11,700,825-11,763.811*	*ACO1 aconitase 1, soluble*
	17	*56,466,582-56,498,348,;* *56,790,348-56,488,450*	*CAMKK2 calcium/calmodulin-dependent protein kinase kinase 2, beta;* *ATP2A2 ATPase, Ca++ transporting, cardiac muscle, slow twitch 2*
	18	47,597,196-47,605,452	*KCNK6 potassium channel, subfamily k,member 6*
	19	56,790,348-56,844,450; 24,594,778-24,623,204	*ATP2A3 ATPase, Ca++ transporting, ubiquitous;* *CAMKK1 calcium/calmodulin-dependent protein kinase kinase 1, alpha*
	24	31,254,115-31,532,051	*KCTD1 potassium channel tetramerisation domain containing 1*
	26	22,854,587-22,857,882; 24,558,695-24,564,440	*KCNIP2 Kv channel interacting protein 2;* *CALHM3 calcium homeostasis modulator 3*
Lipid metabolism	2	6,192,072-6,348,621	*HIBCH 3 hydroxibutirril o idrolase*
	5	53,700,174-53,700,270	*ACAT2 AcetylCoA acetyltransferase 2*
	10	59,440,432-59,504,627	*CYP19A1 cytochrome P450, family 19, subfamily A, polypeptide1*
	13	48,423,438-48,446,513	*CRLS1 cardiolipin synthase 1*
	15	55,827,654-56,160,380	*ACER3 alkaline ceramidase 3*
	18	13,212,190-13,250,827	*ACSF3 Acyl-CoA synthase family member 3*
	19	35,671,152-35,687,188	*SREFB1 sterol regulatory element*

			bibdingtranscription factor 1
mammary gland metabolism	2	10,226,975-10,322,817	*ITGA V integrin alpha V*
	6	37,351,167-37,421,683 37,431,966-37,490,645 37,511,673-37,518,636 72,298,906-72,346,677 72,741,252-72,828,528	*ABCG2 ATP-binding cassette, sub-family G, (WHIT), member 2;* *PKD2 polycistic kidney disease 2;* *SPP1 secreted phosphoprotein 1;* *PDGFRA platlet-derived growth factor receptor, alpha poypetide;* *KIT V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog*
	7	62,635,246-62,657,995	*SPARC secreted protein,acidic, cysteine-rich (osteonectin)*
	10	29,529,387-29,541,874 36,595,794-36,596,071	*GREM1 gremlin1, DAN family BMP antagonist;* *IGFBP3 insulin-like growth factor binding protein 3*
	14	12,506,878-12,583,201	*MTMR2 myotubularin related protin 2;*
	15	20,478,802-20,482,029 50,442,087-50,753,021	*CRYAB crystalline alpha B;* *STIM1 stromal interaction molecule 1*
	18	14,699,407-14,998,970	*ITGF1 integrin alpha FG-GAP repeat containing 1*
	19	35,122,081-35,124,619 35,823,315-35,854,048	*MFAP4 microfibrillar associated protein 4;* *PEMT phosphatidylethanolamine N-methyltransferase*
	24	30,845,569-30,860,104	*AQP4 aquaporin 4*
	27	48,475,540-48,478,931	*OXSM 3-oxyacyl-ACP synthase, mitochondrial*
bone and muscle metabolism	1	77,682,355-77,718,578 155,717,664-155,777,449	*OSTN osteocrin;* *CAPN7 calpain 7*
	2	5,595,799-5,652,801	*BIN1 bridging integrator1;*

	6,532,697-6,539,265 7,066,569-7,148,685 7,740,061-7,779,695	*MSTN myostatin;* *SLC40A1 solute carrier family 40(iron regulated transporter) member1;* *COL3A1 collagen type (III) alpha 1*
7	5,824,715-5,935,402	*MYO9B myosin IXB*
8	11,291,512-11,308,875 105,221,050-105,315,564	*CLU clusterin* *MUSK muscle, skeletal, receptor tyrosine kinase*
10	19,387,377-19,414,041	*PKM pyruvate kinase, muscle*
11	69,145,567-69,152,285 70,648,036-70,648,340 71,029,777-71,105,164	*BMP10 bone morphogenetic protein 10;* *CAPN14 calpain 14;* *CAPN13 calpain 13*
13	48,488,115-48,536,904	*FERMT1 fermitin family member 1;*
15	11,852,140-11,854,278 56,045,818-56,103,271	*PPP1R14C protein phosphatase 1, regulatory (inhibitor) subunit 14C ;* *CAPN5 calpain 5*
16	24,021,217-24,065,788 24,108,032-24,166,355	*CAPN8 calpain 8;* *CAPN2 calpain 2*
17	56,905,068-56,915,878 57,330,762-57,338,500	*ARPC3 actin related protein 2/3 complex, subunit3 21kDa;* *MYL2 myosin light chain 2,regulatory, cardiac, slow*
18	12,908,122-12,913,750 13,260,964-13,279,948 47,527,738-47,531,970 47,701,775-47,875,177	*SNAI3 snail homolog 3;* *CDH15 cadherin 15, type1,M-cadherin (myotubule);* *PPP1R14A protein phosphatase 1, regulatory (inhibitor) subunit 14A;* *RYR1ryanodine receptor 1(skeletal)*
20	23,624,160-23,688,918 27,297,146-27,302,564	*GPBP1 GC-rich promoter binding protein 1;* *FST follistatin*

	21	45,895,690-45,898,343	*CFL2 cofilin 2 (muscle)*
	26	12,908,235-12,917,607 23,540,685-23,557,026	*ANKRD1 ankyrin repeat domain 1 (cardiac muscle);* *ACTR1A ARP1 actinn related protein 1 homolog a, centractyn alpha (yeast)*
others	10	19,817,179-19,849,769	*ADPGK ADP-dependent glukonidase*
	11	68,612,764-68,639,385 69,642,777-69,707,857	*CNRIP1 cannabinoid receptor interacting protein 1;* *GFTP1 glutamine-fructose-6 phosphate transaminase 1*
	15	20,576,533-20,611,864	*DLAT dihydrolipoamide S-acetyltransferase*
	18	13,776,888-13,778,639	*MC1R melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)*
	19	45,226,420-45,227,150 45,325,106-45,329,822	*PPY pancreatic polypeptide* *G6PC3 glucose 6 phosphatase, catalytic, 3*

In this study few putative candidate genes were detected for lipid metabolism (Table 2). This is probably due to the fact that intramuscular fat deposition not only depends on the genetic background but also by other factors such as age, sex, nutrition and farm conditions.

An interesting result was the identification of numerous putative candidate genes involved in the reproductive function (Table 2). Among them, the specific ligand (*KITLG*) for the *Kit* receptor was identified on BTA5. Actually the interaction between *kit* and its ligand is crucial for fertility (Mithraprabhn and Loveland, 2009). Such results suggest a further deepening of the genetic basis of relationships between production and fertility traits (Bello et al., 2012).

Finally, this genome wide analysis highlighted the presence of selection signatures for a group of similar genes. Six genes belonging to the Calpains gene family were detected in four different chromosomes: *CAPN 7* on BTA1, *CAPN 13* and *14* on BTA11, *CAPN 5* on BTA15, and *CAPN 2* and *8* on BTA16, respectively. Several studies indicate

calpains as regulators of apoptosis and suggest an involvement of the calpain system during the muscle postmortem apoptotic pathway (Mohanty et al., 2010). The interaction among calpains and other proteases is considered a foundamental step for after slaughtering meat tenderization (Koohmaraie, 1992). A multi gene family is formed by duplication of a single original gene. In cattle, 3.1% of the genome is composed of duplicated genes, most of which encoding proteins involved in innate immunity, sensory receptors and reproduction (Elsik et al., 2009). Generally, the expansion or contraction of gene families can be due to chance or is the result of natural selection. Gene gain or loss are so considered to be an incentive for evolutionary change and as a common advantageous response to selective regimes (Demuth et al., 2006).

A detailed list of putative genes for all 29 bovine chromosomes highlighted by the Control Chart outliers values is summarized in Table 2. All gene content information presented was derived from the UCSC Genome Broswer Gateway (http://genome.ucsc.edu/) using the fourth draft of bovine genome sequence assembly (Btau 4.0) and from NCBI or Swiss ProtK Source consultation.

Results obtained in this study on the comparison between two cattle breeds with different production aptitude, beef and dairy, agree with previous report on milk QTL (Cohen et al., 2004) and transcriptome analysis (Bionaz and Loor, 2008; Lemay et al., 2009). Moreover it confirmed what observed in previous comparisons between cattle breeds (Flori et al., 2009; Stella et al., 2010; Qanbari et al., 2011). Differences have been found with the QTL analysis carried out by Prasad et al. (2008) on BTA19 and 29, where selection signatures in different chromosomal regions were found. A possible explanation could be represented in the different genomic assemblies used.

CONCLUSIONS

The combined use of a LOWESS regression and a Control Chart approach here proposed was effective in studying the genetic differences between the Piedmontese and the Italian Brown cattle breeds. In particular, the local regression was able to yield a smooth F_{st} pattern, easy to interpret compared to raw data. The Control Chart allowed for a quite simple detection of significant F_{st} values that may indicate selection signatures. The method was validated by comparing results obtained on several chromosomes with previous reports in cattle (Hayes et al., 2008a and b; Flori et al., 2009; Stella et al., 2010; Qanbari et al., 2011). Moreover, some regions harboring genes not yet associated to traits

of economic importance for livestock have been detected. In particular, genes involved in the calcium metabolism and muscle biology have been highlighted. The methodology could be proposed as an easy approach for performing a whole genome scan in studies aimed at identifying selection signatures by using high throughput SNP maps.

Authors' contributions

E.P., S.S. and N.P.P. M. planned and performed analysis and drafted the manuscript. A.A., G.G., C.D., R.S. and G.M., performed the analysis. All authors reviewed the manuscript.

Acknowledgments

This work was funded by the Italian Ministry of Agriculture (grants SELMOL and INNOVAGEN). Authors wish to acknowledge Proff. Alessio Valentini and Paolo Ajmone Marsan for providing bull genotypes.

REFERENCES

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

Band M.R. (2000) An ordered comparative map of the cattle and human genomes. *Genome Research* **10,** 1359-1368.

Barendse W., Harrison B.E., Bunch R.J., Thomas M.B. and Turner L.B. (2009) Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* **10,** 178.

Bello N.M., Stevenson J.S., Tempelman R,J. (2012) Invited review: Milk production and reproductive performances: Modern interdisciplinary insights into an enduring axiom. *Journal of Dairy Science* **95,** 5461_5475.

Bionaz M. and Loor J. (2008) Gene networks driving bovine milk fat synthesis during the lactation cycle. *BMC Genomics* **9,** 366.

Biswas S. and Akey J.M. (2006) Genomic insights into positive selection. T*rends in Genetics* **22 ,** 437-446.

Budayr A.M., Halloran B.P., King J.C., Diep D., Nissenson R:A., Strewler G.j. (1989) High levels of a parathyroid hormone-like protein in milk. *Proceedings of the National Academy of Science U.S.A.* **86,** 7183-7185.

Canalis E., Parker K., Zanotti S. (2012) Gremlin 1 is required for skeletal development and postnatal skeletal homeostasis. *Journal of Cell Physiology* **227,** 269-277.

Cañón J., Alexandrino P., Bessa I., Carleos C., Carretero Y., Dunner S., Ferran N., Garcia D., Jordana J., Laloë D., Pereira A., Sanchez A., Moazami-Goudarzi K. (2001) Genetic diversity measures of local European beef cattle breeds for conservation purposes**.** *Genetic Selection Evolution* **33,** 311-332.

Casas E., Keele J.W., Fahrenkrug S.C., Smith T.P., Cundiff L.V. and Stone R.T. (1999) Quantitative analysis of birth, weaning, and yearling weights and calving difficulty in Piedmontese crossbreds segregating an Inactive Myostatin Allele. *Journal of Animal Science* **77,** 1686-1692.

Ciampolini R., Moazami-Goudarzi K., Vaiman D., Dillmann C., Mazzanti E., Foulley J.L., Leveziel H. and Cianci D. (1995) Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. *Journal of Animal Science* **73,** 3259-3268.

Cleveland W.S. (1979) Robust Locally Weighted Fitting and Smoothing Scatterplots. *Journal of the American Statistical Association*, **74,** 829-836.

Cohen M., Reichenstein M., Everts-van der Wind A., Heon-Lee J., Shani M., Lewin H.A., Weller J.I., Ron M. and Seroussi E. (2004) Cloning and characterization of FAM13A1 a gene near a milk protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holstein. *Genomics* **84,** 374-383.

Cohen R.A. (1999) An introduction to PROC LOESS for local regression. SAS Institute Inc. Cary, North Carolina, USA.

Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., Everts-van der wind A., DrackleyJ.K., Band M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I., Ron M. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research* **15,** 936-44.

Coskun A., Serster M., Fraterman A., Unsal I. (2008) a new international quality control chart based on biological variation. *Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement* **13**, 69-75

Das S., Duggal P., Roy R., Myneedu V.P., Behera D., Prasad H.k., Bhattacharya A. (2012) Identification of hot and cold spots in genome of *Mycobacterium tubercolosis* using Shewart Control Chart. *Nature Scientific Reports* **2,** article number:297 doi:10.1038/srep00297.

Demuth J.P., DeBie T., Stajich J.E., Cristianini N., Hahm M.W. (2006) The Evolution of mammalian gene families. *PLoS ONE* **1,**: e85. doi.10.B71/journal.pone.0000085.

Deng L.B., Tang X.L., Kang J., Wang Q.Y., Zeng C.Q. (2007) Scanning for signature of geographically restricted selection based on population genomics analysis. *Chinese Science Bulletin* **52,** 2649-2656.

Diskin S., Li M., Yang S., Glesseen J., Hakonarson H., Bucan M., Maris J.M., Wang K. (2008) Adjustment of genomic waves in signal intensities from whole- genome SNP genotyping platforms. *Nucleic Acid Research* **36**, e126.

Doebley J.F., Gaut B.S. and Smith B.D. (2006) The Molecular Genetics of Crop Domestication. *Cell* **127,** 1309-1321.

Ellis A.D. (2010) The impact of nutrition on the health and welfare of horses. *EAAP* **128,** 53-74 Wageningen Academic Publishers. The Netherlands.

Elsik C.G., Tellam R.S., Worley K.C., and the Bovine Genome Sequencing and Analysis Consortium (2009) The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science **324,** 522-528.

Ferencakovic M., Hamzic E., Gredler B., Solberg T.R., Klemetsdal G., Curik I., Solkner J. (2012) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics* **ISS0931-2668,** 1-8.

Finucane K.A., McFadden T.B., Bond J.P., Kennelly J.J., Zhao F-Q. (2008) Oneset of lactation in the bovine mammary gland: gene expression profiling indicates a strong inhibition of gene expression in cell proliferation. *Functional Integrative Genomics* doi 10.1007/s10142-008-0074-y.

Flori L., Fritz S., Jaffrézic F., Boussaha M., Gut I., Heath S., Foulley J.L., and Gautier M. (2009) The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. *PloS One* **4**, e6595.

Gibson J., Morton N.E., Colins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15,** 789-795.

Groenveld E.H.J and Burger E.H. (2000) Bone morphogenetic proteins in human bone regulation. *European Journal of Endocrinology* **142,** 1-9.

Hayes B.J., Chamberlain A.J., Maceachern S., Savin K., McPartlan H., MacLeod I. (2008a) A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. *Animal Genetics* **40,** 176–184.

Hayes B.J., Lien S., Nilsen H., Olsen H.G., Berg P. (2008b) the origin of selection signatures on bovine chromosome 6. *Animal Genetics* **39,** 105-111.

Jacoby W.G. (2000) Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies* **19,** 577-613.

Kaupe B., Winter A., Fries R., Erhardt G. (2004) DGAT1 polymorphism in *Bos Indicus* and *Bos Taurus* cattle breeds. *Journal of Dairy Research* **71,** 182-187.

Khatkar M.S., Tier B., Hobbs M., Kharkar D., Cavanagh J.A.L. Crump R., Moser G., Raadsma H.W. (2010) Genome structure in Australian Holstein Frisian cattle revealed by combined analysis of three high density SNP panels. *Proceeding Association Advancement Animal Breeding and Genetics* **18,** 247-250.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan L., Dalrymple B., and International Sheep Genomics Consortium (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PloS Biology* **10,** e 10001258.

Koomharaie M. (1992) The role of (Ca^{2+})-dependent proteases (calpains) in post-mortem proteolysis and meat tenderness. *Biochimie* **74,** 239-245.

Lemay D.G., Lynn D., Martin W.F., et al. (2009) The bovine lactation genome: insigths into the evolutionof mammalian milk. *Genome Biology* **10,** R43 1-18.

Li T., Finch E.A., Graham V., Zhang Z.S., Ding J.D., Burch J., Oh-hora M., Rosenberg P. (2012) STIM1-(Ca^{2+}) signalling is required for the hypertrophic growth of skeletal muscle in mice. *Molecular Cell Biology* **15,** 3009-3017.

Mäki-Tanila A., Fernandez J., Toro M., and Meuwissen T. (2010) Assesment and management of genetic variation. Local Cattle Breeds in Europe. Development of policies and strategies for self-sustaining breeds. Wageningen Academic Publishers 98-117.

Maynard-Smith J. and Haigh J. (1974) The hitch-hiking effect of a fovourable gene. *Genetical Research Cambridge* **23,** 23-35.

McAndrew D., Grice D.M., Peters A.A., Davis F.M., Stewart T., Rice M., Smart C., Brown M.A., Kenny P.A., Roberts-Thomson S.J., Monteith G.R. (2011) ORAI1-Mediated Calcium influx in lactation and breast cancer. *Molecular Cancer Therapeutics* **10,** 448-460.

McPherron A.C. and Lee S.J. (2001) Regulation of myostatin activity and muscle growth. *Proceedings of the National Academy of Science U.S.A.* **98**, 9306-9311.

Mithraprabhu S. and Loveland K.L. (2009) Control of KIT signalling in male germ cells: what can we learn from other systems? *Reproduction* **138,** 743-757.

Mohanty T.R., Park K.M., Pramod A.B., Kim J.H., Choe H.S., Hwang I.H. (2010) Molecular and Biological factors affecting skeletal muscle cells after slaughtering and their impact on meat quality. *Journal of Muscle Foods* **21,** 51-78.

Moran J.L., Li Y.,Hill A.A., Mounts W.M:, Miller C.P. (2002) Gene expression changes during mouse skeletal myoblast differentiation revealed by transcription profiling. *Physiological Genomics* **10,** 103-111.

Morrison L. W. (2008) The Use of Control Chart to Interpret Environmental Monitoring Data. *Natural Areas Journal* **28,** 66-73.

Prasad A., Schnabel R.D., McKay S.D., Murdoch B., Stothard P., Kolbehdari D., Wang Z., Taylor J.F., Moore S.S. (2008) linkage disequilibrium and signature of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal Genetics* **39,** 597-605.

Przeworski M., Graham C., and Wall J.D. (2005) The Signature of Positive Selection on Standing Genetic Variation. *Evolution* **59**, 2312-2323.

Purfield D.C., Berry D.P., McParland S., Bradley D.G. (2012) runs of Homozygosity nd population history in cattle. *BMC Genetics* **13,** 70.

Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S., Moore S., Thaller G., Simianer H. (2011) Application of site haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* **12**, 318.

Qanbari S., Pimentel E.C.G., Tetens J., Thaller G., Lichtner P., Sharifi A.R., Simianer H. (2010) A genome-wide scan for signatures of recent selection in Holstein cattle. *Animal Genetics* **41,** 377–389.

Qing H., Ardeshirpour L., Pjevic P.D., Dusevich V., Jahn K., Kato S., Wyslomerski J., bonewald L.F. (2012) Demonstration of osteocytic perilacunar/canalicular remodeling in mice during lactation. *Journal of Bone and Mineral Research* **27,** 1018-1029.

Questa-Marcos A., Szucs P., Close T.J., Flichking T., Muehlbauer G., Smith K.P., Hayes P.M. (2010)

Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure. *BMC Genomics* **11,** 707.

Ron M. and Weller J.I. (2007) From QTL to QTN identification in livestock- winning points rather than knock-out: a review**.** *Animal Genetics* **38,** 429-439.

Russo V., Fontanesi L., Scotti E., Tazzoli M., Dall'Olio S., Davoli R. (2007) Analysis of melanocortin 1 receptor (MC1R) gene polymorphisms in some cattle breeds.their

usefulness and application for breed traceability and authentication of Parmigiano Reggiano cheese. *Italian Journal of Animal Science* **6,** 257-272.

Schwarzenbacher H., Dolezal M., Flisikowski K., Seefried F., Wurmser C., Schlötterer and Fries R. (2012) Combining evidence of selection with association analysis increases to detect regions influencing complex traits in dairy cattle. *BMC Genomics* **13**, 48.

Sedwick C. (2010) BIN1: A protein with Great Heart. *PLoS Biology* **8,** e1000311.

Sheehy P.A., Riley L.G., Raadsma H.W., Williamson P., and Wynn P.C. (2009) A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Animal Genetics* **40,** 492-498.

Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P. (2010) Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. *Genetics* **185,** 1451-1461.

Stephan W., Song Y.S., Langley C.H. (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* **172,** 1-37.

The Bovine HapMap Consortium (2009) Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* **324,** 528-531.

Thomas G., Moffatt P., Salois P., d M.H., Gingras R., Godin E., Miao D., Goltzman L., Lanctot C. (2003) Osteocrin a novel bone-specific secreted protein that modulates the osteoblast phenotype. *Journal of Biological Chemistry* **278,** 50563-50571.

Weikard R., Widmann P., Buitkamp J., Emmerling R. and Kuehn C. (2012) Revisiting the quantitative trait loci for milk production traits on BTA6. *Animal Genetics* **43**, 318-323.

Weir B.S. and Cockerham C.C. (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370.

Weir B.S., Cardon L.R., Anderson A.D., Nielsen D.M. and Hill W.G. (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Research* **15,** 1468-1476.

Westgard J.O., Barry P.L., Hunt M.R. (1981) A Multi- Rule Shewart Chart for Quality Control in Clinical Chemistry. *Clinical Chemistry* **27,** 493-501.

Wiener P. and Wilkinson S (2011) Deciphering the genetic basis of animal domestication. *Proceedings of The Royal Society B* **doi:10.1098/rspb.2011.1376**

Zhuge **X**, Kataoka H, Tanaka M, Murayama T, Kawamoto T, Sano H, Togi K, Yamauchi R, Ueda Y, Xu Y, Nishikawa S, Kita T, Yokode M. (2005) Expression of the novel Snai-related zinc-finger transcription factor gene Smuc during mouse development. *International Journal of Molecular Medicine* **15,** 945-948.

CHAPTER 3

Detection of selection signatures in five Italian cattle breeds with different productive specialization

Detection of selection signatures in five Italian cattle breeds with different productive specialization

ABSTRACT

The man-made process of domestication and subsequently the practice of artificial selection have led to the current constitution of animals breeds and plants varieties. Objective of positive selection is to improve the biological efficiency of an individual in order to increase production performance. These processes lead to allelic changes which can be found in the genome in the form of selection signatures. The aim of this study was to detect signatures of selection in five Italian cattle breeds selected for different productive attitude (dairy, beef or both) using the Illumina Bovine SNP50 bead-chip. The data consisted of genotypes for 44,325 SNPs from 2093 Italian Holstein bulls, 749 Italian Brown bulls, 479 Italian Simmental bulls, 364 Piedmontese bulls and 410 Marchigiana bulls The statistic test applied was the parametric composite log likelihood (CLL) of the difference in allelic frequencies between the test and the reference population for a sliding window of 1Mb. In relation to the production types two new putative candidate gene clusters (SERPINs and KLKs) were detected in this study.

INTRODUCTION

Genetic drift and selection are the major forces shaping the genetic variation of population and affecting divergence of species. Alongside these two evolutionary forces the action exerted by humans through the domestication led to the change in the phenotypic traits of plants and animals (Andersson, 2012).

After decades of heated debate between neutralists and selectionists, now there is a general consensus on the fundamental role played by neutral drift and positive selection in creating genetic variation and promoting evolutionary changes in living organisms (Bamshad and Wooding, 2003; Kosiol et al., 2008). Examples of the results of the intense artificial selection performed by the man are the large number of existing dog breeds that have originated only from the wild wolf domestication that occurred about 14,000 years ago (Akey, 2009) or the hundreds of cattle breeds known today (Qanbari et al.,

2011).Therefore, domestication and subsequent selective pressure have altered most of the phenotypic and behavioral traits of these animals.

The unraveling of the history of animal domestication may help to understand biological mechanisms underlying the phenotypic variability in livestock and inferences about the effects of selection may provide important functional information (Nielsen, 2005; Wiener and Wilkinson, 2011). Objective of artificial selection in farm populations is to increase the frequency of desired alleles and simultaneously decrease that of unwanted ones in order to improve the biological and productive efficiency of breeds. Cattle breeds have been historically selected for milk and beef production, and this has led to observable phenotypic differences.

If in a population subjected to artificial selection appears a favorable mutation, the frequency of the positive allele may increase over the time until fixation. Generally, also the loci closely near the mutated one are affected by this phenomenon and also they tend to the fixation. This phenomenon is called hitch-hiking effect (Smith and Haigh, 1974). As a consequence of hitch-hiking, in the concerned genomic region there will be a reduction of heterozygosity resulting in a so called selective sweep (Przeworski et al., 2005; Stephan et al., 2006). Consequently, domestication and artificial selection appear to have left detectable signatures within the genome of livestock species (Gibbs et al., 2009). In general, selective sweeps lead to a reduction of variability within the population and an increase of the diversity between population (Weir et al., 2005). This reduction of variability can be measured by comparing the allele frequencies or haplotype structure of individuals with suitable statistical methods and estimators such as fixation index, Linkage Disequilibrium, iHS, EHH (Biswas and Akey, 2006; Lenstra et al., 2012). Currently, through molecular genetic markers such as single nucleotide polymorphisms (SNPs) is now possible to quantify and qualify the genetic variability between and within breeds. In fact, high-throughput platforms guarantee the simultaneous genotyping of thousands of SNPs for hundreds of individuals and provide a powerful tool to investigate the genetic diversity contained in the genome (Andersson and Georges, 2004).

In the present work a dataset of five Italian cattle breeds (Italian Holstein, Italian Brown, Piedmontese, Marchigiana and Italian Simmental) strongly selected for divergent production aptitudes was analysed, dairy, meat or both (dual purpose breeds). The aim was to detect regions of the cattle genome involved in the process of selection and adaptation for dairy and beef production, or that are under selection in all breeds, pointing at more

fundamental processes involved in cattle domestication. In the study a parametric composite log likelihood statistical test (CLL) was used, as same proposed by Stella et al., 2010 (calculation of CLL and smoothing of row data with sliding windows of adjacent loci). Moreover, in the present study an alternative test of significance based on the control chart application (Pintus et al., 2013) was performed. Outlier signals displayed by control chart highlighted chromosomic regions containing selection signatures for milk or beef productive traits.

MATERIAL AND METHODS

Experimental population and genotypic data

The animal population used in this study consisted of 4,095 bulls from five Italian cattle breeds, specifically selected to represent part of the dairy, beef and dual-purpose cattle genetic resources of Italy. A total of 2,093 and 749 bulls were selected from the Holstein (HOL) and Italian Brown (IB) dairy breeds, respectively. On the other hand, 364 and 410 bulls were selected from the Piedmontese (PIE) and Marchigiana (MAR) beef breeds, respectively. In addition, a 479 dual-purpose Italian Simmental bulls (SIM) were included in the dataset. The selection criteria of bull of all breeds was based on pedigree information, aimed at maximizing genetic variability and reducing, as far as possible, close relationships between animals. Biological samples (semen) used in this study were obtained within a collaboration between the research institutions, breed Associations, and Italian certified artificial insemination centers (EU Directive 88/407/CEE) involved in the SELMOL and PROZOO research projects, funded by the Italian Ministry of Agriculture and Fondazione CARIPLO, respectively. All bulls were genotyped with the Illumina Bovine SNP50 bead-chip version 1 containing 54,001 SNPs. Markers belonging to the X chromosome, non mapped, and those with within-breed call rate $\leq$ 97.5% were removed. Moreover, SNPs which were monomorphic across all breeds, and those that have an overall minor allele frequency $\leq$ 0.01 were also removed. No editing for deviations from Hardy-Weinberg equilibrium was applied. Missing values were not considered in the analysis. After editing, 44,325 SNP were retained for the study.

Detection of signatures of selection

According to Stella et al. (2010), we used the composite log likelihood (CLL) of SNP allelic frequencies along "sliding windows" of adjacent loci to detect signatures of selection on the genome of the five cattle breeds. The method is based on the comparison of allele frequencies at each locus between the sub-population of interest (test population, P_t), and the reference population (P_{n-t}, with n being the total population size), formed by all breeds in our dataset excluding the test population t. This comparison against the whole experimental population was repeated for each sub-population (breed) in the dataset. The composite log-likelihood (CLL) of the allele frequencies in the test population, compared to the reference population, was based on the binomial distribution and calculated as follows:

$$CLL_j = -\frac{1}{w}\sum_{i=j}^{j+(w-1)} \log_{10}\left(\binom{n}{k_j}(1-q_j)^{k_j}\left(q_j^{n-k_j}\right)\right)$$

where CLL_j is the composite log-likelihood at locus j,

w is the size of the sliding window;

n is the absolute frequency of both the two alleles at locus j in the test population (i.e. double the number of animals);

k is the absolute frequency of the minor allele (or of the unique allele in case of fixation for that specific breed) at locus j in the test population;

q is the absolute frequency of the minor allele in the reference population.

Thus defined, the CLL measures the logarithm of the probability that the allele frequencies observed in the test population belong to the allele frequency distribution in the reference population. In other words, the CLL indicates how likely it is -at each locus- for an allele of the test population to be sampled from the reference population, and measures the degree of similarity between the two allele frequency distributions. The larger the absolute value of composite log-likelihood the larger the difference between the two distributions (and, therefore, between the two populations).

Table 1. Different cases of log-likelihood value calculation: example 1 monomorphic for population test; example 2 monomorphic for population reference, example 3 polymorphic for both population.

	Reference Population		Population Test			
	Animals	Minor Allele	Animals	Minor Allele	Formula	CLL value
Example 1	500	230	70	139	$-\left(\binom{140}{139}(1-0.77)^{139}\left(0.77^{140-139}\right)\right)$	86.68
Example 2	500	980	70	65	$-\left(\binom{140}{65}(1-0.02)^{65}\left(0.77^{200-65}\right)\right)$	87.17
Example 3	500	470	70	75	$-\left(\binom{140}{75}(1-0.53)^{75}\left(0.53^{200-75}\right)\right)$	1.69

Under the assumption that frequency differences between populations geographically separated and with different productive specialization are due to selection, relevant CLL values could be considered signatures of selection (Stella et al., 2010). The chromosome-wide sliding window approach was adopted in order to reduce the influence of isolated extreme CLL values, and to give stronger support to the identified regions under selection. Each autosome was divided into windows with a fixed size of 1 Mb. The average number of SNP contained in each sliding windows was 19 ± 4,8 SD (min = 2 max = 36).

To identify statistically significant allele frequency differences, a control chart approach was applied as suggested by Pintus et al. (2013). In this procedure a threshold of three standard deviations over the average of whole-chromosome CLL values was defined (Figure 1). Markers exceeding this empirical threshold were considered significant signatures of selection and an accurate analysis to identify putative candidate genes has been conducted. These relevant genes were searched over a range of 500 Kb around each significant signal (250 Kb upstream and 250 Kb downstream the significant region). The dataset of annotated genes in cattle was derived from the UCSC Genome Browser (Baylor 4.0/bostau4) (http://genome-euro.ucsc.edu/index.html).

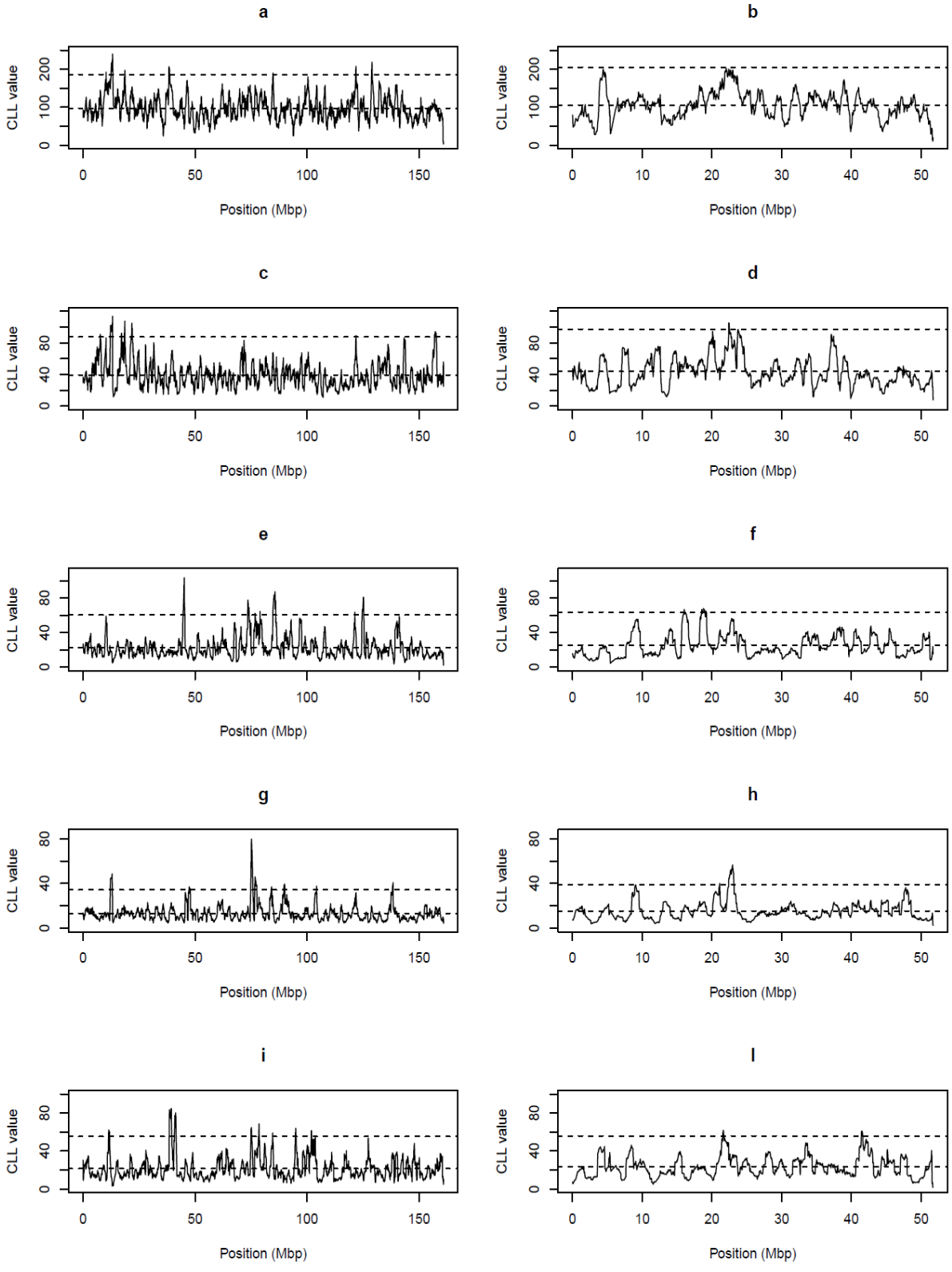


Figure 1 Pattern of the composite log-likelihood in chromosome 1 (graphs a, c, e, g and i) and in chromosome 26 (b, c, f, h and l). The breeds are divided in the following way: Italian Holstein graph a and b, Italian Brown c and d, Italian Simmental e and f, Marchigiana g and h and Piedmontese i and l.

RESULTS AND DISCUSSION

Considering the density of SNPs as a determinant factor in the calculation of CLL, windows of 1 Mb were used (Matukumalli et al., 2009). The chromosome average inter-marker distance in the whole genome and the maximum distance in the single autosome were detected. The average inter marker distance was approximately 57.7 Kb, ranging from 68.6 Kb in BTA 5 to 52.2 Kb in BTA 25, while the maximum inter-marker distance between adjacent markers (13,346,579 bp) was observed in BTA 10 and the lower inter-marker distance (343,9) was on BTA 4 (Table1). This density was considered adequate to represent the 3 billion bases of sequence in the bovine genome (Matukumalli et al., 2009).

Chromosome CLL averages obtained in this study decrease synchronously with the reduction of population size. For example on BTA6 CLL smoothed value was 105.2 for the 2,093 Holstein bulls, 49.5 for the 749 Italian Brown bulls, and quite similar values for Simmental and Marchigiana (22.9 and 21.2 respectively), and 15.4 for 364 Piedmontese bulls (Table 2). This trend was in agreement with Stella et al. (2010) where CLL values decrease with the reduction of subpopulation number.

The highest CLL value (323,3) was obtained for BTA 16, in Holstein (Table2). Moreover other high values were observed in Brown on BTA 5 (155). For beef breeds, a value equal to 200 was found in Marchigiana on BTA19, 126 for Italian Simmental on BTA16 and finally 162 for Piedmontese on BTA5 (Table 2). Furthermore, is interesting to note that the highest value of the maximum CLL obtained in Italian Holstein and in Italian Simmental (126,8) corresponding to the same genomic region (4,656,759 bp) on BTA16. Furthermore, in this chromosomic region, a CLL peak (113,8) was found also in Brown bulls (Table 2). For this breed this value was not the highest, but what intrigues is that this peak was in the same position in all three breeds (Table 2). Unfortunately, no annotated genes were found in this region in cattle genome. Similar results were obtained by other authors (Flori et al., 2009; Qanbari et al., 2011; Ramey et al., 2013).

Table 1. For each chromosome were reported total length in bp, the number of SNP present and the maximum value and the average of the distance between SNP

BTA	Length (bp)	SNP (N)	Mean (bp)	Maximum (bp)
1	161'021'444	2870	56'125	631'945
2	140'672'838	2363	59'557	614'304
3	127'908'629	2255	56'747	796'340
4	124'125'394	2158	57'545	353'978
5	125'804'605	1835	68'596	1'050'481
6	122'543'360	2214	55'374	826'195
7	112'064'213	1939	57'825	632'575
8	116'938'581	2059	56'821	738'260
9	107'962'209	1749	61'763	760'804
10	119'596'824	1875	63'819	13'346'579
11	110'120'689	1936	56'910	580'187
12	85'277'438	1423	59'970	859'164
13	84'344'187	1530	55'163	592'177
14	81'323'942	1480	54'986	575'966
15	84'598'267	1462	57'901	683'257
16	77'895'388	1366	57'065	1'015'396
17	76'454'249	1390	55'043	840'350
18	66'116'595	1171	56'510	896'406
19	65'213'966	1193	54'709	543'535
20	75'705'448	1382	54'819	837'057
21	69'171'298	1184	58'471	849'428
22	61'825'382	1108	55'849	532'751
23	53'329'482	932	57'282	482'114
24	64'945'342	1107	58'721	456'629
25	44'021'516	844	52'216	555'937
26	51'726'098	929	55'733	682'582
27	48'726'297	852	57'258	1'677'820
28	46'020'951	814	56'584	496'079
29	51'979'343	905	57'498	821'936

Table2. Summary of the average, standard deviation and maximum values of CLL for each chromosome within breeds.

	Italian Holstein		Italian Brown		Italian Simmental		Piedmontese		Marchigiana	
BTA	Mean ± SD	Maximum	Mean ± SD	Maximum	Mean ± SD	Maximum	Mean ± SD	Maximum	Mean ± SD	Maximum
1	97.1±29.6	240.2	39.4±15.9	113.6	22.5±12.6	103.7	13.1±7.0	79.6	21.3±11.3	84.8
2	99.3±30.6	263.0	42.5±17.5	131.7	21.4±10.2	66.7	14.2±7.8	80.4	26.0±17.7	174.2
3	103.4±31.5	240.0	44.9±19.2	152.0	20.6±10.9	119.7	14.1±6.7	47.0	21.0±10.1	59.8
4	96.5±32.4	229.3	40.8±16.8	120.4	23.7±13.0	107.7	13.2±6.5	75.2	22.7±12.1	91.5
5	100.1±30.8	247.3	47.9±21.4	155.3	25.0±11.4	100.3	15.8±9.9	162.7	23.1±11.6	90.7
6	105.2±35.2	284.5	49.5±17.5	118.1	23.9±11.2	92.9	15.9±8.5	57.3	24.1±11.7	69.5
7	100.7±32.4	225.4	42.7±19.1	131.7	21.2±10.9	75.7	15.4±7.6	49.8	22.9±12.6	118.7
8	97.8±32.5	225.3	44.2±17.4	118.7	20.4±9.1	70.2	13.6±6.7	48.2	26.1±15.1	110.9
9	94.1±28.9	190.7	40.6±18.6	103.5	20.1±9.7	66.9	12.8±6.1	44.3	23.2±12.3	143.6
10	106.1±34.1	246.0	40.0±13.9	95.1	20.6±9.6	67.7	12.1±4.7	43.9	22.9±12.6	102.5
11	93.3±30.4	210.3	44.9±19.1	128.4	22.6±9.4	65.2	12.9±6.2	43.8	22.7±11.2	68.3
12	92.3±26.1	172.8	42.6±15.9	102.8	21.4±9.3	76.8	13.1±8.0	73.3	22.6±13.0	66.0
13	103.4±30.8	202.7	46.3±17.2	106.0	20.8±10.1	71.9	15.6±7.7	79.6	27.2±14.0	103.6
14	109.7±37.7	271.5	43.7±17.1	152.1	22.7±11.3	85.5	14.9±7.7	56.6	22.4±10.1	62.2
15	87.1±26.1	184.8	36.5±14.7	109.7	21.0±10.2	68.9	11.2±4.9	35.1	20.2±13.7	118.7
16	103.2±33.0	323.3	43.6±17.0	113.8	27.0±13.4	126.8	14.1±5.2	35.7	24.3±12.6	70.6
17	100.1±30.1	203.9	44.0±14.2	109.4	21.9±10.5	79.2	12.7±5.1	33.1	18.9±8.8	59.1
18	96.2±25.1	196.3	43.1±18.5	115.9	23.9±11.7	90.1	12.4±4.8	49.8	23.6±11.7	81.4
19	93.1±29.7	181.5	38.8±13.2	92.1	19.9±7.8	48.5	13.6±6.9	55.3	24.2±20.0	200.4
20	112.5±33.4	225.4	47.1±17.8	114.8	21.9±7.8	50.7	14.9±7.1	62.4	25.8±12.8	97.9
21	95.4±34.3	239.3	37.3±15.4	114.6	21.3±10.6	68.1	14.6±9.0	88.6	23.0±12.6	101.6
22	99.2±33.0	216.1	44.8±16.2	109.0	22.6±9.1	67.9	14.1±5.8	49.3	20.5±12.4	90.1
23	93.5±31.7	221.6	40.5±14.6	109.5	18.7±7.3	43.8	13.1±6.5	41.0	19.9±10.5	61.8
24	96.5±30.2	267.0	42.2±18.1	138.2	20.6±10.2	75.2	14.5±7.4	42.0	24.9±13.9	76.6
25	90.6±24.4	215.0	40.2±14.0	95.3	18.4±7.6	49.5	10.2±3.0	22.5	20.1±8.2	53.9
26	105.5±33.0	202.5	43.8±17.7	105.3	25.3±12.7	68.0	15.2±7.8	56.5	23.5±10.7	61.8
27	86.6±29.3	187.1	37.8±14.6	90.3	19.4±7.6	53.9	11.3±4.6	33.3	21.1±11.4	72.4
28	92.9±23.6	157.5	41.3±17.9	100.5	21.3±10.2	57.6	12.3±5.7	36.7	17.9±8.2	56.5
29	97.6±34.3	192.0	42.1±16.9	108.7	19.0±8.7	56.1	13.4±7.2	51.9	20.9±13.5	182.8

Using the control chart statistical test, the greatest number of upper values (i.e. CLL values the exceed the upper limit of the control chart fixed at three standard deviations from the mean) per chromosome was found for the Italian Simmental population test (n = 60 for both BTA1 and 10) (Table3). This result could be ascribed to its dual purpose nature.

Table 3. Number of uppers identified for each chromosome in the five Italian cattle breeds

BTA	Italian Holstein	Italian Brown	Italian Simmental	Piedmontese	Marchigiana
1	22	49	60	53	57
2	10	32	47	45	54
3	4	27	38	35	27
4	10	21	36	16	31
5	11	20	18	28	34
6	28	22	21	56	38
7	10	25	30	33	35
8	10	28	26	57	44
9	5	8	35	47	19
10	11	20	60	25	36
11	5	21	22	41	27
12	1	4	19	31	13
13	5	5	17	21	26
14	11	24	24	22	7
15	6	22	21	32	33
16	16	6	21	10	15
17	6	14	32	17	16
18	4	27	21	15	30
19	-	15	12	38	24
20	5	7	11	33	21
21	11	18	12	16	21
22	4	4	21	11	32
23	14	10	2	16	12
24	9	15	18	14	5
25	6	10	9	8	7
26	-	3	11	15	6
27	6	7	12	20	19
28	-	6	4	17	16
29	3	18	12	8	20

Comparing the genome-wide number of upper CLL values for all five breeds, t PIE breed showed the largest number of significant regions (780), followed by, MAR (725), SIM (672) and IB (488). HOL shows the smallest number of significant regions (233) (Table3).

Moreover, no significant SNPs were found in chromosomes 19, 26 and 28, (Table 3) in this breed. These results were not in agreement with other similar studies where several significant signatures of selection were found for dairy and beef cattle on these BTAs (Prasad et al., 2008; Stella et al., 2010).

Table 4. Significant markers common among breeds with similar productive aptitude

SNP Name	Position	BTA	Italian Holstein	Italian Brown	Italian Simmental	Piedmontese	Marchigiana
Hapmap43215-BTA-24270	13'106'799	1	×	+			
ARS-BFGL-BAC-685	13'137'364	1	×	+			
BTB-01890165	18'662'822	1	×	+			
ARS-BFGL-NGS-23874	121'900'729	1	×	+			
Hapmap33033-BTA-149123	95'823'711	2	×	+			
Hapmap50837-BTA-98392	50'589'135	5	×	+			
BTA-74203-no-rs	86'756'396	5	×	+			
Hapmap25632-BTA-23506	87'218'180	5	×	+			
Hapmap54103-rs29010895	37'433'108	6	×	+			
BTA-121739-no-rs	37'454'410	6	×	+			
Hapmap27072-BTC-033816	37'524'839	6	×	+			
Hapmap27503-BTC-033786	37'564'380	6	×	+			
Hapmap33288-BTC-033751	37'594'439	6	×	+			
Hapmap26555-BTC-033429	37'624'773	6	×	+			
Hapmap26258-BTC-033509	37'647'259	6	×	+			
Hapmap26259-BTC-033526	37'669'926	6	×	+			
Hapmap26308-BTC-057761	37'963'148	6	×	+			
ARS-BFGL-NGS-112812	38'014'255	6	×	+			
ARS-BFGL-NGS-4595	107'597'162	9	×	+			
ARS-BFGL-NGS-42329	107'622'501	9	×	+			
ARS-BFGL-NGS-101642	107'678'537	9	×	+			
BTB-01953819	22'019'956	14	×	+			
Hapmap45796-BTA-25271	22'056'944	14	×	+			
ARS-BFGL-BAC-8052	22'096'518	14	×	+			
ARS-BFGL-NGS-104268	22'260'372	14	×	+			
BTA-91250-no-rs	22'346'857	14	×	+			
BTB-01417924	22'382'726	14	×	+			
BTB-01530788	22'720'373	14	×	+			
BTB-01199899	4'656'759	16	×	+			
ARS-BFGL-NGS-54181	57'874'904	18	×	+			
ARS-BFGL-BAC-30737	3'830'272	23	×	+			
ARS-BFGL-NGS-33785	64'240'935	24	×	+	●		
Hapmap42596-BTA-58793	64'279'403	24	×	+	●		

BTB-00893432	64'301'665	24	×	+	●		
ARS-BFGL-NGS-11659	64'469'668	24	×	+	●		
BTB-01890193	64'808'820	24	×	+	●		
ARS-BFGL-NGS-109285	57'125'868	18			●	■	▲
BTA-35627-no-rs	75'013'836	1				■	▲
BTA-35628-no-rs	75'042'636	1				■	▲
UA-IFASA-4774	75'104'988	1				■	▲
ARS-BFGL-NGS-114627	75'135'242	1				■	▲
BTA-35631-no-rs	75'180'498	1				■	▲
ARS-BFGL-NGS-64511	75'221'451	1				■	▲
ARS-BFGL-NGS-7099	75'246'290	1				■	▲
ARS-BFGL-NGS-112477	22'831'437	10				■	▲
ARS-BFGL-NGS-20828	26'703'980	11				■	▲
ARS-BFGL-BAC-13009	44'268'740	11				■	▲
ARS-BFGL-NGS-25464	44'309'152	11				■	▲
ARS-BFGL-NGS-117681	44'403'089	11				■	▲
BTA-93084-no-rs	44'458'293	11				■	▲
ARS-BFGL-NGS-33705	29'929'537	25				■	▲
BTA-59806-no-rs	29'965'579	25				■	▲
ARS-BFGL-NGS-106207	35'408'558	27				■	▲

In general, dairy breeds hold less outliers signals than beef breeds and the dual purpose breed is positioned between the two. This trend was probably due to distinct causes: different selection pressure operated in dairy and beef cattle or phenomena such as ascertainment bias, effective population size, artificial insemination and inbreeding (MacEachern et al., 2009).

In order to see if any genomic regions were flagged as selective sweeps for both dairy and beef breeds, a cross-check of the outlier signals was carried out for all five populations.

A total of 31 significant markers distributed on BTAs 1, 2, 5, 6, 9, 14, 16, 18, 23 were found to be specific for dairy breeds (HOL and IB) (Table 4). Moreover, five markers, between 64'240'935-64'808'820 bp on BTA24, were also in common with SIM. A low number of significant regions were found in common for beef breeds (PIE and MAR), where only 16 markers (on BTA 1, 10, 11, 25 and 27) exceed the threshold value. Only one of these SNP (ARS-BFGL_NGS-109285) on BTA18 was in common between the beef and the dual-purpose (Table 4) breed.

In this study, a total of 2,152 genes were identified. Many of these genes were shared between two or more breeds. In Table 5 was reported a list of genes common to breeds that have the same productive aptitude. The large number of genes was found for beef breeds.

In fact 653 genes were found in MAR and 587 in PIE, 559 were detected in SIM, and 488 and 259 in HOL and IB respectively. Just a limited number of these of genes was found to be in common between dairy and beef breeds. Actually, only 32 genes were shared by Italian Holstein and Italian Brown. Most of them were not genes belonging to biological pathways related to the lactation process but were involved in more general functions (i.e. DNA repair, cell cycle). A total of eighteen genes was obtained comparing dairy and dual purpose breeds. Among these genes (located on BTA24) of particular interest were some members of the serpin family. This y is a large group of proteins involved in the regulation of inflammatory reactions primarily isolated from bovine milk (Christensen and Sottrup-Jensen, 1994) and more recently from human breast cancer cells (Tseng et al., 2008).

Table 5. List of genes identified in the races with the same productive attitude.

Gene	Gene Start	Gene End	BTA	Italian Holstein	Italian Brown	Italian Simmental	Piedmontese	Marchigiana
CHODL	18'403'996	18'428'296	1	×	+			
WDR12	95'927'100	95'951'629	2	×	+			
ALS2CR8	95'964'239	96'106'243	2	×	+			
CCDC91	87'381'914	87'781'773	5	×	+			
HERC6	37'128'404	37'185'776	6	×	+			
PPM1K	37'268'108	37'290'154	6	×	+			
ABCG2	37'351'167	37'421'683	6	×	+			
DCAF16	38'141'492	38'142'137	6	×	+			
NCAPG	38'153'046	38'199'153	6	×	+			
LCORL	38'227'954	38'378'385	6	×	+			
SOX17	22'087'981	22'089'782	14	×	+			
RP1	22'193'802	22'202'968	14	×	+			
TMEM68	22'891'754	22'927'546	14	×	+			
TGS1	22'927'619	22'953'424	14	×	+			
PLOD1	38'460'816	38'489'937	16	×	+			
KIAA2013	38'495'183	38'500'811	16	×	+			
NPPB	38'564'774	38'566'162	16	×	+			
NPPA	38'576'503	38'577'584	16	×	+			
CLCN6	38'583'732	38'614'517	16	×	+			
MTHFR	38'614'885	38'629'401	16	×	+			
AGTRAP	38'670'145	38'689'094	16	×	+			
C16H1orf187	38'703'762	38'731'376	16	×	+			
MAD2L2	38'738'475	38'744'041	16	×	+			
FBXO6	38'744'134	38'761'640	16	×	+			
FBXO44	38'763'865	38'767'919	16	×	+			

FBXO2	38'769'087	38'775'125	16	×	+			
MGST2	19'126'124	19'164'411	17	×	+			
KLC1	68'560'789	68'622'104	21	×	+			
XRCC3	68'625'204	68'632'763	21	×	+			
ZFYVE21	68'641'768	68'644'916	21	×	+			
C21H14orf2	68'777'641	68'782'998	21	×	+			
PRIM2	3'723'486	4'228'127	23	×	+			
C9H6orf120	107'457'743	107'460'354	9	×	+	●		
PHF10	107'459'423	107'479'560	9	×	+	●		
TCTE3	107'489'317	107'512'066	9	×	+	●		
C9H6orf70	107'515'366	107'540'924	9	×	+	●		
MRPL41	109'897'287	109'897'928	11	×	+	●		
ARRDC1	109'943'705	109'951'230	11	×	+	●		
EHMT1	109'998'749	110'062'934	11	×	+	●		
CACNA1B	110'097'644	110'171'688	11	×	+	●		
BCL2	64'011'851	64'217'313	24	×	+	●		
KDSR	64'228'468	64'267'079	24	×	+	●		
VPS4B	64'285'230	64'303'567	24	×	+	●		
SERPINB4	64'405'108	64'412'083	24	×	+	●		
SERPINB5	64'641'559	64'642'812	24	×	+	●		
LOC511106	64'710'232	64'752'991	24	×	+	●		
LOC786410	64'726'264	64'732'444	24	×	+	●		
SERPINB7	64'813'913	64'832'561	24	×	+	●		
SERPINB2	64'893'037	64'910'799	24	×	+	●		
SERPINB10	64'917'121	64'940'502	24	×	+	●		
KLK4	56'882'151	56'885'860	18			●	■	▲
KLK5	56'912'103	56'921'495	18			●	■	▲
KLK6	56'924'932	56'933'446	18			●	■	▲
KLK7	56'940'059	56'945'285	18			●	■	▲
KLK8	56'951'636	56'956'819	18			●	■	▲
KLK9	56'959'498	56'965'590	18			●	■	▲
KLK10	56'968'363	56'974'337	18			●	■	▲
KLK12	56'981'756	56'985'707	18			●	■	▲
CTU1	57'046'105	57'054'632	18			●	■	▲
CD33	57'110'782	57'120'384	18			●	■	▲
IGLON5	57'300'652	57'313'996	18			●	■	▲
ETFB	57'333'560	57'346'050	18			●	■	▲
CLDND2	57'346'839	57'348'122	18			●	■	▲
NKG7	57'350'249	57'352'981	18			●	■	▲
LIM2	57'357'208	57'364'348	18			●	■	▲
SIGLEC10	57'372'512	57'380'066	18			●	■	▲
OPA1	75'162'499	75'253'869	1				■	▲
KIT	72'741'252	72'828'528	6				■	▲

MTA3	26'357'081	26'454'698	11				■	▲
HAAO	26'467'962	26'486'655	11				■	▲
ZFP36L2	26'868'769	26'873'023	11				■	▲
ASL	29'676'856	29'685'629	25				■	▲
CRCP	29'684'671	29'723'903	25				■	▲
ERLIN1	21'341'902	21'383'677	26				■	▲
ZNF703	35'200'987	35'203'749	27				■	▲
ERLIN2	35'238'999	35'256'138	27				■	▲
PROSC	35'264'856	35'274'389	27				■	▲
BRF2	35'325'842	35'330'168	27				■	▲
GOT1L1	35'404'416	35'410'154	27				■	▲
ADRB3	35'431'727	35'434'819	27				■	▲
EIF4EBP1	35'470'744	35'494'090	27				■	▲
ASH2L	35'511'961	35'537'791	27				■	▲
STAR	35'539'114	35'546'541	27				■	▲
LSM1	35'564'657	35'574'870	27				■	▲
BAG4	35'575'213	35'600'749	27				■	▲
DDHD2	35'613'436	35'642'363	27				■	▲
WHSC1L1	35'652'529	35'750'914	27				■	▲

A smaller number of genes was found to be in common between beef breeds. A total of 21 putative candidate common genes were shared by PIE and MAR, whereas among the beef breeds and SIM 16 genes were found (Table 5). Among these genes some members of a class of endonucleases, the Kallikreins (KLKs) were highlihted. In mammals these genes are classified as serine peptidases (Yousef and Diamandis, 2001). It is generally assumed that the softening of the myofibrillar structure is performed by proteolitic enzymes such as calpains, cathepsins and serine peptidases (Sentandreu et al., 2002). Meat tenderness is one of the most important traits required by consumers. Therefore, is useful to know and understand the complex biological mechanisms underlying the process of meat tenderization (Warner et al., 2010).

Finally, in order to confirm the validity of the current study, selection signatures for genes known to affect production traits in cattle were searched. In the appendix are reported the plots of control chart of CLL values for each chromosome and breed considered in the study. On BTA 6, the presence of selection signature was detected in the region spanning between 37-38 Mb, where a known gene cluster involved in milk production is located. A polymorphism at the *ATP-binding cassette, sub-family G (WHITE), member 2* (*ABCG2*) locus causes a decrease in milk yield and an in increase in the of protein and fat concentration (Ron et al., 2006). Using CLL this gene was highlighted for the two dairy

cattle breeds, Italian Holstein and Italian Brown. Several studies have already shown the presence of a QTL for milk production traits in Holstein and Brown (Cohen-Zinder et al., 2005; Ron et al., 2006). In the same gene cluster there are also two loci, the *non-SMC condensin I complex, subunit G* (*NCAPG)* and the *ligand dependent nuclear receptor corepressor-like (LCORL)*. The last one have been recently associated with hip axis length variation in human (Soranzo et al., 2009). In cattle Flori et al., 2009 has proposed a role for *LCORL* in the pelvic morphology. In addition Bongioni et al. (2012) found that *NCAPG* is directly involved in calving ease in Piedmontese cattle breeds.

For the two beef breeds (Piedmontese and Marchigiana) on BTA6 at 72 Mb the gene encoding for the v-*kit* Hardy-Zuckerman *4 feline sarcoma viral oncogene homolog (KIT)* was revealed. In animal species *KIT* has been associated in with the coat colour determinism (Marklund et al., 1999; Brenig et al., 2013). In cattle this gene is responsible for the "piebald" spotted coat-colour pattern (Grosz and MacNeil, 1999). Also Stella et al., (2010) reported a strong selection signature in this region but the breeds involved were predominantly selected for milk production. Results proposed for this gene by the literature are conflicting, strengthening the hypothesis of a more complex role for this locus in the biology of bovine (Koch et al., 2009; Lemay et al., 2009). In the present study, strong selection signatures were observed in the initial portion of the chromosome BTA2 in Piedmontese. In this region the *MSTN* (*myostatin*) locus is located. This gene is a component of the transforming growth factor-β super family and plays a role as negative regulator of skeletal muscle mass (McPherron et al., 1997). Animals that possess an alternative form of the gene display a muscular hypertrophy. Polymorphisms in this gene have already been reported for the beef breeds that show a phenotype commonly known as double muscling (Kambadur et al., 1997; Marchitelli et al., 2003; Grisolia et al., 2009). Still analyzing the signs of selection on bovine chromosome 2 another known gene is highlighted. At 64 Mb is contained a gene responsible for the digestion of lactose in mammals. This locus is called *lactase (LCT)* and was found to be under selection in human European populations (Tishkoff et al., 2007). In cattle this selective sweep suggest that mutations in this region may affect energy homeostasis (Barendse et al., 2009; Gibbs et al., 2009).

Expected results that were not obtained in the present study deal with two genes widely studied. The *melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)* (*MC1R*) on BTA18 and the *diacylglycerol O-acyltransferase 1 (DGAT1*) on BTA14. In

cattle, MC1R gene is involved in the coat and the membranes pigmentation being responsible for the production of eu- and pheomelanin (Klungland and Vage, 2000). This polymorphic gene (Olson 1999) is mainly investigated to define the genetic origin of livestock productions (Maudet and Taberlet, 2002; Russo et al., 2007). The absence of significant signals for this gene was unexpected because the five considered breed exhibit different coat color (Crepaldi et al., 2003). Also DGAT1, well known to milk production traits (Hayes et al., 2008; Kaupe et al., 2004; Cohen et al., 2004), was expected. Different hypotheses can be formulated to explain these incongruence such as the complex genetic interactions or the density of the markers used for this type of study (Hayes et al., 2008; Qanbari et al., 2011).

CONCLUSIONS

Our results provide a genome-wide map sight of selection signatures in five Italian cattle breeds selected for different productive capacity. Significant regions depicted after CLL analysis seem to play prominent roles in economically important traits in dairy and beef cattle and may be useful as a starting point for the formulation of biological hypotheses. However, the results obtained in this study lead to the conclusion that for a complete description of the selection signatures in a species is not enough pick out two populations with divergent phenotype but it is necessary to analyze a large number of populations in order to confirm the genomic regions that emerge in common. In addition this study also confirms the need to build high-density maps of markers in order to capture for a trait all the genetic variance and to get a more accurate prediction of the animal genetic value.

REFERENCES

Akey, J. M. 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Research, 19, 711-722.

Andersson, L. 2012. How selective sweeps in domestic animals provide new insight into biological mechanisms. Journal of Internal Medicine, 271, 1-14.

Andersson, L. & Georges, M. 2004. Domestic-animal genomics: Deciphering the genetics of complex traits. Nature Reviews Genetics, 5, 202-212.

Bamshad, M. & Wooding, S. P. 2003. Signatures of natural selection in the human genome. Nature Reviews Genetics, 4, 99-111.

Barendse, W., Harrison, B. E., Bunch, R. J., Thomas, M. B. & Turner, L. B. 2009. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. Bmc Genomics, 10.

Biswas, S. & Akey, J. M. 2006. Genomic insights into positive selection. Trends in Genetics, 22, 437-446.

Bongiorni, S., Mancini, G., Chillemi, G., Pariset, L. & Valentini, A. 2012. Identification of a Short Region on Chromosome 6 Affecting Direct Calving Ease in Piedmontese Cattle Breed. Plos One, 7.

Brenig, B., Beck, J., Floren, C., Bornemann-Kolatzki, K., Wiedemann, I., Hennecke, S., Swalve, H. & Schutz, E. 2013. Molecular genetics of coat colour variations in White Galloway and White Park cattle. Animal Genetics, 44, 450-3.

Christensen, S. & Sottrup-Jensen, L. 1994. Characterization of two serpins from bovine plasma and milk. Biochemical Journal, 303, 383-390.

Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loor, J. J., Everts-Van Der Wind, A., Lee, J. H., Drackley, J. K., Band, M. R., Hernandez, A. G., Shani, M., Lewin, H. A., Weller, J. I. & Ron, M. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Research, 15, 936-944.

Cohen, M., Reichenstein, M., Everts-Van Der Wind, A., Heon-Lee, J., Shani, M., Lewin, H. A., Weller, J. I., Ron, M. & Seroussi, E. 2004. Cloning and characterization of FAM13A1 - A gene near a milk protein QTL on BTA6: Evidence for population-wide linkage disequilibrium in Israeli Holsteins. Genomics, 84, 374-383.

Crepaldi, P., Marilli, M., Gorni, C., Meggiolaro, D., Cicogna, M. & Renieri, C. 2003. Preliminary study on MC1R polymorphism in some cattle breeds raised in Italy. Italian Journal of Animal Science, 2, 13-15.

Flori, L., Fritz, S., Jaffrezic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.-L. & Gautier, M. 2009. The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. Plos One, 4.

Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversoie, K. A., Gill, C. A., Green, R. D., Hamernik, D. L., Kappes, S. M., Lien, S., Matukumalli, L. K., McEwan, J. C., Nazareth, L. V., Schnabel, R. D., Weinstock, G. M., Wheeler, D. A., Ajmone-Marsan, P., Boettcher, P. J., Caetano, A. R., Garcia, J. F., Hanotte, O., Mariani, P., Skow, L. C., Williams, J. L., Diallo, B., Hailemariam, L., Martinez, M. L., Morris, C. A., Silva, L. O. C., Spelman, R. J., Mulatu, W., Zhao, K., Abbey, C. A., Agaba, M., Araujo, F. R., Bunch, R. J., Burton, J., Gorni, C., Olivier, H., Harrison, B. E., Luff, B., Machado, M. A., Mwakaya, J., Plastow, G., Sim, W., Smith, T., Sonstegard, T. S., Thomas, M. B., Valentini, A., Williams, P., Womack, J., Wooliams, J. A., Liu, Y., Qin, X., Worley, K. C., Gao, C., Jiang, H., Moore, S. S., Ren, Y., Song, X.-Z., Bustamante, C. D., Hernandez, R. D., Muzny, D. M., Patil, S., Lucas, A. S., Fu, Q., Kent, M. P., Vega, R., Matukumalli, A., McWilliam, S., Sclep, G., Bryc, K., Choi, J., Gao, H., Grefenstette, J. J., Murdoch, B., Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Goddard, M. E., Hayes, B. J., Bradley, D. G., da Silva, M. B., Lau, L. P. L., Liu, G. E., Lynn, D. J., Panzitta, F. & Dodds, K. G. 2009. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. Science, 324, 528-532.

Grisolia, A. B., D'Angelo, G. T., Porto Neto, L. R., Siqueira, F. & Garcia, J. F. 2009. Myostatin (GDF8) single nucleotide polymorphisms in Nellore cattle. Genetics and Molecular Research, 8, 822-830.

Grosz, M. D. & MacNeil, M. D. 1999. The 'spotted' locus maps to bovine chromosome 6 in a Hereford-cross population. Journal of Heredity, 90, 233-236.

Hayes, B. J., Lien, S., Nilsen, H., Olsen, H. G., Berg, P., Maceachern, S., Potter, S. & Meuwissen, T. H. E. 2008. The origin of selection signatures on bovine chromosome 6. Animal Genetics, 39, 105-111.

Kambadur, R., Sharma, M., Smith, T. P. L. & Bass, J. J. 1997. Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. Genome Research, 7, 910-916.

Kaupe, B., Winter, A., Fries, R. & Erhardt, G. 2004. DGAT1 polymorphism in Bos indicus and Bos taurus cattle breeds. Journal of Dairy Research, 71, 182-187.

Klungland, H. & Vage, D. I. 2000. Molecular genetics of pigmentation in domestic animals. Current Genomics, 1, 223-242.

Koch, D., Sakurai, M., Hummitzsch, K., Hermsdorf, T., Erdmann, S., Schwalbe, S., Stolzenburg, J. U., Spanel-Borowski, K. & Ricken, A. M. 2009. KIT variants in bovine ovarian cells and corpus luteum. Growth Factors, 27, 100-113.

Kosiol, C., Vinař, T., Da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R. & Siepel, A. 2008. Patterns of positive selection in six mammalian genomes. Plos Genetics, 4.

Lemay, D. G., Lynn, D. J., Martin, W. F., Neville, M. C., Casey, T. M., Rincon, G., Kriventseva, E. V., Barris, W. C., Hinrichs, A. S., Molenaar, A. J., Pollard, K. S., Maqbool, N. J., Singh, K., Murney, R., Zdobnov, E. M., Tellam, R. L., Medrano, J.

F., German, J. B. & Rijnkels, M. 2009. The bovine lactation genome: Insights into the evolution of mammalian milk. Genome Biology, 10.

Lenstra, J. A., Groeneveld, L. F., Eding, H., Kantanen, J., Williams, J. L., Taberlet, P., Nicolazzi, E. L., Soelkner, J., Simianer, H., Ciani, E., Garcia, J. F., Bruford, M. W., Ajmone-Marsan, P. & Weigend, S. 2012. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. Animal Genetics, 43, 483-502.

MacEachern, S., Hayes, B., McEwan, J. & Goddard, M. 2009. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (Bos taurus) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. Bmc Genomics, 10.

Marchitelli, C., Savarese, M. C., Crisà, A., Nardone, A., Marsan, P. A. & Valentini, A. 2003. Double muscling in Marchigiana beef breed is caused by a stop codon in the third exon of myostatin gene. Mammalian Genome, 14, 392-395.

Marklund, S., Moller, M., Sandberg, K. & Andersson, L. 1999. Close association between sequence polymorphism in the KIT gene and the roan coat color in horses. Mammalian Genome, 10, 283-288.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P., Sonstegard, T. S. & Van Tassell, C. P. 2009. Development and characterization of a high density SNP genotyping assay for cattle. Plos One, 4, e5350.

Maudet, C. & Taberlet, P. 2002. Holstein's milk detection in cheeses inferred from melanocortin receptor 1 (MC1R) gene polymorphism. Journal of Dairy Science, 85, 707-715.

McPherron, A. C., Lawler, A. M. & Lee, S. J. 1997. Regulation of skeletal muscle mass in mice by a new TGF-β superfamily member. Nature, 387, 83-90.

Nielsen, R. 2005. Molecular signatures of natural selection. Annual Review of Genetics.

Olson T.A. (1999) Genetics of colour variation. In: R. Fries, A. Ruvinsky (Eds); The genetics of the cattle. CABI Publishing, Wallingford, UK, pp. 33-53.

Pintus, E., Sorbolini, S., Albera, A., Gaspa, G., Dimauro, C., Steri, R., Marras, G. & Macciotta, N. P. P. 2013. Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds. Animal Genetics.

Prasad, A., Schnabel, R. D., McKay, S. D., Murdoch, B., Stothard, P., Kolbehdari, D., Wang, Z., Taylor, J. F. & Moore, S. S. 2008. Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. Animal Genetics, 39, 597-605.

Przeworski, M., Coop, G. & Wall, J. D. 2005. The signature of positive selection on standing genetic variation. Evolution, 59, 2312-2323.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G. & Simianer, H. 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. Bmc Genomics, 12.

Ramey, H. R., Decker, J. E., McKay, S. D., Rolf, M. M., Schnabel, R. D. & Taylor, J. F. 2013. Detection of selective sweeps in cattle using genome-wide SNP data. Bmc Genomics, 14.

Ron, M., Cohen-Zinder, M., Peter, C., Weller, J. I. & Erhardt, G. 2006. Short communication: A polymorphism in ABCG2 in Bos indicus and Bos taurus cattle breeds. Journal of Dairy Science, 89, 4921-4923.

Russo, V., Fontanesi, L., Scotti, E., Tazzoli, M., Dall'Olio, S. & Davoli, R. 2007. Analysis of melanocortin 1 receptor (MC1R) gene polymorphisms in some cattle breeds: Their usefulness and application for breed traceability and authentication of Parmigiano Reggiano cheese. Italian Journal of Animal Science, 6, 257-272.

Sentandreu, M. A., Coulis, G. & Ouali, A. 2002. Role of muscle endopeptidases and their inhibitors in meat tenderness. Trends in Food Science & Technology, 13, 400-421.

Smith, J. M. & Haigh, J. 1974. The hitch-hiking effect of a favourable gene. Genetical research, 23, 23-35.

Soranzo, N., Rivadeneira, F., Chinappen-Horsley, U., Malkina, I., Richards, J. B., Hammond, N., Stolk, L., Nica, A., Inouye, M., Hofman, A., Stephens, J., Wheeler, E., Arp, P., Gwilliam, R., Jhamai, P. M., Potter, S., Chaney, A., Ghori, M. J. R., Ravindrarajah, R., Ermakov, S., Estrada, K., Pols, H. A. P., Williams, F. M., McArdle, W. L., Van Meurs, J. B., Loos, R. J. F., Dermitzakis, E. T., Ahmadi, K. R., Hart, D. J., Ouwehand, W. H., Wareham, N. J., Barroso, I., Sandhu, M. S., Strachan, D. P., Livshits, G., Spector, T. D., Uitterlinden, A. G. & Deloukas, P. 2009. Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. Plos Genetics, 5.

Stella, A., Ajmone-Marsan, P., Lazzari, B. & Boettcher, P. 2010. Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. Genetics, 185, 1451-U498.

Stephan, W., Song, Y. S. & Langley, C. H. 2006. The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics, 172, 2647-2663.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A. & Deloukas, P. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genetics, 39, 31-40.

Tseng, I. C., Chou, F.-P., Su, S.-F., Oberst, M., Madayiputhiya, N., Lee, M.-S., Wang, J.-K., Sloane, D. E., Johnson, M. & Lin, C.-Y. 2008. Purification from human milk of matriptase complexes with secreted serpins: mechanism for inhibition of matriptase other than HAI-1. American Journal of Physiology-Cell Physiology, 295, C423-C431.

Warner, R. D., Greenwood, P. L., Pethick, D. W. & Ferguson, D. M. 2010. Genetic and environmental effects on meat quality. Meat Science, 86, 171-183.

Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M. & Hill, W. G. 2005. Measures of human population structure show heterogeneity among genomic regions. Genome Research, 15, 1468-1476.

Wiener, P. & Wilkinson, S. 2011. Deciphering the genetic basis of animal domestication. Proceedings of the Royal Society B-Biological Sciences, 278, 3161-3170.

Yousef, G. M. & Diamandis, E. P. 2001. The new human tissue kallikrein gene family: Structure, function, and association to disease. Endocrine Reviews, 22, 184-204.

CHAPTER 4

Analysis Of Distribution Of Runs Of Homozygosity And Of Their Relationship With Inbreeding In Five Cattle Breeds Farmed In Italy

Analysis Of Distribution Of Runs Of Homozygosity And Of Their Relationship With Inbreeding In Five Cattle Breeds Farmed In Italy

ABSTRACT

The increase of inbreeding levels in selected livestock populations somehow unavoidable. Genotyping animals at tens of thousands Single Nucleotide Polymorphism (SNP) permitted to identify long and uninterrupted stretches of homozygous genotypes called Run of Homozygosity (ROH) can be used to measure autozygosity. The objective of this work is to describe the distribution of ROH of different length classes in five Italian cattle breed. A total of 4095 bulls of 5 cattle breed were genotyped at 54 K SNP: 2093 Holstein, 749 Brown, 364 Piedmontese, 410 Marchigiana and 479 Simmental. ROH data were then used to compute molecular inbreeding coefficients and to compare them to the pedigree based coefficients.

A total of 335,985 ROH were detected across five breeds. The average number of ROH per animals ranged from 54 of Piedmontese to 94.6 of Brown. The high number of short ROH were found in Piedmontese, followed by Simmental. The opposite was found for dairy Breeds. Moreover Simmental presented the highest percentage of genome covered by short ROH (in Mb) which are related to ancient consanguinity event. Conversely, the genome of Brown and Holstein was covered by an higher proportion of longer ROH, related to recent inbreeding. The inbreeding coefficients computed using ROH data (FROH) were moderately correlated with pedigree based inbreeding coefficient (up to 0.765 in Simmental). The regression between traditional inbreeding coefficient of FROH of different minimum length (from >1 Mb to >16 Mb) seem to confirm the hypothesis that including ROH < 4 Mb in the inbreeding calculation generate an overestimation of this coefficient using 54 K SNP panel

INTRODUCTION

Inbreeding definition, its measure and consequences

The detrimental effects of parental relatedness on progeny performances and fitness have long been identified both in natural and selected animal populations (Keller 2002, Adams et al. 2006, Charlesworth and Willis 2009). The inbreeding coefficient (F) is defined as the probability that in a locus sampled randomly in a population a pair of alleles is *identical by descend* (IBD). Pedigree based definition of inbreeding dates long time ago, when Sewall Wright (1922) proposed to compute the abovementioned probability based on path coefficient method.

The increase of inbreeding levels in livestock populations results in loss of genetic variation and in the occurrence of inbreeding depression, with the main consequences of an increased prevalence of rare lethal genetic disorders (e.g. BLAD, DUMPS, in cattle) and a reduction of profitability of farm animals (Smith et al., 1998, Bjelland et al., 2013). Whereas the first aspect deals with the viability of the animals, the second one is mainly related to the decrease of animal productive performances and to a worsening of the reproductive efficiency (Smith et al., 1998, González-Recio et al., 2007). In a population under selection (especially in closed or small populations) inbreeding is somehow unavoidable. Thus, suitable breeding strategies must be implemented to limit the increase of inbreeding in the newborn animals. An approach is represented by the optimal contribution selection, based on classical tools of quantitative genetics theory: the objective is to optimize the response to selection while constraining the inbreeding at an acceptable level per generation (Meuwissen, 1997; Kearney et al., 2004).

Currently, genotyping animals up to hundreds of thousands Single Nucleotide Polymorphism (SNP) has become relatively cheap, due to the recent development of high-throughput genomic tools such as 7K, 54K and 800K SNP Bead Chip for cattle (Illumina Inc., San Diego, Ca). The availability of such technologies opened new scenarios in the genetic evaluation system, through the implementation of genomic selection procedures (Meuwissen et al., 2001). Furthermore, these new tools may be used to manage the inbreeding rate in livestock population, implementing optimal contribution selection on genomic basis (Sonesson et al., 2010, Nielsen et al., 2011; Pryce et al., 2012; Clark et al., 2013). In particular, dense SNP panels can be used to calculated molecular inbreeding coefficients in different fashions. For instance, VanRaden et al., (2011) computed

inbreeding coefficients from the genomic relationship matrix (GRM) based on the variance of genotypic values across the whole genome. Alternatively, the adoption of the concept of Run of Homozigosity (ROH) may allow to obtain an estimates of the inbreeding coefficient for an individual. Genomic based inbreeding metrics are particularly useful in case of missing, incomplete or not particularly depth pedigrees, but also they may help to provide an improved estimate of the inbreeding coefficients.

A Measure of Autozigosity derived from SNP data: Run of Homozigosity

ROH are defined as DNA segments that harbor uninterrupted sequences of homozygous genotypes. They are interpreted as a measure of autozigosity at genome-wide level (Gibson et al., 2006). It is called autozigosity the condition that occur at a specific locus when the two copies of an ancestral haplotype, that come together in an individual, are IBD (Hart and Clark 1997). The occurrence of ROH in an individual may be the result of mating of closely related individuals (Inbreeding components) but they may also the result of Linkage Disequilibrium (LD) in the population (i.e. homozygous fragments are identical by state (IBS) in the population). Thus, for one individual the inbreeding coefficient can also be seen as the proportion of genome that is IBD. Broman and Weber (1999) first identified that most of ROH are likely to be autozyous and proposed a statistical method based on LOD score to detect them using about 8,000 Short Tandem-Repeat polymorphism (STRP) on 134 subject in human populations. According to Clark (1999) an homozygous segment may have been originated in different fashions. Hence, he defined: *i*) "unrecombined autozygous segment" the DNA segments passed without undergo any recombination event from a common ancestor: *ii*) "autozygous segment" a chunk of DNA that may have recombined apart and recombined back. *iii*) "homozygous segment" is a run of DNA that is homozygous in the two homologous chromosomes, that may derive from multiple recombination or mutation and not necessarily is autozygous; *iv*) "apparent homozygous segments", those segments that are homozygous due to scoring error during genotyping process. The identification of the *unrecombined autozygous segments* is required to measure the inbreeding of an individual using whole genome-wide marker information, given that not all the homozygous segments in the genome are necessarily IBD. A possible limitation derived from analyzing ROH data is the lack of guideline to defining ROH (Howrigan et., 2011). In fact, different softwares have been proposed to map ROH in the genome, based on slightly different input parameters and

procedures. Howrigan et al. (2011), used three of these software on simulated data to test their ability to correctly detect the *unrecombined autozygous segments* excluding the fragments that are not autozygous.

Once the ROH have been detected, they can be used in different ways. One of the tasks is to compute genome based inbreeding coefficients. Some metrics have been proposed to detect IBD regions in the genome of individuals using SNP data. However, the simplest approach is summing across the whole genome the length of DNA that is in ROH and dividing this summation by the total length of the genome (McQuillan et al., 2008). While using pedigree information the probability that a pairs of allele is IBD is measured by its expectation, that is derived from the relationships among the individuals in the pedigree, in the case of SNP data this probability can be derived in a more precise way, and theoretically the actual values of inbreeding can be fully detected from sequence data. Keller et al., (2011) suggested to use metrics derived from ROH to predict accurately autozygosity and inbreeding effect.

Usefulness of ROH in animal breeding and formulas to calculate molecular inbreeding

In human genetics field ROH have been associated to higher prevalence of complex disease (Lencz et al., 2007, Szpiech, et al., 2013) and they have been used to map the recessive variant of many other disorders with high density SNP panel. Ku et al., (2011) reviewed the most recent paper on the use of ROH in human genetics.

In cattle genetics, few works dealt with this topic using both 54K and HD SNP chip panels. In particular, ROH have been used: to analyze population history of recent selection (Purfield et al., 2012); to estimate inbreeding coefficients (Ferencakovic et al., 2011; 2012; 2013); to study the detrimental effect of inbreeding on trait affecting farm profitabiliy (Bjelland et al 2013) and to control the inbreeding rate in genomic breeding scheme (Pryce et al., 2012).

As far as the inbreeding level is concerned, regions of autozyosity are expected to harbor in a higher frequency genes associated to inbreeding depression. These regions have been reported to have negative effects on productive and reproductive performances in the US dairy cattle population (Bjelland et al., 2013) and in pig growth performances (Siliò et al., 2013). The inbreeding depression is a very well-known phenomenon in animal breeding. The relation between classical measures of inbreeding and productive and

reproductive traits were provided by literature (Miglior et al., 1995, McParland et al., 2007, Gonzales-Recio et al., 2007). However, the development of molecular measure of inbreeding is particular appealing due to the possibility to determine the age of inbreeding and to read the history or recent selection at molecular level**.** In fact, the extent and frequency of ROHs may be seen as a measure of the age of inbreeding: the longer the homozygous segments the more recent is the inbreeding. In other words, because recombination events broke down long chromosome segments, it is expected that long autozygous segments in an individual derived from a common recent ancestor. In the opposite way, shorter autozygous segments trace back to remote common ancestor and could also include some IBS segment (Ferencakovic, et al., 2012). In cosmopolitan cattle breeds, the level of homozigosity traced by ROH can be related to the intensive artificial selection carried out in the recent times with intensive use of few sires in artificial insemination (IA). At the same time, this process led to a reduction in population size and genetic diversity. The signal of ancient and recent selection can be traced back analyzing ROH data, as shown by Purfield et al., (2012) using several cattles breed and HD SNP panel.

Different measure of genomic inbreeding have been proposed, such as proportion of homozygous (F_{PH}) and those one derived from genomic relationship matrix (F_{GRM}). The use of F calculated from ROH data (F_{ROH}) seem to present some advantage over the F calculated from pedigree data (F_{PED}) or F_{GRM}. One of this advantage is to take into account of the autozigosity in the founder, that in general even for large and depth pedigrees are considered unrelated. In fact, even very precise estimation of inbreeding from pedigree do not face the problem of ancient relatedness (Ferencakovic et. al., 2011), and estimates of inbreeding can deviate from the pedigree based expectation as shown by Carothers et al., (2006). As far F_{GRM} is concerned, for its calculation a assumption is made on the allele frequency in the base population: p=0.5 was used for the computation of F_{GRM} as described by VanRaden et al., (2011). Alternatively, the estimation of allele frequency in the base population is required, but this is computationally tricky and resulted in a worse correlation with F_{PED}.

OBJECTIVE OF THE WORK

The objective of this work is to describe the distribution of ROH of different length classes in five Italian cattle breed: 2 dairy breeds (Holstein and Brown), 2 beef breeds

(Piedmontese and Marchigiana) and 1 double purpose breed (Simmental) for a total of 4,095 animals. Furthermore the association between number of ROH and length of ROH have been studied. Finally molecular inbreeding coefficients were derived from ROH data and compared to the pedigree based coefficients.

MATERIALS AND METHODS

Data

A total of 4,095 bulls were assayed using Illumina BovineSNP50 v.1 beadchip (www.illumina.com). Five Italian breeds were used for the ROH analysis: Italian Holstein (n=2,093), Italian Brown (n=749), Italian Simmental (n=479), Piedmontese (n=364) and Marchigiana breed (n=410). All of them were genotyped in the framework of two Italian National projects (SELMOL, PROZOO).

Data editing were performed both on animals and SNP. All the animal with mendelian inconsistency or with more than 1000 missing genotypes were removed from the dataset. Data quality control on SNP were performed for all breeds together. SNP that did not map to any chromosome or that were in the X chromosome were excluded from the analysis. In addition, SNP with more than 2.5% of missing data, with MAF < 1% were removed. After data editing, 44,325 SNP were retained in the final dataset. No pruning based on LD were performed in these dataset (Ferencakovic et al., 2013) and the whole set of selected marker were used to calculated ROH.

Criteria used for ROH detection

ROHs were detected using a SAS script (SAS Institute, Inc; Cary NC). This script was designed to find stretches of DNA with a certain number of homozygous SNP in a row. Some constrains were put to limit the number of spurious ROH detected (Howrigan et al., 2011): *i*) the minimum number of SNP included in the ROH was fixed at 15 SNP; *ii*) ROH were called only when stretches of DNA longer than 1 Mb were detected; *iii*) the maximum distance between adjacent SNP to be considered in the same row was equal to 1 Mb; *iv*) neither heterozygous nor missing genotypes were allowed in a ROH; *v*) no sliding windows were used to assess the presence of a ROH. We have chosen to call ROH longer than 1 Mb given that the distribution of DNA IBD fragments is function of the number of generation since common ancestor. Thus, smaller homozygous fragment are more likely to

be not IBD but generated from LD. The expected length of DNA segments that is IBD was derived by Fisher (1954) and follow an exponential distribution with mean equal to ½ *g Morgan* where *g* is the number of generation since the common ancestor (Bjelland et al., 2012). Given that, there is a rough correspondence between ROH length and the number of generations that separates the individual analyzed from the common ancestor that hold the original IBD fragment. Since the recombination broke down the original segment over time it is possible to date the IBD fragment using the length of the ROH. Fragments of 1 Mb and 16 Mb long date around ~50 and ~6 generation since common ancestor, respectively. To find more details on this calculation see Howrigan et al., (2011). According to the criterion abovementioned, five different classes of ROH length were defined following Kirin et al., (2010) and Ferencakovic et al., (2012): with $ROH_{1\text{-}2Mb}$ is defined the class of length that include ROH which length is < 2 Mb; $ROH_{2\text{-}4Mb}$, $ROH_{4\text{-}8Mb}$, $ROH_{8\text{-}16Mb}$, ROH_{16Mb} are the classes of length whose boundaries are indicates in the lower scripts.

Basic statistics and genomic inbreeding calculation

To characterize the autozygosity within and across breeds a descriptive statistical analysis was carried out. In particular, the average number of SNP within each ROH length category (N_{ROH}) and for each breed was calculated. Moreover, the average length of ROH (L_{ROH}, in Mb) was computed for each animals, as well as the summation across the whole genome of the length of all segment in ROH for each animals (S_{ROH}, in Mb). These coefficients (N_{ROH}, L_{ROH} and S_{ROH}) were calculated separately for each breed in the whole dataset, or considering the division in classes of different ROH length. Box plots for different class length and breeds were provided to assess the distribution of the ROH coefficients.

The Sum of all ROH that each animal carried was identified as the best metrics to characterize the different breed (McQuillan et al., 2008; Ferencakovic et al., 2012). S_{ROH} was the used in two different way: for each animals S_{ROH} was first plotted against the total numbers of ROH detected in each breed, to analyze simultaneously the differences in length and number of ROH among breeds; S_{ROH} was then used to computes genomic based inbreeding coefficients (F_{ROH}).

Genomic measure of autozygosity was calculated following McQuillan et al., (2008), who defined for the *i*-th individual F_{ROH} as the proportion of the autosomal genome in ROH:

$$F_{ROH\,i} = \frac{S_{ROH\,i}}{L}$$

where S_{ROH} is the sum of all the ROH (Mb) for an individual (here, centromeric region were included) and L is the genome length covered by SNP, in our case L corresponded to ~2,556.437 Mb. Furthermore, the F_{ROH} were computed for each class of ROH length (<2, 2-4, 4-8, 8-16 and >16 Mb) and they were indicated as $F_{ROHCLASS}$. Pedigree based inbreeding coefficient F_{PED} were available for three out of 5 breed analyzed provided by national breeding associations: Holstein, Brown and Simmental. For these three breed basic statistics were computed for all these coefficients (genomically determined and pedigree based) and correlation analysis was carried out. Finally the regression of F_{ROH} on F_{PED} was carried out.

RESULTS AND DISCUSSION

Number of ROH

The total number of ROH (N_{ROH}) detected by our algorithm sum up 335,985 across 4,095 individuals of 5 cattle breeds, ranging from 19,657 ROH identified in Piedmontese to 171,047 in Holstein (Table 1). The N_{ROH} identified within 364 animals of Piedmontese is similar to what obtained by Ferencakovic et al., (2013) (19,392) in a sample of similar size.

To get rid of sample size effect N_{ROH} were averaged by animal (Table 1). Brown and Simmental showed the higher N_{ROH} per animals (average of 94.6±11.6 and 94.3±12.2 respectively) whereas the lower N_{ROH} per animals were found in Piedmontese breed (54.0±7.2). These Figures are in good agreement with the finding of Ferencakovic et al., (2012) who observed an average N_{ROH} of 98.9±10.2 and 94.5±13.2 analyzing ~300 Brown Swiss and ~500 double purpose Austrian Simmental bulls, respectively. The N_{ROH} per animals for each class of ROH length and their relative frequency by breed were reported in Table 1 and Figure 1. The N_{ROH} per animal varied both among breeds and across classes of ROH length. For example, the first two classes of ROH length (ROH<4 Mb in total) represented from 74% (in Brown) to 94% (in Piedmontese) of the total number of ROH detected. Looking at Figure 1 a general pattern can be observed: animals from Brown and Holstein exhibit lower number of ROH <2Mb in comparison to Piedmontese, Simmental

and Marchigiana, whereas the opposite happened for classes of length >4 Mb. Furthermore, animals with at least one ROH>16 Mb are 4% (Brown) or less (other breed) of the total.

The average number of SNP per ROH for each classes of length were provided in Table 2. The average number of SNP slightly changes among different breeds, especially for the three classes of ROH <8 Mb ($F_{ROH1\text{-}2Mb}$; $F_{ROH2\text{-}4Mb}$; $F_{ROH4\text{-}8Mb}$) whereas, the two class ROH >8 Mb seem to be more variable and to present more outliers. For instance, in the class $ROH_{1\text{-}2Mb}$ the average number of SNP ranged from 21.6±5.6 (Piedmontese) to 23.5±6.9 (Brown). In the class $ROH_{>16Mb}$ the average number of SNP per ROH varied from 405.7±150.5 (Brown) to 476.3±241.3 (Marchigiana). In particular, the upper limit for Marchigiana was 2,231 SNP, it means that nearly an entire chromosome (BTA12) is homozygous for an animal. The Figures reported in Table 2 are quite in agreement with the average number of SNP obtained by Ferencakovic et al., (2013) using 54K on three different cattle breed ($ROH_{1\text{-}2Mb}$ = 21.69±5.68, $ROH_{>16Mb}$ = 399.39±156.13).

The number of SNP that go into a ROH also depend on the software setting used to ROH discovery. In particular, the number of ROH detected relies on the allowance of some heterozygous SNP inside a ROH. For instance, Howrigan et al., (2011) tested three popular ROH detection softwares on simulated data, reporting error rates for incorrectly identification of autozygous segments. He did not advise to allow some SNP to be heterozygous in a ROH as we did in this work. Despite of that, Ferencakovic et al., (2013) demonstrated that some heterozygous need to be allowed, especially when dealing with HD SNP chip panels, to take into account of genotyping error.

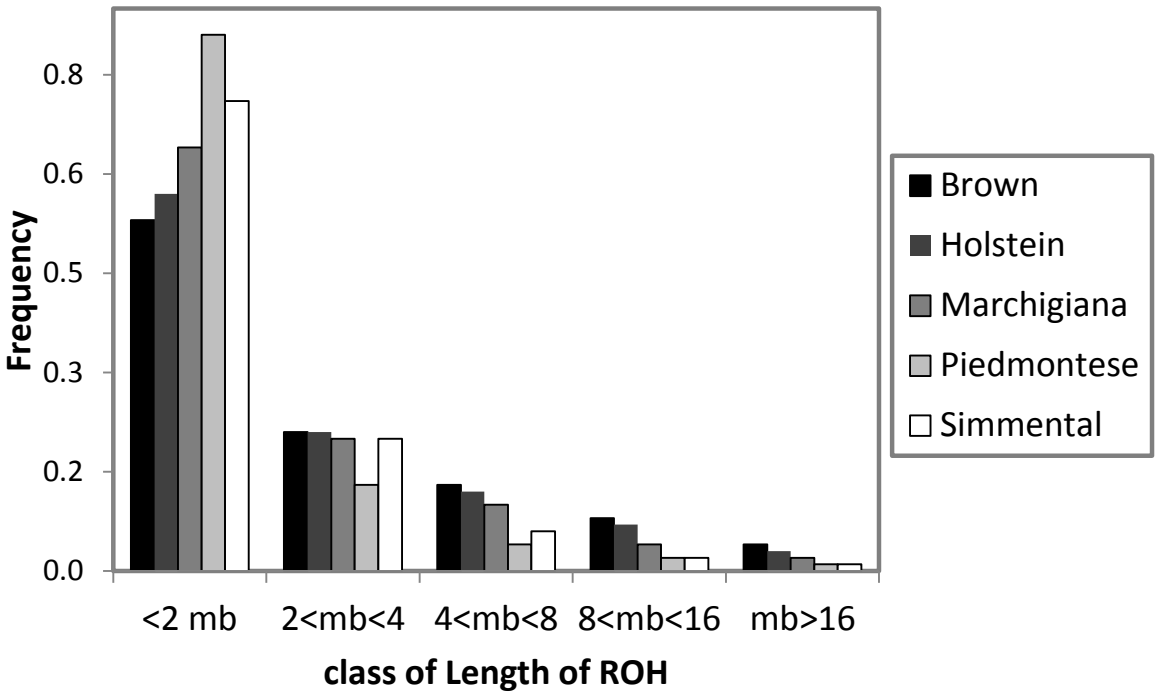


Figure 1. Distribution of the number of ROH of different class of length for each breed analyzed.

Table 1. Total Number of Runs of Homozigosity (N_{ROH}) and average number of ROH per animals (± sd) detected for each class of length in five Italian cattle breed.

	Breed[1]									
	Brown		Holstein		Marchigiana		Piemontese		Simmental	
ROH class	N_{ROH}	N_{ROH} animal	N_{ROH}	N_{ROH} Animal	N_{ROH}	N_{ROH} animal	N_{ROH}	N_{ROH} animal	N_{ROH}	N_{ROH} animal
$ROH_{1\text{-}2Mb}$	37,465	50.0 (*7.2*)	97,379	46.5 (*6.8*)	18,641	45.5 (*7.1*)	15,836	43.5 (*6.2*)	31,919	66.6 (*9.2*)
$ROH_{2\text{-}4Mb}$	15,191	20.3 (*5.1*)	35,586	17.0 (*4.2*)	5,803	14.2 (*4.6*)	2,646	7.3 (*2.8*)	9,174	19.2 (*5.2*)
$ROH_{4\text{-}8Mb}$	9,544	12.7 (*4.5*)	20,315	9.7 (*3.5*)	2,932	7.2 (*3.2*)	742	2.3 (*1.6*)	2,840	5.9 (*2.8*)
$ROH_{8\text{-}16Mb}$	5,972	8.1 (*3.3*)	12,258	5.9 (*2.7*)	1,248	3.0 (*2.2*)	302	1.9 (*1.5*)	882	2.3 (*1.5*)
ROH_{16Mb}	2,676	3.8 (*2.2*)	5,509	3.0 (*1.8*)	647	1.6 (*3.4*)	131	1.4 (*0.9*)	347	2.0 (*1.5*)
Tot ROH	70,847	94.6(*11.6*)	171,047	81.7 (*9.7*)	29,271	71.4 (*11.1*)	19,657	54.0 (*7.21*)	45,162	94.3 (*12.2*)

1) Number of effective observations used for N_{ROH} computation varied from class of length ranging from 706-749 (mean=738.8) in Brown, 1839-2093 (mean=2039) in Holstein, 237-401 (mean=366) din Marchigiana, 88-364 (mean=257.8) in Piedmontese and from 176-479 (mean=398.2) in Simmental.

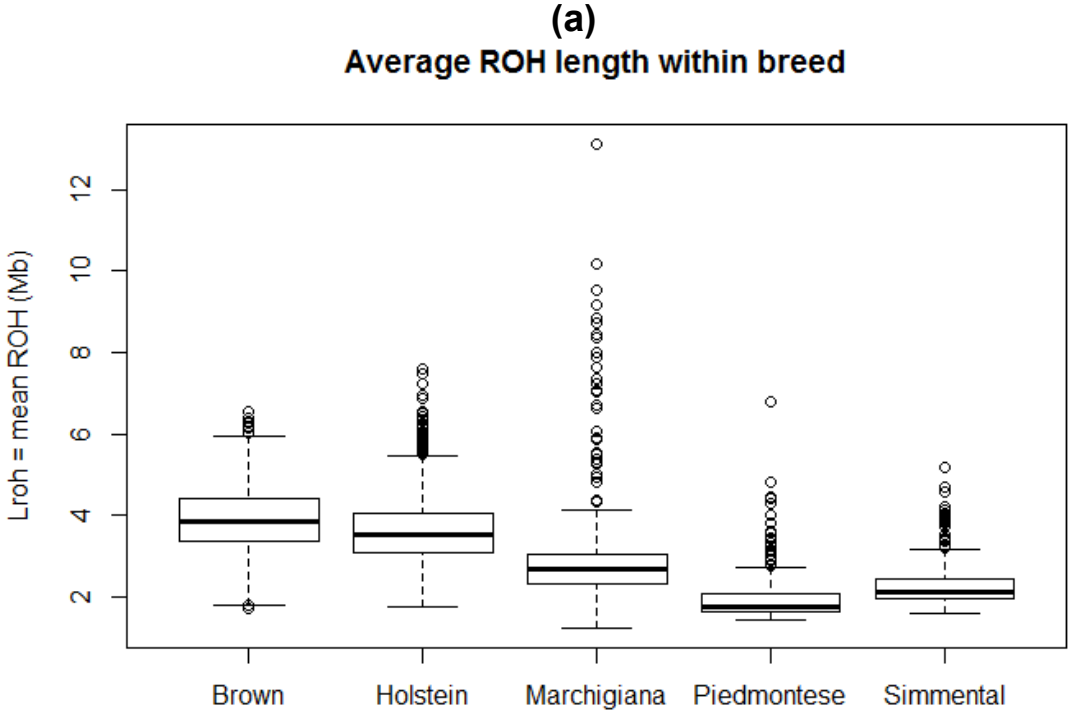
Table 2. Summary statistics about the average number of SNP in each ROH of different length category for 5 Italian Cattle breed.

Breed	item	$ROH_{1\text{-}2}$	$ROH_{2\text{-}4}$	$ROH_{4\text{-}8}$	$ROH_{8\text{-}16}$	ROH_{16}
Brown	*mean*	23.5	48.2	99.6	191.9	405.7
	st dev	6.9	15.7	26.4	45.6	150.5
	range	*(15-58)*	*(15-112)*	*(32-221)*	*(76-158)*	*(158-1,854)*
Holstein	*mean*	23.4	47.6	99.8	192.6	431.3
	st dev	6.7	15.8	27.1	47.1	170.3
	range	*(15-58)*	*(15-113)*	*(32-214)*	*(68-376)*	*(169-1,745)*
Marchigiana	*mean*	22.6	46.1	94.0	184.6	476.3
	st dev	6.4	15.4	26.2	45.6	241.3
	range	*(15-57)*	*(15-105)*	*(31-213)*	*(72-326)*	*(176-2,231)*
Piedmontese	*mean*	21.6	40.4	92.1	189.0	444.5
	st dev	5.6	15.6	26.5	51.3	185.0
	range	*(15-57)*	*(15-97)*	*(36-178)*	*(72-341)*	*(225-1,225)*
Simmental	*mean*	23.5	44.7	90.6	182.4	443.1
	st dev	6.7	14.9	25.8	47.5	173.3
	range	*(15-59)*	*(15-108)*	*(33-188)*	*(80-336)*	*(165-1,145)*

Average Length of ROH and Sum of ROH

Figure 2 reports the average ROH length (L_{ROH}) and the average sum of ROH (S_{ROH}) across all the animals and for each breeds. The distribution of the L_{ROH} for each breed is shown in Figure 2a. On average, Brown and Holstein showed longer L_{ROH} (about 3.9 and 3.6 Mb), conversely, Piedmontese and Simmental presented shorter L_{ROH} (1.9 and 2.2 Mb respectively) with the lowest variability. Marchigiana placed in the middle and presents several outliers. This results was corroborate by the work of Ferencakovic et al., (2012) who found and average L_{ROH} of 4.01 and 2.26 Mb for Brown Swiss and Austrian Simmental respectively.

In general, the average L_{ROH} showed much less variation compared to the sum of ROH by animals (Figure 2b). L_{ROH} does not seem to have great utility to analyze ROH data because of excessive data compression. In many papers S_{ROH} was used in its place (Kirin et al., 2011; Purfield et al., 2012; Siliò et al., 2013). In the current paper, Brown showed the highest average S_{ROH} (371.8±96.2 Mb), followed by Holstein (296.3±76.5), Simmental (212.8±56.7), Marchigiana (216.9±120) and Piedmontese (105.86±40.72). Average S_{ROH} present similar pattern of L_{ROH} with less overlaps, offering a much more clear picture. Moreover, S_{ROH} have a genetic interpretation as measure of the length of autozigous segments (McQuillan et al., 2008).



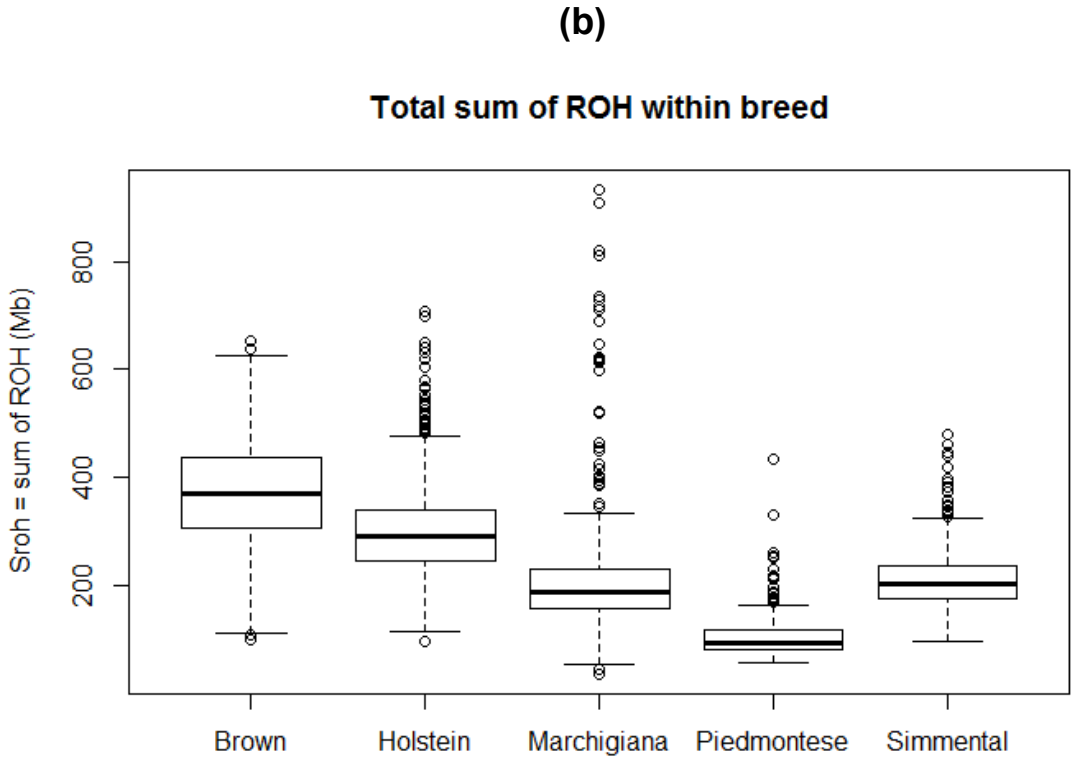
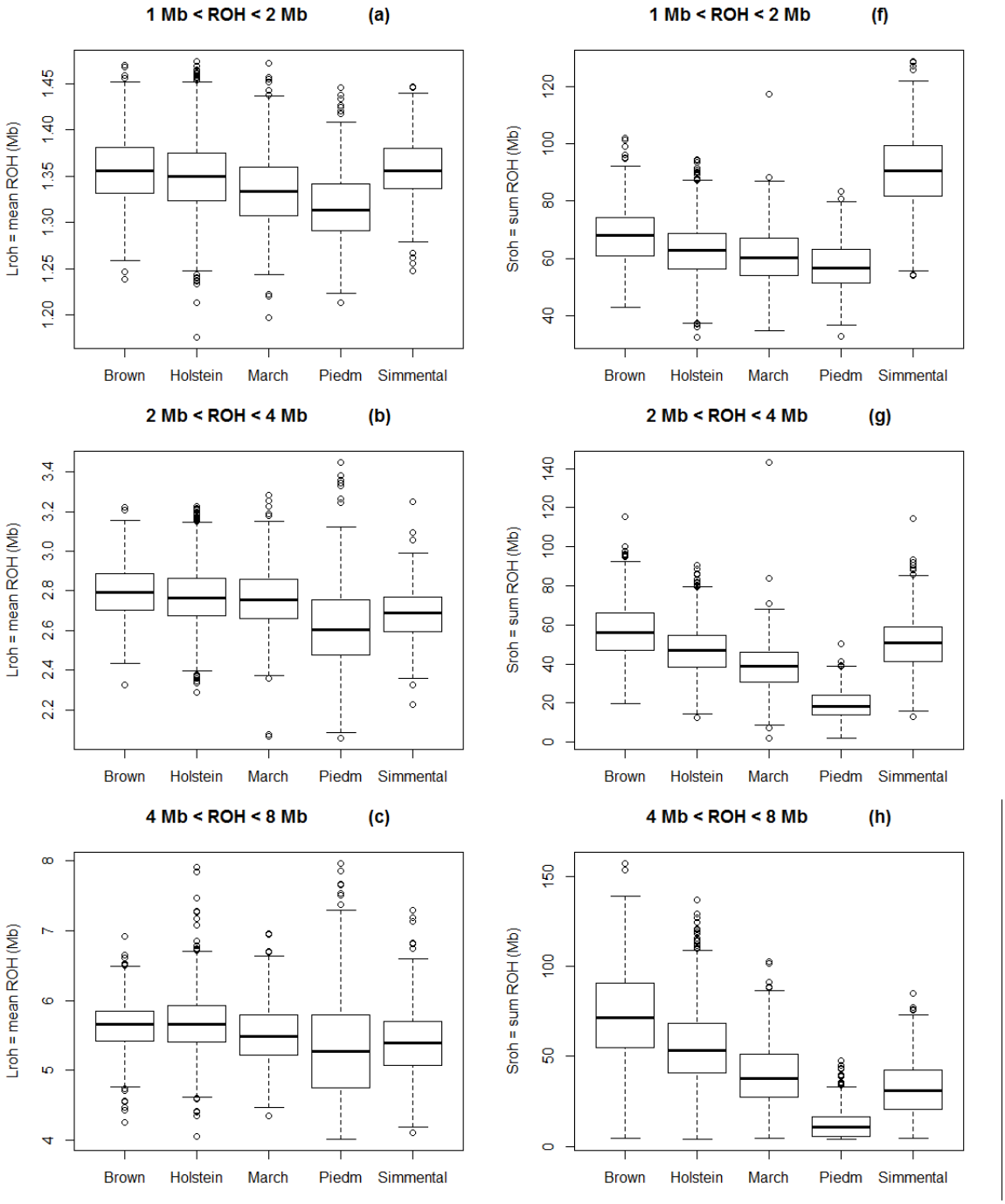


Figure 2. Box Plots of within Breed (a) average Length of ROH (LROH, in Mb) and (b) Sum of Length of ROH (SROH, in Mb) across all the animals.

The pattern observed for S_{ROH} closely confirm what found in a work of Ferencakovic et al., (2012) that reported an average S_{ROH} of 396.8±82.9 and 223.1±52.5 Mb in Brown Swiss and Austrian Simmental respectively. For the other three breeds the average S_{ROH} was close to the finding of Purfield et al., (2012) in Holstein and Piedmontese with very high and very low average sum of ROH respectively. As far Marchigiana concerned, we have noticed that the average S_{ROH} for the individuals of this breed were similar of those ones obtained in Romagnola breed (Purfield et al., 2012) and this results might be due to the reduced genetic distance from Marchigiana and Romagnola deriving from their common recent selection (Ciampolini et al., 1995; Blott et al., 1998).

The repartition of L_{ROH} and S_{ROH} based on the their classification in different length category is showed in Figure 3. As for the case of pooled data, the average L_{ROH} were not able to distinguish the breeds according to their ROH composition. In particular the average L_{ROH} values did not vary very much within each class of ROH length. For instance, average L_{ROH} ranged from 1.32 Mb (Piedmontese) to 1.35 Mb in $ROH_{1\text{-}2Mb}$ (Brown) and from 24.09 Mb (Brown) to 28.13 Mb (Marchigiana) in $ROH_{>16Mb}$, (see Figure

3a to 3e). Despite that, an increases in the dispersion of S_{ROH} can be observed, passing from shorter ROH to longer ones. This fact is probably due to the different sample sizes, being the shorter ROH (<4 Mb) at least 80% of the total number of ROH detected.



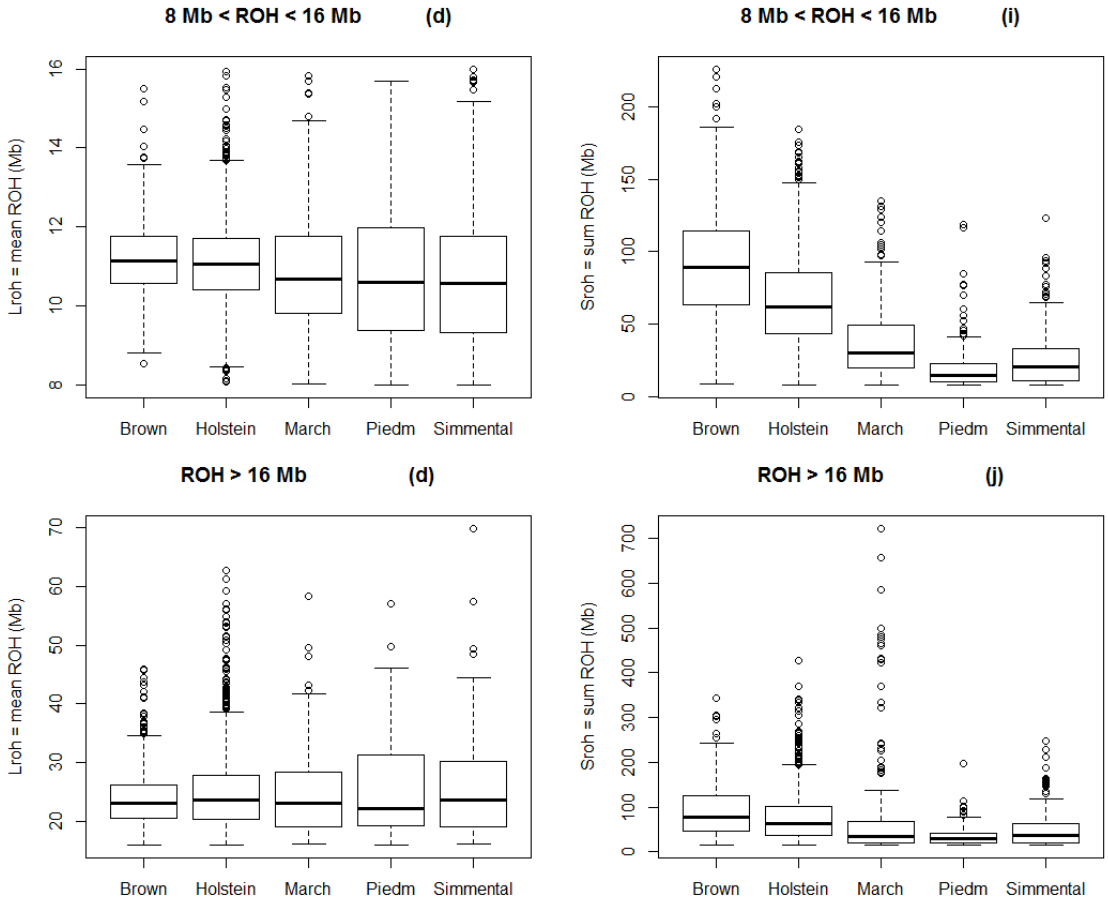


Figure 3. Box Plots of within Breed (on *x* axis) Length of ROH (L_{ROH}, in Mb) for each class of ROH length in Mb on the left side (a) [<2], (b) [2-4] , (c) [4-8], (d)[8-16] and (e) [>16]. On the right side box plots of sum of length of ROH (S_{ROH}, in Mb) within each class of length in Mb (f) [<2], (g) [2-4], (h) [4-8], (i)[8-16] and (j) [>16]

Figure 3 (from f to j) show the S_{ROH} for different ROH classes. In these box whisker plots it is possible to notice some differences among breeds within the same ROH length class. For instance, in the class of shorter segment ($ROH_{1\text{-}2Mb}$), Simmental presented a values of S_{ROH} of 90.4 Mb, whereas the other breeds approached 60 Mb of genome covered by ROH<2 Mb. The shorter average S_{ROH} in Simmental (and the higher N_{ROH} <2Mb) compared to the other breeds, can be due to ancestral relatedness which characterize this breed. Conversely, an opposite trend was observed with longer ROH length class (>8 Mb), average S_{ROH} of Simmental were up to 3.6 and 2.6 fold shorter than average S_{ROH} of Brown and Holstein, respectively. Very similar results was also found by Ferencakovic et al., (2012) for ROH detected in Austrian Simmental compared to Brown Swiss.

For longer ROH length classes (>4 Mb) Brown and Holstein presented the higher average S_{ROH} (Figure 3). In Brown, on average 71.9, 90.1 and 91.3 Mb of genome was covered by $ROH_{4\text{-}8Mb}$, $ROH_{8\text{-}16Mb}$ and $ROH_{>16Mb}$ respectively. Similarly, in Holstein the average S_{ROH} lengths resulted equal to 55.1, 65.5 and 75.4 Mb for the class of length $ROH_{4\text{-}8Mb}$, $ROH_{8\text{-}16Mb}$ and $ROH_{>16Mb}$ respectively. These findings are comparable to the results of Ferencacovic et al. (2012) and Purfiled et al. (2012) in Brown and Holstein cattle, where the higher prevalence of longer S_{ROH} were related to recent consanguineous mating due to the intensive selection and the reduced number of IA sire massively used in these two breed.

Differently from the other breeds, Piedmontese showed the lowest average S_{ROH} for all the class of length with exception to the shorter $ROH_{1\text{-}2Mb}$, denoting trace of ancient relatedness (see also Purfield et al., 2012).

Length of ROH vs Number of ROH

The total length of ROH was plotted against the number of ROH for all breeds analyzed (Figure 4). The Scatter Plot in Figure 4 demonstrate that there is a strong relationship between the number of ROH and the total length of genome of an individual covered by ROH. However, the entity of this relationship change from breed to breed. In particular, the number of ROH with the shortest length is recorded for Piedmontese followed by Marchigiana and Simmental. Probably, the difference observed within beef cattle can be related to the genetic distance between Piedmontese and Marchigiana (Blott et al.,1998). The separation of the Piemontese from the other Italian beef cattle has been previously observed by Astolfi et al. (1983) and Ciampolini et al. (1995).

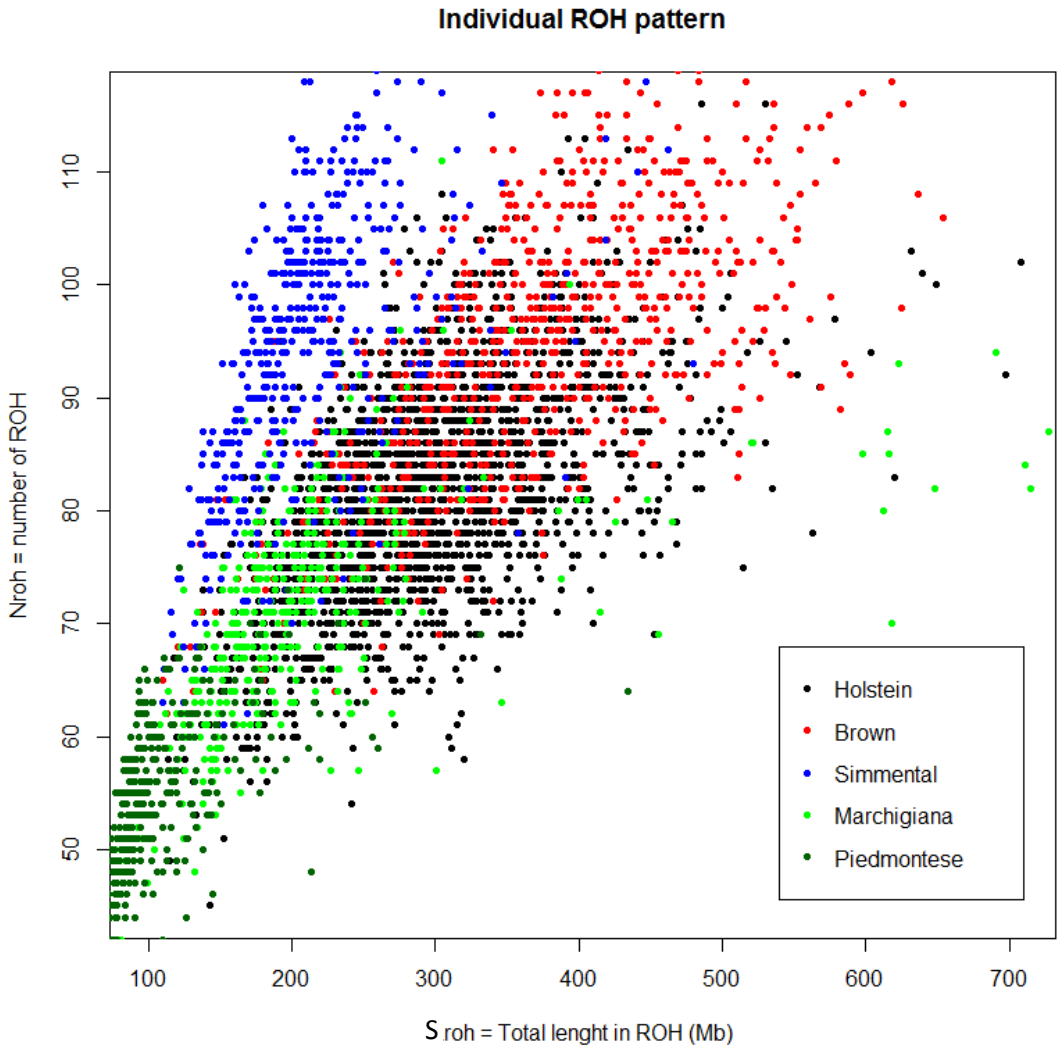


Figure 4 Individual Pattern of run of Homozigosity (ROH). Scatter plot of Number of the ROH (N_{ROH}) *vs* the total length of ROH (S_{ROH}): dots of different colors are animals of different breeds.

The Brown and Holstein breeds generally presented an higher number segments with average and total length much higher than the other 3 breeds. From the comparison of 2 dairy breed (Holstein and Brown) with Simmental (Figure 4), can be highlighted that the total length of ROH for Simmental is mainly composed of Shorter ROH segment, while especially for the brown, and secondly for Holstein, the total number of ROH is represented from a higher number of larger segments.

Inbreeding measure based on ROH

The pedigree inbreeding coefficients were available for Brown, Holstein and Simmental bulls. They were calculated using all the pedigree information available (the minimum number of generation was 4). The highest inbreeding level was observed for Holstein (0.044) followed by brown (0.026) and Simmental (0.008) bulls (Table 3). Looking at ROH based inbreeding coefficient they ranged from 0.034 ($F_{ROH>16Mb}$) to 0.145 ($F_{ROH>1Mb}$) in Brown, from 0.026 to 0.116 in Holstein and from 0.007 to 0.083 in Simmental. In general, F_{ROH} for were higher than F_{PED} coefficients, and in particular considering the F_{ROH} calculated using the whole ROH information. $F_{ROH>16Mb}$ approached the values of F_{PED} with very small difference (<0.008).

Table 3. Summary statistics (mean, standard deviation and range) of pedigree based inbreeding coefficients (F_{PED}) and ROH based inbreeding coefficients (F_{ROH}) for three Italian Cattle breed which pedigree data were available. Both F_{ROH} (ROH >1Mb) and F_{ROH} greater than a specific class length $F_{ROH->class}$ (>2, >4, >8, and >16 Mb) were reported.

Breed	F_{PED} ± *sd*	F_{ROH} ± *sd*	$F_{ROH>2}$ ± *sd*	$F_{ROH>4}$ ± *sd*	$F_{ROH>8}$ ± *sd*	$F_{ROH>16}$ ± *sd*
Brown	0.026±0.021	0.145±0.038	0.119±0.038	0.097±0.036	0.068±0.032	0.034±0.023
	(0 –.141)	*(.039 –.256)*	*(.016 –.227)*	*(.002–.203)*	*(0 –.173)*	*(0 –.134)*
Holstein	0.044±0.022	0.116±0.030	0.092±0.030	0.073±0.029	0.051±0.026	0.026±0.021
	(0 –.179)	*(.038 –.277)*	*(.015 –.254)*	*(.006–.233)*	*(0 –.197)*	*(0 –.167)*
Simmental	0.008±0.013	0.083±0.022	0.048±0.022	0.028±0.021	0.015±0.018	0.007±0.014
	(0 –.078)	*(.038 –.188)*	*(.010 –.161)*	*(0–.142)*	*(0 –.118)*	*(0 –.967)*

F_{ROH} calculated considering both shorter and longer ROH together ($F_{ROH>1Mb}$, $F_{ROH>2Mb}$) was higher than F_{PED}. This was probably due to the fact that ROH are able to capture both ancient and recent relatedness explained by shorted and longer IBD fragment respectively. Conversely, pedigree is only able to track the recent IBD segments. In fact, the coefficients derived for longer ROH ($F_{ROH>16Mb}$), that track recent IBD fragment, are very similar in magnitude to the F_{PED}. Similar upward shift between genomic based coefficients and classical pedigree based coefficients were also highlighted in different works, which dealt with genomic inbreeding estimation both in Human and in Cattle genetics (McQuillan et al., 2008; Van Raden et al., 2011; Ferencakovic et al., 2012; 2013). In particular, our results confirm the results of Ferencakovic et al. (2012) who found very similar F_{PED} and F_{ROH} in a different Brown Swiss population (F_{PED}=0.024 $F_{ROH>1Mb}$=0.156 $F_{ROH>16Mb}$ = 0.037) and Austrian Simmental (F_{PED}=0.009; $F_{ROH>1Mb}$= 0.088 $F_{ROH>16Mb}$=0.017).

Table 4 reports the F_{ROH} calculated using the sum of ROH of a particular category of ROH length. The values of F_{ROH} calculated in such way, presumably, take into account of the specific length of IBD fragments. The values of these coefficients were much lower than those showed in the Table 3, ranging from 0.027 ($F_{ROH1\text{-}2Mb}$) to 0.034 ($F_{ROH>16Mb}$) in Brown, from 0.027 to 0.034 in Holstein and from 0.027 to 0.034 in Simmental. The inbreeding derived from short IBD segments (deriving from recent common ancestor) were higher for Simmental, in comparison to Brown and Holstein, due to the high N_{ROH} in the class of short ROH. The opposite happened for the longest ROH.

Table 4. Summary statistics (mean, standard deviation and range) ROH based inbreeding coefficients of different ROH class length ($F_{ROH\text{-}class}$: 1-2, 2-4, 4-8, 8-16 and >16 Mb) for three Italian Cattle breed which pedigree data were available

Breed	$F_{ROH1\text{-}2} \pm sd$	$F_{ROH2\text{-}4} \pm sd$	$F_{ROH4\text{-}8} \pm sd$	$F_{ROH8\text{-}16} \pm sd$	$F_{ROH16} \pm sd$
Brown	0.027±0.004	0.022±0.006	0.028±0.010	0.035±0.015	0.034±0.023
	(.017–.039)	*(.008–.045)*	*(.002–.062)*	*(.–.088)*	*(.–.134)*
Holstein	0.025±0.004	0.018±0.005	0.022±0.008	0.025±0.012	0.026±0.021
	(.013 –.037)	*(.005–.036)*	*(.–.054)*	*(.–.072)*	*(.–.167)*
Simmental	0.035±0.005	0.020±0.006	0.013±0.006	0.008±0.007	0.007±0.014
	(.021–.051)	*(.005–.044)*	*(.–.033)*	*(.–.048)*	*(.–.967)*

The correlations among F_{ROH}, calculated according to their ROH length category, and $F_{ROH>1Mb}$ and F_{PED} were reported in Tables 5, 6 and 7 for Brown, Holstein and Simmental, respectively. These Tables highlight that F_{ROH} derived from very Short (1-2 Mb) and Short (2-4 Mb) ROH are negatively low correlated (or not correlated) with F_{PED} with a similar pattern, regardless of the breed considered. Conversely, the F_{PED} is moderately correlated with $F_{ROH>16Mb}$ (up to 0.712 in Simmental). This may be a further confirmation of the fact the longer ROH track the information that is already present in the pedigree, hence the recent inbreeding. The $F_{ROH>1Mb}$ are those one with the highest correlation with F_{PED} in all the breed with the exception of Simmental. Further confirmation of this pattern was given by the higher correlation between $F_{ROH>1Mb}$ and $F_{ROH>16Mb}$ (up to 0.828 with the lower values for Simmental = 0.781), that can possibly suggest the ROH are able to discriminate breeds from recent consanguinity history (Purfield et al., 2012). The absence of linear relation between F_{ROH} derived from very short (1-2 Mb) and short (2-4 Mb) ROH can be related to the origins of this ROH segments.

Table 5. Correlation matrix between pedigree based inbreeding coefficients (F_{PED}) and ROH based inbreeding coefficients ($F_{ROHclass}$) for different class of ROH length or for the raw $F_{ROH>1Mb}$ calculated using all the ROH >1 Mb for Brown.

	F_{PED}	F_{ROH}	F_{ROH1-2}	F_{ROH2-4}	F_{ROH4-8}	$F_{ROH8-16}$	F_{ROH16}
F_{PED}	*	0.662	-0.099	0.234	0.354	0.447	0.588
F_{ROH}		*	-0.125	0.423	0.625	0.700	0.811
F_{ROH1-2}			*	-0.023	-0.069	-0.140	-0.243
F_{ROH2-4}				*	0.294	0.220	0.172
F_{ROH4-8}					*	0.363	0.280
$F_{ROH8-16}$						*	0.295
F_{ROH16}							*

Table 6. Correlation matrix between pedigree based inbreeding coefficients (F_{PED}) and ROH based inbreeding coefficients ($F_{ROHclass}$) for different class of ROH length or for the raw $F_{ROH>1Mb}$ calculated using all the ROH information together for Holstein.

	F_{PED}	$F_{ROH>1Mb}$	F_{ROH1-2}	F_{ROH2-4}	F_{ROH4-8}	$F_{ROH8-16}$	F_{ROH16}
F_{PED}	*	0.700	-0.075	0.173	0.386	0.460	0.561
$F_{ROH>1Mb}$		*	-0.027	0.270	0.473	0.627	0.828
F_{ROH1-2}			*	0.022	-0.038	-0.064	-0.166
F_{ROH2-4}				*	0.150	0.073	0.061
F_{ROH4-8}					*	0.210	0.145
$F_{ROH8-16}$						*	0.238
F_{ROH16}							*

Table 7. Correlation matrix between pedigree based inbreeding coefficients (F_{PED}) and ROH based inbreeding coefficients ($F_{ROHclass}$) for different class of ROH length or for the raw $F_{ROH>1Mb}$ calculated using all the ROH information together for Simmental

	F_{PED}	$F_{ROH>1Mb}$	F_{ROH1-2}	F_{ROH2-4}	F_{ROH4-8}	$F_{ROH8-16}$	F_{ROH16}
F_{PED}	*	0.669	-0.108	-0.027	0.249	0.539	0.712
$F_{ROH>1Mb}$		*	0.141	0.375	0.566	0.670	0.781
F_{ROH1-2}			*	0.161	0.037	-0.053	-0.186
F_{ROH2-4}				*	0.245	0.123	-0.032
F_{ROH4-8}					*	0.251	0.224
$F_{ROH8-16}$						*	0.397
F_{ROH16}							*

In particular, it is possible that those ones are not truly IBD, and this clue seem to be confirmed by some works that compared the results of ROH detection using 54K and 800K SNPs chips. Purfield et al., (2012) analyzing ROH data from 9 cattle breeds, concluded that most of the ROH <5Mb found with a 54K panel, were not confirmed using the 800 K panels. Ferencakovic et al. (2013) concluded that when shorter ROH were included for the calculation of F_{ROH}, the inbreeding coefficients derived were systematically overestimated using 54 K panel.

To test the abovementioned hypothesis, correlation analysis and regression analysis were carried out between F_{PED} and F_{ROH} calculated including ROH with the minimum length here reported (>1, >2, >4, >8 and >16 Mb). Result of correlation analysis were reported in Table 8, whereas the results of the regressions of F_{ROH} on F_{PED} were showed in Table 9 and Figure 5, 6 and 7 for Brown, Holstein and Simmental respectively.

Table 8 Correlation between pedigree based inbreeding coefficients (F_{PED}) and ROH based inbreeding coefficients ($F_{ROH>1,2,4,8,>16}$) for different class of ROH length in three Italian cattle breed.

Inbreeding Coefficient	Brown	Holstein	Simmental
$F_{ROH>1Mb}$	0.662	0.700	0.669
$F_{ROH>2\ Mb}$	0.661	0.702	0.698
$F_{ROH>4\ Mb}$	0.661	0.696	0.747
$F_{ROH>8\ Mb}$	0.654	0.651	0.765
$F_{ROH>16\ Mb}$	0.588	0.561	0.712

The Pearson correlation between F_{PED} and F_{ROH} of different minimum ROH length follow a downward trend when increasing the minimum length, passing from 0.662 (> 1Mb) to 0.558 (>16 Mb) and from 0.7 (>1 Mb) to 0.561 (>16 Mb) in Brown and Holstein, respectively (Table 8). Conversely, for the Simmental the correlation between F_{PED} and F_{ROH} showed a different pattern, increasing up to $F_{ROH>8Mb}$. The difference in the observed pattern among breeds can be due to several causes. Firstly, analyzing the results from the dairy breed, it seems that the inclusion of shorter ROH in the calculation of genomic inbreeding coefficients does not improve that much the correlation between F_{PED} and F_{ROH} both in Brown and Holstein. Secondly, the effect of inclusion of ROH <4 Mb in the calculation of F_{ROH} seem to prove the different origin of this short homozygous segments (not truly IBD) provided that the hypothesis of overestimation of the correlation between

F_{ROH} calculated with ROH <4 Mb and F_{PED} is true. As far as Simmental concerned, the observed pattern is the opposite of what found in the two dairy breeds. This may be related to the particular genetic structure of this population. In particular, very high number of ROH shorter than <4 Mb (91%) or 8 Mb (97%) make possible that the F_{ROH} calculated with this minimum length are not reliable. The downward trend in the correlation observed for the other 2 dairy breeds was also observed in Simmental, but from ROH >8 Mb (0.765) and >16 Mb (0.712). Similar results were obtained by Ferencakovic et al., (2012) who found decreasing correlation from $F_{ROH>1Mb}$ (0.66) to $F_{ROH>16Mb}$ (0.60) in Brown and increasing correlation from $F_{ROH>1Mb}$ (0.66) to $F_{ROH>16Mb}$ (0.64) in Austrian Simmental. Purfield et al. (2012) found correlation between F_{PED} and $F_{ROH>1Mb}$ or $F_{ROH>10Mb}$ equal to 0.73 and 0.70 respectively.

Table 9. Intercept (b_0), Slope (b_1) ± relative standard error of the regression of ROH based inbreeding coefficients (F_{ROH}) on pedigree based inbreeding coefficient (F_{PED}) for the breed which pedigree data were available.

Breed	Coeff.	$ROH_{>1Mb}$	$ROH_{>4Mb}$	$ROH_{>8Mb}$	$ROH_{>16Mb}$
Brown	b_0	0.114 ± 0.002	0.066 ± 0.002	0.043 ± 0.002	0.016 ± 0.001
	b_1	1.202 ± 0.050	1.156 ± 0.048	0.983 ± 0.042	0.657 ± 0.033
Holstein	b_0	0.073 ± 0.001	0.030 ± 0.001	0.015 ± 0.001	0.002 ± 0.001
	b_1	0.972 ± 0.022	0.948 ± 0.021	0.804 ± 0.021	0.549 ± 0.017
Simmental	b_0	0.075 ± 0.001	0.019 ± 0.001	0.007 ± 0.001	0.002 ± 0.001
	b_1	1.146 ± 0.058	1.199 ± 0.049	1.084 ± 0.042	0.781 ± 0.035

Table 9 and Figures 5-7 showed the results of the regression analysis of F_{ROH} on F_{PED}. In particular, the intercept for of the regression class of inbreeding $F_{ROH>1Mb}$ was greater than zero for the all three breeds analyzed: 0.114 (se. 0.002) in Brown (Figure 5); 0.073 (se. 0.001) in Holstein (Figure 6); 0.075 (se. 0.001) in Simmental (Figure 7). Moreover, looking at the intercept of the same regression for $F_{ROH>8Mb}$ this was much close to zero: 0.043 (se. 0.002) in Brown; 0.015 (se. 0.001) in Holstein and 0.007 (se. 0.001) in Simmental. The same wqas pattern found for the slope that approach the unity in $F_{ROH>8Mb}$. The intercept greater than zero suggest that F_{PED} coefficient may underestimate the genomic ancestral relatedness on the base population which is captured by shorter ROH segments instead. When F_{ROH} was calculated with longer ROH the intercept of the regression approached closely to zero, and there was no tendency of F_{PED} to underestimate F_{ROH} because the recent inbreeding explained by longer ROH was also captured also by

pedigree. This concept was developed by Purfield et al. (2012) but fit in our case looking at the regression in Figures 5-7.

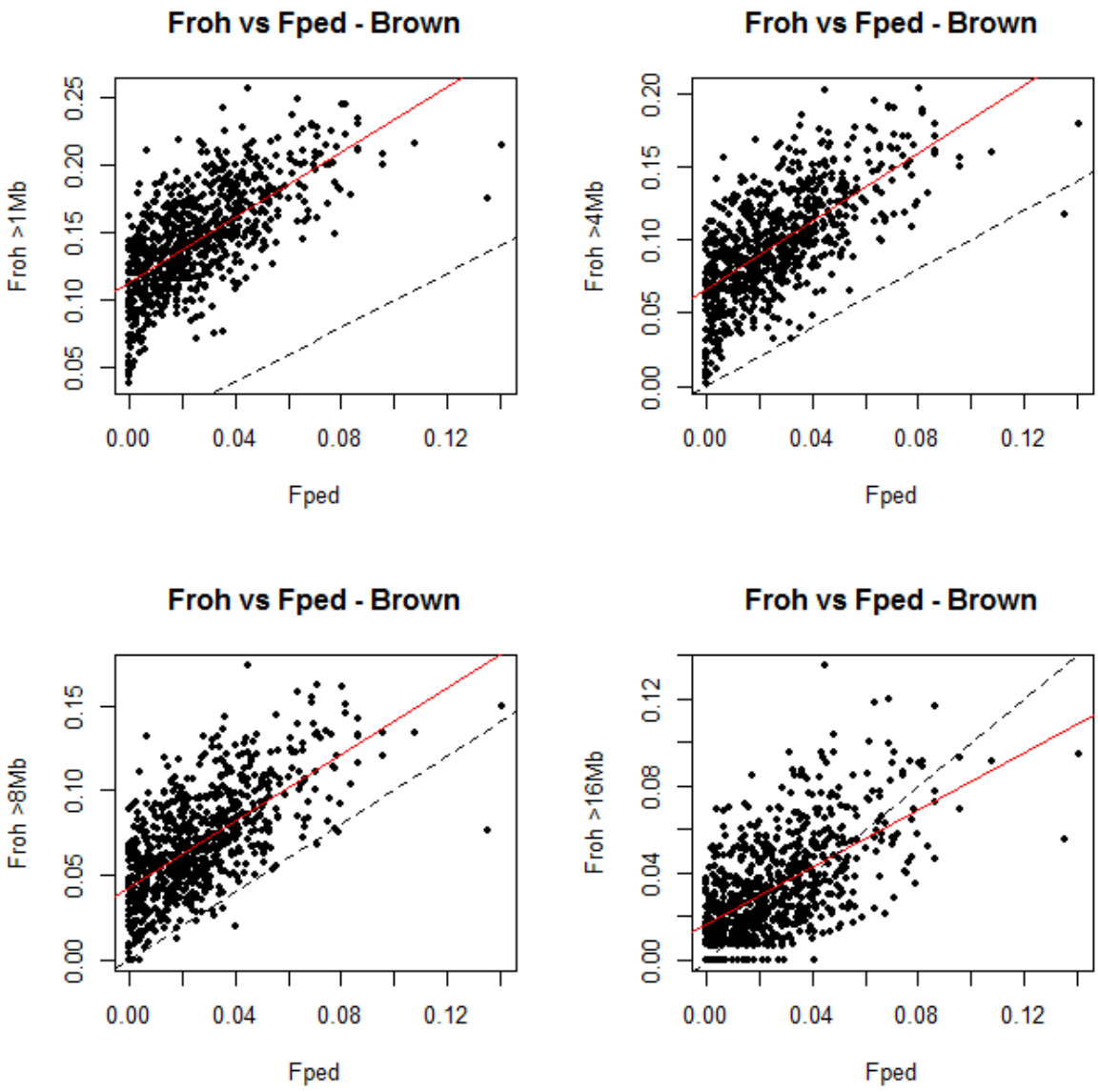


Figura 5. Regression between molecular inbreeding coefficient derived from ROH of difference length class ($F_{ROH>1Mb}$;$F_{ROH>4Mb}$;$F_{ROH>8\ Mb}$; $F_{ROH>16\ Mb}$) and inbreeding coefficient derived from pedigree (F_{PED}) for Brown. The black dashed lines are the lines of the equivalence ($y = x$). (R^2 =0.439 for $F_{ROH>1Mb}$; R^2 =0.436 for $F_{ROH>4Mb}$; R^2 =0.428 for $F_{ROH>8Mb}$; R^2 = 0.346 $F_{ROH>16Mb}$)

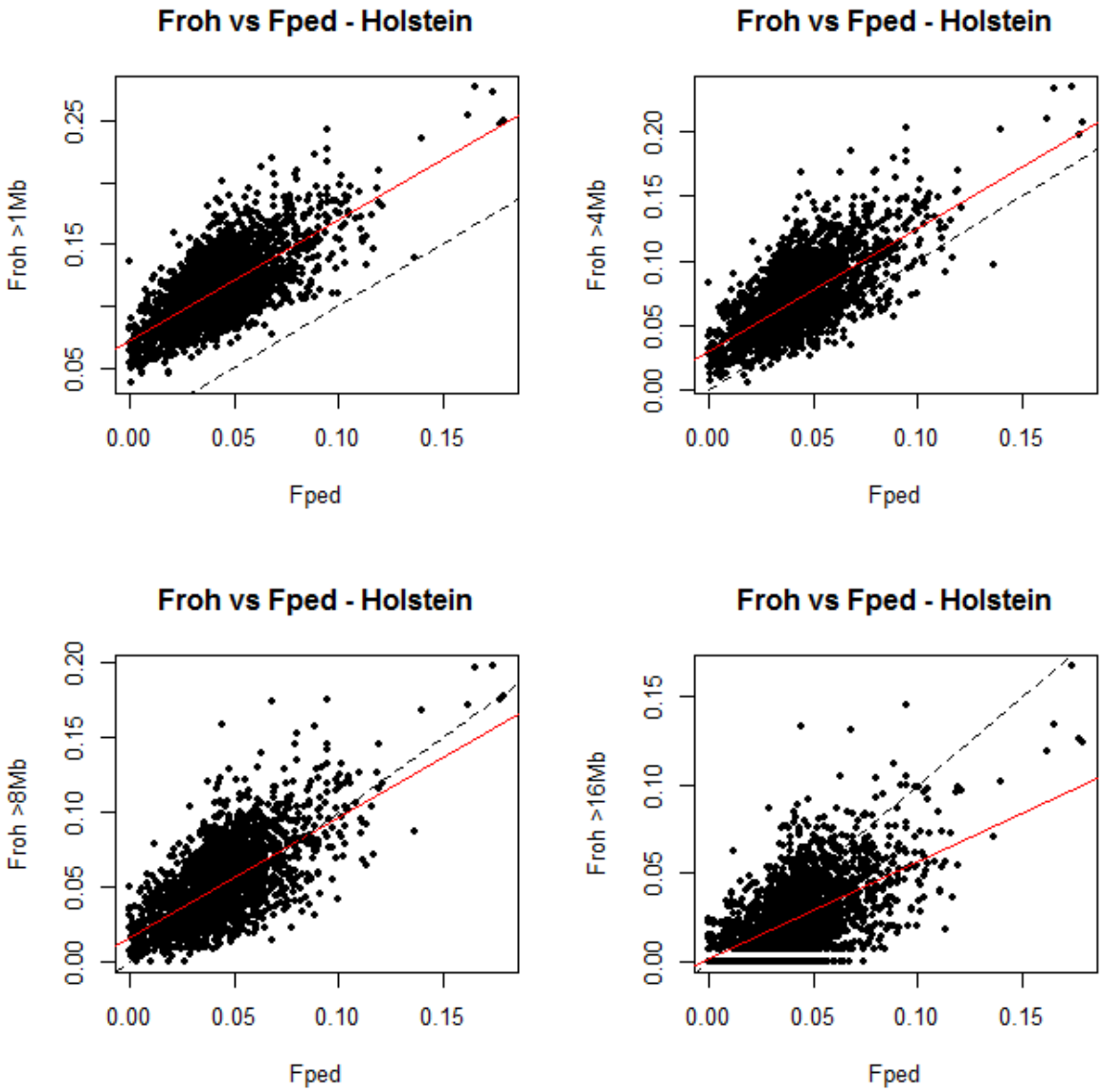


Figura 6. Regression between molecular inbreeding coefficient derived from ROH of difference length class ($F_{ROH>1Mb}$;$F_{ROH>4Mb}$; $F_{ROH>8\ Mb}$; $F_{ROH>16\ Mb}$) and inbreeding coefficient derived from pedigree (F_{PED}) for Holstein. The black dashed lines are the lines of the equivalence ($y = x$). (R^2 =0.491 for $F_{ROH>1Mb}$; R^2 =0.484 for $F_{ROH>4Mb}$; R^2 =0.424 for $F_{ROH>8Mb}$; R^2 = 0.314 $F_{ROH>16Mb}$)

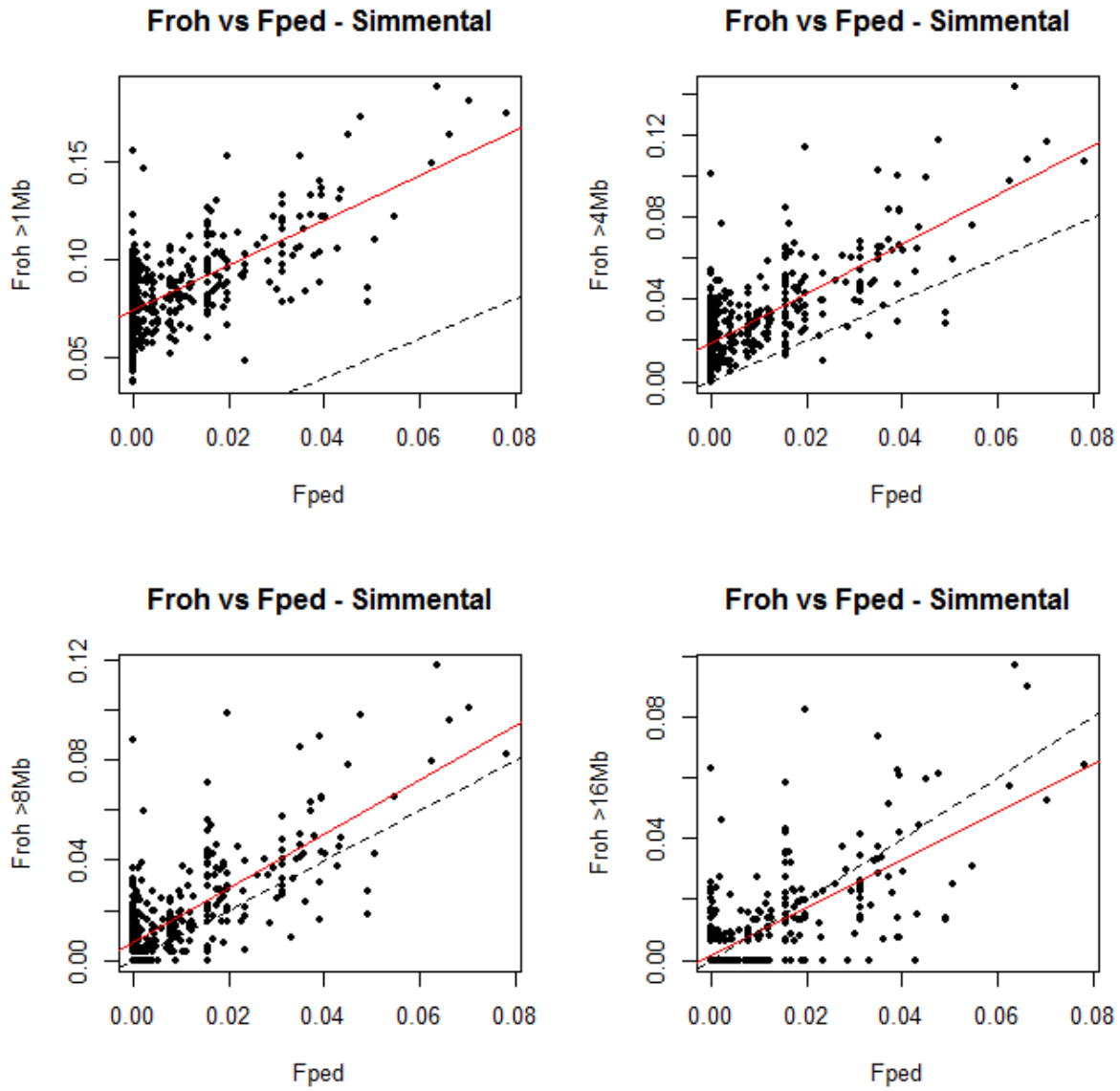


Figura 7. Regression between molecular inbreeding coefficient derived from ROH of difference length class ($F_{ROH>1Mb}$;$F_{ROH>4Mb}$;$F_{ROH>8\ Mb}$; $F_{ROH>16\ Mb}$) and inbreeding coefficient derived from pedigree (F_{PED}) for Simmental. The black dashed lines are the lines of the equivalence ($y = x$). (R^2 =0.448 for $F_{ROH>1Mb}$; R^2 =0.557 for $F_{ROH>1Mb}$; R^2 =0.585 for $F_{ROH>1Mb}$; R^2 =0.507 $F_{ROH>1Mb}$)

CONCLUSION

ROHs were calculated from medium density SNP panel using five Italian cattle breeds. The average sum of ROH in the genome was smaller for Piedmontese and Marchigiana that also showed the smaller variability, followed by Simmental, Holstein and Brown. On

average, Simmental had either the higher percentage or the higher sum of shorter ROH than other dairy breeds. Brown and Holstein hold a greater average number and average sum of longer ROH. A strong relationship between the number of ROH of definite size and the length of ROH was also found. Furthermore, an association between pedigree based inbreeding coefficient and ROH based inbreeding was highlighted remarking the recent selective history of these breeds. In particular, a positive relationship between inbreeding coefficients and length of ROH observed found for Simmental, Holstein and Brown. Lower inbreeding coefficients seem to be associated to the reduced number of longer ROH in Simmental. Higher inbreeding coefficient were related to longer ROH in Holstein and Brown. The regression between F_{ROH} of different minimum length (from >1 Mb to >16 Mb) on traditional inbreeding coefficient seem to confirm the hypothesis that including ROH < 4 Mb in the inbreeding calculation generate an overestimation of this coefficient using a 54 K chip.

REFERENCES

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. Journal of Dairy Science, 96, 4697-4706.

Blott, S. C., Williams, J. L. & Haley, C. S. 1998. Genetic relationships among European cattle breeds. Animal Genetics, 29, 273-282.

Broman, K. W. & Weber, J. L. 1999. Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain. American Journal of Human Genetics, 65, 1493-1500.

Carothers, A. D., Rudan, I., Kolcic, I., Polasek, O., Hayward, C., Wright, A. F., Campbell, H., Teague, P., Hastie, N. D. & Weber, J. L. 2006. Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. Annals of Human Genetics, 70, 666-676.

Ciampolini, R., Moazami-Goudarzi, K., Vaiman, D., Dillmann, C., Mazzanti, E., Foulley, J. L., Leveziel, H. & Cianci, D. 1995. Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. Journal of Animal Science, 73, 3259-3268.

Clark, A. G. 1999. The size distribution of homozygous segments in the human genome. American Journal of Human Genetics, 65, 1489-1492.

Clark, S. A., Kinghorn, B. P., Hickey, J. M. & van der Werf, J. H. J. 2013. The effect of genomic information on optimal contribution selection in livestock breeding programs. Genetics Selection Evolution, 44.

Ferencakovic, M., Hamzic, E., Gredler, B., Curik, I. & Sölkner, J. 2011. Runs of homozygosity reveal genomewide autozygosity in the Austrian Fleckvieh cattle. Agriculturae Conspectus Scientificus, 76, 325-328.

Ferencakovic, M., Hamzic, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I. & Soelkner, J. 2012. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. Journal of Animal Breeding and Genetics, 130, 286-293.

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

Gibson, J., Morton, N. E. & Collins, A. 2006. Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics, 15, 789-795.

González-Recio, O., López De Maturana, E. & Gutiérrez, J. P. 2007. Inbreeding depression on female fertility and calving ease in spanish dairy cattle. Journal of Dairy Science, 90, 5744-5752.

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.

Kearney, J. F., Wall, E., Villanueva, B. & Coffey, M. P. 2004. Inbreeding trends and application of optimized selection in the UK Holstein population. Journal of Dairy Science, 87, 3503-3509.

Keller, M. C., Visscher, P. M. & Goddard, M. E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics, 189, 237-249.

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M. & Wilson, J. F. 2010. Genomic Runs of Homozygosity Record Population History and Consanguinity. Plos One, 5.

Ku, C. S., Naidoo, N., Teo, S. M. & Pawitan, Y. 2011. Regions of homozygosity and their impact on complex diseases and traits. Human Genetics, 129, 1-15.

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., Kucherlapati, R. & Malhotra, A. K. 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proceedings of the National Academy of Sciences of the United States of America, 104, 19942-19947.

Mc Parland, S., Kearney, J. F., Rath, M. & Berry, D. P. 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. Journal of Dairy Science, 90, 4411-4419.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, 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., Campbell, H. & Wilson, J. F. 2008. Runs of homozygosity in European populations. American Journal of Human Genetics, 83, 359-372.

Meuwissen, T. H. E. 1997. Maximizing the Response of Selection with a Predefined Rate of Inbreeding. Journal of Animal Science, 75, 934-940.

Meuwissen, T. H. E. & Goddard, M. E. 2001. Prediction of identity by descent probabilities from marker-haplotypes. Genetics Selection Evolution, 33, 605-634.

Miglior, F., Burnside, E. B. & Kennedy, B. W. 1995. Production traits of Holstein cattle: estimation of nonadditive genetic variance components and inbreeding depression. Journal of Dairy Science, 78, 1174-1180.

Nielsen, H. M., Sonesson, A. K. & Meuwissen, T. H. E. 2011. Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. Journal of Animal Science, 89, 630-638.

Pryce, J. E., Hayes, B. J. & Goddard, M. E. 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. Journal of Dairy Science, 95, 377-388.

Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. 2012. Runs of homozygosity and population history in cattle. Bmc Genetics, 13.

SAS Institute. 2008. SAS/STAT Version 9.2. SAS Inst. Inc., Cary, NC.

Silió, L., Rodríguez, M. C., Fernández, A., Barragán, C., Benítez, R., Óvilo, C. & Fernández, A. I. 2013. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. Journal of Animal Breeding and Genetics, 130, 349-360.

Smith, L. A., Cassell, B. G. & Pearson, R. E. 1998. The Effects of Inbreeding on the Lifetime Performance of Dairy Cattle. Journal of Dairy Science, 81, 2729-2737.

Sonesson, A. K., Woolliams, J. A. & Meuwissen, T. H. E. 2012. Genomic selection requires genomic control of inbreeding. Genetics Selection Evolution, 44.

Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zoellner, S., Rosenberg, N. A. & Li, J. Z. 2013. Long Runs of Homozygosity Are Enriched for Deleterious Variation. American Journal of Human Genetics, 93, 90-102.

VanRaden, P. M., Olson, K. M., Wiggans, G. R., Cole, J. B. & Tooker, M. E. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. Journal of Dairy Science, 94, 5673-5682.

CHAPTER 5

General Conclusions

GENERAL CONCLUSIONS

The genome of living organisms is characterised by an high degree of complexity. The huge development of sequencing technologies allows for a continuous upgrading of the status of knowledge on DNA structure and organisation. Such a dynamic situation makes it difficult also to try to build a defined classification of genetic material structures, due to the continuous overlapping of functions, structures, regulation mechanisms. The evolution that has occurred in the last thirty-forty years in the field of markers used to study genetic variation between individuals represents a clear example of such a continuously evolving situation. Thus the genome should be then regarded as something very flexible, not easy classifiable with rigid criteria and whose elements could be analysed from different perspectives.

All these achievements are deeply changing the approaches used to study complex phenotypes as main physiological processes, diseases and, for livestock species, traits of economic interest. An important consequence is represented the new opportunities to study differences between animal populations under different points of view. In particular, genotypes at tens of thousands Single Nucleotide Polymorphisms that are currently generated by using high throughput platforms, can be used in different ways to investigate the components of genetic diversity. The analysis of the main cause of genetic variation, i.e. the difference between allelic frequencies at specific loci, can be now integrated by other structural aspects related to the duplication of DNA regions, as the copy number variations, or to uninterrupted successions of homozygous genotypes as the case of Runs of Homozygosity. All these features may allow to study some of the different elements that contribute to the genetic diversity between groups: phylogenesis, artificial selection, inbreeding.

Cattle represent an appealing opportunity to test different methods for disentangling specific features of genetic variation. Due to the effect of the intense artificial selection operated by man, especially in the last eighty years, various breeds can be found with markedly different phenotypes. Differences in the genetic structure could be detected also between breeds within the same production aptitude due to their history, specific breeding goals, environmental conditions where they evolved. SNP are of great help for the unravelling of differences in traits of economic importance because, due to their high

density along the genome, some of them are likely to be in linkage disequilibrium with causal mutations with effect on the phenotype of interest.

In the present thesis, three approaches have been presented for studying the genetic differences between five cattle breeds farmed in Italy. They were based on three metrics: the classical fixation index of Wright, a composite log likelihood measure of allelic differences, and the length and distribution of runs of homozygosity. The considered cattle breeds are the most important in Italy as far as the population size is concerned. The choice has been motivated by the need for comparing breeds with different selection goals (dairy, beef and dual purpose), but also for having populations with different genetic history within each production aptitude (i.e. Holstein vs Brown, or Piedmontese vs Marchigiana). Expected results were the detection of expected regions in which important genes that affect dairy and beef traits are located, but also new undetected regions responsible of genetic variations were not excluded, being some of the considered Italian populations not included in previous studies using SNP beadchips.

In the first experimental contribution, a classical comparison between a dairy and a beef breed was performed. Differences of allelic frequencies between the two populations were tested using the F statistics. This metrics is widely used in studies of genetic variation, even though there are some unresolved problems in its implementation, as the occurrence of random noise that makes patterns difficult to read, and the absence of a statistical test to identify significant values. In the present thesis, the pattern of raw Fst was smoothed using a local regression. This technique allowed to obtain more defined signals, also in comparison with the traditional averaging for adjacent windows. In particular, the LOWESS tended to enhance the group of peaks, i.e. regions were several adjacent high Fst signals were detected, whereas isolated peaks tended to be shortened. Thus the effect of hitchhiking, that should be interpreted as an evidence of the existence of a selective sweep in a specific region, is somewhat highlighted. On the other hand, the random variation (that could be the cause of isolated Fst values) is, at least partly, removed. The identification of significant Fst values was carried out using a control chart. This analysis is very rapid to perform and its interpretation is intuitive. The goodness of the approach was confirmed by the detection of genes that are well known to affect milk and beef traits, as the ABCG2 or the Myostatin, and loci involved in the determinism of the coat colour as MC1R and KIT. However, this approach allowed to detect genomic regions not previously associated to

dairy or beef aptitudes as the GREM1 gene involved in the bone metabolism or the SLC40A1 which codes for a protein that regulates the body levels of iron. Another interesting aspect of the genetic diversity between the two considered breeds has been the flagging of the calpain gene family. In spite of its relative simplicity and easiness to use, the combined LOESS-Control chart approach was effective in discriminating salient selection signatures signals from random variation providing an user friendly tool for genetic difference studies.

In the second research, five breeds were considered simultaneously. The measure of diversity used was the composite log likelihood, which compares the allele frequency of the different populations using the binomial distribution for the construction of the null hypothesis. Such a multi comparison yielded interesting results. Also in this case, well studied candidate genes were detected, as the ABCG2 for dairy traits and MSTN for beef trait. However, apart from the sample composition effect, it has to be deeply considered whether evidences of the existence of selection signatures in a species could be adequately provided by the comparison of two populations with divergent phenotype. Genome Wide Analysis studies carried out in cattle have highlighted the poor repeatability of significant SNP across different experiments and populations. The use of HD panels (770K) should provide an higher LD between marker and favorable mutations. In any case, the adopt of a multi population strategy seems a reasonable option. An interesting result of this research has been also the quantification of the signals shared between breeds grouped according to their production aptitude

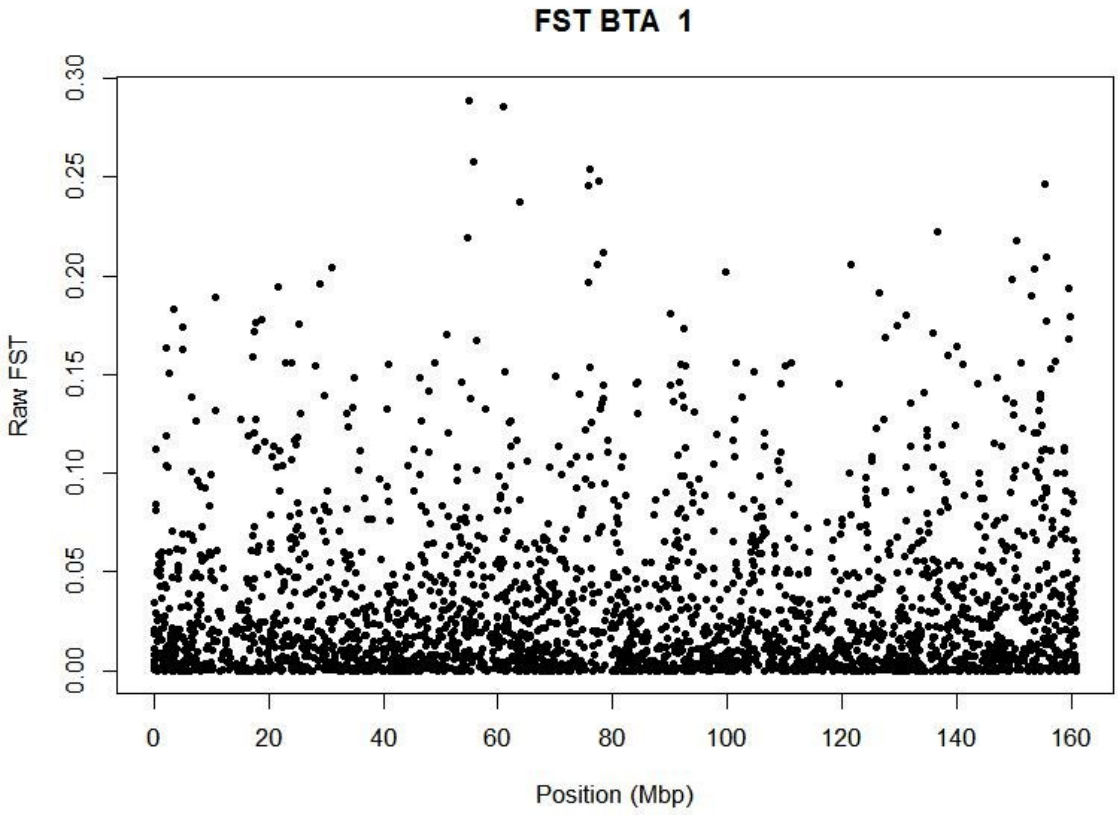
The last experimental contribution was aimed at analysing relationships between a structural element of the genome and the genetic diversity of the five considered breeds, especially as far as their evolution history and their level of inbreeding is concerned. Runs of homozygosity are a powerful tool for estimating inbreeding coefficients, evaluating its effect on traits of economic importance and controlling its level in breeding programs. The distribution of ROH found in the present thesis was in agreement with previous studies carried out in populations of the same breeds farmed in other countries. In particular dairy breeds exhibited a larger occurrence of longer ROH compared to beef and to the dual purpose cattle, respectively, evidencing a more recent inbreeding event. Moreover relationships between pedigree based inbreeding coefficients and ROH based inbreeding were confirmed as their ability to provide indications on the recent selective history of

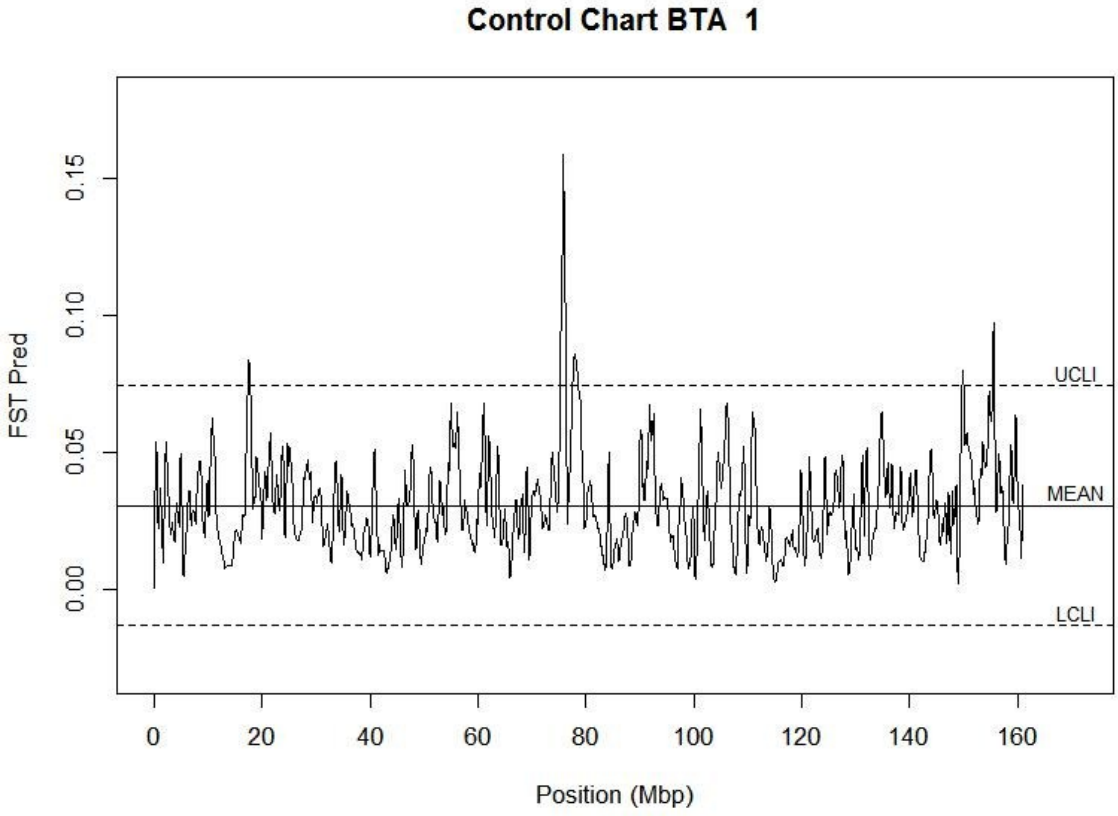
these breeds. One of the advantages of using ROH base inbreeding instead of the pedigree-based is the ability of the former to trace back to ancient inbreeding events, that is constrained in the latter by the depth of the pedigree structure. However, the regression between traditional inbreeding coefficient and the F_{ROH} calculated different sets of ROH according to their minimum length (from >1 Mb to >16 Mb) seem to confirm the hypothesis that including ROH < 4 Mb in the inbreeding calculation generate an overestimation of this coefficient using 54 K SNP panel.

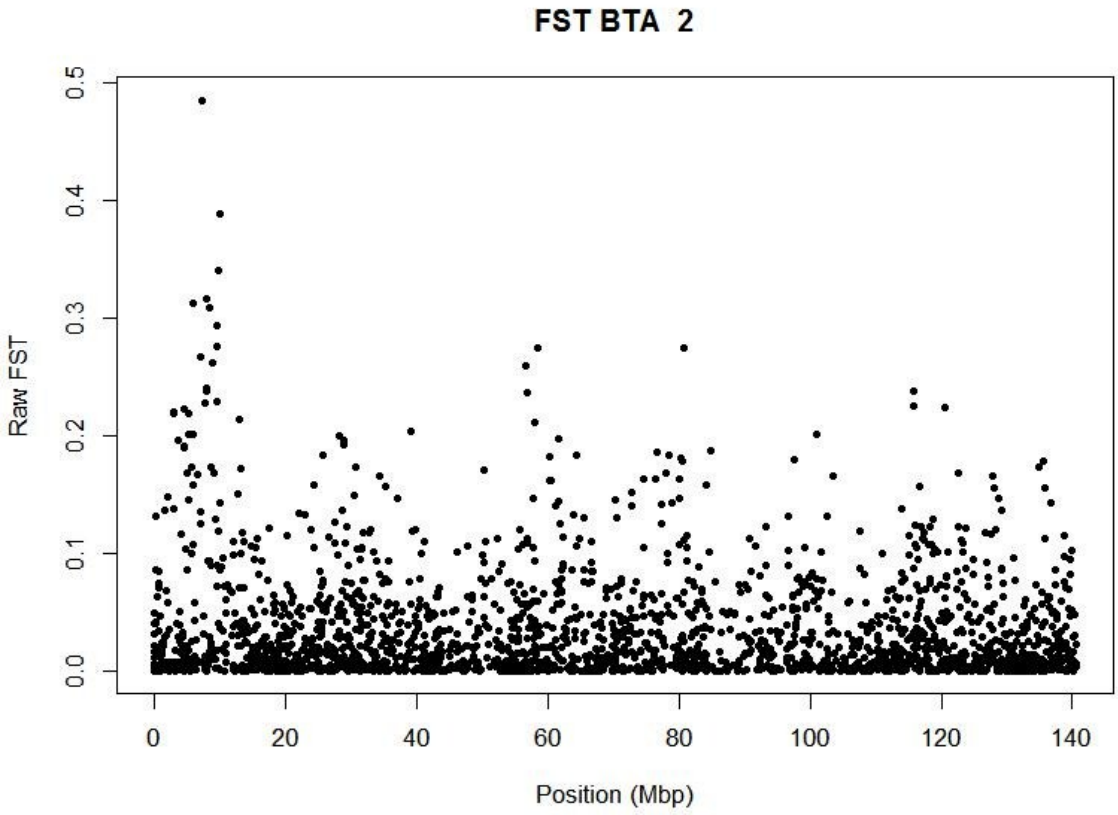
Supplemental Material

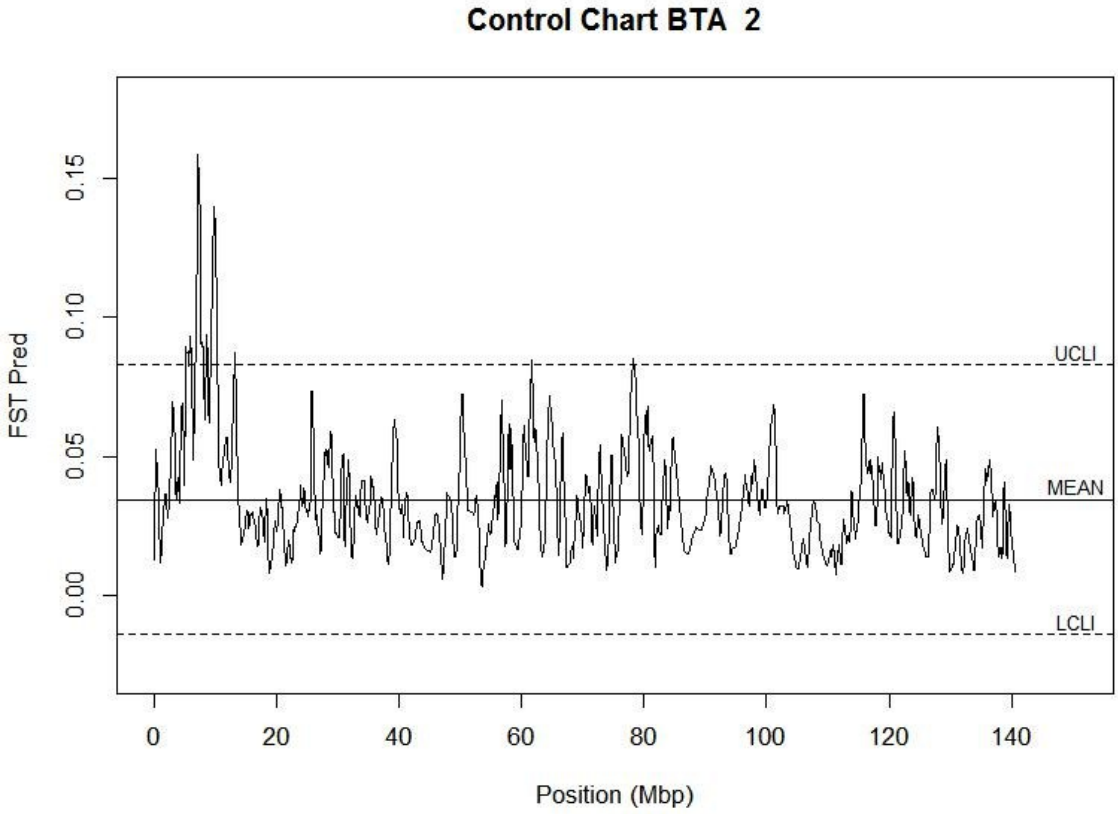
"Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds"

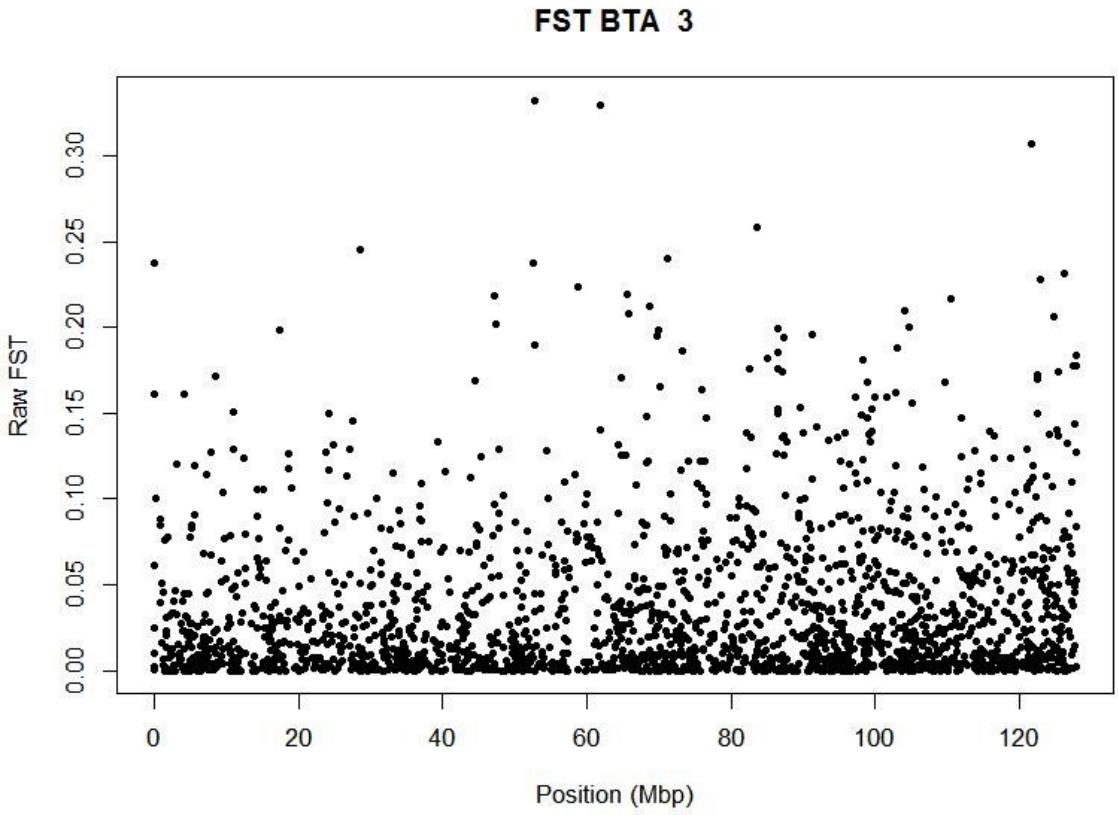
(**Chapter 2**)

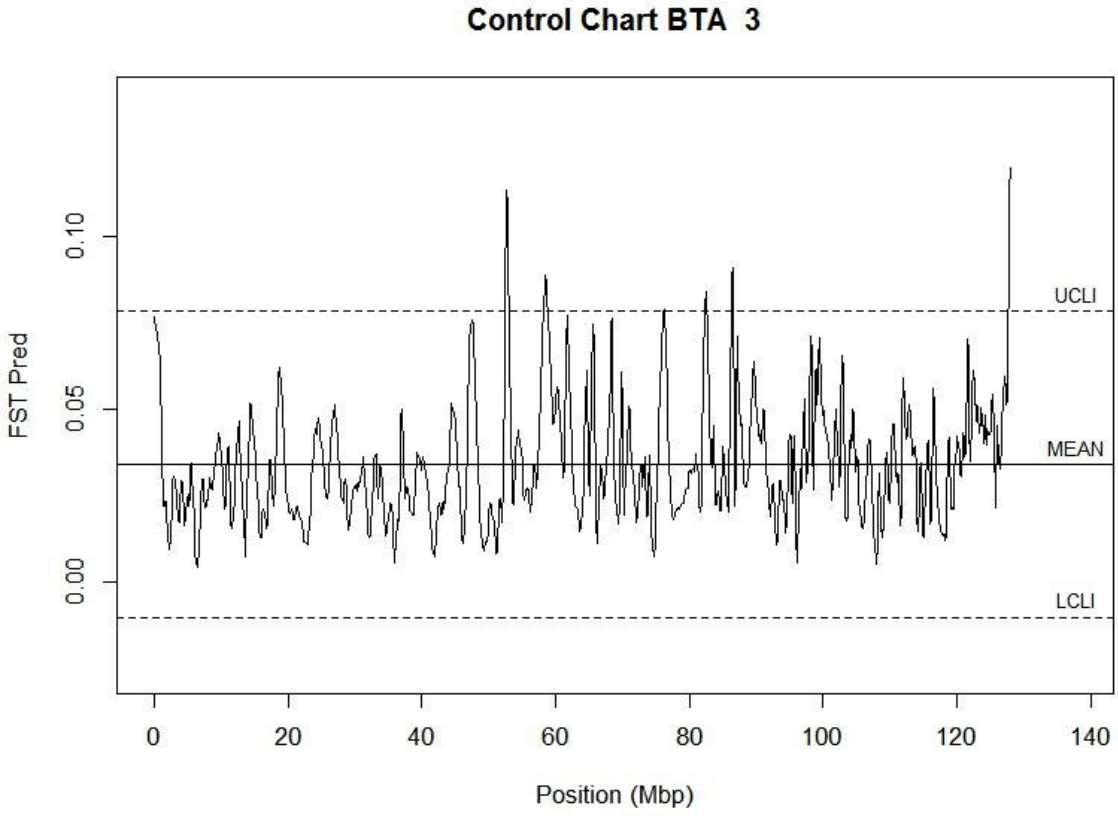

FST BTA 1
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
20
40
60
80
100
120
140
160
Position (Mbp)

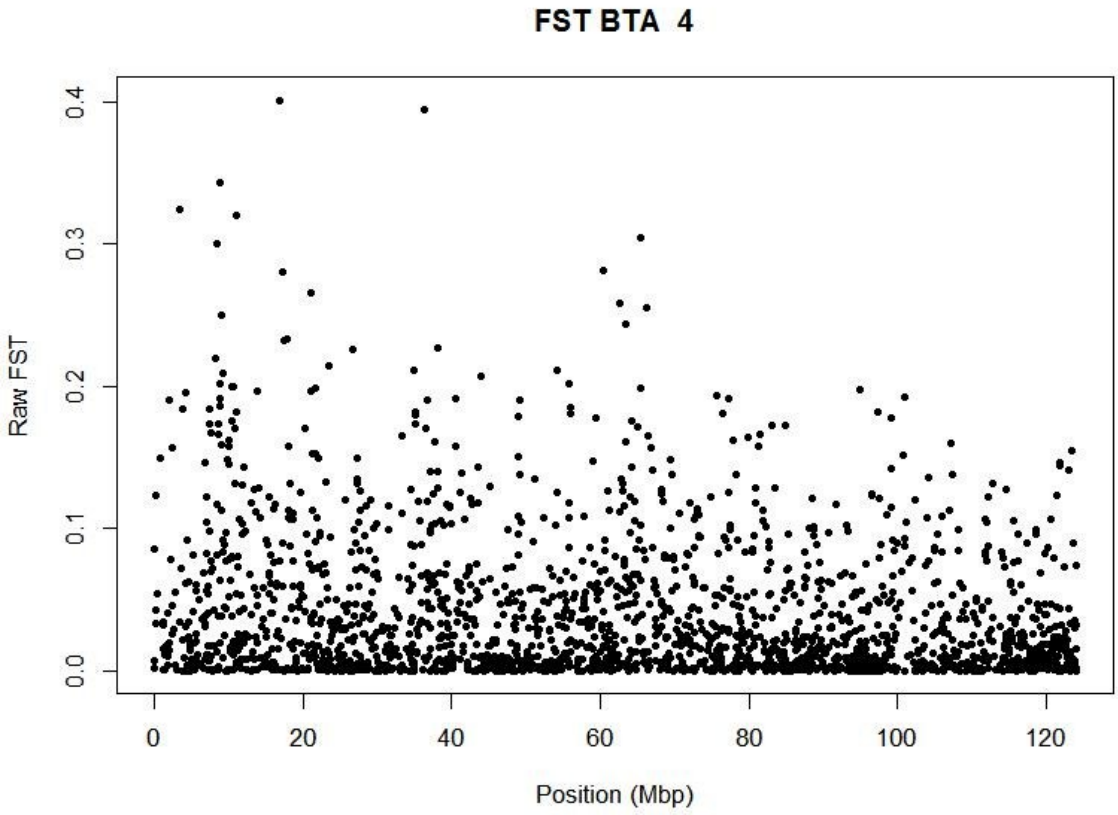

Control Chart BTA 1
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
140
160
Position (Mbp)

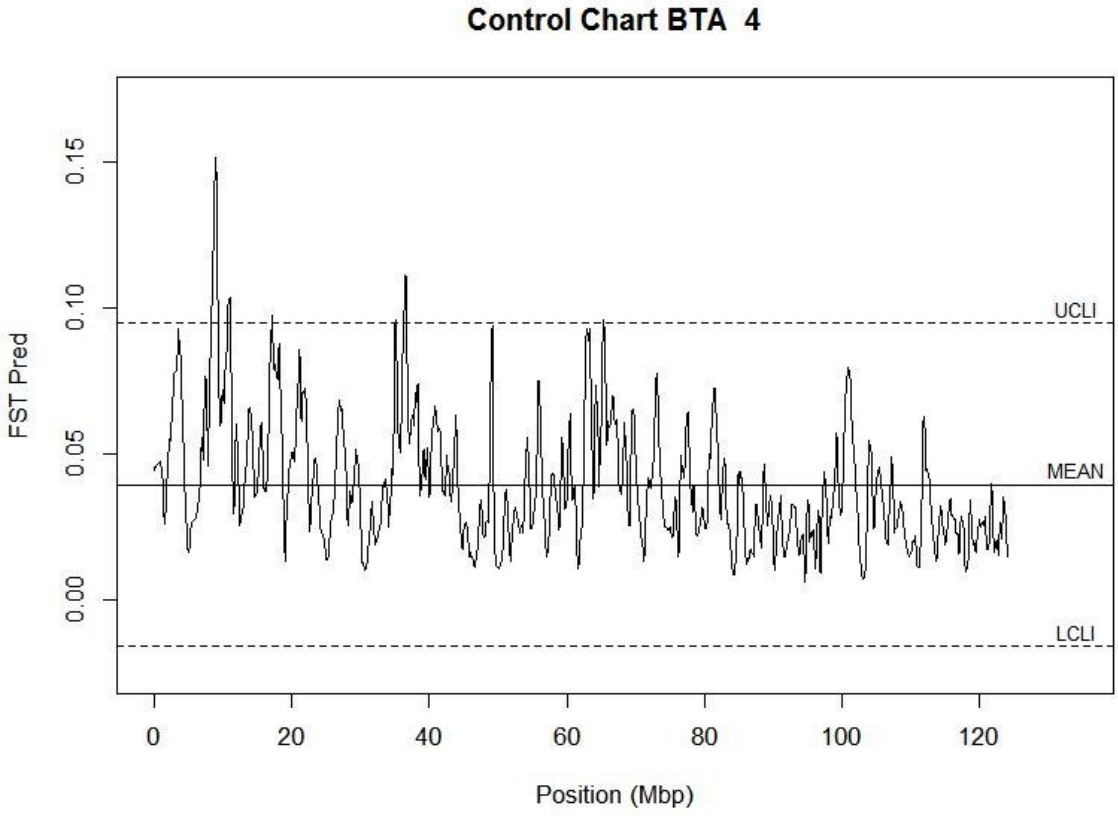

FST BTA 2
Raw FST
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100
120
140
Position (Mbp)

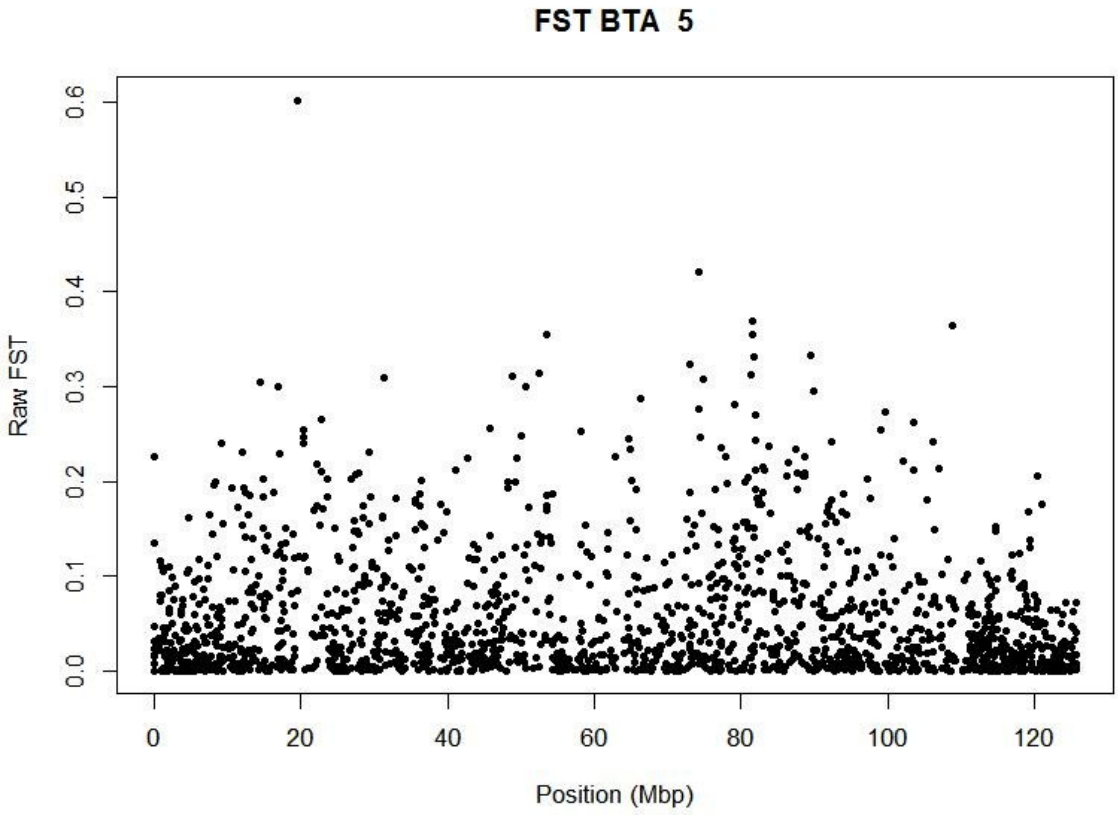

Control Chart BTA 2
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
140
Position (Mbp)

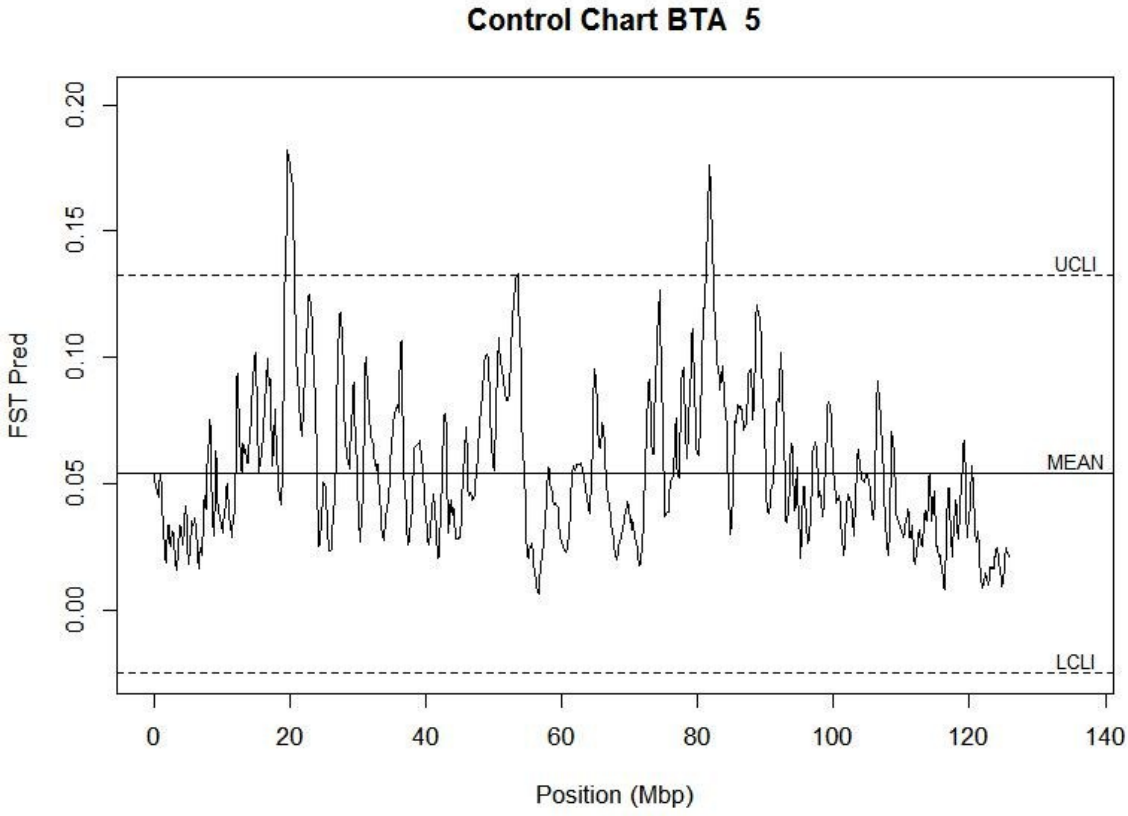

FST BTA 3
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
20
40
60
80
100
120
Position (Mbp)

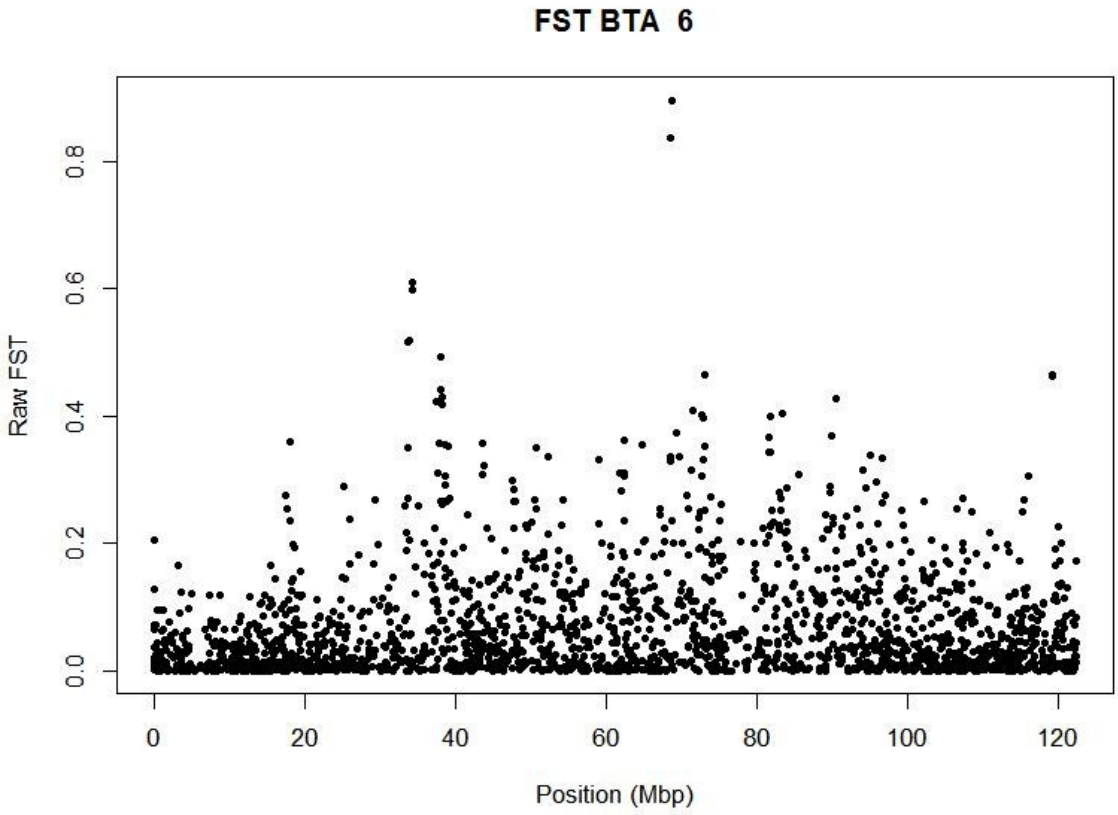

Control Chart BTA 3
FST Pred
0.00
0.05
0.10
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
140
Position (Mbp)

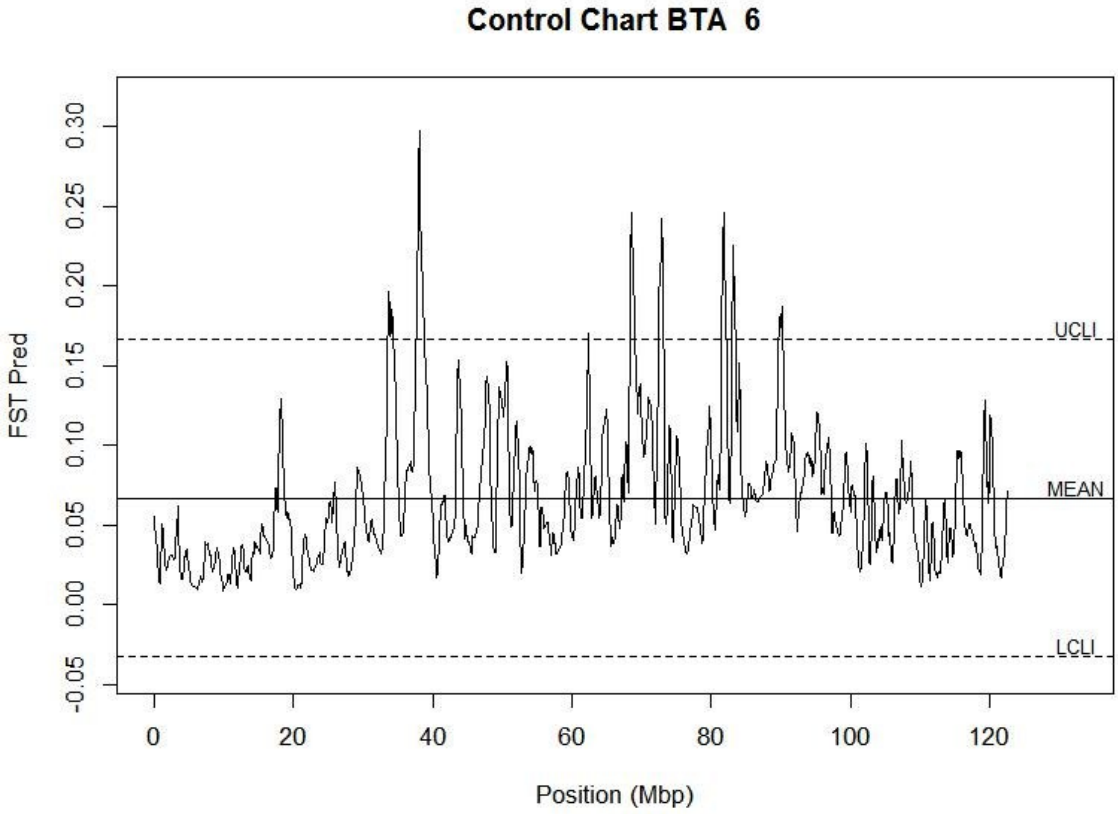

FST BTA 4
Raw FST
0.0
0.1
0.2
0.3
0.4
0
20
40
60
80
100
120
Position (Mbp)

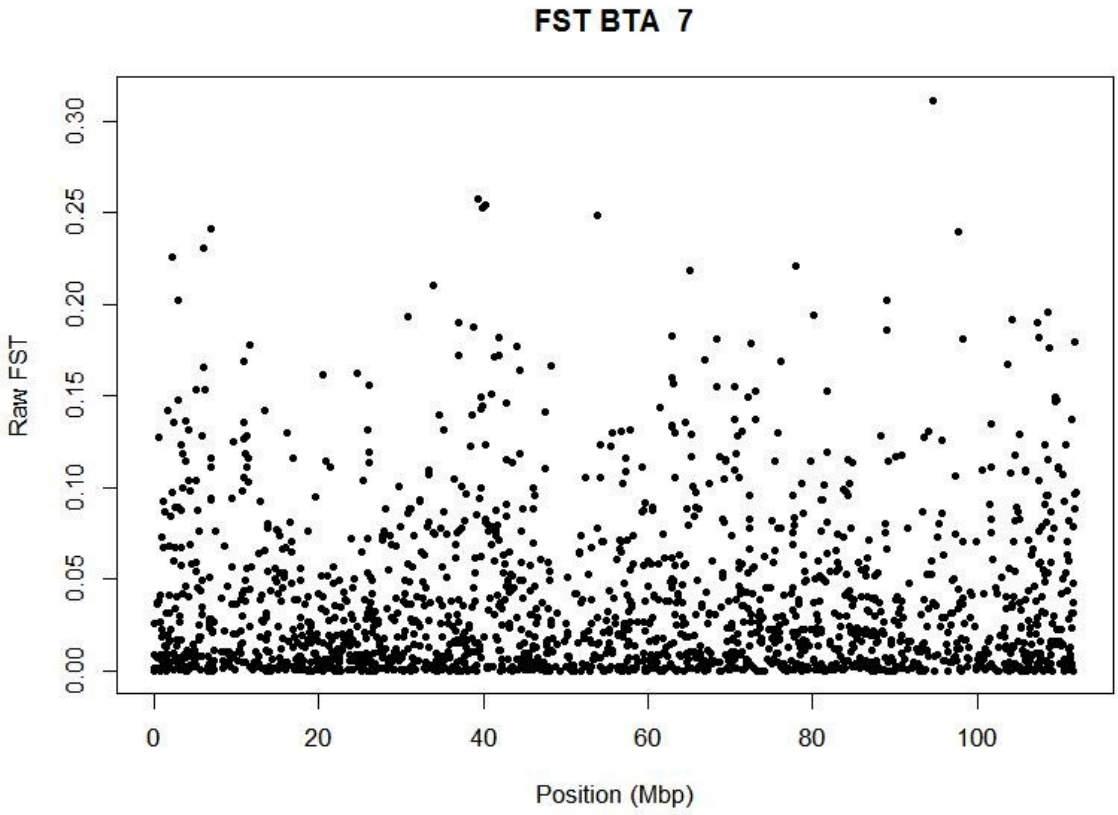

Control Chart BTA 4
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
Position (Mbp)

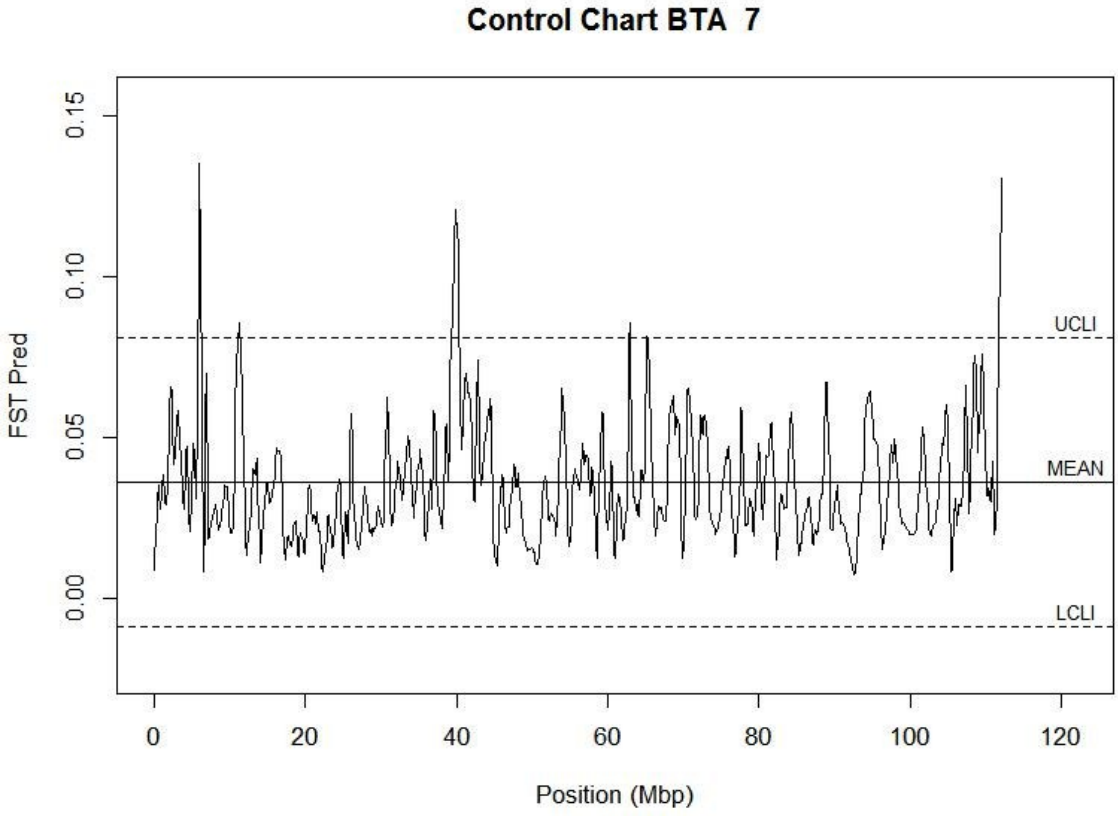

FST BTA 5
Raw FST
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
120
Position (Mbp)

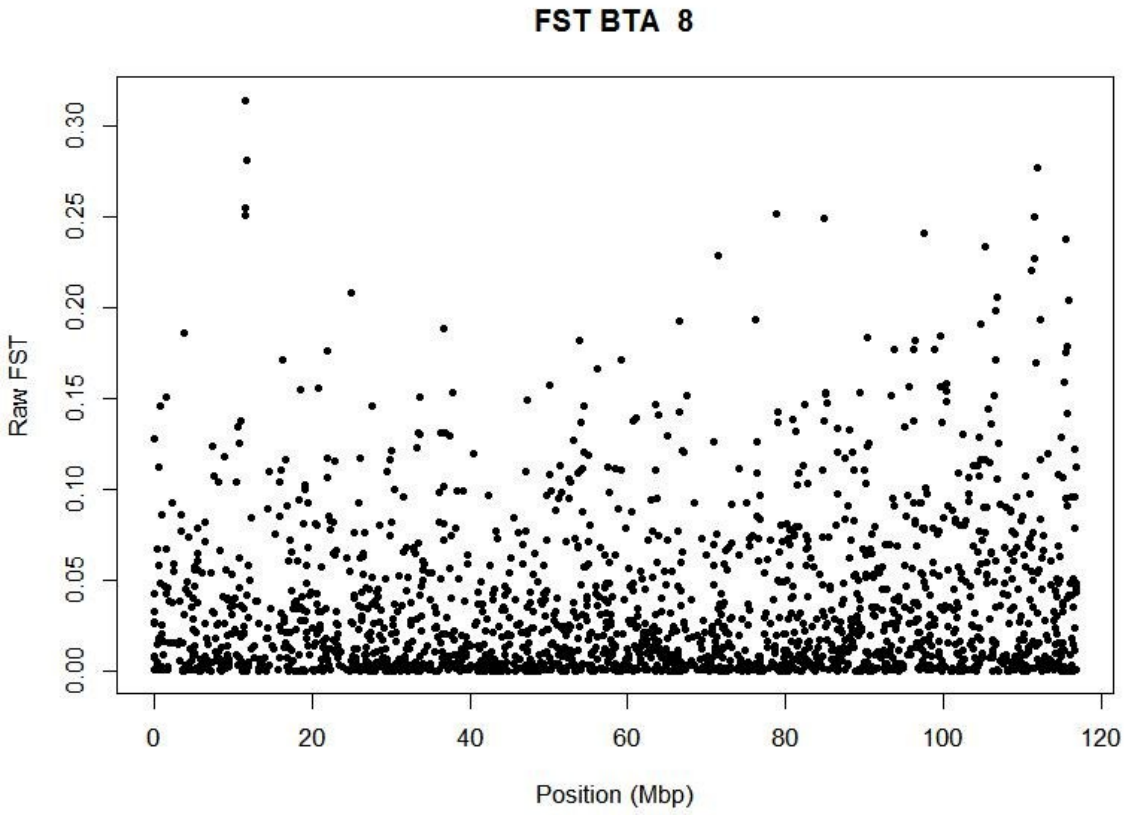

Control Chart BTA 5
FST Pred
0.00
0.05
0.10
0.15
0.20
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
140
Position (Mbp)

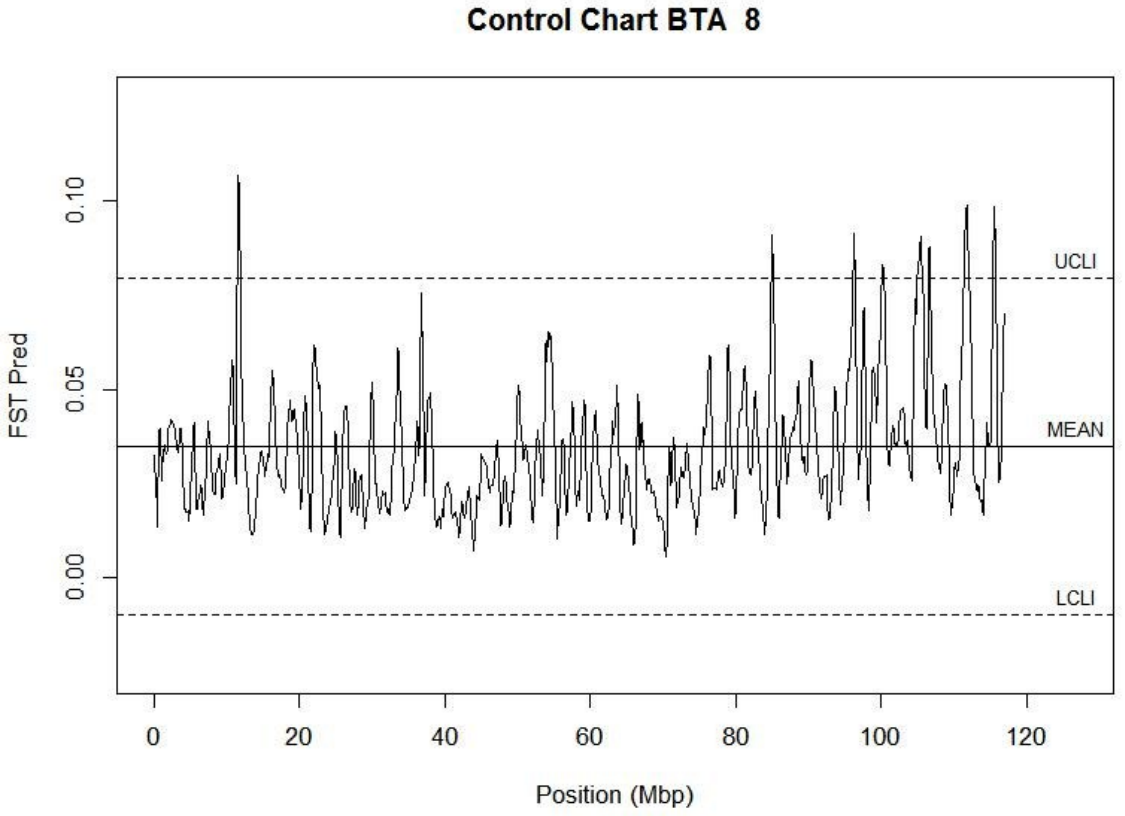

FST BTA 6
Raw FST
0.0
0.2
0.4
0.6
0.8
0
20
40
60
80
100
120
Position (Mbp)

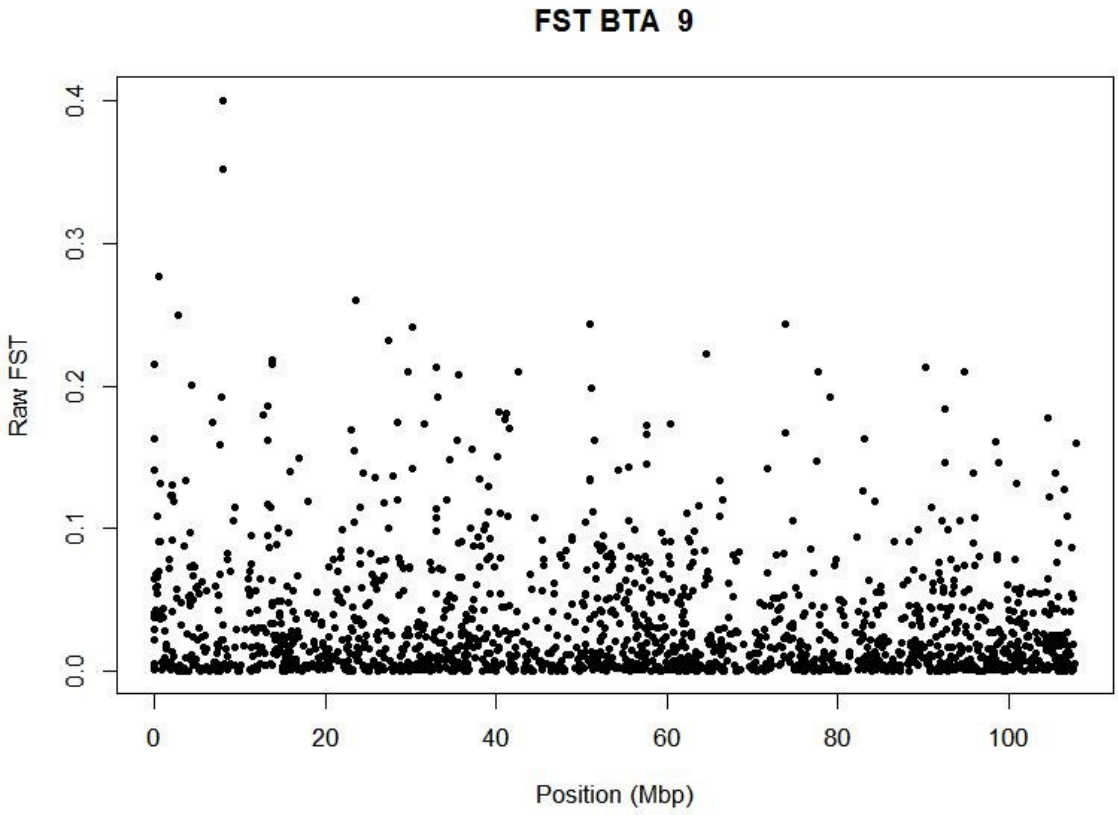

Control Chart BTA 6
FST Pred
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
0.30
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
Position (Mbp)

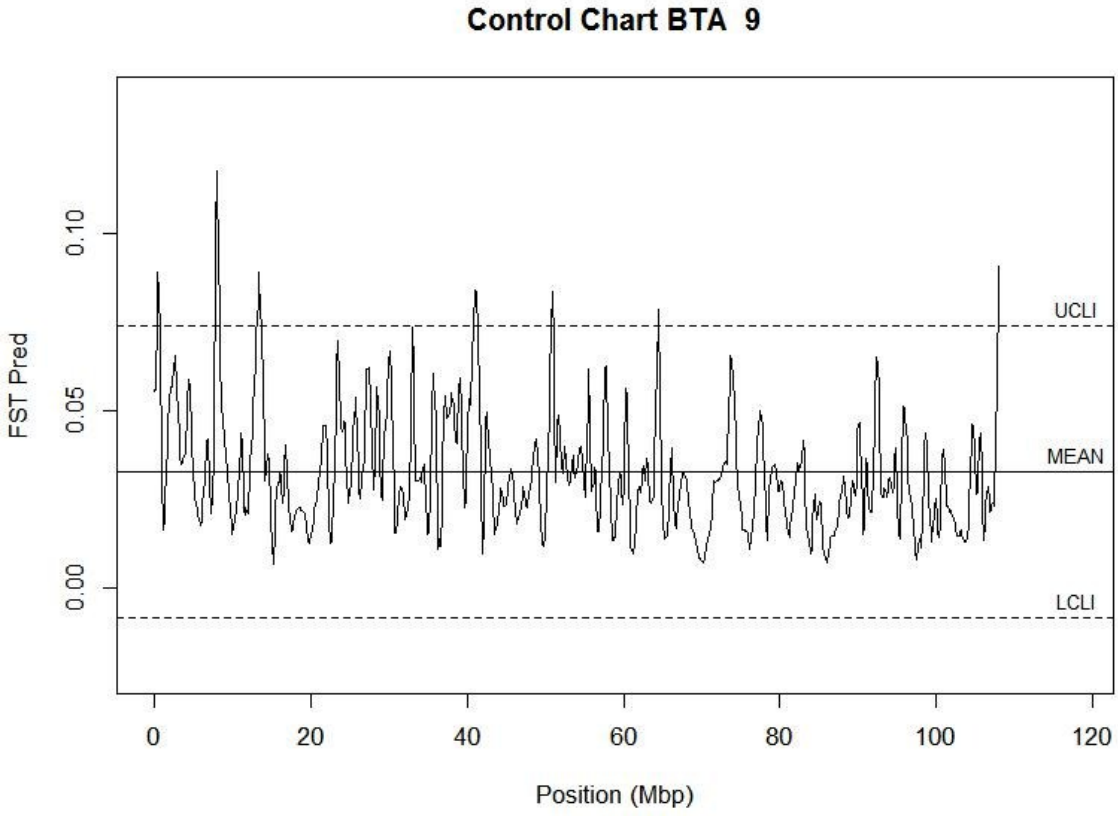

FST BTA 7
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
20
40
60
80
100
Position (Mbp)

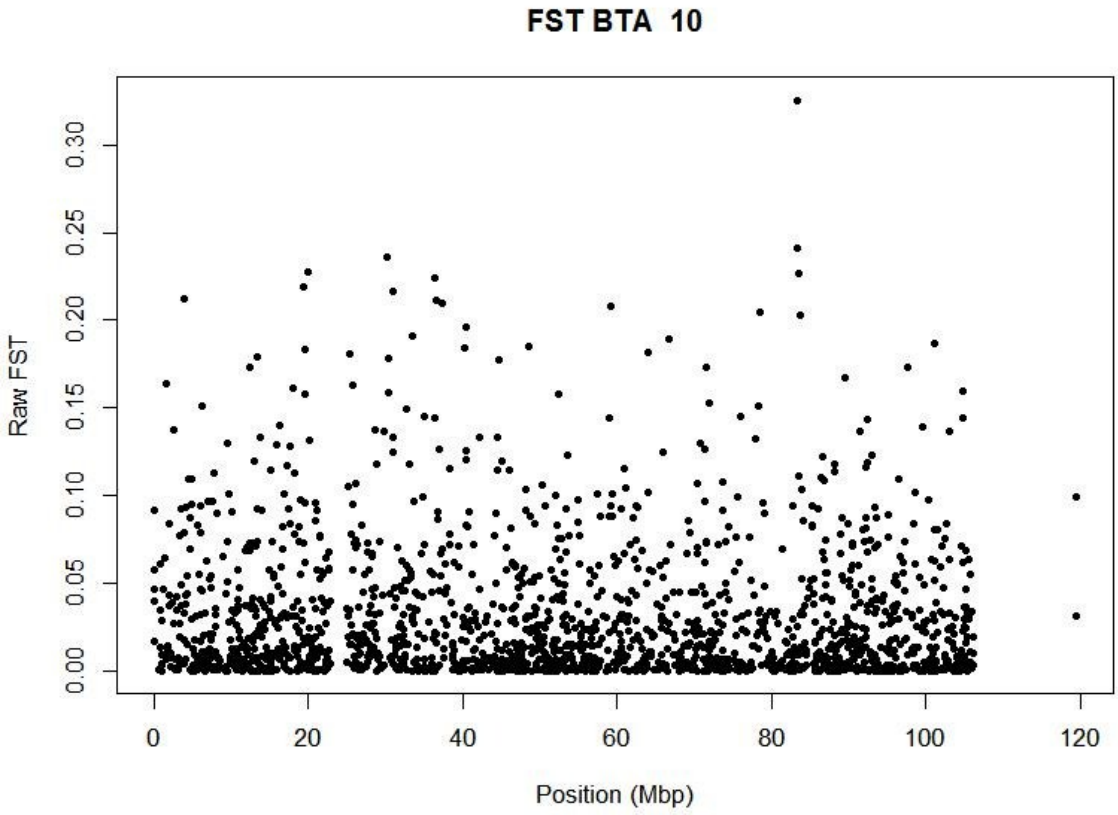

Control Chart BTA 7
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
Position (Mbp)

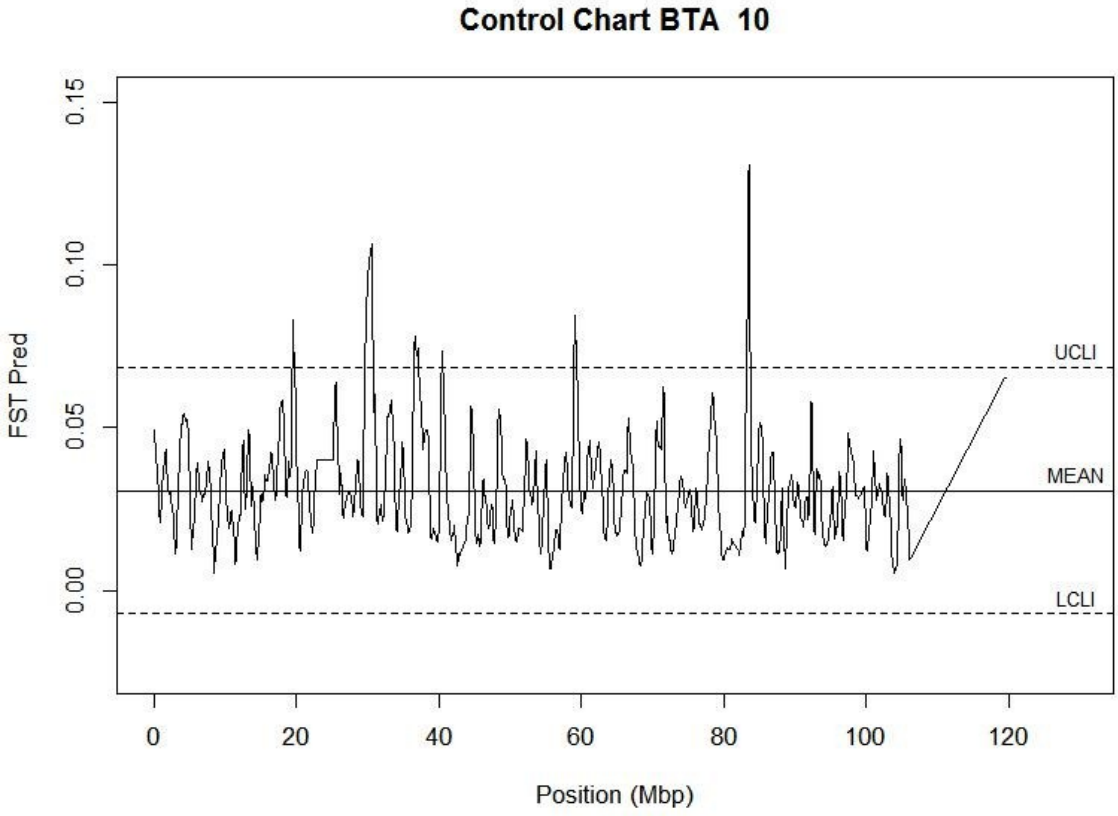

FST BTA 8
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
20
40
60
80
100
120
Position (Mbp)

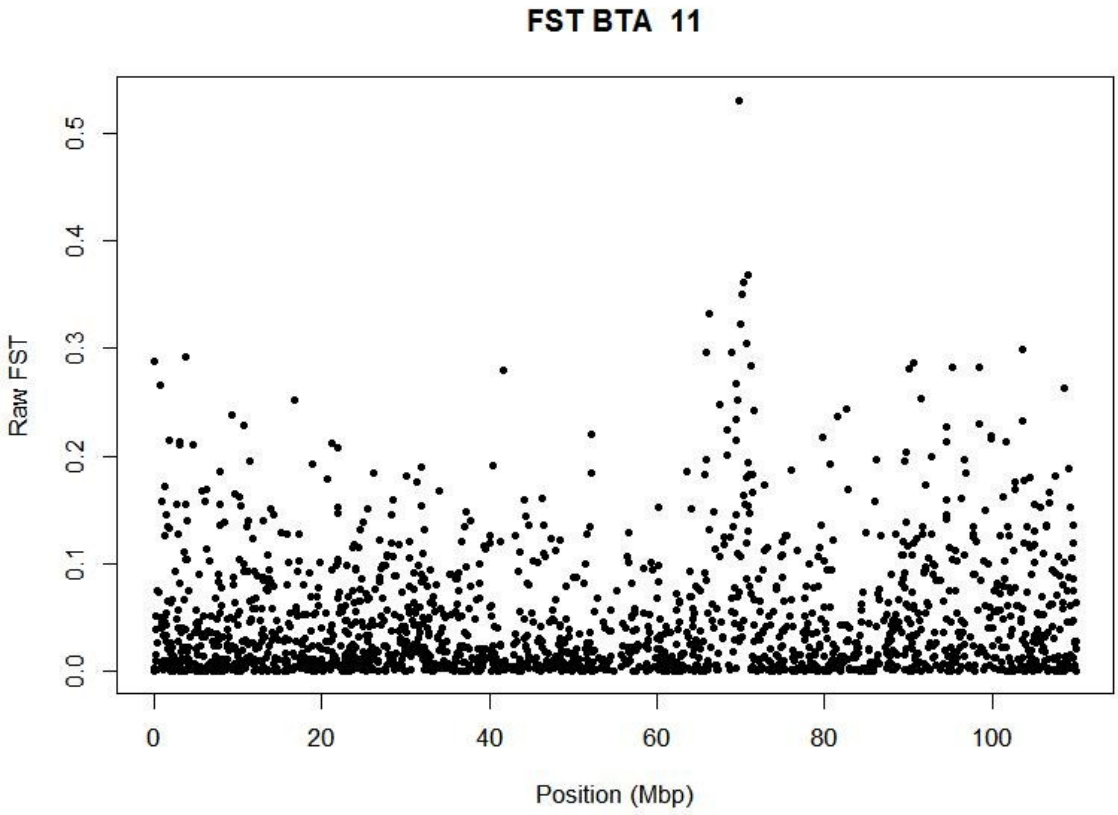

Control Chart BTA 8
FST Pred
0.00
0.05
0.10
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
Position (Mbp)

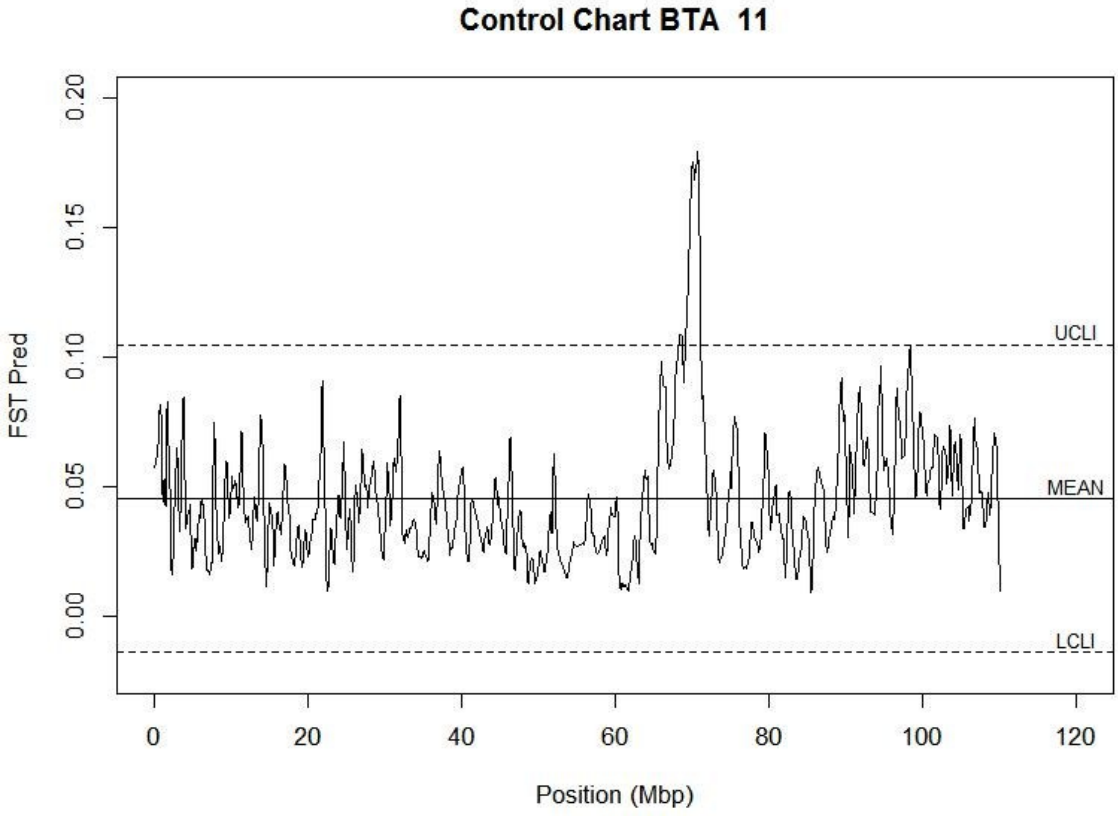

FST BTA 9
Raw FST
0.0
0.1
0.2
0.3
0.4
0
20
40
60
80
100
Position (Mbp)

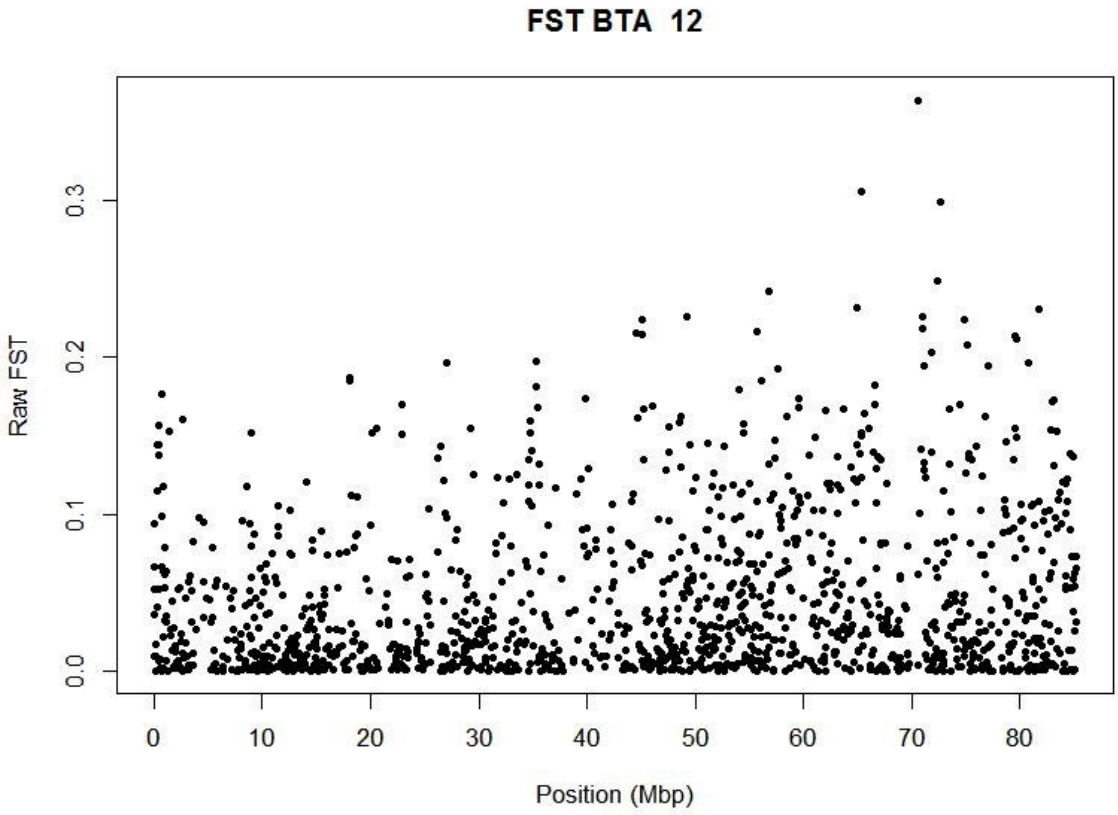

Control Chart BTA 9
FST Pred
0.00
0.05
0.10
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
Position (Mbp)

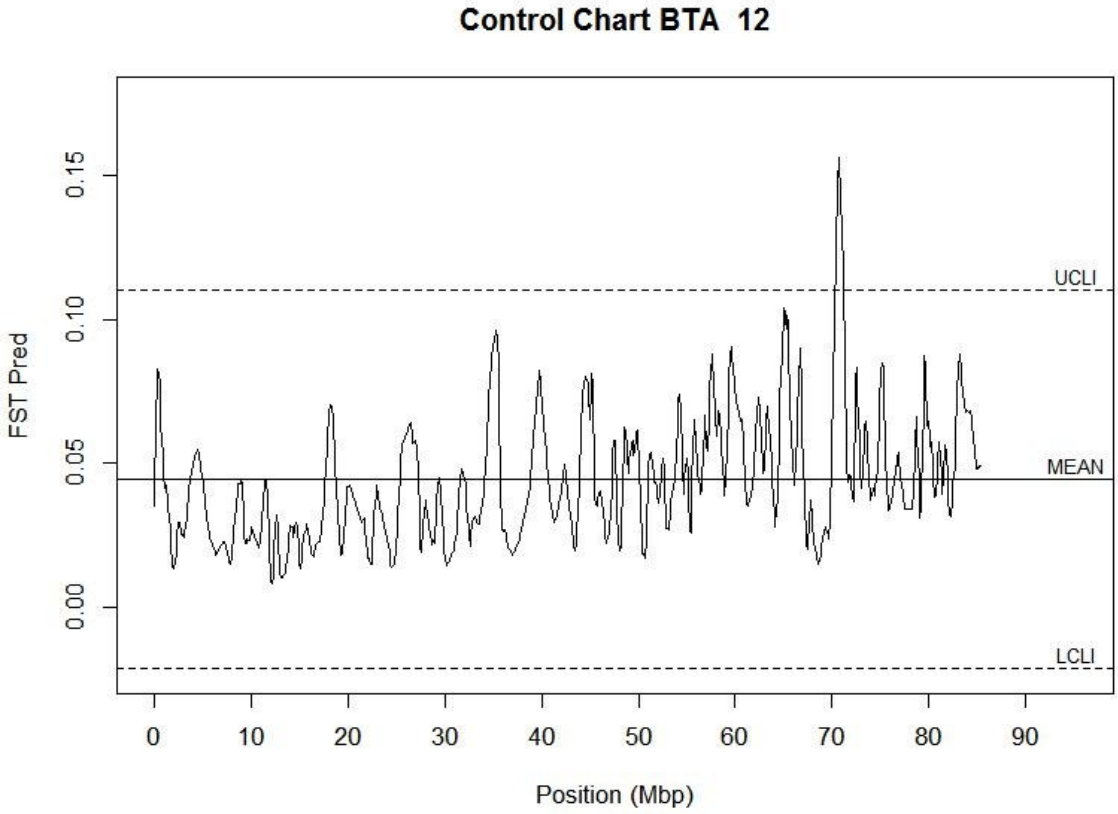

FST BTA 10
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
20
40
60
80
100
120
Position (Mbp)

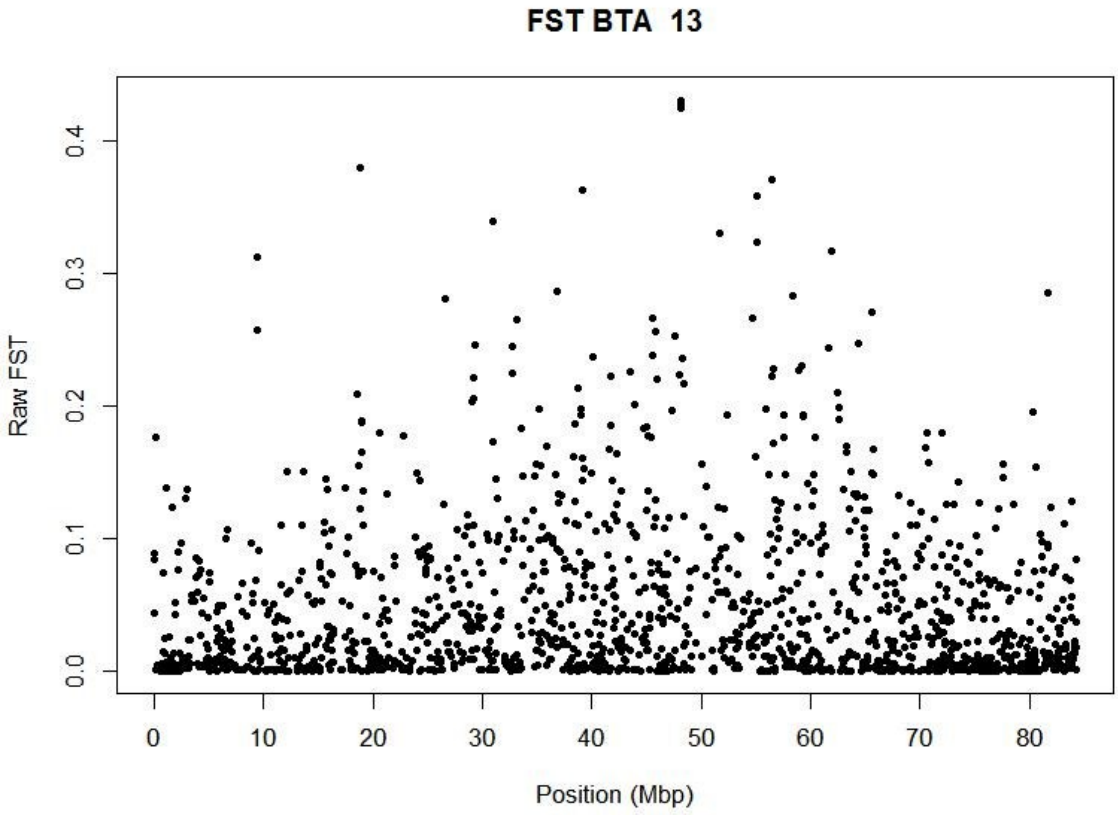

Control Chart BTA 10
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
Position (Mbp)

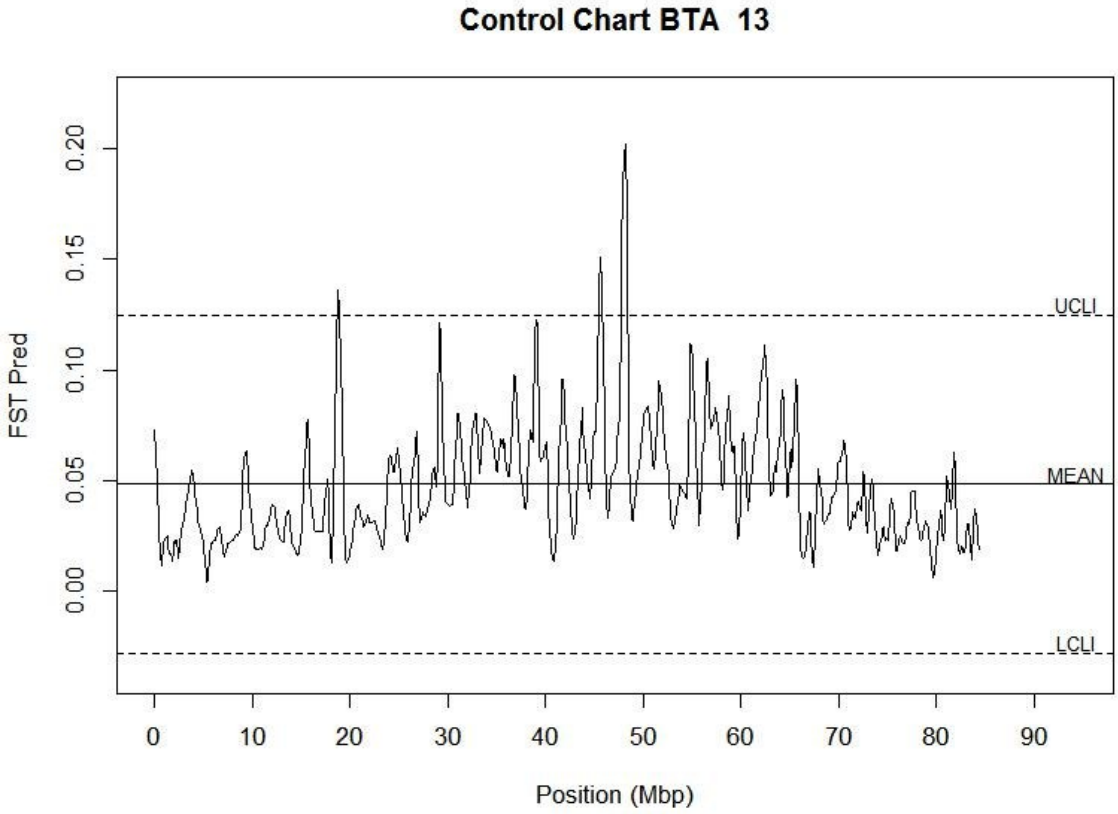

FST BTA 11
Raw FST
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100
Position (Mbp)

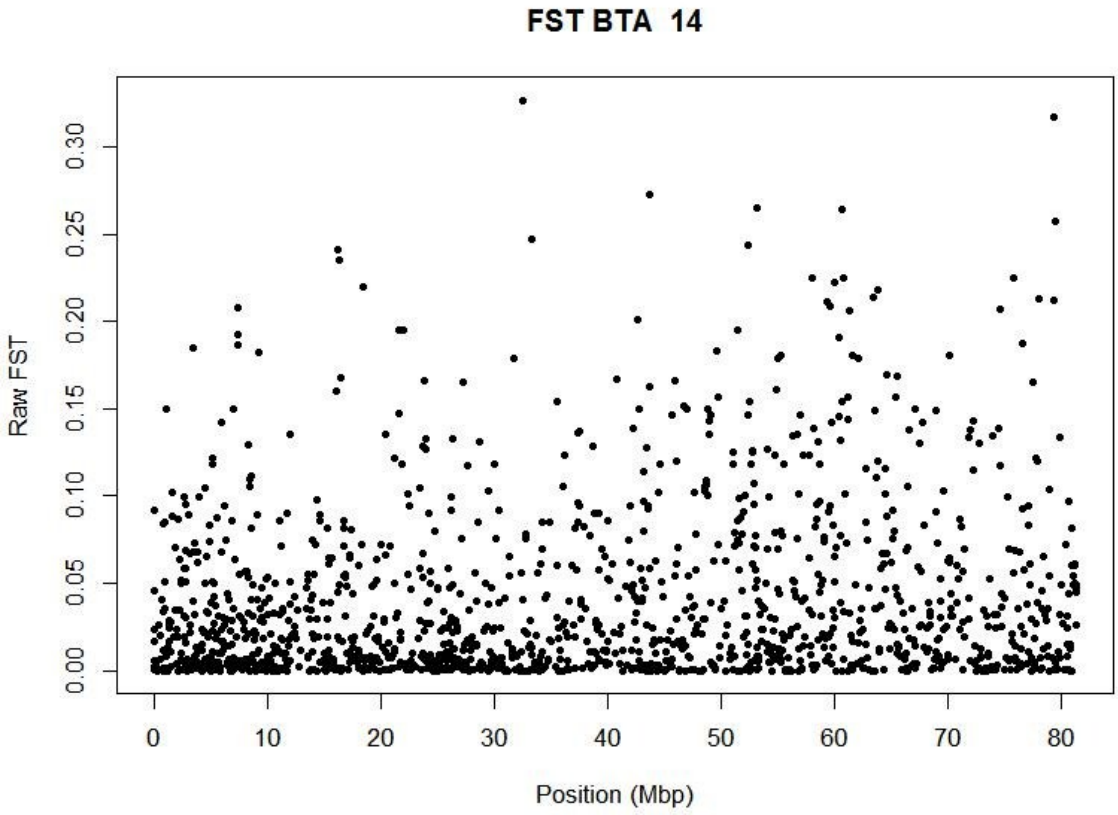

Control Chart BTA 11
FST Pred
0.00
0.05
0.10
0.15
0.20
UCLI
MEAN
LCLI
0
20
40
60
80
100
120
Position (Mbp)

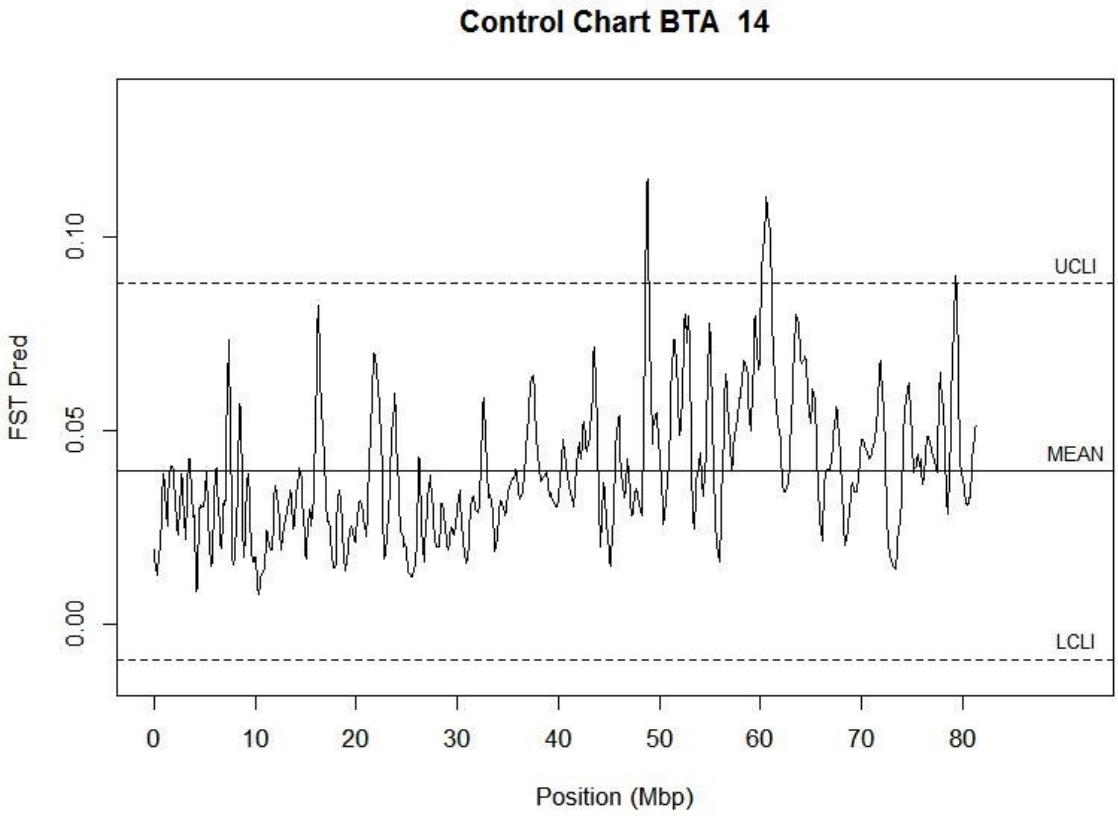

FST BTA 12
Raw FST
0.0
0.1
0.2
0.3
0
10
20
30
40
50
60
70
80
Position (Mbp)

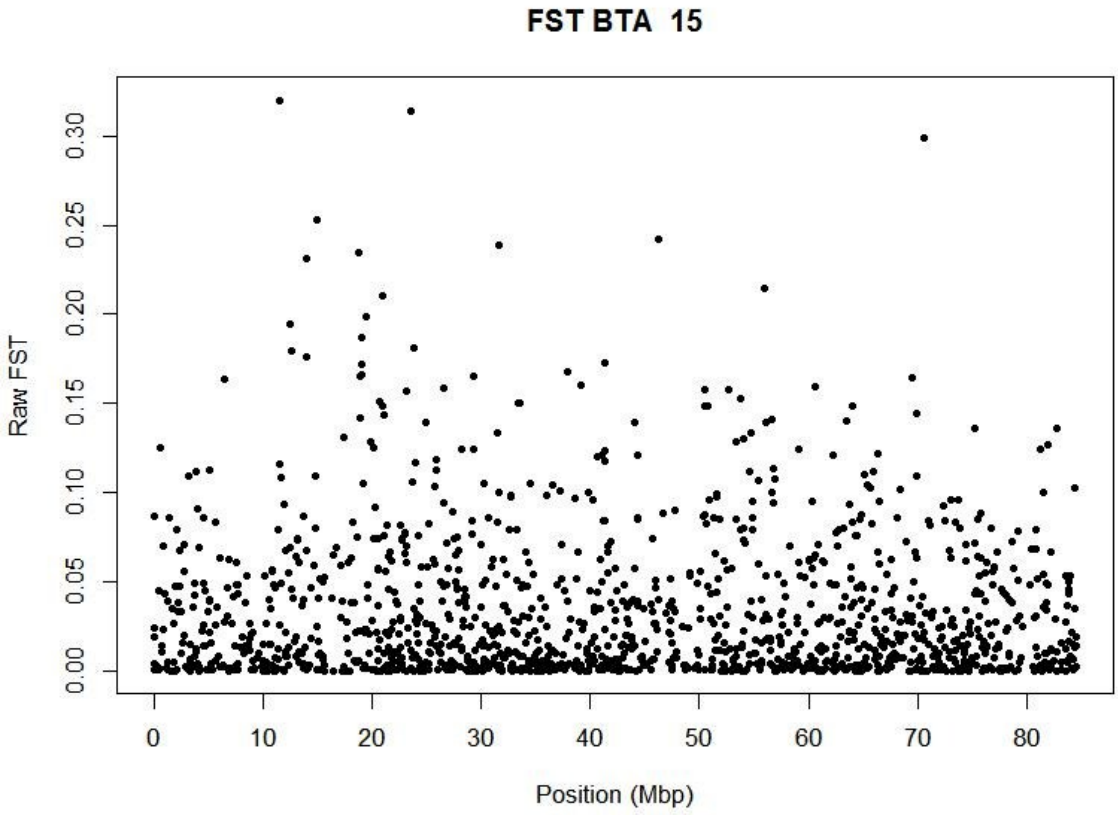

Control Chart BTA 12
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
80
90
Position (Mbp)

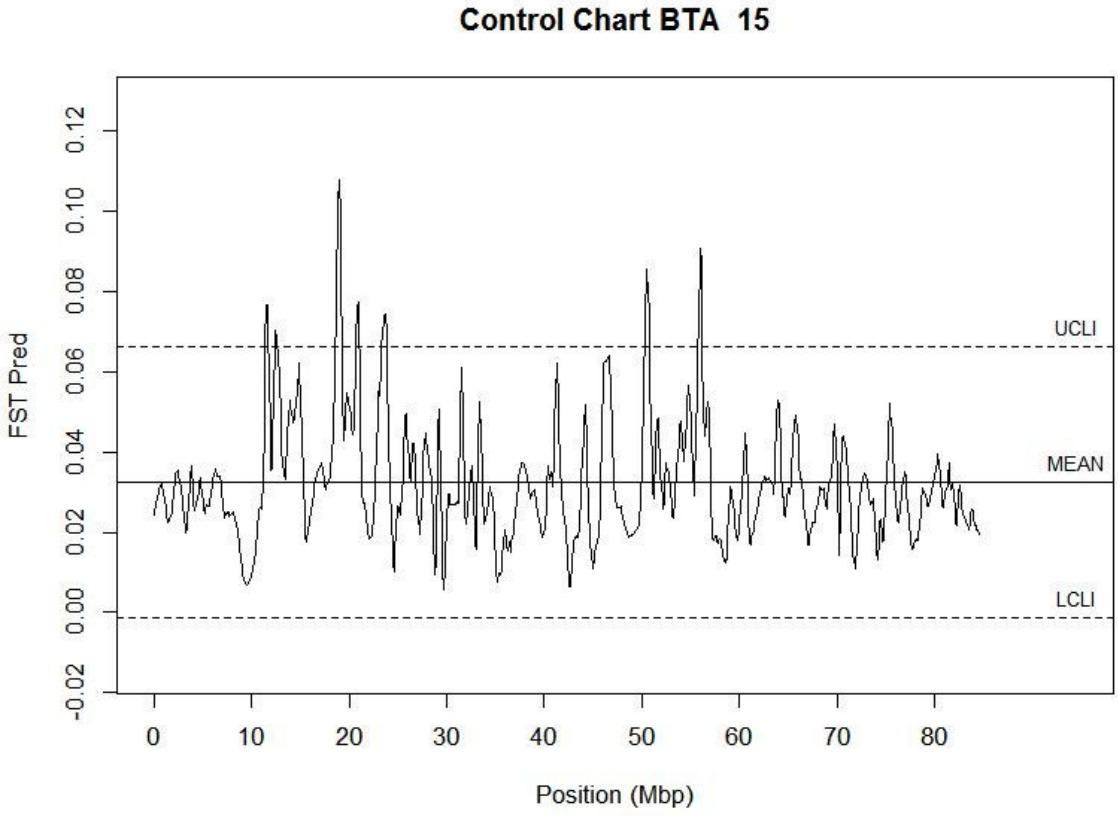

FST BTA 13
Raw FST
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
60
70
80
Position (Mbp)

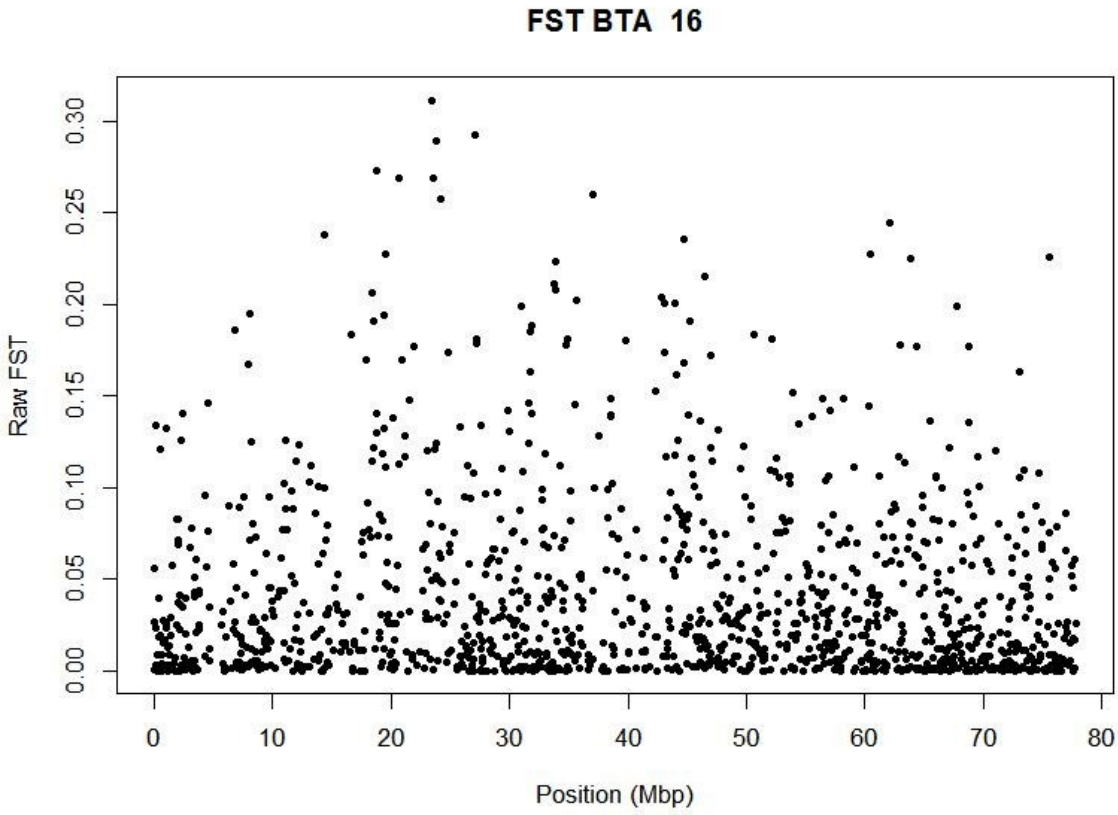

Control Chart BTA 13
FST Pred
0.00
0.05
0.10
0.15
0.20
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
80
90
Position (Mbp)

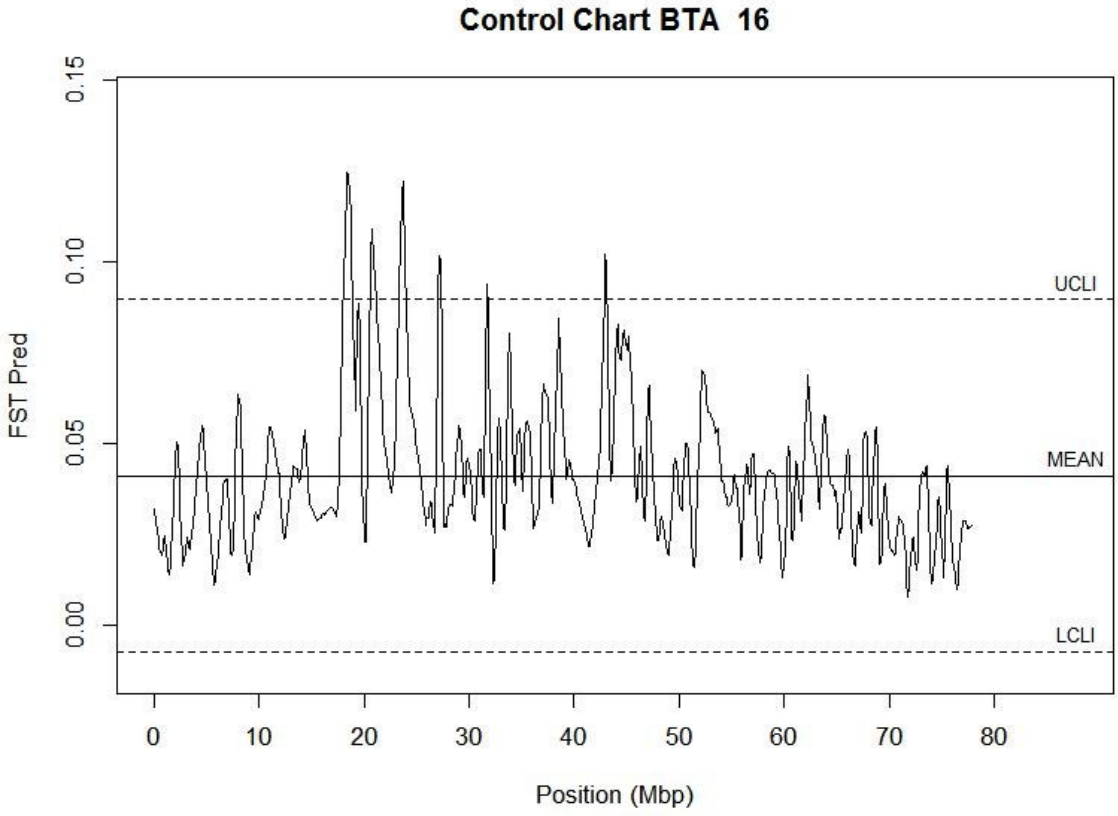

FST BTA 14
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
60
70
80
Position (Mbp)

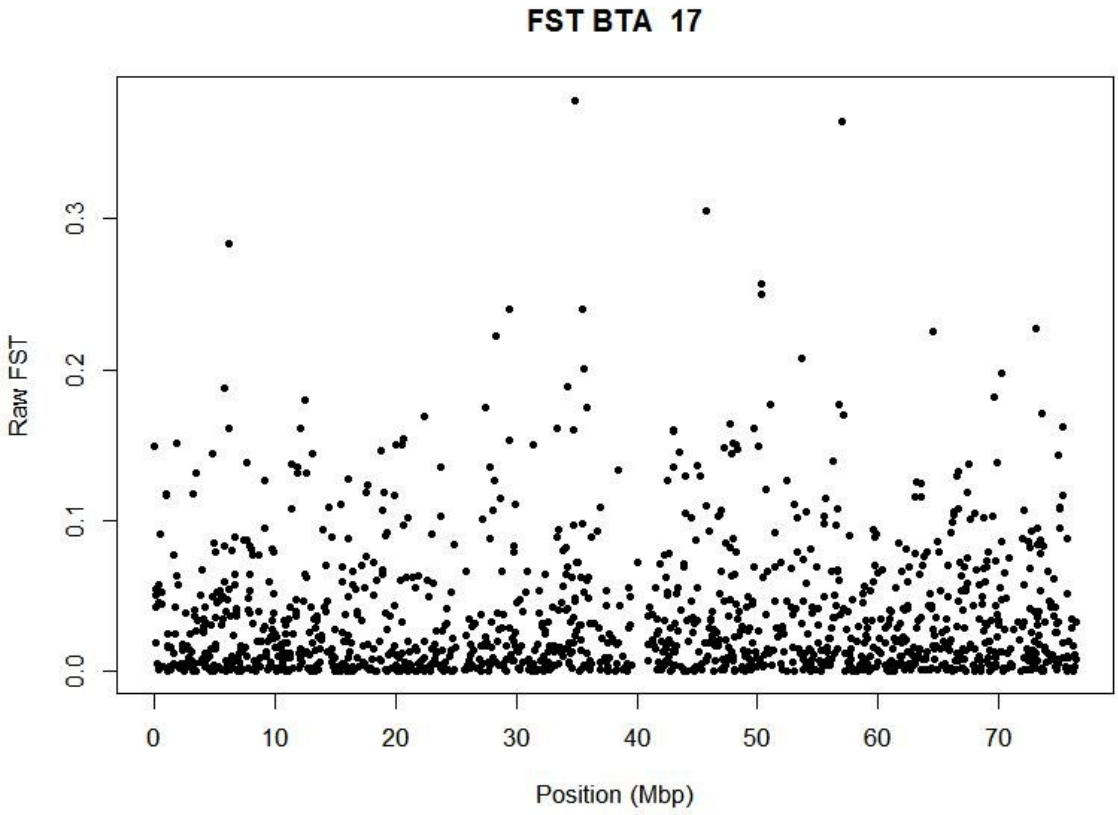

Control Chart BTA 14
FST Pred
0.00
0.05
0.10
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
80
Position (Mbp)

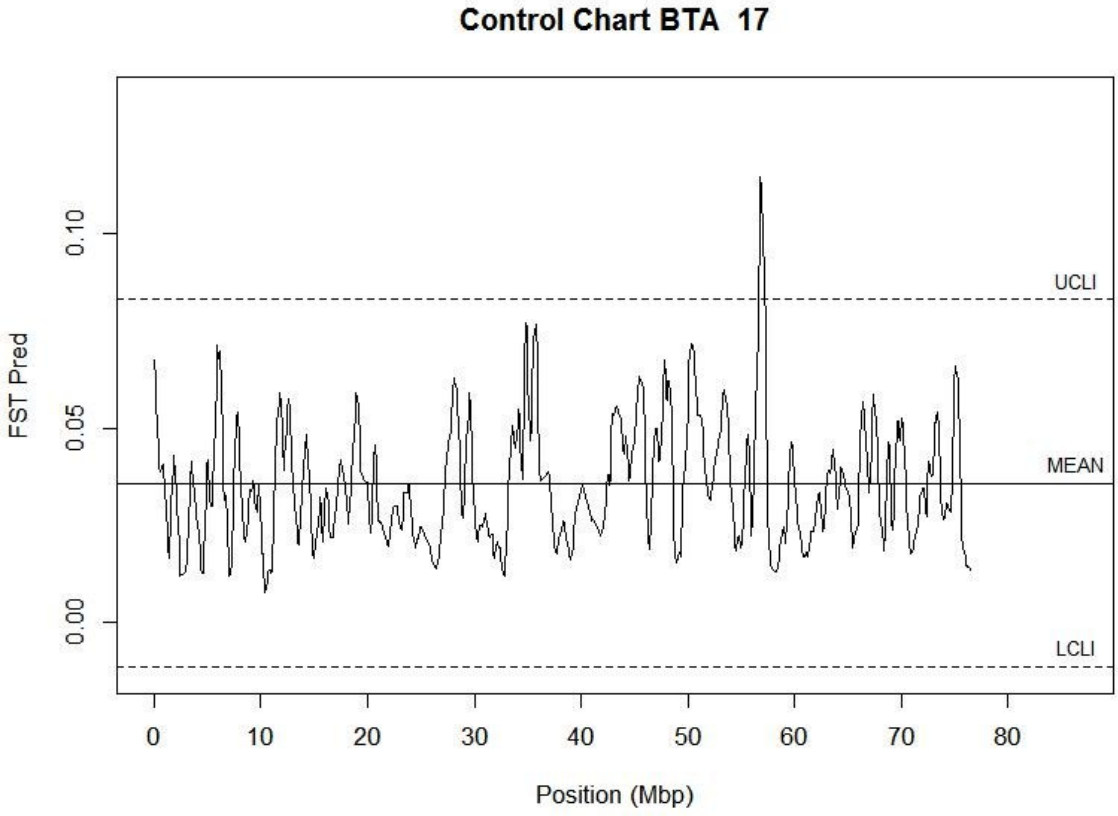

FST BTA 15
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
60
70
80
Position (Mbp)

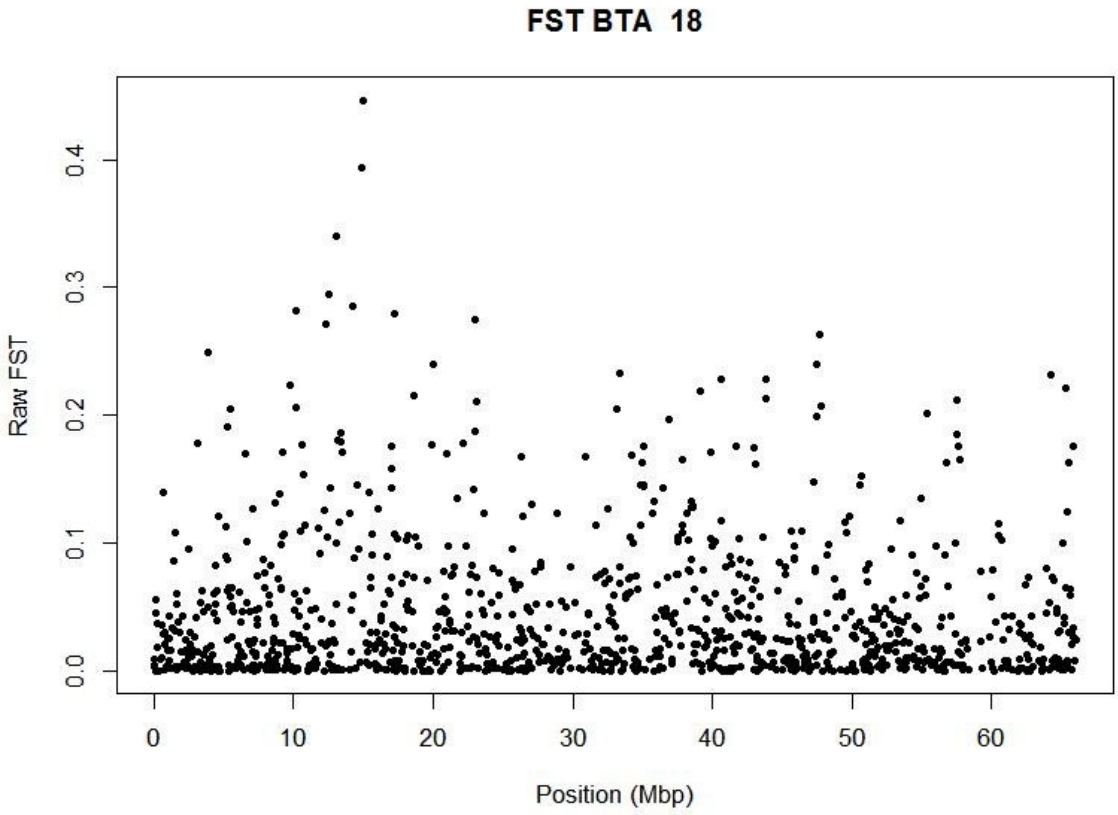

Control Chart BTA 15
FST Pred
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
0.12
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
80
Position (Mbp)

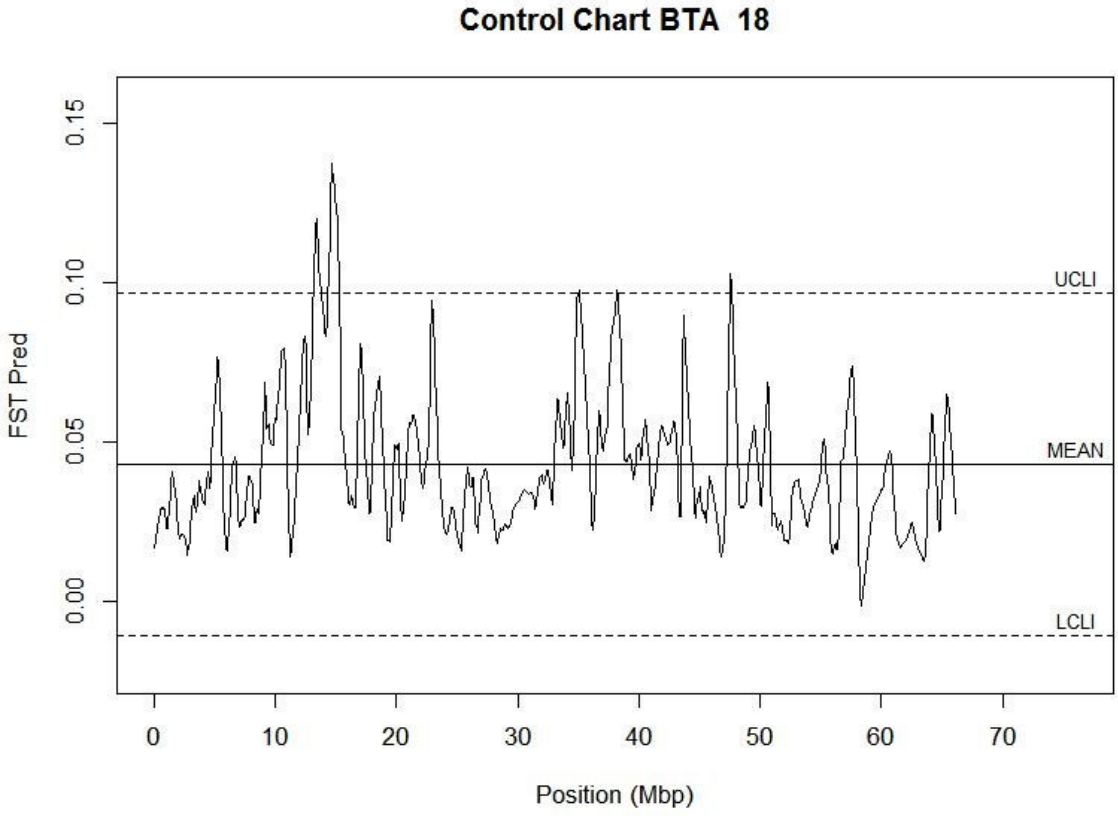

FST BTA 16
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
60
70
80
Position (Mbp)

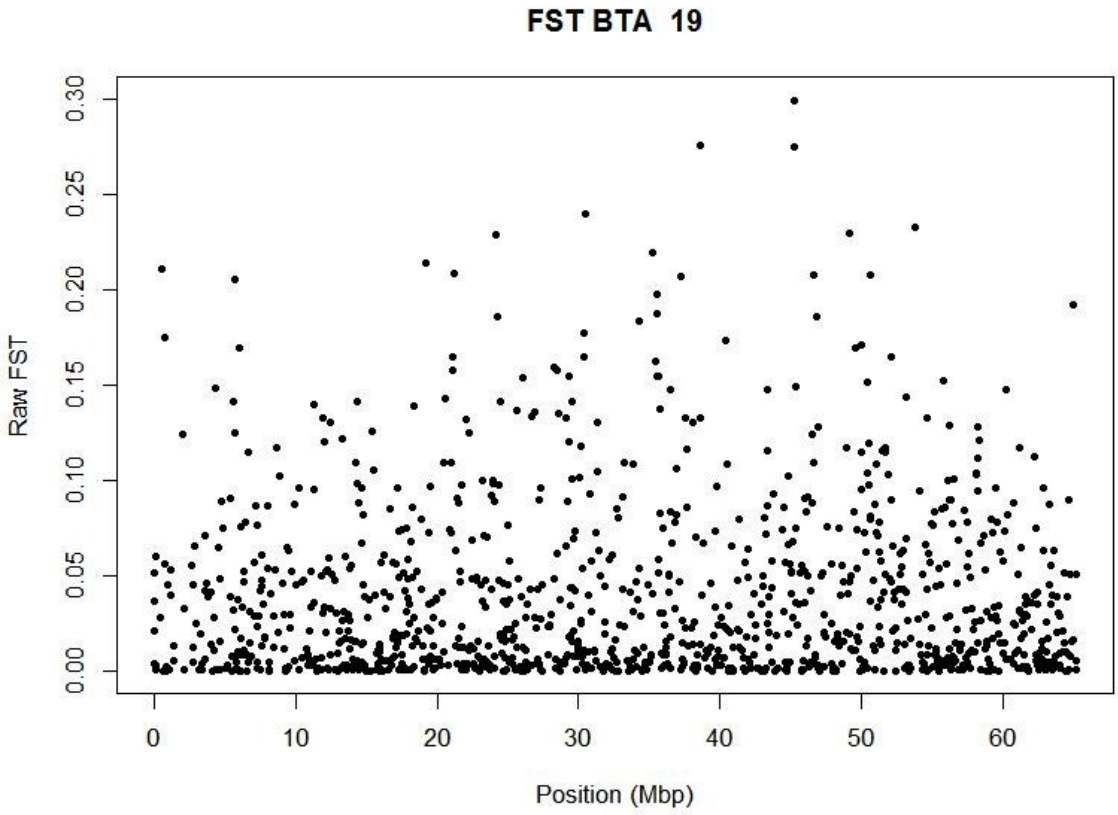

Control Chart BTA 16
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
80
Position (Mbp)

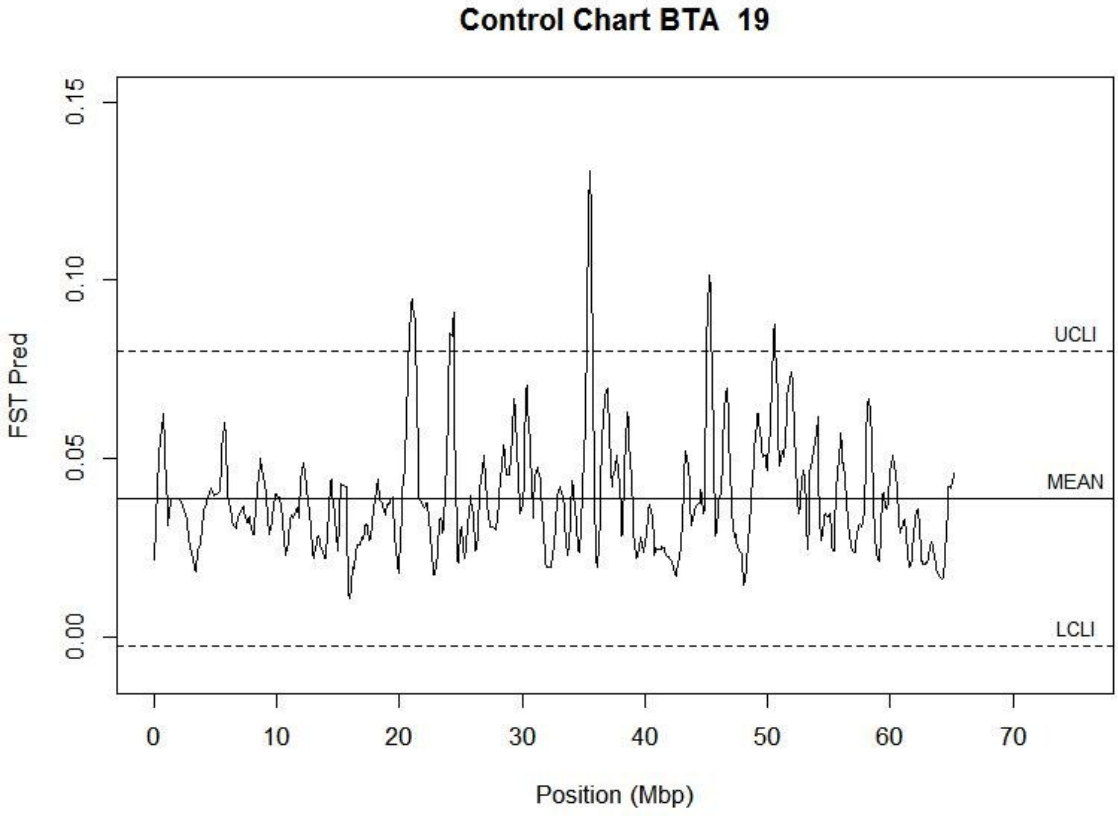

FST BTA 17
Raw FST
0.0
0.1
0.2
0.3
0
10
20
30
40
50
60
70
Position (Mbp)

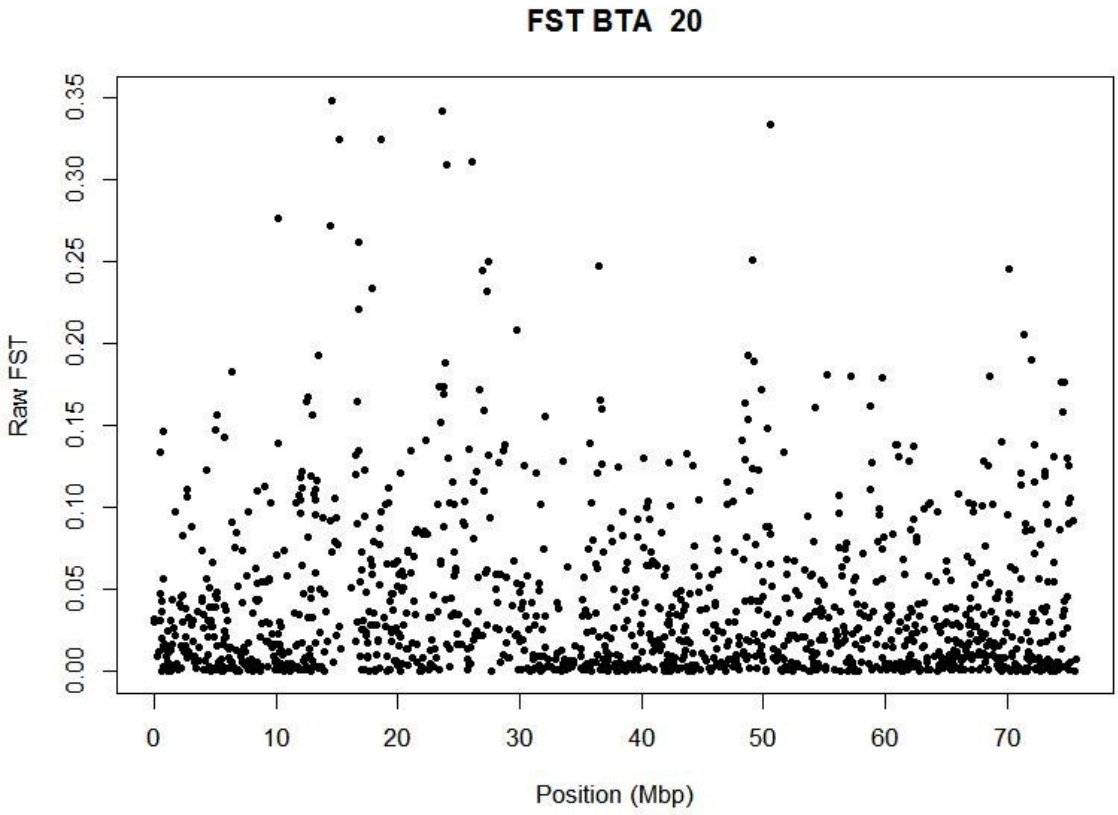

Control Chart BTA 17
FST Pred
0.00
0.05
0.10
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
80
Position (Mbp)

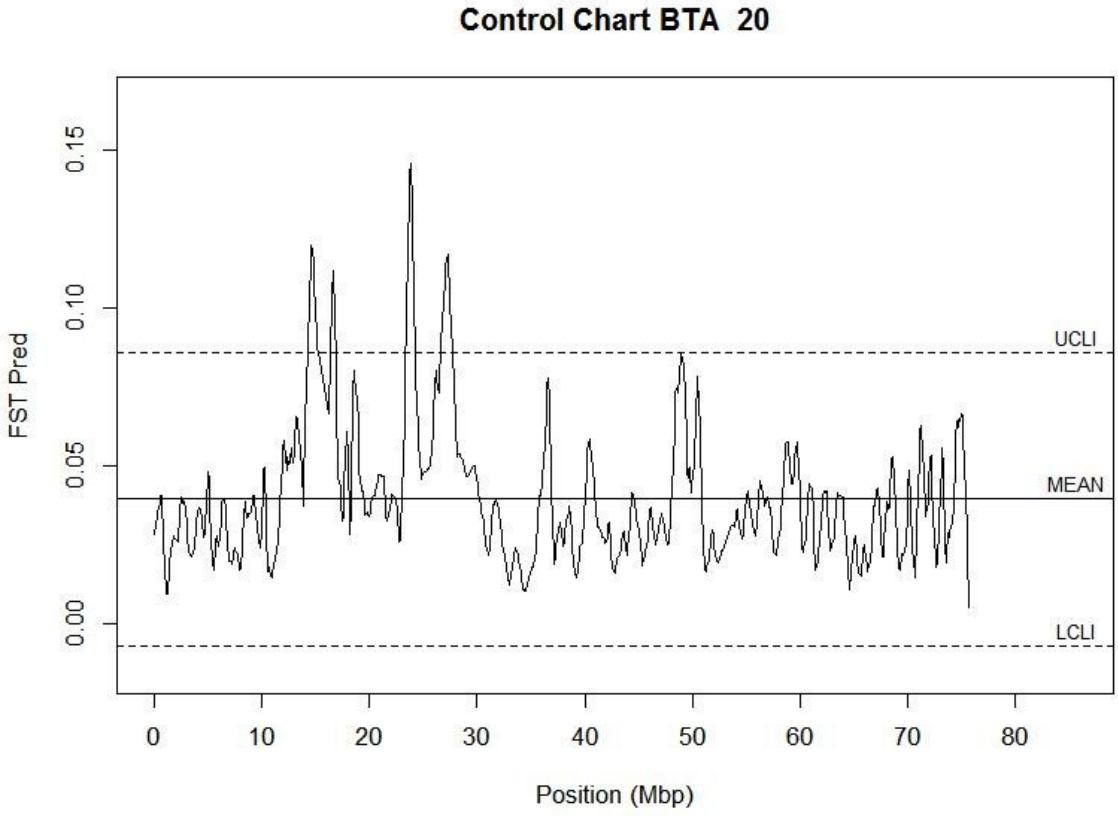

FST BTA 18
Raw FST
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
60
Position (Mbp)

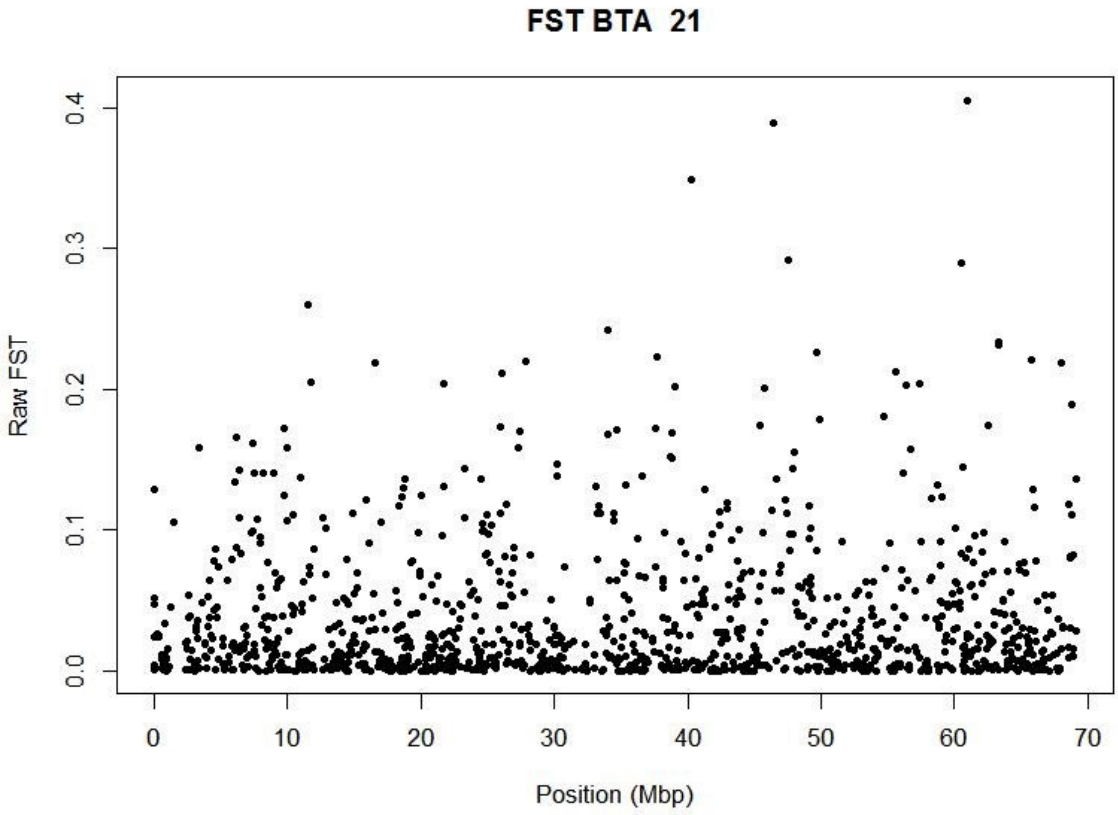

Control Chart BTA 18
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
Position (Mbp)

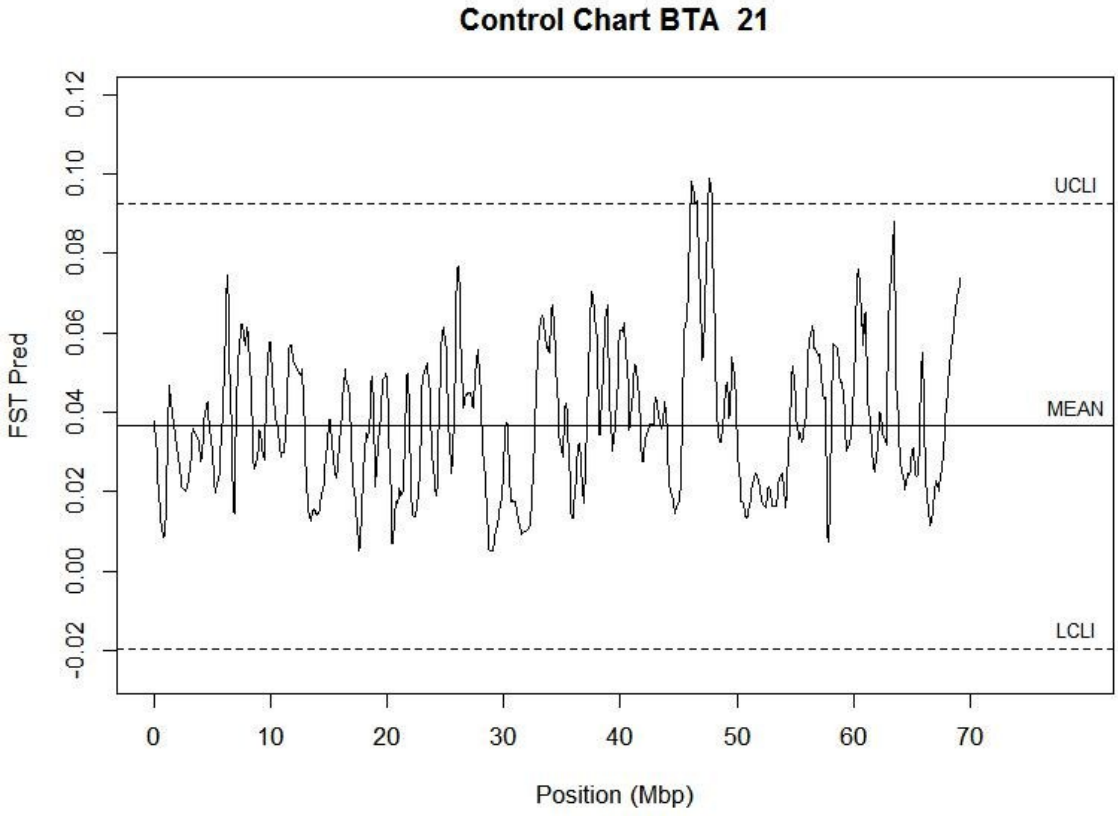

FST BTA 19
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
60
Position (Mbp)

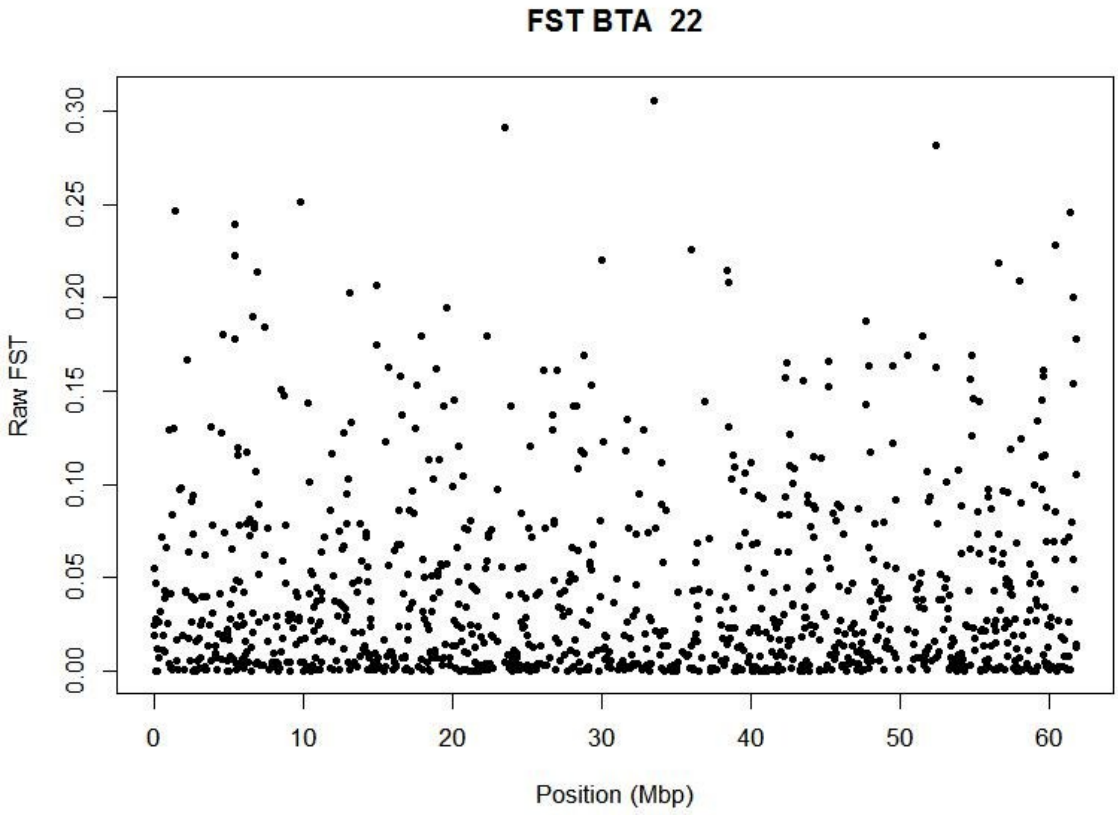

Control Chart BTA 19
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
Position (Mbp)

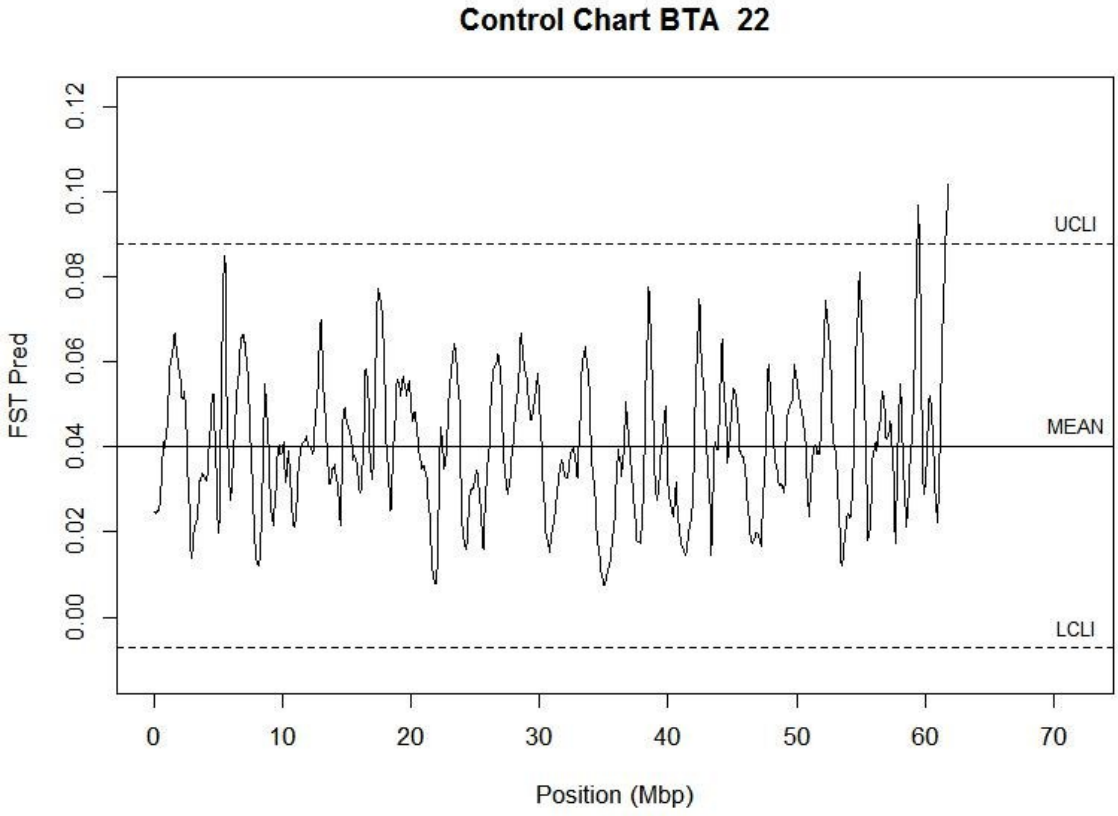

FST BTA 20
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
50
60
70
Position (Mbp)

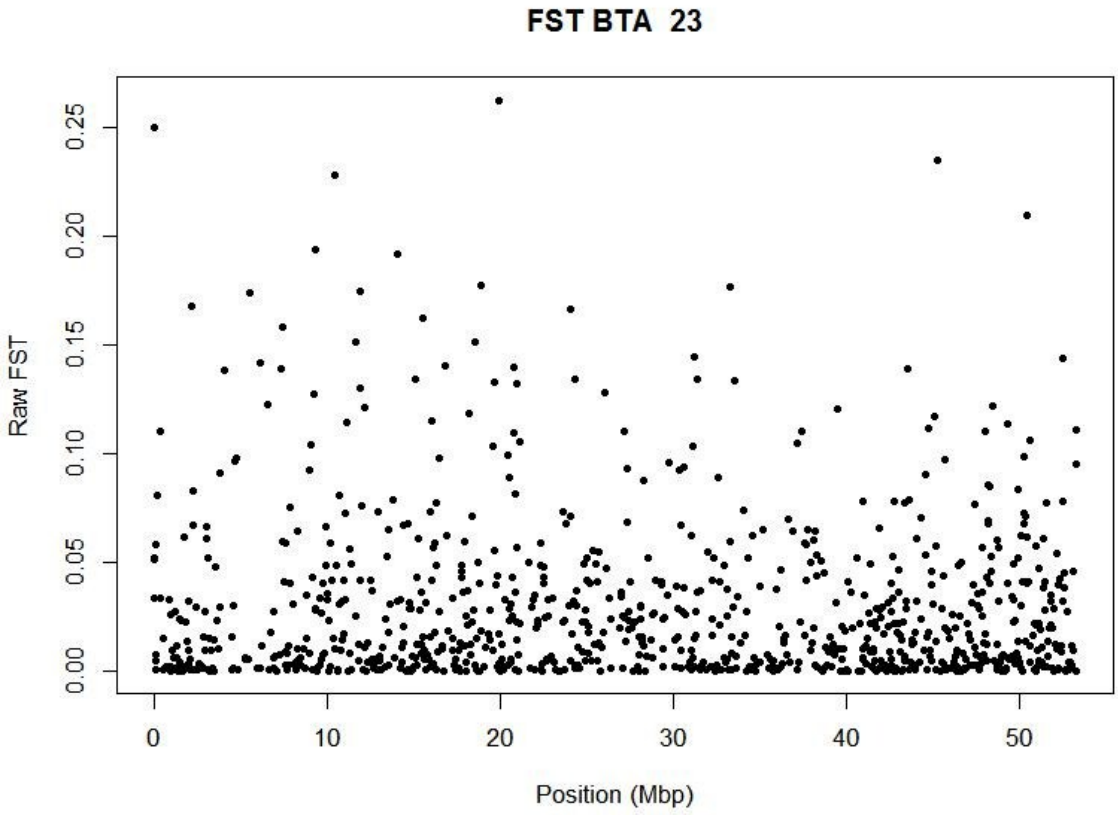

Control Chart BTA 20
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
80
Position (Mbp)

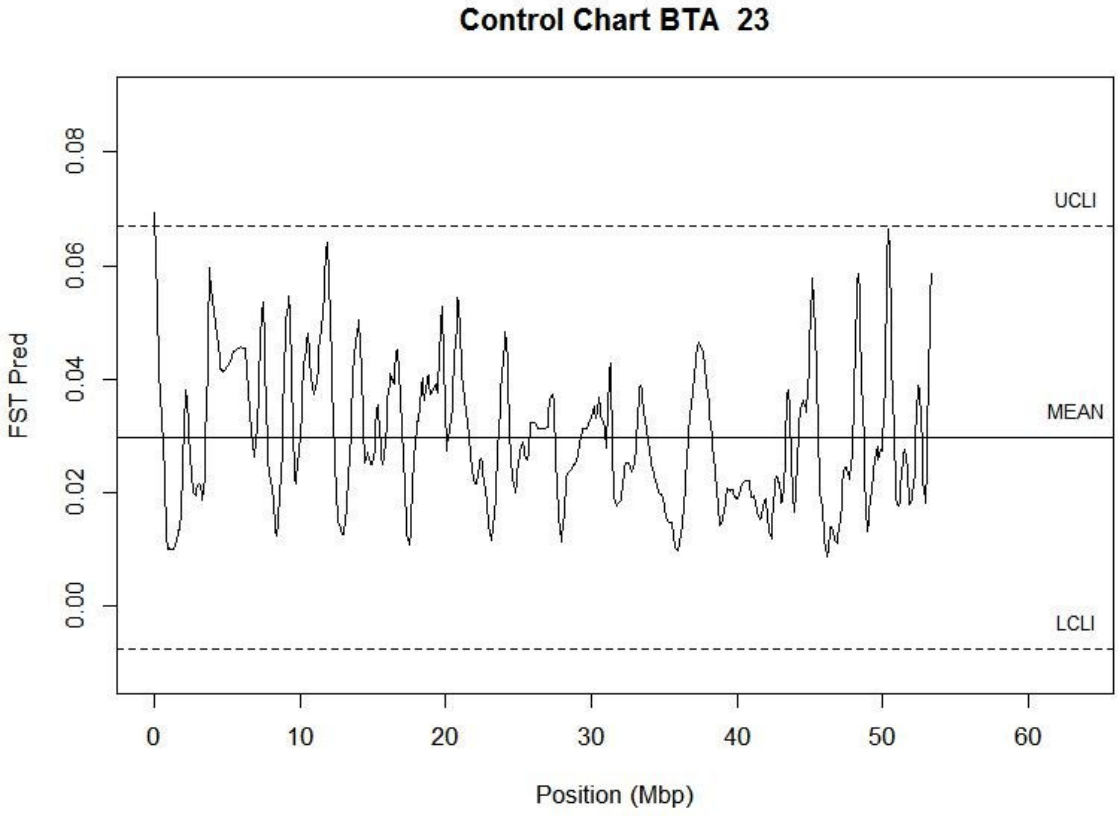

FST BTA 21
Raw FST
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
60
70
Position (Mbp)

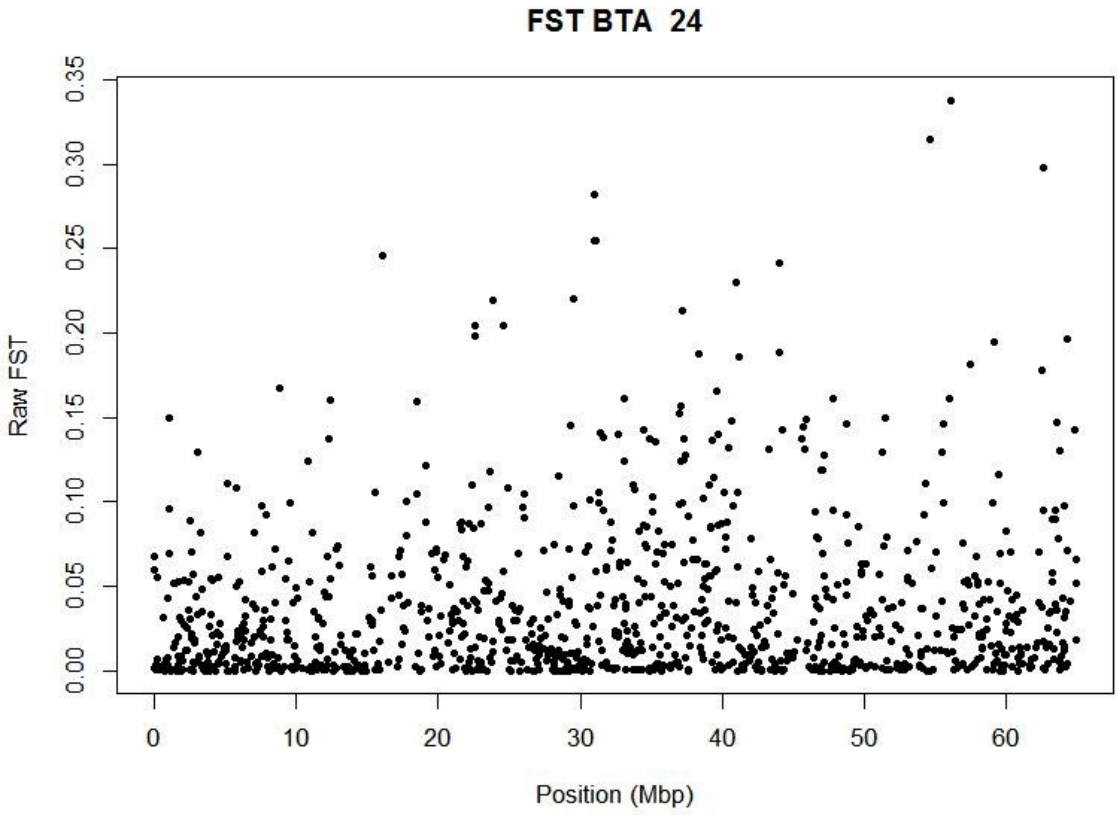

Control Chart BTA 21
FST Pred
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
0.12
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
Position (Mbp)

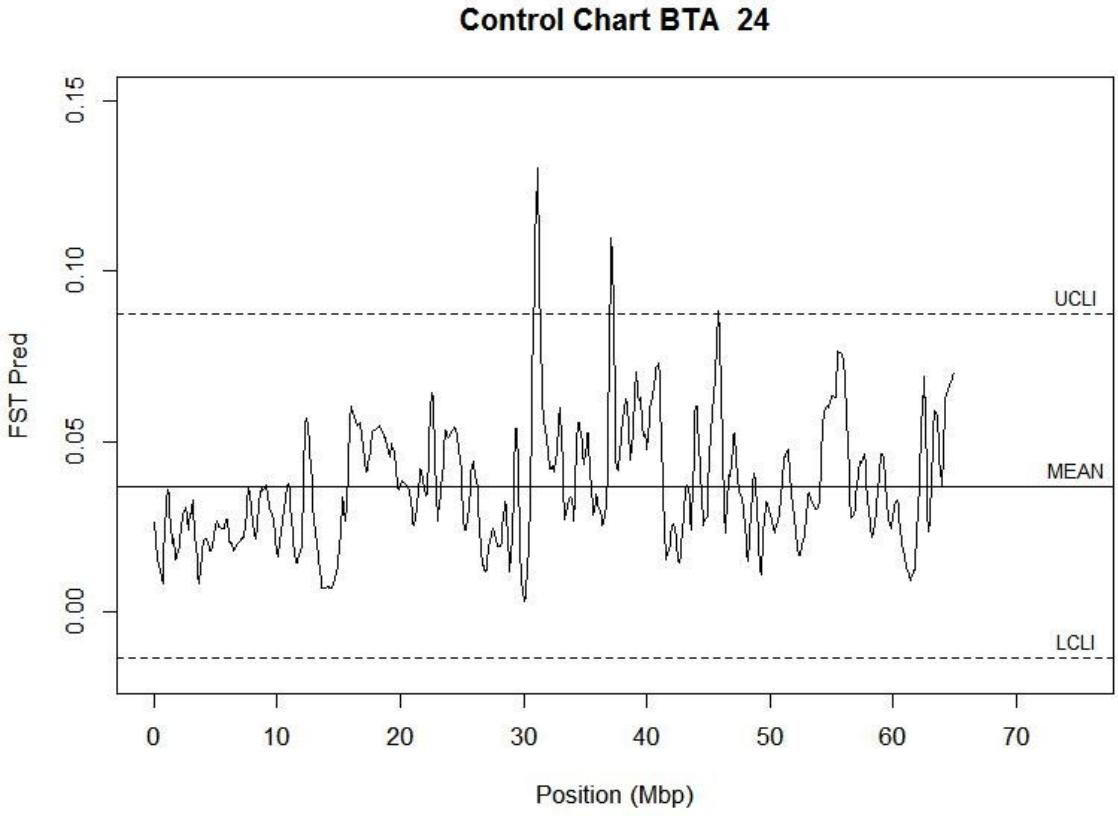

FST BTA 22
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
60
Position (Mbp)

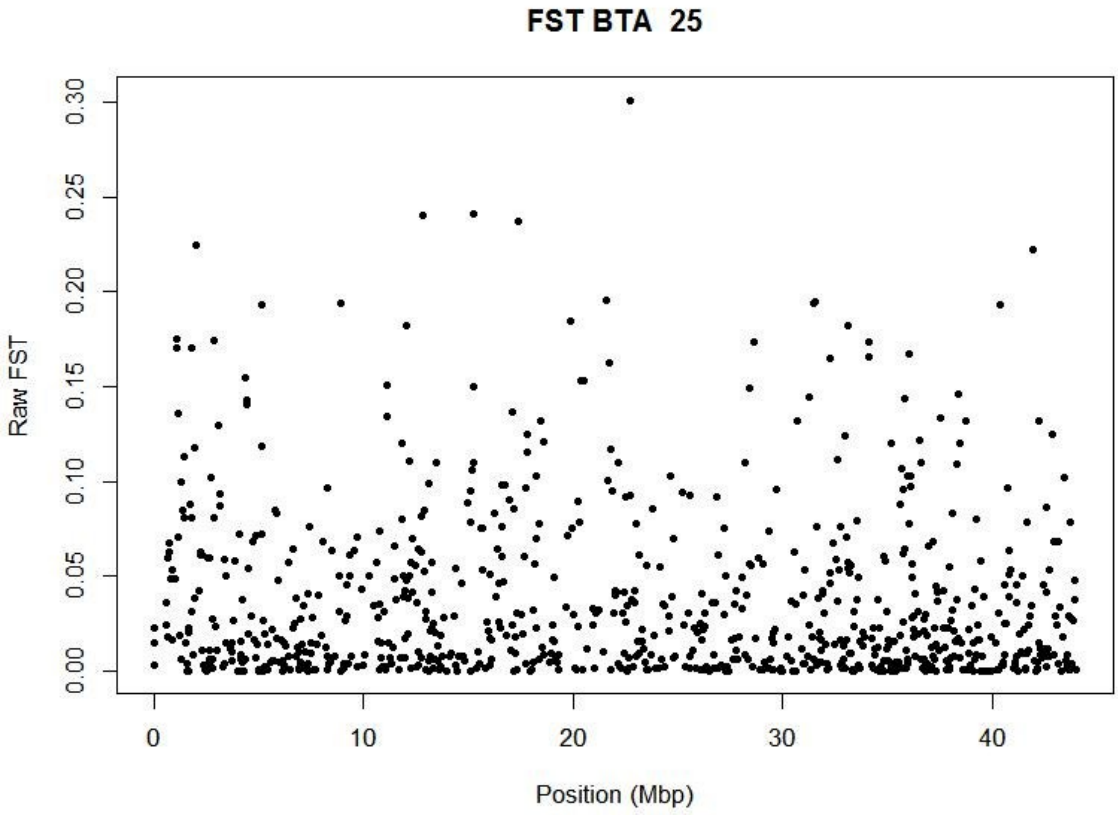

Control Chart BTA 22
FST Pred
0.00
0.02
0.04
0.06
0.08
0.10
0.12
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
Position (Mbp)

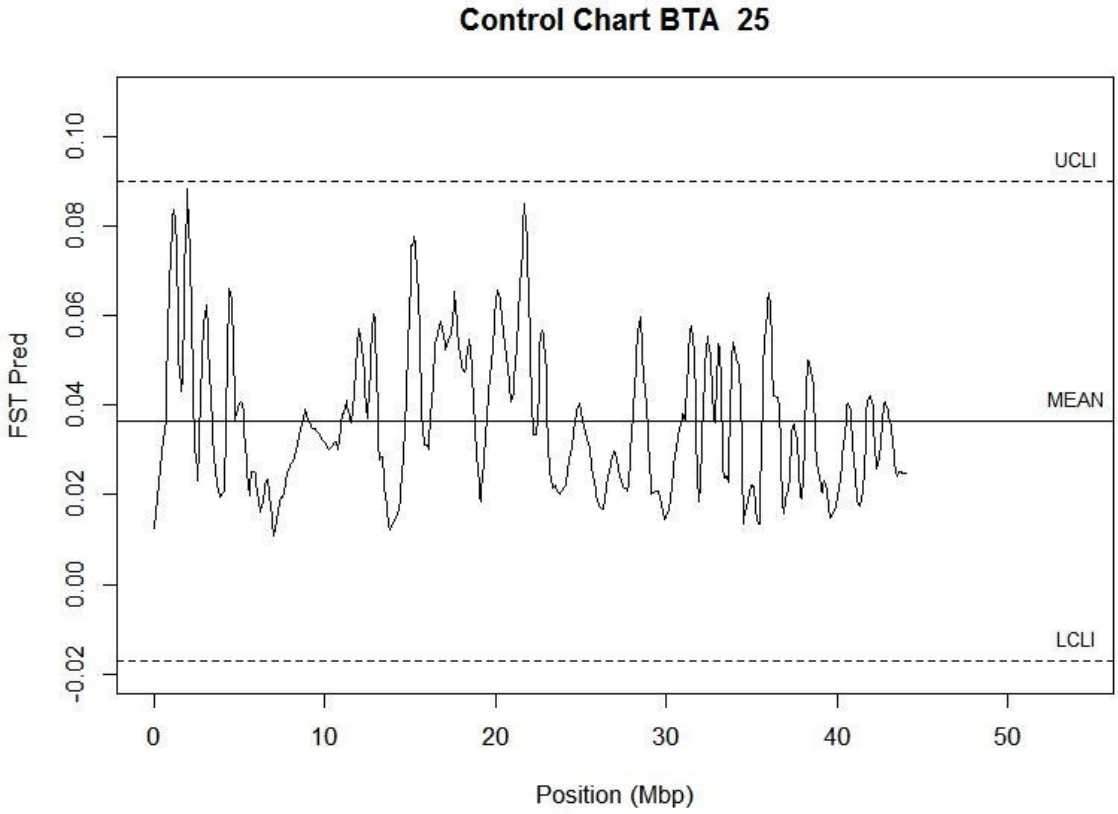

FST BTA 23
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
50
Position (Mbp)

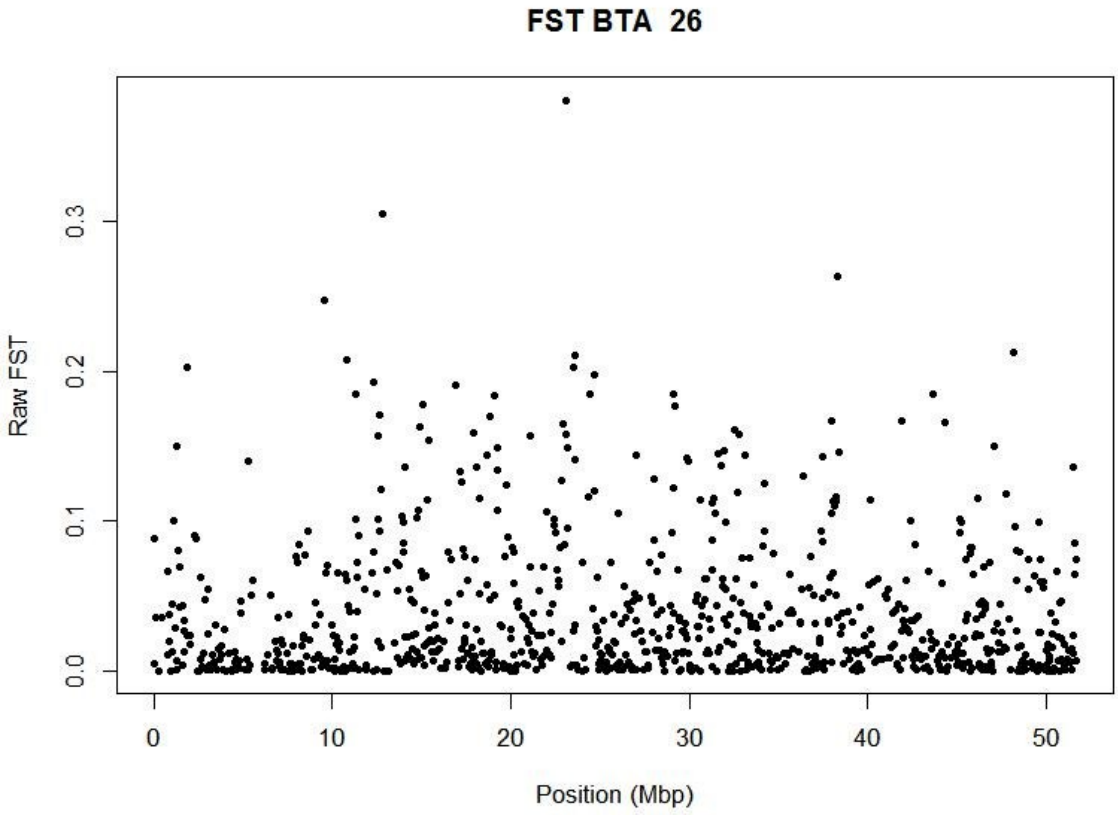

Control Chart BTA 23
FST Pred
0.00
0.02
0.04
0.06
0.08
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
Position (Mbp)

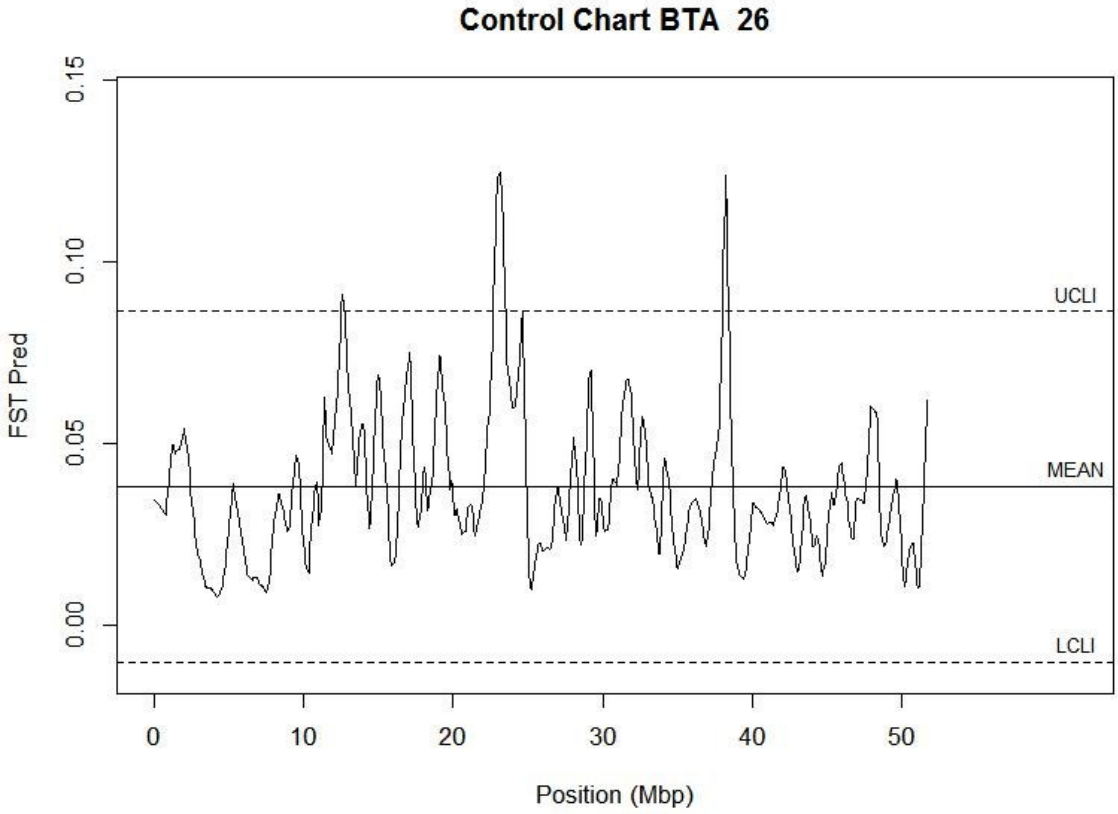

FST BTA 24
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
50
60
Position (Mbp)

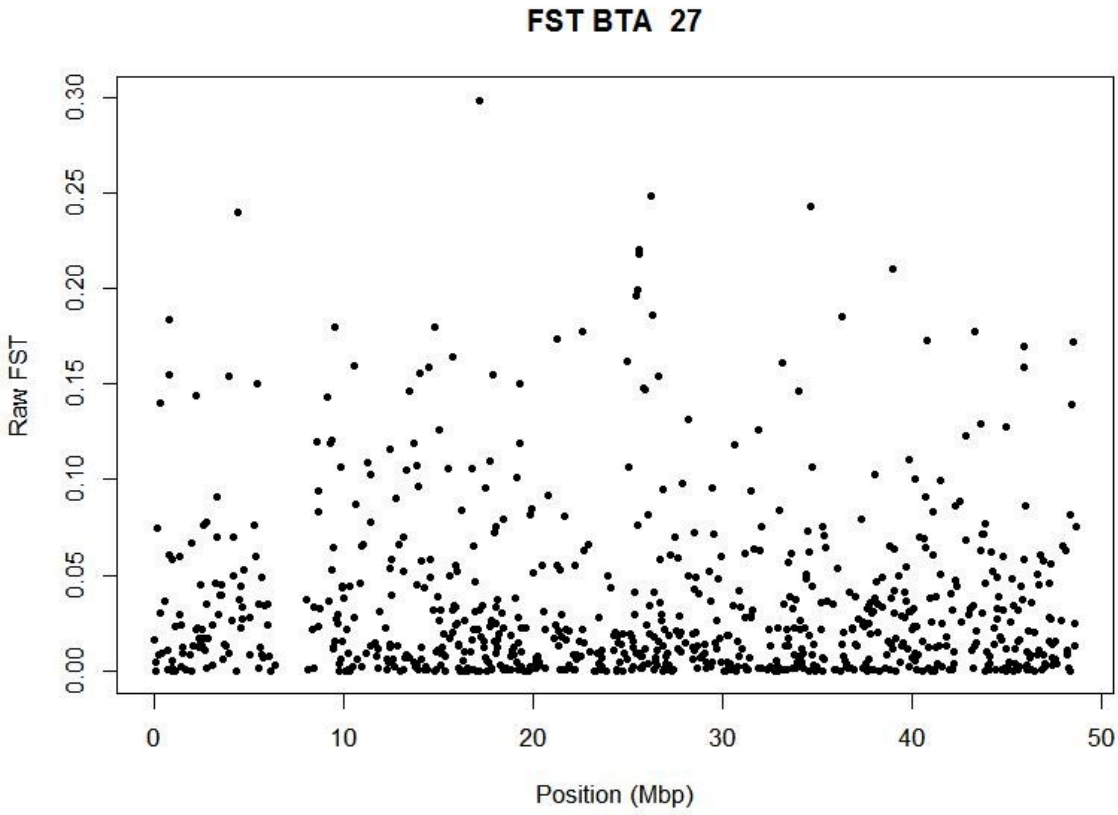

Control Chart BTA 24
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
70
Position (Mbp)

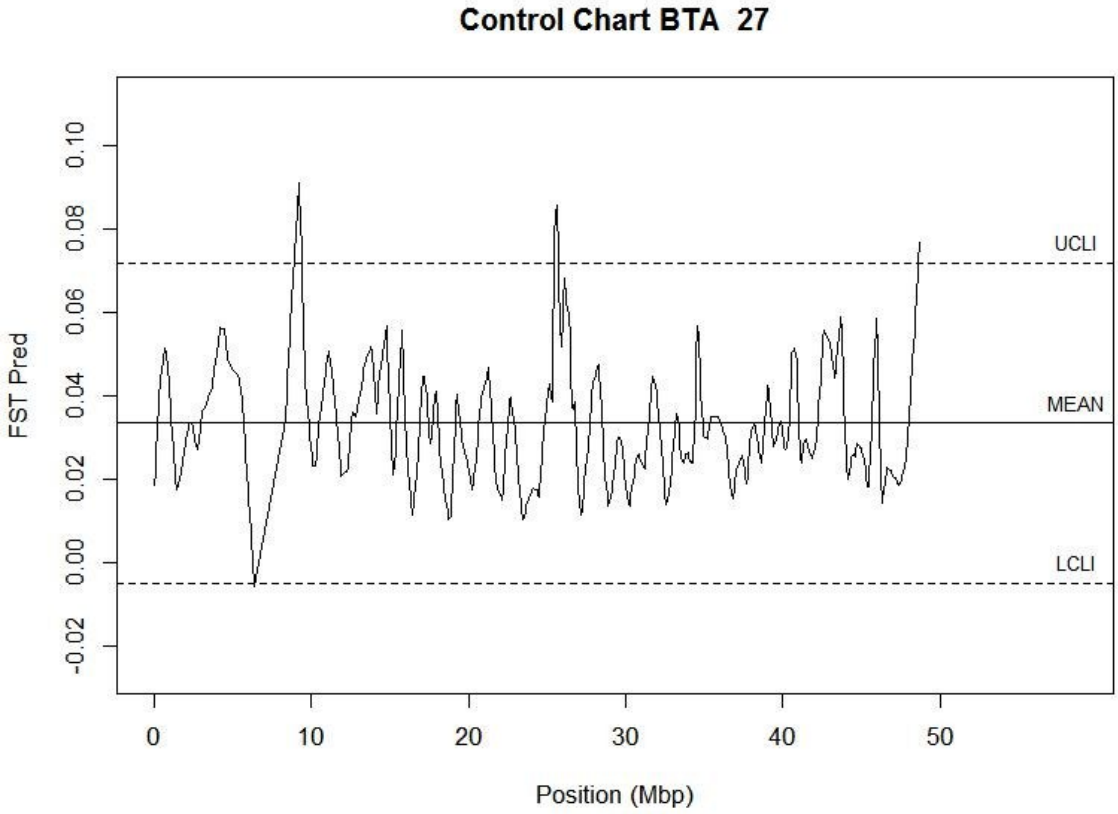

FST BTA 25
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
Position (Mbp)

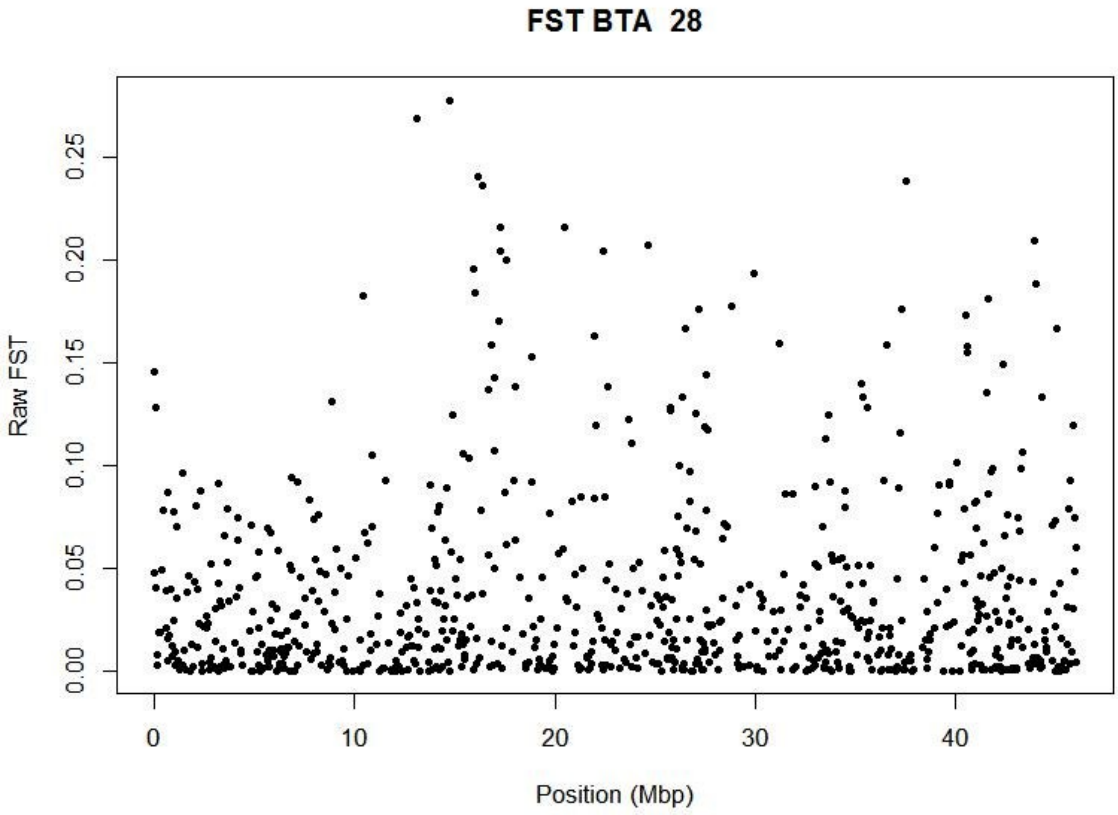

Control Chart BTA 25
FST Pred
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
UCLI
MEAN
LCLI
0
10
20
30
40
50
Position (Mbp)

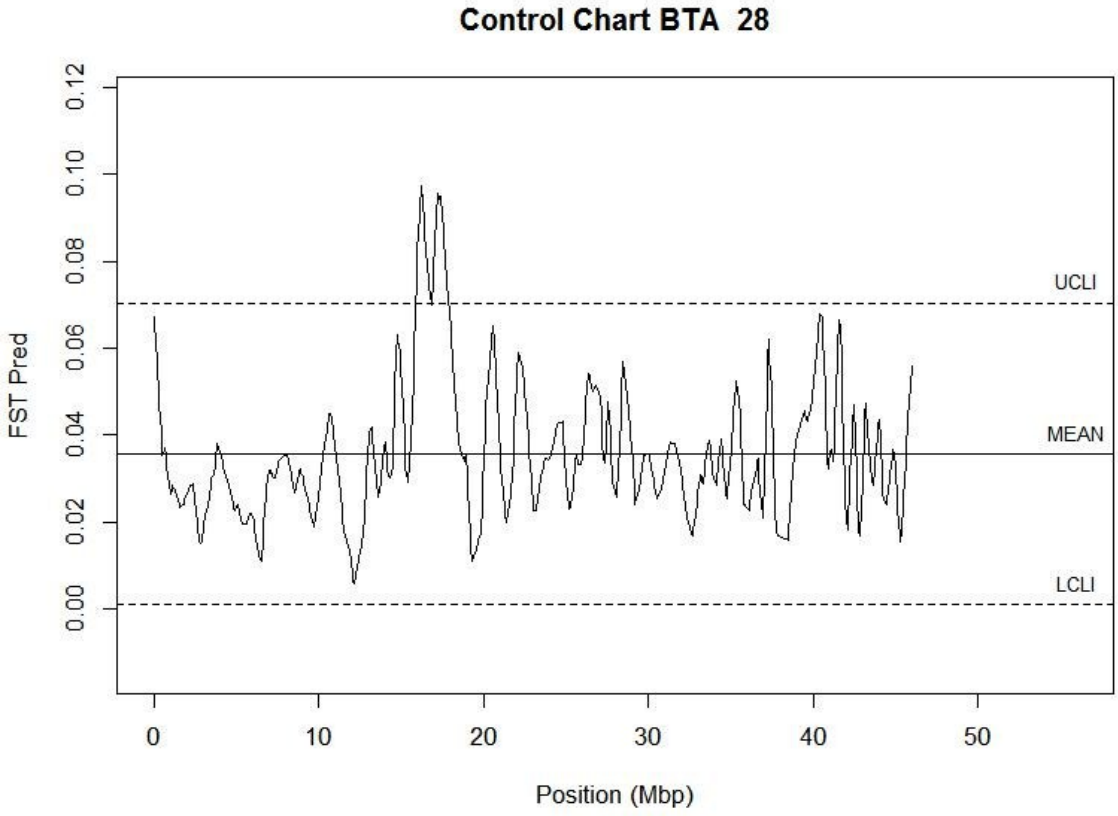

FST BTA 26
Raw FST
0.0
0.1
0.2
0.3
0
10
20
30
40
50
Position (Mbp)

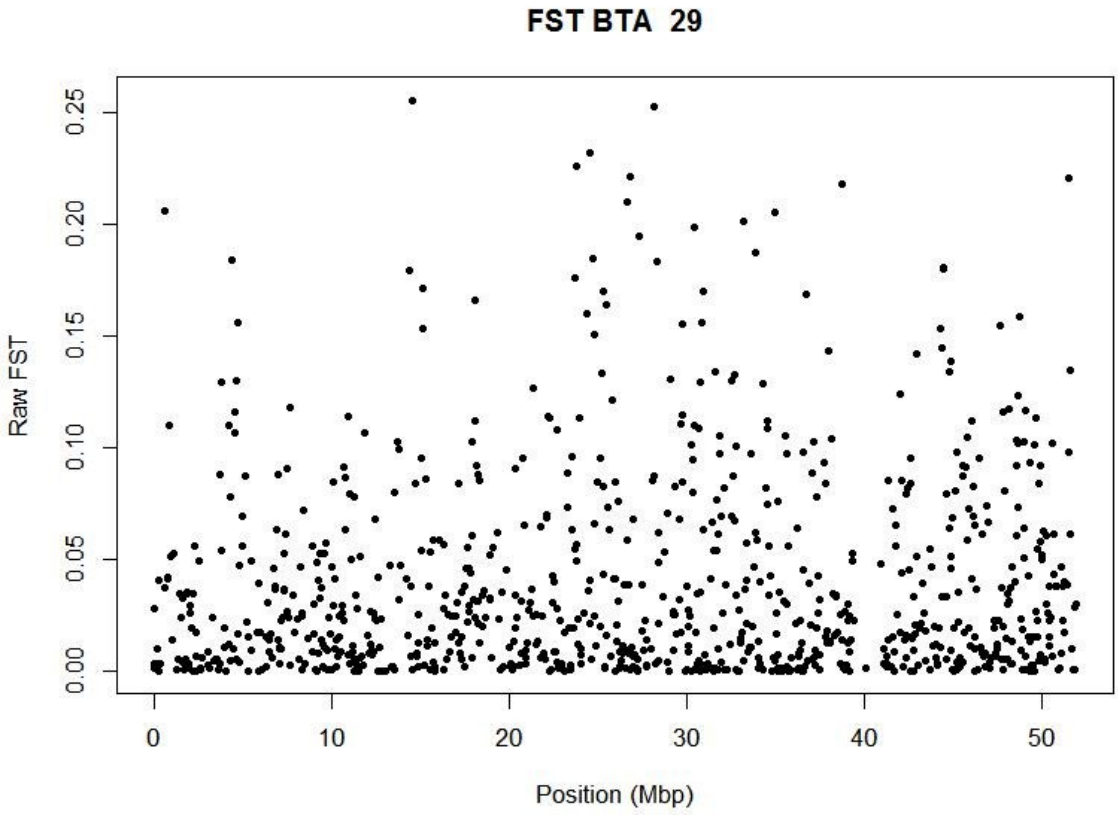

Control Chart BTA 26
FST Pred
0.00
0.05
0.10
0.15
UCLI
MEAN
LCLI
0
10
20
30
40
50
Position (Mbp)

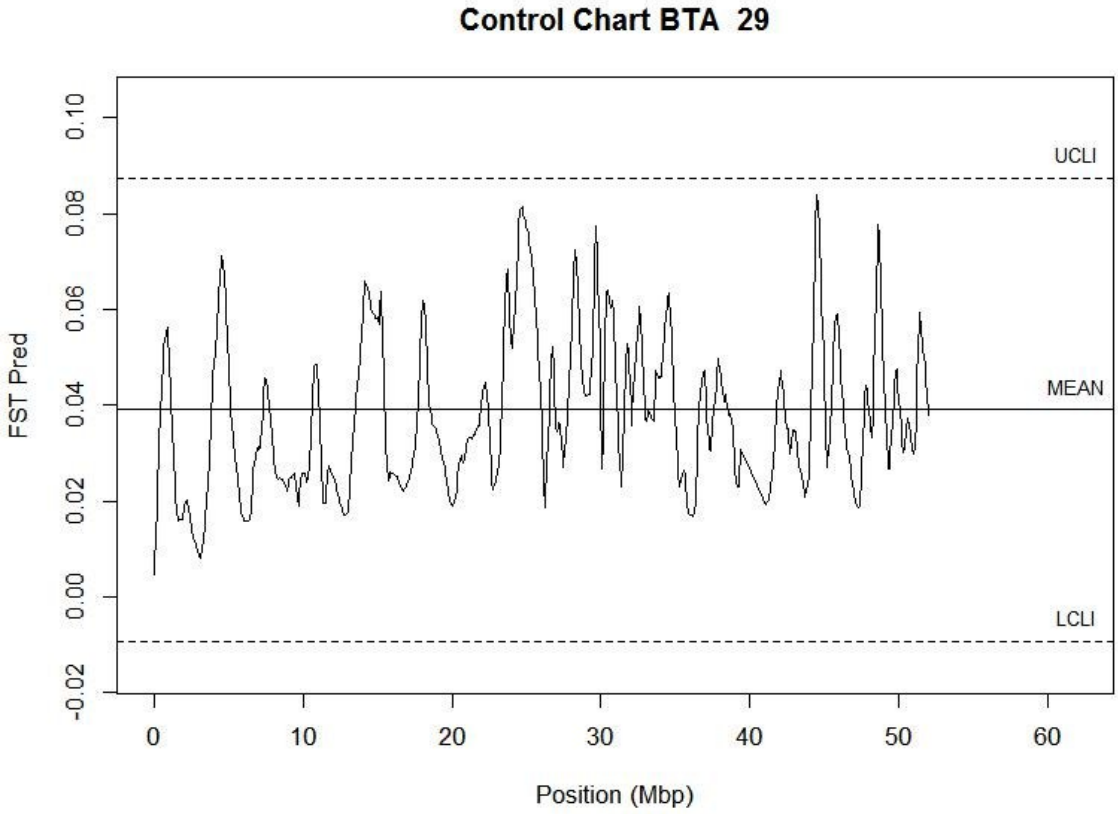

FST BTA 27
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
Position (Mbp)


Control Chart BTA 27
FST Pred
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
UCLI
MEAN
LCLI
0
10
20
30
40
50
Position (Mbp)


FST BTA 28
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
Position (Mbp)


Control Chart BTA 28
FST Pred
0.00
0.02
0.04
0.06
0.08
0.10
0.12
UCLI
MEAN
LCLI
0
10
20
30
40
50
Position (Mbp)

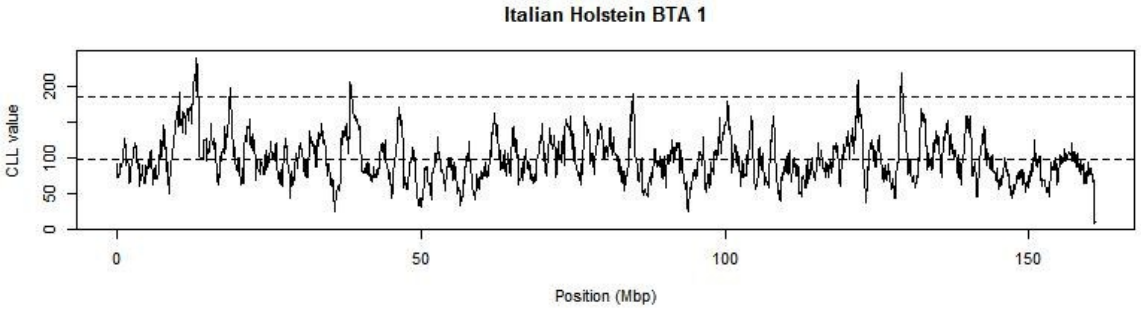

FST BTA 29
Raw FST
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
50
Position (Mbp)

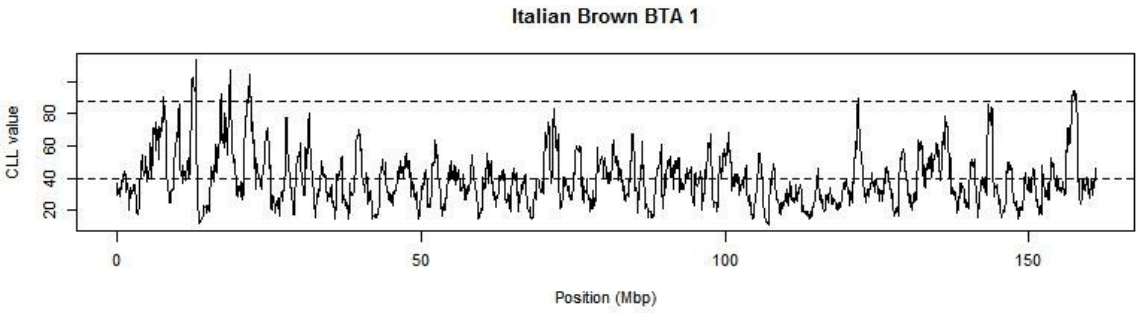

Control Chart BTA 29
FST Pred
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
UCLI
MEAN
LCLI
0
10
20
30
40
50
60
Position (Mbp)

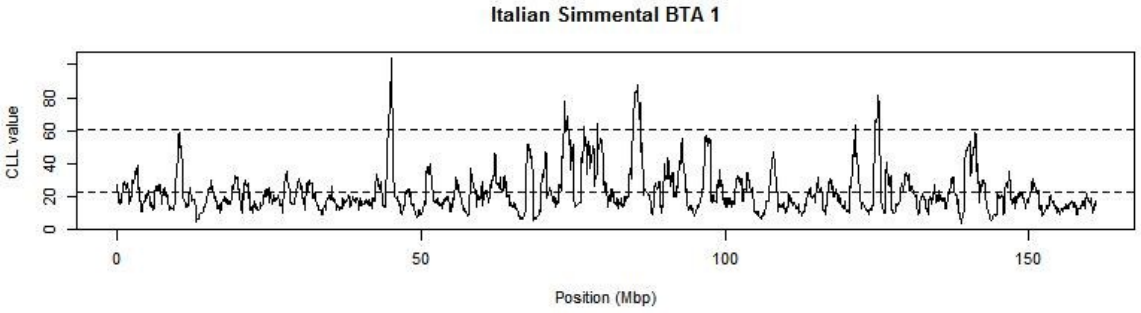
Supplemental Material

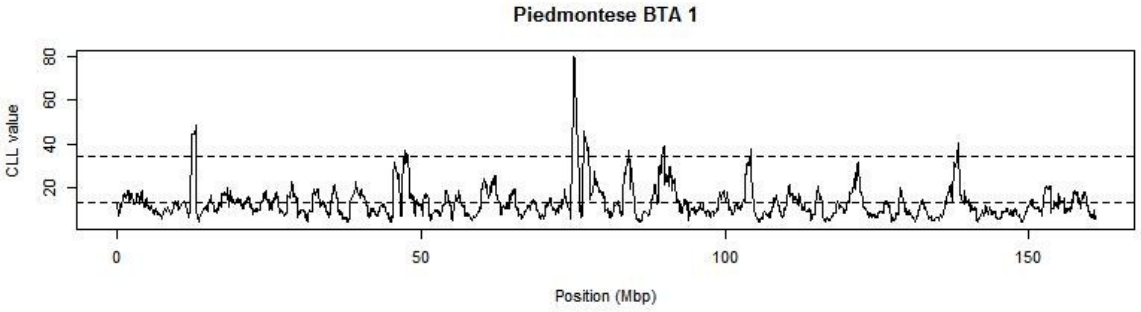
"Detection of selection signatures in five Italian cattle breeds with different productive specialization"

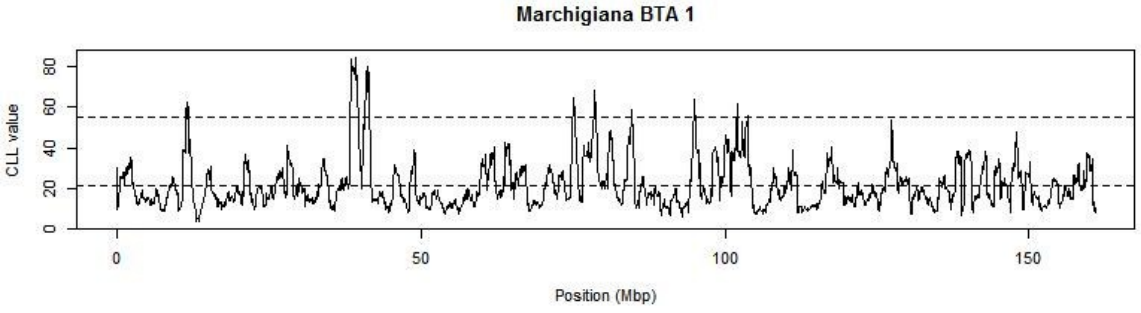
(**Chapter 3**)

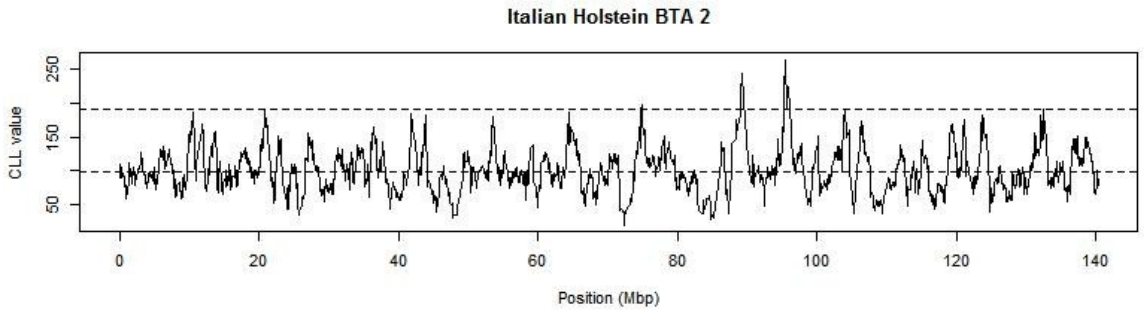

Italian Holstein BTA 1
CLL value
0
50
100
200
0
50
100
150
Position (Mbp)

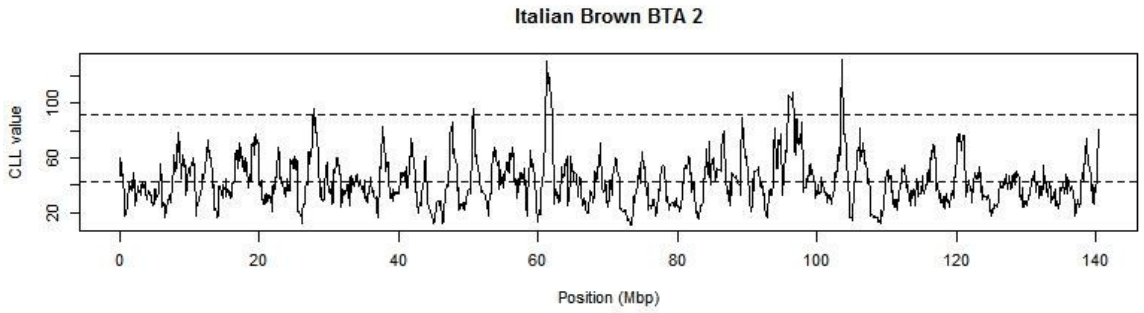

Italian Brown BTA 1
CLL value
20
40
60
80
0
50
100
150
Position (Mbp)

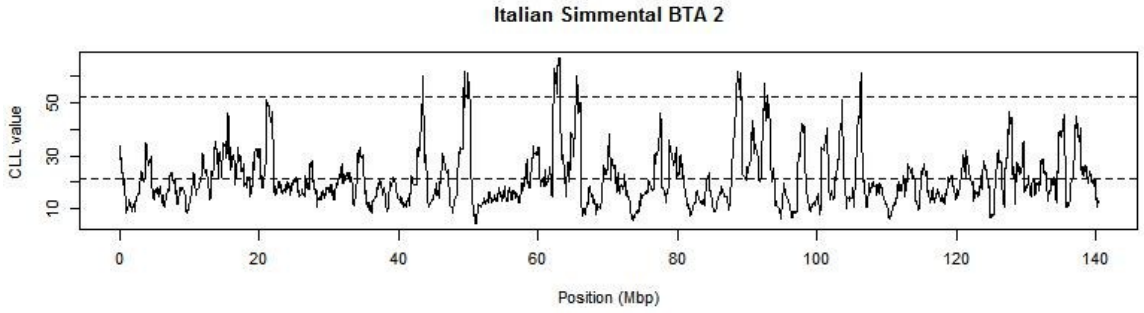

Italian Simmental BTA 1
CLL value
0
20
40
60
80
0
50
100
150
Position (Mbp)

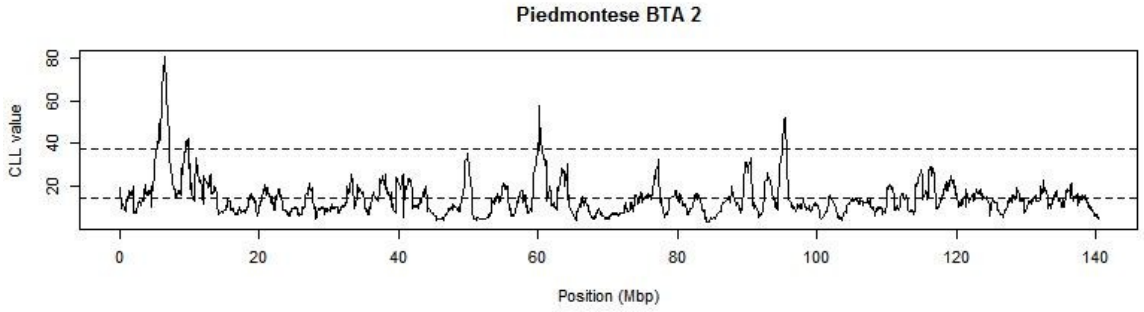

Piedmontese BTA 1
CLL value
20
40
60
80
0
50
100
150
Position (Mbp)

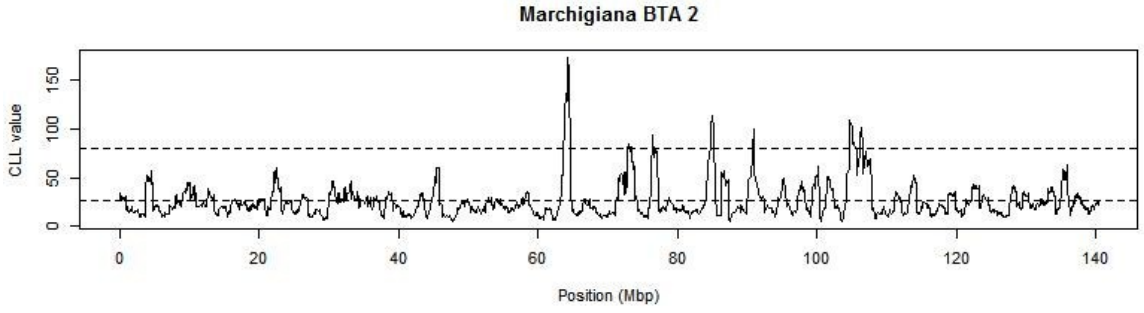

Marchigiana BTA 1
CLL value
0
20
40
60
80
0
50
100
150
Position (Mbp)

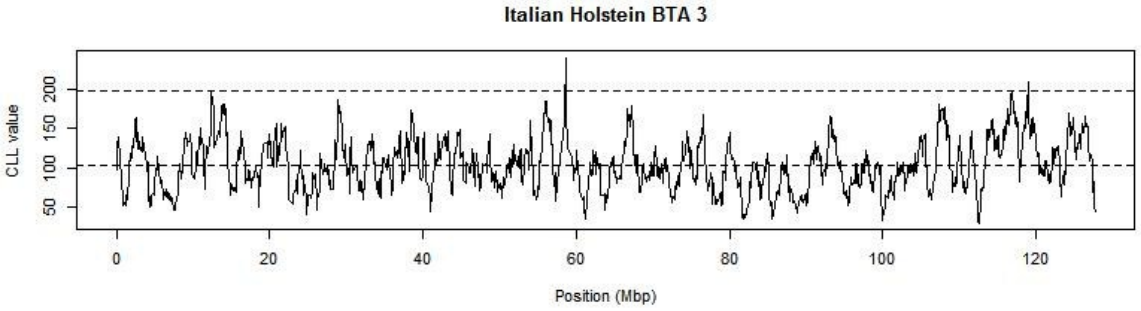

Italian Holstein BTA 2
CLL value
50
150
250
0
20
40
60
80
100
120
140
Position (Mbp)

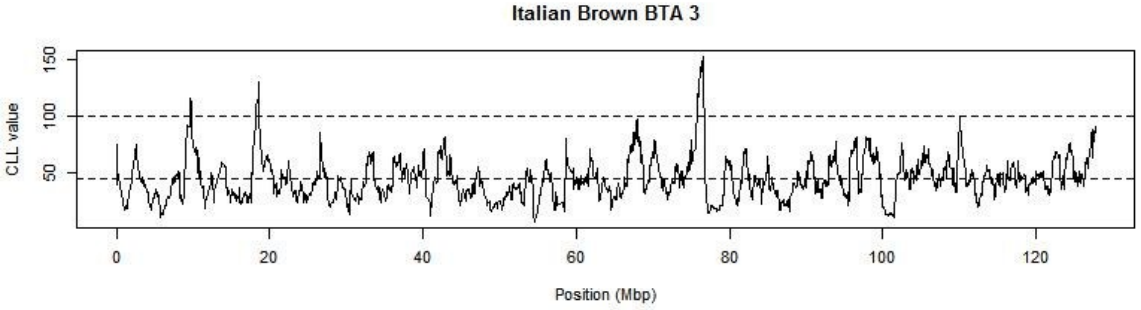

Italian Brown BTA 2
CLL value
20
60
100
0
20
40
60
80
100
120
140
Position (Mbp)

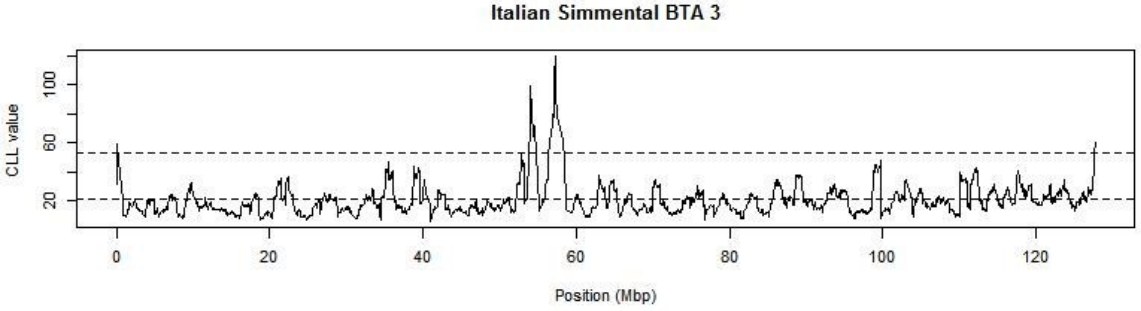

Italian Simmental BTA 2
CLL value
10
30
50
0
20
40
60
80
100
120
140
Position (Mbp)

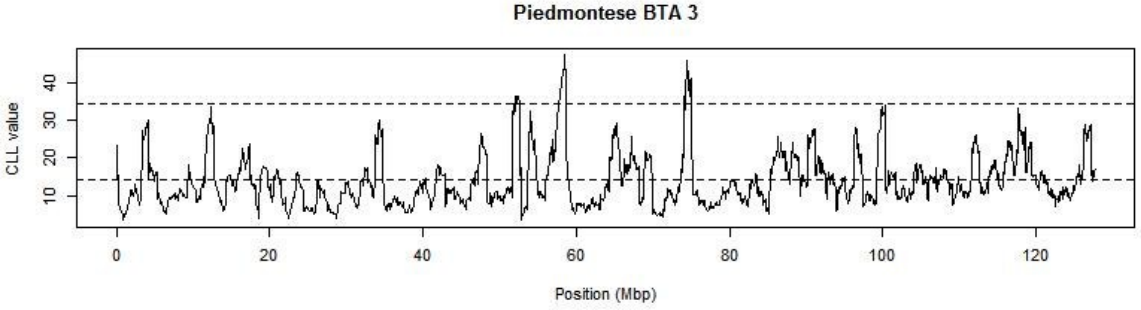

Piedmontese BTA 2
CLL value
20
40
60
80
0
20
40
60
80
100
120
140
Position (Mbp)

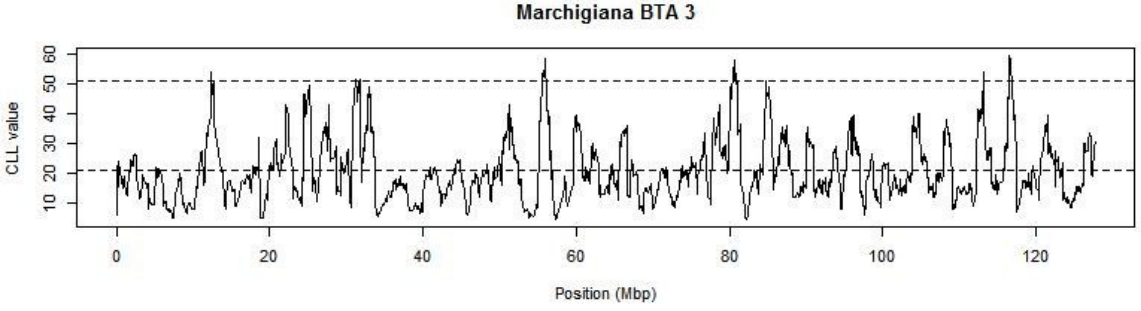

Marchigiana BTA 2
CLL value
0
50
100
150
0
20
40
60
80
100
120
140
Position (Mbp)

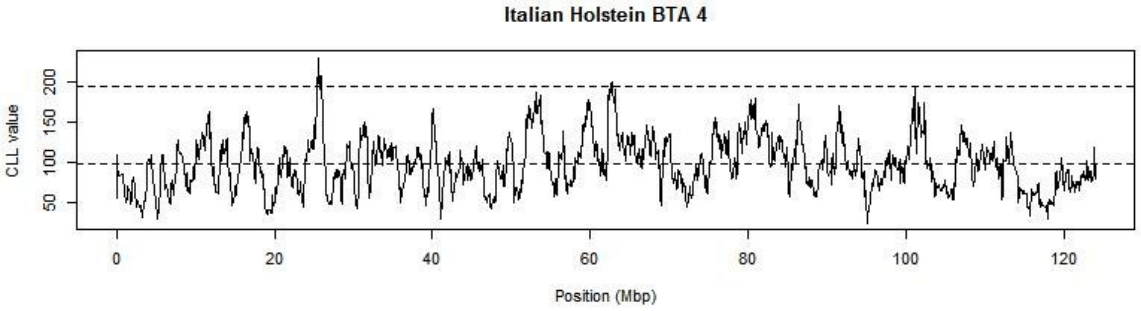

Italian Holstein BTA 3
CLL value
50
100
150
200
0
20
40
60
80
100
120
Position (Mbp)

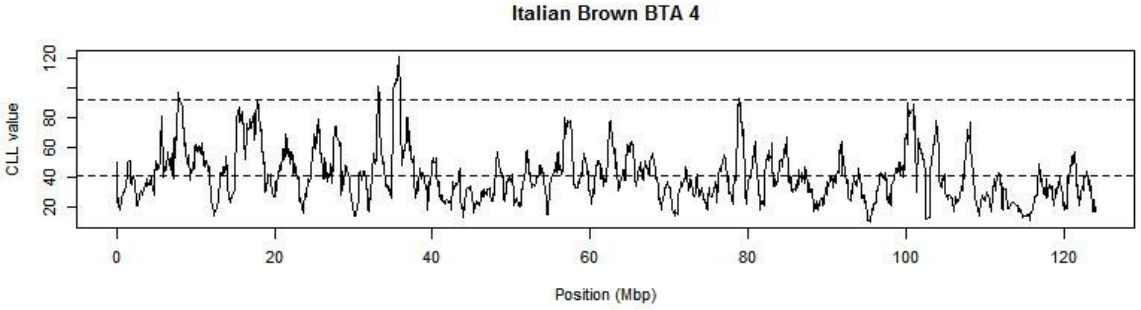

Italian Brown BTA 3
CLL value
50
100
150
0
20
40
60
80
100
120
Position (Mbp)

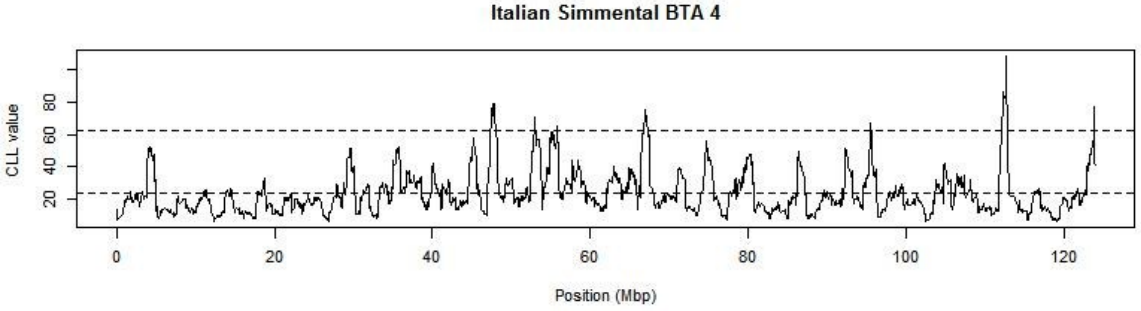

Italian Simmental BTA 3
CLL value
20
60
100
0
20
40
60
80
100
120
Position (Mbp)

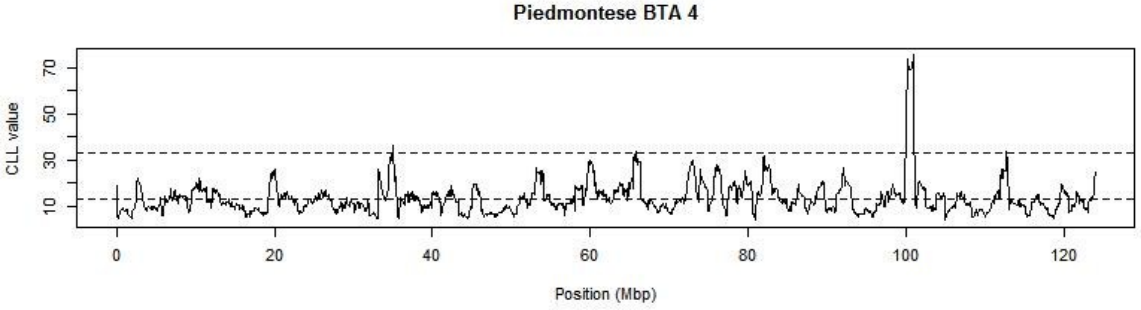

Piedmontese BTA 3
CLL value
10
20
30
40
0
20
40
60
80
100
120
Position (Mbp)

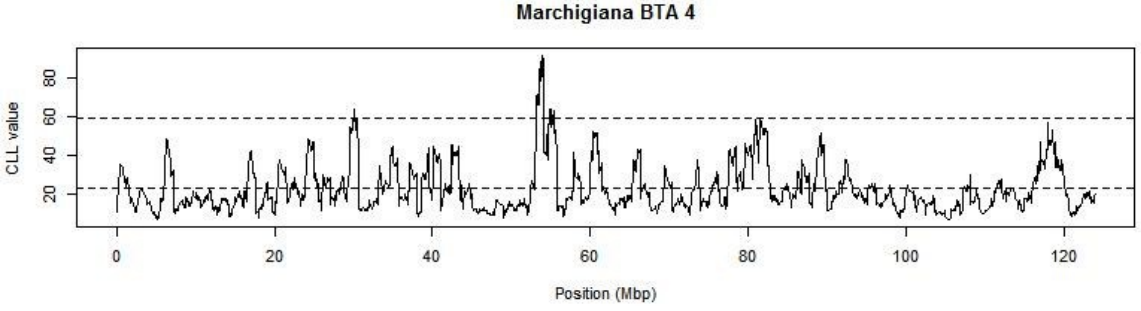

Marchigiana BTA 3
CLL value
10
20
30
40
50
60
0
20
40
60
80
100
120
Position (Mbp)

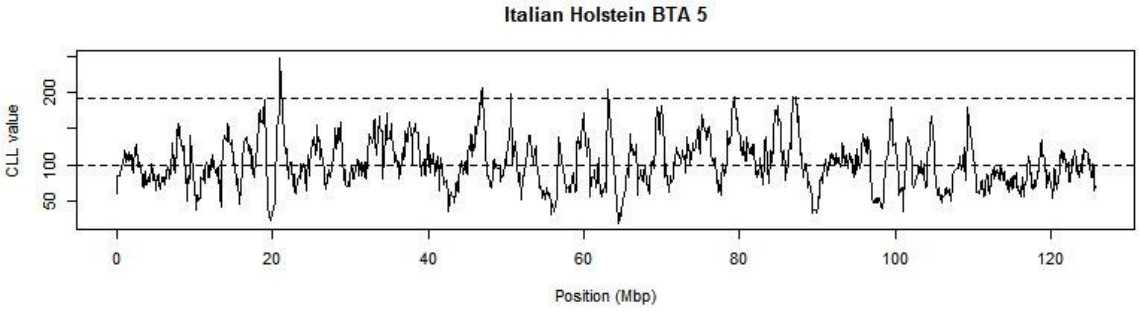

Italian Holstein BTA 4
CLL value
50
100
150
200
0
20
40
60
80
100
120
Position (Mbp)

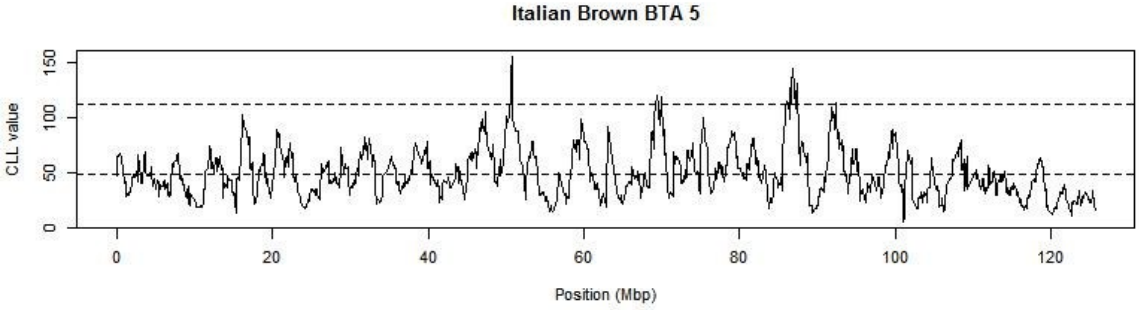

Italian Brown BTA 4
CLL value
20
40
60
80
120
0
20
40
60
80
100
120
Position (Mbp)

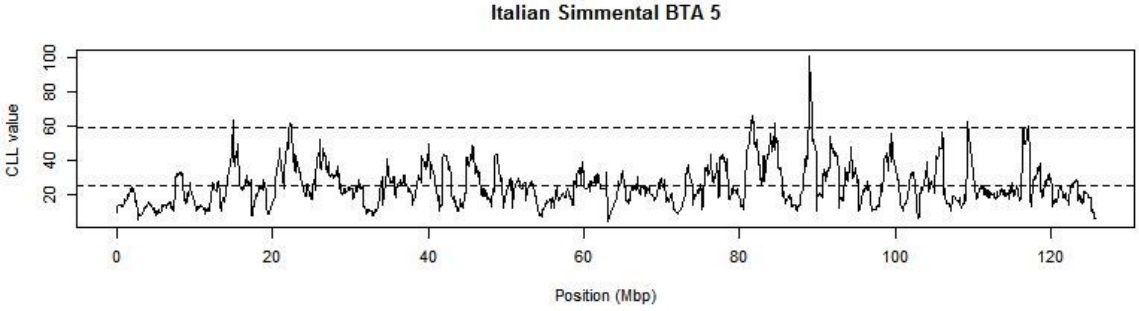

Italian Simmental BTA 4
CLL value
20
40
60
80
0
20
40
60
80
100
120
Position (Mbp)

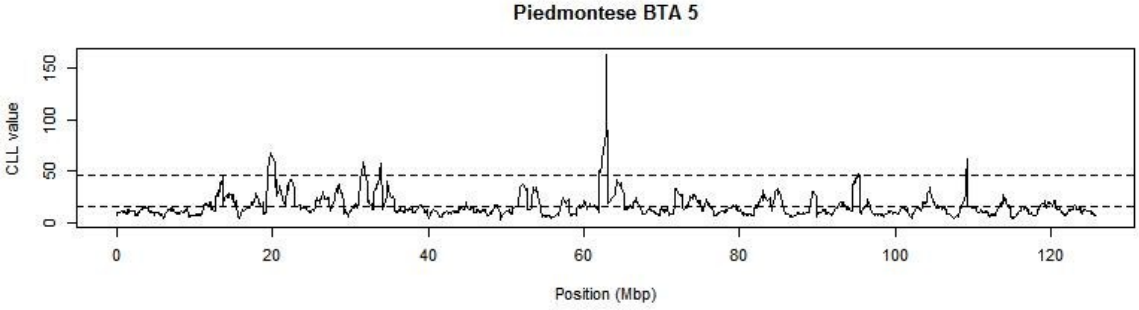

Piedmontese BTA 4
CLL value
10
30
50
70
0
20
40
60
80
100
120
Position (Mbp)

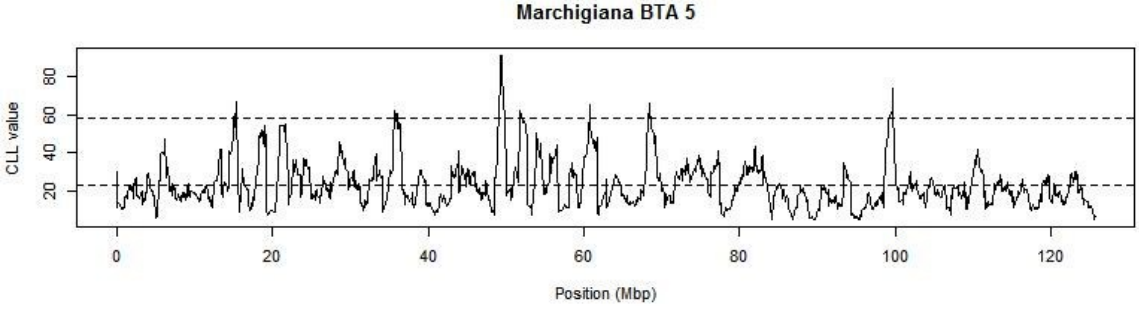

Marchigiana BTA 4
CLL value
20
40
60
80
0
20
40
60
80
100
120
Position (Mbp)

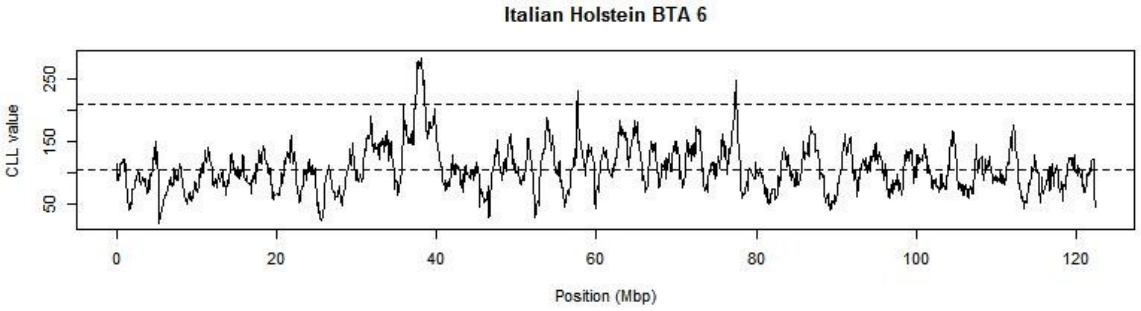

Italian Holstein BTA 5
CLL value
50
100
200
0
20
40
60
80
100
120
Position (Mbp)

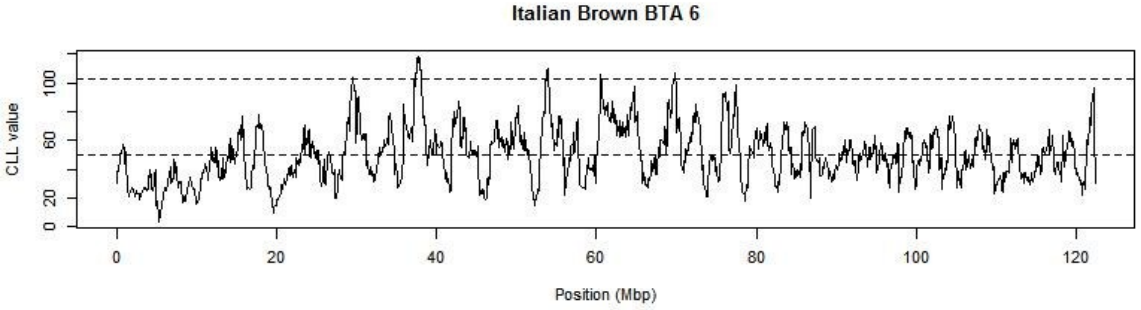

Italian Brown BTA 5
CLL value
0
50
100
150
0
20
40
60
80
100
120
Position (Mbp)

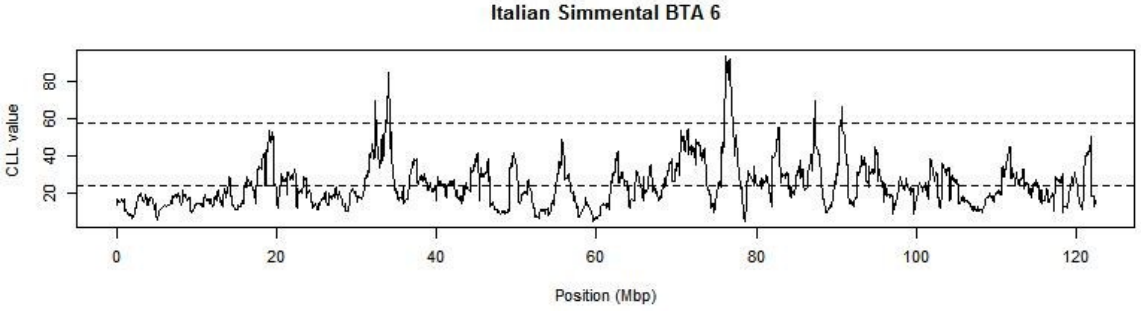

Italian Simmental BTA 5
CLL value
20
40
60
80
100
0
20
40
60
80
100
120
Position (Mbp)

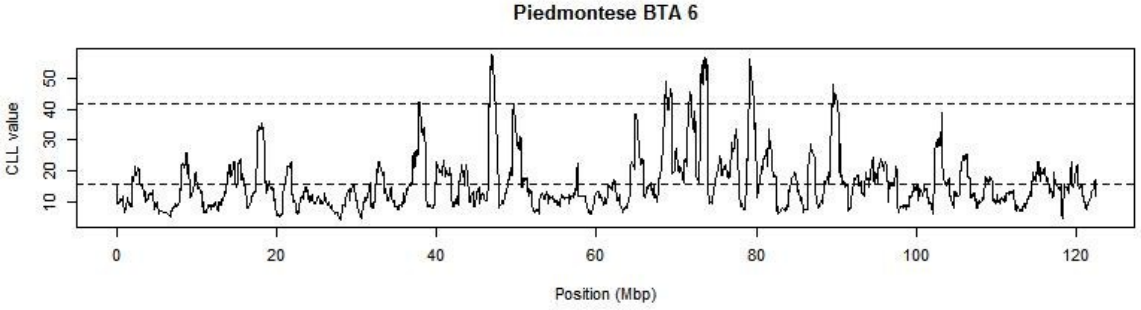

Piedmontese BTA 5
CLL value
0
50
100
150
0
20
40
60
80
100
120
Position (Mbp)

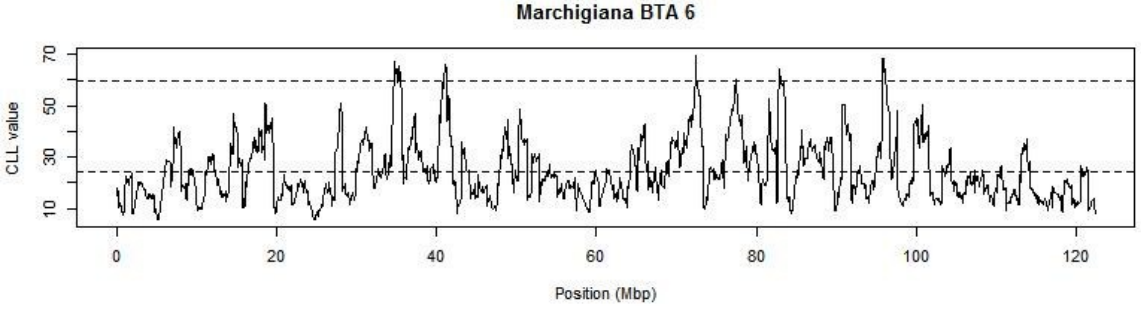

Marchigiana BTA 5
CLL value
20
40
60
80
0
20
40
60
80
100
120
Position (Mbp)

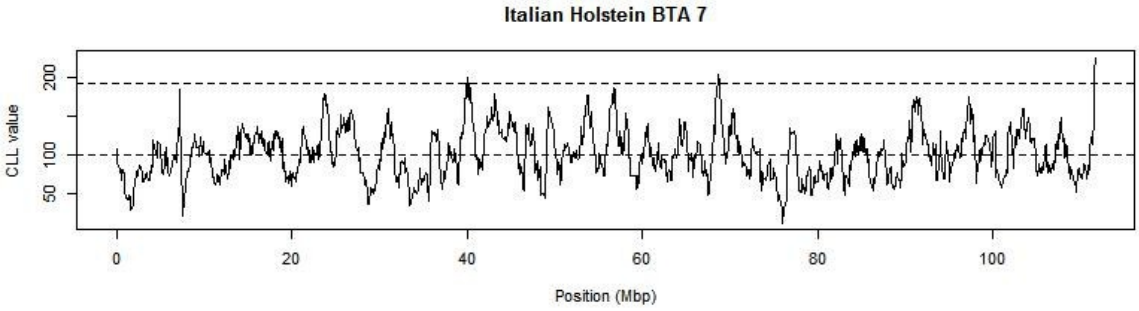

Italian Holstein BTA 6
CLL value
50
150
250
0
20
40
60
80
100
120
Position (Mbp)

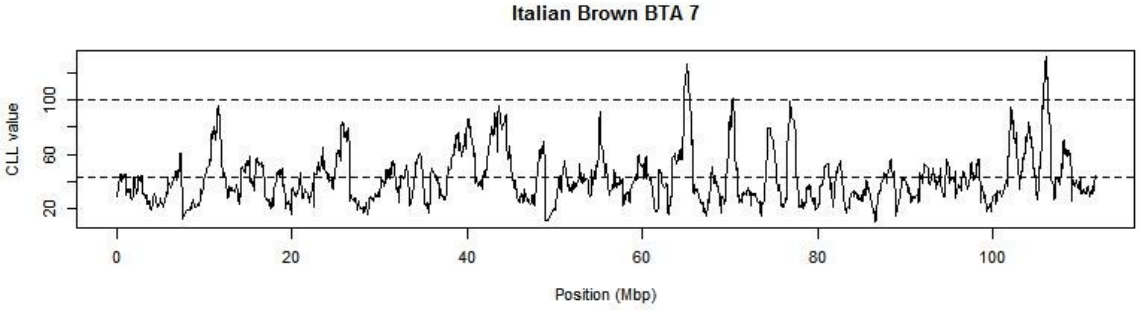

Italian Brown BTA 6
CLL value
0
20
60
100
0
20
40
60
80
100
120
Position (Mbp)

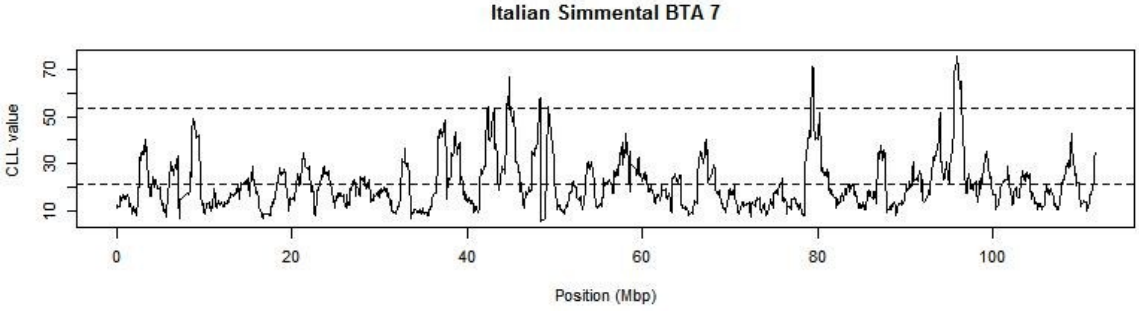

Italian Simmental BTA 6
CLL value
20
40
60
80
0
20
40
60
80
100
120
Position (Mbp)

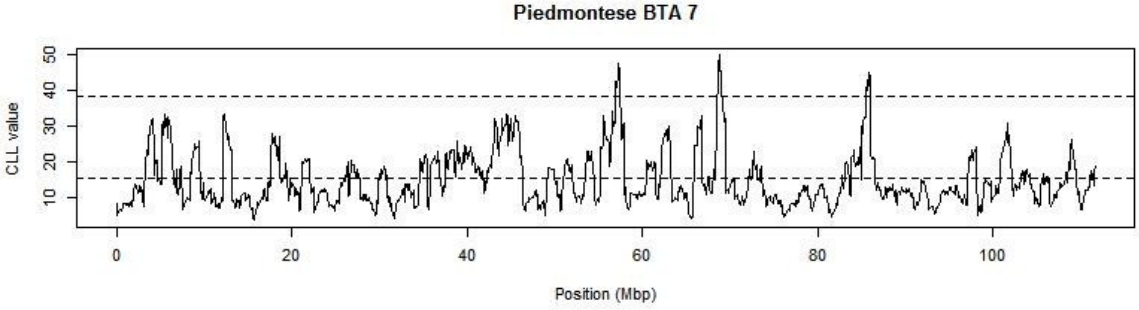

Piedmontese BTA 6
CLL value
10
20
30
40
50
0
20
40
60
80
100
120
Position (Mbp)

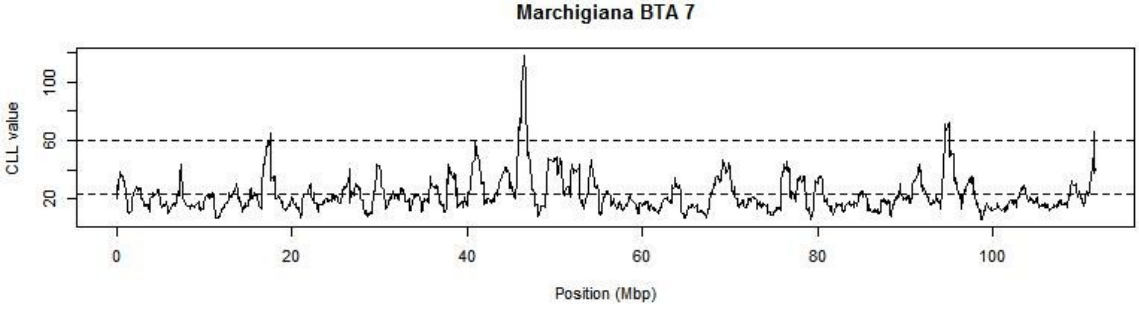

Marchigiana BTA 6
CLL value
10
30
50
70
0
20
40
60
80
100
120
Position (Mbp)

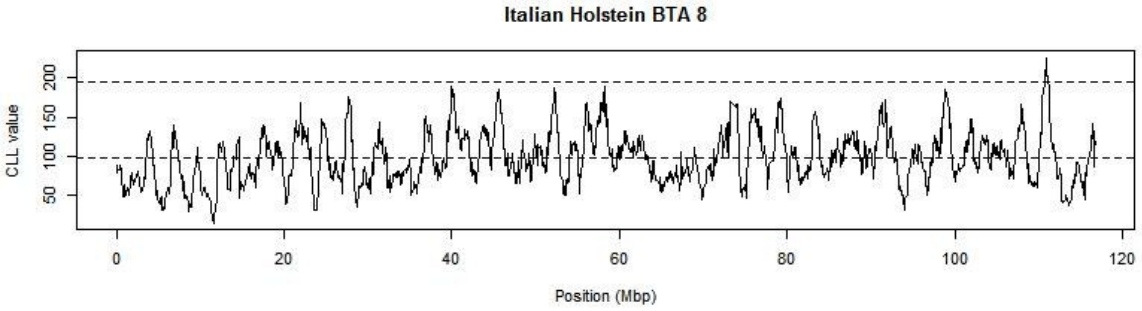

Italian Holstein BTA 7
CLL value
50
100
200
0
20
40
60
80
100
Position (Mbp)

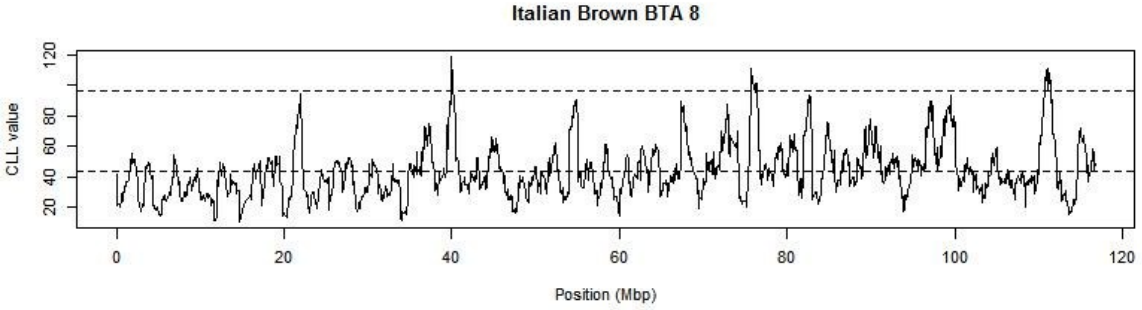

Italian Brown BTA 7
CLL value
20
60
100
0
20
40
60
80
100
Position (Mbp)

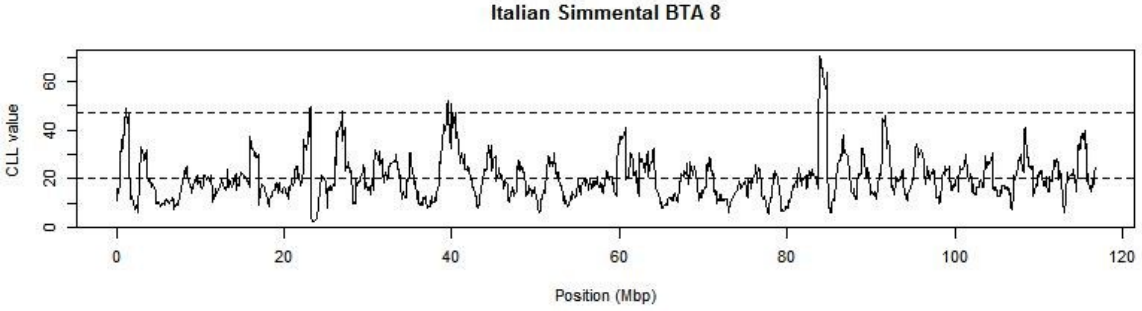

Italian Simmental BTA 7
CLL value
10
30
50
70
0
20
40
60
80
100
Position (Mbp)

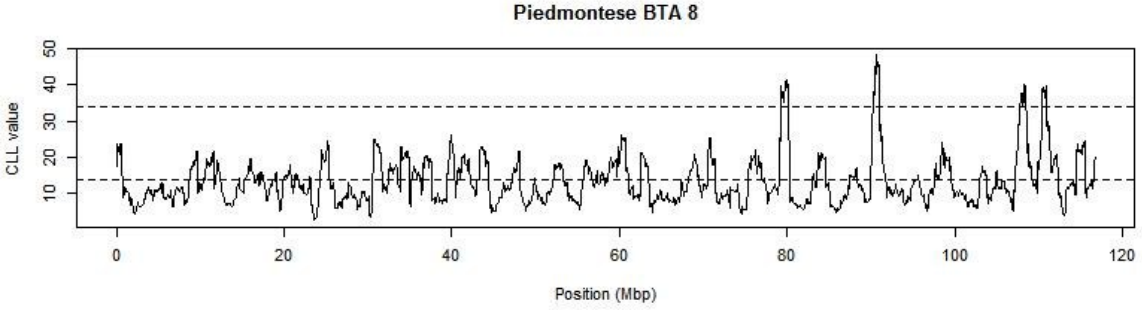

Piedmontese BTA 7
CLL value
10
20
30
40
50
0
20
40
60
80
100
Position (Mbp)

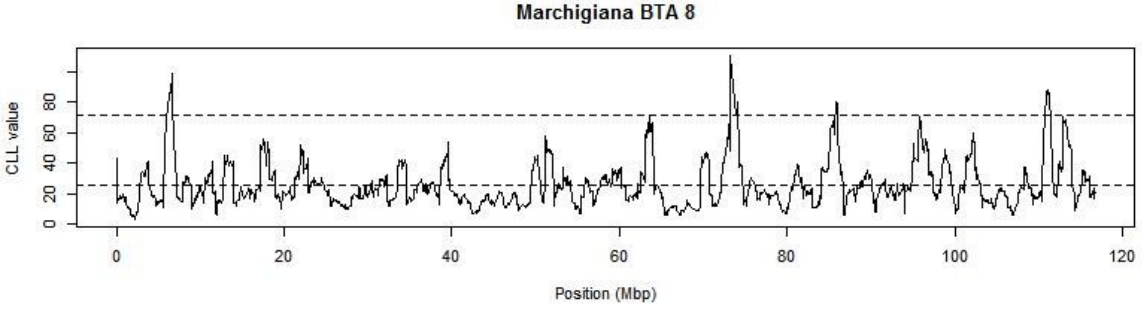

Marchigiana BTA 7
CLL value
20
60
100
0
20
40
60
80
100
Position (Mbp)

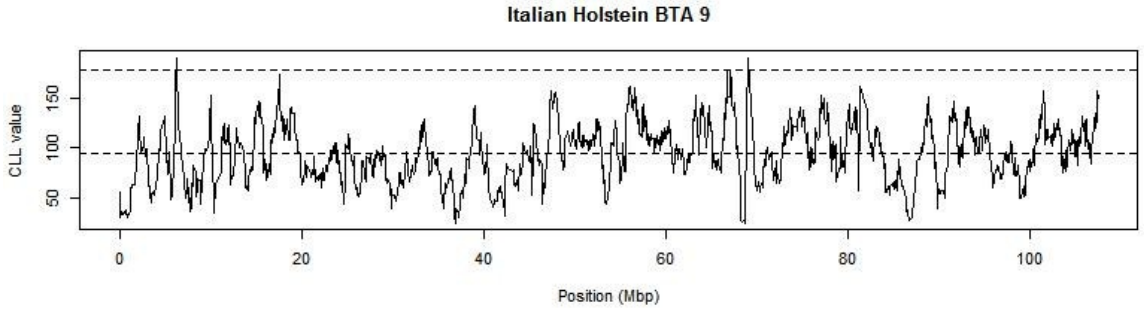

Italian Holstein BTA 8
CLL value
50
100
150
200
0
20
40
60
80
100
120
Position (Mbp)

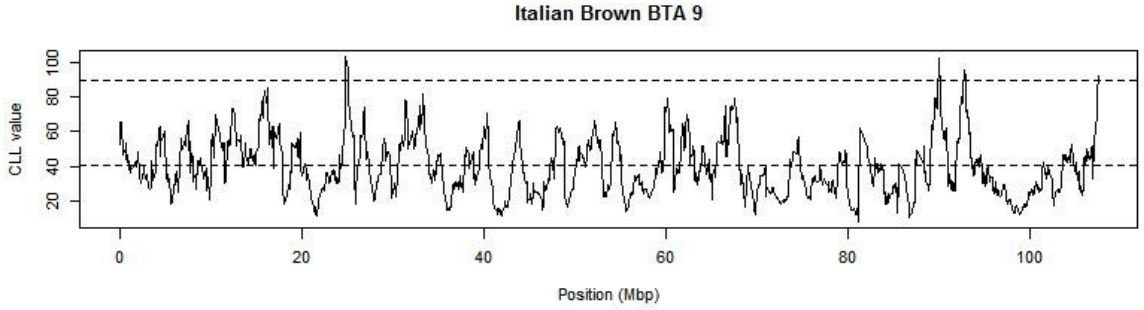

Italian Brown BTA 8
CLL value
20
40
60
80
120
0
20
40
60
80
100
120
Position (Mbp)

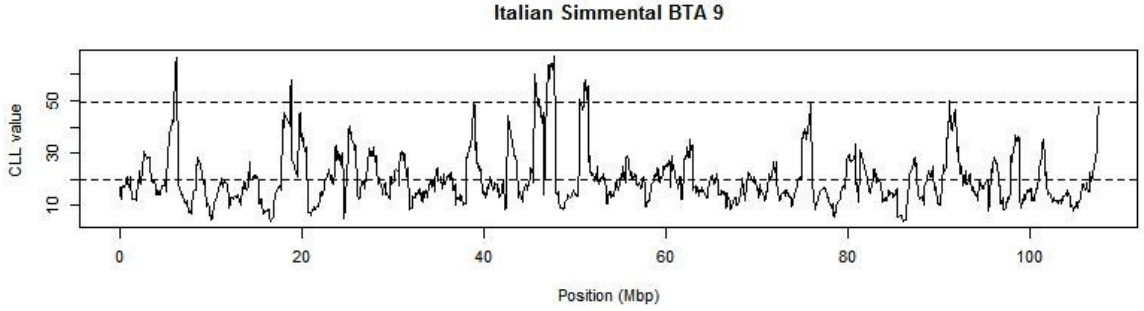

Italian Simmental BTA 8
CLL value
0
20
40
60
0
20
40
60
80
100
120
Position (Mbp)

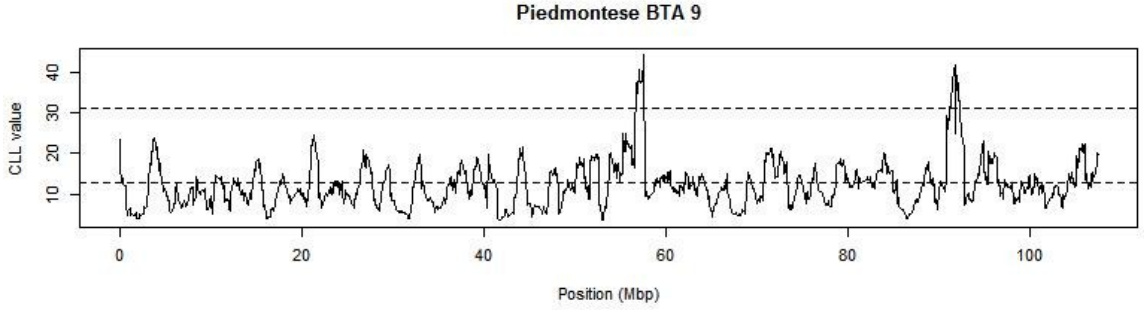

Piedmontese BTA 8
CLL value
10
20
30
40
50
0
20
40
60
80
100
120
Position (Mbp)

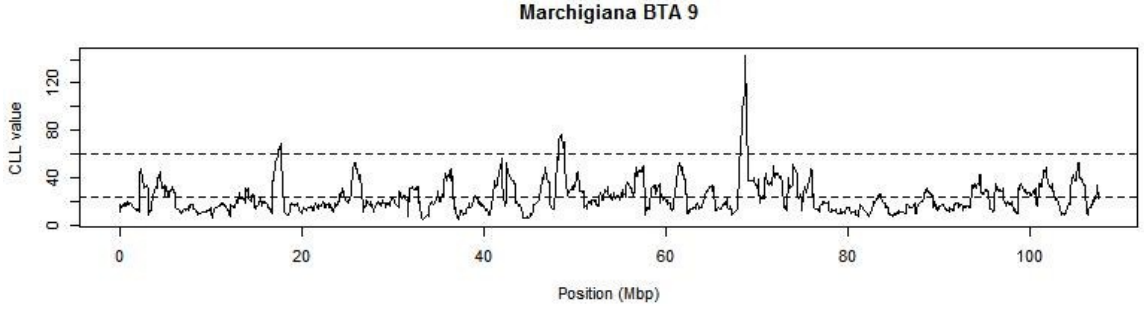

Marchigiana BTA 8
CLL value
0
20
40
60
80
0
20
40
60
80
100
120
Position (Mbp)

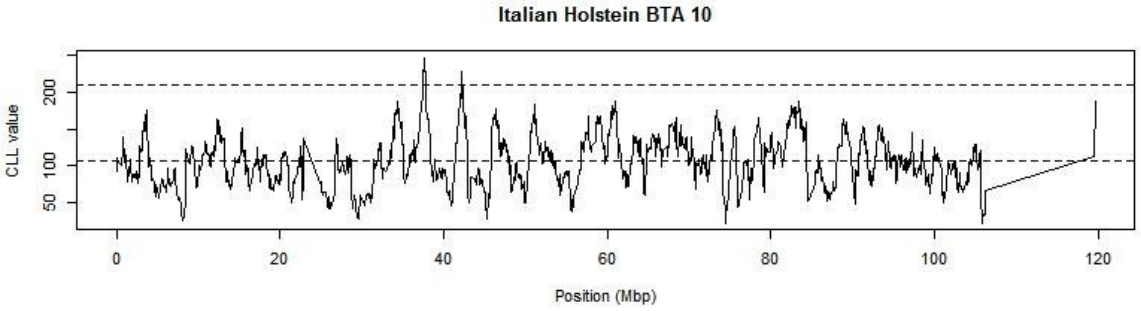

Italian Holstein BTA 9
CLL value
50
100
150
0
20
40
60
80
100
Position (Mbp)

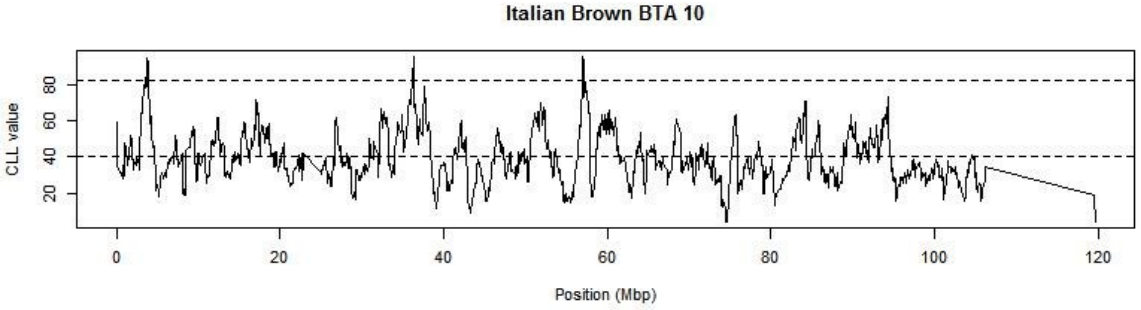

Italian Brown BTA 9
CLL value
20
40
60
80
100
0
20
40
60
80
100
Position (Mbp)

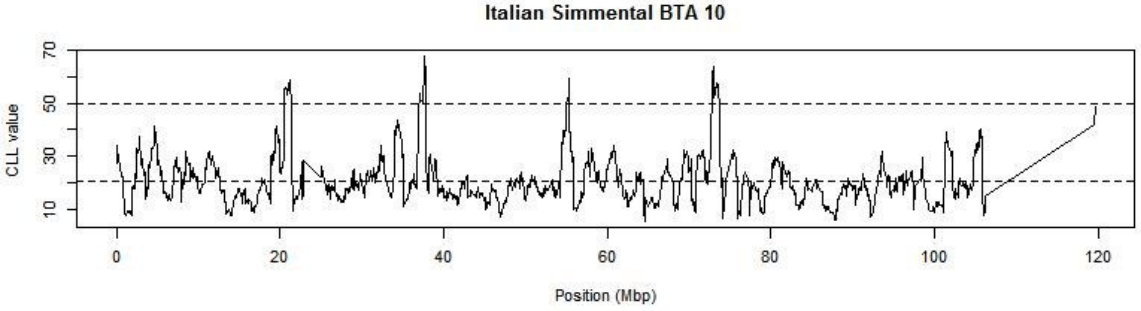

Italian Simmental BTA 9
CLL value
10
30
50
0
20
40
60
80
100
Position (Mbp)

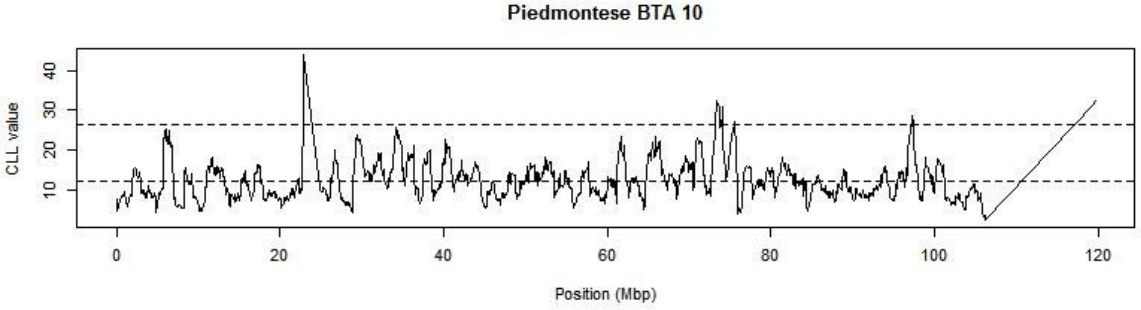

Piedmontese BTA 9
CLL value
10
20
30
40
0
20
40
60
80
100
Position (Mbp)

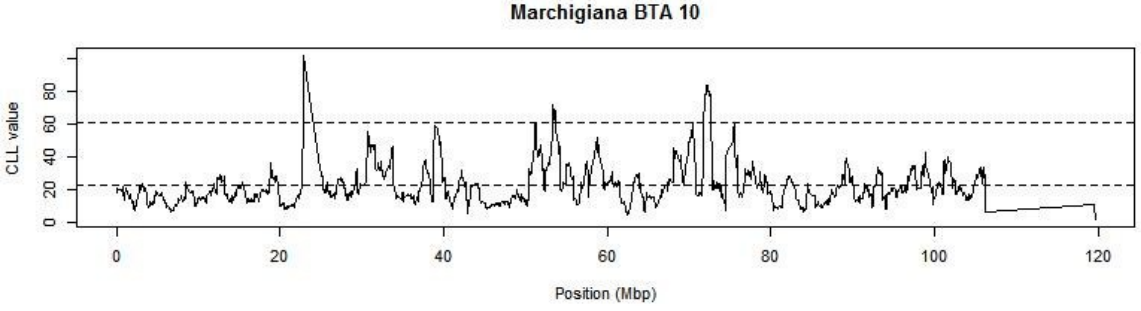

Marchigiana BTA 9
CLL value
0
40
80
120
0
20
40
60
80
100
Position (Mbp)

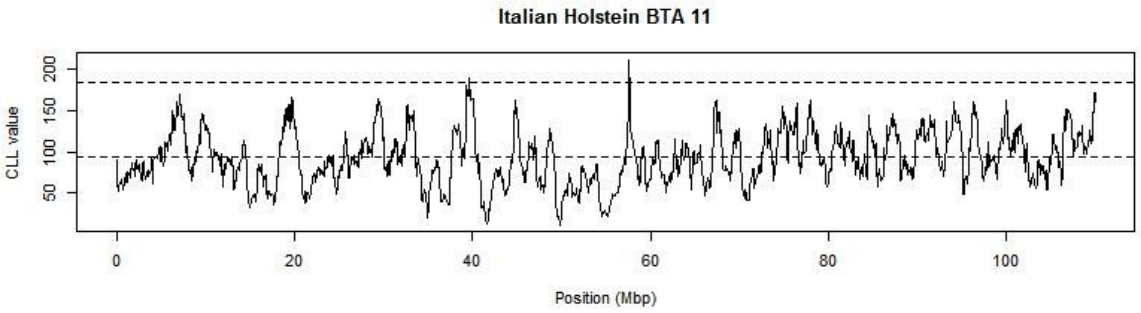

Italian Holstein BTA 10
CLL value
50
100
200
0
20
40
60
80
100
120
Position (Mbp)

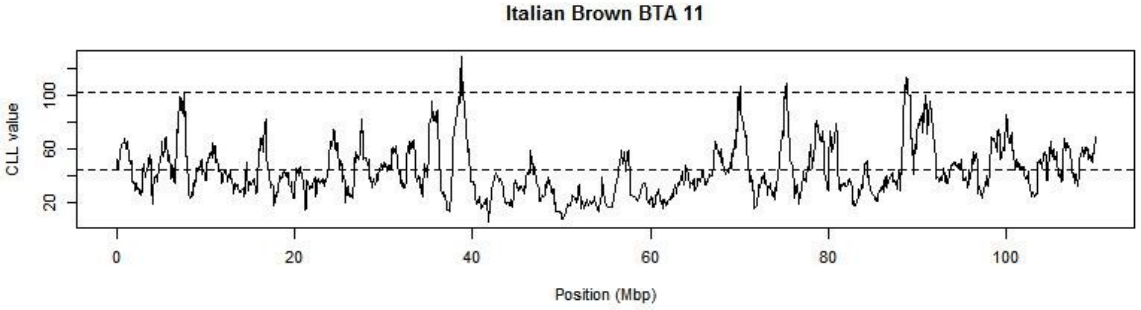

Italian Brown BTA 10
CLL value
20
40
60
80
0
20
40
60
80
100
120
Position (Mbp)

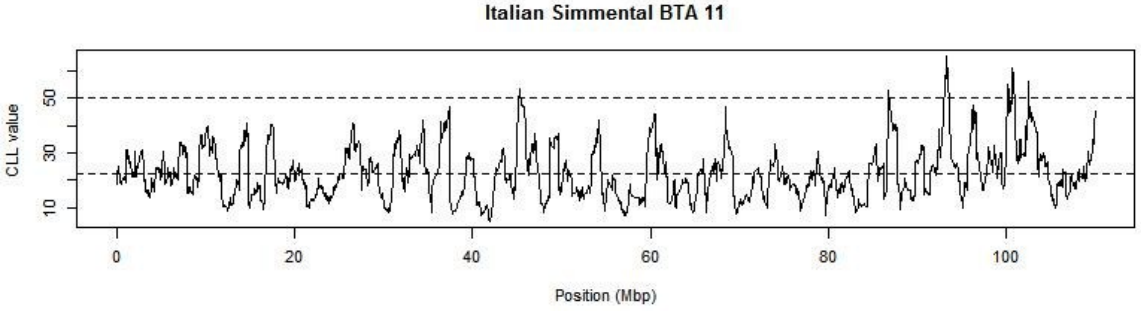

Italian Simmental BTA 10
CLL value
10
30
50
70
0
20
40
60
80
100
120
Position (Mbp)

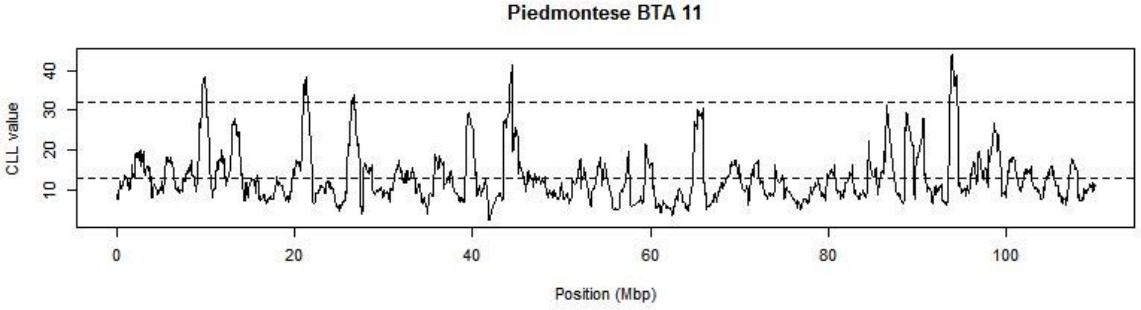

Piedmontese BTA 10
CLL value
10
20
30
40
0
20
40
60
80
100
120
Position (Mbp)

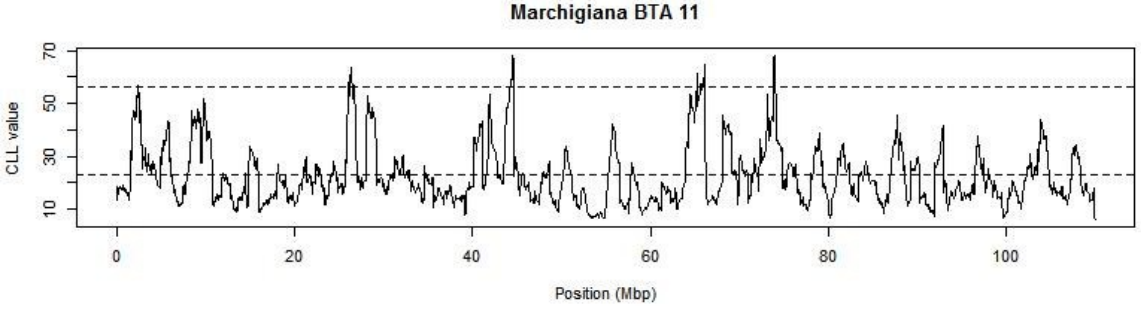

Marchigiana BTA 10
CLL value
0
20
40
60
80
0
20
40
60
80
100
120
Position (Mbp)

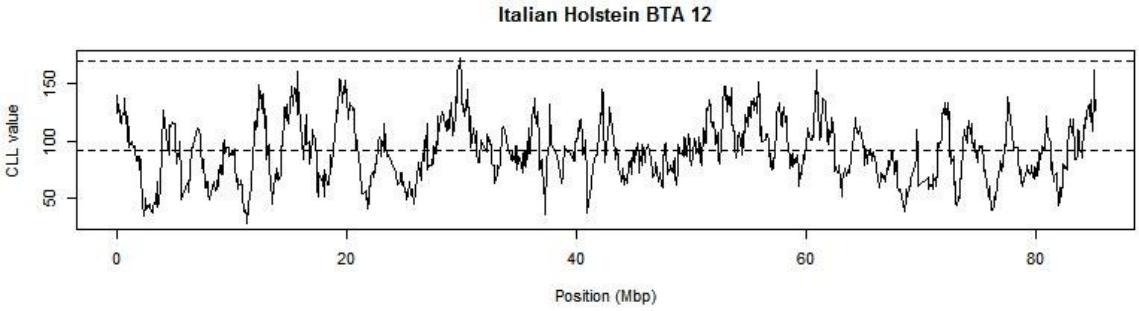

Italian Holstein BTA 11
CLL value
50
100
150
200
0
20
40
60
80
100
Position (Mbp)

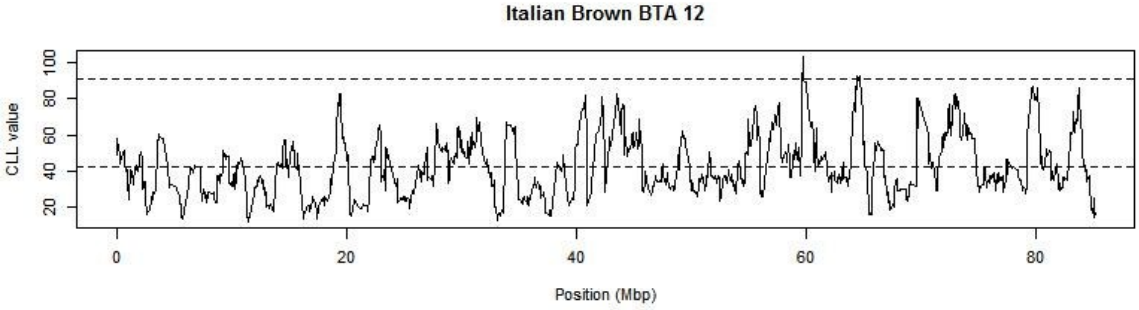

Italian Brown BTA 11
CLL value
20
60
100
0
20
40
60
80
100
Position (Mbp)

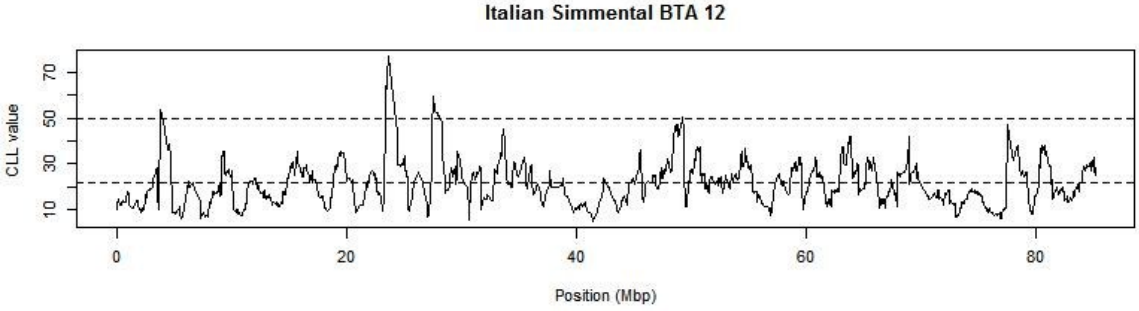

Italian Simmental BTA 11
CLL value
10
30
50
0
20
40
60
80
100
Position (Mbp)

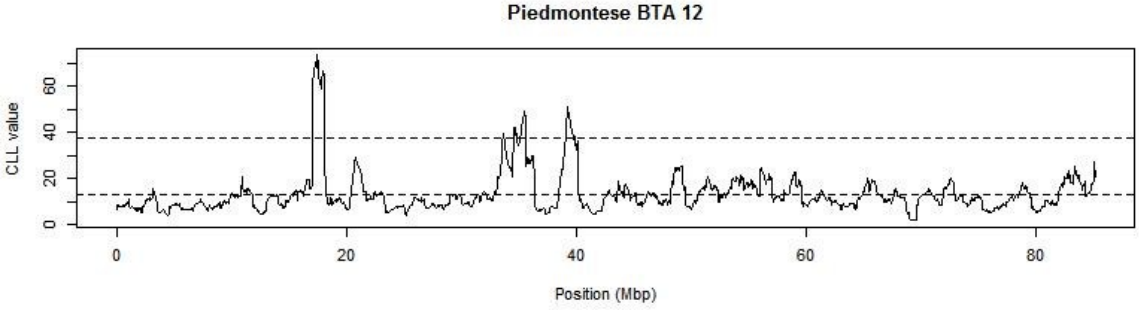

Piedmontese BTA 11
CLL value
10
20
30
40
0
20
40
60
80
100
Position (Mbp)

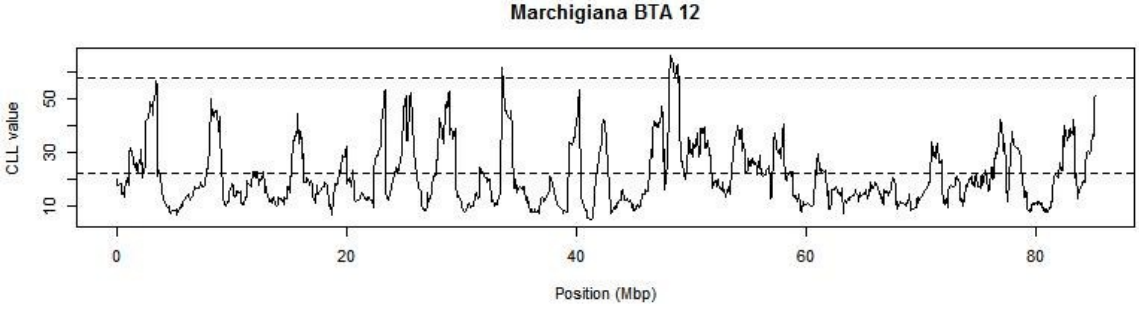

Marchigiana BTA 11
CLL value
10
30
50
70
0
20
40
60
80
100
Position (Mbp)

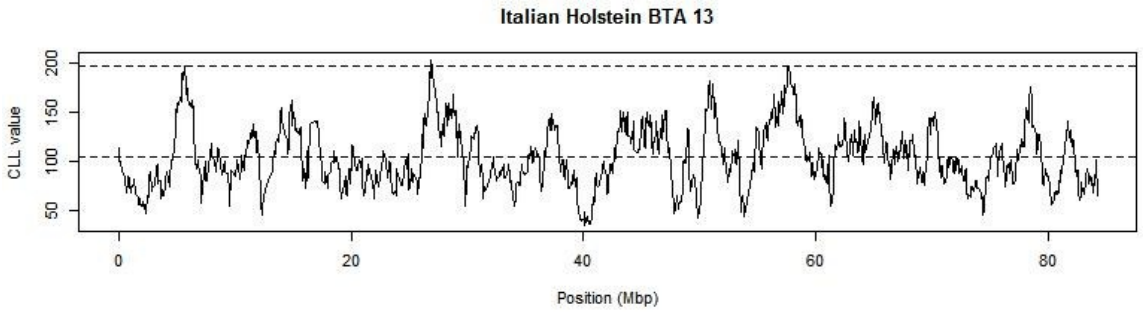

Italian Holstein BTA 12
CLL value
50
100
150
0
20
40
60
80
Position (Mbp)

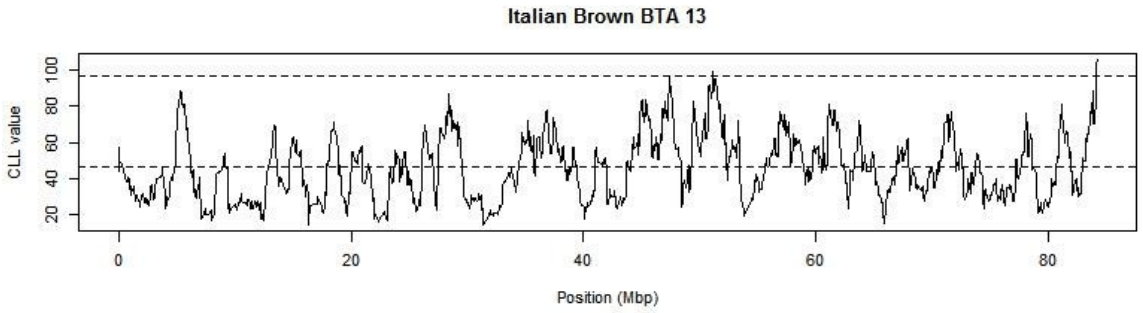

Italian Brown BTA 12
CLL value
20
40
60
80
100
0
20
40
60
80
Position (Mbp)

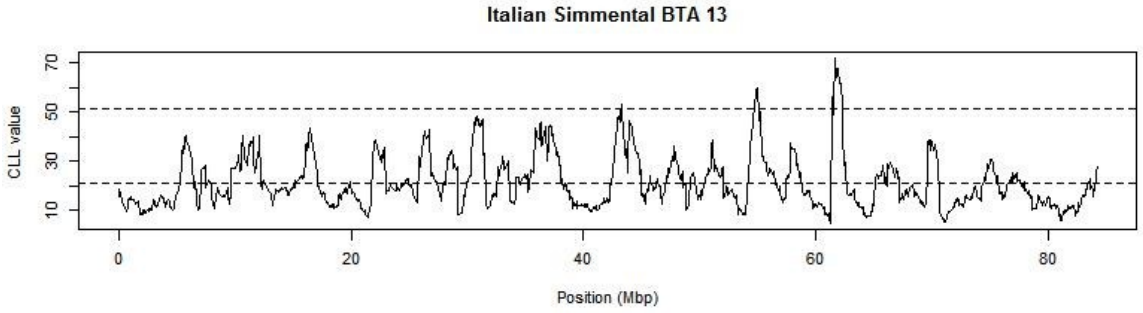

Italian Simmental BTA 12
CLL value
10
30
50
70
0
20
40
60
80
Position (Mbp)

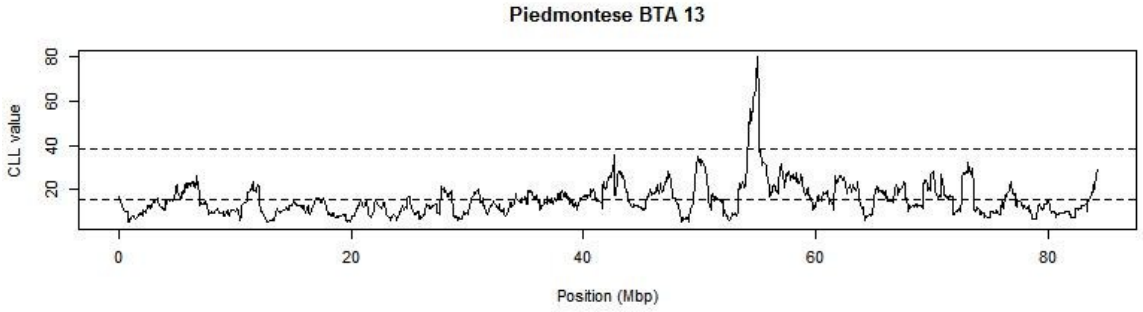

Piedmontese BTA 12
CLL value
0
20
40
60
0
20
40
60
80
Position (Mbp)

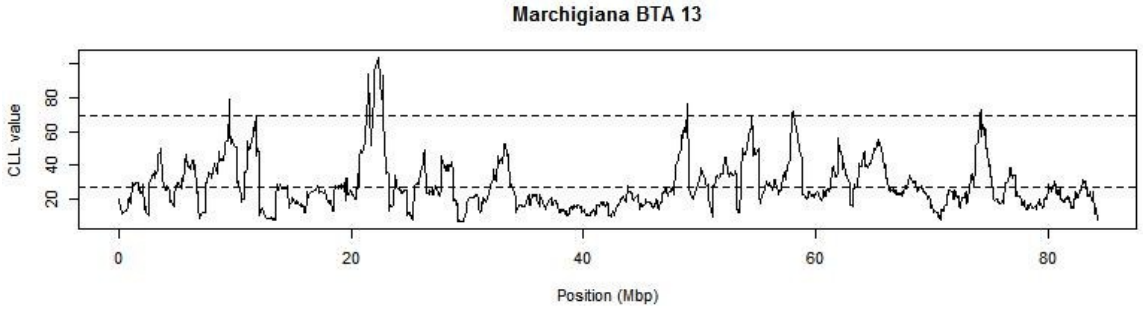

Marchigiana BTA 12
CLL value
10
30
50
0
20
40
60
80
Position (Mbp)

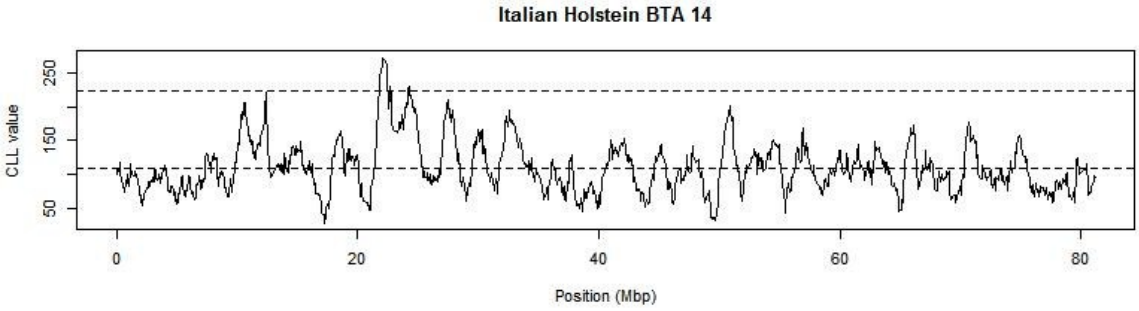

Italian Holstein BTA 13
CLL value
50
100
150
200
0
20
40
60
80
Position (Mbp)

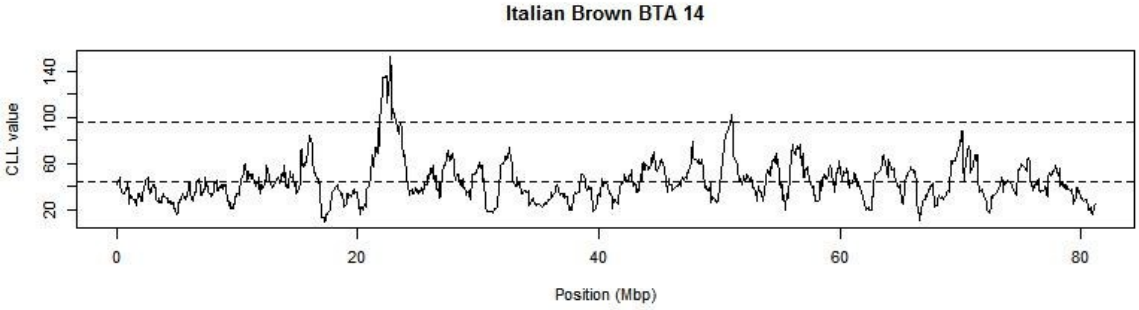

Italian Brown BTA 13
CLL value
20
40
60
80
100
0
20
40
60
80
Position (Mbp)

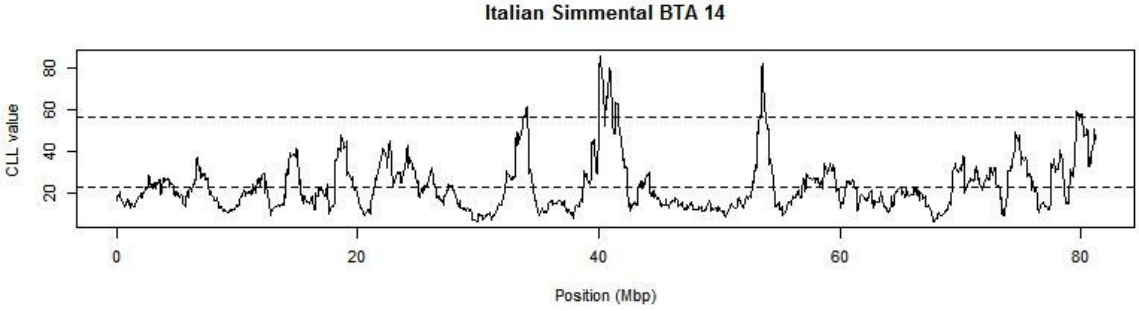

Italian Simmental BTA 13
CLL value
10
30
50
70
0
20
40
60
80
Position (Mbp)

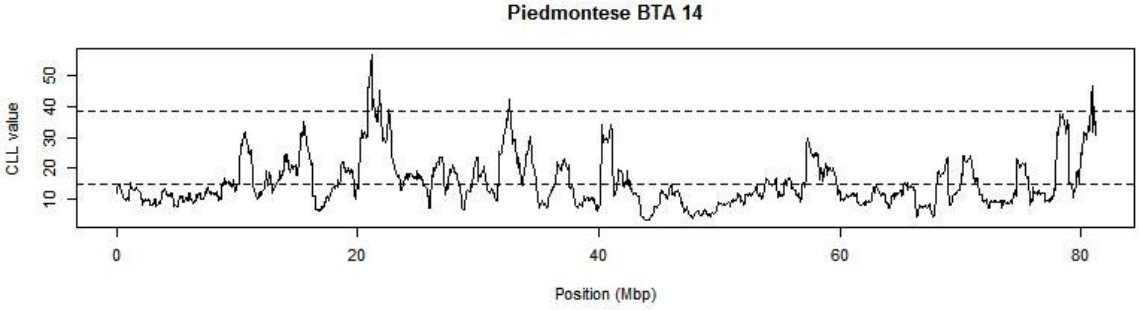

Piedmontese BTA 13
CLL value
20
40
60
80
0
20
40
60
80
Position (Mbp)

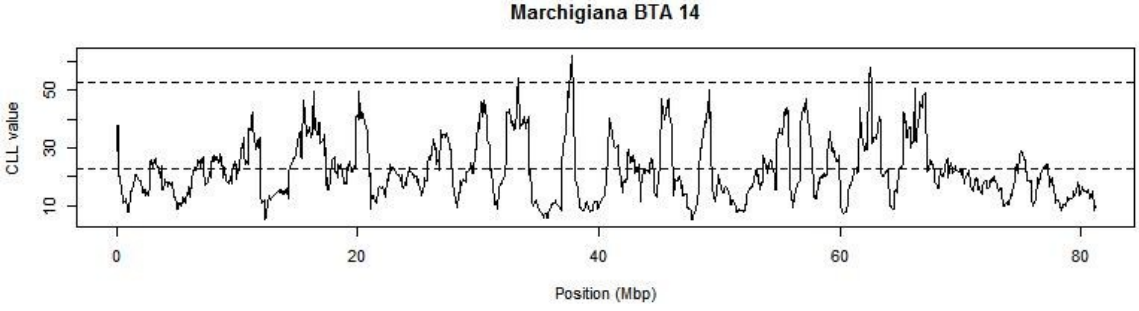

Marchigiana BTA 13
CLL value
20
40
60
80
0
20
40
60
80
Position (Mbp)

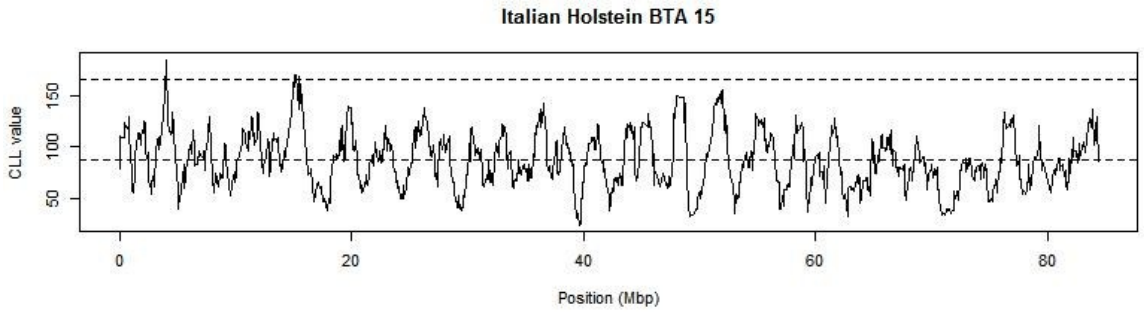

Italian Holstein BTA 14
CLL value
50
150
250
0
20
40
60
80
Position (Mbp)

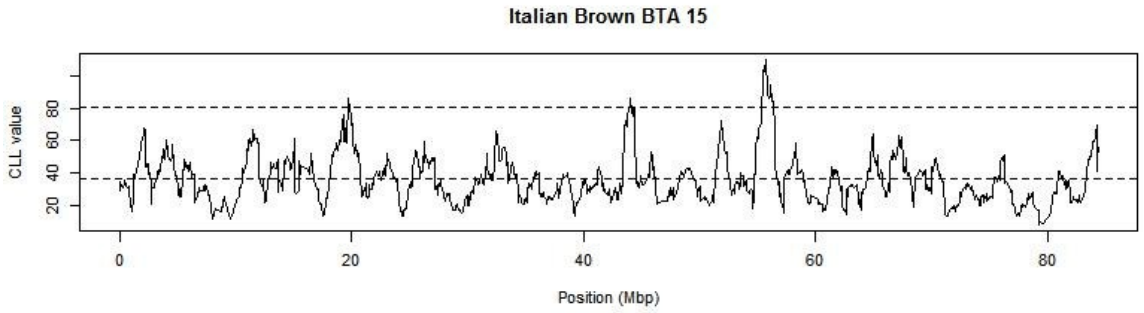

Italian Brown BTA 14
CLL value
20
60
100
140
0
20
40
60
80
Position (Mbp)

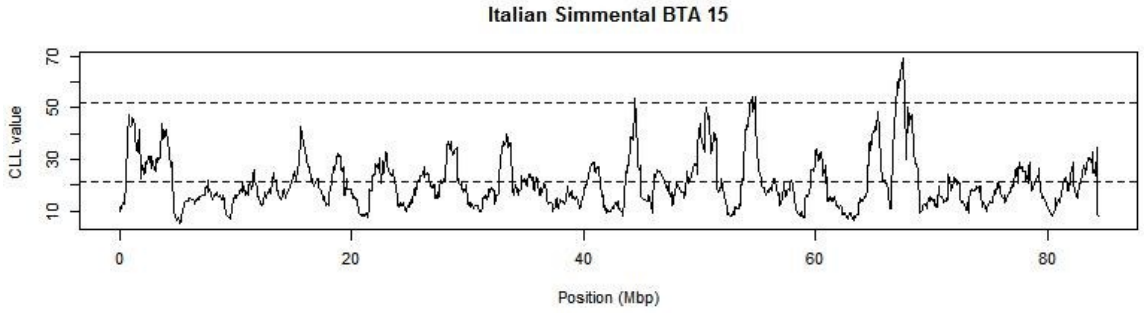

Italian Simmental BTA 14
CLL value
20
40
60
80
0
20
40
60
80
Position (Mbp)

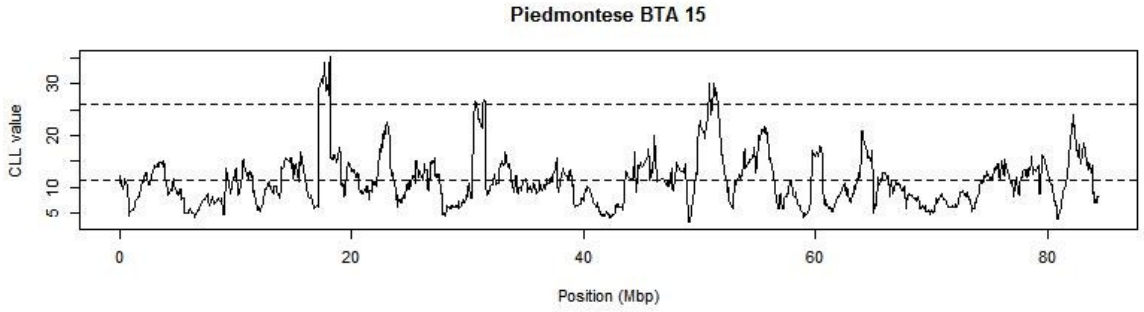

Piedmontese BTA 14
CLL value
10
20
30
40
50
0
20
40
60
80
Position (Mbp)

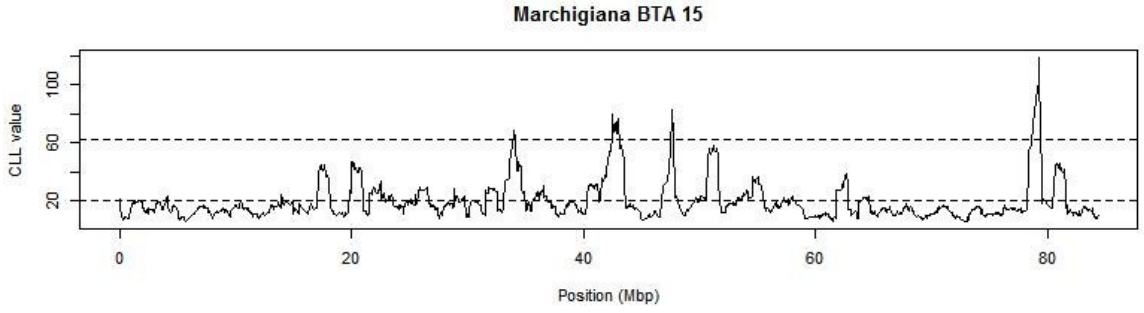

Marchigiana BTA 14
CLL value
10
30
50
0
20
40
60
80
Position (Mbp)

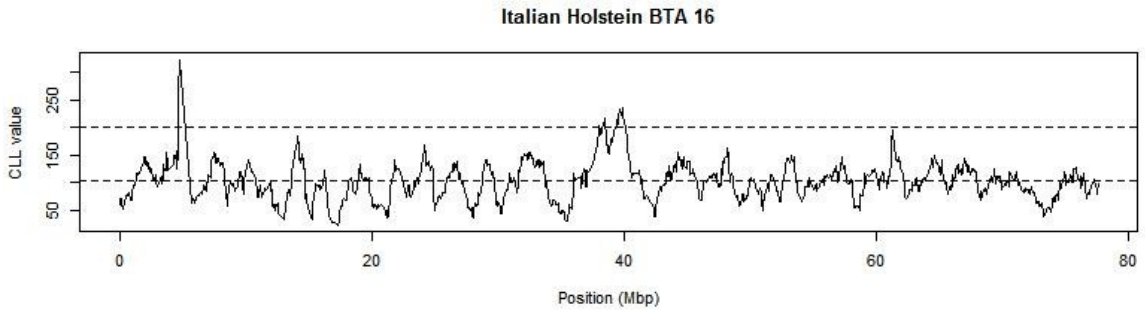

Italian Holstein BTA 15
CLL value
50
100
150
0
20
40
60
80
Position (Mbp)

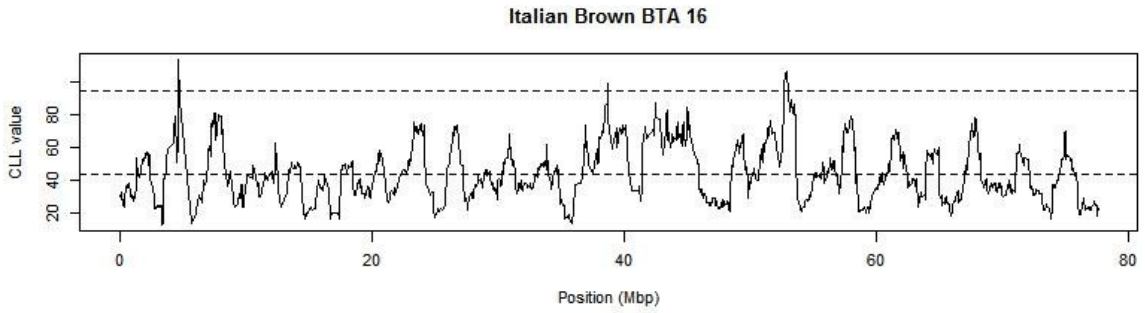

Italian Brown BTA 15
CLL value
20
40
60
80
0
20
40
60
80
Position (Mbp)

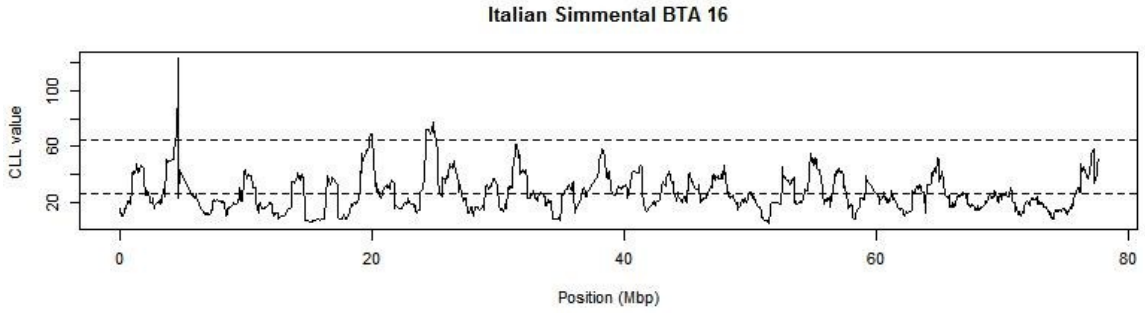

Italian Simmental BTA 15
CLL value
10
30
50
70
0
20
40
60
80
Position (Mbp)

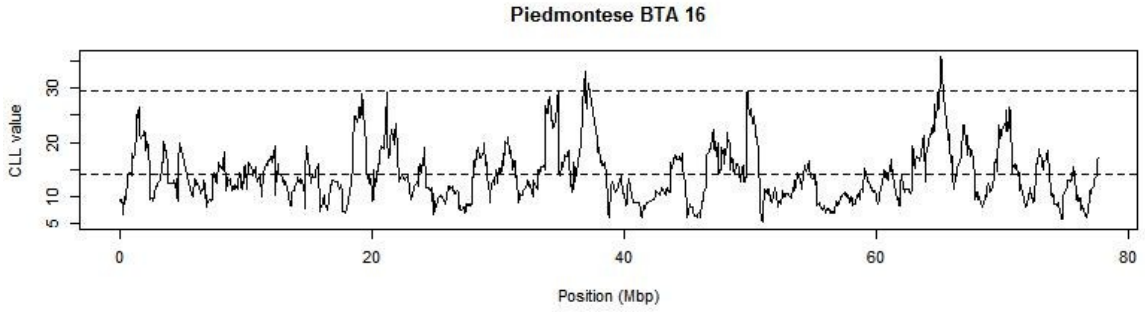

Piedmontese BTA 15
CLL value
5
10
20
30
0
20
40
60
80
Position (Mbp)

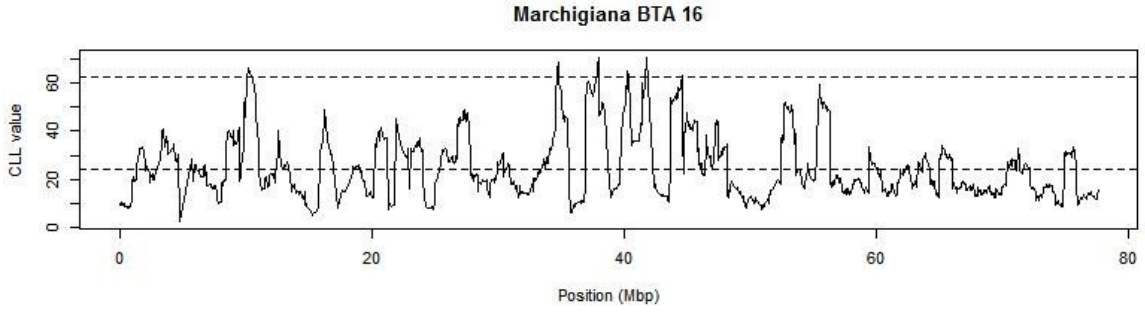

Marchigiana BTA 15
CLL value
20
60
100
0
20
40
60
80
Position (Mbp)

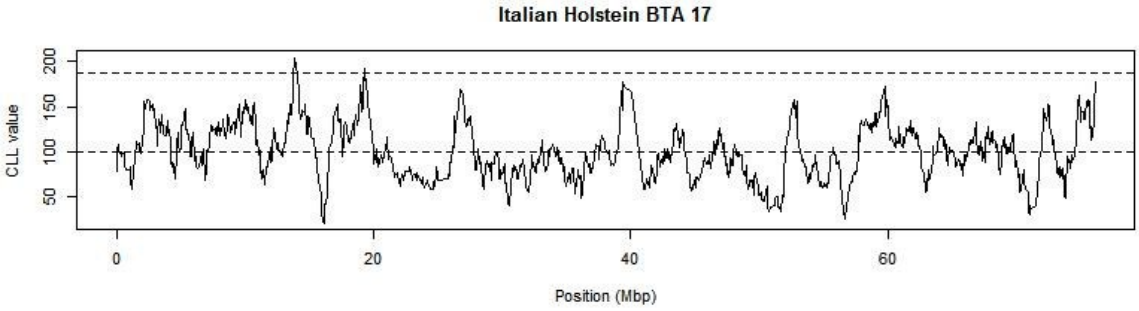

Italian Holstein BTA 16
CLL value
50
150
250
0
20
40
60
80
Position (Mbp)

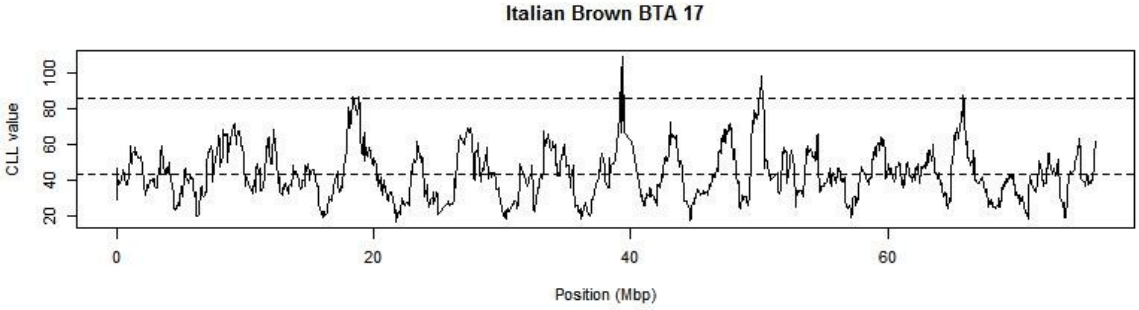

Italian Brown BTA 16
CLL value
20
40
60
80
0
20
40
60
80
Position (Mbp)

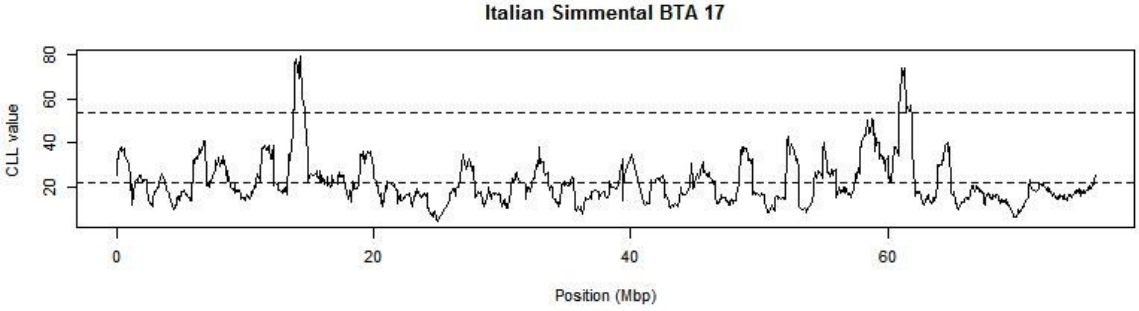

Italian Simmental BTA 16
CLL value
20
60
100
0
20
40
60
80
Position (Mbp)

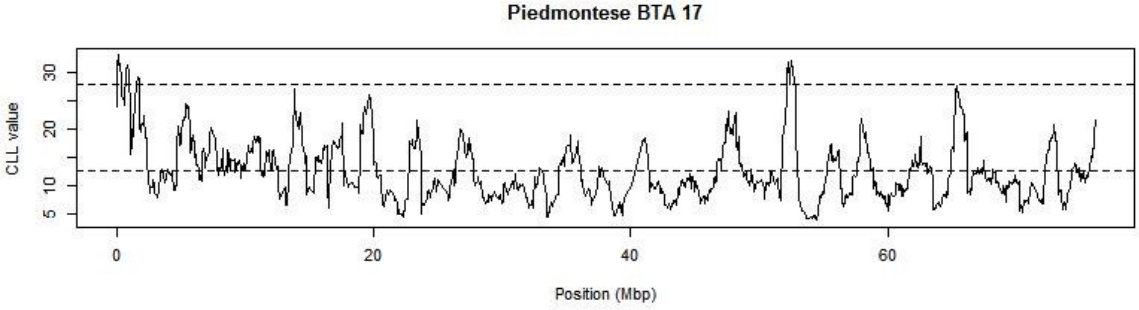

Piedmontese BTA 16
CLL value
5
10
20
30
0
20
40
60
80
Position (Mbp)

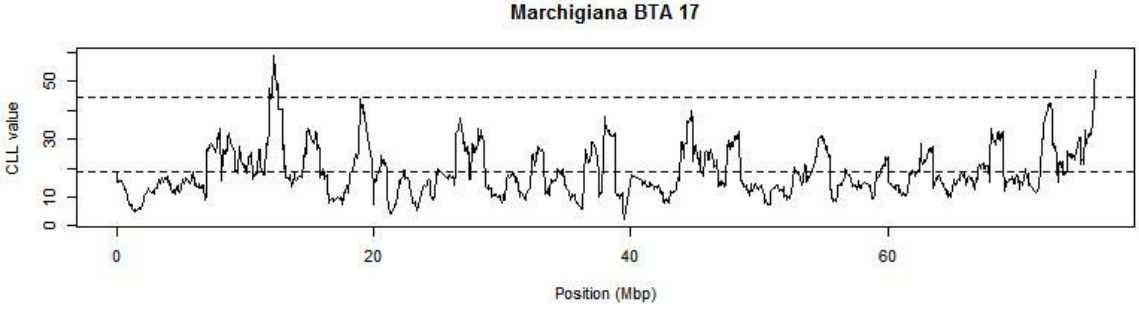

Marchigiana BTA 16
CLL value
0
20
40
60
0
20
40
60
80
Position (Mbp)

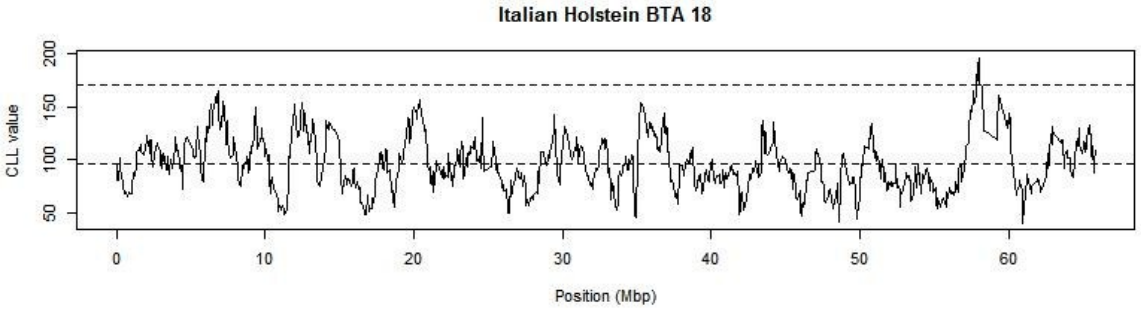

Italian Holstein BTA 17
CLL value
50
100
150
200
0
20
40
60
Position (Mbp)

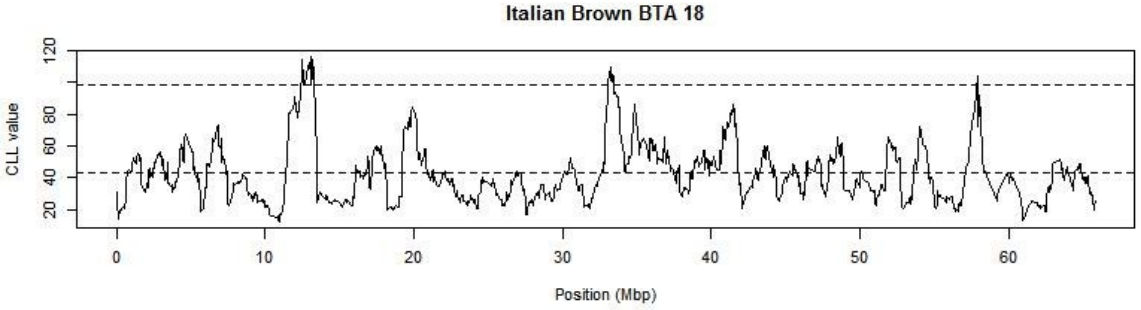

Italian Brown BTA 17
CLL value
20
40
60
80
100
0
20
40
60
Position (Mbp)

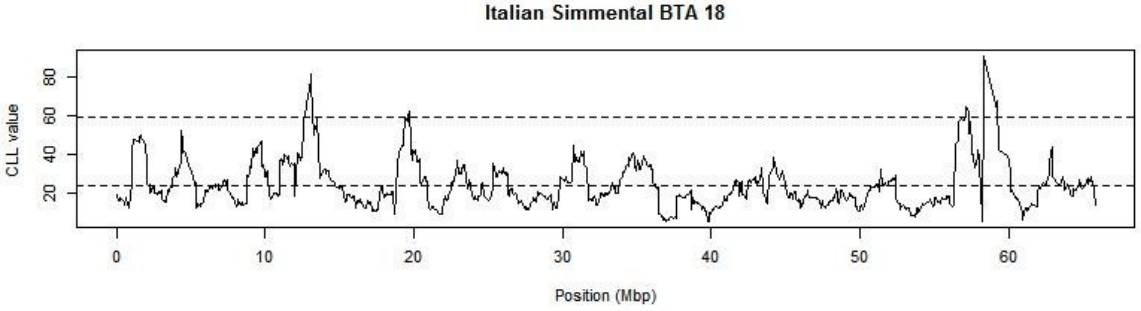

Italian Simmental BTA 17
CLL value
20
40
60
80
0
20
40
60
Position (Mbp)

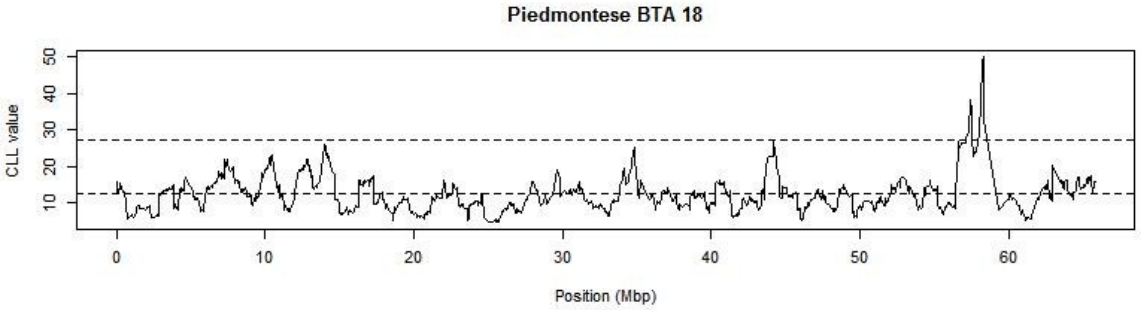

Piedmontese BTA 17
CLL value
5
10
20
30
0
20
40
60
Position (Mbp)

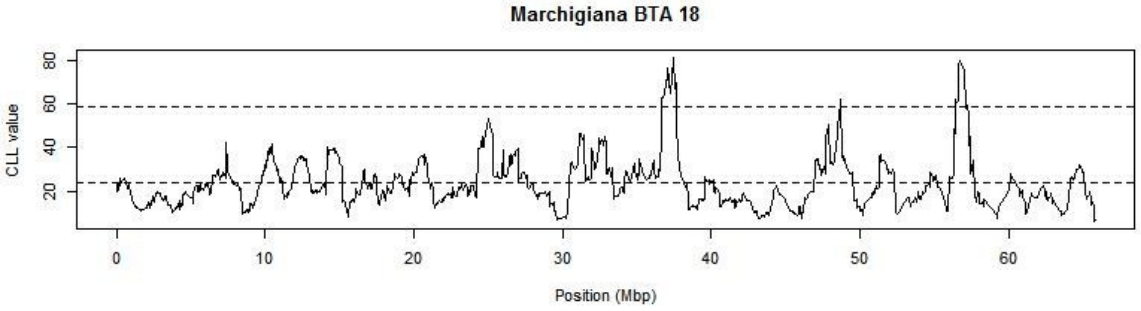

Marchigiana BTA 17
CLL value
0
10
30
50
0
20
40
60
Position (Mbp)

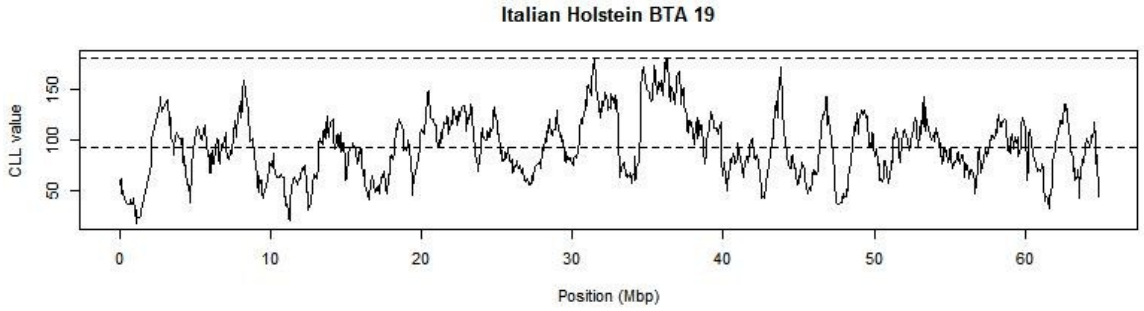

Italian Holstein BTA 18
CLL value
50
100
150
200
0
10
20
30
40
50
60
Position (Mbp)

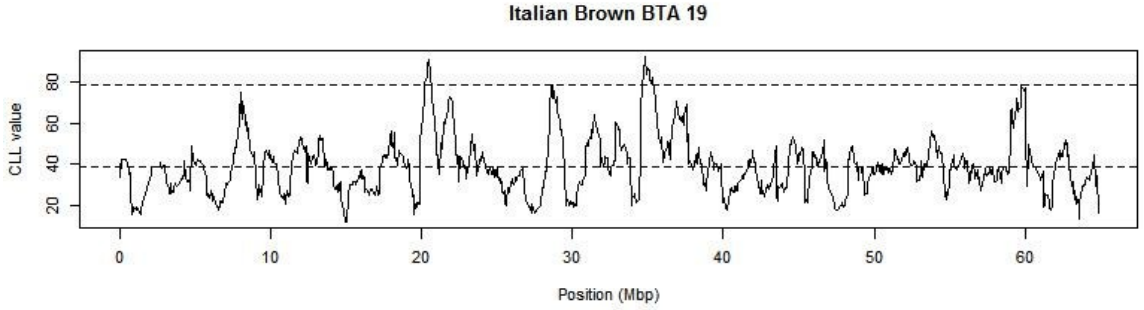

Italian Brown BTA 18
CLL value
20
40
60
80
120
0
10
20
30
40
50
60
Position (Mbp)

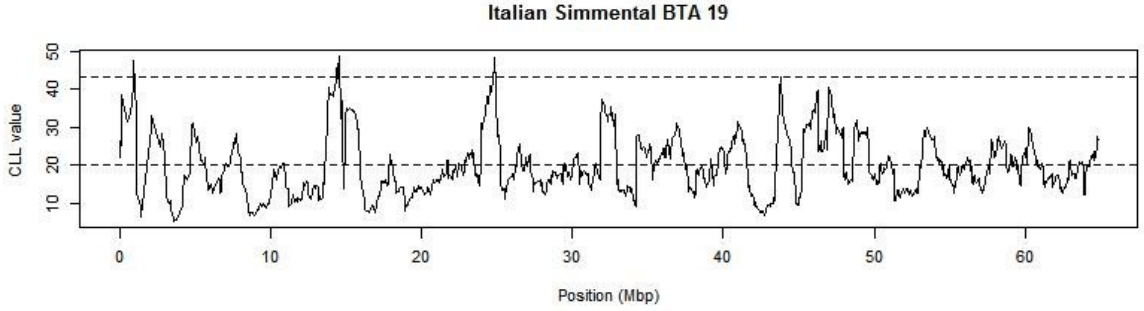

Italian Simmental BTA 18
CLL value
20
40
60
80
0
10
20
30
40
50
60
Position (Mbp)

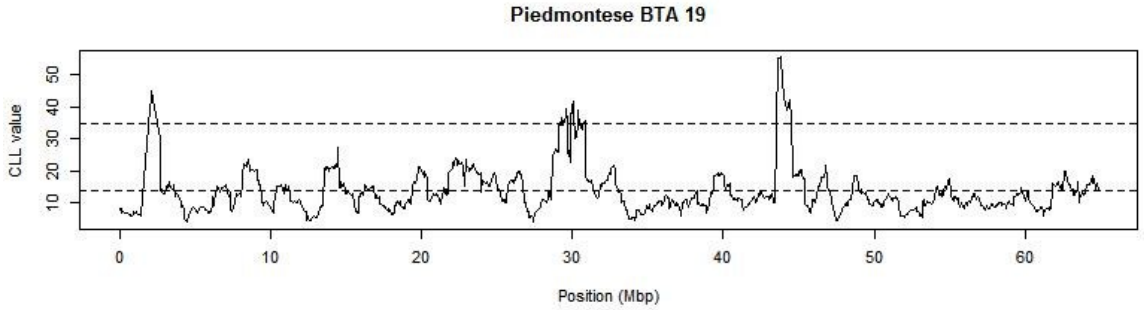

Piedmontese BTA 18
CLL value
10
20
30
40
50
0
10
20
30
40
50
60
Position (Mbp)

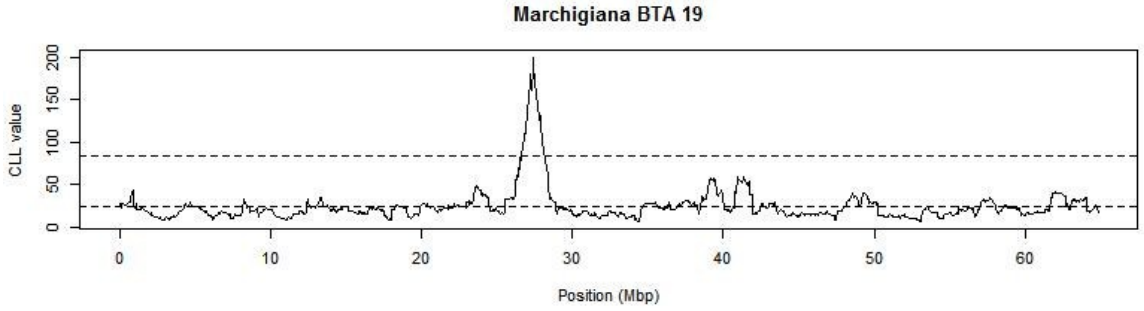

Marchigiana BTA 18
CLL value
20
40
60
80
0
10
20
30
40
50
60
Position (Mbp)

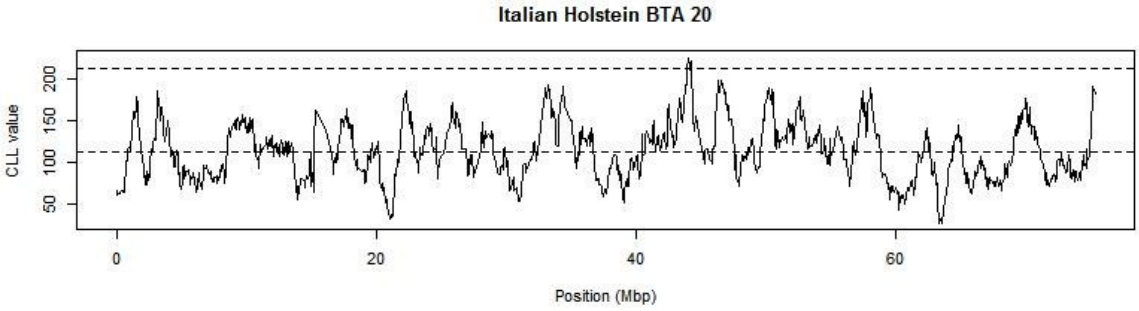

Italian Holstein BTA 19
CLL value
50
100
150
0
10
20
30
40
50
60
Position (Mbp)

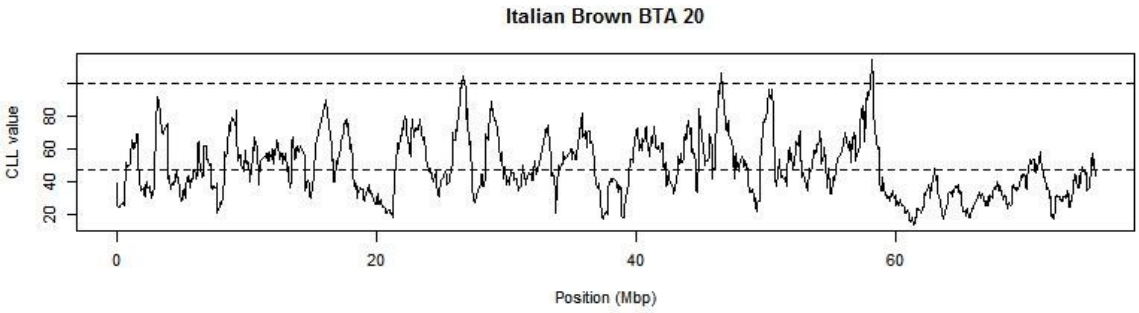

Italian Brown BTA 19
CLL value
20
40
60
80
0
10
20
30
40
50
60
Position (Mbp)

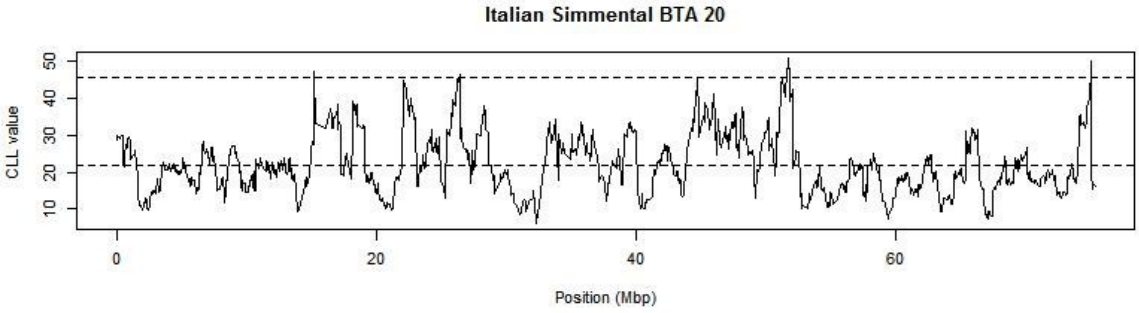

Italian Simmental BTA 19
CLL value
10
20
30
40
50
0
10
20
30
40
50
60
Position (Mbp)

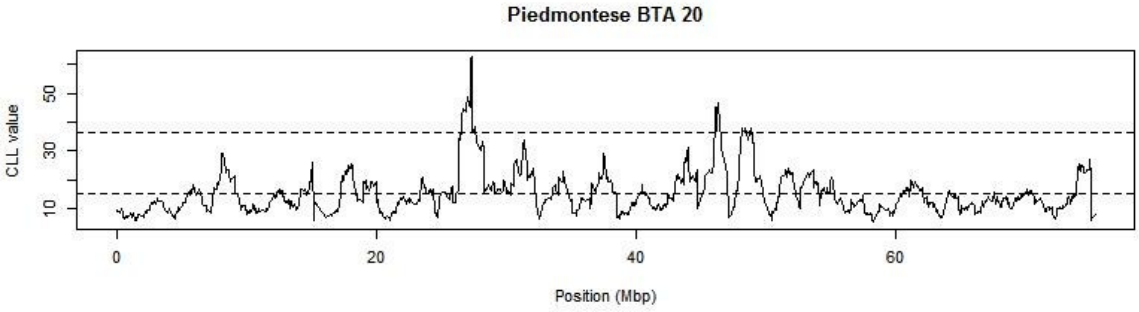

Piedmontese BTA 19
CLL value
10
20
30
40
50
0
10
20
30
40
50
60
Position (Mbp)

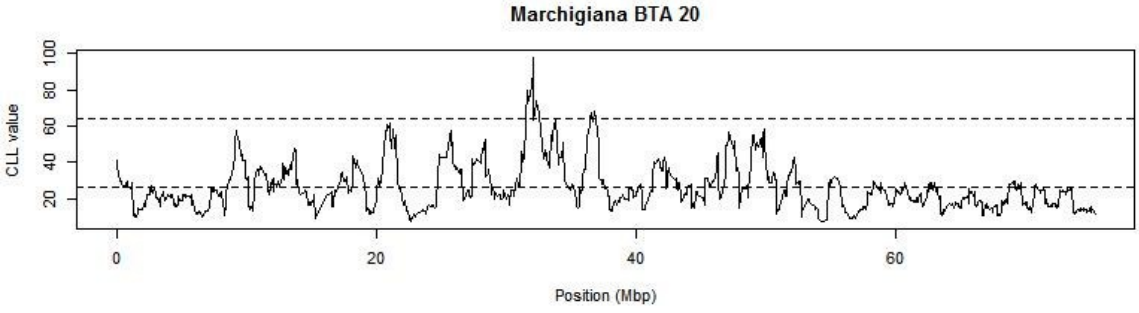

Marchigiana BTA 19
CLL value
0
50
100
150
200
0
10
20
30
40
50
60
Position (Mbp)

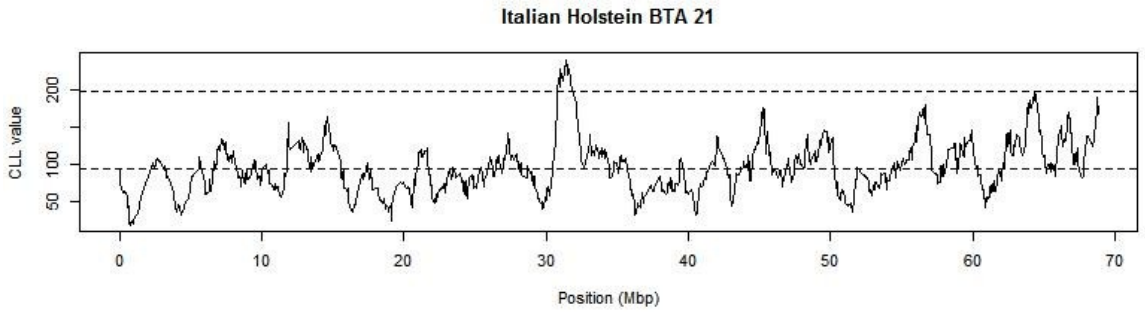

Italian Holstein BTA 20
CLL value
50
100
150
200
0
20
40
60
Position (Mbp)

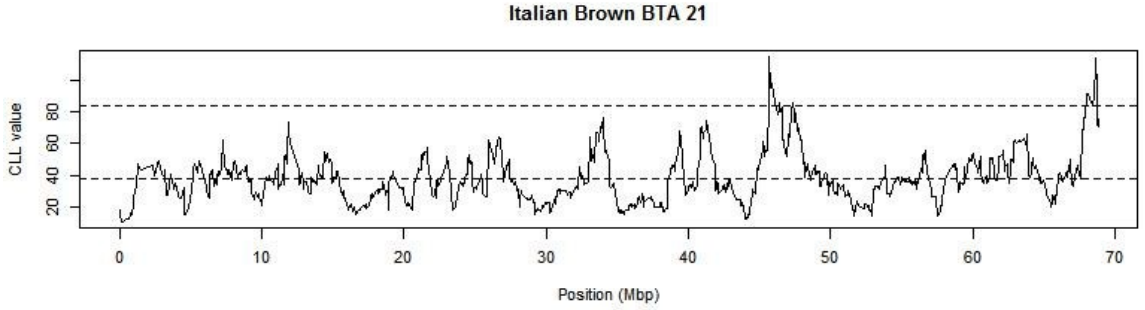

Italian Brown BTA 20
CLL value
20
40
60
80
0
20
40
60
Position (Mbp)

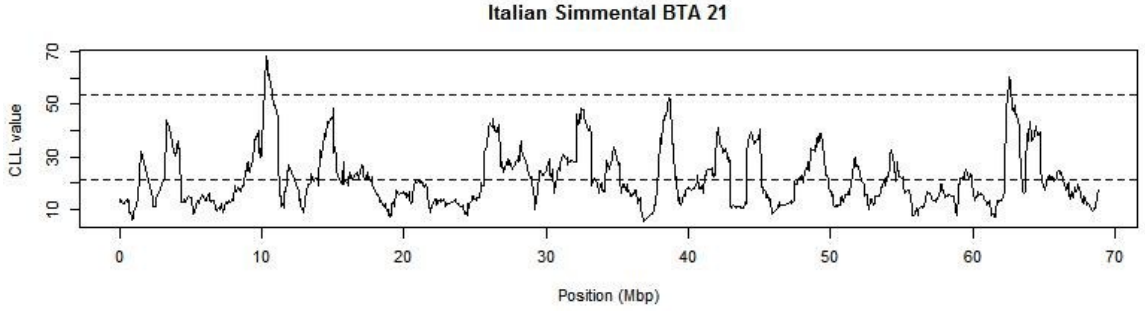

Italian Simmental BTA 20
CLL value
10
20
30
40
50
0
20
40
60
Position (Mbp)

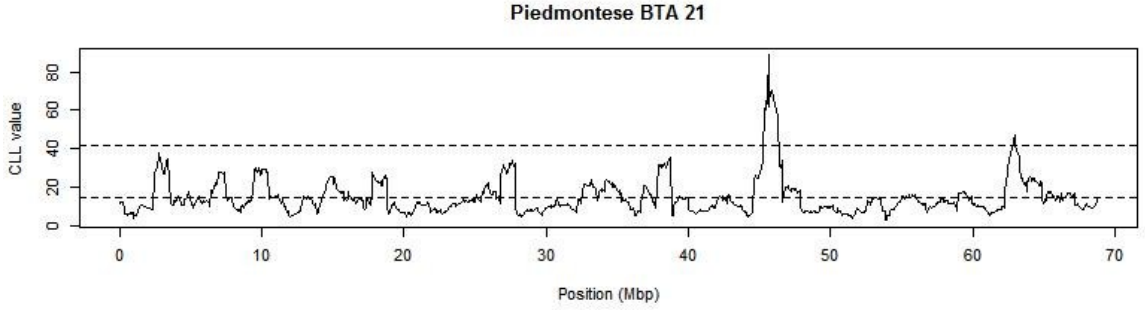

Piedmontese BTA 20
CLL value
10
30
50
0
20
40
60
Position (Mbp)

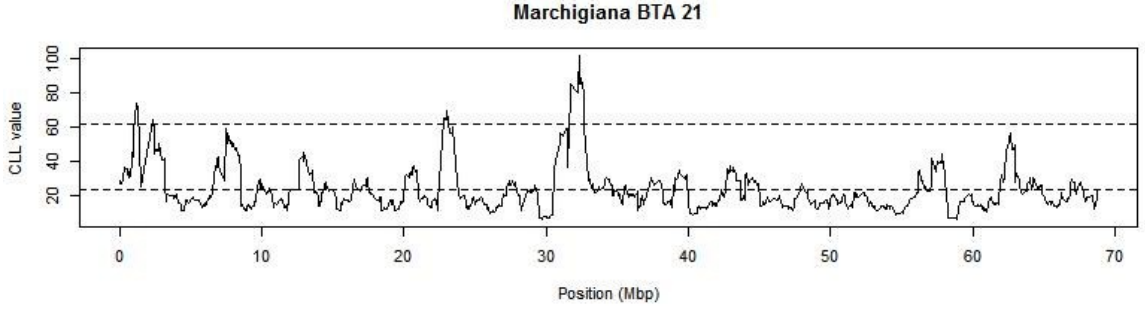

Marchigiana BTA 20
CLL value
20
40
60
80
100
0
20
40
60
Position (Mbp)

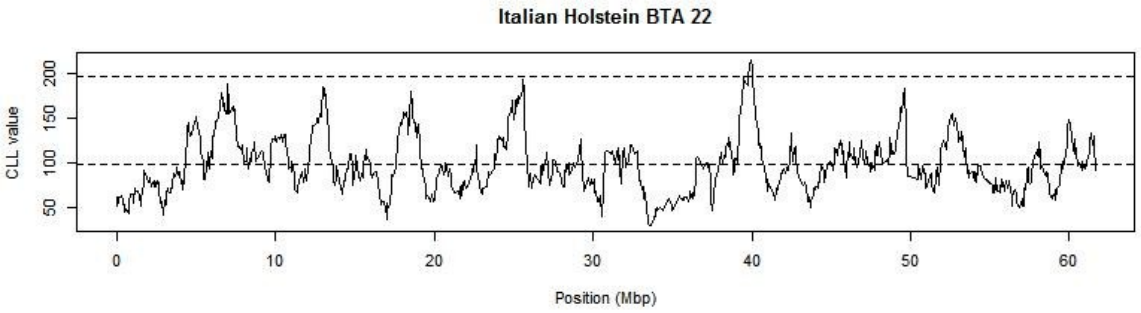

Italian Holstein BTA 21
CLL value
50
100
200
0
10
20
30
40
50
60
70
Position (Mbp)

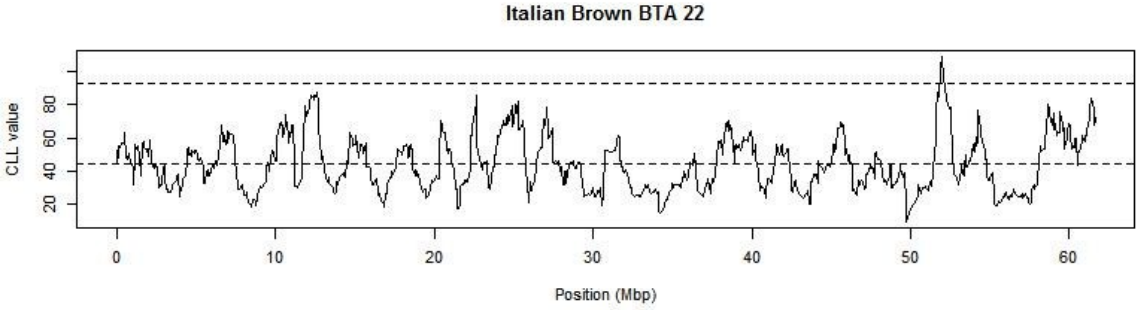

Italian Brown BTA 21
CLL value
20
40
60
80
0
10
20
30
40
50
60
70
Position (Mbp)

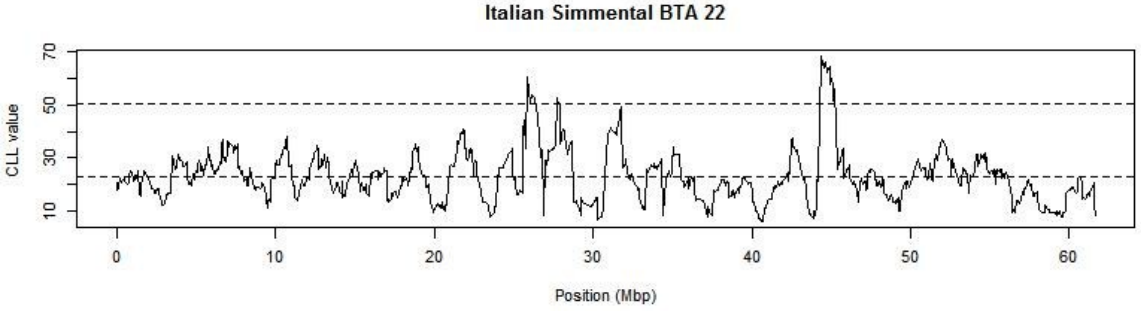

Italian Simmental BTA 21
CLL value
10
30
50
70
0
10
20
30
40
50
60
70
Position (Mbp)

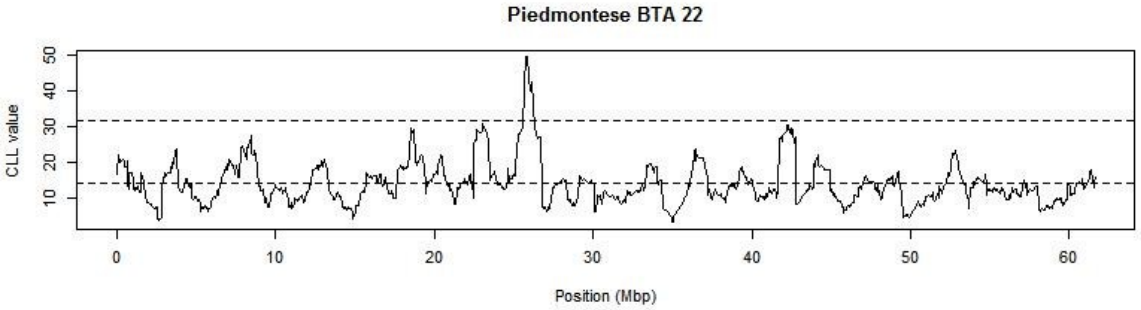

Piedmontese BTA 21
CLL value
0
20
40
60
80
0
10
20
30
40
50
60
70
Position (Mbp)

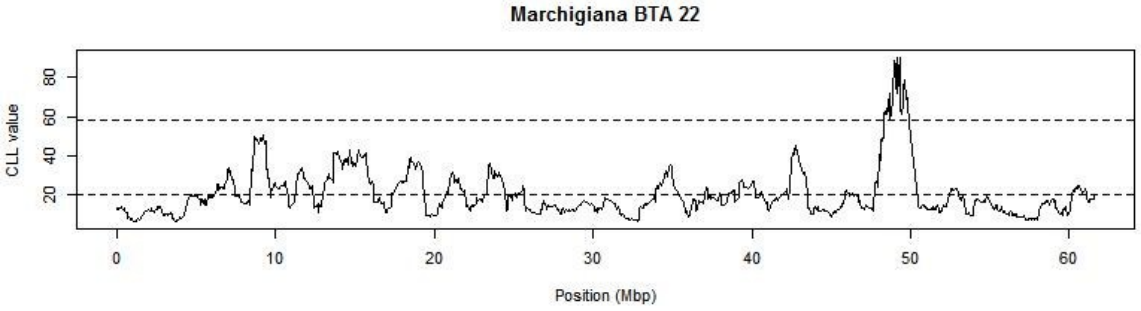

Marchigiana BTA 21
CLL value
20
40
60
80
100
0
10
20
30
40
50
60
70
Position (Mbp)

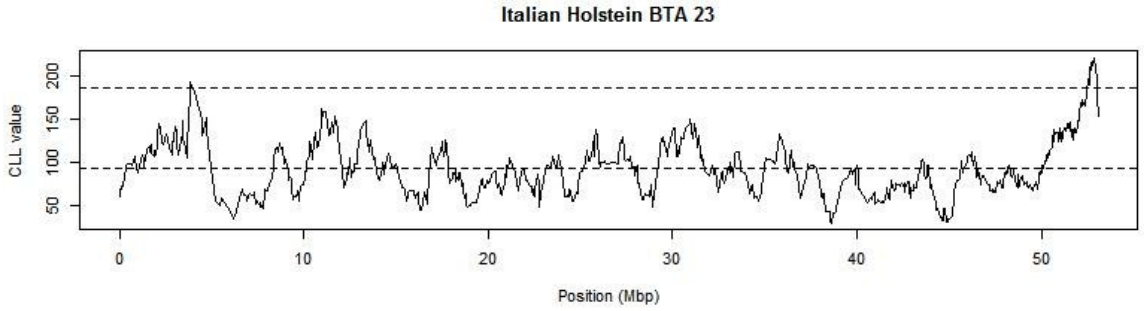

Italian Holstein BTA 22
CLL value
50
100
150
200
0
10
20
30
40
50
60
Position (Mbp)

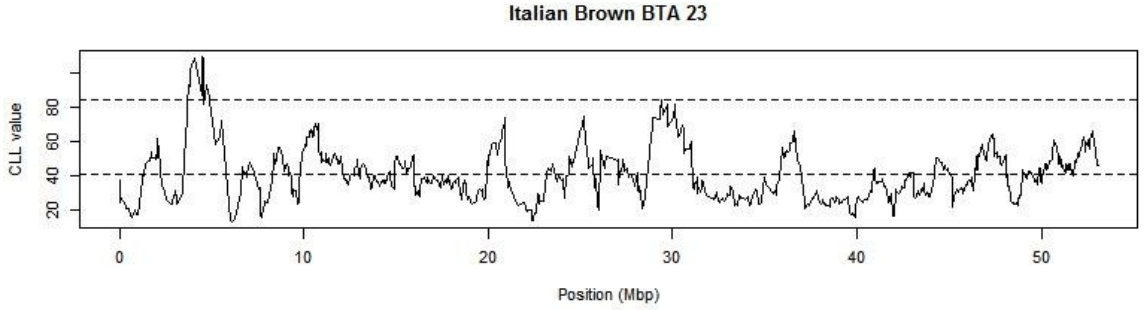

Italian Brown BTA 22
CLL value
20
40
60
80
0
10
20
30
40
50
60
Position (Mbp)

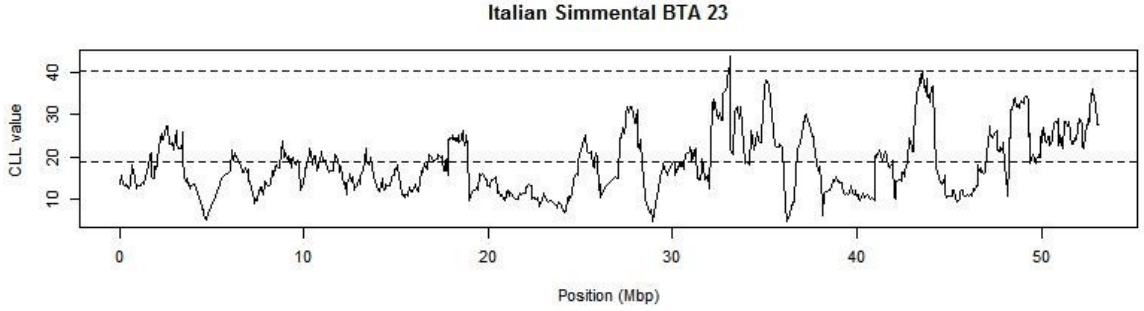

Italian Simmental BTA 22
CLL value
10
30
50
70
0
10
20
30
40
50
60
Position (Mbp)

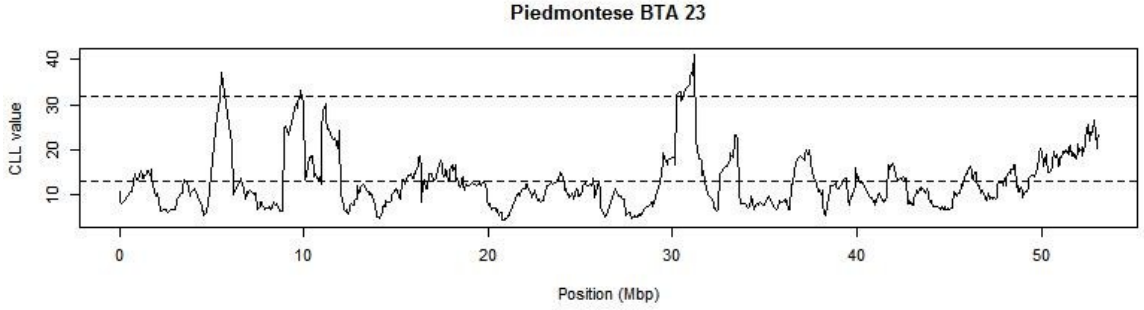

Piedmontese BTA 22
CLL value
10
20
30
40
50
0
10
20
30
40
50
60
Position (Mbp)

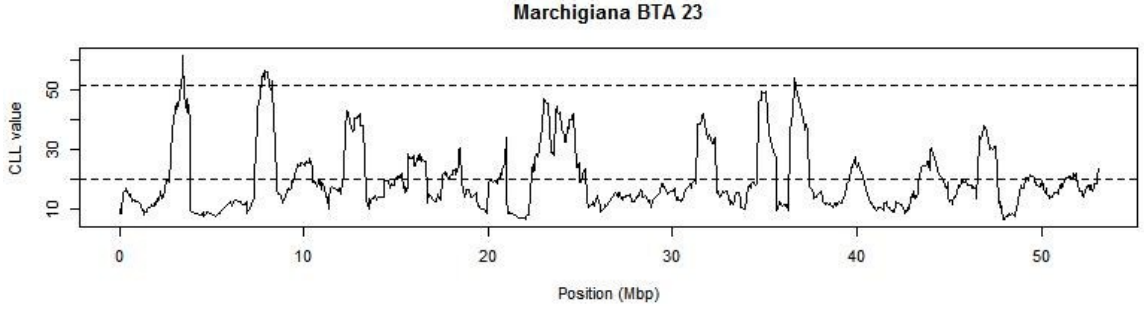

Marchigiana BTA 22
CLL value
20
40
60
80
0
10
20
30
40
50
60
Position (Mbp)

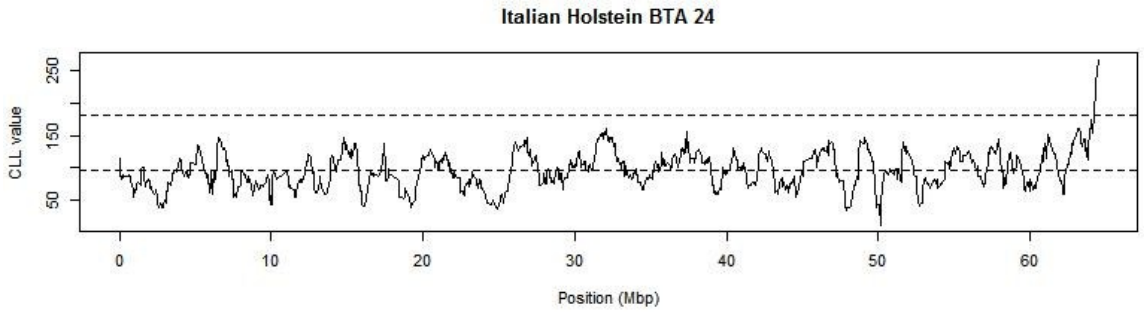

Italian Holstein BTA 23
CLL value
50
100
150
200
0
10
20
30
40
50
Position (Mbp)

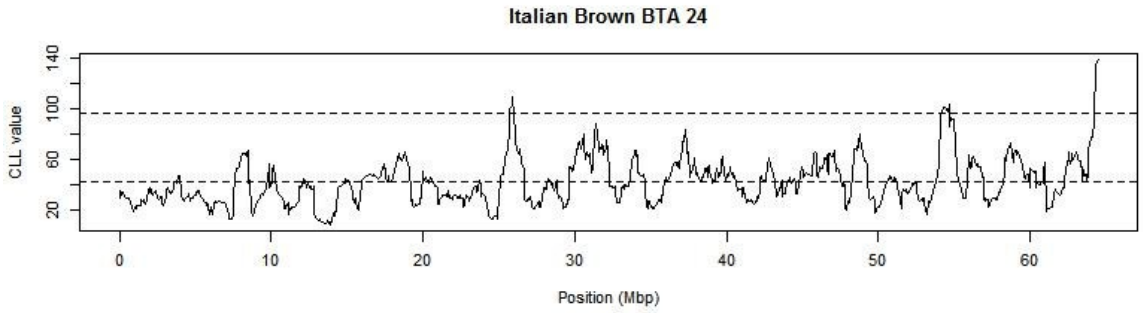

Italian Brown BTA 23
CLL value
20
40
60
80
0
10
20
30
40
50
Position (Mbp)

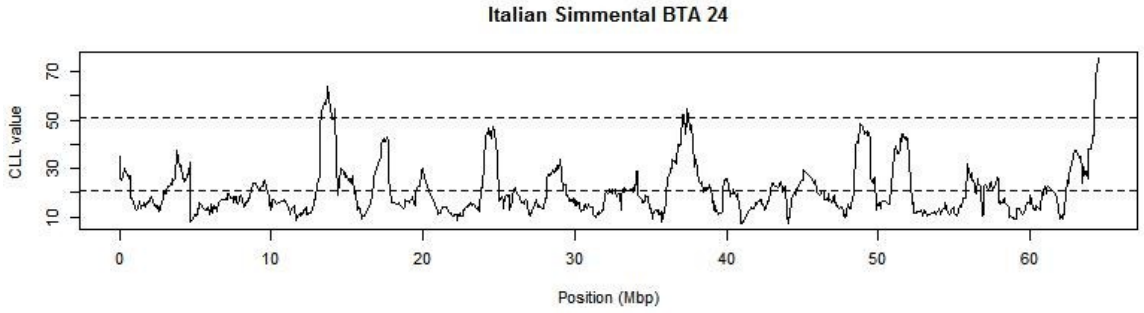

Italian Simmental BTA 23
CLL value
10
20
30
40
0
10
20
30
40
50
Position (Mbp)

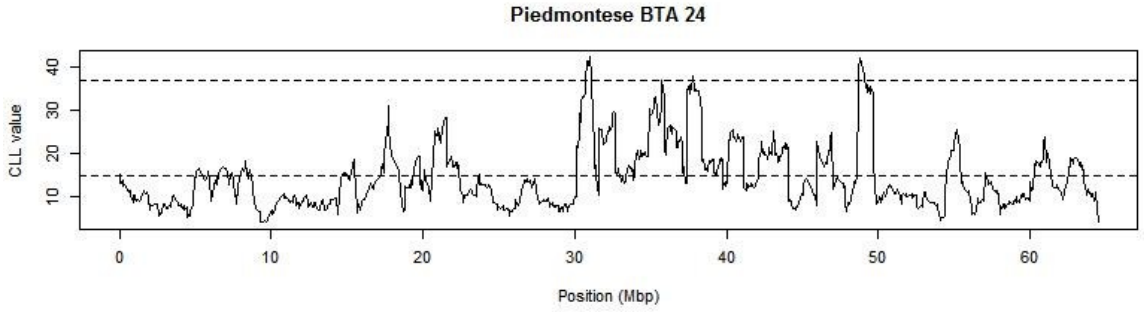

Piedmontese BTA 23
CLL value
10
20
30
40
0
10
20
30
40
50
Position (Mbp)

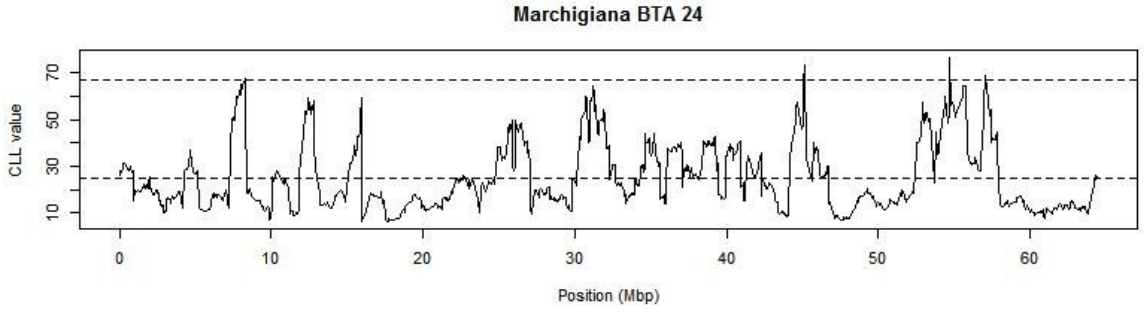

Marchigiana BTA 23
CLL value
10
30
50
0
10
20
30
40
50
Position (Mbp)

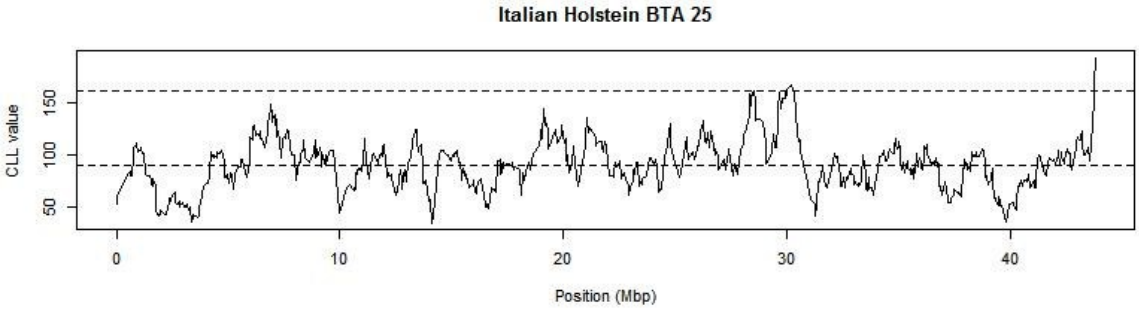

Italian Holstein BTA 24
CLL value
50
150
250
0
10
20
30
40
50
60
Position (Mbp)

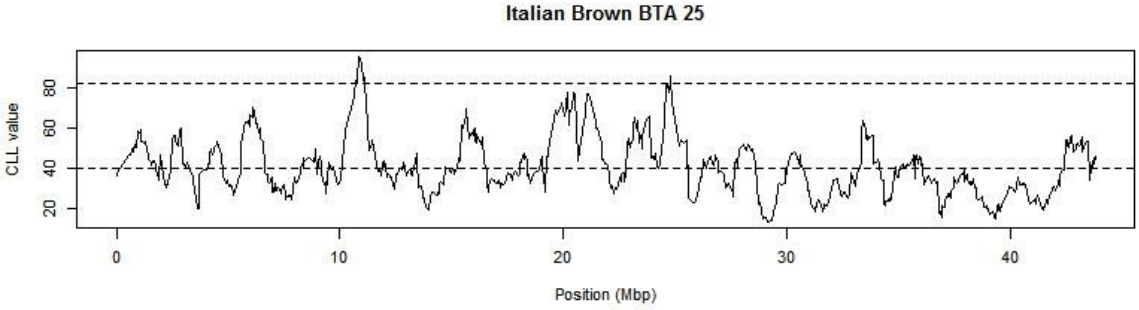

Italian Brown BTA 24
CLL value
20
60
100
140
0
10
20
30
40
50
60
Position (Mbp)

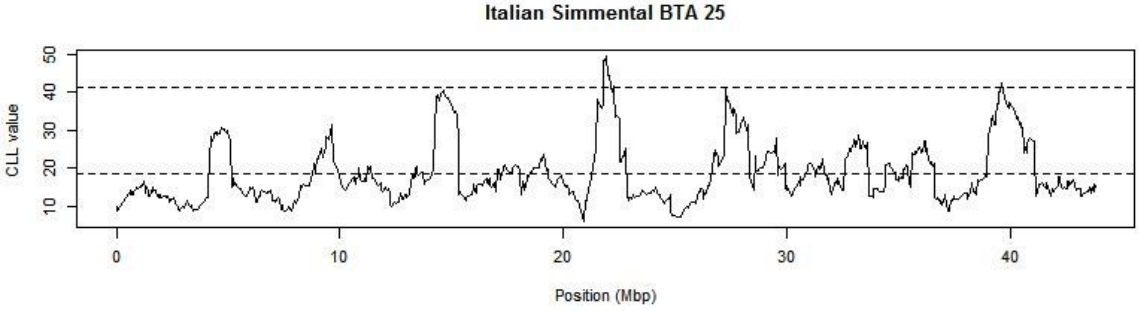

Italian Simmental BTA 24
CLL value
10
30
50
70
0
10
20
30
40
50
60
Position (Mbp)

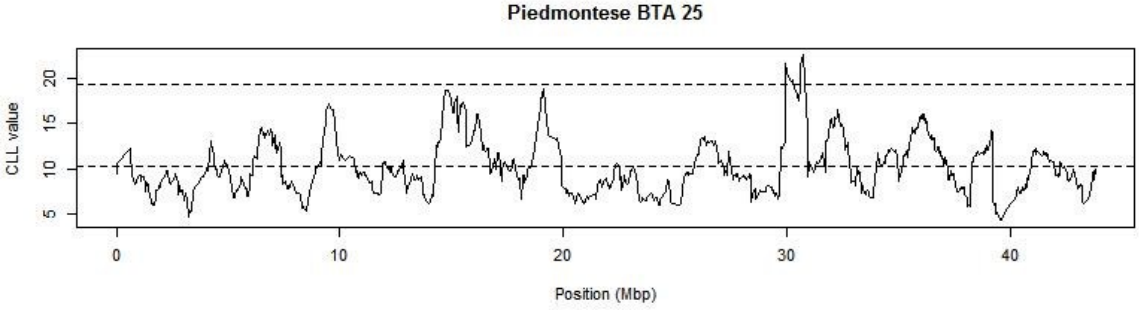

Piedmontese BTA 24
CLL value
10
20
30
40
0
10
20
30
40
50
60
Position (Mbp)

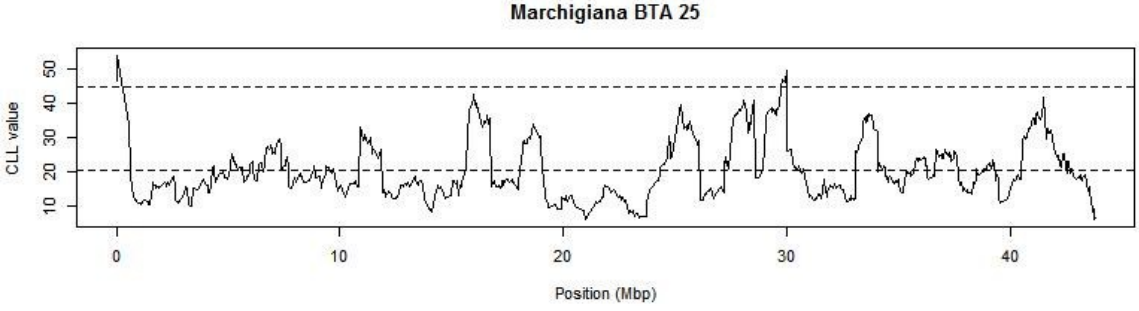

Marchigiana BTA 24
CLL value
10
30
50
70
0
10
20
30
40
50
60
Position (Mbp)

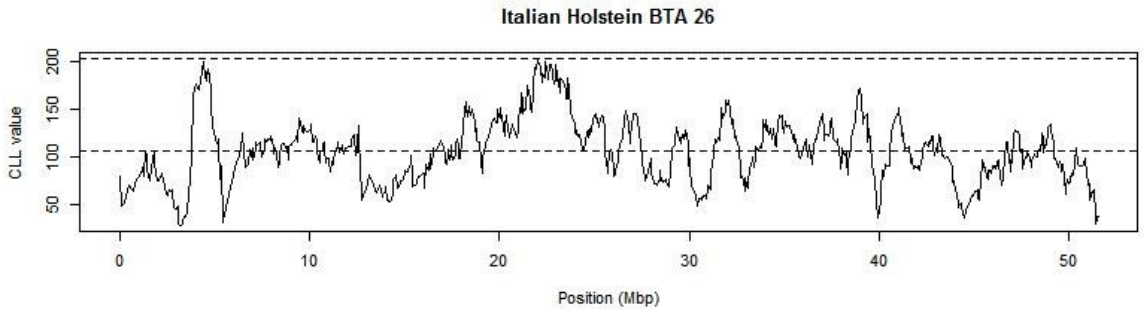

Italian Holstein BTA 25
CLL value
50
100
150
0
10
20
30
40
Position (Mbp)

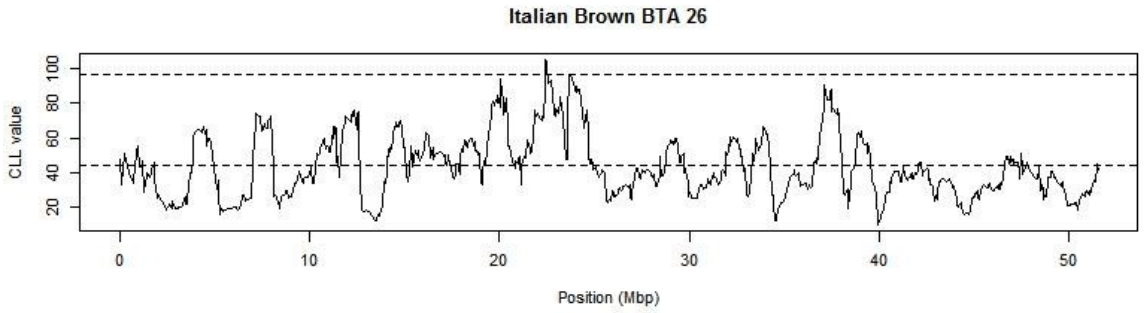

Italian Brown BTA 25
CLL value
20
40
60
80
0
10
20
30
40
Position (Mbp)

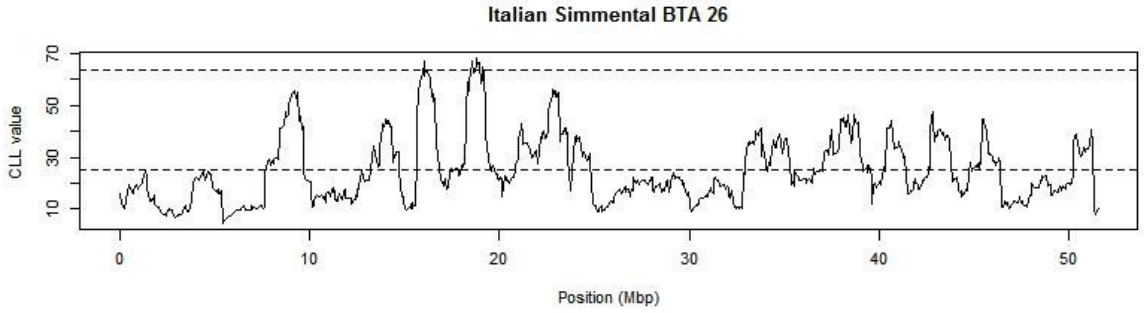

Italian Simmental BTA 25
CLL value
10
20
30
40
50
0
10
20
30
40
Position (Mbp)

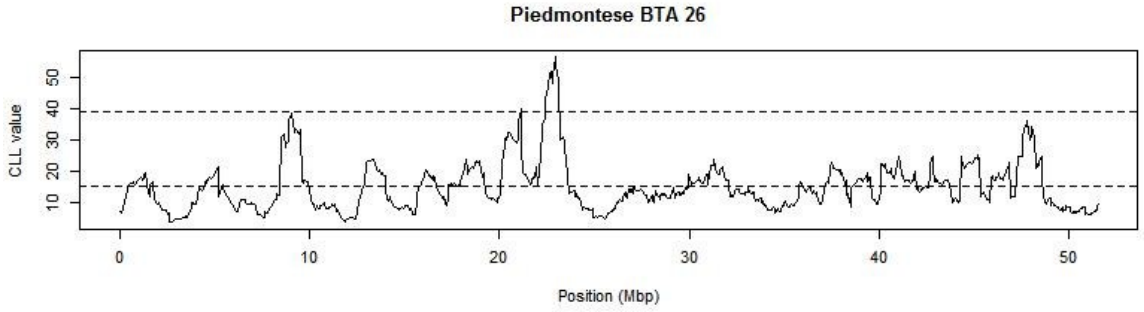

Piedmontese BTA 25
CLL value
5
10
15
20
0
10
20
30
40
Position (Mbp)

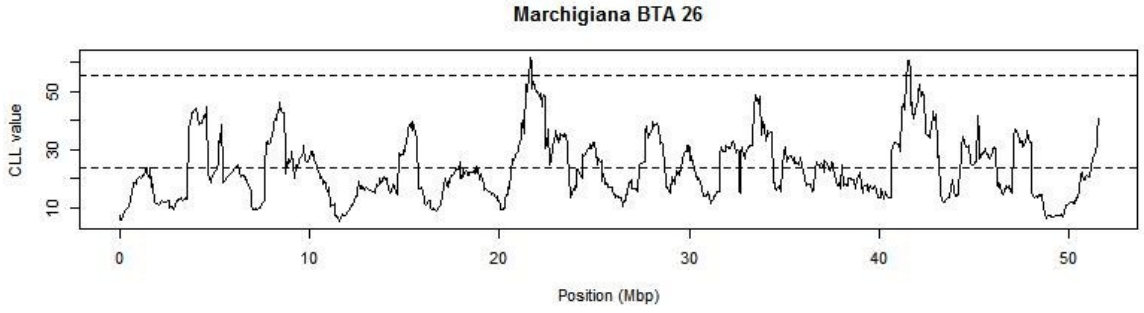

Marchigiana BTA 25
CLL value
10
20
30
40
50
0
10
20
30
40
Position (Mbp)

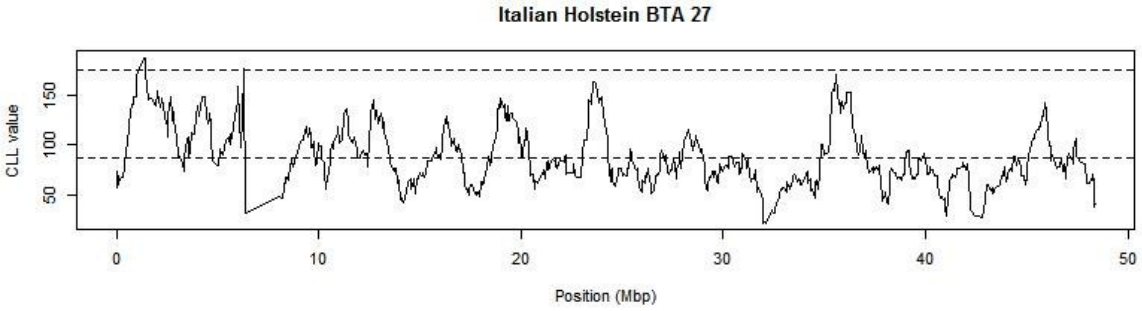

Italian Holstein BTA 26
CLL value
50
100
150
200
0
10
20
30
40
50
Position (Mbp)

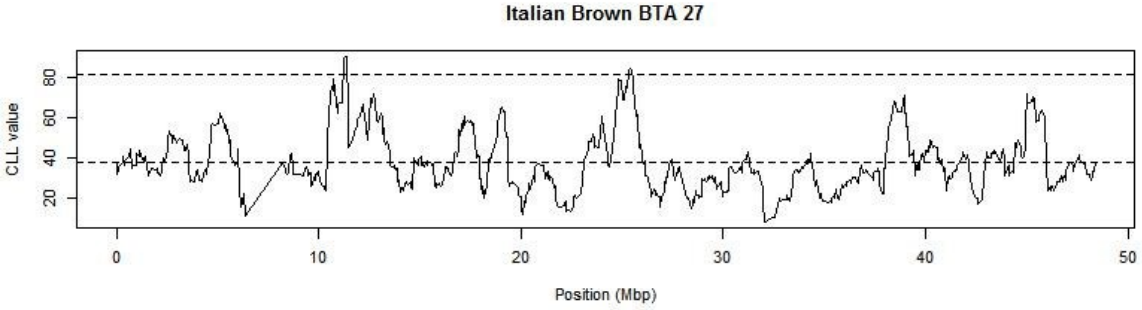

Italian Brown BTA 26
CLL value
20
40
60
80
100
0
10
20
30
40
50
Position (Mbp)

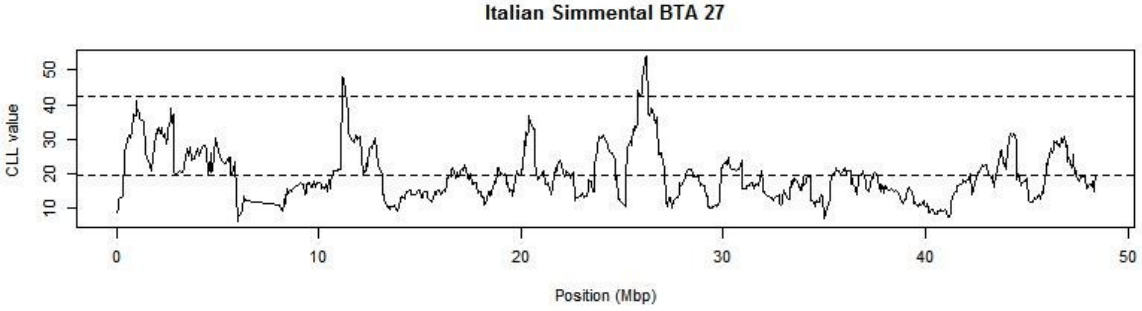

Italian Simmental BTA 26
CLL value
10
30
50
70
0
10
20
30
40
50
Position (Mbp)

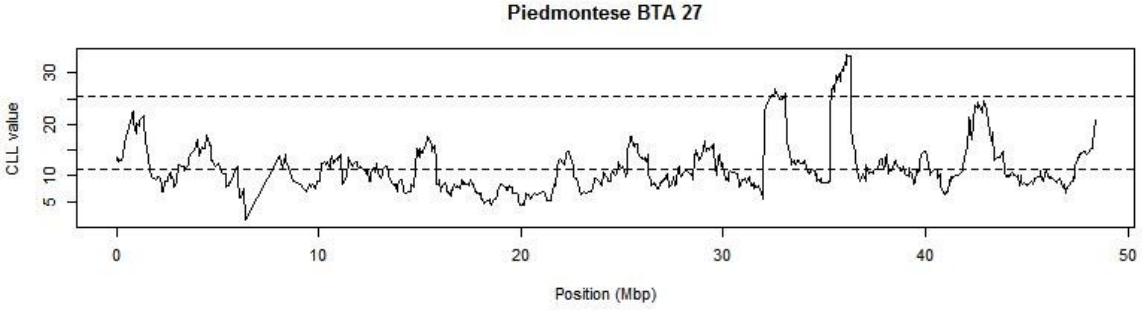

Piedmontese BTA 26
CLL value
10
20
30
40
50
0
10
20
30
40
50
Position (Mbp)

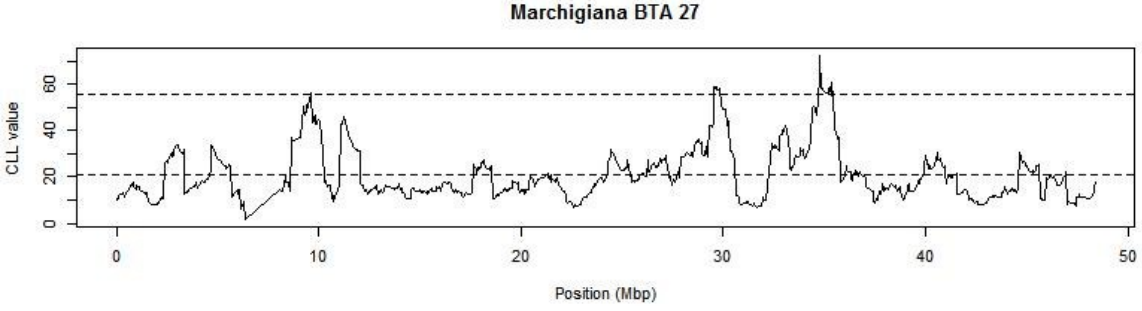

Marchigiana BTA 26
CLL value
10
30
50
0
10
20
30
40
50
Position (Mbp)

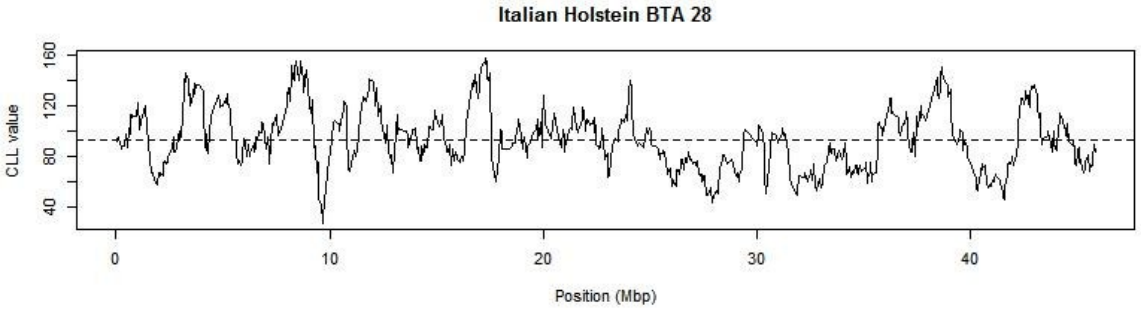

Italian Holstein BTA 27
CLL value
50
100
150
0
10
20
30
40
50
Position (Mbp)

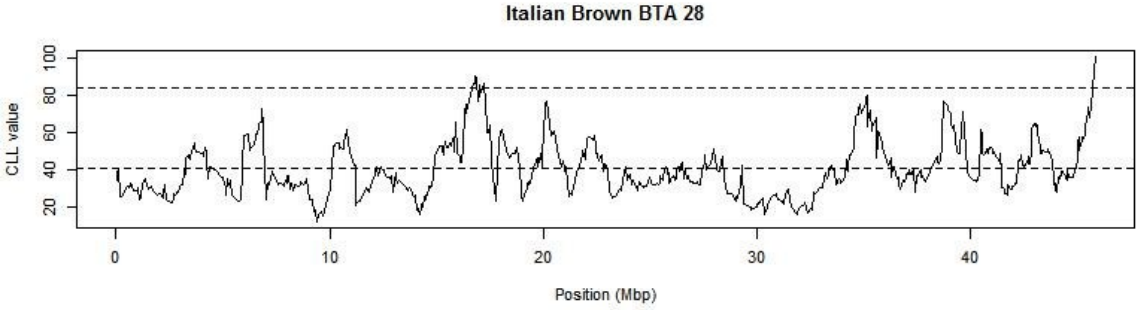

Italian Brown BTA 27
CLL value
20
40
60
80
0
10
20
30
40
50
Position (Mbp)

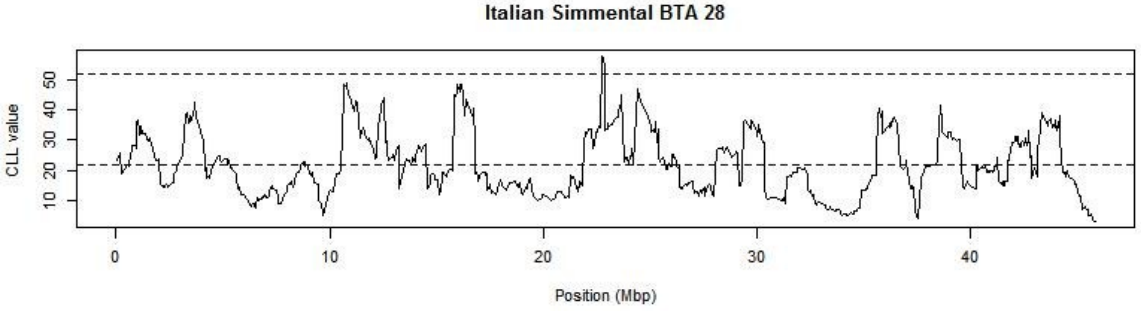

Italian Simmental BTA 27
CLL value
10
20
30
40
50
0
10
20
30
40
50
Position (Mbp)

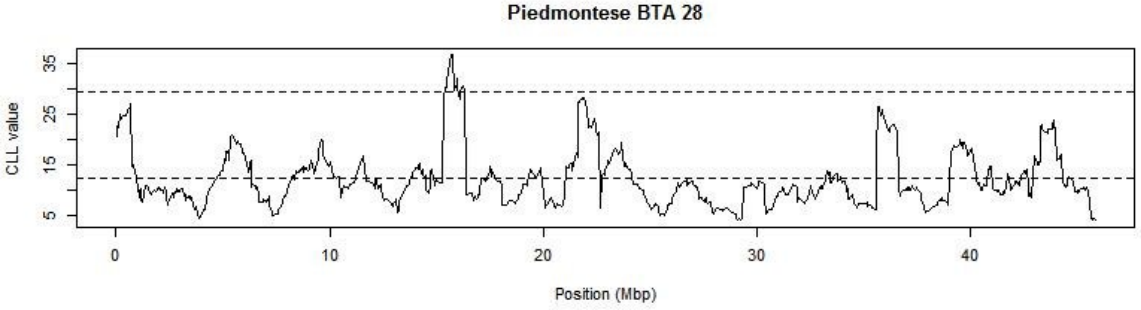

Piedmontese BTA 27
CLL value
5
10
20
30
0
10
20
30
40
50
Position (Mbp)

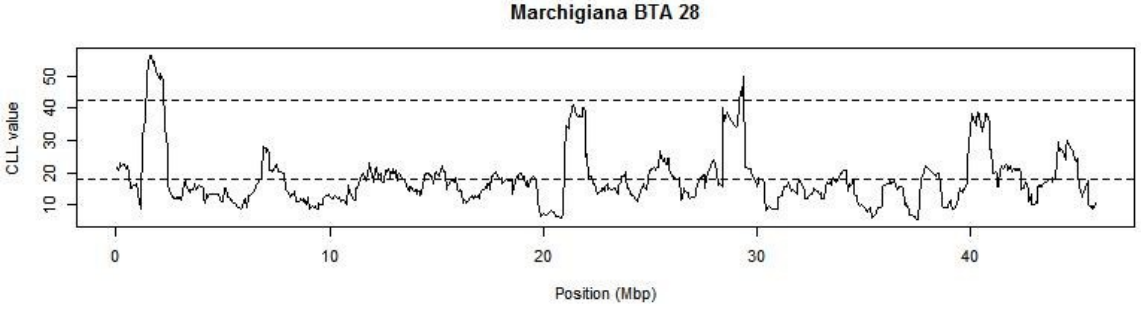

Marchigiana BTA 27
CLL value
0
20
40
60
0
10
20
30
40
50
Position (Mbp)

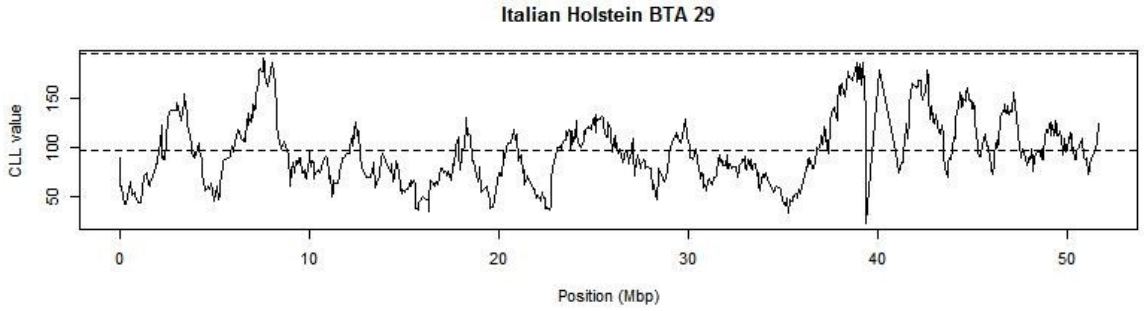

Italian Holstein BTA 28
CLL value
40
80
120
160
0
10
20
30
40
Position (Mbp)

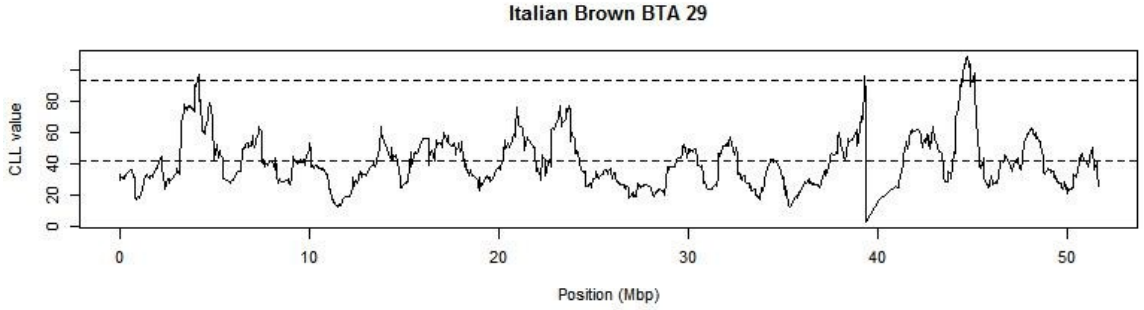

Italian Brown BTA 28
CLL value
20
40
60
80
100
0
10
20
30
40
Position (Mbp)

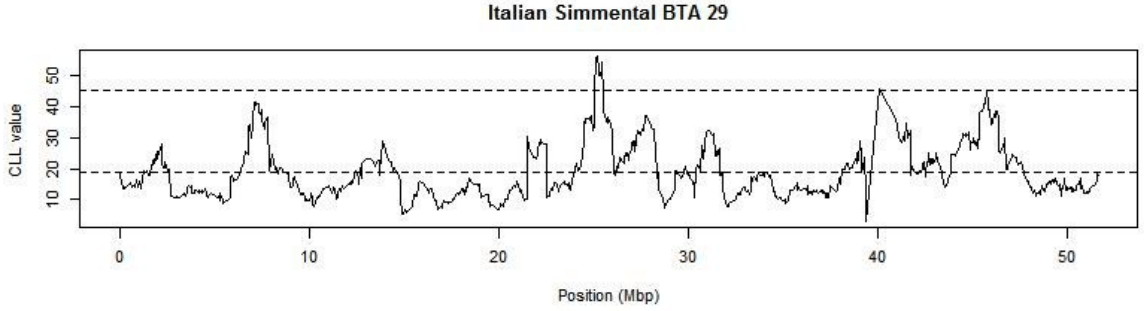

Italian Simmental BTA 28
CLL value
10
20
30
40
50
0
10
20
30
40
Position (Mbp)

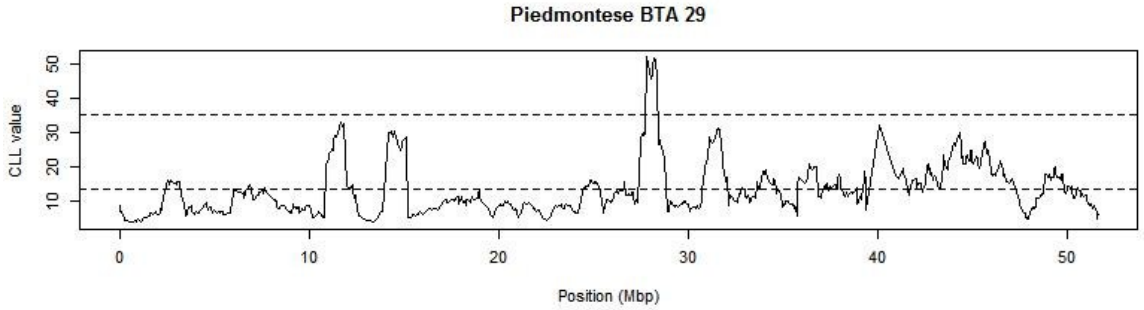

Piedmontese BTA 28
CLL value
5
15
25
35
0
10
20
30
40
Position (Mbp)

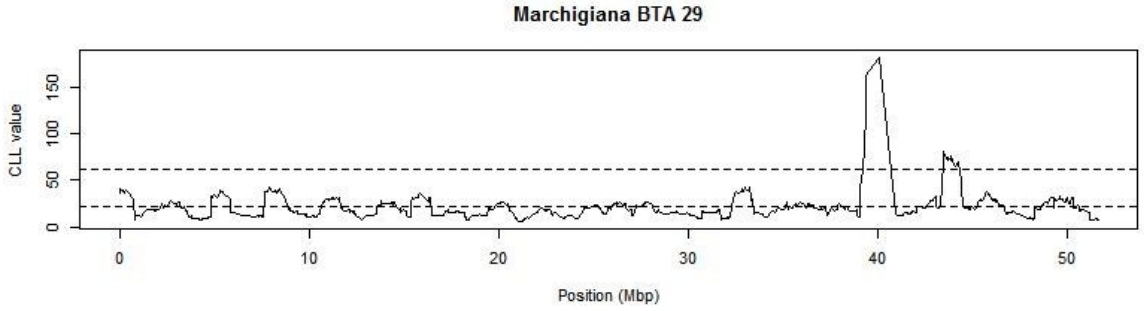

Marchigiana BTA 28
CLL value
10
20
30
40
50
0
10
20
30
40
Position (Mbp)


Italian Holstein BTA 29
CLL value
50
100
150
0
10
20
30
40
50
Position (Mbp)


Italian Brown BTA 29
CLL value
0
20
40
60
80
0
10
20
30
40
50
Position (Mbp)


Italian Simmental BTA 29
CLL value
10
20
30
40
50
0
10
20
30
40
50
Position (Mbp)


Piedmontese BTA 29
CLL value
10
20
30
40
50
0
10
20
30
40
50
Position (Mbp)


Marchigiana BTA 29
CLL value
0
50
100
150
0
10
20
30
40
50
Position (Mbp)

Acknowledgments

Vorrei ringraziare il prof. Macciotta per tutta la disponibilità e le conoscenze che mi ha dato in questi tre anni, e a tutto il gruppo di lavoro che mi ha sempre aiutato.

Un grazie anche al gruppo di bioinformatica di Lodi, in particolare la dott.ssa Stella, Ezequiel, Filippo e Nelson, per tutto l'aiuto che mi hanno dato.

Infine un grande grazie a Carla e alla mia famiglia per avermi sostenuto nei momenti di difficoltà.