IDENTIFICATION OF SIGNATURES OF SELECTION IN *BOS TAURUS* BEEF AND DAIRY CATTLE USING GENOME-WIDE SNP GENOTYPES

A Dissertation

by

JUNG WOO CHOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Animal Science

IDENTIFICATION OF SIGNATURES OF SELECTION IN *BOS TAURUS* BEEF AND DAIRY CATTLE USING GENOME-WIDE SNP GENOTYPES

A Dissertation

by

JUNG WOO CHOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Clare A. Gill
Committee Members,	James O. Sanders
	Andy D. Herring
	Konstantin V. Krutovsky
Head of Department,	Gary Acuff

August 2009

Major Subject: Animal Science

ABSTRACT

Identification of Signatures of Selection in *Bos taurus* Beef and Dairy Cattle Using Genome-wide SNP Genotypes. (August 2009)

Jung Woo Choi, B.S., Kang-Won National University

Chair of Advisory Committee: Dr. Clare A. Gill

The objectives of this study were to identify signatures of selection in *Bos taurus* beef and dairy cattle populations and to annotate regions of selection with gene, function and QTL information. Differences in minor allele frequencies, population-average F_{ST}, population-specific F_{ST}, and integrated extended haplotype homozygosity scores were applied to a subset of the bovine HapMap data to characterize signatures of selection in 7 *Bos taurus* beef and 5 *Bos taurus* dairy cattle populations.

Numerous single nucleotide polymorphisms (SNP) exhibited evidence of selection across the genome and regions of BTA2 and BTA14 that are considered to be under positive selection in beef and dairy cattle, respectively, were highlighted. The current density of SNP limited our ability to annotate regions putatively under selection because most SNP in the assay were intergenic. This is likely because of the between-breed SNP discovery method that was used, which typically identifies SNP with higher allele frequencies.

DEDICATION

To my parents,

Kyung-Hee Choi and Young-Boo Choi

ACKNOWLEDGEMENTS

First, I am deeply indebted to my committee chair, Dr. Clare Gill, for giving me the chance to get wonderful research experience and for her patience, knowledge, and dedication toward finishing my Ph.D. I appreciate Dr. Jim Sanders and Dr. Andy Herring for serving as mentors and, in particular, for opening my eyes to cattle breeds, the dynamics of animal breeding, and quantitative genetics theories. Especially, the memory of the tutoring I received from them and the discussion of cattle on the world map in my preliminary examination will be unforgettable. I would also like to thank Dr. Konstantin Krutovsky for his dedication in guiding me with his solid population genetics experience. Thanks to Dr. Penny Riggs for her encouragement and advice.

Thanks also go to all of the members of the Animal Genomics Laboratory: Colette Abbey, Sarah Vintanza, and Lauren Hulsman, for enjoyable communication about research and our lives. I would also like to thank Dr. Hsin-I Chiang for being a friend, exchanging perspective on genetics and encouraging me to survive in the Kleberg Center. Deep appreciation goes to Dr. C. Michael Dickens for great computational assistance and for taking my hand once when I was lost and alone, as a true friend would. My appreciation also extends to members in the Korean Aggie Soccer Association for helping to keep my mind and body healthy.

In the International Bovine HapMap Project, my research would have been almost impossible without cooperation of the project members. Therefore, my gratitude goes to: Dr. Rafael Villa-Angulo, Koni Wright, Dr. Carlos Bustamante, Dr. Lakshmi

Matukumalli, Dr. Curt Van Tassell and Dr. John Grefenstette for their valuable and sincere help in my research.

Thanks to my friends in Korea for their unchangeable encouragement. I deeply appreciate my fiancée, Sun-Young Lee for her patience and love and my family for their encouragement, love and sacrifices, with special mention of my lovely nephew Jong-Hun Lee.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

.

LIST OF TABLES

1. INTRODUCTION

One of the goals of bovine genomics research is to identify genes contributing to variation in economically important traits in cattle. The recently completed Bovine Genome Sequencing Project (The Bovine Genome Sequencing and Analysis Consortium, 2009) and International Bovine HapMap Project (The Bovine HapMap Consortium, 2009) generated resources that will accelerate progress towards identifying these genes. Genome-wide single nucleotide polymorphisms (**SNP**) will enable us to identify novel regions that are under selection in the bovine genome.

For more than 100 years, cattle breeders have been selecting for specific characteristics in beef and dairy cattle, which has resulted in substantial increases in productivity. For instance, genetic merit for milk yield in Holstein cows is improving at a rate of 83 liters (~1%) per year (AIPL, 2009). In the United States, kilograms of beef per animal harvested have increased by over 80% in 50 years (Elam and Preston, 2004). Whereas much of the improved performance is due to changes in management, some is due in part to selection of underlying genotypes, which in turn leaves signatures of selection in the genome.

The objectives of this study were to identify signatures of selection in *Bos taurus* beef and dairy cattle populations and to annotate regions of selection with gene, function and QTL information. Population genomic approaches were applied to a subset of the bovine HapMap data to characterize signatures of selection in 7 *Bos taurus* beef and 5 *Bos taurus* dairy cattle populations.

This dissertation follows the style of Journal of Animal Science.

2. LITERATURE REVIEW

According to the neutral theory proposed by Kimura (1968), most mutations are selectively neutral and are maintained or lost by random genetic drift in finite populations. For studies that aim to identify signatures of selection, this theory is important because it provides the null hypothesis that the sequence or genotype under consideration is expected to be neutral (Duret, 2008).

Both natural selection and artificial selection also can affect allele frequencies in populations. Artificial selection, which is of interest in this project, has been practiced intensively in cattle industries (Dekkers and Hospital, 2002). Artificial selection can be categorized as directional selection (positive or negative selection) and balancing selection. Positive selection results in an increase in frequency of alleles with higher fitness (or higher selective value in artificial selection) and these will eventually become fixed in the population. Negative (purifying) selection causes alleles that have deleterious effects (lower fitness or lower selective value) to become reduced in frequency and, theoretically, the process will continue until the allele is eliminated from the population. In the case of balancing selection, fixation of alleles does not occur, but instead the alleles tend to be maintained at intermediate frequencies in the population (Hurst, 2009).

There will also tend to be a concordant change in the frequency of alleles at linked loci in proximity to the locus under selection. As a result of positive selection, alleles at neutral loci that are strongly associated (i.e. in linkage disequilibrium) with the positive mutation will be maintained over time along with the alleles at the selected

locus. Variations that are associated with the disadvantageous allele will be removed over time, resulting in a selective sweep (Nielsen et al., 2007). Positive selection leaves patterns of alleles in the genome different from the pattern expected under a neutral model, and evidence of positive selection includes a skewed allele frequency distribution, lower genetic variation and elevated linkage disequilibrium (Biswas and Akey, 2006; Nielsen et al., 2007).

Population demographic history (e.g. population bottlenecks and migration) can also generate a skew in allele frequencies, cause reduced levels of genetic variation, and elevate the level of linkage disequilibrium (Biswas and Akey, 2006; Nielsen et al., 2007). When considering populations of domesticated animals, it is difficult to separate the effects of selection and population demographic history because often the two events are interrelated (e.g. selection of specific animals to form breeds creates a population bottleneck). Selection acts on specific loci within the genome, whereas population demographic history is a genome-wide force that affects all loci equally (Akey, 2009). Therefore, a challenge associated with the detection of signatures of selection is how to separate effects that are due to systematic selection from those due to population demographic history (Stinchcombe and Hoekstra, 2007).

Several different methods have been proposed to identify signatures of selection. One commonly used approach is to compare the levels of synonymous (dS, K_S) and non-synonymous substitutions (dN, K_A) in genes within and between species (Nei and Gojobori, 1986; Suzuki and Gojobori, 1999). The ratio expressed as dN/dS (or K_A/K_S) provides information on evolutionary forces affecting a protein-coding region. Under

neutrality, the ratio is expected to be 1, whereas for a positively selected gene, dN/dS > 1 and for a negatively selected gene, dN/dS < 1. However, this interpretation assumes the samples are from divergent populations. Kryazhimskiy and Plotkin (2008) demonstrated that the relationship does not hold if the sequences are drawn from a single population because dN/dS is less sensitive to selection and dN/dS < 1 can occur under both positive and negative selection.

In the Mouse (*Mus musculus*) sequencing project (Mouse Genome Sequencing Consortium, 2002), evidence for positive selection in protein coding genes was investigated using the K_A/K_S ratio on 12,845 orthologs between human and mouse. In the study, domain families with enzymatic activity had lower K_A/K_S ratio than non-enzymatic domains indicating fewer substitutions are tolerated in catalytic regions. Furthermore, there were higher values of K_A/K_S for a domain family of secreted proteins implicated in reproduction, host defense and immune response, indicative of positive selection (Mouse Genome Sequencing Consortium, 2002).

In a subsequent study in the Dog (*Canis familiaris*) sequencing project (Lindblad-Toh et al., 2005), evidence for positive selection in protein-coding genes across 3 mammalian orders was investigated using a subset of 4,950 genes from 13,816 human, mouse and dog orthologs for which there was either microarray data or other functional annotations. The number of the lineage-specific synonymous (K_S) and non-synonymous (K_A) substitutions was inferred and the K_A/K_S ratio was calculated. Sets of genes with elevated K_A/K_S ratio relative to the other lineages were identified. Overall, there were small deviations among the three lineages, but there was greater relative

variation in human-mouse and dog-mouse comparisons than in human-dog comparisons. Genes encoding subunits of mitochondrial electron transport chain complexes and genes expressed in the testis showed evidence of significantly accelerated evolution in humans relative to both mouse and dog (Lindblad-Toh et al., 2005; Ostrander and Wayne, 2005).

One of the disadvantages of using among species comparisons to identify regions affected by selection is that identification of orthologs is hindered by lineage-specific duplication and deletion events that are difficult to resolve in incomplete (or incompletely annotated) genomes (Roth et al., 2007). Because of this, there has been recent interest in other methodologies that are not reliant on protein data to identify signatures of selection.

Several approaches to detect signatures of selection are based on within species analysis of SNP data. Recently, researchers have begun to take advantage of the large numbers of SNP from genome sequencing projects (Mouse Genome Sequencing Consortium, 2002; Rat Genome Sequencing Project Consortium, 2004; Lindblad-Toh et al., 2005; The International HapMap Consortium, 2005, 2007). Examining numerous loci simultaneously throughout the genome to identify regions that are outliers is an appealing solution for detection of signatures of selection (Biswas and Akey, 2006; Stinchcombe and Hoekstra, 2007). Some of the methods that have been proposed include examining minor allele frequency (**MAF**) differences (Prasad et al., 2008; Hayes et al., 2009), Wright's population average F_{ST} (Wright, 1943a; Wright, 1943b), population-specific F_{ST} (Weir and Hill, 2002; Weir et al., 2005), and integrated extended haplotype homozygosity score (**iHS**) (Voight et al., 2006).

Examining differences in MAF between pairs of populations (e.g. breeds) is one simple way to identify SNP outliers that may be associated with selection. The basis for this approach is that selection will increase the frequency of favorable alleles characteristic of one population compared to the other. In this method, the allele that has the lowest frequency across both breeds is typically nominated as the minor allele, without regard to the ancestral allele state. To reduce some of the noise in the data due to differences in the age of SNP, averages from sliding windows (e.g. 5-SNP or 10-SNP) have been used (Prasad et al., 2008; Hayes et al., 2009). These MAF are then plotted by subtracting the mean for one breed from the other with respect to the genomic position or by plotting the absolute value of MAF. Whether these values differ significantly from zero is determined by permutation (Prasad et al., 2008) or simulation (Hayes et al., 2009).

Wright (1943a, b) proposed the fixation index ($\mathbf{F_{ST}}$) to measure the degree of population differentiation due to random genetic drift and the inbreeding effect in a subpopulation. The fixation index has been widely used to detect loci subjected to selection. The basic theory is that positive selection tends to reduce the heterozygosity of specific loci in a population. If positive selection increases the frequency of an allele in just one population (e.g. one breed), a higher fraction of the variation will be noticed in comparisons between populations than within a population. Akey et al. (2002) identified genomic regions that are targets of selection in humans by calculating population-average F_{ST} for each locus. By examining the distribution of the population-average F_{ST} values mapped to gene-associated regions, they identified 174 candidate

genes affected by selection: 156 had high values of F_{ST} indicative of divergent selection and 18 had low F_{ST} values indicative of balancing selection (Akey et al., 2002).

Weir et al. (2005) demonstrated that population-specific F_{ST} estimates can be better indicators of selection because when F_{ST} values are averaged across populations they can mask past evolutionary events. Population-specific F_{ST} was applied to the human Perlegen and the Phase I HapMap data sets. In particular, the sensitivity of this approach was demonstrated in a study that identified an unusually long haplotype indicative of strong selection surrounding the lactase (*LCT*) gene on HSA2 (Bersaglieri et al., 2004). The ability to digest milk as an adult is a phenotype predominant in Caucasians of European descent but absent in many other populations. Population-specific F_{ST} revealed that there was a strong signature of recent positive selection in *LCT* for Caucasians of European descent and European Americans, whereas population-average F_{ST} did not distinguish this region from others (Weir et al., 2005).

It is also possible to detect selective sweeps that are still in progress. Voight et al. (2006) developed iHS to detect loci where selection has driven new (derived) alleles to intermediate frequencies in the population. Using this approach, the first genome-wide map of incomplete selective sweeps in humans was produced (Voight et al., 2006). As in the study of Bersaglieri et al. (2004) using population-specific F_{ST}, extreme iHS values were found in the region of *LCT* in European samples, reinforcing that this gene is a target of selection in Caucasians. Although iHS provided a profile of widespread selective incomplete sweeps throughout the human genome, it could not detect alleles that were approaching fixation or that were already fixed.

There have been several recent efforts to identify signatures of selection in the bovine genome using SNP from the bovine genome sequencing and HapMap projects (The Bovine Genome Sequencing and Analysis Consortium, 2009; The Bovine HapMap Consortium, 2009). Some of these studies have been restricted to the characterization of individual chromosomes and specific breeds (Angus, Holstein, and Norwegian Red) rather than an analysis of the entire genome. Prasad et al. (2008) used SNP markers for BTA 19 and 29 and calculated rolling average minor allele frequencies using a 5-marker sliding window to identify SNP outliers indicative of selection in Angus and Holstein. Some of the regions that were identified corresponded to QTL for marbling, structural soundness, and milk fat.

In an investigation of BTA6 in Norwegian Red cattle, evidence of positive selection was detected by the iHS method with a cluster of SNP representing a partial selective sweep at the distal end of BTA6. This region coincides with QTL affecting milk yield, protein yield, and protein percentage.

Genome-wide analyses in cattle have used population-average F_{ST} (Barendse et al., 2009; The Bovine HapMap Consortium, 2009), the allele frequency difference approach (Hayes et al., 2009), and iHS (The Bovine HapMap Consortium, 2009; Hayes et al., 2009) to detect genes that are fixed or approaching fixation.

3. MATERIALS AND METHODS

3.1 Genotype Data

The Bovine HapMap Consortium (2009) genotyped 37,470 SNP in 497 cattle from 19 geographically and biologically distinct *Bos taurus* and *Bos indicus* cattle breeds. Two assays of 22,608 ('25K' set) and 11,790 SNP ('11.5K' set) were genotyped on the Affymetrix platform and 2 assays each containing 1,536 SNP were genotyped on the Illumina GoldenGate platform. The majority of the SNP were derived by comparison of the draft sequence from a Hereford cow to skim sequence from 6 additional breeds (Angus, Limousin, Jersey, Norwegian Red, Holstein, and Brahman). More than 4,500 SNP in the 11.5K set were derived by comparison of the draft Hereford sequence to sequence generated from a series of bacterial artificial chromosomes constructed from Holstein DNA that represented regions from BTA 6, 14, and 25. A pair of samples from 2 outgroups (*Bubalus quarlesi* and *Bubalus bubalis*) were genotyped using the 25K SNP set.

Genotypes were downloaded from the Bovine HapMap database (http://bfgl.anri.barc.usda.gov) and data for the 12 *Bos taurus* beef and dairy breeds sampled were extracted using a Perl script. This subset of the HapMap data included genotypes for 331 animals representing 7 *Bos taurus* beef (Angus, Hereford, Red Angus, Charolais, Limousin, Romagnola, and Piedmontese) and 5 *Bos taurus* dairy (Norwegian Red, Brown Swiss, Guernsey, Jersey, and Holstein) breeds. Animals were chosen to be as unrelated as possible based on 5-generation pedigrees, except that there was at least one trio (sire, dam and calf) sampled for each of the breeds.

3.1.1 Filtering Genotypes. These data (~11.2 million genotypes) were filtered using Perl scripts to remove any monomorphic markers, markers or animals with poor completion rates (<90%), and markers that were discordant in multiple trios. Furthermore, markers that violated Hardy Weinberg equilibrium (**HWE**) proportions ($P < 0.05$) in multiple breeds, indicative of genotyping errors, were removed. The offspring of trios (n = 34) also were removed to avoid overrepresentation of the haplotypes they received from their parents in the final dataset. Finally, markers assigned to the X chromosome or to unassigned scaffolds (Chr. Un) were filtered.

3.2 Calculating Estimators to Detect Signatures of Selection

Three population genomic approaches were applied to this subset of the bovine HapMap data to characterize signatures of selection in *Bos taurus* beef and dairy populations. For each of the approaches, values were plotted with respect to the SNP coordinates on the autosomes from the draft assembly of the bovine genome sequence (BTAu4.0).

3.2.1 Differences in Minor Allele Frequency. The first approach was to investigate minor allele frequency (**MAF**) differences that may be shaped by selection that has been practiced on beef and dairy cattle populations, respectively. Minor allele frequency of each SNP for each of the 12 beef and dairy breeds was calculated using a Perl script. The allele that had the lowest combined frequency across breeds was designated the minor allele. This meant that in some cases the same allele was a minor allele in one breed but the common allele in another breed. For each SNP, the MAF for the 7 beef breeds was averaged and compared to the average MAF of the 5 dairy breeds

to identify differences. To establish whether the average allele frequency differences between beef and dairy cattle differed significantly from zero, permutation tests by chromosome to identify the 5% significance level were performed as in Prasad et al. (2008).

3.2.2 Wright's F_{ST}. The second approach was to estimate Wright's F_{ST} index using both population-average (Wright, 1943a; Wright, 1943b) and population-specific (Weir et al., 2005) measures for F_{ST}. For population-average F_{ST}, the following equation was used for each SNP:

$$F_{ST} = \frac{(H_T - H_S)}{H_T}$$

where H_S is the average expected heterozygosity assuming HWE among organisms within a random mating subpopulation (breed), and H_T is the average expected heterozygosity assuming HWE among organisms within the total population. These values were interpreted using the qualitative guidelines proposed by Wright (1978) where $F_{ST} > 0.25$ means very great differentiation, 0.15 to 0.25 means great differentiation, 0.05 to 0.15 means moderate differentiation, and $F_{ST} < 0.05$ means little differentiation among the populations.

Weir et al. (2005) using human SNP data demonstrated that the distribution of F_{ST} values calculated for individual SNP tends to approximate χ^2, whereas averaging F_{ST} values for SNP across a 5Mb window better approximates a normal distribution. Because the bovine dataset has a lower SNP density than human data, various window sizes for averaging were investigated. The distribution of individual F_{ST} values as well

as overlapping 1Mb, 5Mb, and 10Mb averages, plus non-overlapping 8 SNP averages were determined. Each overlap was centered on a SNP so there were as many windows generated as there were SNP.

Population-specific F_{ST} (βi) was estimated for each of the 12 breeds using the following equation from Weir and Hill (2002):

$$\beta_i = \frac{\theta_i - \theta_A}{1 - \theta_A} = 1 - \frac{\left(\sum_{i=1}^{r} n_{ic}\right)\sum_{u=1}^{m} \frac{n_i}{n_i - 1} \tilde{p}_{iu}(1 - \tilde{p}_{iu})}{\sum_{u=1}^{m}\sum_{i=1}^{r}[n_i(\tilde{p}_{iu} - \bar{p}_u)^2 + n_{ic}\,\tilde{p}_{iu}(1 - \tilde{p}_{iu})]}$$

where θ_i is the average within-population coancestry, θ_A is the average between-population-pair coancestry, r is the number of breeds, m indicates the number of alleles for a locus, n_i are the alleles sampled from the ith population, $\tilde{p}_{iu}$ is the frequency of the u allele in the ith population, $\bar{p}_{iu}$ is the average allele frequency weighted for sample size, and $n_{ic} = n_i - n_i^2 / \sum_{i=1}^{r} n_i$.

3.2.3 Integrated Extended Haplotype Homozygosity Score. The third approach was to apply iHS (Voight et al., 2006), which provides a measure of recent positive selection based on the decay of extended haplotype homozygosity as a function of distance. The iHS computing tool (kindly provided by Jonathon Pritchard and William Wen) was used to calculate unstandardized iHS for each breed. The program requires 3 parameters: estimated haplotypes from fastPhase, estimated recombination rate (Rho) and ancestral allele state. Resolved haplotypes were provided by John Grefenstette and Rafael Villa-Angulo (George Mason University). Population scaled estimates of recombination rates generated for each breed using overlapping windows of 10 Mb

along each chromosome by the composite likelihood method implemented in the program "pairwise" from the LDHat package (McVean et al., 2002), were provided by Carlos Bustamante and Koni Wright (Cornell University). Genotypes from 2 individuals from 2 outgroups (*Bubalus quarlesi* and *Bubalus bubalis*) were used to assign ancestral allele state using a Perl script. Ancestral allele state was determined for SNP that amplified in Anoa or Water buffalo and was defined as the allele that was homozygous in both species or in one or other species. Any markers that segregated in one species or that were fixed for alternate alleles in the Anoa and Water Buffalo samples were not considered. In practice, the Perl script considered whether the allele frequency differed from 0.5 and then assigned the allele with the highest frequency as the ancestral allele. This approach failed to eliminate 27 segregating markers that only amplified in Buffalo, 75 markers that only amplified in Anoa, and 94 markers that were segregating at different frequencies in both species, and these were filtered manually.

The iHS computing tool uses estimated haplotypes, estimated recombination rate and ancestral allele state to compute unstandardized iHS values:

$$\text{unstandardized iHS} = \ln\left(\frac{iHH_A}{iHH_D}\right)$$

where iHH_A and iHH_D refer to the integrated extended haplotype homozygosity score (EHH; Sabeti et al., 2002) for the ancestral and derived alleles, respectively. To adjust for the age of the SNP, the iHS values were standardized as in Voight et al. (2006) to obtain a final statistic with mean 0 and variance 1, regardless of the allele frequency of the SNP:

$$iHS = \frac{\ln\left(\frac{iHH_A}{iHH_D}\right) - E_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}{SD_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}$$

where the expectation and standard deviation are estimated from the empirical distribution at SNPs whose allele frequency *p* matches the frequency at the core SNP. This was done recursively for each breed and each chromosome using a series of Perl scripts.

3.3 Annotation

Coordinates identifying untranslated regions (**UTR**), introns and exons for the bovine RefSeq and GLEAN gene prediction sets were obtained from the University of California-Santa Cruz genome browser (http://genome.ucsc.edu/) and from the bovine genome database coordinator, respectively. Markers were classified according to their functional category (coding, intronic, UTR, and noncoding) for each of the gene prediction sets and then MAF differences, F_{ST} and |iHS| values were averaged with respect to each category, as in Akey et al. (2002). Associations between functional category and the values obtained for each of the population genomic methods for markers in regions that were identified as outliers also were evaluated. These markers were further interrogated by comparing them to QTL for beef and dairy traits (Polineni et al., 2006). Finally, to better understand potential functions of exonic SNP identified in signatures of selection, gene ontogeny (GO) terms that describe the biological process, cellular components and molecular functions for the associated gene were assigned (Ashburner et al., 2000).

4. RESULTS AND DISCUSSION

4.1 Summary of Filtered Genotypes

Of the markers that were genotyped in the HapMap project, 93.7% produced data that were released to the consortium (Figure 4.1). Some markers had been pre-filtered by the database coordinator because of quality-control issues. There were an additional 27 markers that were removed for poor completion rate, 3 that were monomorphic, 2,804 with MAF < 0.05, 22 where multiple trios were discordant, 384 that violated HWE, 268 with completion rates <90%, and 2,247 markers assigned to the X chromosome or Chr. Un. This filtering process left 29,131 markers and 8,174,204 genotypes for analysis. Genotypes analyzed in this study correspond to 25,332 SNP from genome-wide assays and 3,799 SNP from densely sampled regions of BTA 6, 14, and 25. Approximately 2.5 Gb of the genome was represented in this dataset with an average intermarker spacing of ~100 kb based on genome-wide markers.

4.2 Differences in Minor Allele Frequency

In this study, the allele with the lowest frequency across all breeds was designated the minor allele and then averages were obtained for beef and dairy. This differs from previous studies, which only considered pairwise breed comparisons (Prasad et al., 2008; Hayes et al., 2009). The range of differences (beef – dairy) in MAF was from -0.375 to 0.314 (Figure 4.2). When 0.5 Mb sliding windows were used, the range of differences in MAF was from -0.212 to 0.239 (Figure 4.2, Table 4.1 and Appendix A).

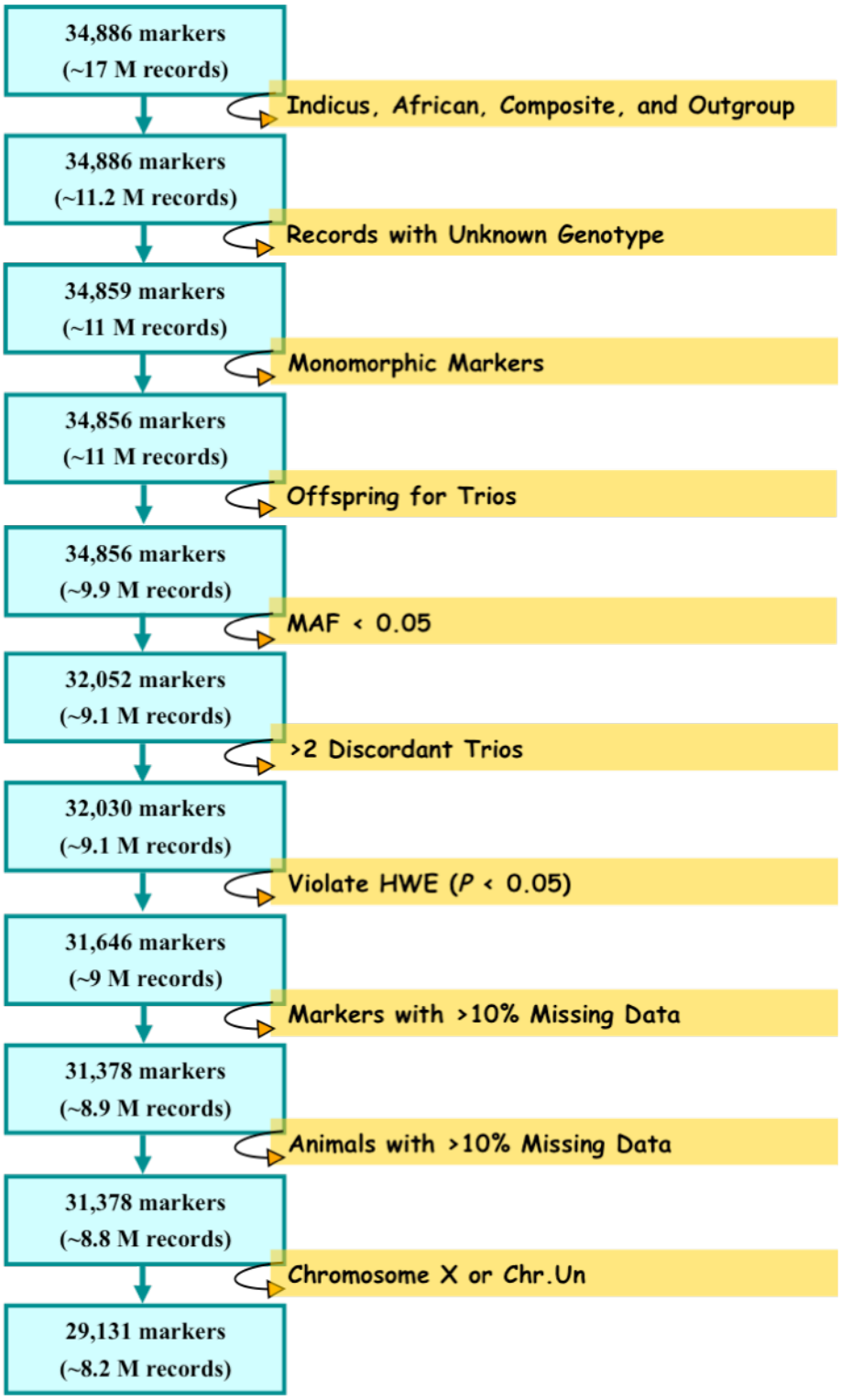


Figure 4.1. Flow chart of genotype filtering procedure.

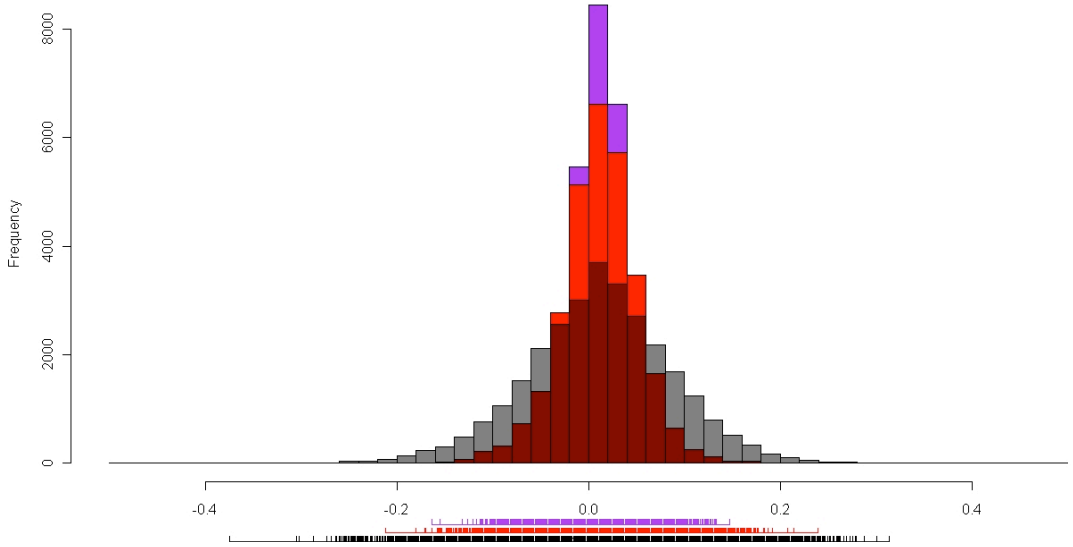


Figure 4.2. Histogram of the empirical differences in minor allele frequency between beef and dairy. Black indicates differences calculated for individual SNP, red is the distribution for 0.5 Mb overlapping sliding windows, and purple is 1 Mb overlapping sliding windows. Negative numbers indicate that the dairy average MAF was higher than beef. Positive numbers indicate that the beef average MAF was higher than dairy.

Table 4.1. Summary statistics for differences in minor allele frequency between beef and dairy based on 0.5 Mb overlapping sliding windows

					Permutation Thresholds[1]		MAF Difference[2]	
BTA	Mean	Std. Dev.	Windows	SNP per Window	5th Percentile	95th Percentile	No. Sign. Negative	No. Sign. Positive
1	0.017	0.040	1,554	7.44	-0.136	0.171	0	0
2	0.012	0.042	1,446	7.73	-0.153	0.172	0	0
3	0.009	0.041	1,278	7.46	-0.131	0.147	3	3
4	0.009	0.041	1,215	7.63	-0.135	0.152	3	2
5	0.008	0.048	1,214	7.10	-0.135	0.154	7	12
6	0.019	0.034	2,367	34.92	-0.090	0.132	12	5
7	0.012	0.040	1,034	7.15	-0.139	0.164	0	0
8	0.012	0.041	1,154	7.44	-0.131	0.156	0	2
9	0.014	0.042	963	6.83	-0.142	0.169	2	2
10	0.005	0.038	1,045	7.89	-0.141	0.149	0	2
11	0.021	0.041	1,159	7.84	-0.117	0.153	3	1
12	0.010	0.036	837	7.24	-0.140	0.159	0	0
13	0.015	0.037	923	8.51	-0.124	0.158	0	2
14	-0.003	0.044	2,662	64.01	-0.114	0.106	46	10
15	0.012	0.039	770	7.17	-0.143	0.163	2	1
16	0.017	0.042	793	8.28	-0.119	0.152	3	5
17	0.011	0.040	794	7.60	-0.134	0.159	1	4
18	0.012	0.044	653	7.95	-0.141	0.167	0	0
19	0.014	0.037	683	7.78	-0.130	0.154	1	0
20	0.018	0.039	807	7.88	-0.122	0.158	1	0
21	0.011	0.036	662	7.59	-0.143	0.163	0	2
22	0.017	0.039	642	7.88	-0.131	0.159	0	0
23	0.004	0.039	592	7.77	-0.144	0.151	1	1
24	0.017	0.042	678	7.32	-0.141	0.164	1	1
25	0.009	0.029	1,233	38.79	-0.088	0.109	0	5
26	0.012	0.040	532	7.73	-0.111	0.141	1	1
27	0.019	0.038	457	7.20	-0.130	0.172	0	0
28	0.007	0.031	486	7.82	-0.124	0.135	1	0
29	0.016	0.037	490	7.68	-0.142	0.167	0	0
All	0.012	0.039	29,123	7.44	-0.130	0.154	61	88

[1]Upper and lower thresholds based on 1,000 permutations of the data.
[2]Number of markers with significant differences in minor allele frequency between beef and dairy.

These windows corresponded to an average of 7.6 markers per 0.5 Mb except for BTA6, 14 and 25, which had 34.9, 64.0, and 38.8 markers per 0.5 Mb, respectively. Based on the genome-wide SNP assay, the average falls between the 5-marker windows used by Prasad et al. (2008), and the 10-marker windows used by Hayes et al. (2009). Use of 1 Mb sliding windows (12.7 markers per window) was observed to reduce too much of the variability about the mean (Figure 4.2). The range in differences in MAF was narrower than in the 2 previously reported studies (Prasad et al., 2008; Hayes et al., 2009), probably because an average of several breeds was used. Averaging can mask differences within beef and within dairy such as when the minor allele for a breed is opposite to what was designated across the breed type. Furthermore, averaging this average across markers further reduced the observed variability. For example on BTA6, in a comparison of the differences in MAF between Angus and Holstein there were 125 significant markers, whereas for the comparison between beef and dairy there were only 17 significant markers (Figure 4.3). The profile of differences in MAF generated for Angus and Holstein is similar to that reported by Hayes et al. (2009) for a subset of these data.

Markers that were significantly different between beef and dairy may be indicative of regions that affect milk or meat production. For example, the cluster of SNP that have higher MAF in dairy than in beef at ~6.2 Mb on BTA6 are associated with phosphodiesterase 5A. This gene encodes a cGMP-binding, cGMP-specific phosphodiesterase that is involved in the regulation of intracellular concentrations of

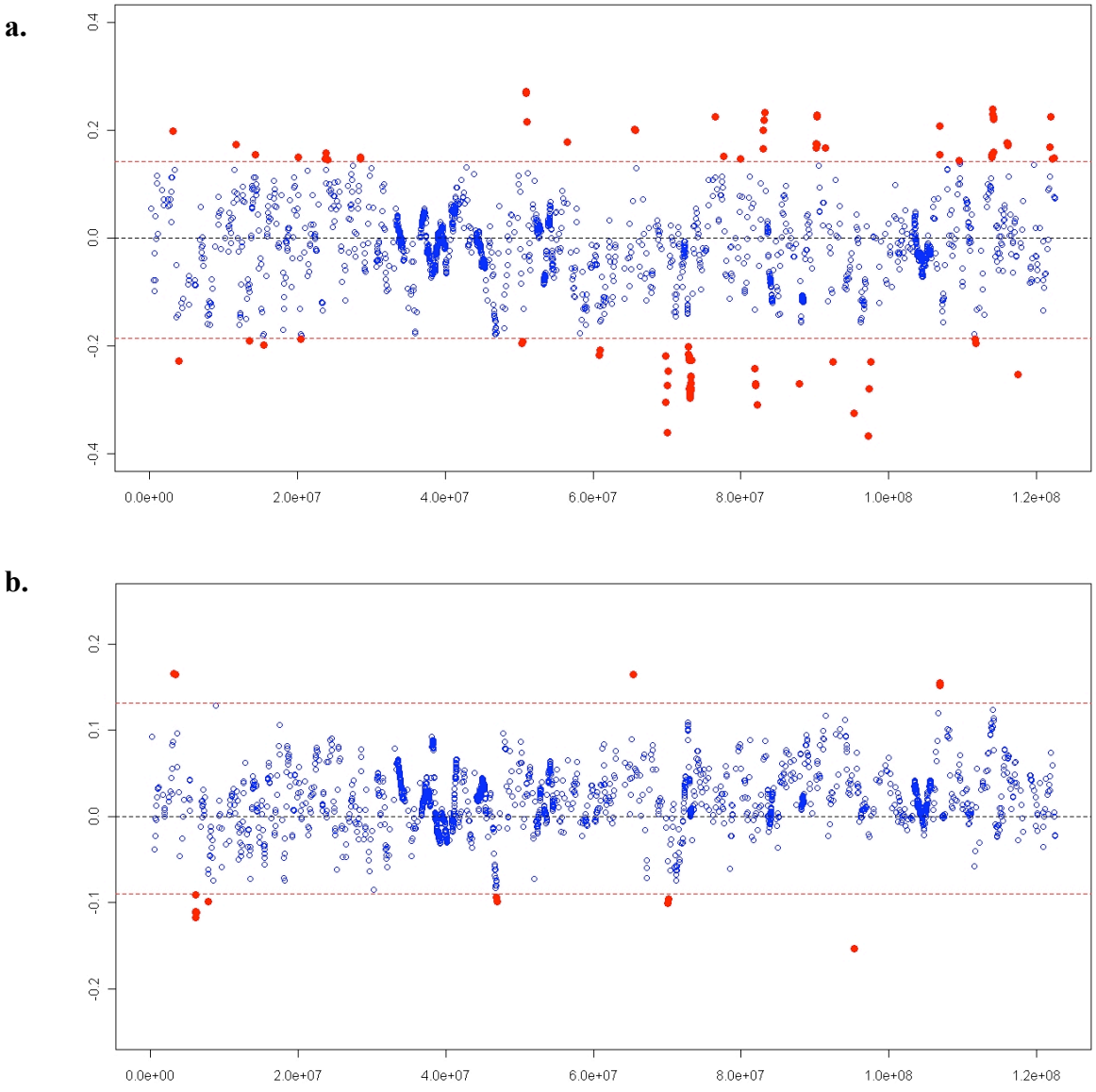


Figure 4.3. Differences in minor allele frequency across BTA6 based on 0.5 Mb overlapping sliding windows. (a) Difference in MAF between Angus and Holstein, (b) Difference in MAF between beef and Dairy. Horizontal dotted lines indicate the upper and lower 5% thresholds determined by 1,000 permutations of the data. Markers in red are significantly different between breeds or breed types. Markers are plotted based on coordinates from build Btau4.0 of the bovine genome sequence.

cyclic nucleotides and is important for smooth muscle relaxation in the cardiovascular system. There is no obvious role for this gene in milk production. However, this region does correspond to a QTL for milk yield, fat percentage and protein percentage and further investigation of PDE5A as a candidate gene affecting this trait is therefore warranted. Additional regions were significant in the comparison between Angus and Holstein and this suggests that it may be necessary to perform all pairwise breed comparisons to better evaluate differences in MAF, rather than averaging across breeds.

4.3 Population-Average F_{ST}

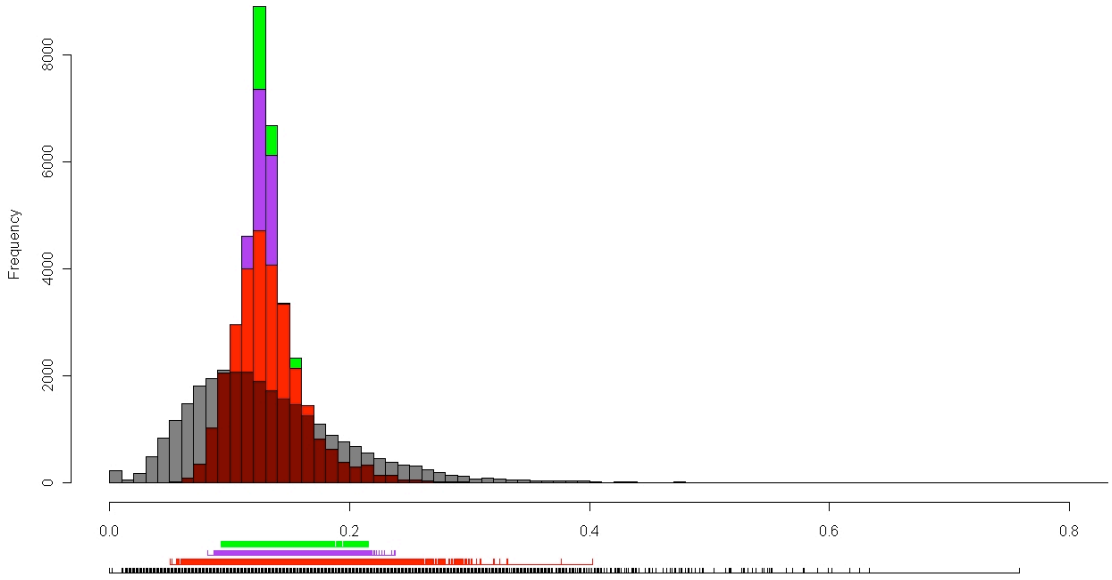
Wright's F_{ST} (Wright 1943a, b) was calculated for every marker with MAF > 0.05 in at least one breed (Table 4.2). This is similar to the approach of Akey et al. (2002) for human SNP data. The average value of F_{ST} was 0.131 across the autosomes when only the genome-wide SNP were considered. The mean F_{ST} was not affected by the addition of the densely sampled SNP on BTA 6, 14 and 25. This average is similar to the mean F_{ST} of 0.123 that was observed in humans using 25,549 autosomal SNP (Akey et al., 2002). These data approximated a χ^2 distribution (Figure 4.4). Large standard deviations were associated with the mean value of F_{ST} for each chromosome because there is substantial variation in F_{ST} values throughout the genome, even for closely associated markers (Weir et al., 2005). Only 0.80% of markers had $F_{ST} = 0$ and 0.42% had $F_{ST} > 0.40$. The proportions of markers with extreme values are much less than the 11% and 6%, respectively, reported by Akey et al. (2002). However, the human study only considered 3 populations, whereas in this study 12 breeds were considered.

Table 4.2. Summary statistics for F_{ST} for combined SNP sets

BTA	Length (bp)	# SNP	Avg. F_{ST}	Std. Dev.	No. Markers[1] Very Great	Great	Mod.	Little
1	160,907,802	1,555	0.133	0.068	91	428	951	85
2	140,356,784	1,446	0.130	0.067	76	376	905	89
3	127,652,798	1,279	0.134	0.070	78	365	763	73
4	124,039,900	1,215	0.136	0.068	64	356	732	63
5	125,784,649	1,214	0.148	0.077	105	410	638	61
6	122,295,181	2,367	0.147	0.078	256	682	1,341	88
7	111,450,679	1,034	0.139	0.070	73	320	582	59
8	116,646,425	1,154	0.125	0.060	50	283	744	77
9	107,350,408	963	0.134	0.070	58	263	603	39
10	106,098,797	1,046	0.125	0.059	40	270	673	63
11	110,099,902	1,159	0.144	0.069	83	394	624	58
12	85,206,304	837	0.126	0.066	33	210	535	59
13	84,107,162	924	0.149	0.075	91	292	498	43
14	81,080,327	2,662	0.135	0.065	134	791	1,615	122
15	84,423,077	770	0.122	0.056	20	199	500	51
16	77,570,437	793	0.127	0.062	28	202	516	47
17	76,127,165	795	0.134	0.072	52	217	481	45
18	65,707,717	653	0.135	0.068	43	186	390	34
19	65,063,234	683	0.132	0.071	36	173	436	38
20	75,458,338	807	0.133	0.069	49	227	479	52
21	68,877,573	662	0.118	0.058	19	141	456	46
22	61,746,535	643	0.126	0.058	19	169	422	33
23	53,228,442	592	0.122	0.064	30	128	383	51
24	64,932,885	679	0.133	0.066	34	196	416	33
25	43,444,595	1,233	0.118	0.054	30	290	818	95
26	51,000,868	532	0.131	0.069	30	141	332	29
27	48,747,412	457	0.110	0.052	6	79	330	42
28	46,014,400	486	0.115	0.052	11	105	338	32
29	51,649,444	491	0.124	0.060	17	129	304	41
All	**2,537,069,240**	**29,131**	**0.131**	**0.065**	**1,656**	**8,022**	**17,805**	**1,648**

[1]Number of markers in each category based on the qualitative guidelines for interpretation of F_{ST} (Wright, 1978); very great differentiation ($F_{ST} > 0.25$), great differentiation (0.15 to 0.25), moderate differentiation (0.05 to 0.15) and little differentiation ($F_{ST} < 0.05$).

a.



b.

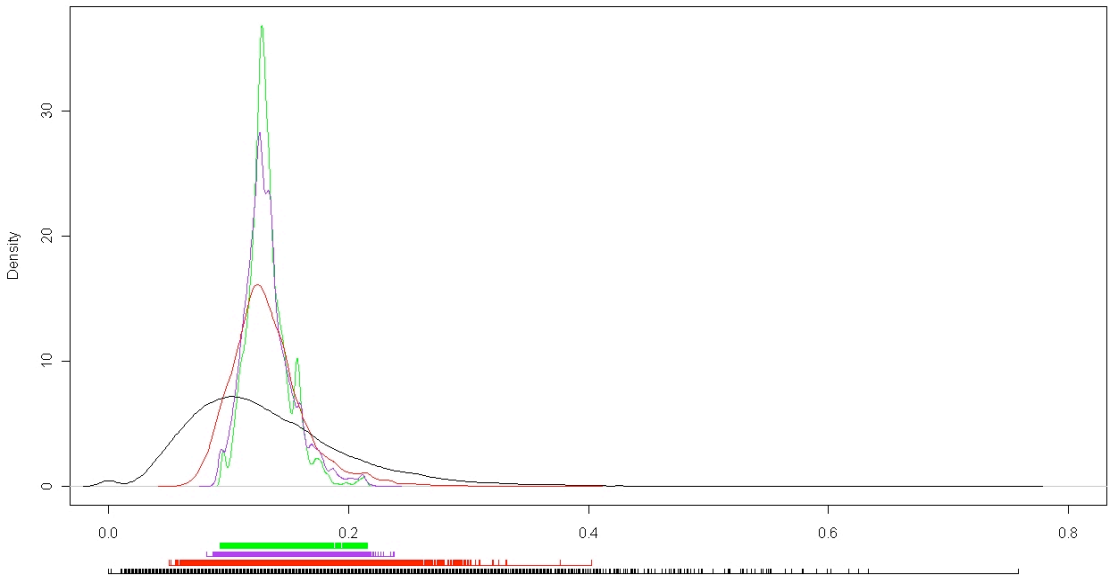


Figure 4.4. Comparison of different sliding windows for F_{ST} calculations. Histogram (a) and density plot (b) for 29,131 autosomal SNP. Individual SNP are plotted in black, 1 Mb sliding windows are in red, 5 Mb sliding windows are in purple, and 10 Mb sliding windows are in green.

As in Weir et al. (2005), a distribution that was closer to normality was obtained by averaging values of F_{ST} for adjacent markers (Figure 4.4). However, all of the distributions failed the Kolmogorov-Smirnov test for normality ($P < 0.05$). The Bovine Hapmap Consortium (2009) used 8-marker non-overlapping sliding windows for the calculation of F_{ST}. A comparison of the distribution of FST calculated using 3,642 8-marker non-overlapping windows was approximately equivalent to that found using 1 Mb overlapping sliding windows (Figures 4.4 and 4.5). To display population average F_{ST} by chromosomal coordinates, estimates for individual SNP (Figure 4.6) and 1 Mb sliding windows (Figure 4.7 and Appendix B) were used.

When individual SNP were considered, there were 1,656 markers (5.68 %) across the genome with extremely high values of F_{ST} (>0.25), suggestive of divergent selection (Figure 4.6). Conversely, there were 1,648 markers (5.65 %) across the genome with extremely low values of F_{ST} (<0.05), suggestive of balancing selection. When 1Mb overlapping sliding windows were used, there were 162 windows (0.56%) with extremely high values of F_{ST} (>0.25) suggestive of divergent selection (Figure 4.7 and Table 4.3). There were no windows observed with very low values of F_{ST} (<0.05) identified by this approach. This is possibly because the distribution is skewed towards higher values of F_{ST}. The average F_{ST} was not affected by using overlapping sliding windows, but the variation about the mean was greatly reduced (Table 4.3).

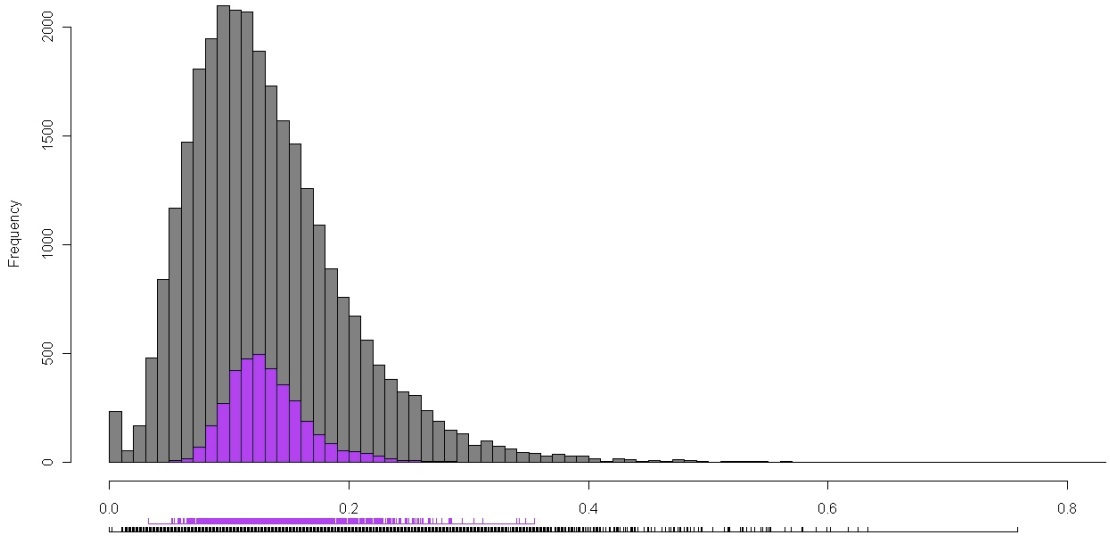


Figure 4.5. Distribution of F_{ST} values for 8-marker non-overlapping sliding windows. Histogram of (a) 29,131 individual autosomal SNP in black, and (b) 3,642 non-overlapping windows in purple.

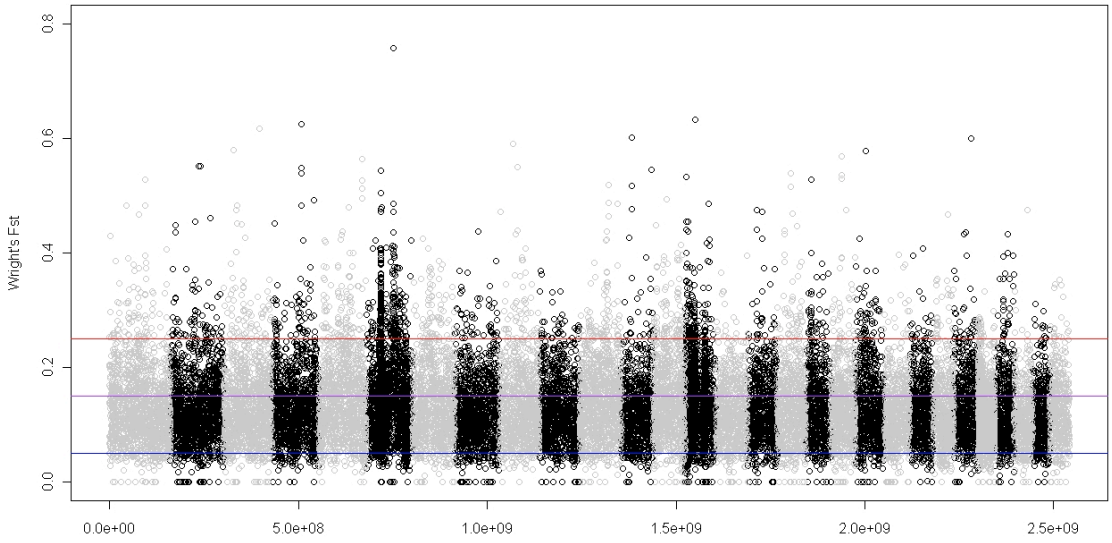


Figure 4.6. Distribution of individual F_{ST} values across the 29 bovine autosomes. Values for each marker were plotted against the coordinates in Mb from build Btau 4.0 of the bovine genome sequence. Odd numbered chromosomes are presented in gray and even numbered chromosomes are black. Horizontal lines indicate F_{ST} thresholds of 0.05, 0.15 and 0.25 that are commonly used for qualitative interpretation of Wright's fixation index. Very high values of F_{ST} are suggestive of divergent selection and very low values are suggestive of balancing selection.

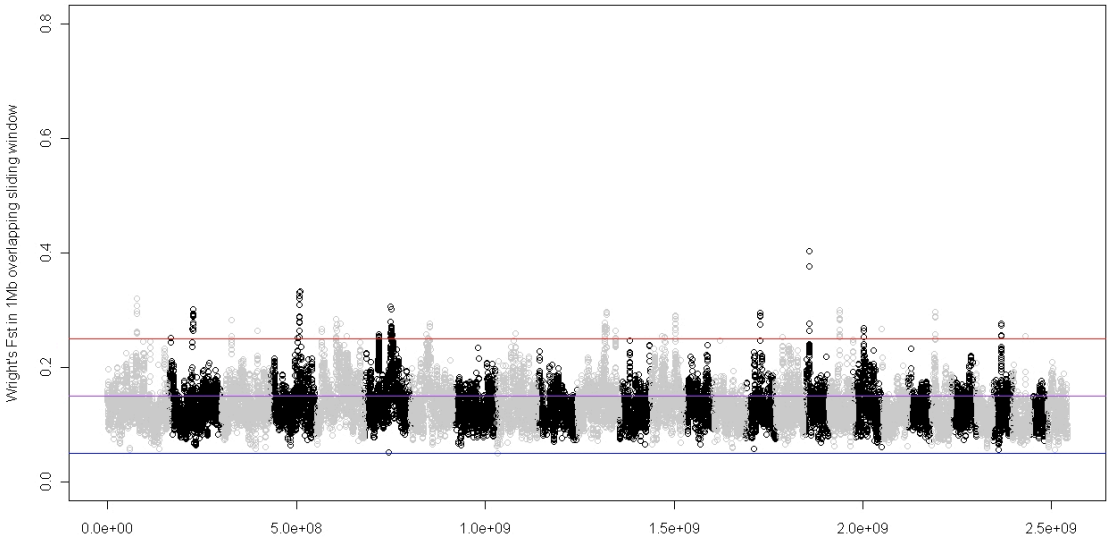


Figure 4.7. Distribution of F_{ST} values in 1 Mb overlapping sliding windows across the 29 bovine autosomes. Values for each window were plotted against the coordinates in Mb from build Btau4.0 of the bovine genome sequence. Odd numbered chromosomes are presented in gray and even numbered chromosomes are black. Horizontal lines indicate F_{ST} thresholds of 0.05, 0.15 and 0.25 that are commonly used for qualitative interpretation of Wright's fixation index. Very high values of F_{ST} are suggestive of divergent selection and very low values are suggestive of balancing selection.

Table 4.3. Summary statistics for estimates of F_{ST} generated using 1 Mb overlapping sliding windows

					No. Windows[1]			
BTA	No. Windows	Avg. SNP/Mb	Avg. F_{ST}	Std. Dev.	Very Great	Great	Mod.	Little
1	1,555	12.5	0.133	0.032	7	337	1,211	0
2	1,446	13.2	0.131	0.031	16	300	1,130	0
3	1,279	12.7	0.134	0.029	3	353	923	0
4	1,215	13.0	0.136	0.033	13	310	892	0
5	1,214	12.0	0.148	0.039	12	504	698	0
6	2,367	57.4	0.147	0.038	36	737	1,594	0
7	1,034	11.8	0.140	0.035	6	320	708	0
8	1,154	12.2	0.125	0.024	0	186	968	0
9	963	11.4	0.134	0.031	1	256	706	0
10	1,046	12.6	0.125	0.026	0	156	890	0
11	1,159	13.2	0.144	0.033	23	378	758	0
12	837	12.3	0.126	0.027	0	132	705	0
13	924	14.3	0.149	0.035	10	377	537	0
14	2,662	117.9	0.135	0.019	0	488	2,174	0
15	770	11.8	0.122	0.022	0	75	695	0
16	793	14.0	0.127	0.029	4	134	655	0
17	795	12.7	0.135	0.033	1	230	564	0
18	653	12.9	0.135	0.035	4	148	501	0
19	683	12.9	0.132	0.032	12	111	560	0
20	807	13.4	0.133	0.034	3	212	592	0
21	662	12.8	0.118	0.024	1	52	609	0
22	643	13.3	0.127	0.021	0	77	566	0
23	592	13.2	0.122	0.035	5	89	498	0
24	679	12.4	0.133	0.023	0	147	532	0
25	1,233	67.4	0.118	0.018	0	24	1,209	0
26	532	12.8	0.131	0.033	4	133	395	0
27	457	11.8	0.110	0.025	1	21	435	0
28	486	13.2	0.115	0.019	0	27	459	0
29	491	12.6	0.124	0.027	0	72	419	0
All	29,131	19.8	0.131	0.029	162	6,386	22,583	0

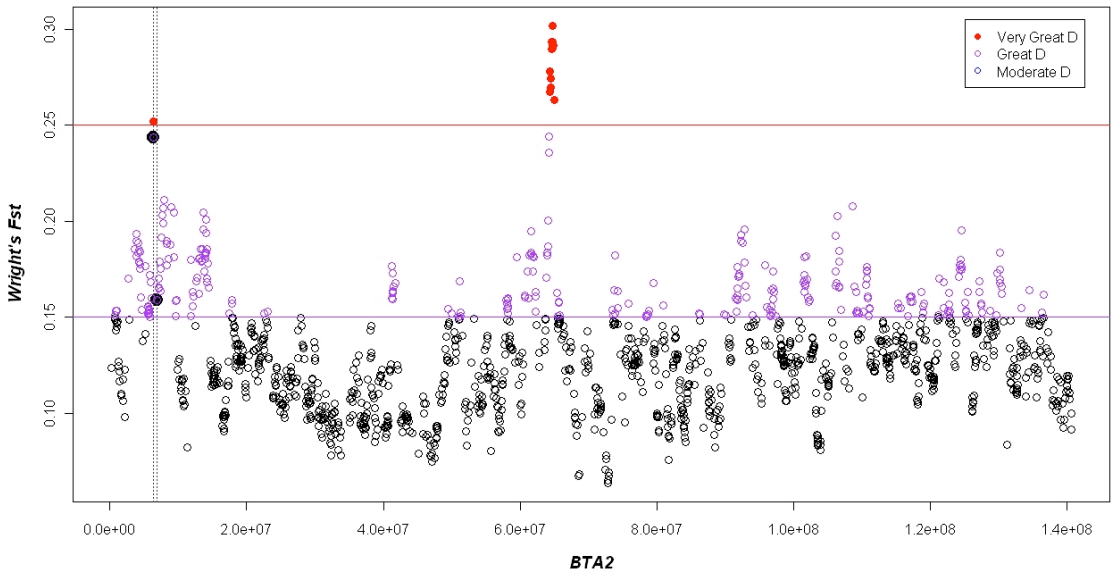
[1]Number of markers in each category based on the qualitative guidelines for interpretation of F_{ST} (Wright, 1978); very great differentiation ($F_{ST} > 0.25$), great differentiation (0.15 to 0.25), moderate differentiation (0.05 to 0.15) and little differentiation ($F_{ST} < 0.05$).

Clusters of markers with high F_{ST} may allow genes under directional selection to be identified. This is exemplified for SNP in the vicinity of myostatin (*MSTN*) on BTA2 and diacylglycerol O-acyltransferase 1 (*DGAT1*) on BTA14 that are considered to be under positive selection in some beef (Bellinge et al., 2005) and dairy breeds (Grisart et al., 2004), respectively (Figures 4.8 and 4.9).

Mutations that inactivate MSTN cause muscular hypertrophy, also known as double muscling (McPherron and Lee, 1997). The phenotype was first documented in 1807 and has become increasingly widespread in European cattle (Culley, 1807 cited by Bellinge et al., 2005). The double muscling phenotype is observed at moderate to high frequencies in the Piedmontese, Limousin and Charolais breeds (Grobet et al., 1998), but at low frequencies in the other breeds that were sampled for the HapMap study. In the vicinity of *MSTN* at ~6.53 Mbp on BTA2 there is a cluster of 8 markers with $F_{ST} > 0.25$ (Figure 4.9a). Although these SNP with extreme F_{ST} values are not in the MSTN gene, it is likely that haplotypes for these SNP have hitchhiked (Maynard Smith and Haigh, 1974) with the *MSTN* mutation and thus mark this as a region under positive selection for the double muscling phenotype.

A mutation (K232A) in DGAT1 has been shown to influence milk percentage, milk yield, and intramuscular fat content in cattle (Grisart et al., 2002; 2004). The K allele increases fat percentage, which has been a major breeding objective of the dairy industry and therefore this mutation has been under strong positive selection (Grisart et al., 2004). Grisart et al. (2004) suggested that because a limited number of K-carrying

a.



b.

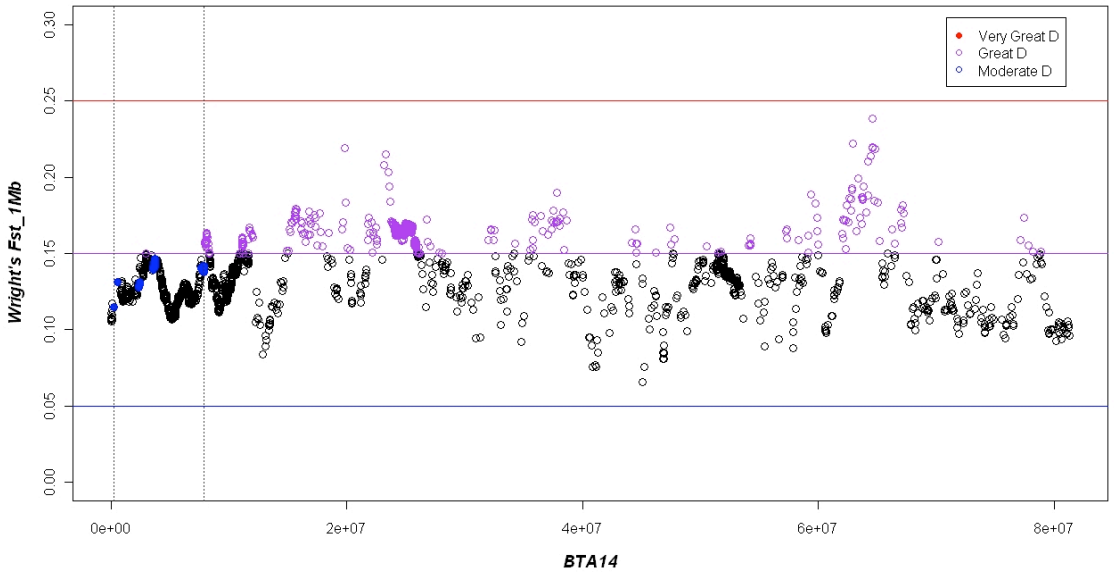


Figure 4.8. Distribution of F_{ST} values in 1Mb overlapping sliding windows across individual chromosomes. (a) BTA2. Dotted vertical lines and solid black circles indicate the coordinates and SNP that flank *MSTN*. (b) BTA14. Dotted vertical lines and solid blue circles indicate the coordinates and SNP that flank the *DGAT1, AGO2, COL22A1* and *TG*, respectively. An expanded view of these regions is in Figure 4.12 below. Horizontal lines indicate F_{ST} thresholds of 0.05, 0.15 and 0.25 that are commonly used for qualitative interpretation of Wright's fixation index. Values for each window were plotted against the coordinates in Mb from build Btau4.0 of the bovine genome sequence.

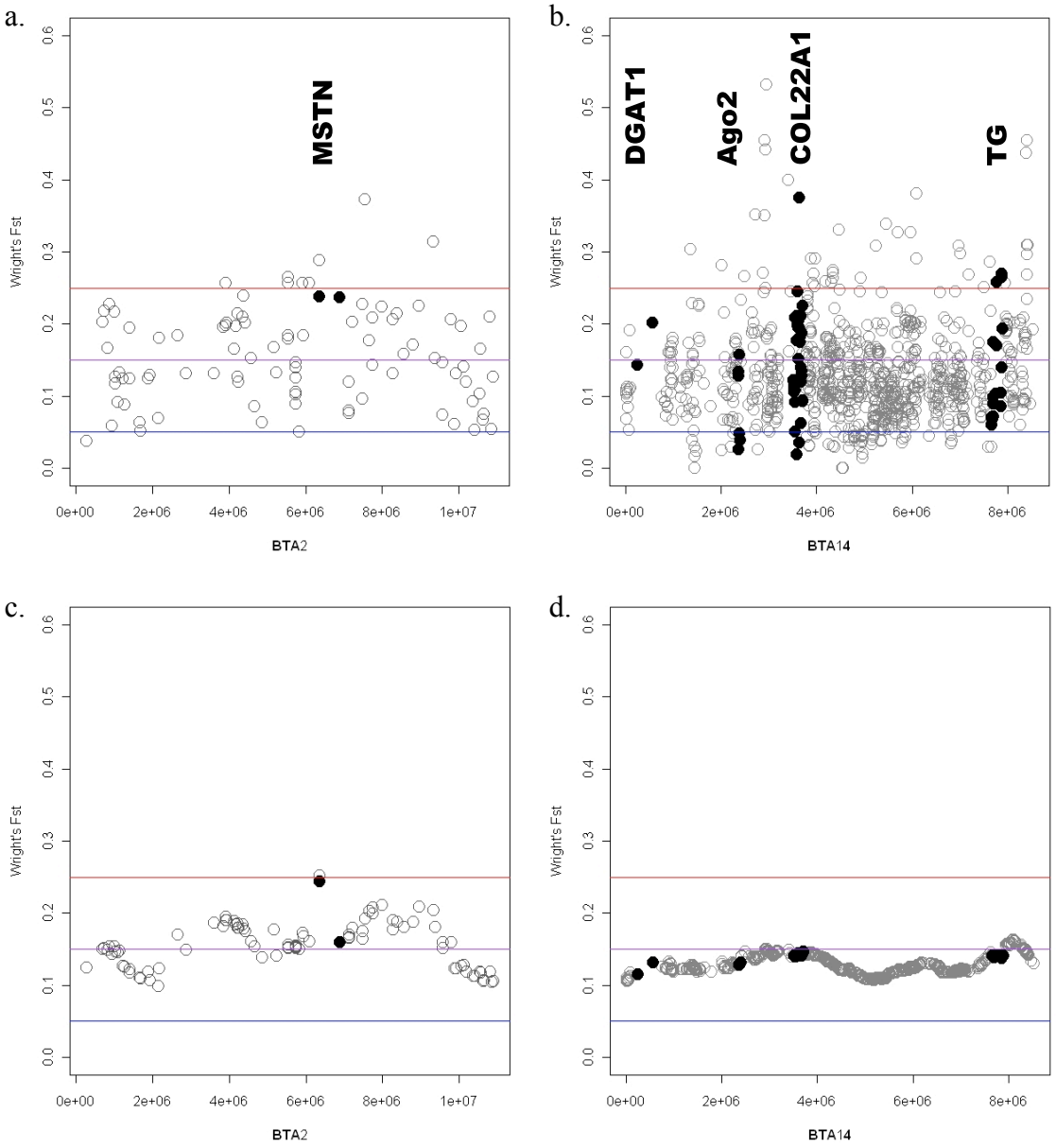


Figure 4.9. Distribution of F_{ST} values in regions known to be under positive selection. (a-b) Individual FST values and (c-d) 1Mb overlapping sliding windows in the proximal region of BTA2 (a, c) that includes the gene for myostatin and the proximal region of BTA14 (b, d) that includes the genes encoding diacylglycerol O-acyltransferase 1, thyroglobulin, eukaryotic translation initiation factor 2C, 2 and collagen, type XXII, alpha 1. Single nucleotide polymorphisms that flank or fall within the named genes are shown in black. Values were plotted against the coordinates in Mb from build Btau4.0 of the bovine genome sequence.

chromosomes would have initially existed, there would be considerable linkage disequilibrium with surrounding markers due to the selective sweep for the K allele. However, in the current study, when average F_{ST} for the 12 *Bos taurus* breeds was used as a measure of population differentiation, there was limited evidence of this sweep (Figure 4.9b, d). Few of the markers adjacent to DGAT1 had high values (>0.15) for F_{ST}. However, it should be noted that no markers in this region from 0 to ~1.5 Mb had very low F_{ST} (<0.05).

In both dairy and beef cattle, the confidence interval for the QTL containing DGAT1 extends for 10 to 20 cM. Thus, other genes underlying this QTL may also contribute to the variation in fat composition attributed to DGAT1. For example, thyroglobulin (TG) at ~7.7 Mb (Figure 4.9b, d) is associated with marbling and quality grade in beef cattle and SNP in TG are part of the GeneStar Quality Grade marker panel (Barendse, 1999; Van Eenennaam et al., 2007). For the SNP tested in this study, several that flanked TG had extreme values of F_{ST} and very few markers had very low F_{ST} values, suggesting strong population differentiation in the vicinity of TG.

One of the long-term goals of this study is to identify novel genes that exhibit signatures of recent positive selection. However, this remains challenging because of the relatively low density of SNP markers currently available and the relatively poor annotation of the bovine genome. Often, SNP with extreme values of F_{ST} lie several kilobases from the nearest known gene. In such cases, the question remains whether the SNP are in linkage disequilibrium with other gene-associated SNP or whether they mark as yet undetermined DNA regulatory regions, such as transcription factor binding sites,

enhancers or silencers. As shown in Figure 4.9, SNP in the vicinity of argonaute 2 (AGO2), which is responsible for microRNA cleavage in RNA interference (Morita et al., 2007), had high values of F_{ST}. In mice, knockout of AGO2 causes embryonic lethality early in development. In cattle, it is therefore possible that AGO2 contributes to variation in fertility associated with early embryo losses.

In Figure 4.9, it was also observed that there were many SNP with high F_{ST} values in the vicinity of the collagen, type XXII, alpha 1 (COL22A1) gene (Koch et al., 2004), which encodes a component of collagen XXII. This protein interacts with components of microfibrils (Koch et al., 2004) and therefore it is possible that COL22A1 contributes to variation in meat quality in beef cattle.

4.4 Population-Specific F_{ST}

To be included in calculations of population-specific F_{ST}, at least two breeds needed to be segregating for the SNP and only those breeds that were segregating (MAF > 0.05) were used. This differs from the approach of Weir et al. (2005) who required that the markers be segregating in all 4 human populations. This criterion would have reduced the number of markers for the 12 bovine breeds to only 14,103 SNP.

One issue with the calculation of population-specific F_{ST} using the approach of Weir and Hill (2002) is that negative values can be obtained, but these have no biological meaning. Negative estimates occur when the minor allele frequency in the specific population being considered is close to 0.5 (Figure 4.10 for Holstein). This makes the numerator greater than the denominator, which when subtracted from 1 results in a negative β_i value.

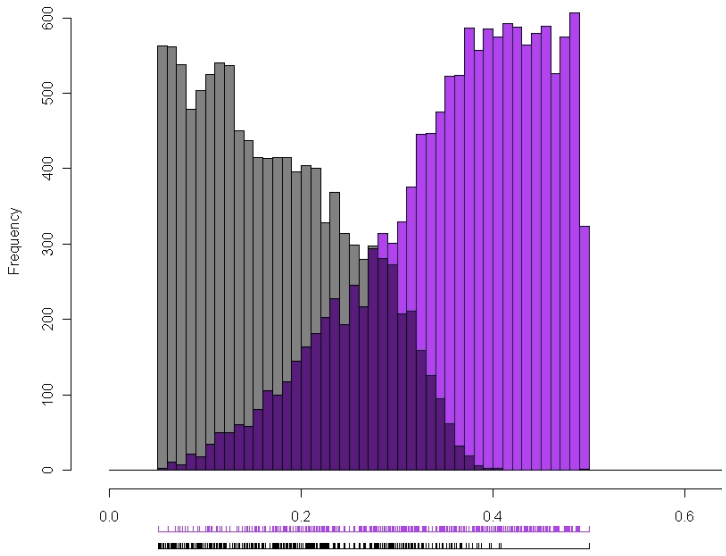


Figure 4.10. Effect of minor allele frequency on population specific F_{ST} calculations. Distribution of MAF that produced positive or zero F_{ST} values are in grey and those that produced negative F_{ST} values are in purple.

There are two possible explanations for the large number of SNP with negative F_{ST} in Holstein. The first is ascertainment bias because Holstein was one of the breeds from which the majority of the SNP were derived. If this was the case, we would expect the number of negative values to be significantly lower in breeds from which the SNP were not derived and this was not observed (data not shown). The second possibility is that the between breed discovery of SNP that was used in the bovine HapMap project identifies ancient SNP, which tend to have higher allele frequencies than recently derived SNP.

All of the negative values were converted to $F_{ST} = 0$. Because of this, the distributions of estimates of population-specific F_{ST} for individual SNP were very skewed (e.g. Figure 4.11 for Holstein). For example, for Holstein there were 13,925 markers (54.6%) with $F_{ST} = 0$ and 4,671 (12.1%) with $F_{ST} > 0$. As previously demonstrated by Weir et al. (2005) for human data, there were huge standard deviations associated with single marker estimates (Table 4.4). Averaging estimates of F_{ST} across overlapping sliding windows did not affect the mean, but greatly reduced the variation about the mean and 5 Mb windows were considered optimal for these data (Figure 4.11, Table 4.5 and Appendix C). The mean values and standard deviations obtained using the 5 Mb overlapping sliding windows were similar to those obtained by Weir et al. (2005) for human data. As observed for population average F_{ST}, the region containing *MSTN* on BTA2 (Figure 4.12) associated with double muscling in Piedmontese and Limousin cattle exhibited a strong signature of selection in those breeds. Surprisingly, the region

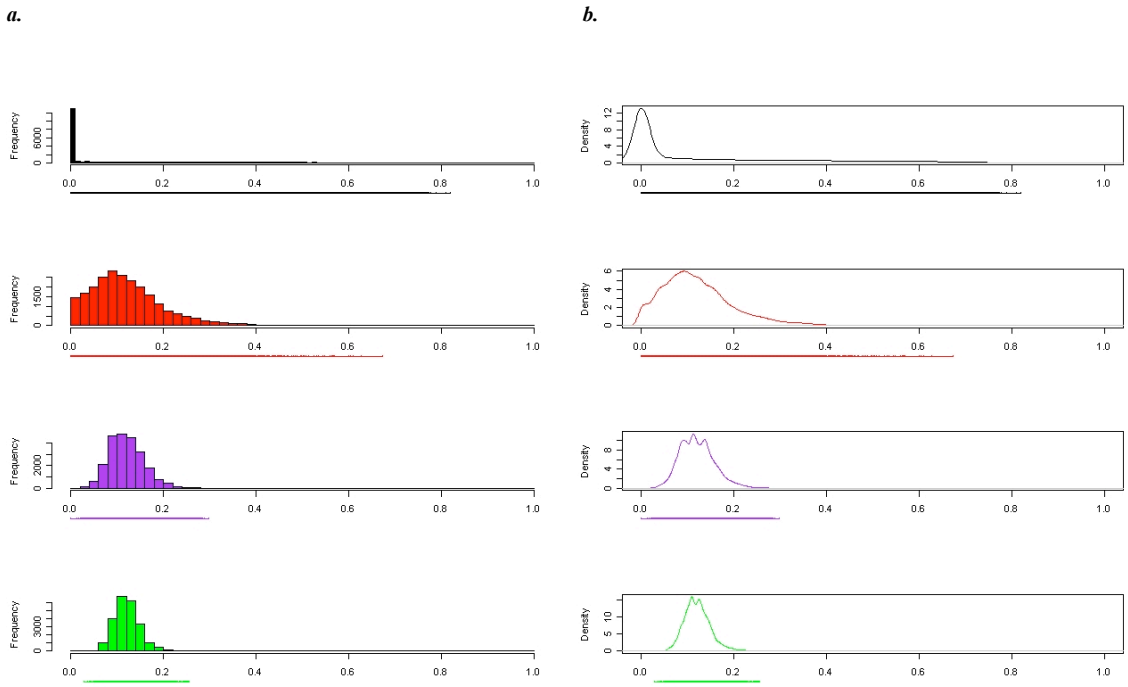


Figure 4.11. Comparison of different sliding windows for population-specific F_{ST} calculations in Holstein. Histogram (a) and density plot (b) for 25,782 autosomal SNP. Individual SNP are plotted in black, 1 Mb sliding windows are in red, 5 Mb sliding windows are in purple, and 10 Mb sliding windows are in green.

Table 4.4. Mean (std. dev.) for population-specific F_{ST} based on individual SNP

	Beef Breeds							Dairy Breeds				
BTA	Angus	Red Angus	Charolais	Hereford	Limousin	Piedmontese	Romagnola	Brown Swiss	Guernsey	Holstein	Jersey	Norwegian Red
1	0.14	0.10	0.13	0.11	0.12	0.10	0.11	0.11	0.13	0.13	0.17	0.13
	(0.20)	(0.16)	(0.20)	(0.19)	(0.18)	(0.17)	(0.19)	(0.18)	(0.19)	(0.20)	(0.22)	(0.20)
2	0.13	0.08	0.12	0.12	0.14	0.11	0.14	0.11	0.13	0.13	0.16	0.12
	(0.19)	(0.15)	(0.19)	(0.21)	(0.19)	(0.18)	(0.19)	(0.18)	(0.20)	(0.20)	(0.22)	(0.18)
3	0.17	0.12	0.11	0.11	0.12	0.10	0.13	0.12	0.12	0.12	0.17	0.12
	(0.22)	(0.18)	(0.18)	(0.19)	(0.19)	(0.17)	(0.19)	(0.19)	(0.19)	(0.19)	(0.23)	(0.19)
4	0.14	0.11	0.13	0.11	0.12	0.11	0.13	0.13	0.12	0.13	0.15	0.12
	(0.20)	(0.17)	(0.19)	(0.19)	(0.17)	(0.17)	(0.19)	(0.20)	(0.19)	(0.20)	(0.20)	(0.20)
5	0.14	0.11	0.14	0.12	0.13	0.11	0.12	0.15	0.12	0.15	0.13	0.14
	(0.21)	(0.17)	(0.20)	(0.20)	(0.19)	(0.17)	(0.19)	(0.20)	(0.18)	(0.20)	(0.21)	(0.21)
6	0.12	0.11	0.14	0.11	0.12	0.13	0.18	0.16	0.14	0.12	0.15	0.15
	(0.20)	(0.17)	(0.20)	(0.18)	(0.18)	(0.18)	(0.23)	(0.23)	(0.20)	(0.19)	(0.21)	(0.21)
7	0.13	0.11	0.12	0.13	0.13	0.10	0.13	0.12	0.13	0.12	0.19	0.13
	(0.19)	(0.18)	(0.19)	(0.20)	(0.19)	(0.17)	(0.19)	(0.19)	(0.19)	(0.19)	(0.23)	(0.21)
8	0.14	0.08	0.13	0.12	0.14	0.11	0.11	0.12	0.13	0.11	0.17	0.13
	(0.19)	(0.15)	(0.19)	(0.20)	(0.20)	(0.17)	(0.18)	(0.19)	(0.19)	(0.18)	(0.21)	(0.20)
9	0.12	0.09	0.12	0.13	0.12	0.11	0.13	0.14	0.13	0.12	0.15	0.11
	(0.18)	(0.16)	(0.19)	(0.20)	(0.18)	(0.17)	(0.19)	(0.20)	(0.19)	(0.18)	(0.21)	(0.19)
10	0.13	0.10	0.13	0.13	0.13	0.12	0.14	0.13	0.13	0.12	0.13	0.12
	(0.19)	(0.17)	(0.19)	(0.20)	(0.19)	(0.18)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)	(0.20)
11	0.12	0.13	0.13	0.11	0.13	0.11	0.11	0.16	0.17	0.12	0.15	0.12
	(0.19)	(0.18)	(0.19)	(0.19)	(0.19)	(0.17)	(0.18)	(0.21)	(0.22)	(0.19)	(0.21)	(0.19)
12	0.14	0.10	0.12	0.12	0.12	0.10	0.12	0.14	0.15	0.13	0.15	0.14
	(0.20)	(0.17)	(0.19)	(0.19)	(0.18)	(0.17)	(0.18)	(0.19)	(0.20)	(0.19)	(0.21)	(0.20)
13	0.15	0.12	0.13	0.14	0.12	0.11	0.12	0.16	0.11	0.13	0.16	0.13
	(0.20)	(0.17)	(0.20)	(0.21)	(0.19)	(0.18)	(0.19)	(0.21)	(0.18)	(0.19)	(0.23)	(0.20)
14	0.14	0.10	0.12	0.11	0.15	0.13	0.16	0.12	0.13	0.11	0.13	0.16
	(0.20)	(0.16)	(0.19)	(0.19)	(0.20)	(0.18)	(0.22)	(0.19)	(0.20)	(0.19)	(0.20)	(0.22)
15	0.13	0.09	0.12	0.11	0.12	0.11	0.13	0.14	0.15	0.11	0.13	0.12
	(0.19)	(0.16)	(0.19)	(0.18)	(0.19)	(0.17)	(0.19)	(0.20)	(0.21)	(0.17)	(0.19)	(0.19)

Table 4.4 Continued.

	Beef Breeds							Dairy Breeds				
BTA	Angus	Red Angus	Charolais	Hereford	Limousin	Piedmontese	Romagnola	Brown Swiss	Guernsey	Holstein	Jersey	Norwegian Red
16	0.13	0.09	0.11	0.10	0.13	0.11	0.14	0.17	0.14	0.12	0.13	0.11
	(0.20)	(0.16)	(0.18)	(0.19)	(0.18)	(0.17)	(0.21)	(0.21)	(0.20)	(0.19)	(0.19)	(0.18)
17	0.13	0.10	0.13	0.15	0.13	0.11	0.11	0.14	0.14	0.12	0.15	0.13
	(0.19)	(0.17)	(0.20)	(0.22)	(0.19)	(0.17)	(0.18)	(0.20)	(0.21)	(0.18)	(0.21)	(0.19)
18	0.12	0.10	0.13	0.12	0.10	0.10	0.12	0.14	0.15	0.11	0.20	0.12
	(0.20)	(0.16)	(0.19)	(0.19)	(0.16)	(0.17)	(0.19)	(0.20)	(0.21)	(0.19)	(0.23)	(0.20)
19	0.15	0.10	0.10	0.15	0.11	0.10	0.12	0.15	0.14	0.12	0.15	0.13
	(0.20)	(0.17)	(0.17)	(0.21)	(0.17)	(0.16)	(0.18)	(0.21)	(0.20)	(0.18)	(0.21)	(0.19)
20	0.13	0.09	0.12	0.12	0.13	0.10	0.14	0.15	0.12	0.14	0.16	0.12
	(0.20)	(0.16)	(0.18)	(0.19)	(0.18)	(0.17)	(0.20)	(0.21)	(0.19)	(0.20)	(0.22)	(0.19)
21	0.10	0.10	0.13	0.11	0.14	0.09	0.11	0.12	0.12	0.12	0.14	0.14
	(0.17)	(0.17)	(0.20)	(0.18)	(0.19)	(0.16)	(0.19)	(0.19)	(0.19)	(0.18)	(0.20)	(0.21)
22	0.10	0.10	0.12	0.11	0.12	0.12	0.14	0.15	0.13	0.11	0.15	0.13
	(0.17)	(0.17)	(0.19)	(0.18)	(0.18)	(0.18)	(0.21)	(0.21)	(0.19)	(0.18)	(0.22)	(0.21)
23	0.13	0.12	0.13	0.10	0.12	0.12	0.13	0.14	0.12	0.10	0.13	0.15
	(0.18)	(0.19)	(0.19)	(0.18)	(0.18)	(0.18)	(0.19)	(0.20)	(0.18)	(0.17)	(0.19)	(0.22)
24	0.11	0.09	0.13	0.11	0.13	0.12	0.11	0.14	0.15	0.14	0.15	0.12
	(0.18)	(0.17)	(0.20)	(0.18)	(0.19)	(0.18)	(0.18)	(0.20)	(0.20)	(0.21)	(0.22)	(0.19)
25	0.12	0.09	0.11	0.10	0.13	0.10	0.13	0.13	0.12	0.10	0.12	0.13
	(0.18)	(0.17)	(0.18)	(0.17)	(0.19)	(0.16)	(0.18)	(0.20)	(0.18)	(0.18)	(0.19)	(0.19)
26	0.14	0.11	0.13	0.13	0.12	0.12	0.12	0.12	0.13	0.12	0.16	0.13
	(0.20)	(0.18)	(0.20)	(0.20)	(0.19)	(0.19)	(0.18)	(0.19)	(0.19)	(0.19)	(0.22)	(0.20)
27	0.14	0.11	0.12	0.09	0.11	0.10	0.11	0.12	0.14	0.10	0.19	0.13
	(0.20)	(0.18)	(0.18)	(0.17)	(0.17)	(0.16)	(0.18)	(0.19)	(0.19)	(0.17)	(0.23)	(0.20)
28	0.13	0.12	0.10	0.13	0.11	0.11	0.13	0.12	0.13	0.13	0.15	0.13
	(0.18)	(0.18)	(0.16)	(0.20)	(0.17)	(0.17)	(0.20)	(0.19)	(0.19)	(0.19)	(0.21)	(0.20)
29	0.10	0.10	0.12	0.13	0.13	0.10	0.10	0.12	0.12	0.12	0.19	0.13
	(0.17)	(0.17)	(0.19)	(0.20)	(0.19)	(0.17)	(0.17)	(0.18)	(0.19)	(0.19)	(0.23)	(0.20)
All	0.10	0.10	0.12	0.13	0.13	0.10	0.10	0.12	0.12	0.12	0.19	0.13
	(0.17)	(0.17)	(0.19)	(0.20)	(0.19)	(0.17)	(0.17)	(0.18)	(0.19)	(0.19)	(0.23)	(0.20)

Table 4.5. Mean (std. dev.) for population-specific F_{ST} based on 5 Mb overlapping sliding windows

	Beef Breeds							Dairy Breeds				
BTA	Angus	Red Angus	Charolais	Hereford	Limousin	Piedmontese	Romagnola	Brown Swiss	Guernsey	Holstein	Jersey	Norwegian Red
1	0.14	0.10	0.13	0.11	0.12	0.10	0.11	0.11	0.13	0.13	0.17	0.13
	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.07)	(0.04)
2	0.12	0.08	0.12	0.12	0.14	0.11	0.14	0.11	0.14	0.13	0.16	0.12
	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)
3	0.18	0.12	0.11	0.11	0.12	0.10	0.13	0.12	0.13	0.12	0.17	0.12
	(0.07)	(0.05)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.08)	(0.04)
4	0.14	0.11	0.13	0.11	0.12	0.11	0.13	0.13	0.13	0.13	0.15	0.12
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)
5	0.14	0.11	0.13	0.12	0.13	0.11	0.12	0.15	0.12	0.15	0.13	0.13
	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.07)	(0.04)	(0.05)	(0.04)	(0.05)
6	0.13	0.11	0.15	0.11	0.12	0.13	0.19	0.16	0.14	0.12	0.15	0.15
	(0.06)	(0.04)	(0.06)	(0.03)	(0.02)	(0.03)	(0.12)	(0.10)	(0.05)	(0.03)	(0.05)	(0.04)
7	0.13	0.11	0.12	0.13	0.13	0.10	0.13	0.12	0.13	0.13	0.19	0.13
	(0.05)	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.09)	(0.03)
8	0.14	0.08	0.13	0.12	0.14	0.11	0.11	0.12	0.13	0.11	0.17	0.13
	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.06)
9	0.12	0.09	0.12	0.13	0.12	0.11	0.13	0.14	0.13	0.12	0.15	0.11
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.06)	(0.04)	(0.05)	(0.03)	(0.05)	(0.06)
10	0.14	0.10	0.13	0.12	0.14	0.12	0.14	0.13	0.12	0.12	0.13	0.12
	(0.04)	(0.03)	(0.03)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
11	0.12	0.13	0.13	0.11	0.13	0.11	0.11	0.15	0.17	0.13	0.15	0.12
	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.06)	(0.03)	(0.04)	(0.04)
12	0.14	0.10	0.12	0.12	0.12	0.10	0.12	0.14	0.15	0.13	0.14	0.14
	(0.06)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
13	0.15	0.12	0.13	0.13	0.12	0.11	0.12	0.16	0.11	0.13	0.16	0.13
	(0.07)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)	(0.05)	(0.03)
14	0.14	0.09	0.12	0.11	0.15	0.13	0.16	0.12	0.14	0.11	0.13	0.16
	(0.04)	(0.03)	(0.04)	(0.02)	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)
15	0.13	0.10	0.12	0.11	0.12	0.11	0.13	0.14	0.15	0.11	0.13	0.12
	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)

Table 4.5 Continued.

	Beef Breeds							Dairy Breeds				
BTA	Angus	Red Angus	Charolais	Hereford	Limousin	Piedmontese	Romagnola	Brown Swiss	Guernsey	Holstein	Jersey	Norwegian Red
16	0.13	0.09	0.11	0.10	0.13	0.11	0.14	0.17	0.14	0.12	0.13	0.11
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.03)
17	0.13	0.10	0.13	0.16	0.13	0.11	0.11	0.14	0.14	0.12	0.15	0.13
	(0.04)	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.05)	(0.04)
18	0.12	0.10	0.13	0.12	0.10	0.10	0.12	0.14	0.15	0.11	0.20	0.12
	(0.05)	(0.05)	(0.03)	(0.05)	(0.03)	(0.03)	(0.06)	(0.04)	(0.05)	(0.04)	(0.06)	(0.03)
19	0.15	0.10	0.10	0.15	0.11	0.10	0.12	0.15	0.14	0.12	0.15	0.13
	(0.05)	(0.04)	(0.03)	(0.06)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.03)	(0.04)	(0.05)
20	0.13	0.09	0.12	0.12	0.13	0.10	0.14	0.16	0.12	0.14	0.15	0.12
	(0.05)	(0.03)	(0.05)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)
21	0.10	0.10	0.13	0.11	0.14	0.09	0.11	0.12	0.12	0.12	0.14	0.14
	(0.04)	(0.05)	(0.04)	(0.03)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
22	0.10	0.10	0.12	0.11	0.12	0.12	0.14	0.15	0.13	0.11	0.15	0.13
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.06)	(0.05)	(0.03)	(0.03)	(0.05)	(0.03)
23	0.13	0.12	0.13	0.10	0.12	0.12	0.13	0.14	0.12	0.10	0.13	0.14
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
24	0.11	0.09	0.13	0.11	0.13	0.11	0.11	0.14	0.15	0.14	0.15	0.12
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)	(0.06)	(0.04)
25	0.12	0.09	0.11	0.09	0.13	0.10	0.13	0.13	0.12	0.10	0.12	0.13
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
26	0.13	0.11	0.13	0.13	0.12	0.12	0.12	0.12	0.13	0.12	0.16	0.13
	(0.05)	(0.05)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)
27	0.14	0.11	0.12	0.09	0.11	0.10	0.11	0.12	0.14	0.10	0.19	0.13
	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.05)	(0.04)
28	0.13	0.12	0.10	0.14	0.11	0.11	0.13	0.12	0.13	0.13	0.15	0.13
	(0.02)	(0.05)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)
29	0.10	0.10	0.12	0.13	0.13	0.10	0.10	0.12	0.12	0.12	0.18	0.13
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)	(0.06)	(0.06)
All	0.10	0.10	0.12	0.13	0.13	0.10	0.10	0.12	0.12	0.12	0.18	0.13
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)	(0.06)	(0.06)

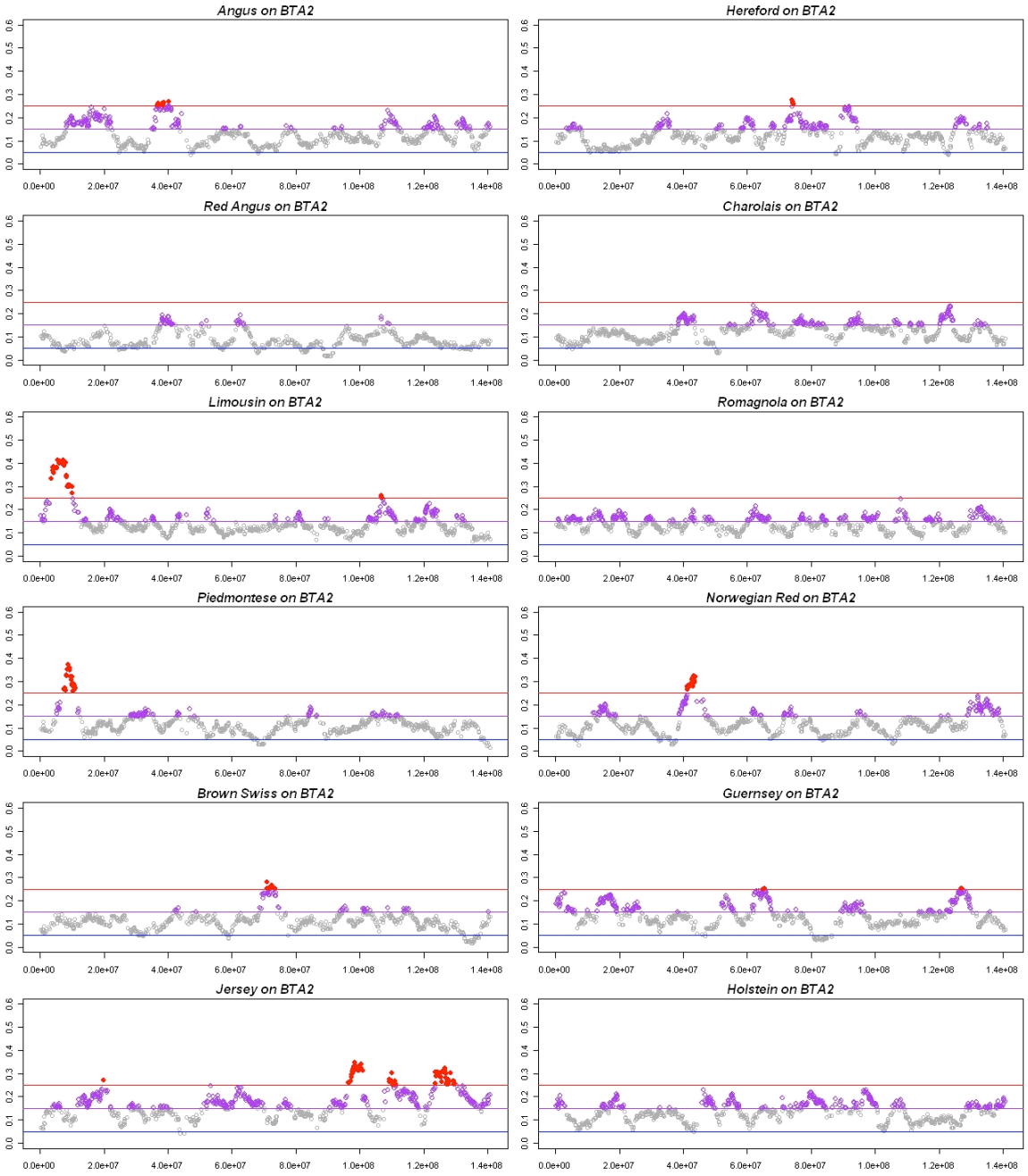


Figure 4.12. Population-specific F_{ST} on BTA2 generated using 5Mb overlapping sliding windows. Horizontal lines indicate F_{ST} thresholds of 0.05, 0.15 and 0.25 that are commonly used for qualitative interpretation of Wright's fixation index.

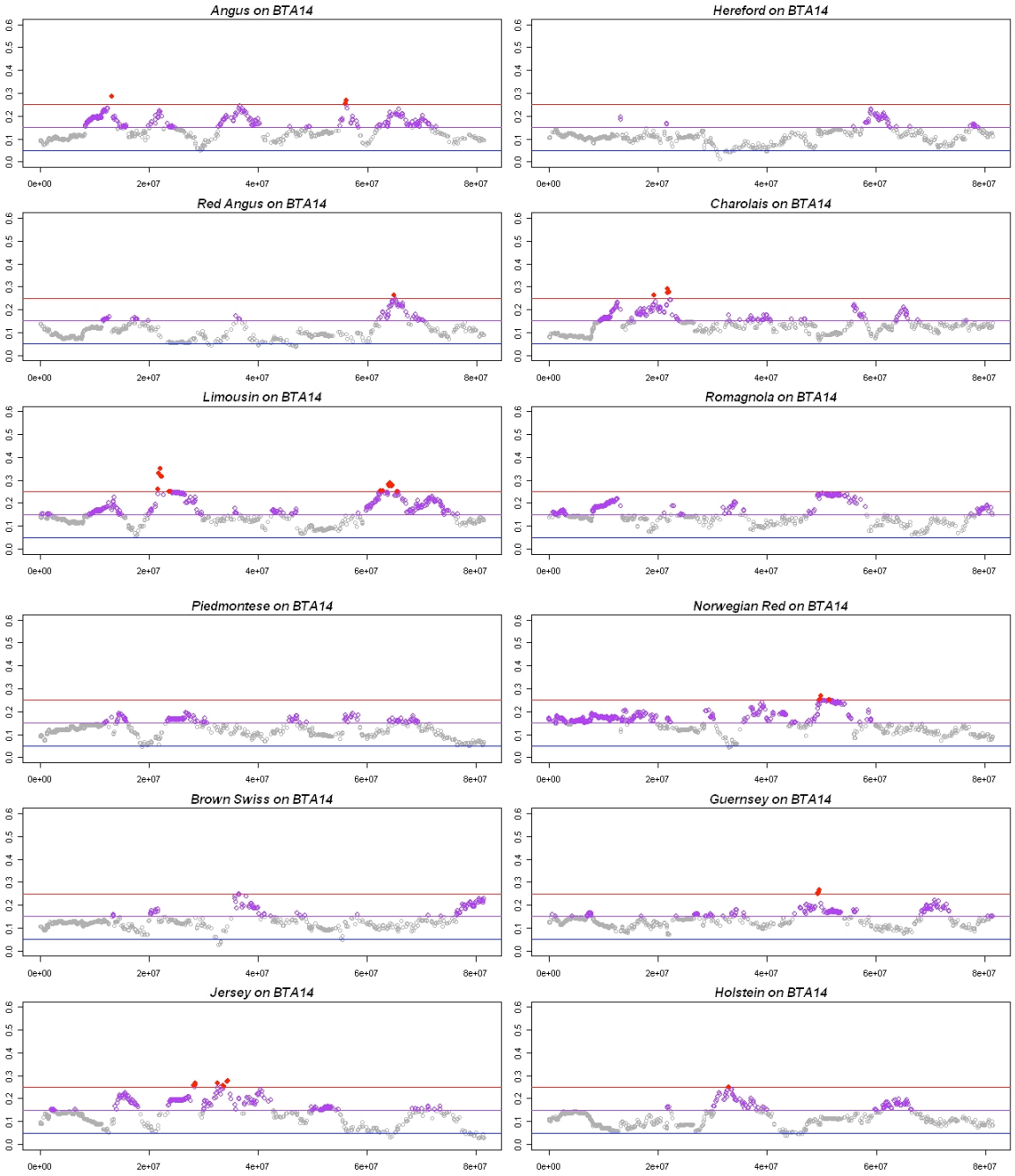


Figure 4.13. Population-specific F_{ST} on BTA14 generated using 5Mb overlapping sliding windows. Horizontal lines indicate F_{ST} thresholds of 0.05, 0.15 and 0.25 that are commonly used for qualitative interpretation of Wright's fixation index.

containing *DGAT1* on BTA14 (Figure 4.13) that is considered to be under positive selection in some dairy breeds did not have extreme F_{ST} values in any breed. A region towards the middle of BTA14 had a cluster of SNP with high population-specific F_{ST} values in several breeds.

4.5 Ancestral Allele State

The 2 DNA samples for Anoa (ANO00001 and ANO00002) produced 11,201 and 11,379 genotypes, respectively, representing 12,123 markers. The 2 DNA samples for Water buffalo (BUF00001 and BUF00002) produced 11,003 genotypes and 10,714 genotypes, respectively, representing 11,742 markers. A genotype was obtained from at least one of the 4 outgroup individuals for 12,810 markers (60.4% of the 21,207 valid markers considered). As expected, the Anoa and Water buffalo samples were monomorphic for the majority of filtered markers that were successfully genotyped. There were 48 markers that were fixed for alternate alleles in the Anoa and Water buffalo samples so the ancestral state could not be determined. A heterozygous genotype was produced by at least one animal for 2,497 markers. For 16 segregating markers, alternate homozygotes were observed and 1,789 markers were segregating in both species so the ancestral state could not be assigned. The ancestral allele was assigned for 11,366 markers of which 10,193 were assigned to autosomes (Table 1).

4.6 Integrated Extended Haplotype Homozygosity Score

The average spacing between markers where the ancestral state was assigned is 240 kb and the largest gap is 2,535 kb (Table 4.6). This spacing has implications for

Table 4.6. Assignment of ancestral allele by chromosome and spacing between markers

BTA	No. Valid Markers[1]	Avg. Spacing	Min. Spacing	Max. Spacing	No. Ancestral Known	Ancestral Avg. Spacing	Ancestral Min. Spacing	Ancestral Max. Spacing
1	1,489	96,692	7	941,423	598	241,000	25	2,438,107
2	1,464	89,633	8	730,260	580	225,420	12	1,587,630
3	1,284	92,737	3	712,246	521	226,764	3	1,823,586
4	1,284	91,340	6	904,416	524	223,121	9	1,733,402
5	1,217	95,435	20	756,584	470	245,975	21	1,726,378
6	2,420	48,138	11	903,871	536	217,229	23	1,977,165
7	1,058	97,809	4	986,266	398	257,240	21	1,576,673
8	1,179	92,399	9	1,019,449	459	236,783	29	1,411,567
9	976	104,787	2	940,400	386	261,034	21	2,091,272
10	1,103	93,008	7	1,104,187	454	226,258	20	2,052,176
11	1,205	89,337	5	875,942	483	223,156	28	1,680,469
12	864	92,536	1	946,290	336	238,253	26	1,459,213
13	968	84,357	5	890,466	346	234,274	7	1,558,751
14	2,722	29,578	5	623,416	370	217,768	5	1,454,769
15	818	97,649	2	1,944,406	291	264,361	23	2,074,585
16	858	86,291	8	867,493	333	221,505	8	1,769,846
17	803	88,922	1	569,866	326	219,432	36	1,256,942
18	640	95,311	15	1,019,891	273	223,550	17	2,328,223
19	665	93,815	4	972,841	252	246,760	16	1,366,682
20	860	87,167	7	1,421,140	334	222,701	36	1,920,696
21	650	99,322	15	799,431	231	276,120	25	1,740,887
22	678	88,235	4	1,123,980	264	220,842	12	2,059,685
23	567	86,520	8	580,792	211	231,122	8	1,424,476
24	675	89,824	21	680,079	265	229,322	15	1,272,328
25	1,184	34,673	1	879,814	169	242,198	50	1,041,640
26	581	84,460	16	1,075,463	238	205,827	25	1,462,344
27	480	92,917	17	881,457	177	252,882	26	1,620,904
28	501	88,724	20	670,421	192	232,262	19	1,632,138
29	467	101,204	10	1,189,809	176	267,392	6	2,373,272

Table 4.6 Continued.

BTA	No. Valid Markers[1]	Avg. Spacing	Min. Spacing	Max. Spacing	No. Ancestral Known	Ancestral Avg. Spacing	Ancestral Min. Spacing	Ancestral Max. Spacing
X	565	159,461	20	2,255,471	234	382,459	2	2,535,130
UN	2,802	-	-	-	939	-	-	-
Total	33,027	89,076	1	2,255,471	11,366	240,434	2	2,535,130

[1]Includes markers from the 4.5K and 7.5K sets that were not genotyped in the ancestral species

Table 4.7. Effect of window size (Mb) on the number (%) of markers for which iHS values were determined

Breed	1Mb	2.5Mb	5Mb	10Mb	15Mb
Angus	138 (1.3%)	5,621 (52.3%)	7,607 (70.8%)	8,065 (75.0%)	8,172 (76.0%)
Hereford	113 (1.1%)	5,576 (51.9%)	7,902 (73.5%)	8,471 (78.8%)	8,565 (79.7%)
Red Angus	115 (1.1%)	5,224 (48.6)	6,906 (64.3%)	7,249 (67.4%)	7,337 (68.3%)
Charolais	177 (1.6%)	7,394 (68.8%)	8,461 (78.7%)	8,586 (79.9%)	8,592 (79.9%)
Limousin	176 (1.6%)	7,717 (71.8%)	8714 (81.1%)	8,780 (81.7%)	8,822 (82.1%)
Romagnola	115 (1.1%)	5,396 (50.2%)	7,087 (65.9%)	7,715 (71.8%)	7,848 (73.0%)
Piedmontese	190 (1.8%)	7,701 (71.7%)	8,574 (79.8%)	8,597 (80.0%)	8,627 (80.3%)
Norwegian Red	140 (1.3%)	6,718 (62.5%)	8,245 (76.7%)	8,555 (79.6%)	8,633 (80.3%)
Brown Swiss	73 (0.7%)	3,244 (30.2%)	5,473 (50.9%)	6,956 (64.7%)	7,290 (67.8%)
Guernsey	130 (1.2%)	5,175 (48.1%)	7,009 (65.2%)	7,492 (69.7%)	7588 (70.6%)
Jersey	88 (0.8%)	3,457 (32.2%)	5,771 (53.7%)	6,944 (64.6%)	7,199 (67.0%)
Holstein	142 (1.3%)	5,820 (54.1%)	7,678 (71.4%)	8,576 (79.8%)	8,787 (81.8%)

analysis of signatures of selection by iHS (Voight et al., 2006) partly because the software was developed to handle much denser data from the human Hapmap project. Using the SNP positions associated with build Btau3.1 of the bovine genome sequence, we empirically determined that the optimum window size for these bovine data was 10Mb with a gap size of 200kb (Table 4.7). On average, these parameters enabled us to obtain standardized iHS values for 68% of the SNP for which the ancestral allele was derived, ranging from 64.6% for Jersey to 81.7% for Limousin.

Following Voight et al. (2006), we plotted |iHS| by position (Appendix D). The outlier approach is often used to identify selected loci in genome-wide studies, because the empirical null distribution can only be obtained by simulation (Akey, 2009). The threshold for determining an outlier is arbitrary and we followed the example of Voight et al. (2006) and chose the top 1% of |iHS| values (>2.694 across breeds). As observed with population-specific F_{ST}, there was evidence of a sweep in progress in Limousin and Piedmontese near *MSTN* on BTA2 (Figure 4.14). Unlike for population-specific F_{ST}, the region on BTA14 that contains *DGAT1, TG1, AGO2*, and *COL22A1* that was previously described for population-average F_{ST} did show evidence of a selective sweep in some breeds using iHS (Figure 4.15). Because iHS detects sweeps in progress and this region has only recently become emphasized in selection programs in the dairy industry, it is likely that this region is continuing to undergo strong artificial selection.

There were 51 regions identified where multiple markers within a 1 Mb interval had extreme |iHS| values in at least one breed (Table 4.8).

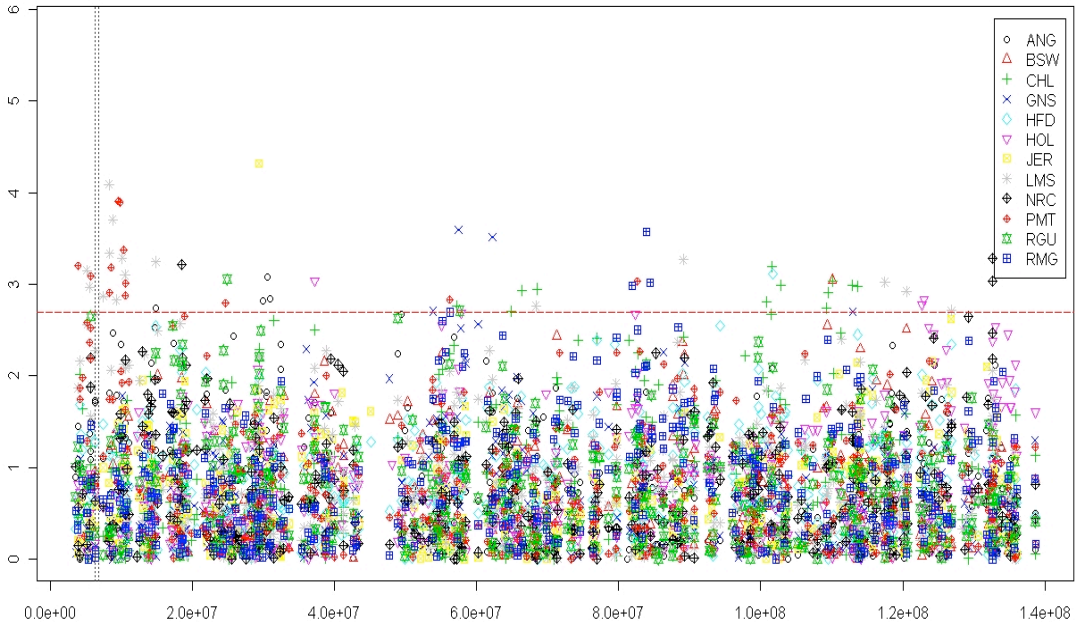


Figure 4.14. iHS values on BTA2 for each of the 12 breeds. Red horizontal lines indicate top 1% of |iHS| values (2.694) and vertical grey lines indicate that the proximal region of BTA2 that includes the gene for myostatin. ANG = Angus, BSW = Brown Swiss, CHL = Charolais, GNS = Guernsey, HFD = Hereford, HOL = Holstein, JER = Jersey, LMS = Limousin, NRC = Norwegian Red, PMT = Piedmontese, RGU = Red Angus, and RMG = Romagnola.

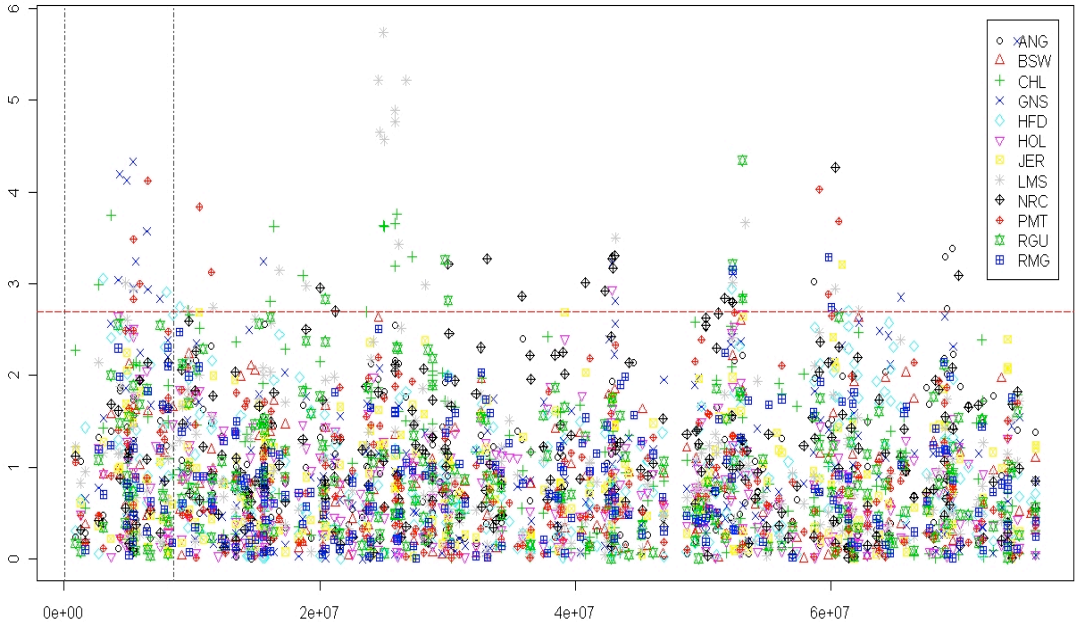


Figure 4.15. iHS values on BTA14 for each of the 12 breeds. Red horizontal lines indicate top 1% of |iHS| values (2.694) and vertical grey lines indicate that the proximal region of BTA14 that includes the genes encoding diacylglycerol O-acyltransferase 1, thyroglobulin, eukaryotic translation initiation factor 2C, 2 and collagen, type XXII, alpha.

Table 4.8. Regions with evidence of positive selection detected by iHS[1]

BTA	Position (Mb)	Breeds[2]
1	68.5-70.2	CHL PMT
	83.4-84.5	LMS
	109.7-110.6	NRC PMT
2	5.1-10.7	LMS PMT
	112.9-114.1	CHL GNS
3	79.0-79.4	GNS RGU
	96.5-98.8	ANG HFD
	102.1-104.7	ANG HOL LMS RGU RMG
4	95.1-96.6	PMT RGU
5	32.7-33.8	LMS NRC
	79.4-80.6	LMS NRC RMG
	104.5-105.6	HOL LMS PMT
6	33.4-34.4	BSW CHL PMT RMG
	36.8-37.8	ANG HFD JER NRC RMG
	44.2-45.3	GNS HOL LMS NRC
7	14.6-15.3	JER
	37.4-39.1	JER RMG
8	57.1-57.7	BSW CHL JER
9	55.1-55.7	HFD PMT
10	13.6-15.2	PMT RMG
	53.3-53.4	CHL HOL LMS
11	27.7-28.1	NRC
	61.0-63.9	GNS HFD PMT
	68.9-71.6	GNS LMS RGU
13	9.6-10.5	CHL
	18.7-19.9	CHL HFD HOL PMT
	25.0-27.0	BSW GNS NRC RMG RGU
	41.8-42.8	BSW GNS NRC RGU
	71.5-73.8	BSW HOL NRC PMT
14	2.6-10.4	CHL GNS HFD LMS PMT
	15.5-16.9	CHL GNS LMS
	23.9-30.1	CHL LMS NRC RGU
	42.3-43.2	GNS HOL LMS NRC
	51.6-53.3	ANG CHL GNS HFD LMS NRC RGU RMG
	58.6-62.2	ANG HFD JER LMS NRC PMT RMG
16	3.0-3.1	CHL
	43.9-44.7	ANG HFD RGU
	66.2-66.3	LMS PMT
17	29.0-29.8	RGU
	63.0-63.5	ANG BSW GNS HFD

Table 4.8. Continued.

BTA	Position (Mb)	Breeds
18	14.9-15.0	RGU RMG
19	22.3-22.5	HFD
	24.8-25.4	GNS
20	18.2-19.6	BSW GNS LMS
	22.7-24.1	BSW HOL JER RMG
	32.0-33.1	HOL RMG
	44.3-44.7	HOL
22	29.1-29.2	LMS PMT
25	6.9-7.8	CHL HFD HOL JER LMS
	12.9-13.3	JER PMT RGU
28	26.6-28.4	CHL HFD NRC

[1] Summary includes those regions where multiple SNPs separated by <1Mb had |iHS| > 2.694 (top 1%).
[2]ANG = Angus, BSW = Brown Swiss, CHL = Charolais, GNS = Guernsey, HFD = Hereford, HOL = Holstein, JER = Jersey, LMS = Limousin, NRC = Norwegian Red, PMT = Piedmontese, RGU = Red Angus, and RMG = Romagnola.

More than half of the regions were identified in at most 2 breeds. Two regions on BTA 14 were identified in at least 7 breeds. There were 15 regions represented by only beef breeds and 4 were represented by only dairy breeds. Although the outlier approach has low power and a high false discovery rate (Akey et al., 2009), these regions where consecutive SNP have extreme values may be indicative of regions under divergent selection.

4.7 Annotation

The majority of the SNP that were assayed in this study were intergenic (Table 4.9). The sparcity of data associated with genes limits the amount of annotation that is possible. Gene ontology terms were obtained for all exonic SNP (Figure 4.16). Genes associated with metabolic processes were overrepresented in the dataset. An attempt was made to annotate the SNP with extreme values of population-average FST. Unfortunately, none of the SNP with extreme values coincided with genes. This makes automated characterization of the signatures of selection problematic. Each of the regions detected must be manually annotated, and this is yet to be done.

Population-specific F_{ST} and iHS are performed on a breed-by-breed basis so annotation of these regions is more computationally intensive. Given the lack of association of extreme values of population average F_{ST} with genes, characterization of extreme SNP for these other measures was not attempted. In order to be able to better characterize the signatures of selection in cattle, denser marker assays will be needed and ongoing annotation of the bovine genome sequence will be required. Villa-Angulo et al. (2009) suggested that ~580,000 SNP would be necessary to characterize the haplotype

Table 4.9. Categorization of SNP by association with RefSeq genes

Category	No. SNP
3’ UTR	214
5’ UTR	273
Exon	119
Intron	1,570
Intergenic	26,955

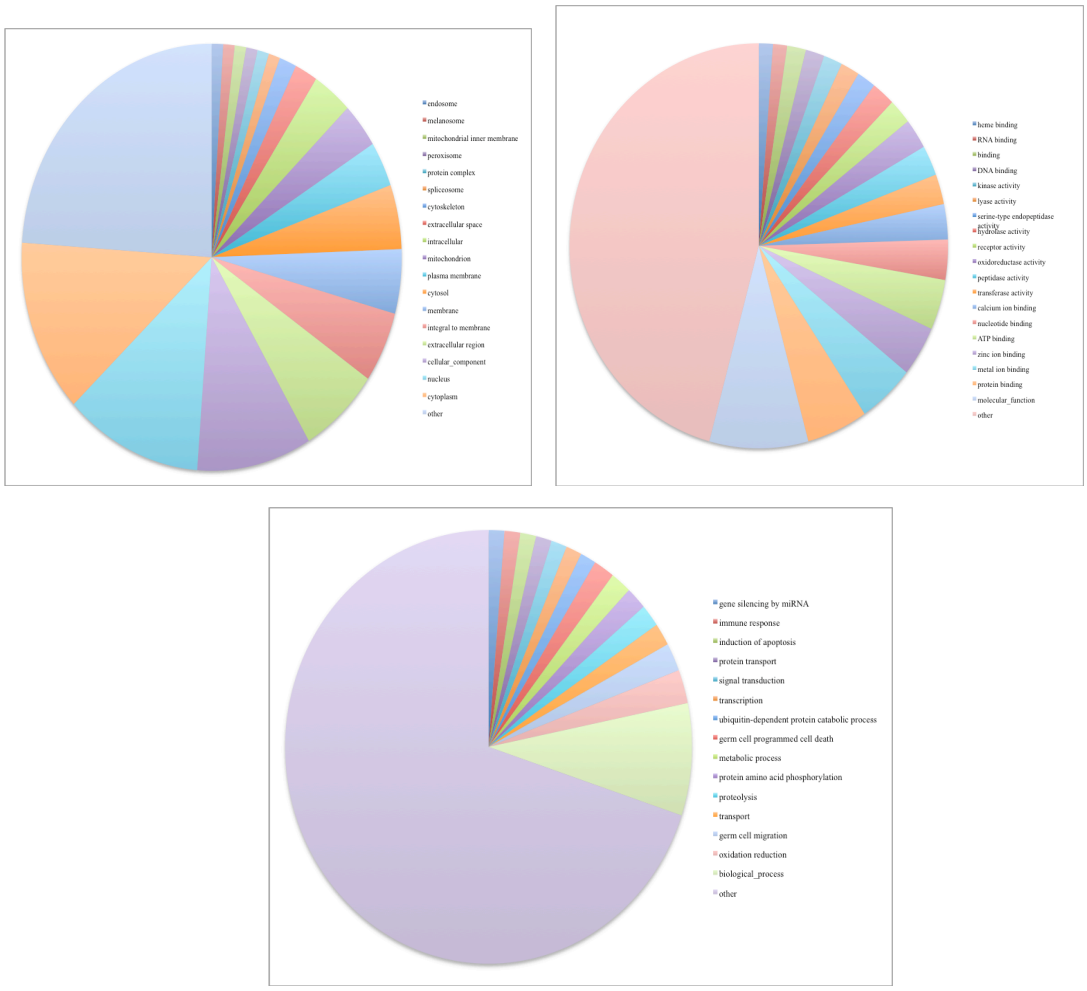


Figure 4.16. Gene ontology for all exonic SNP assayed. (a) Cellular Component, (b) Biological Process, and (c) Biological Function. GO terms for bovine genes were extracted from Entrez Gene (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene), which links to the GO Annotation database (http://www.ebi.ac.uk/GOA/).

block structure across the cattle genome. This 10-fold increase in density would make interpretation of results from a future genome-wide scan of signatures of selection easier.

5. SUMMARY AND CONCLUSIONS

The objectives of this study were to identify signatures of selection in *Bos taurus* beef and dairy cattle populations and to annotate regions of selection with gene, function and QTL information. Differences in minor allele frequencies, population-average F_{ST}, population-specific F_{ST}, and integrated extended haplotype homozygosity scores were applied to a subset of the bovine HapMap data to characterize signatures of selection in 7 *Bos taurus* beef and 5 *Bos taurus* dairy cattle populations.

Numerous SNP exhibited evidence of selection across the genome and we highlighted regions of BTA2 and BTA14 that are considered to be under positive selection in beef and dairy cattle, respectively. The current density of SNP limited our ability to annotate regions putatively under selection because most SNP in the assay were intergenic. This is likely because of the between-breed SNP discovery method that was used, which typically identifies SNP with higher allele frequencies.

A challenge of performing genome-wide scans for signatures of selection is determining the thresholds for significance. For F_{ST}, Wright's qualitative guidelines for interpretation was used to identify outliers, whereas for difference in MAF permutation was employed and for iHS the top 1% of values were taken. As discussed by the Bovine Hapmap Consortium (2009) statistical significance for a genome-wide scan for signatures of selection can only be assessed by generating an empirical null distribution from simulations that capture features of the data that are unrelated to selection. This requires sophisticated statistical models for simulation and as discussed by Akey (2009), none of the existing models adequately account for all of the parameters that would contribute to

the null distribution. This is an area of research that must be pursued in the future. Because the outlier approach that was used herein is known to suffer from low power and high false discovery rates, our identification of genomic regions that have been subjected to recent selection should be considered tentative. However, integrating the data from the various genome-wide scans for signatures of selection with other QTL, mapping or gene expression data we could, in future studies, begin to characterize those regions where multiple sources of evidence suggest that the region is functionally important for adaptive phenotypes.

LITERATURE CITED

AIPL. 2009. Trend in milk bv for holstein or red & white: Calculated April 2009. http://aipl.Arsusda.Gov/eval/summary/trend.cfm Accessed April 25, 2009.

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

Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver. 2002. Interrogating a high-density snp map for signatures of natural selection. Genome Res. 12: 1805-1814.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: Tool for the unification of biology. Nat. Genet. 25: 25-29.

Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich, and J. N. Hirschhorn. 2004. Genetic signatures of strong recent positive selection at the lactase gene. The American Journal of Human Genetics 74: 1111-1120.

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

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.

The Bovine HapMap Consortium. 2009. Genome-wide survey of snp variation uncovers the genetic structure of cattle breeds. Science 324: 528-532.

Dekkers, J. C. M., and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3: 22-32.

Duret, L. 2008. Neutral theory: The null hypothesis of molecular evolution. Nature Education 1: [5].

Elam, T. E., and R. L. Preston. 2004. Fifty years of pharmaceutical technology and its impact on the beef we provide to consumers. http://www.Feedstuffsfoodlink.Com/media/mediamanager/whitepaper-summary.pdf Accessed April 25, 2009.

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

Hurst, L. D. 2009. Fundamental concepts in genetics: Genetics and the understanding of selection. Nat. Rev. Genet. 10: 83-93.

The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437: 1299-1320.

The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million snps. Nature 449: 851-861.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217: 624-626.

Kryazhimskiy, S., and J. B. Plotkin. 2008. The population genetics of dn/ds. PLoS Genet. 4: e1000304.

Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, M. Kamal, M. Clamp, J. L. Chang, E. J. Kulbokas, M. C. Zody, E. Mauceli, X. Xie, M. Breen, R. K. Wayne, E. A. Ostrander, C. P. Ponting, F. Galibert, D. R. Smith, P. J. deJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C.-W. Chin, A. Cook, J. Cuff, M. J. Daly, D. DeCaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K.-P. Koepfli, H. G. Parker, J. P. Pollinger, S. M. J. Searle, N. B. Sutter, R. Thomas, C. Webber, and E. S. Lander. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803-819.

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

McVean, G., P. Awadalla, and P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160: 1231-1241.

Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418-426.

Nielsen, R., I. Hellmann, M. Hubisz, C. Bustamante, and A. G. Clark. 2007. Recent and ongoing selection in the human genome. Nat. Rev. Genet. 8: 857-868.

Ostrander, E. A., and R. K. Wayne. 2005. The canine genome. Genome Res. 15: 1706-1716.

Polineni, P., P. Aragonda, S. Xavier, R. Furuta, and D. Adelson. 2006. The bovine QTL viewer: A web accessible database of bovine quantitative trait loci. BMC Bioinformatics 7: 283.

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

Rat Genome Sequencing Project Consortium. 2004. Genome sequence of the brown norway rat yields insights into mammalian evolution. Nature 428: 493-521.

Roth, C., S. Rastogi, L. Arvestad, K. Dittmar, S. Light, D. Ekman, and D. A. Liberles. 2007. Evolution after gene duplication: Models, mechanisms, sequences, systems, and organisms. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 308B: 58-73.

Stinchcombe, J. R., and H. E. Hoekstra. 2007. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. Heredity 100: 158-170.

Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 16: 1315-1328.

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. PLoS Biology 4: e72.

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

Weir, B. S., and W. G. Hill. 2002. Estimating f-statistics. Annual Review of Genetics 36: 721-750.

Wright, S. 1943a. An analysis of local variability of flower color in linanthus parryae. Genetics 28: 139-156.

Wright, S. 1943b. Isolation by distance. Genetics 28: 114-138.

Wright, S. 1978. Evolution and the genetics of population. Vol. 4. Variability among and within natural populations. The University of Chicago Press, Chicago.

APPENDIX A

Differences in minor allele frequencies generated using 0.5 Mb sliding windows for each autosome are presented in Figures A.1 to A.4. Horizontal red lines indicate upper and lower 5th percentile determined by 1,000 permutations. Each SNP is plotted with respect to the coordinates in build BTAu4.0.

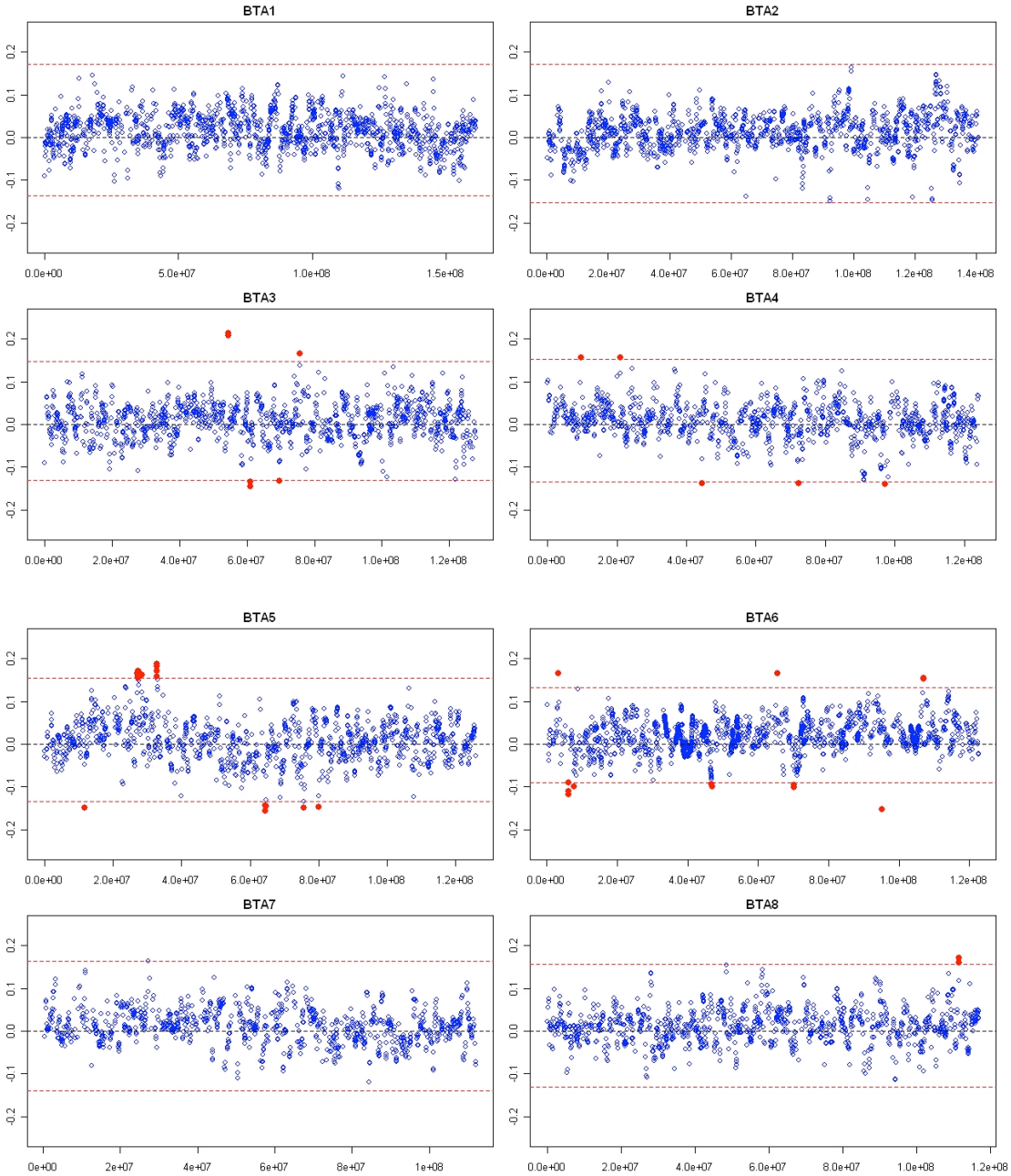


Figure A.1. Differences in MAF between beef and dairy on BTA1 to BTA8.

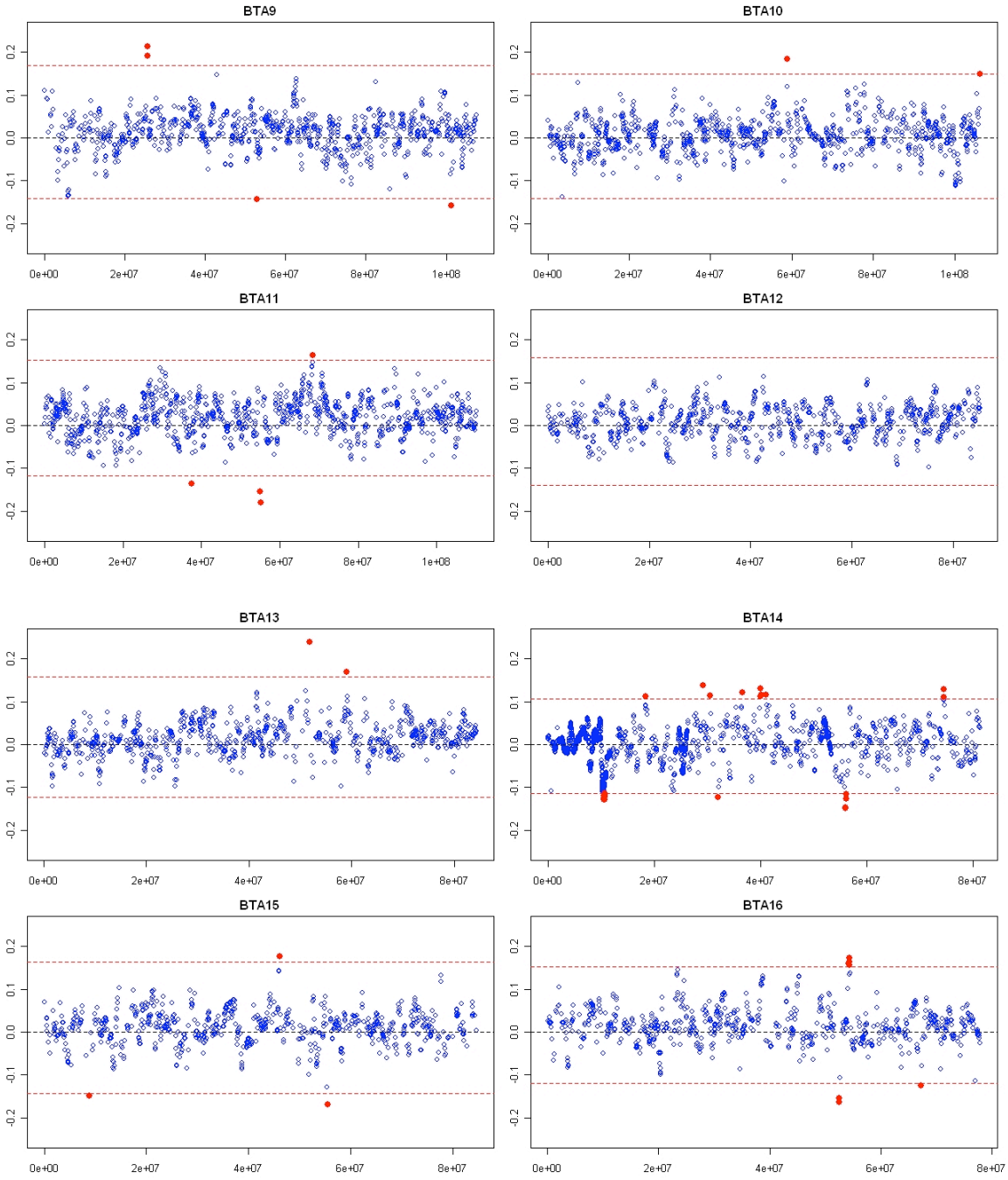


Figure A.2. Differences in MAF between beef and dairy on BTA9 to BTA16.

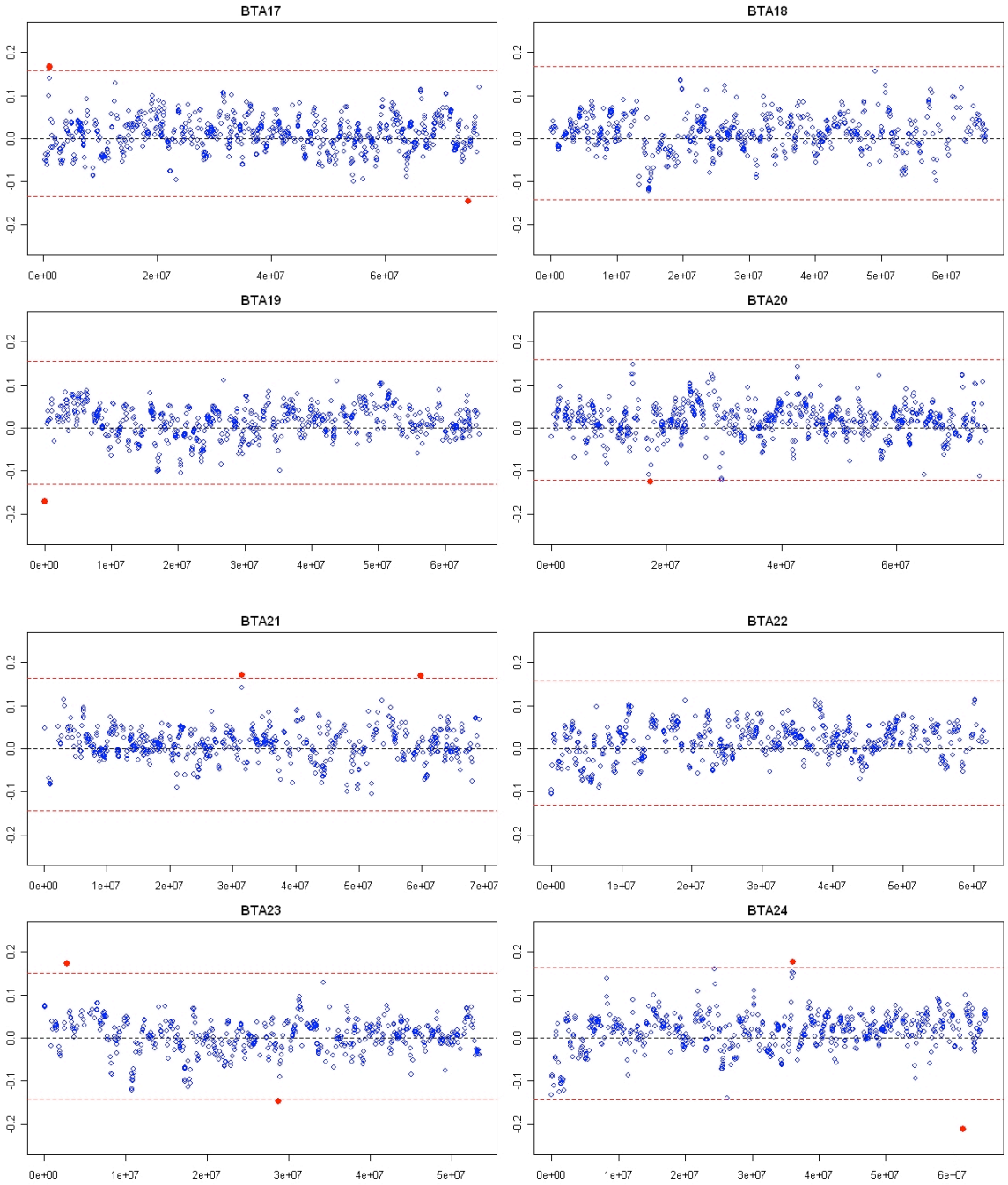


Figure A.3. Differences in MAF between beef and dairy on BTA17 to BTA24.

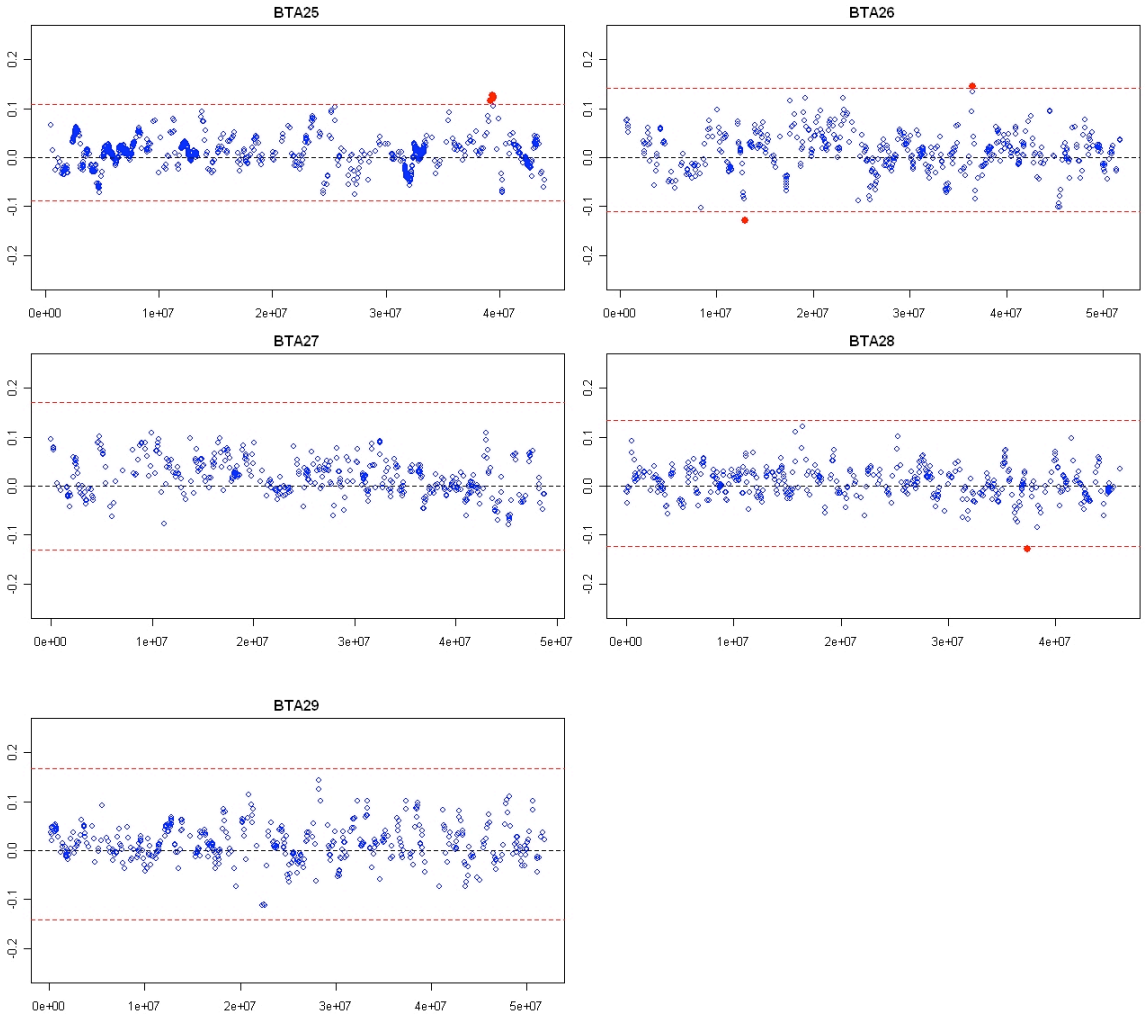


Figure A.4. Differences in MAF between beef and dairy on BTA25 to BTA29.

APPENDIX B

Population-average F_{ST} values generated using 1 Mb sliding windows for each autosome are presented in Figures B.1 to B.8. Horizontal lines indicate the 0.05 (moderate differentiation), 0.15 (great differentiation) and 0.25 (very great differentiation) thresholds that are commonly used for qualitative interpretation of Wright's fixation index. Each SNP is plotted with respect to the coordinates in build BTAu4.0.

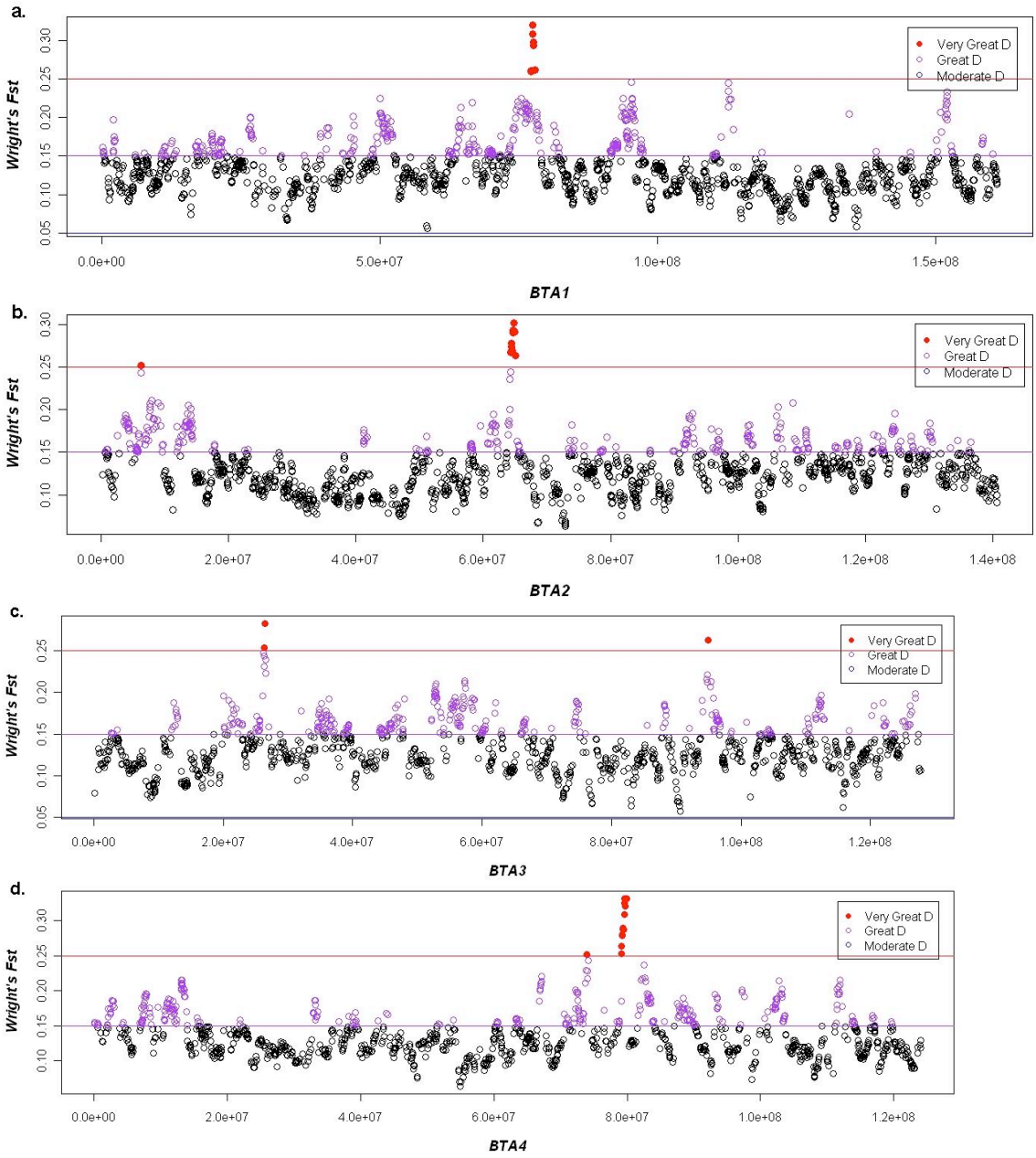


Figure B.1. Distribution of F_{ST} values across BTA1, 2, 3, and 4.

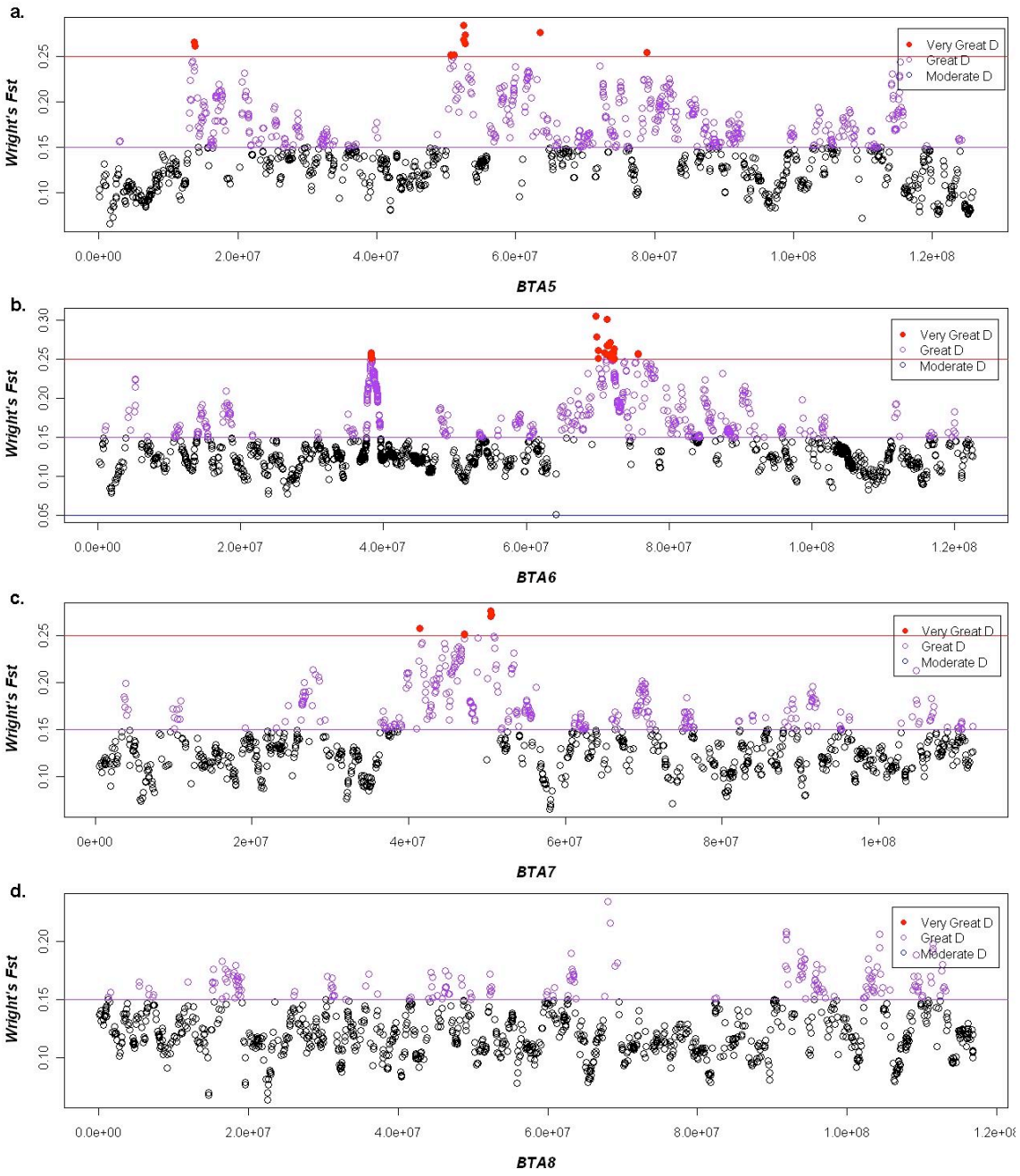


Figure B.2. Distribution of F_{ST} values across BTA5, 6, 7, and 8.

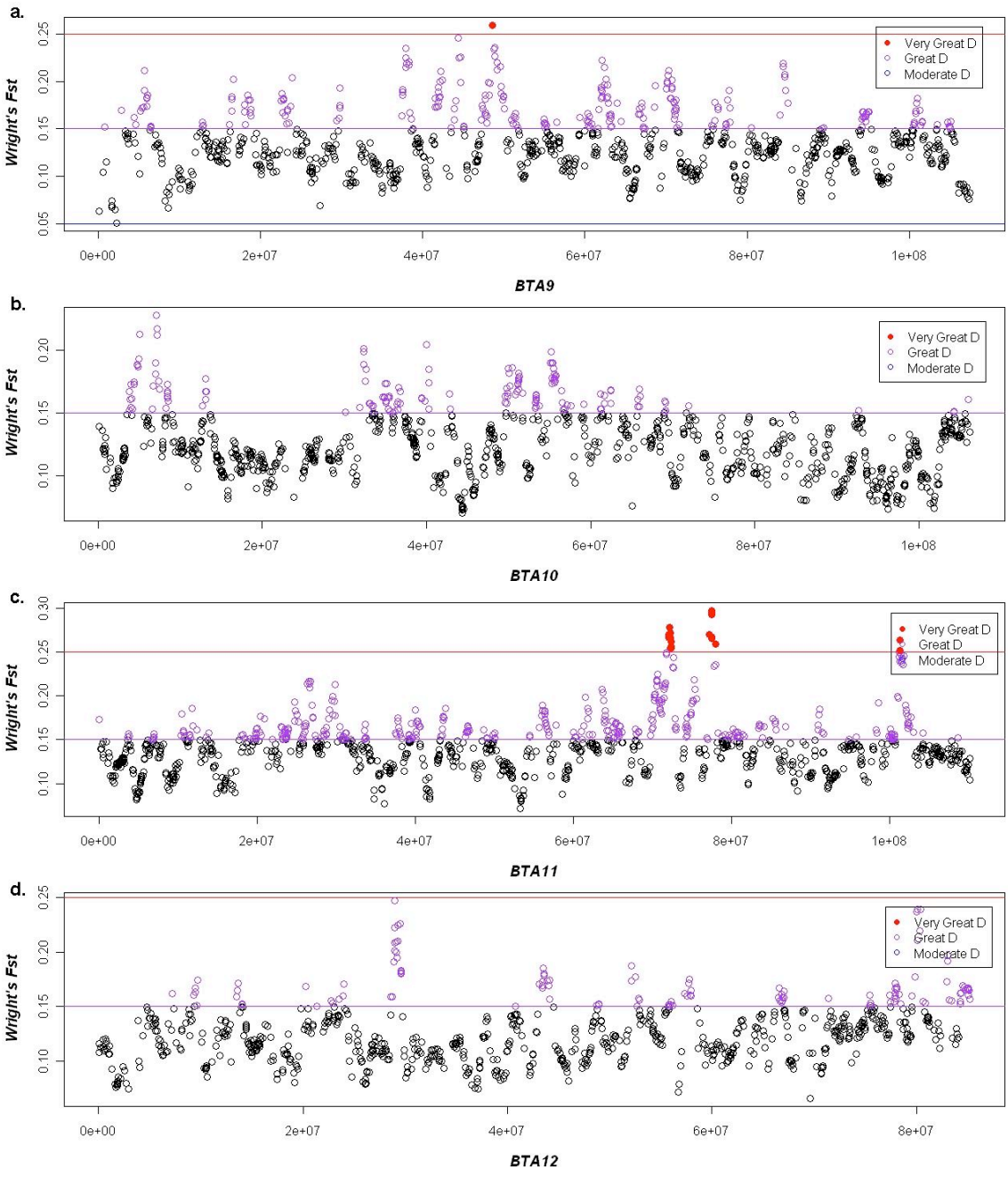


Figure B.3. Distribution of F_{ST} values across BTA9, 10, 11, and 12.

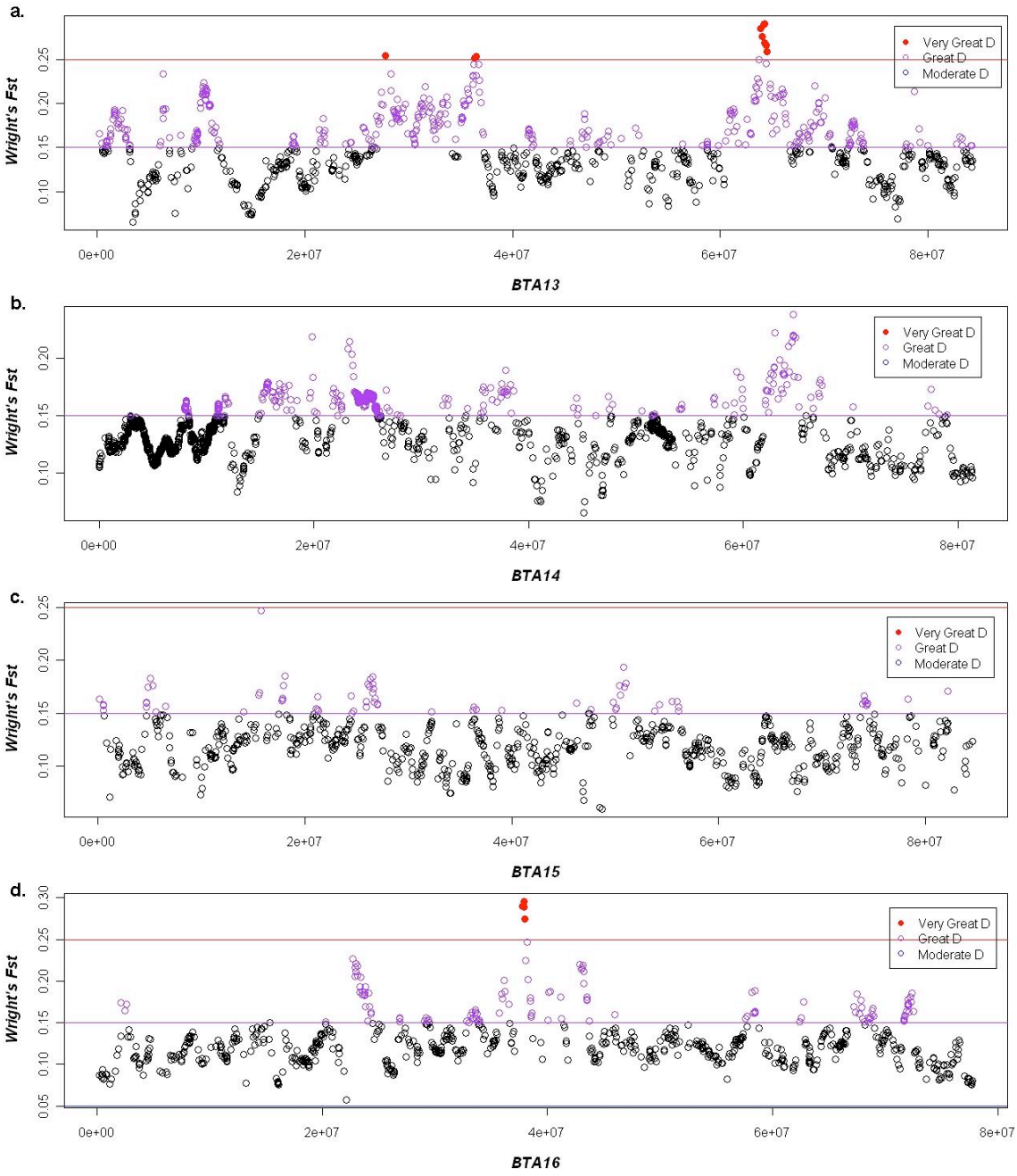


Figure B.4. Distribution of F_{ST} values across BTA13, 14, 15, and 16.

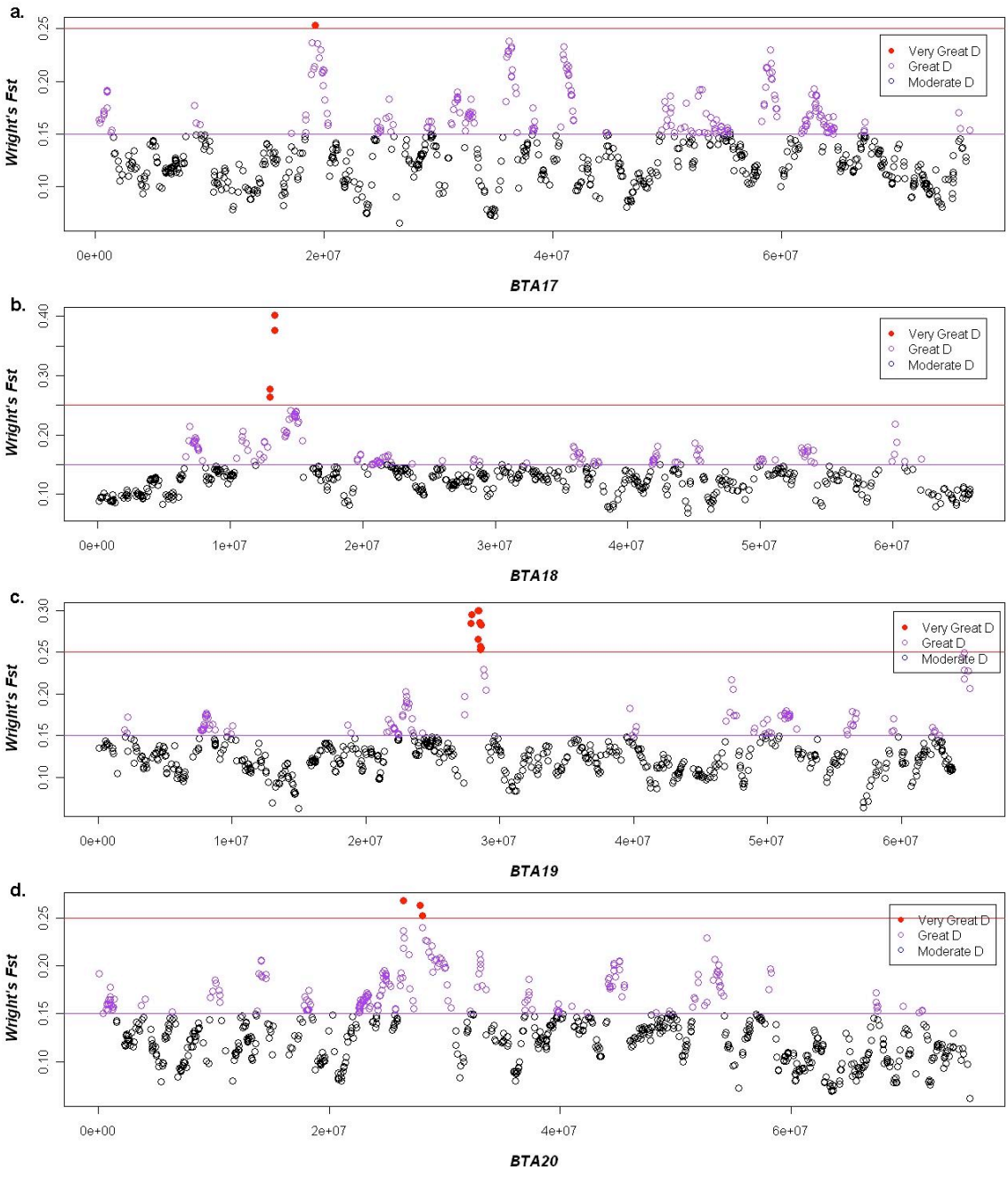


Figure B.5. Distribution of F_{ST} values across BTA17, 18, 19, and 20.

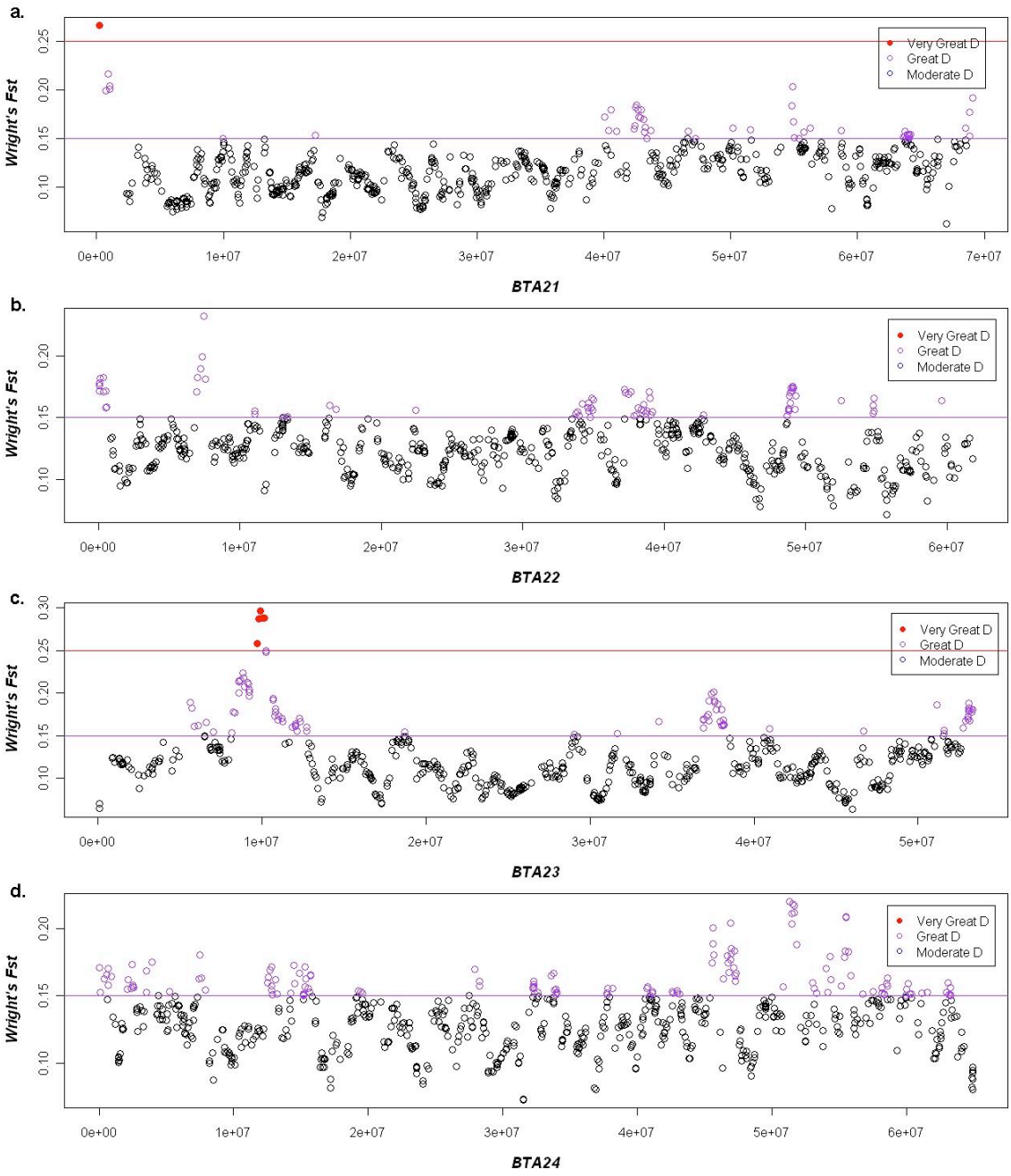


Figure B.6. Distribution of F_{ST} values across BTA21, 22, 23, and 24.

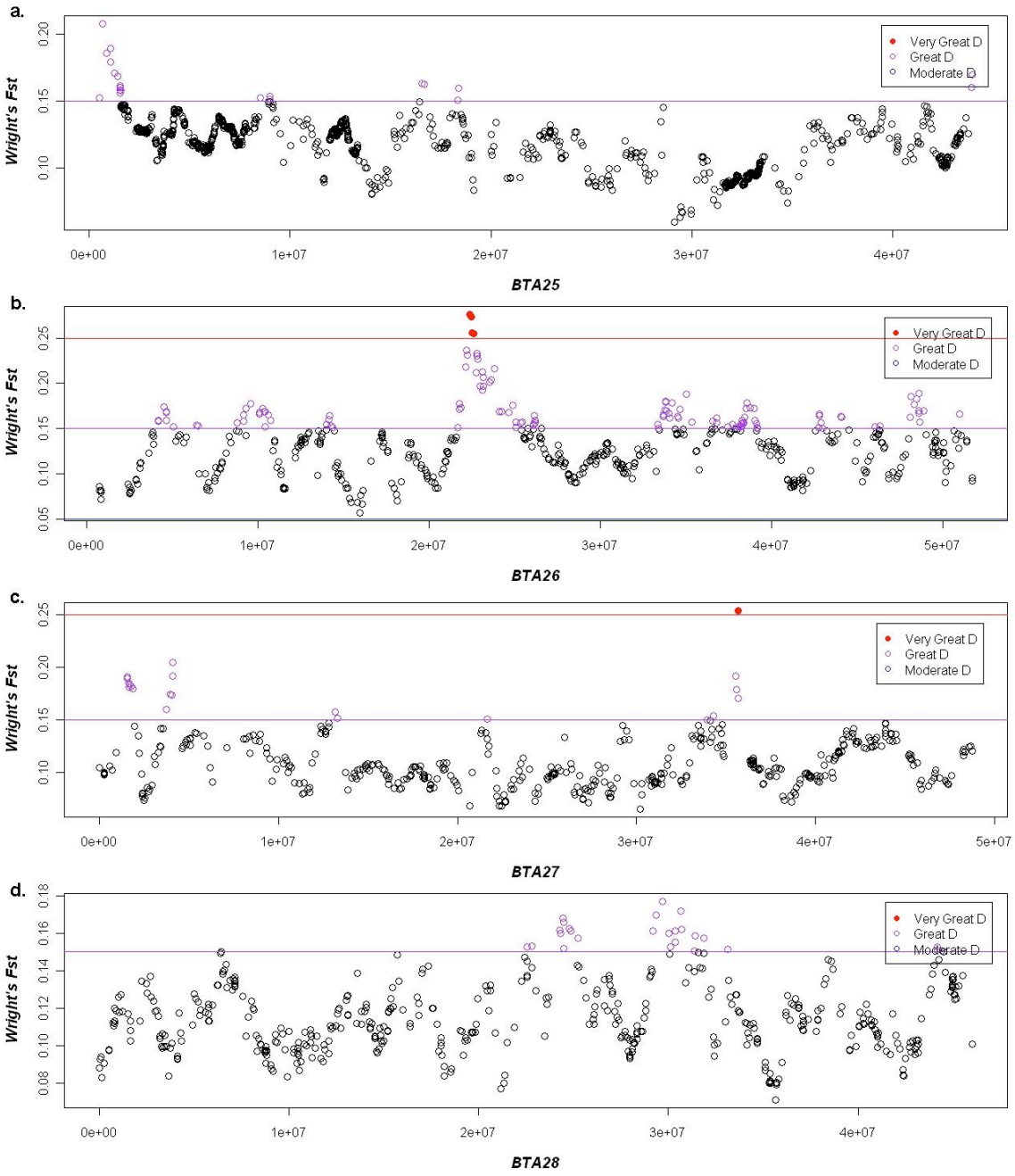


Figure B.7. Distribution of F_{ST} values across BTA25, 26, 27, and 28.

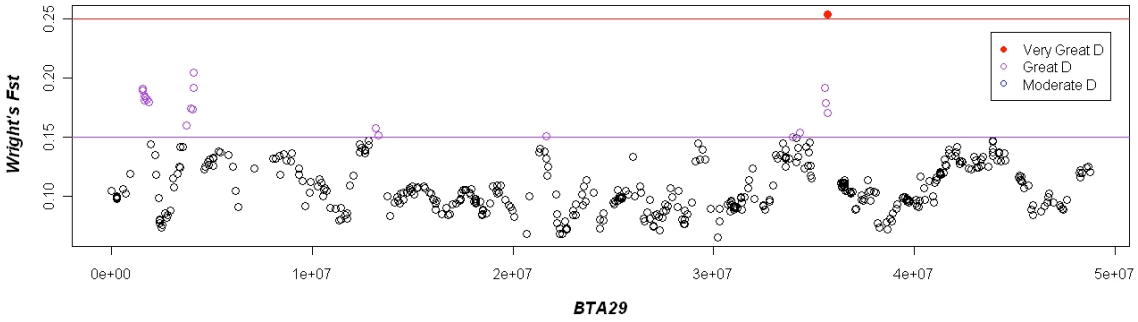


Figure B.8. Distribution of F_{ST} values across 29.

APPENDIX C

Population-specific F_{ST} values generated using 5 Mb sliding windows for each of the 12 breeds are presented in Figures C.1 to C.29. Horizontal lines indicate the 0.05, 0.15 and 0.25 thresholds that are commonly used for qualitative interpretation of Wright's fixation index. Each SNP is plotted with respect to the coordinates in build BTAu4.0.

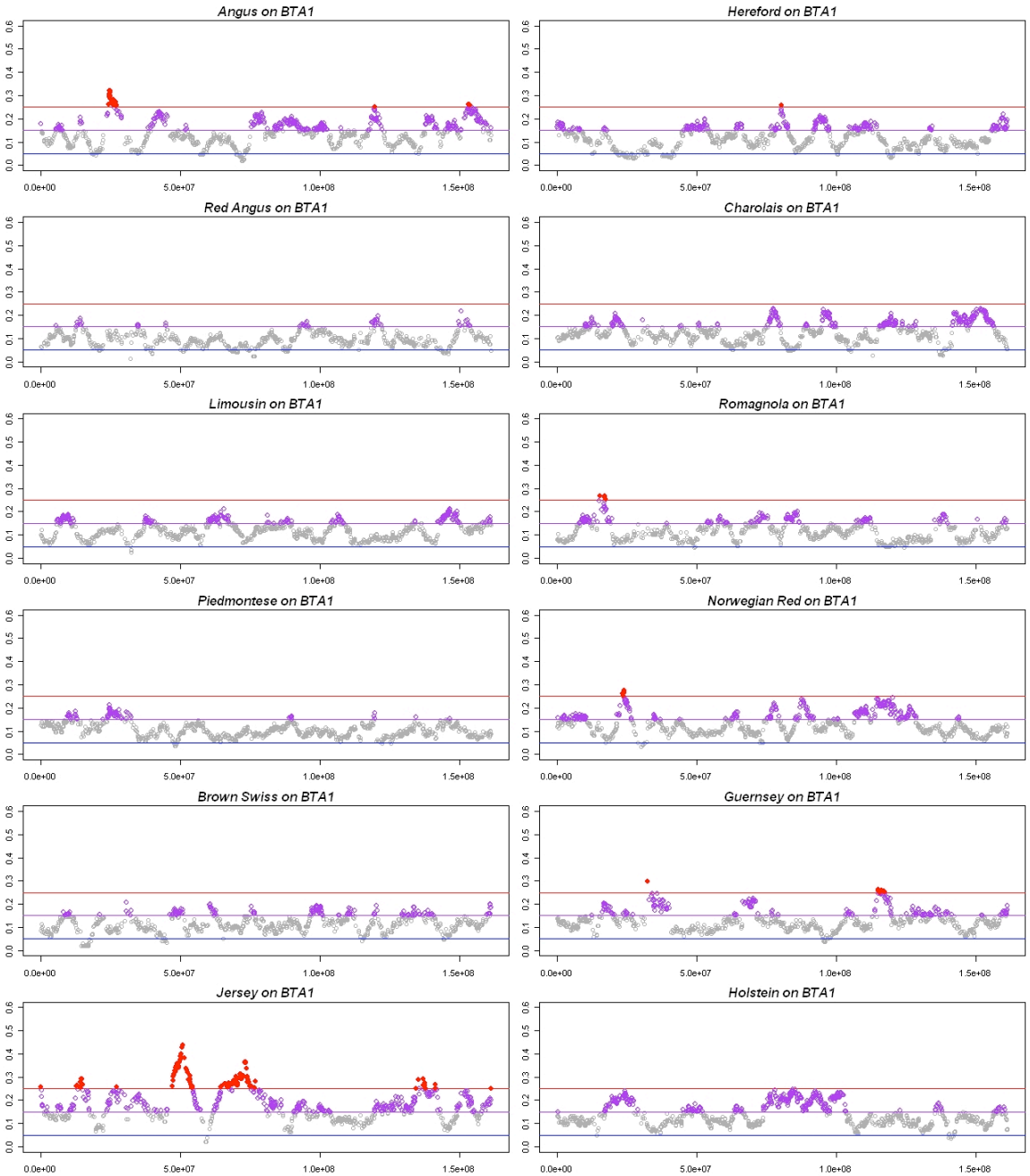


Figure C.1. Population-specific F_{ST} values for BTA1.

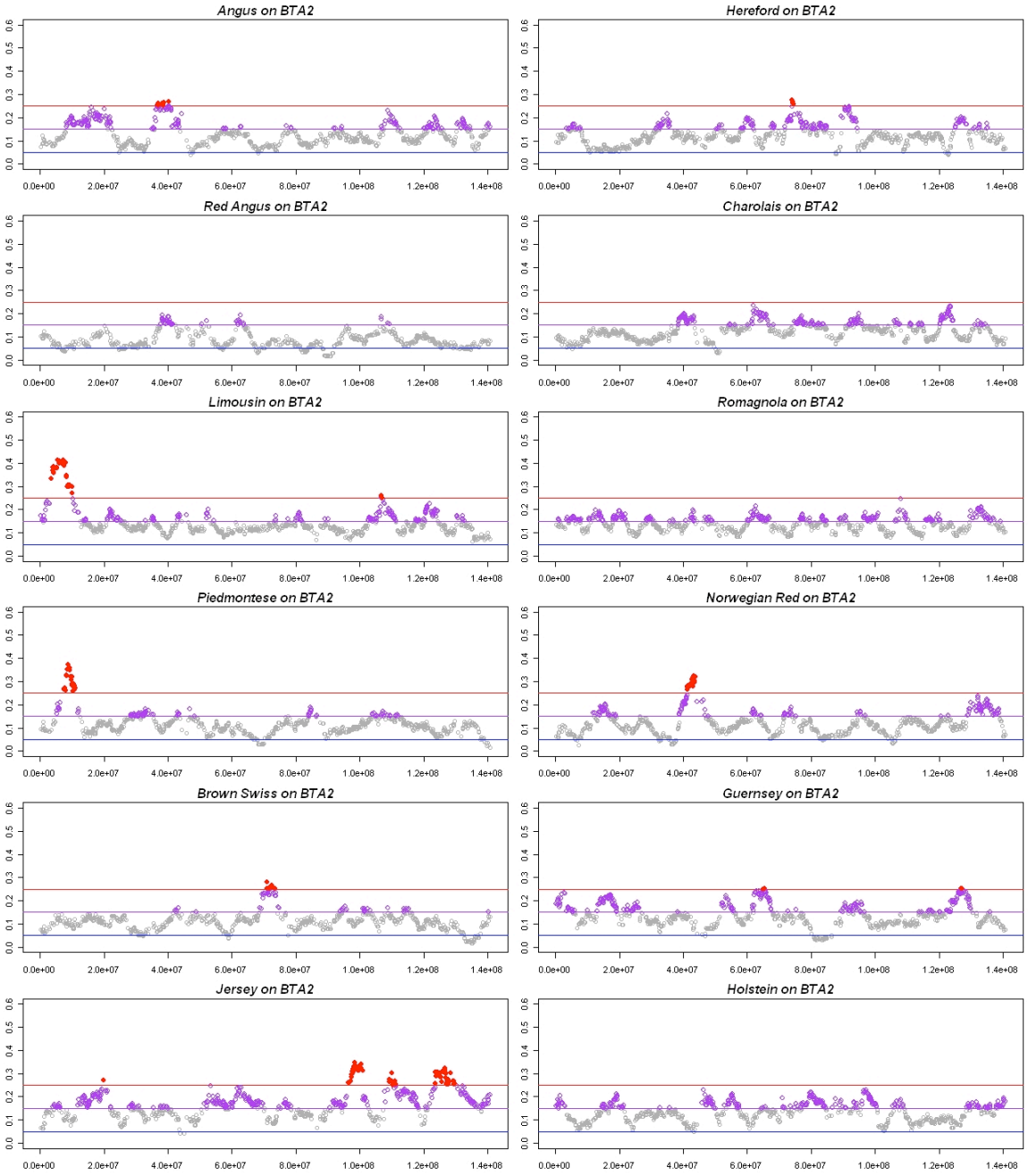


Figure C.2. Population-specific F_{ST} values for BTA2.

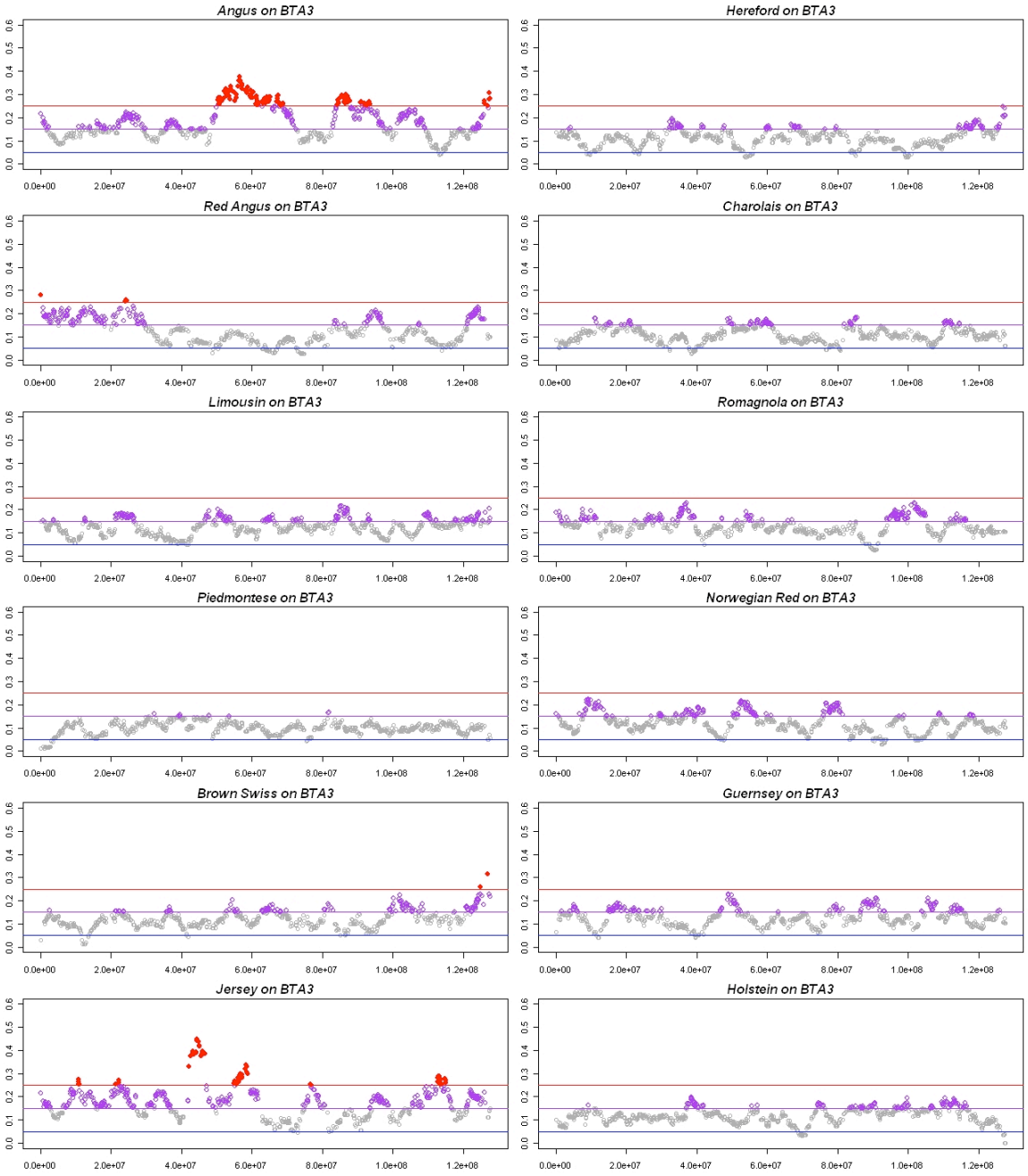


Figure C.3. Population-specific F_{ST} values for BTA3.

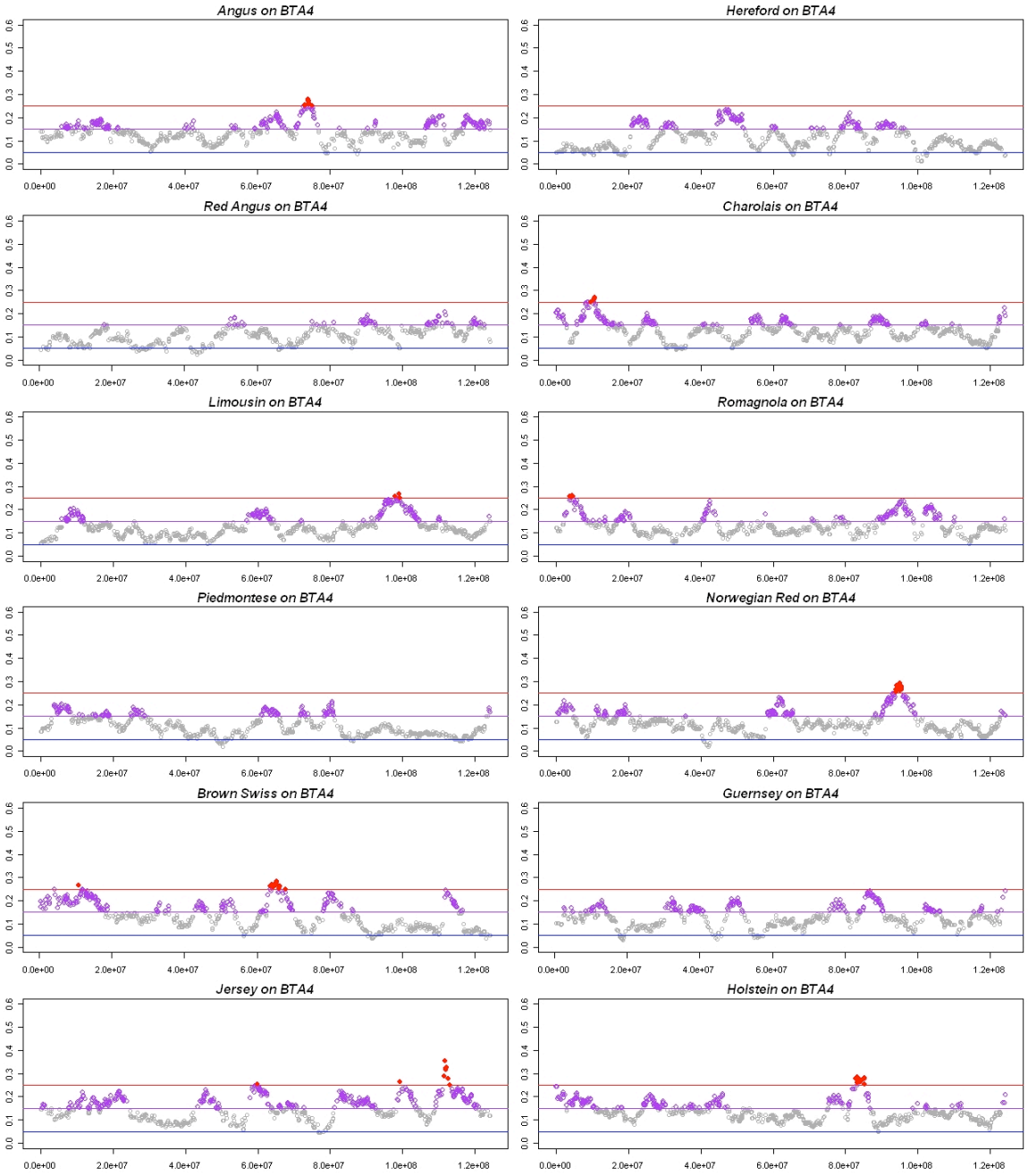


Figure C.4. Population-specific F_{ST} values for BTA4.

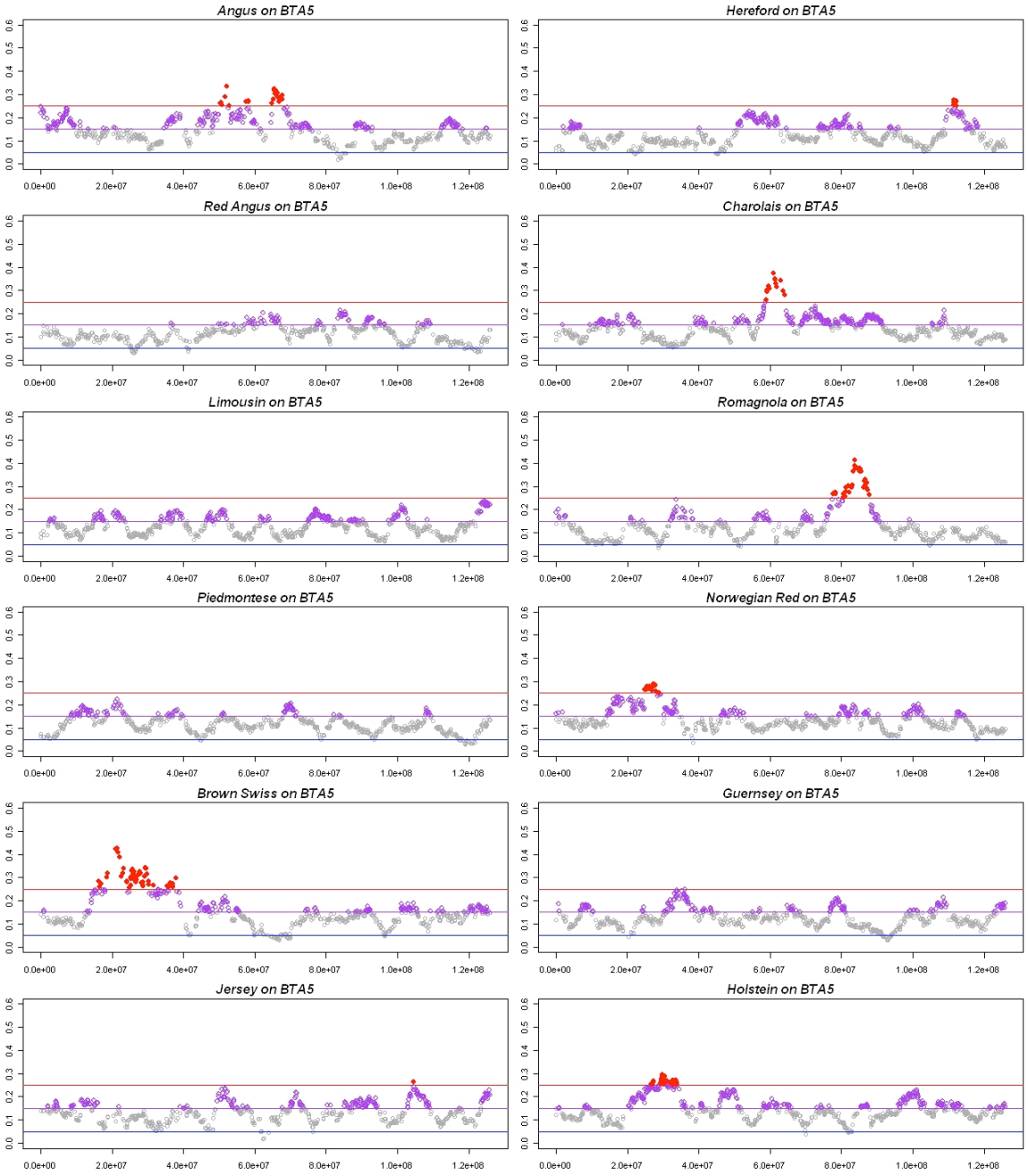


Figure C.5. Population-specific F_{ST} values for BTA5.

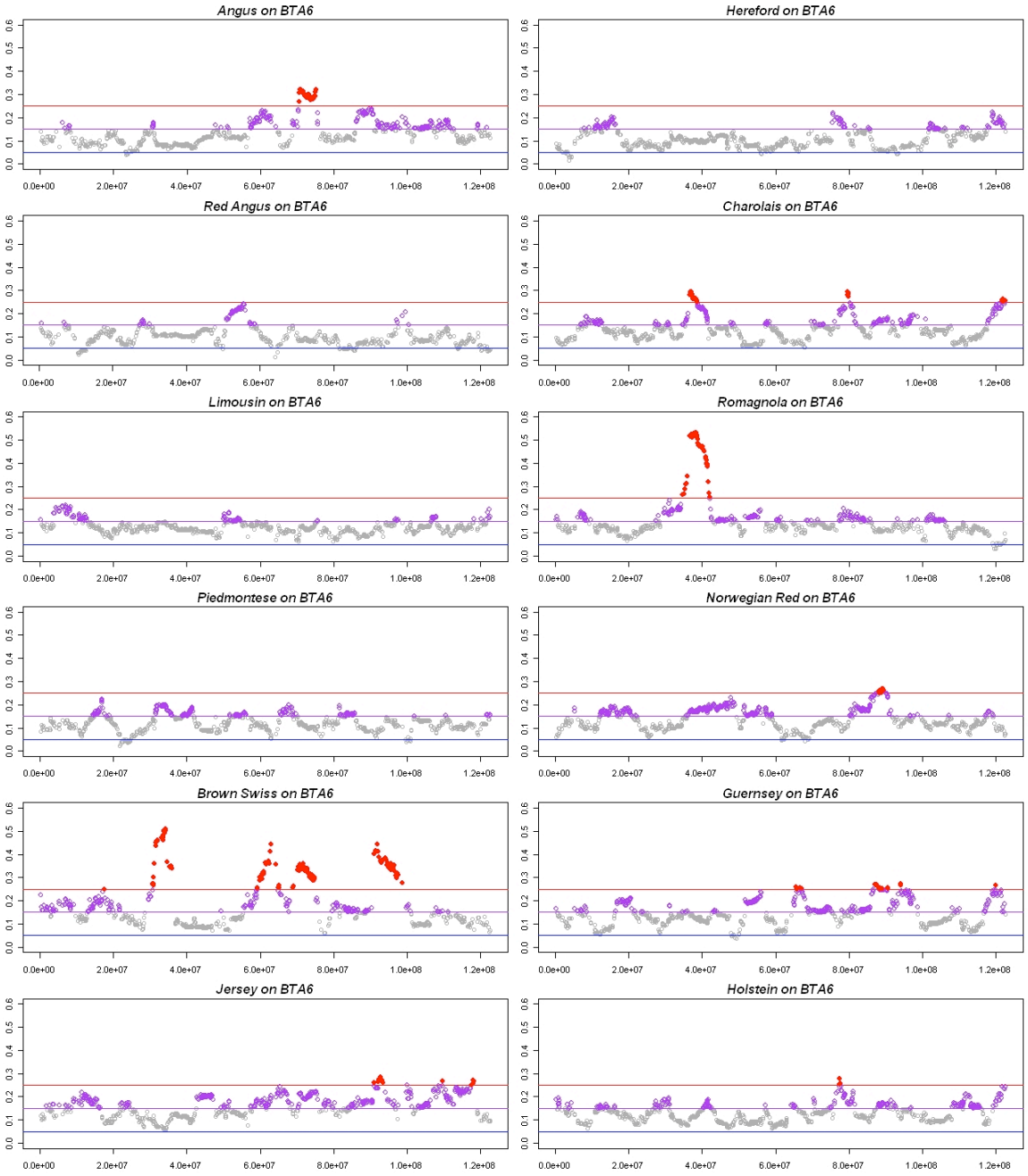


Figure C.6. Population-specific F_{ST} values for BTA6.

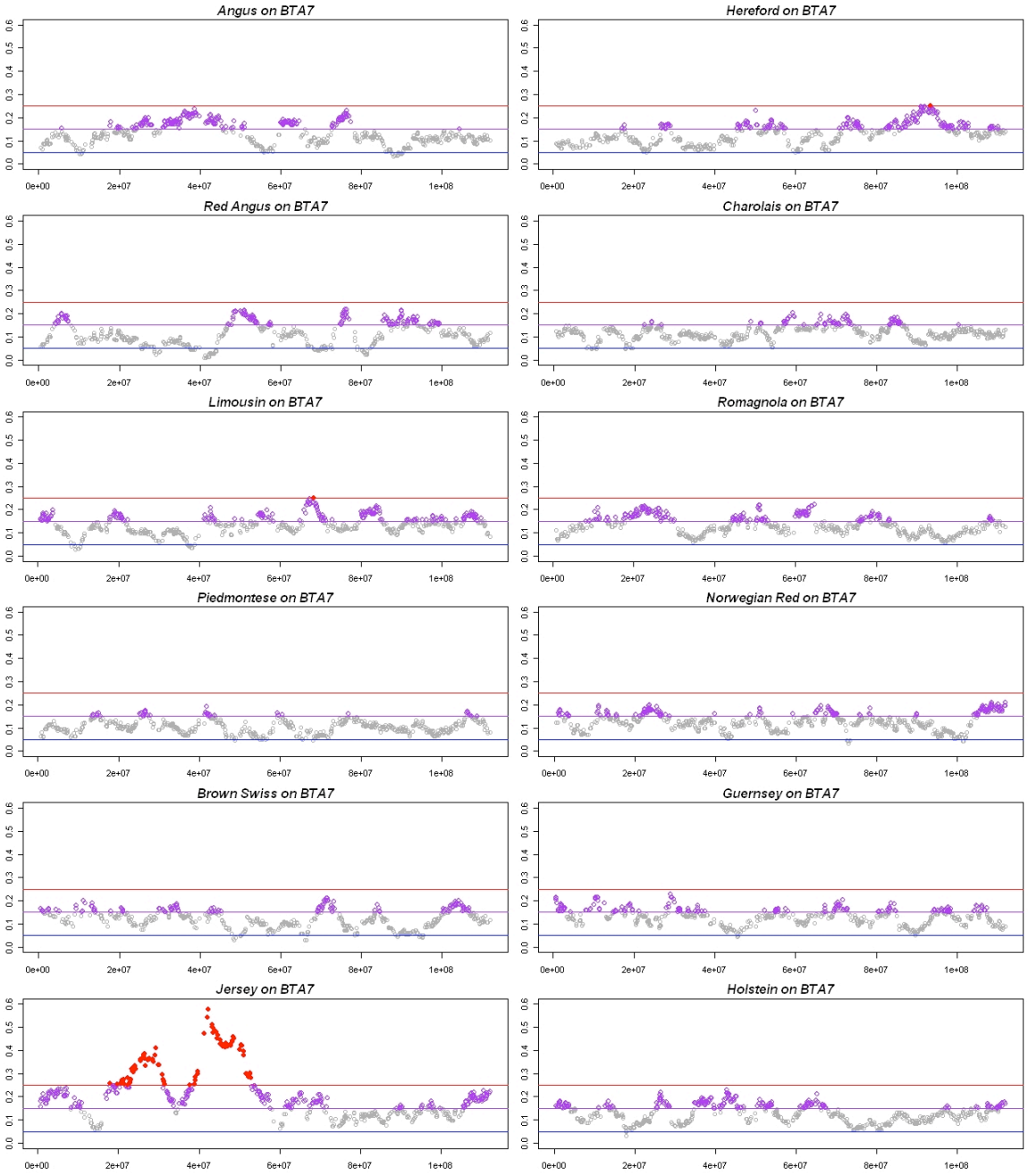


Figure C.7. Population-specific F_{ST} values for BTA7.

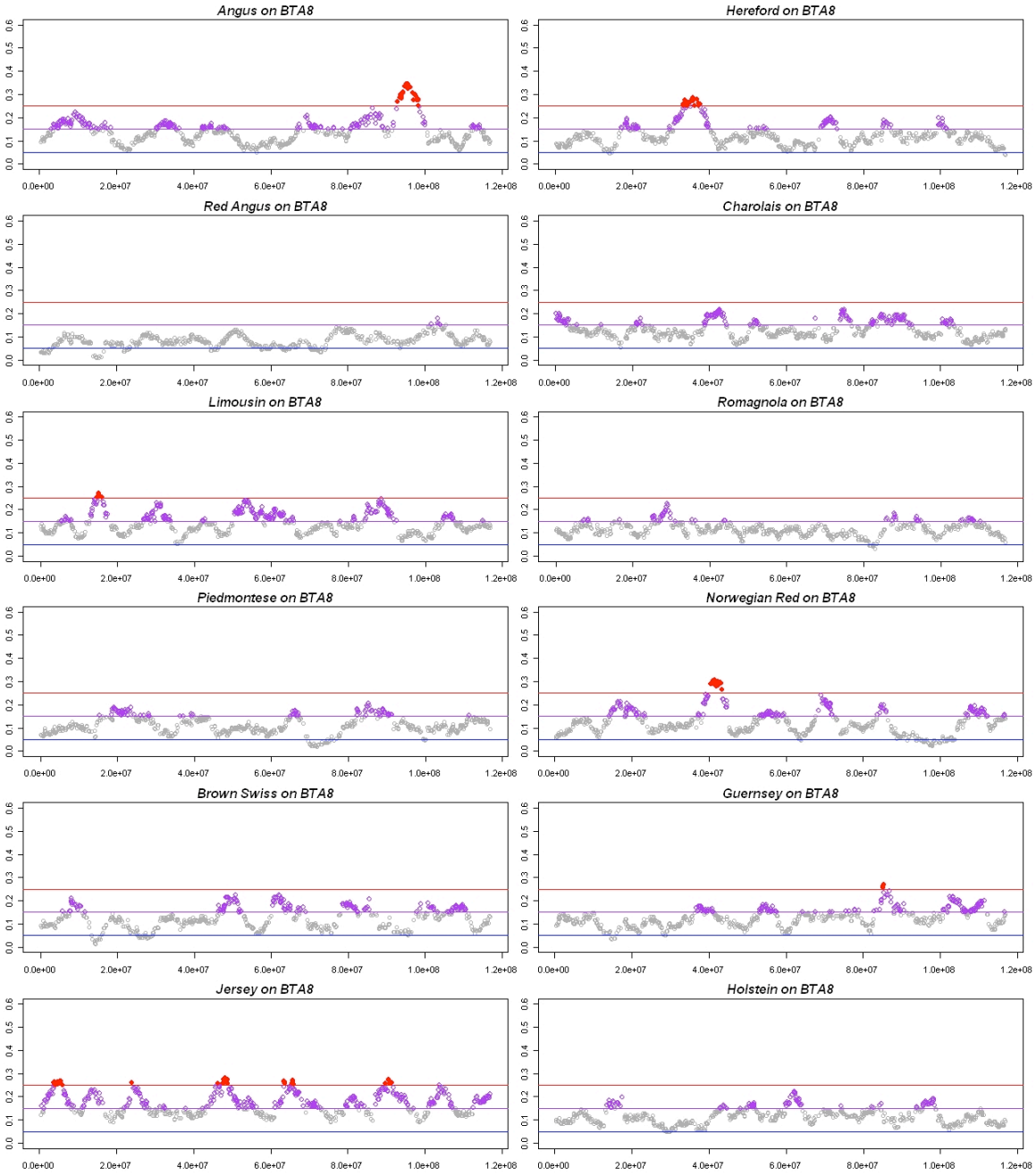


Figure C.8. Population-specific F_{ST} values for BTA8.

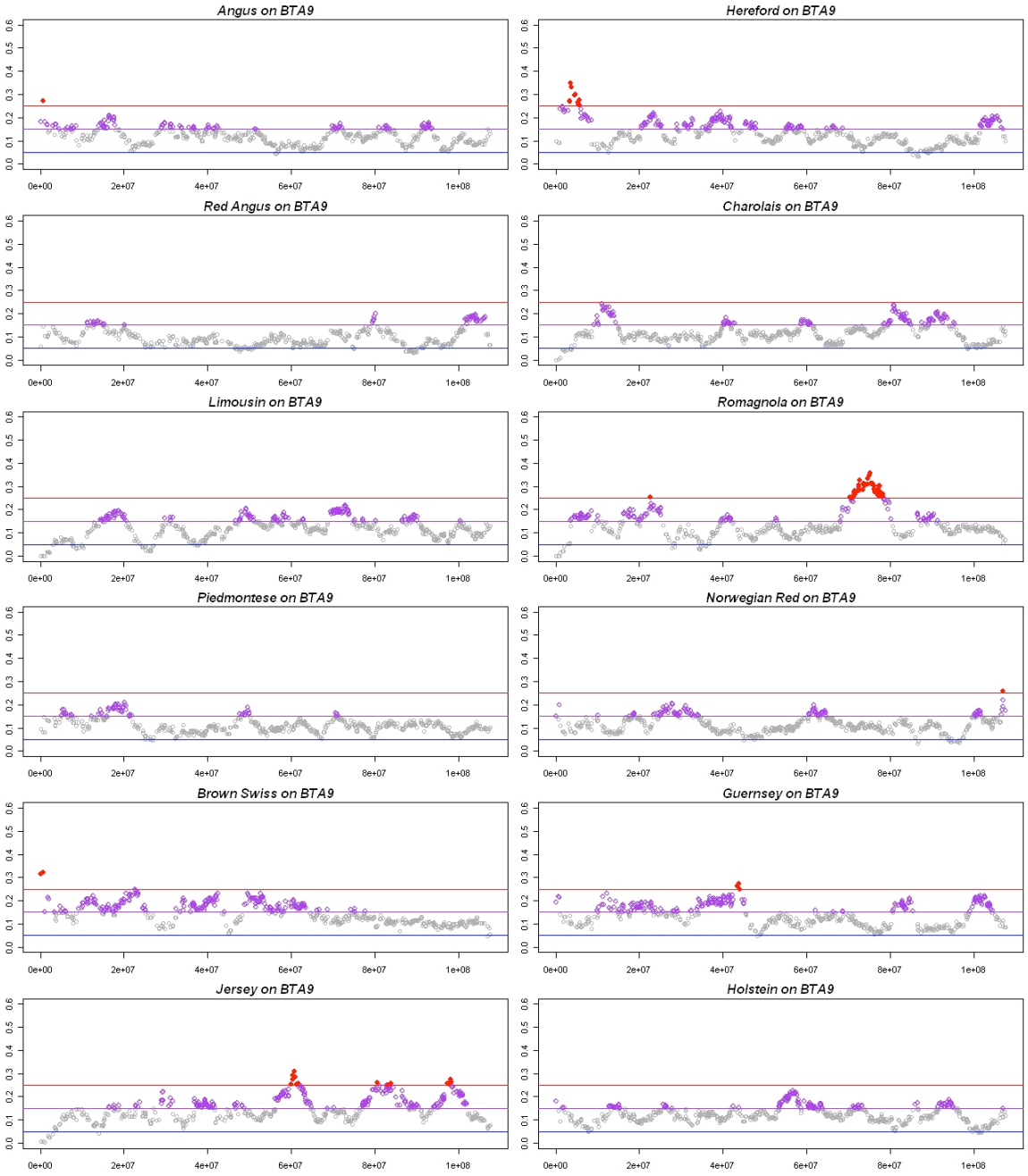


Figure C.9. Population-specific F_{ST} values for BTA9.

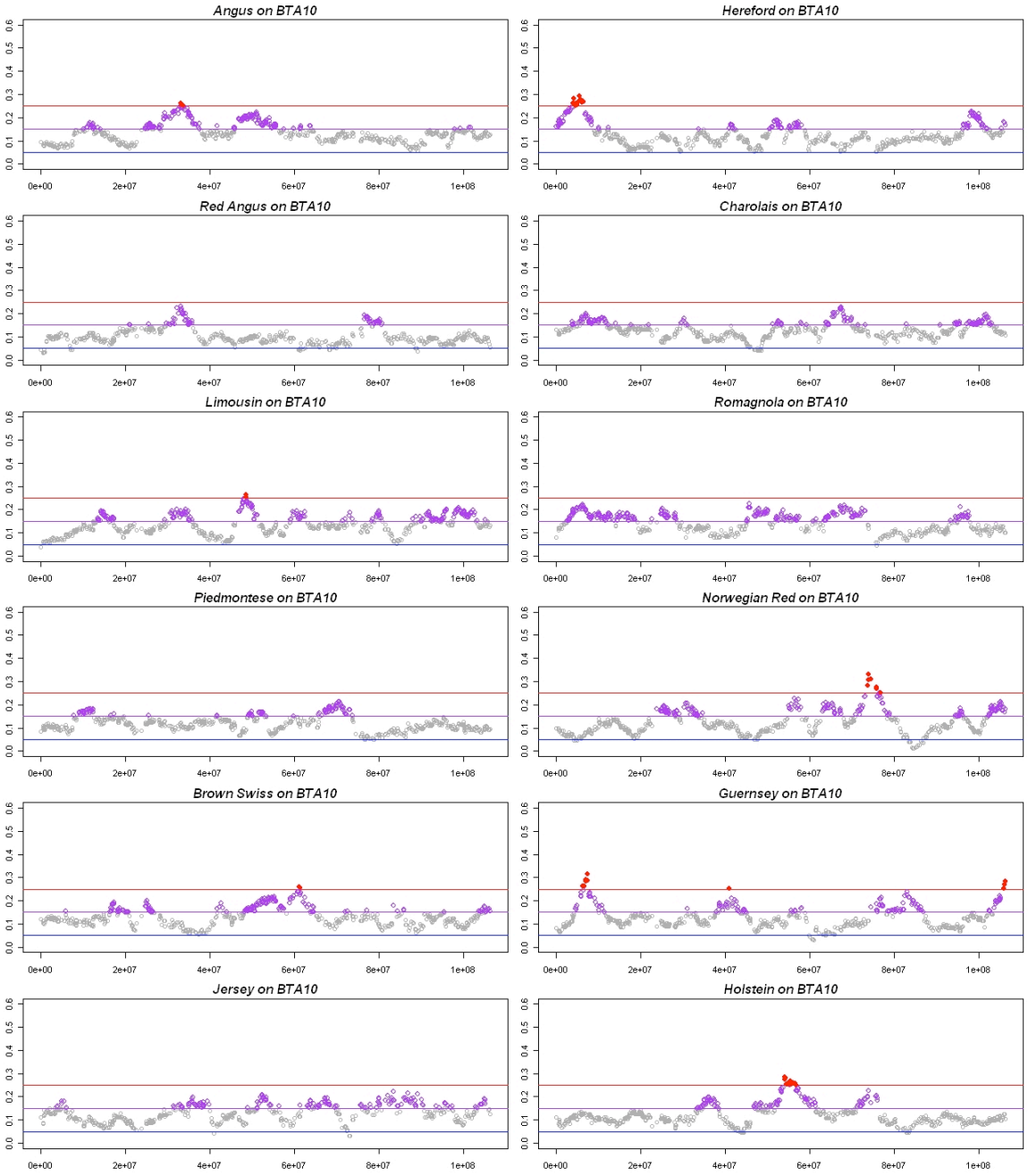


Figure C.10. Population-specific F_{ST} values for BTA10.

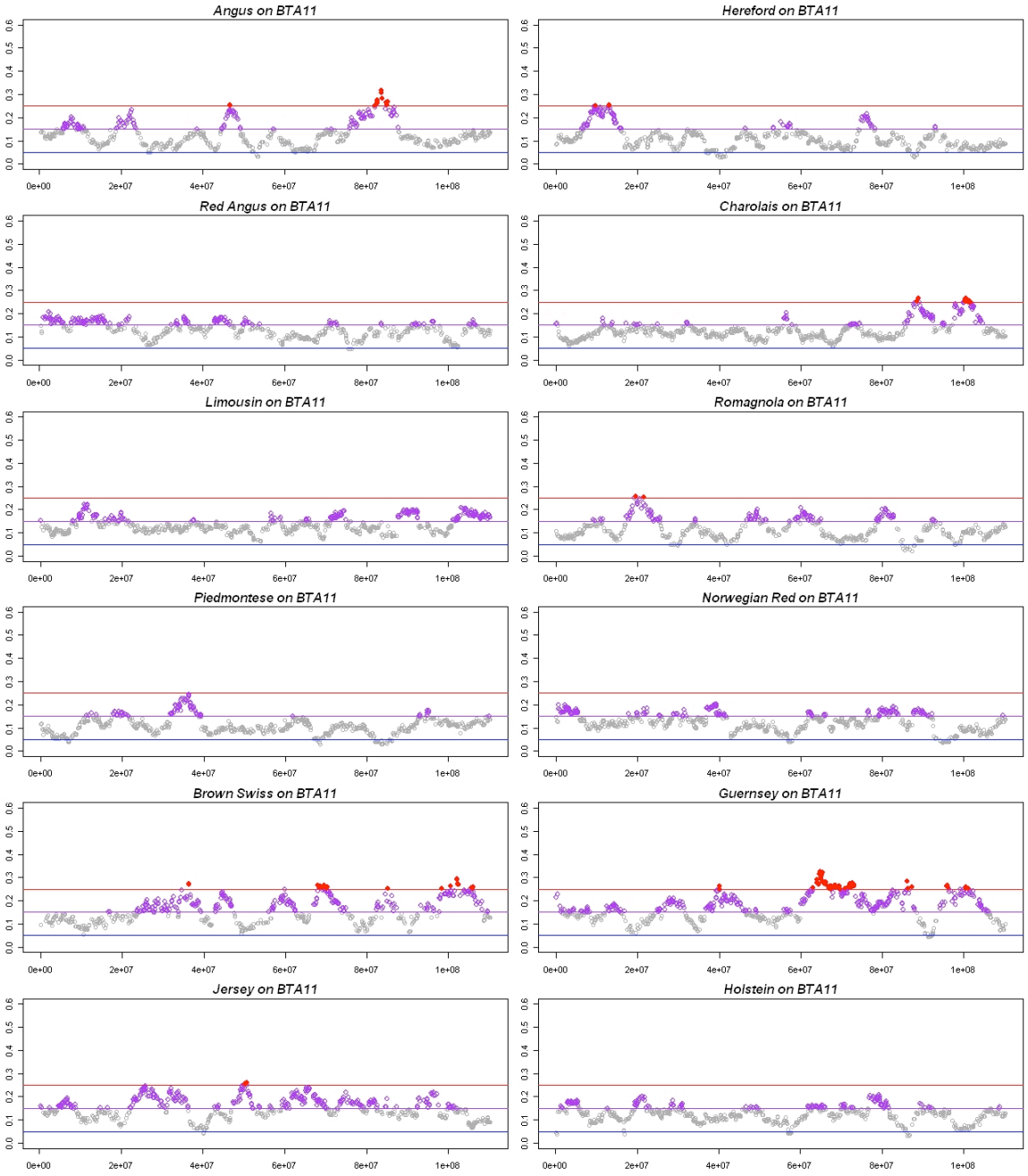


Figure C.11. Population-specific F_{ST} values for BTA11.

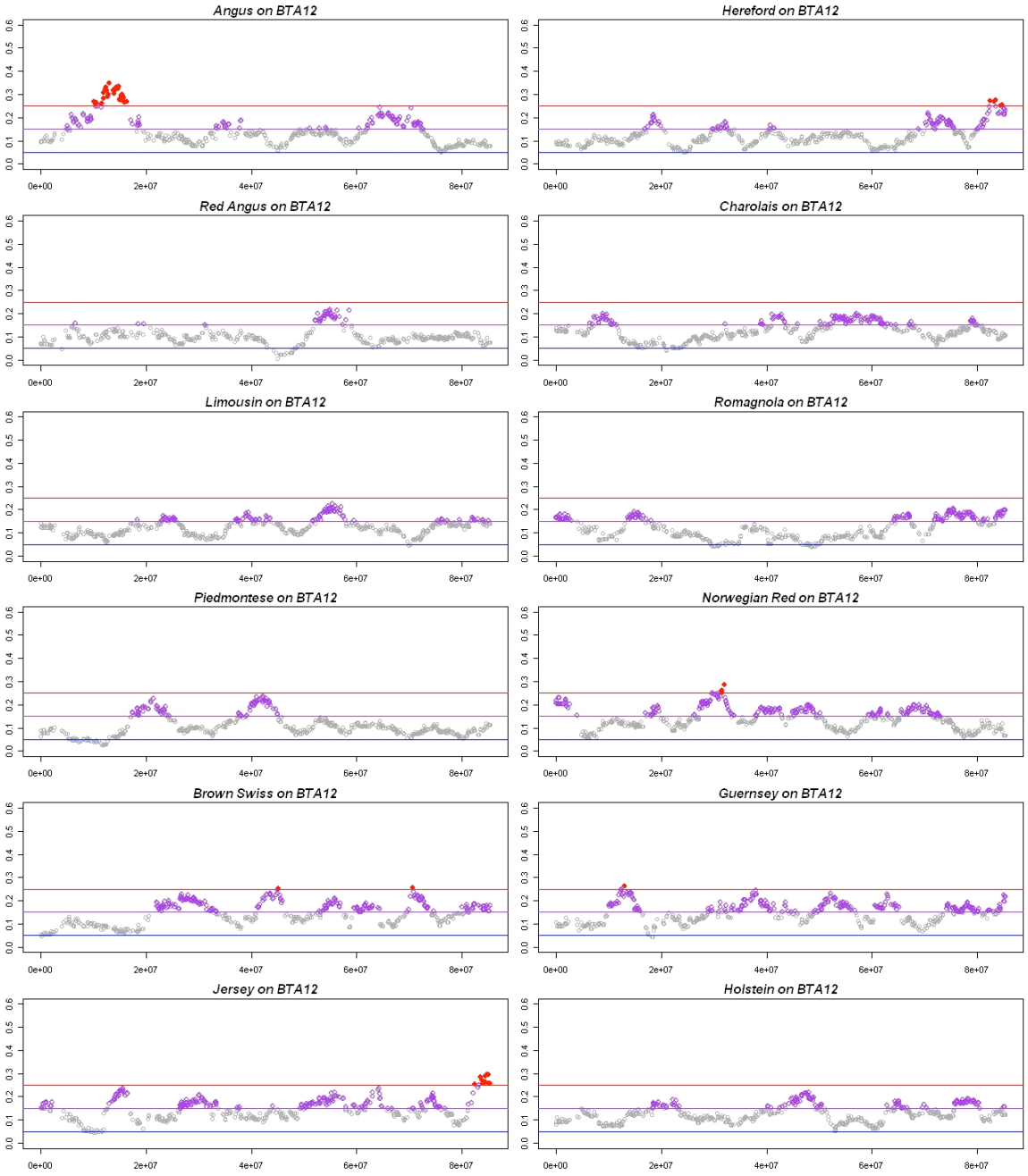


Figure C.12. Population-specific F_{ST} values for BTA12.

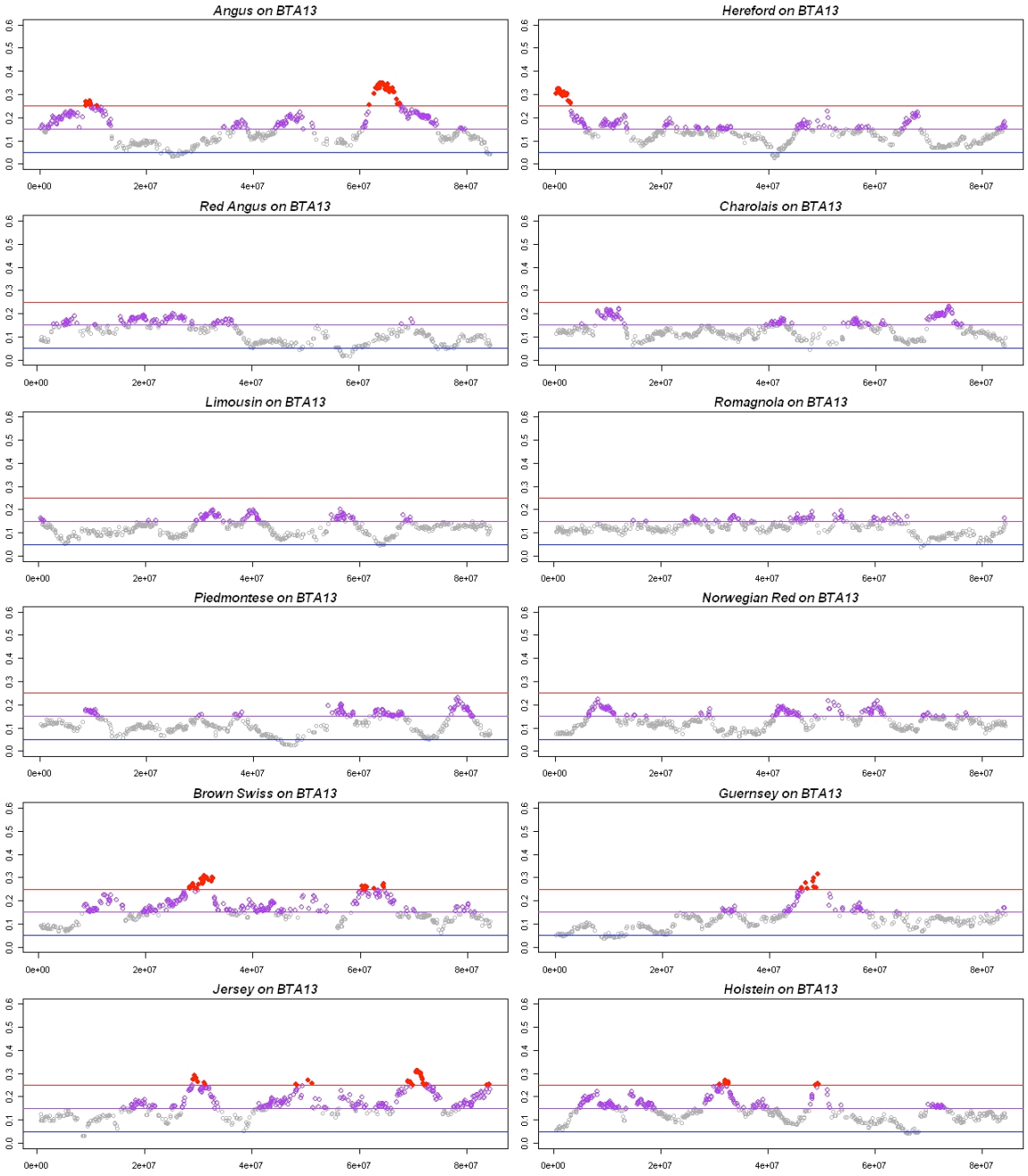


Figure C.13. Population-specific F_{ST} values for BTA13.

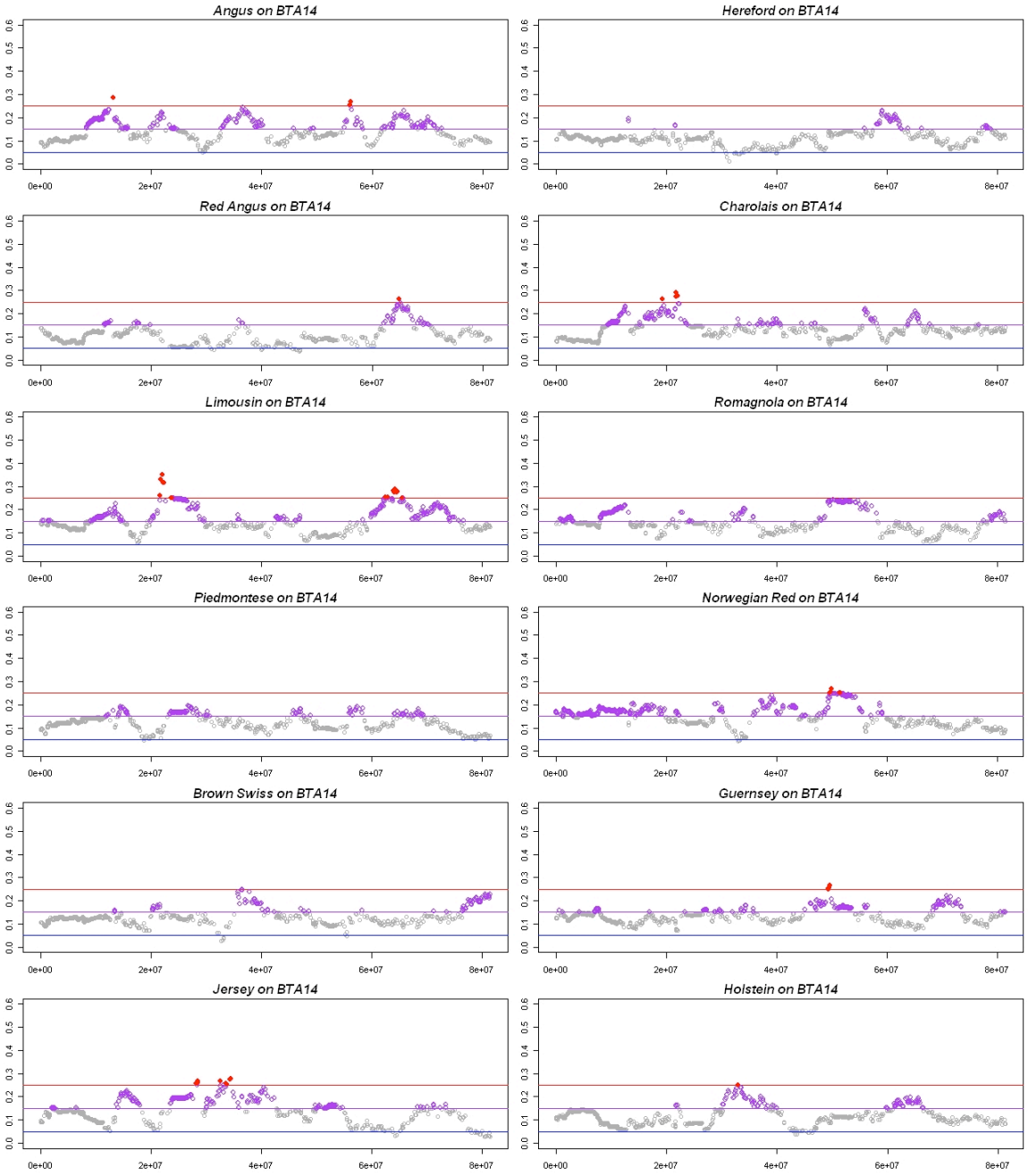


Figure C.14. Population-specific F_{ST} values for BTA14.

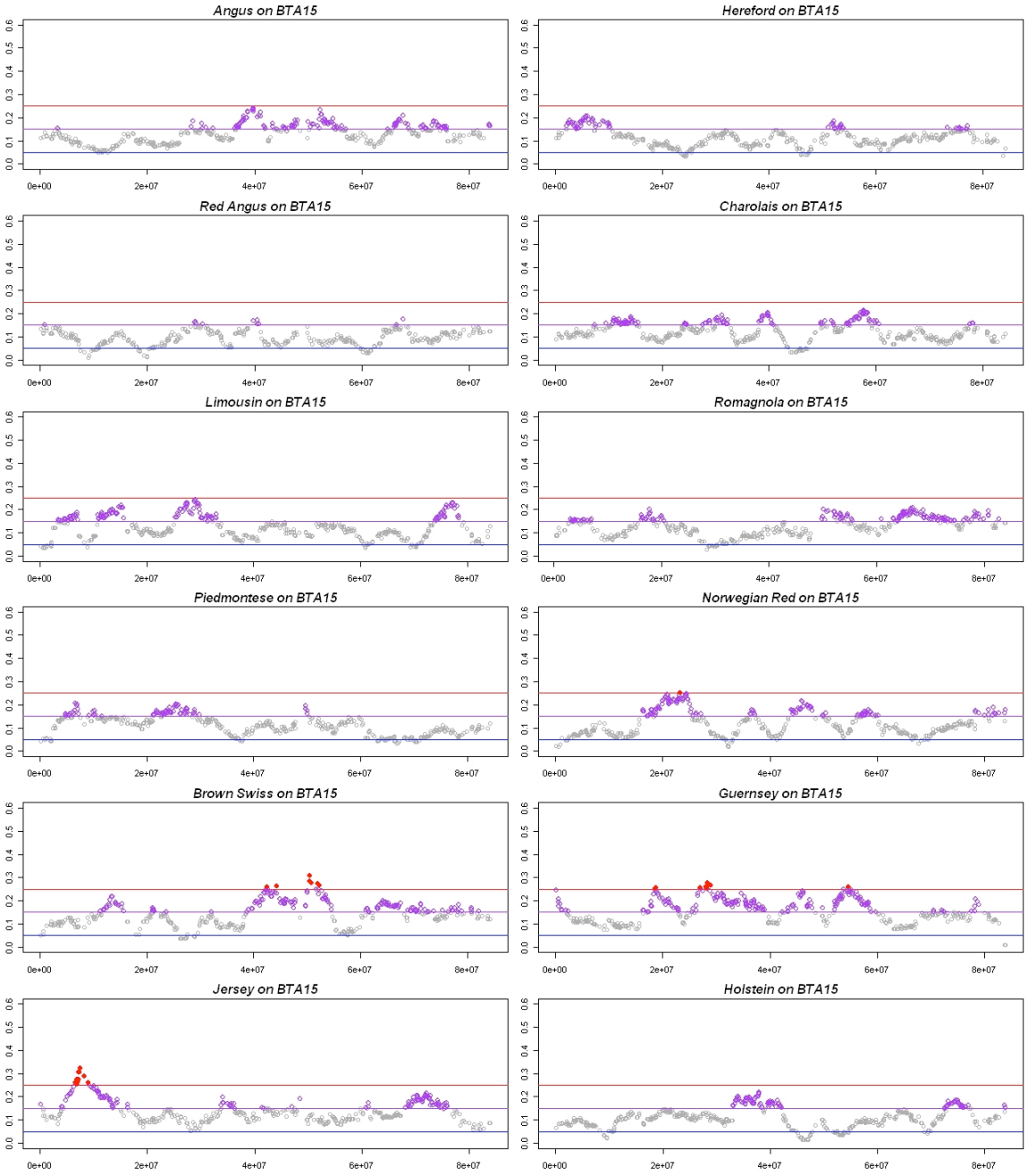


Figure C.15. Population-specific F_{ST} values for BTA15.

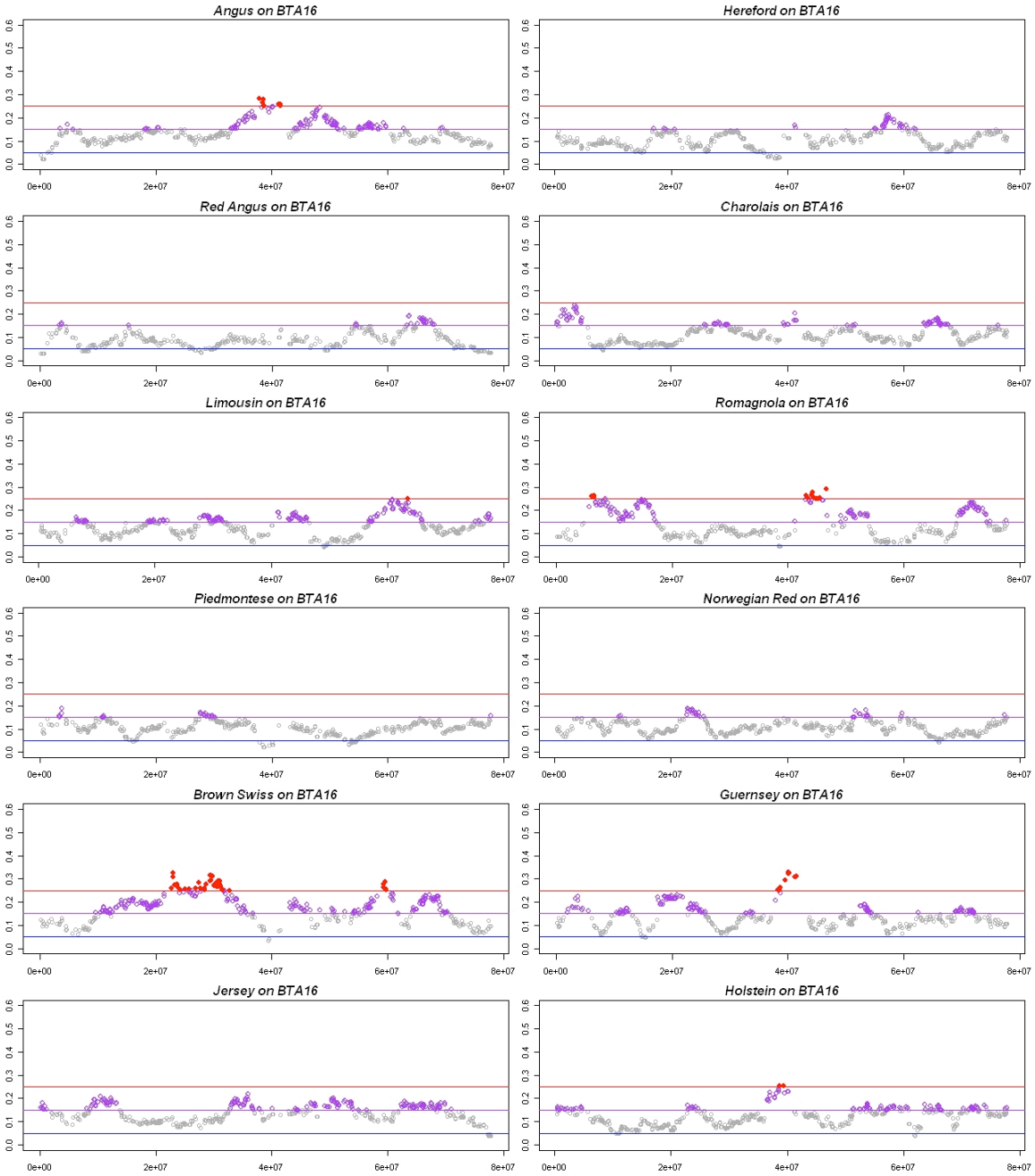


Figure C.16. Population-specific F_{ST} values for BTA16.

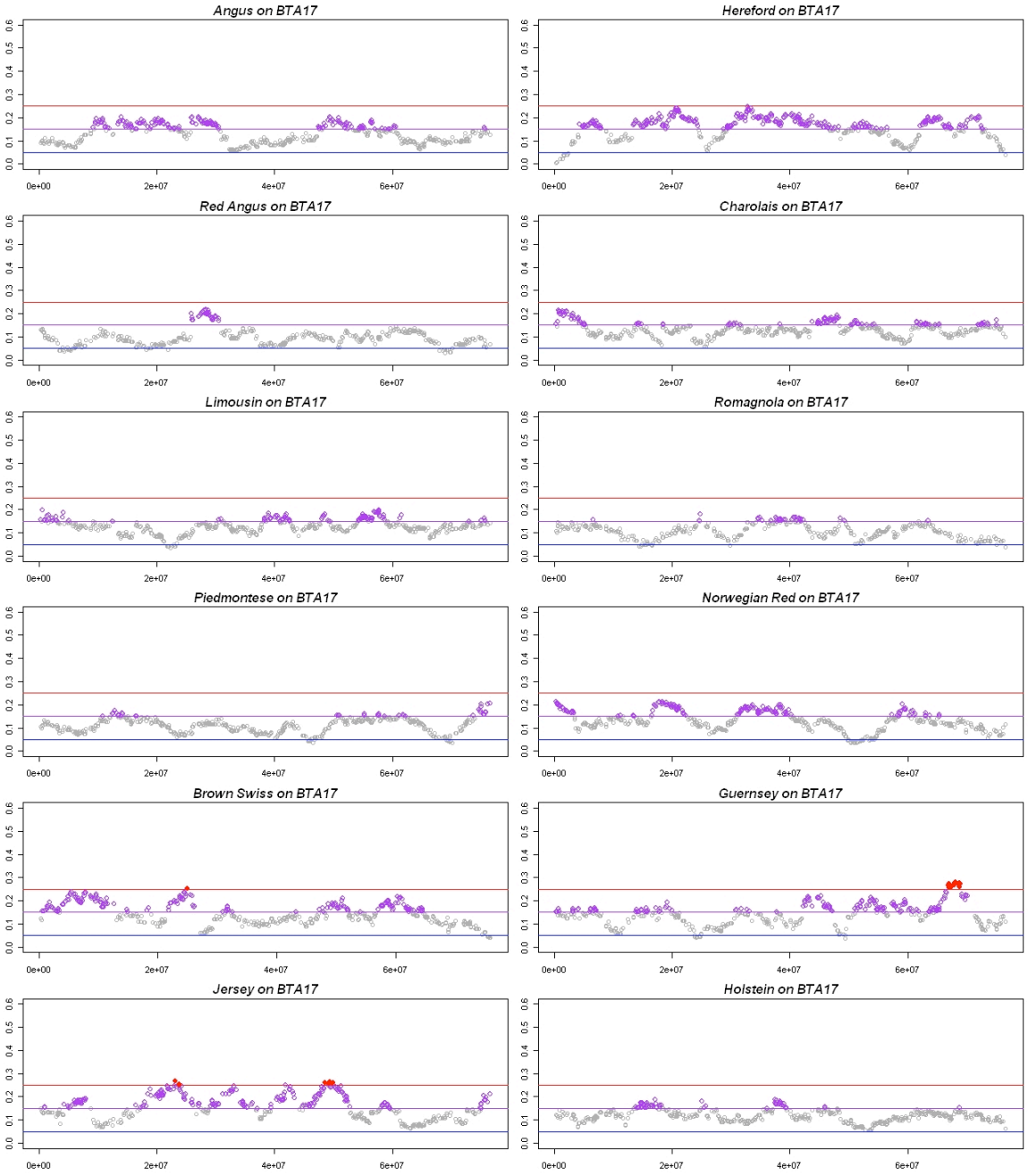


Figure C.17. Population-specific F_{ST} values for BTA17.

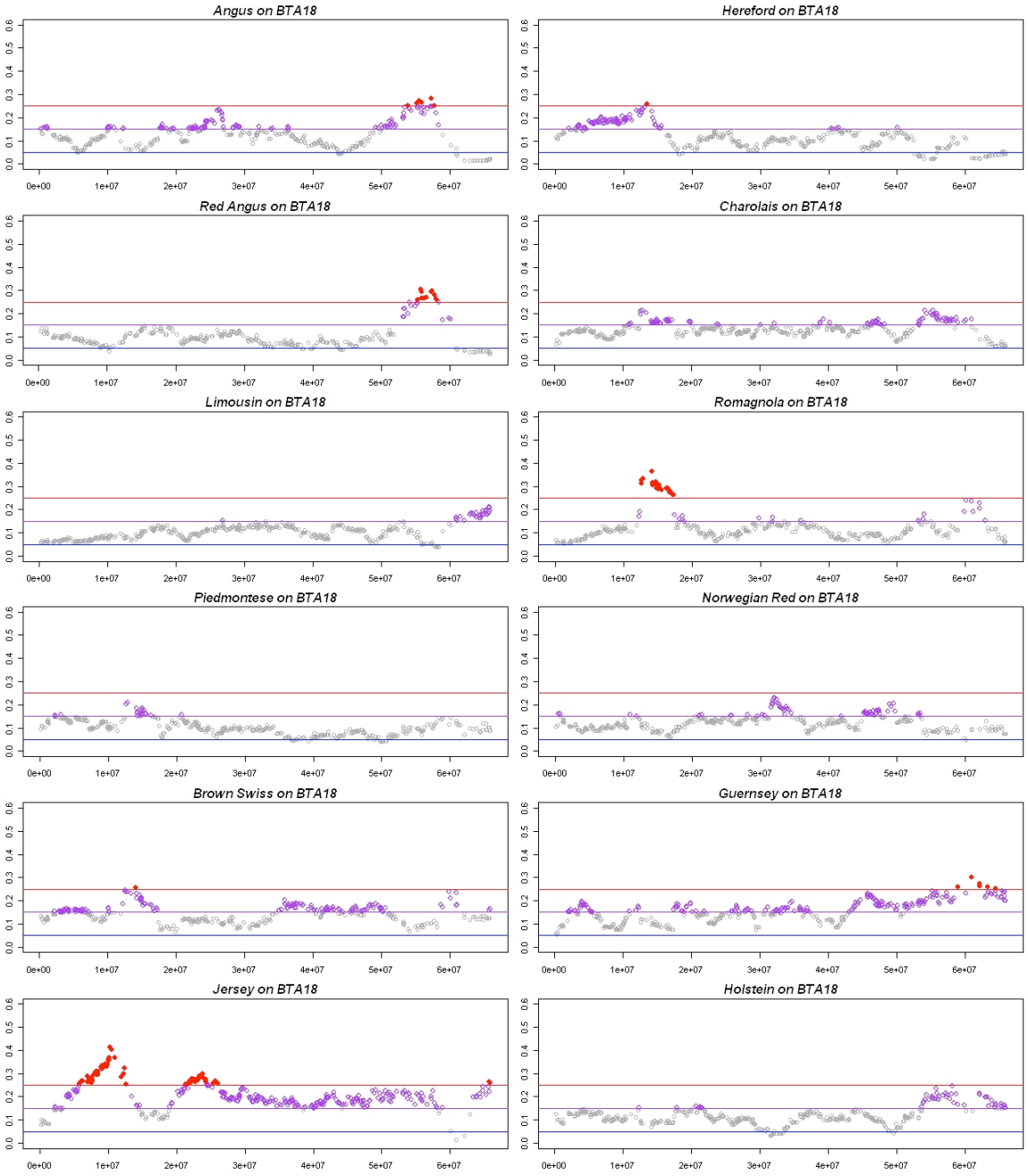


Figure C.18. Population-specific F_{ST} values for BTA18.

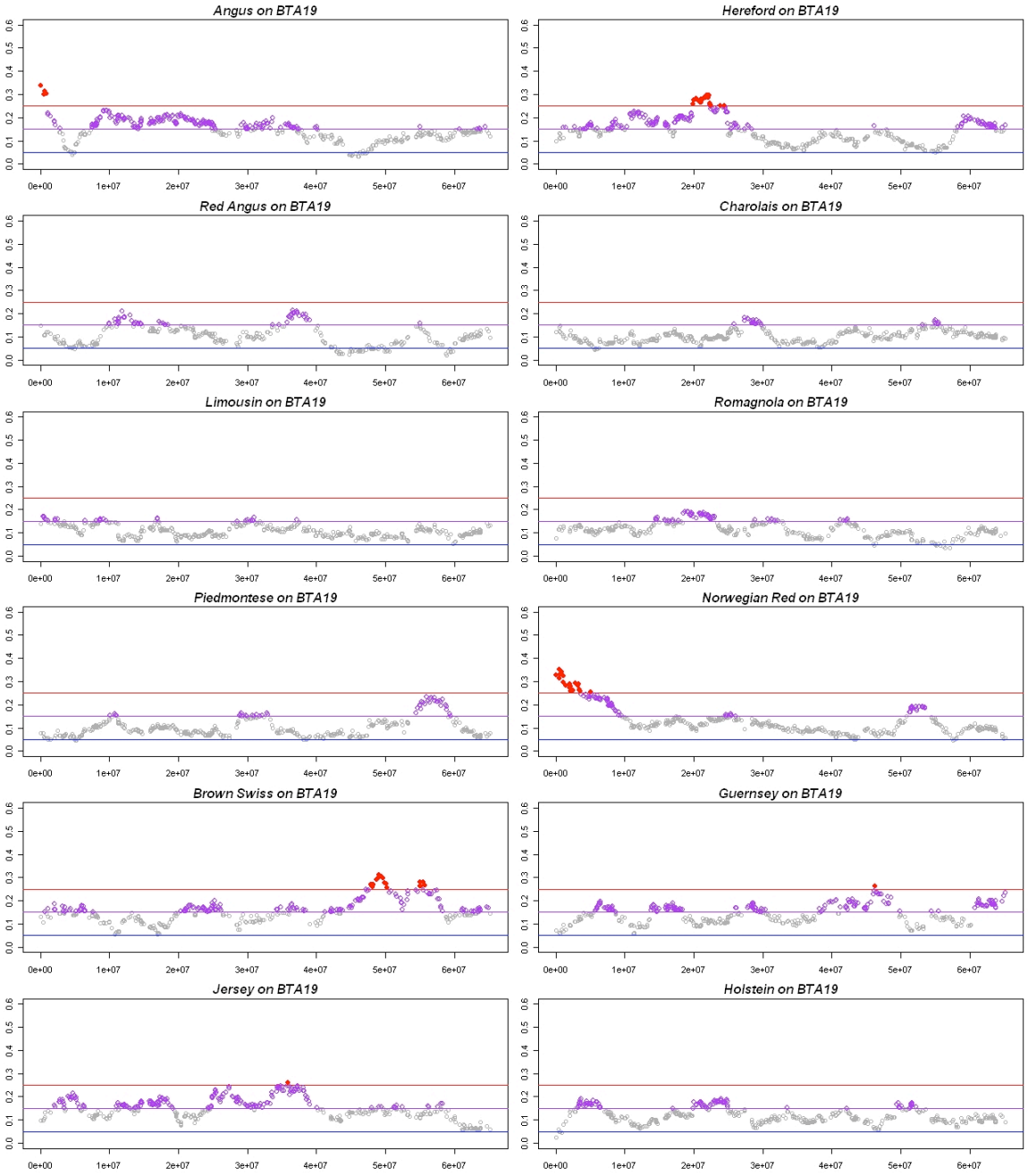


Figure C.19. Population-specific F_{ST} values for BTA19.

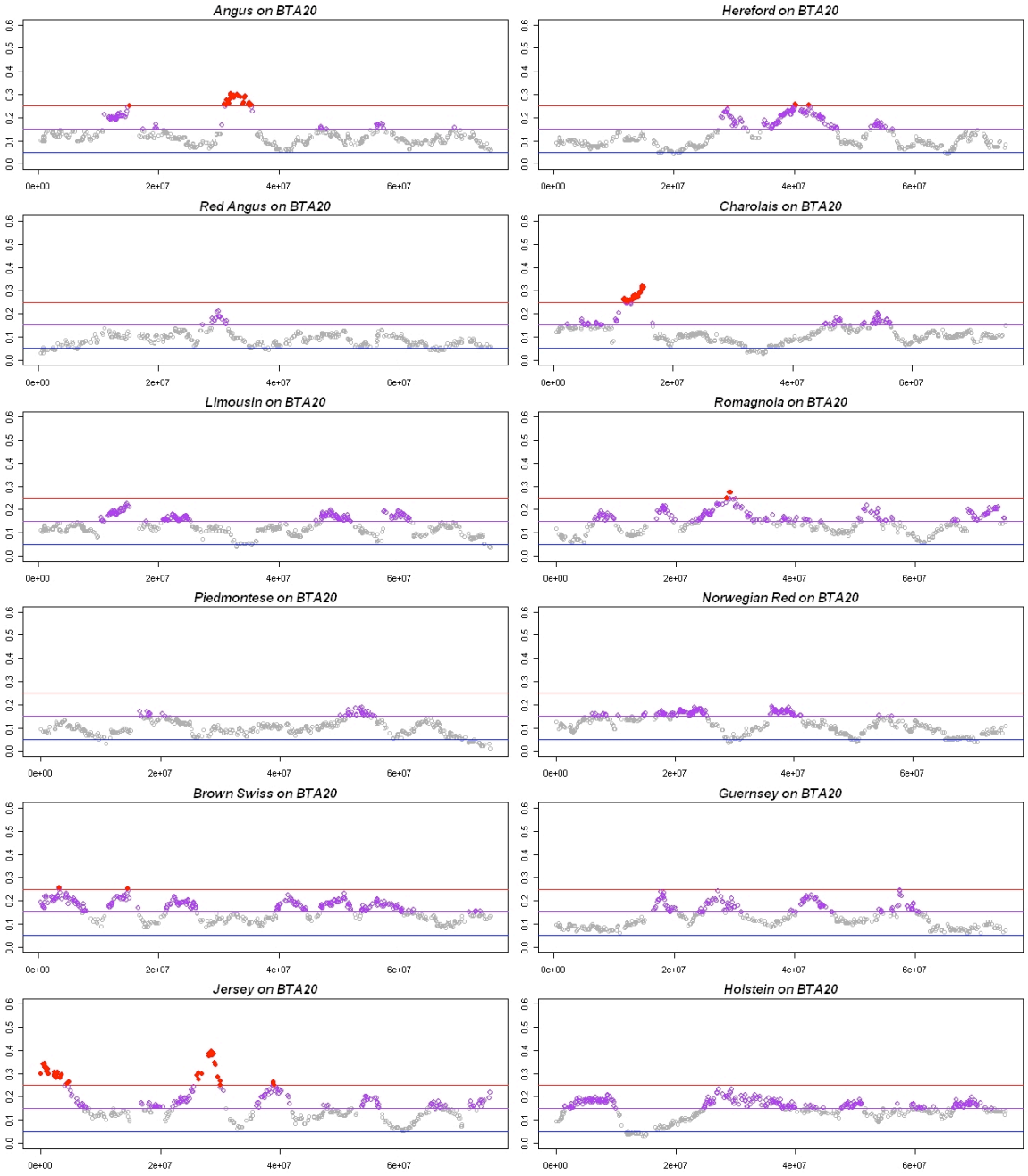


Figure C.20. Population-specific F_{ST} values for BTA20.

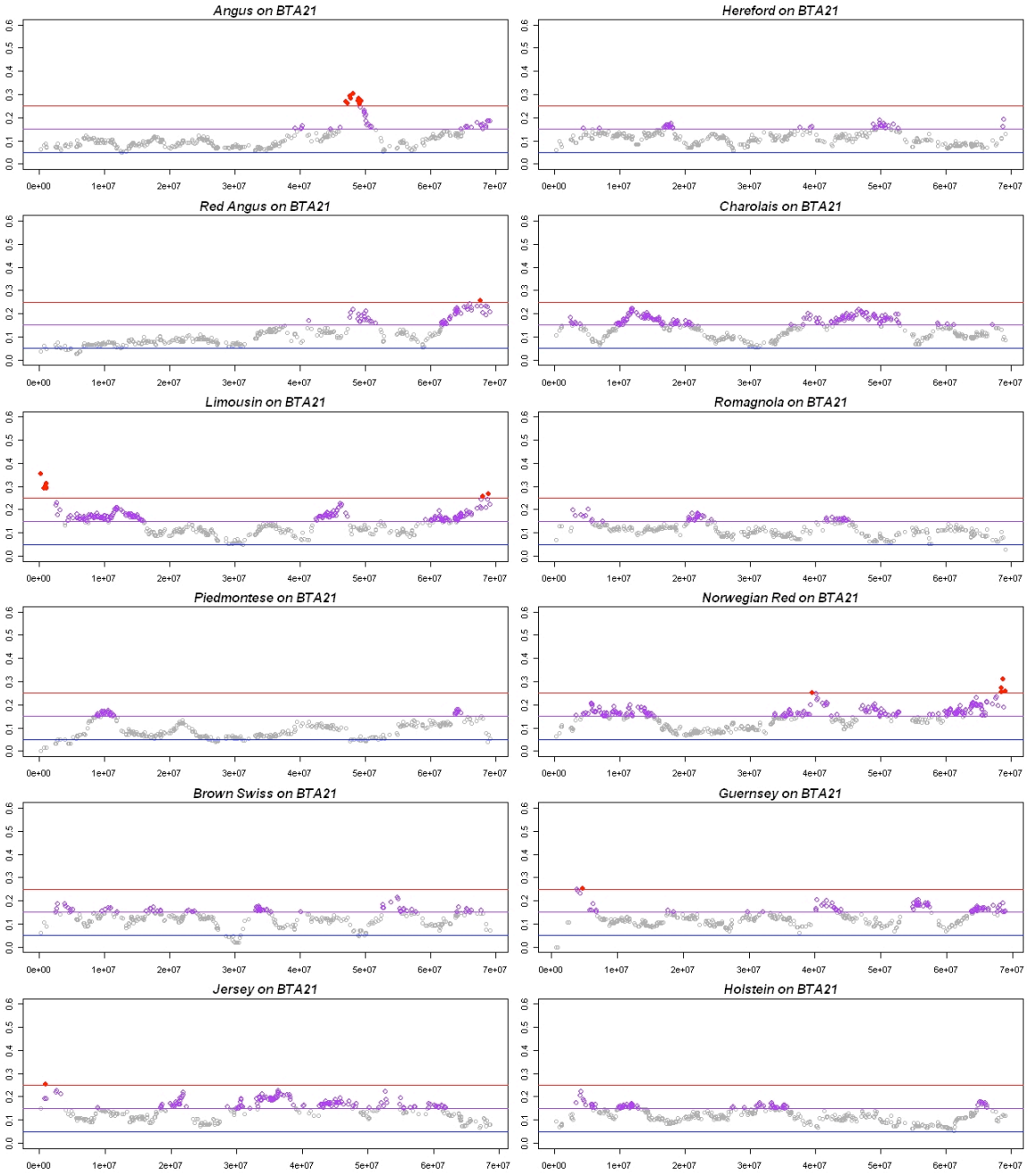


Figure C.21. Population-specific F_{ST} values for BTA21.

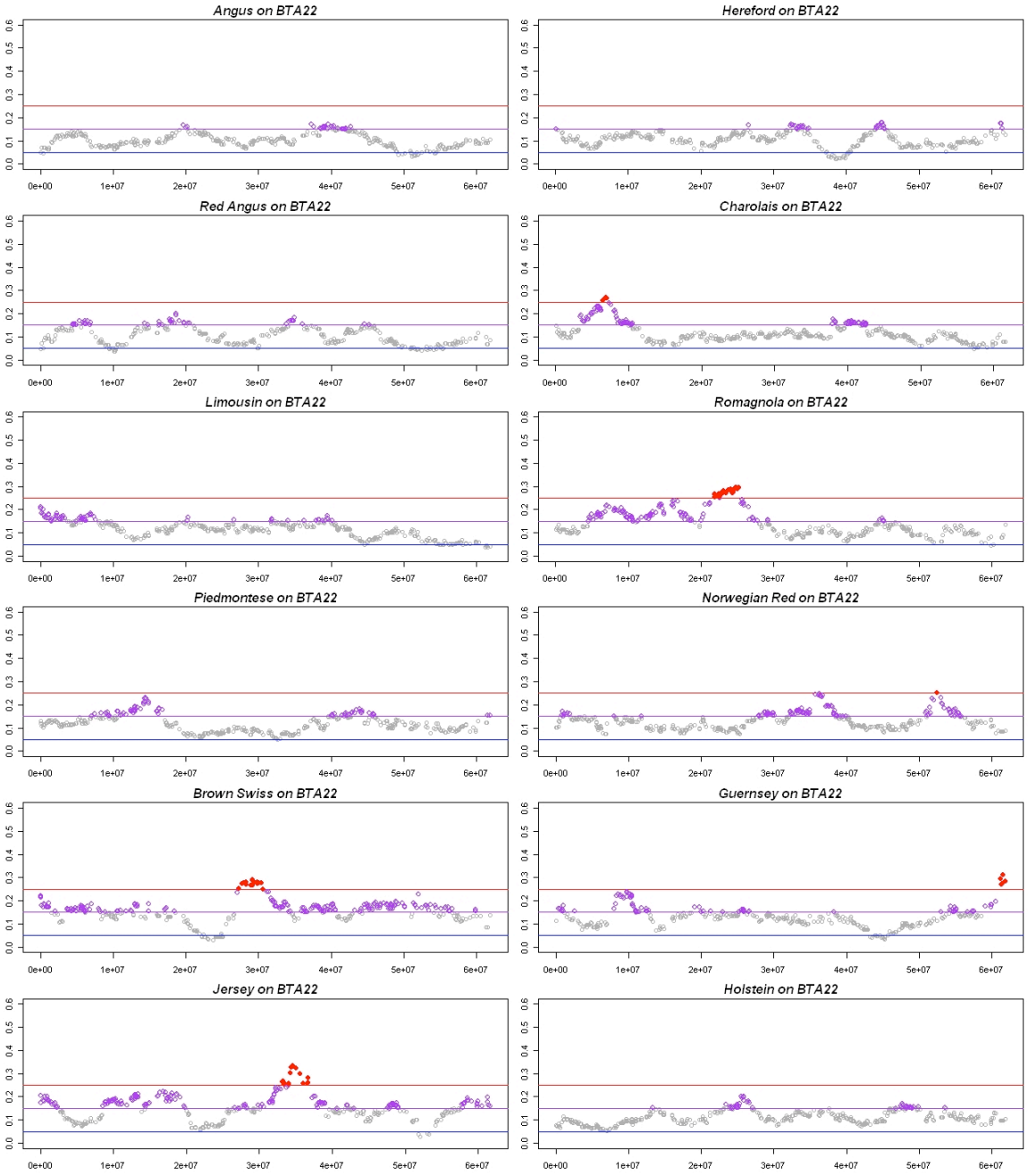


Figure C.22. Population-specific F_{ST} values for BTA22.

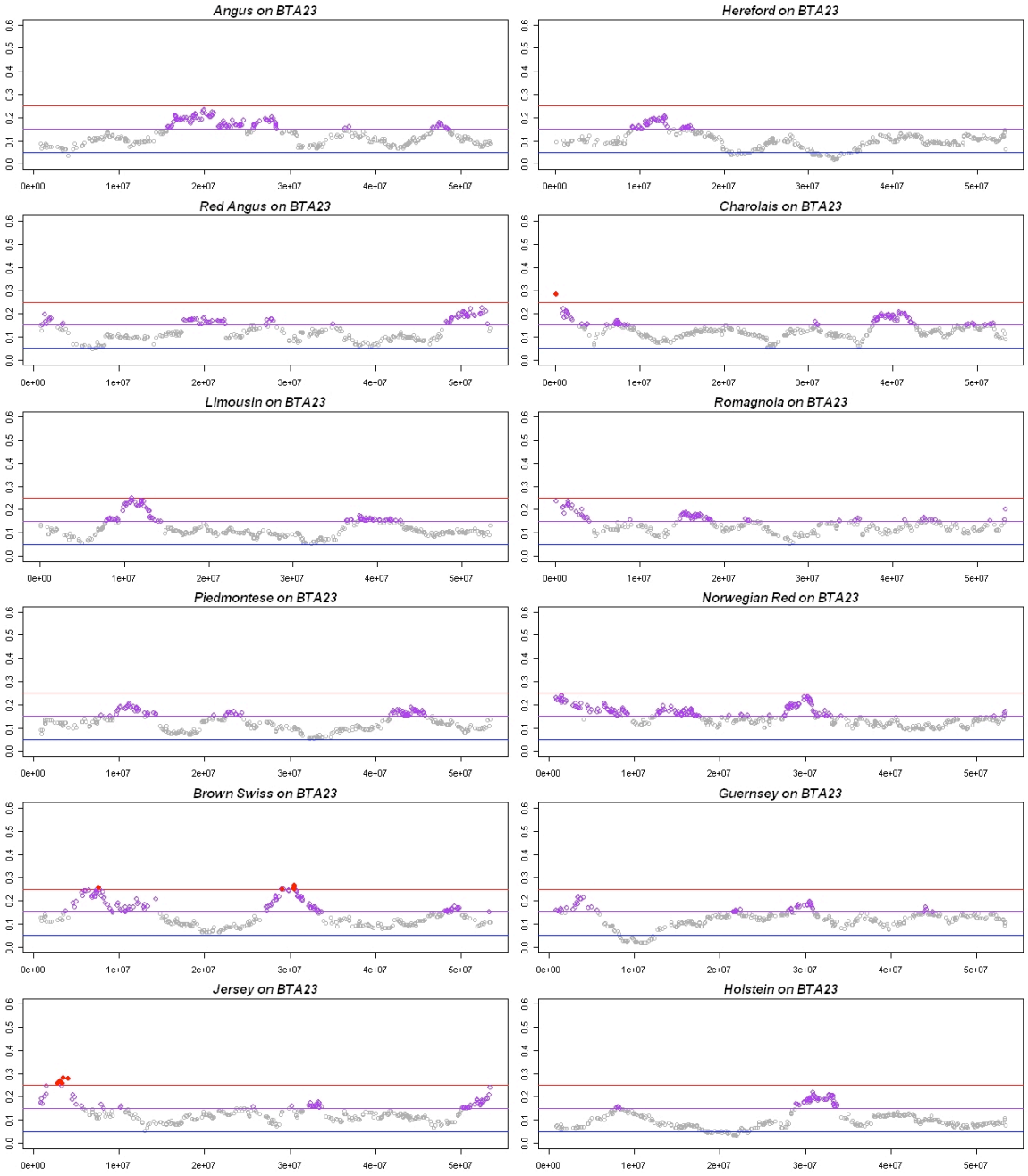


Figure C.23. Population-specific F_{ST} values for BTA23.

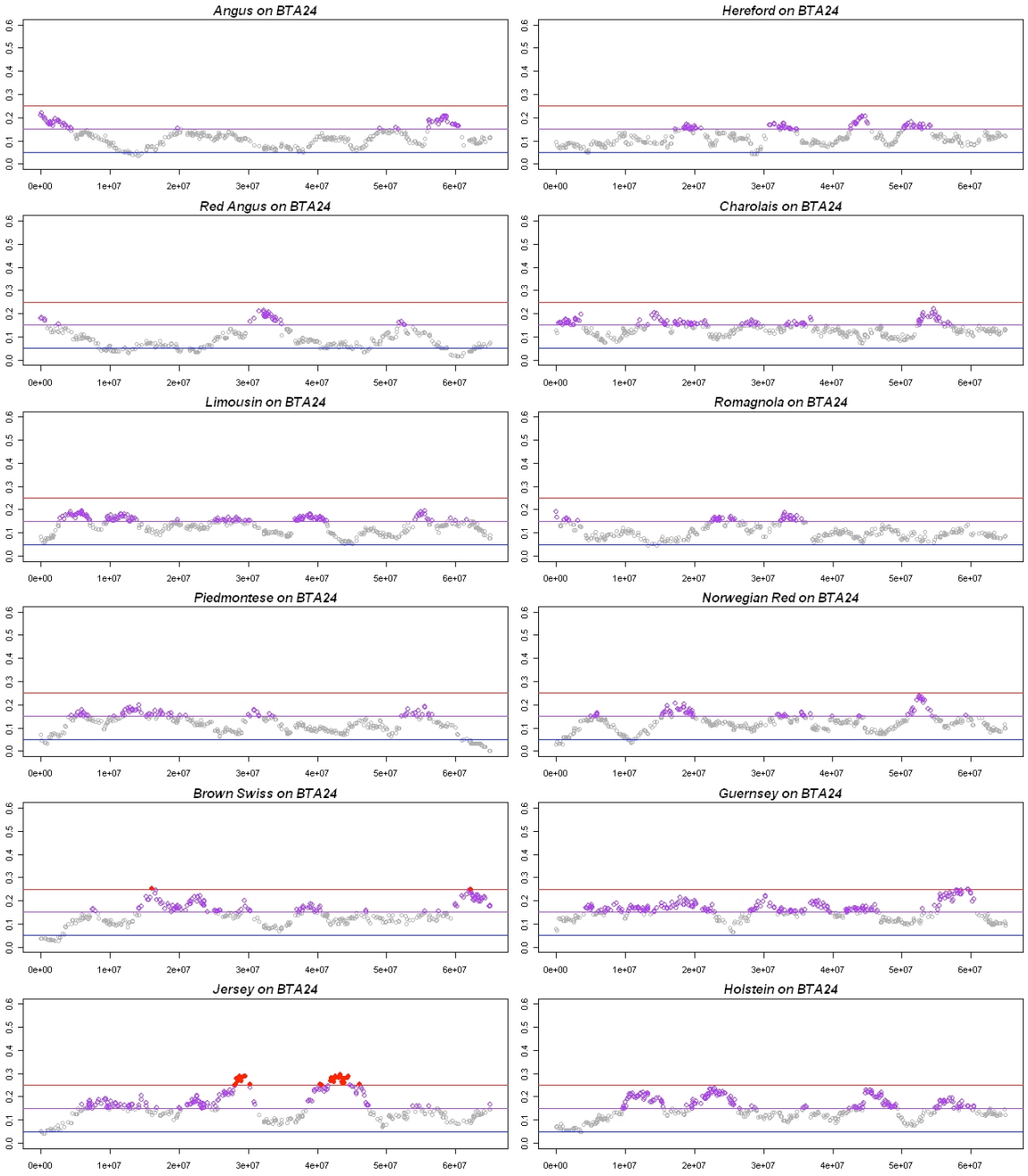


Figure C.24. Population-specific F_{ST} values for BTA24.

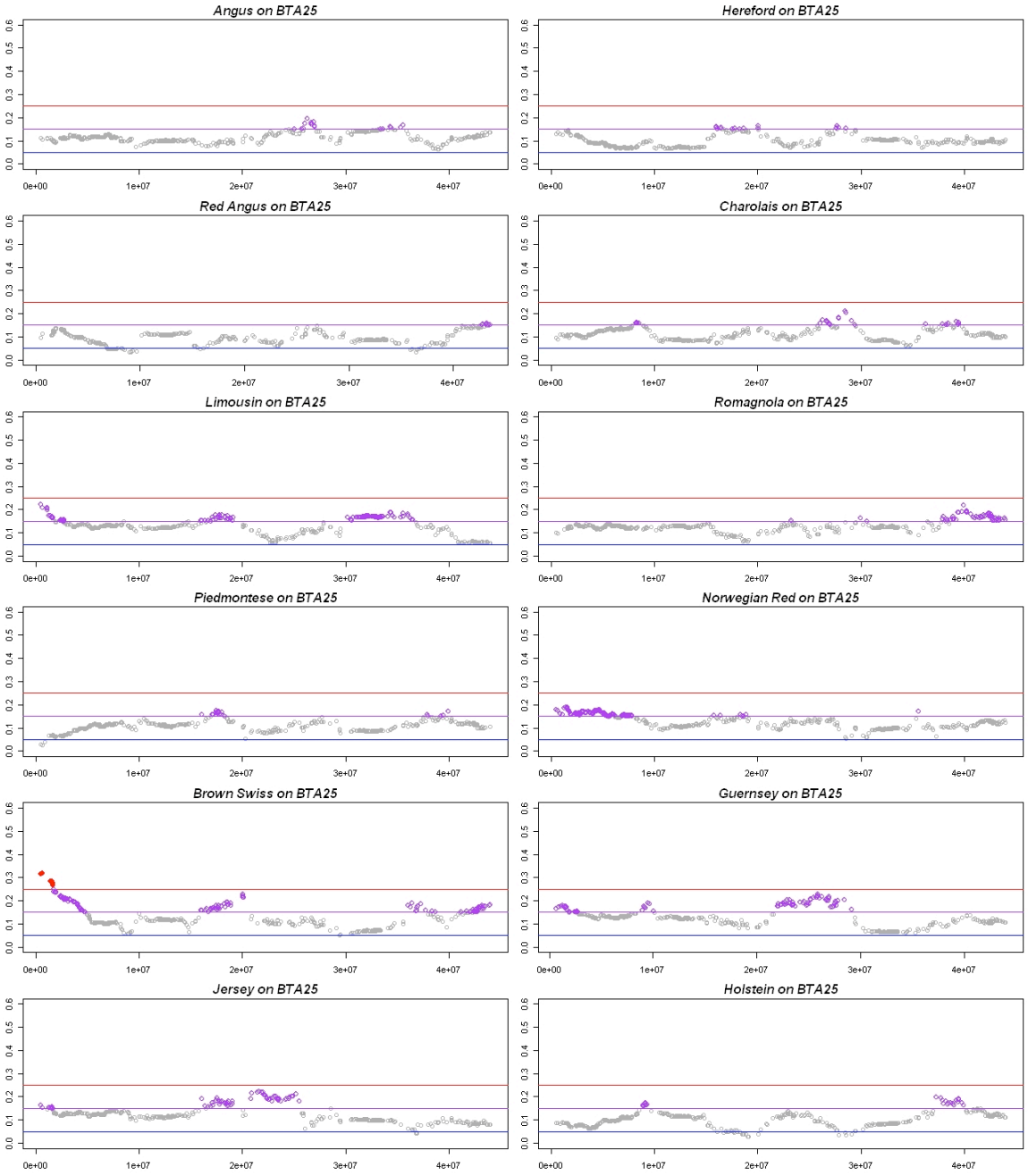


Figure C.25. Population-specific F_{ST} values for BTA25.

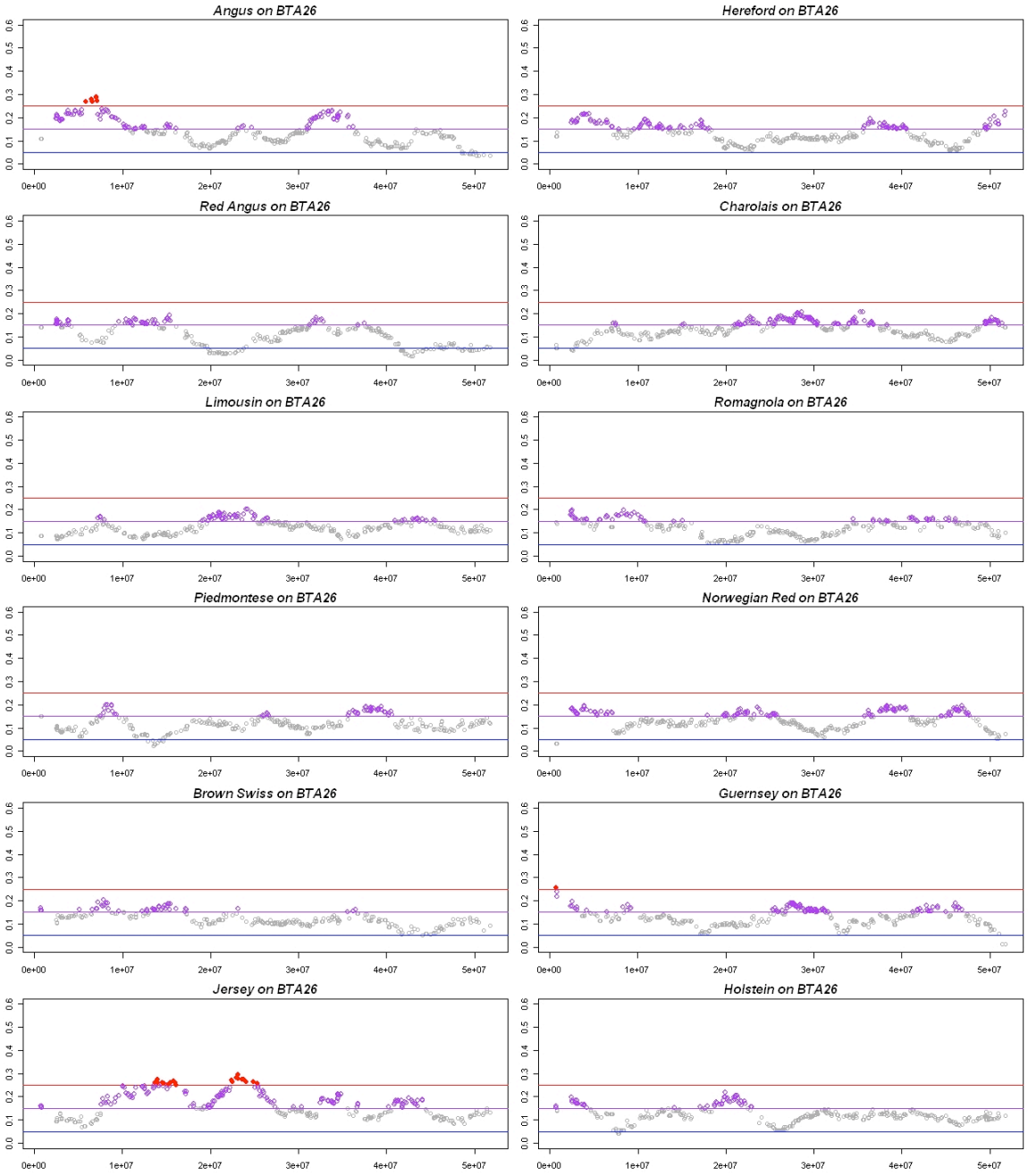


Figure C.26. Population-specific F_{ST} values for BTA26.

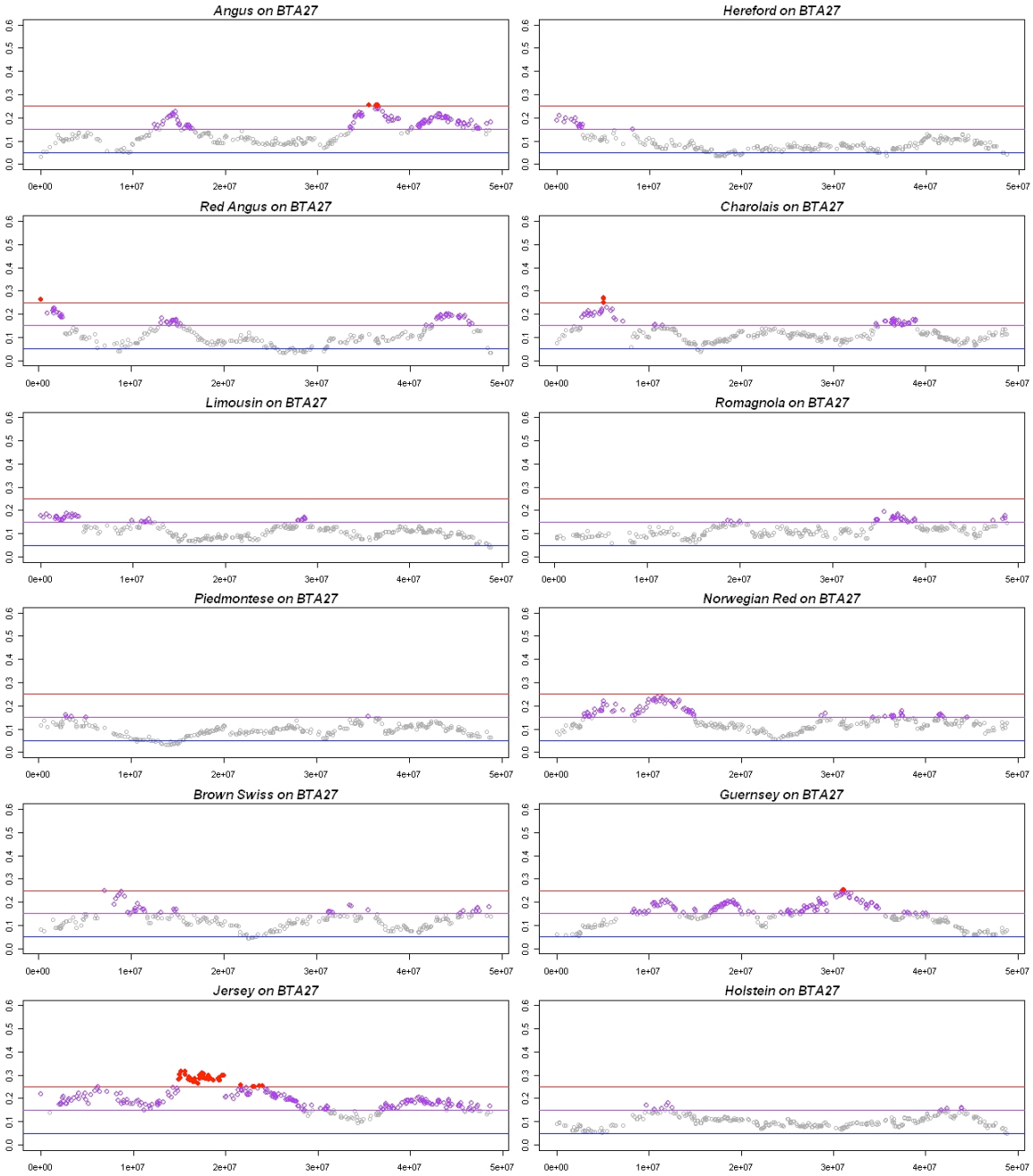


Figure C.27. Population-specific F_{ST} values for BTA27.

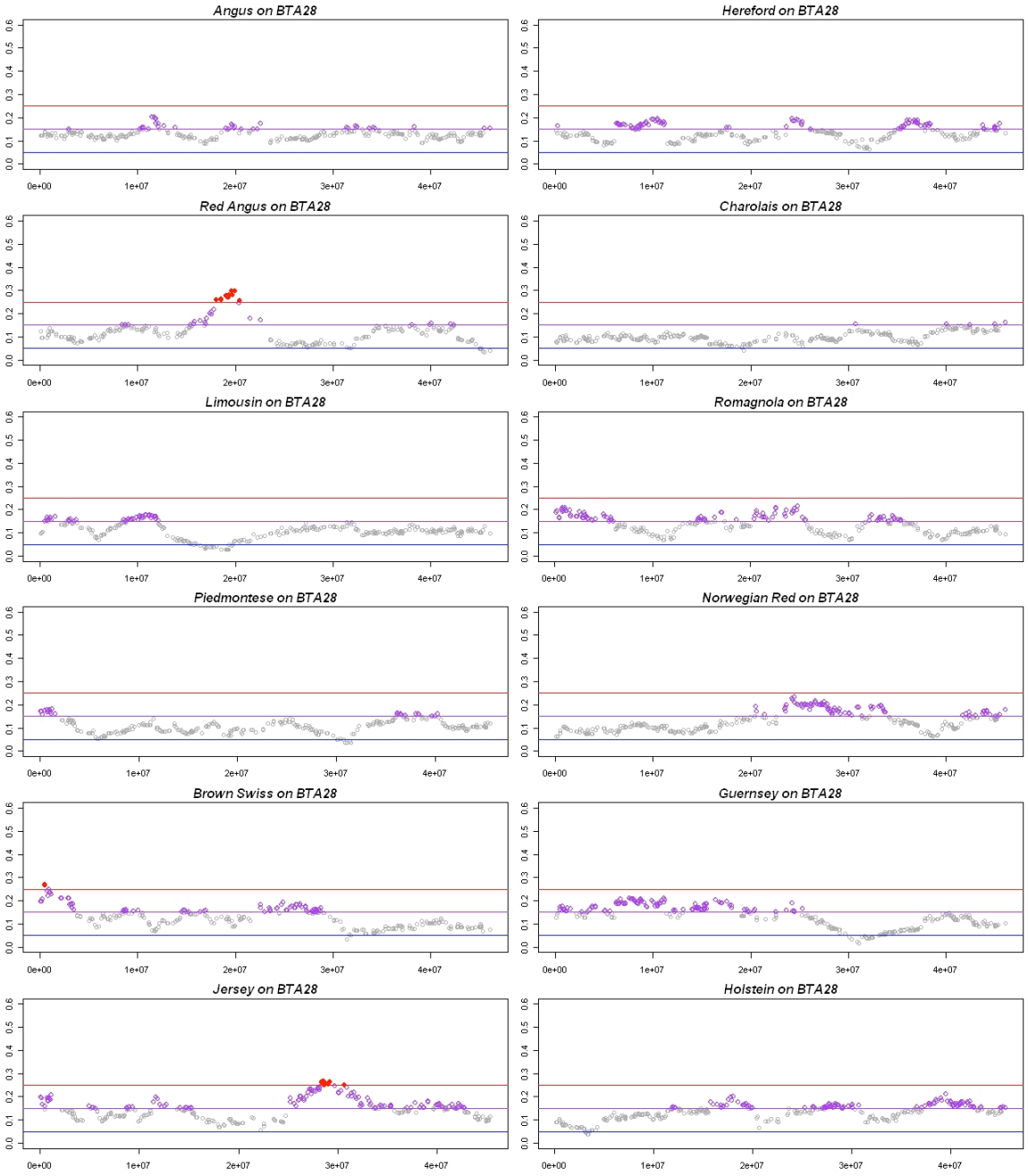


Figure C.28. Population-specific F_{ST} values for BTA28.

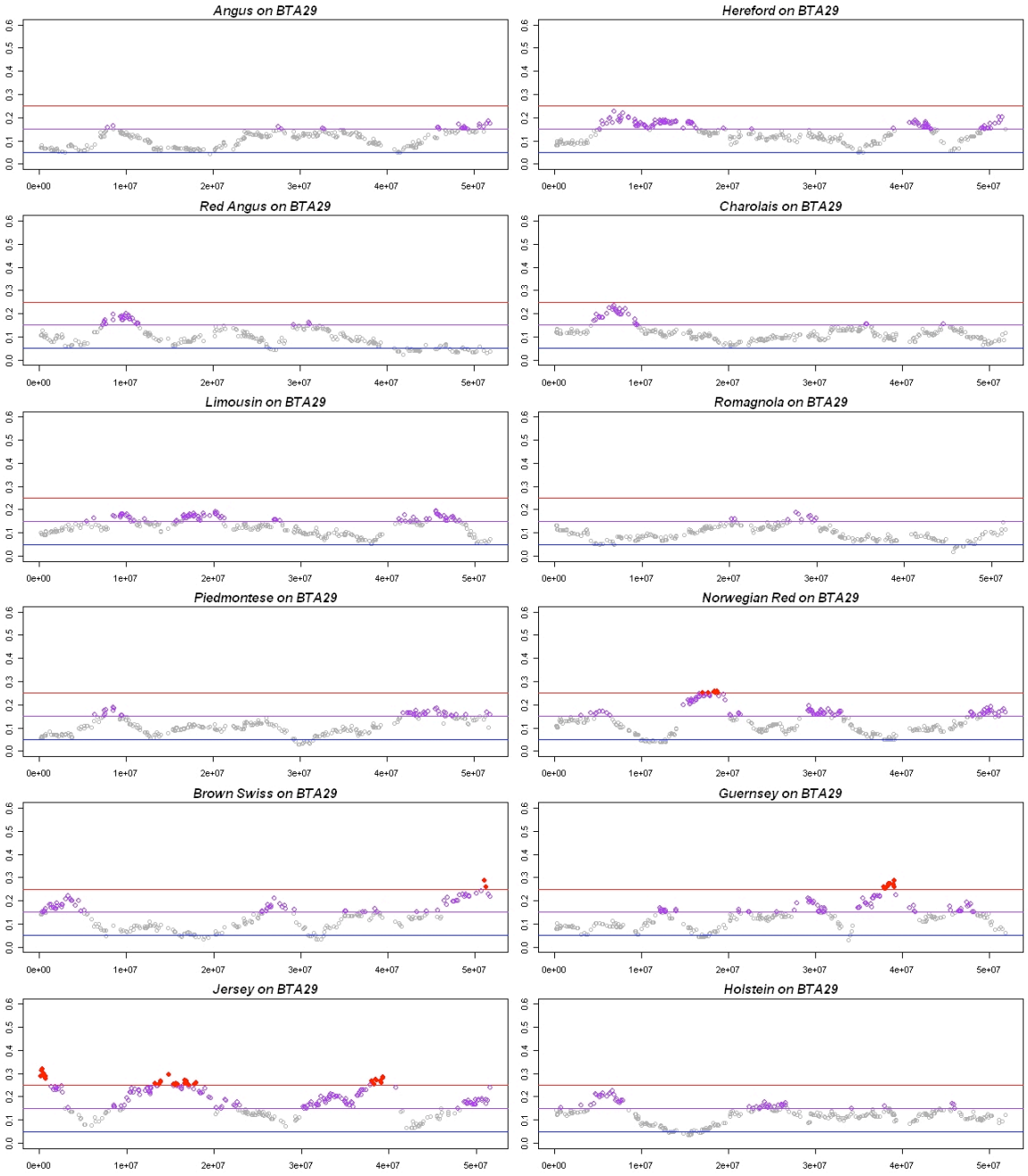


Figure C.29. Population-specific F_{ST} values for BTA29.

APPENDIX D

Integrated extended haplotype homozygosity scores (|iHS|) for each of the bovine autosomes are presented in Figures D.1 to D.29. The breeds sampled were Angus (ANG), Brown Swiss (BSW), Charolais (CHL), Guernsey (GNS), Hereford (HFD), Holstein (HOL), Jersey (JER), Limousin (LMS), Norwegian Red (NRC), Piedmontese (PMT), Red Angus (RGU), Romagnola (RMG). The horizontal line indicates the top 1% of |iHS| values (2.694). Each SNP is plotted with respect to the coordinates in build BTAu4.0.

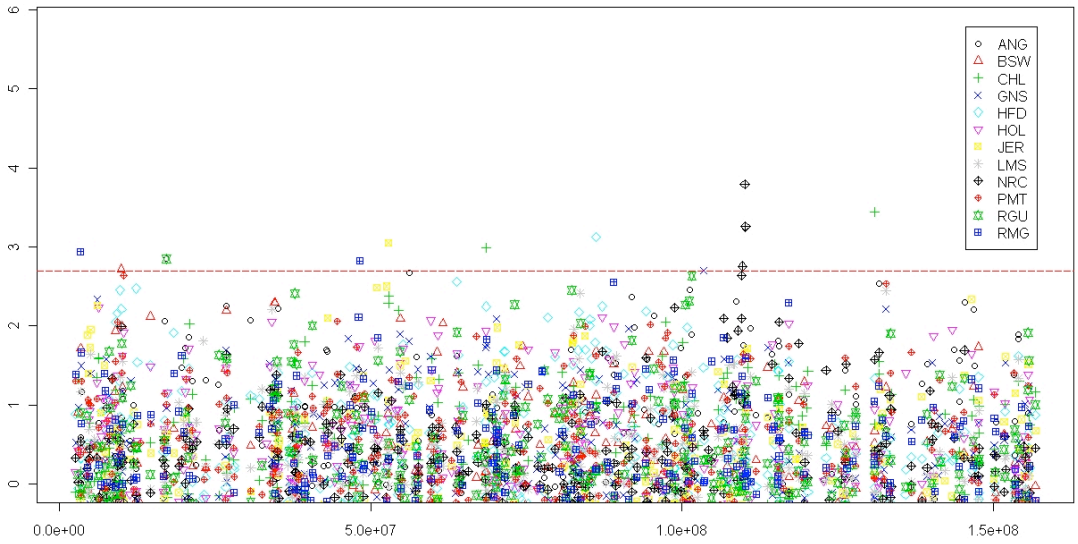


Figure D.1. Integrated extended haplotype homozygosity scores (|iHS|) on BTA1.

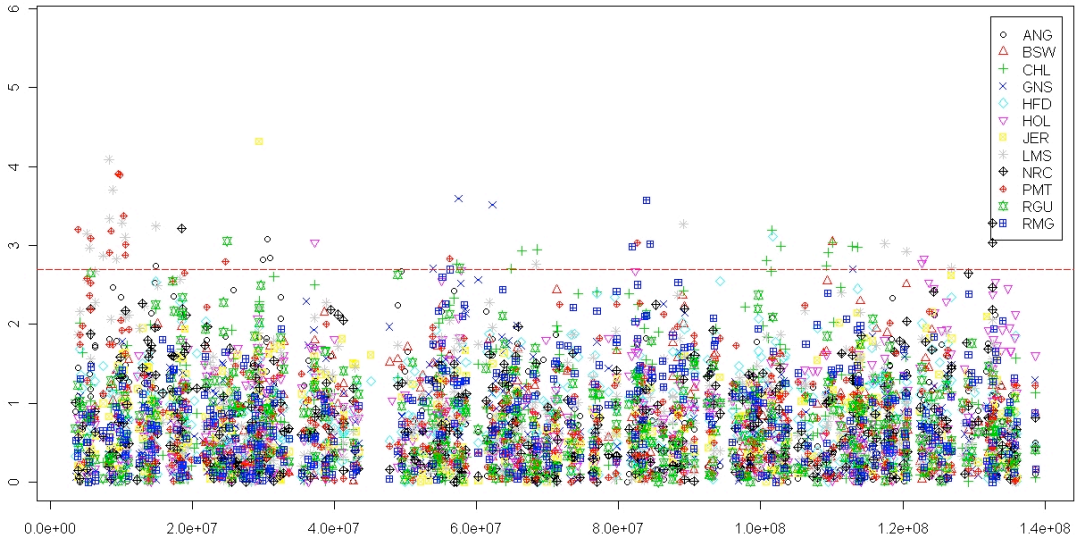


Figure D.2. Integrated extended haplotype homozygosity scores (|iHS|) on BTA2.

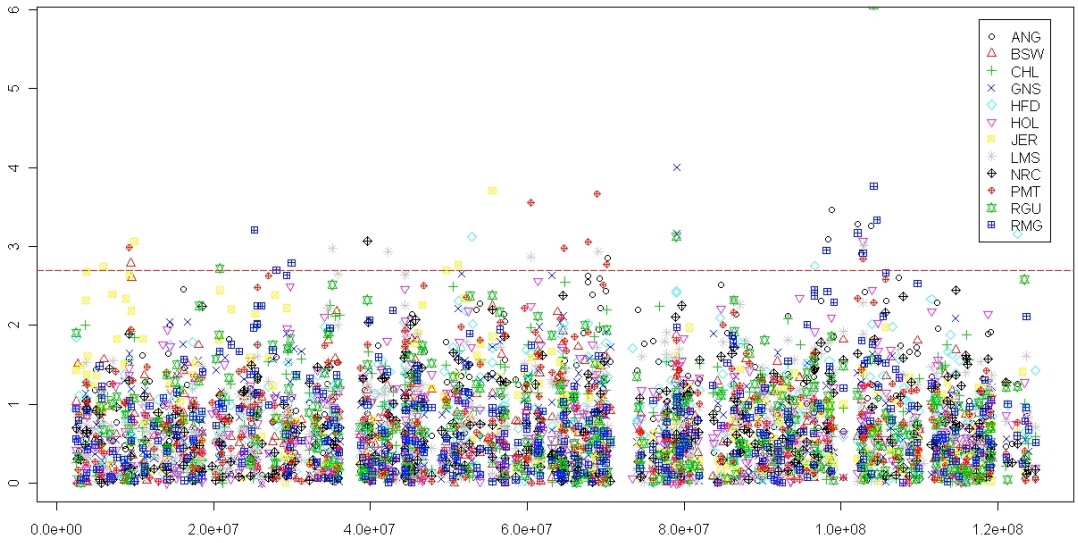


Figure D.3. Integrated extended haplotype homozygosity scores (|iHS|) on BTA3.

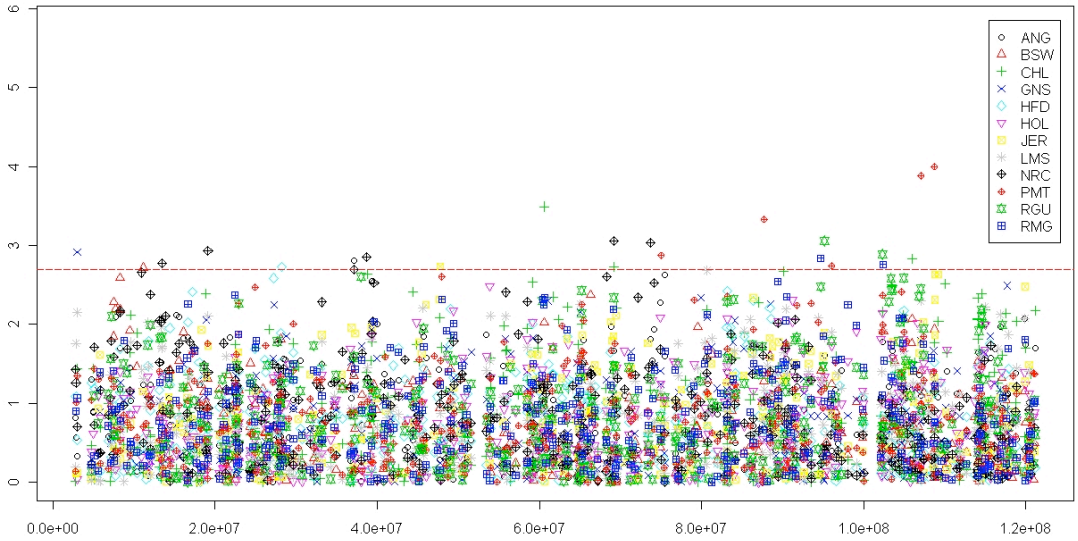


Figure D.4. Integrated extended haplotype homozygosity scores (|iHS|) on BTA4.

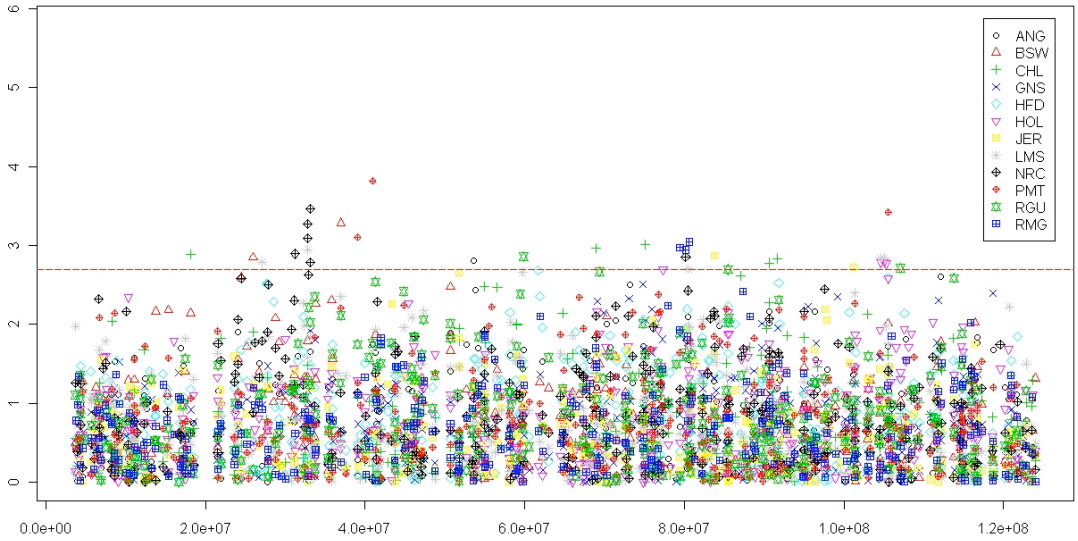


Figure D.5. Integrated extended haplotype homozygosity scores (|iHS|) on BTA5.

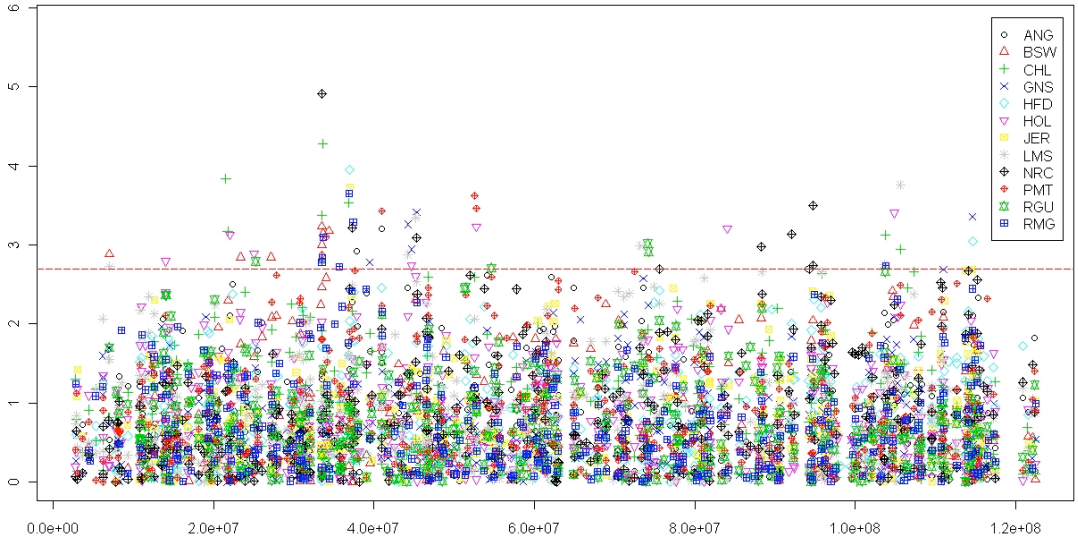


Figure D.6. Integrated extended haplotype homozygosity scores (|iHS|) on BTA6.

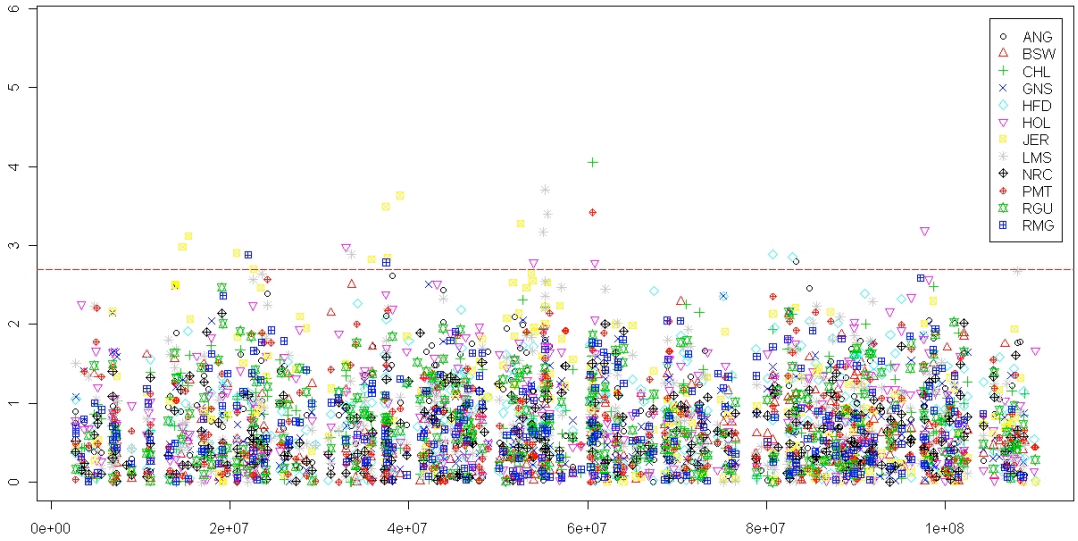


Figure D.7. Integrated extended haplotype homozygosity scores (|iHS|) on BTA7.

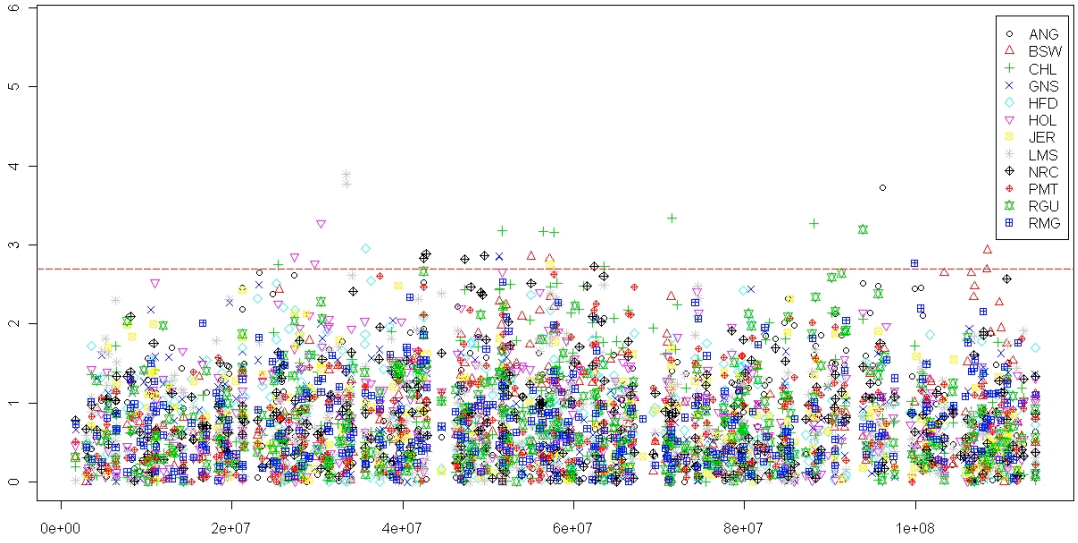


Figure D.8. Integrated extended haplotype homozygosity scores (|iHS|) on BTA8.

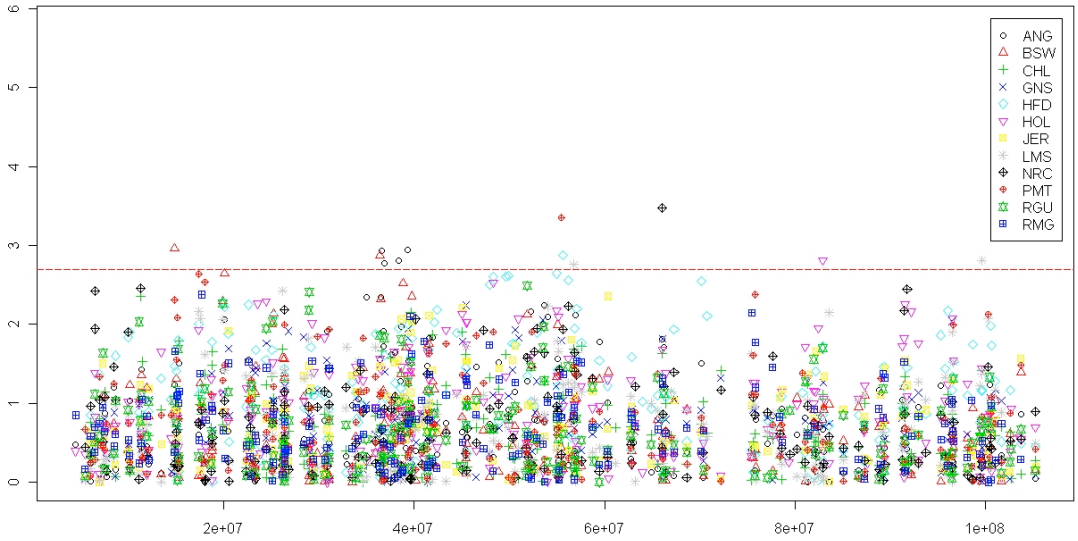


Figure D.9. Integrated extended haplotype homozygosity scores (|iHS|) on BTA9.

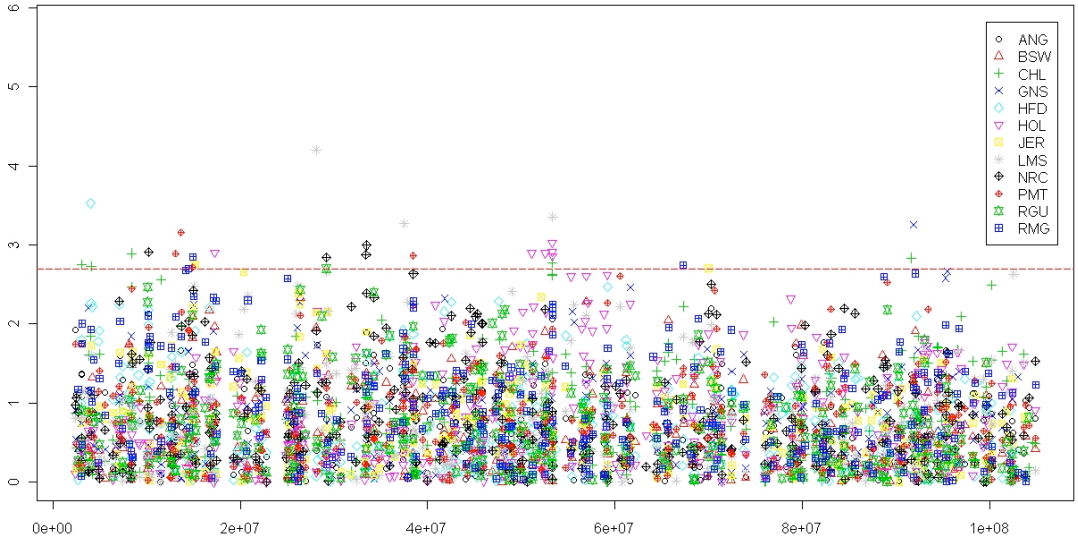


Figure D.10. Integrated extended haplotype homozygosity scores (|iHS|) on BTA10.

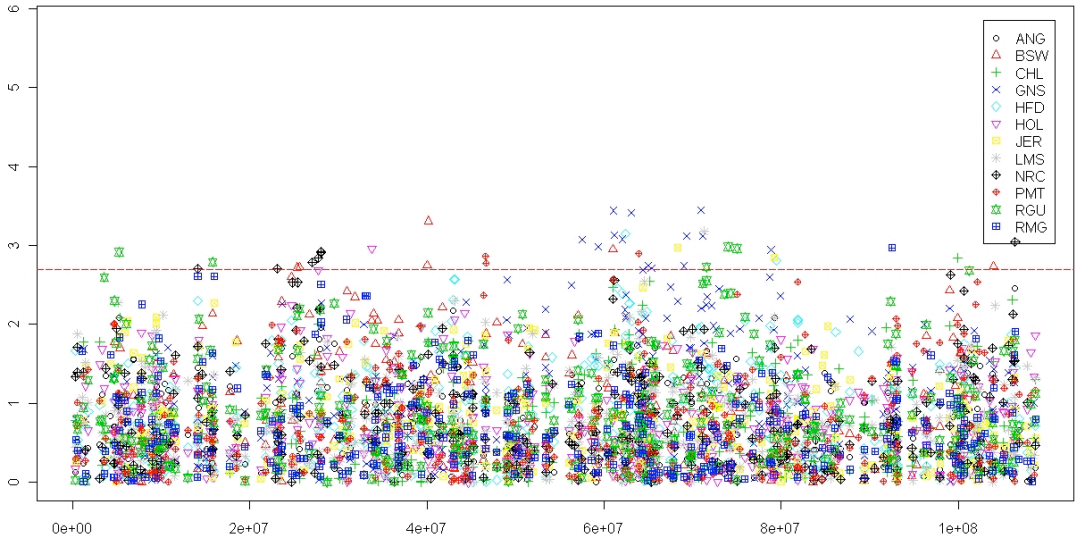


Figure D.11. Integrated extended haplotype homozygosity scores (|iHS|) on BTA11.

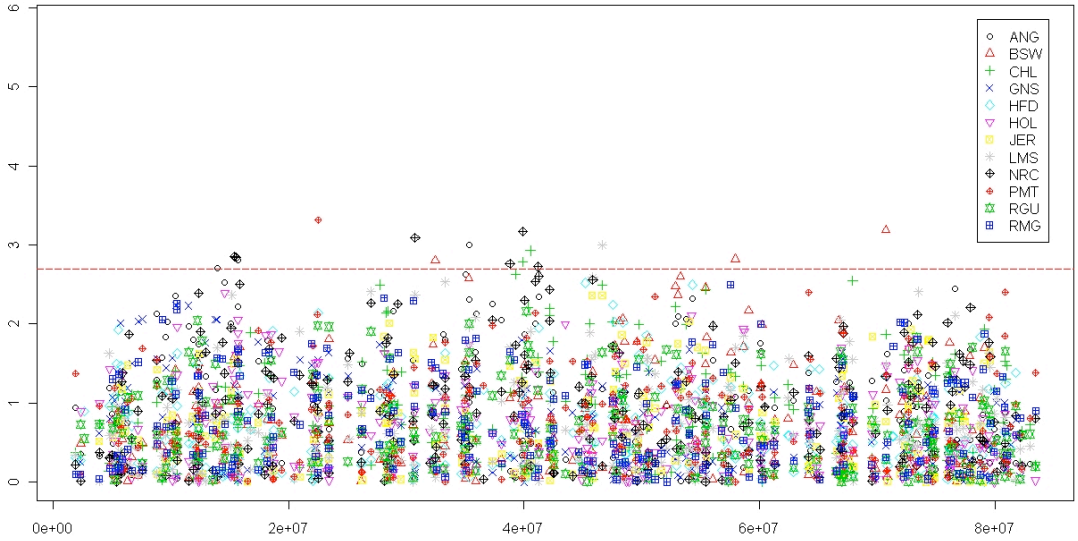


Figure D.12. Integrated extended haplotype homozygosity scores (|iHS|) on BTA12.

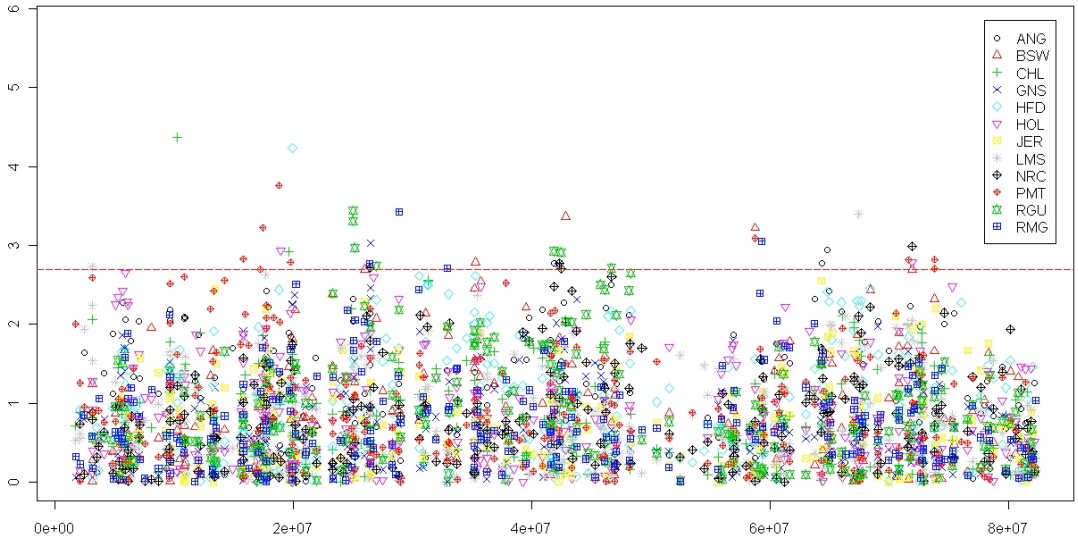


Figure D.13. Integrated extended haplotype homozygosity scores (|iHS|) on BTA13.

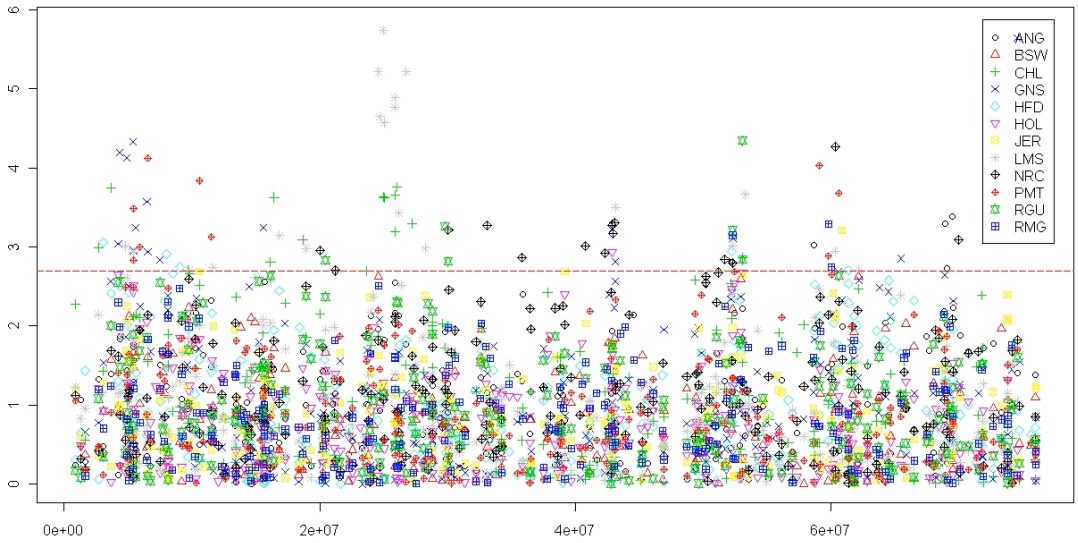


Figure D.14. Integrated extended haplotype homozygosity scores (|iHS|) on BTA14.

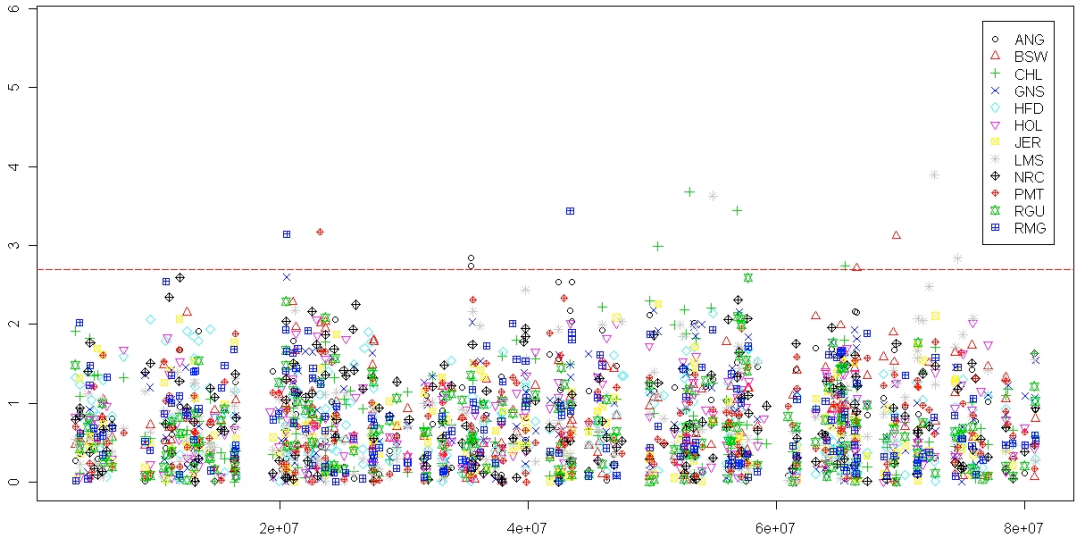


Figure D.15. Integrated extended haplotype homozygosity scores (|iHS|) on BTA15.

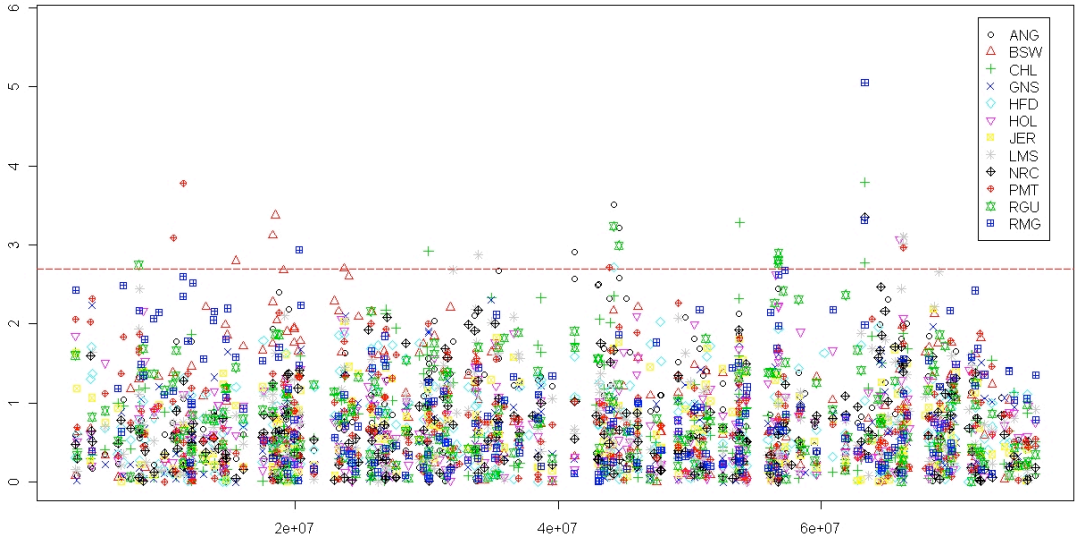


Figure D.16. Integrated extended haplotype homozygosity scores (|iHS|) on BTA16.

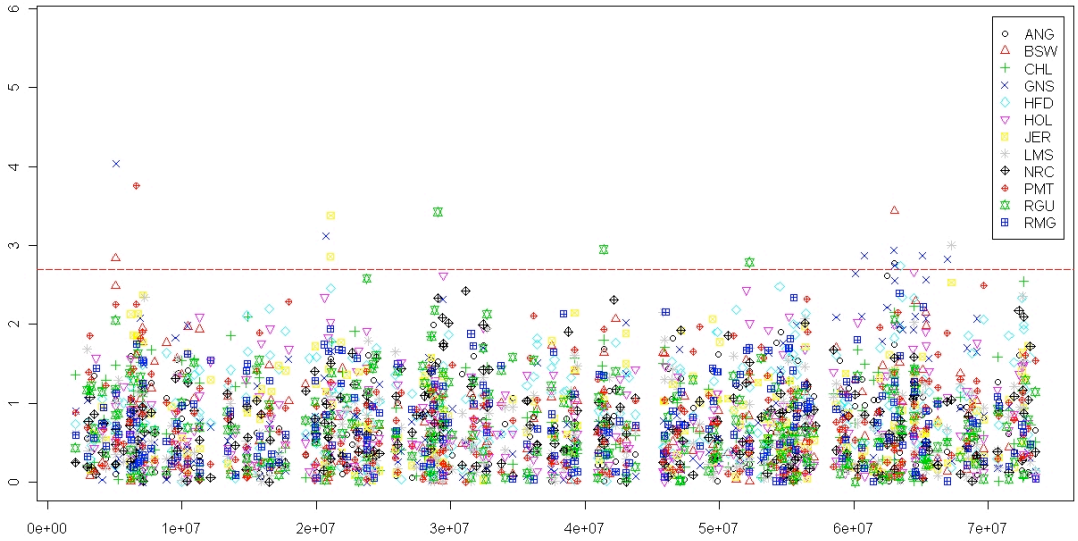


Figure D.17. Integrated extended haplotype homozygosity scores (|iHS|) on BTA17.

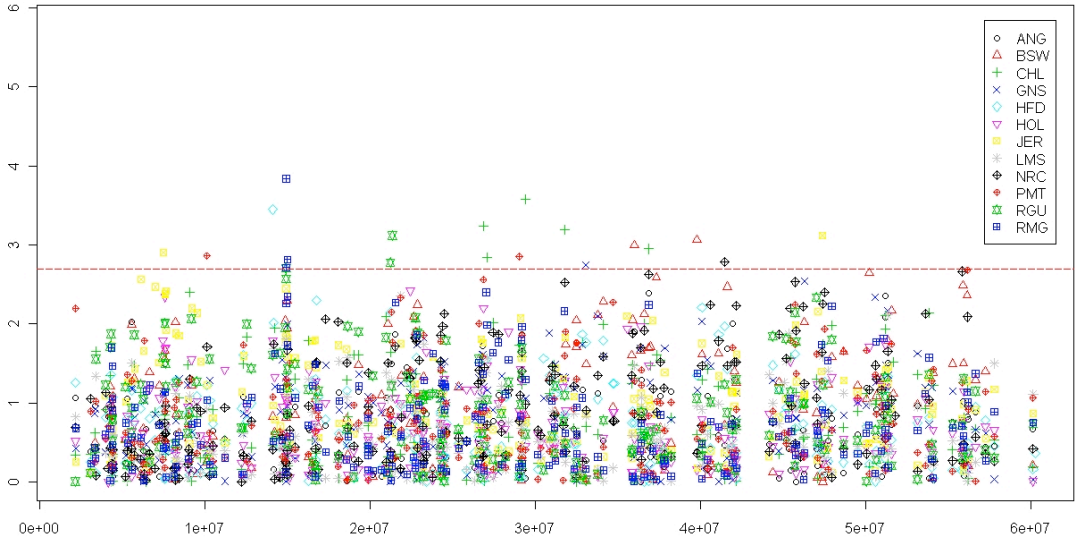


Figure D.18. Integrated extended haplotype homozygosity scores (|iHS|) on BTA18.

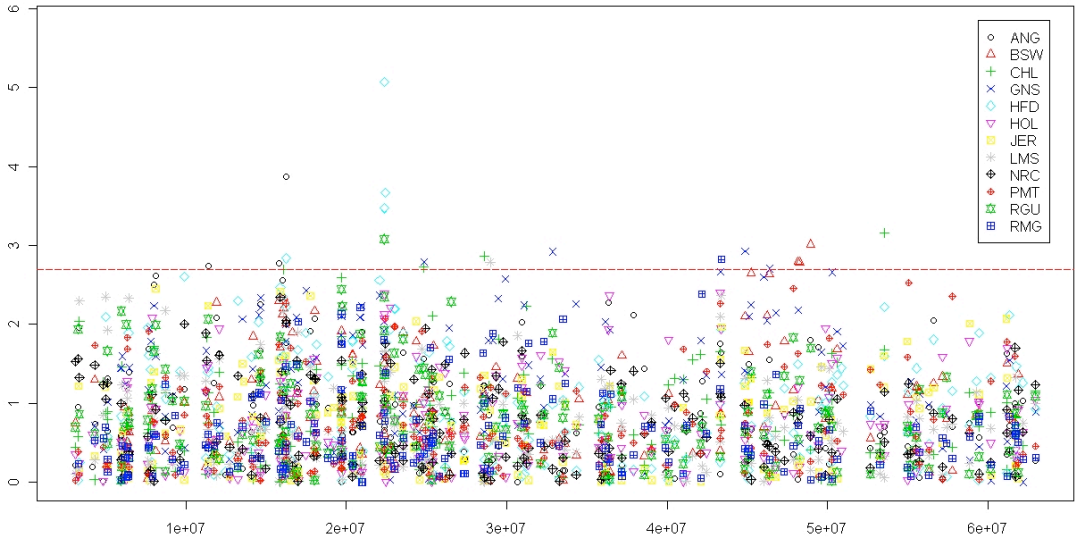


Figure D.19. Integrated extended haplotype homozygosity scores (|iHS|) on BTA19.

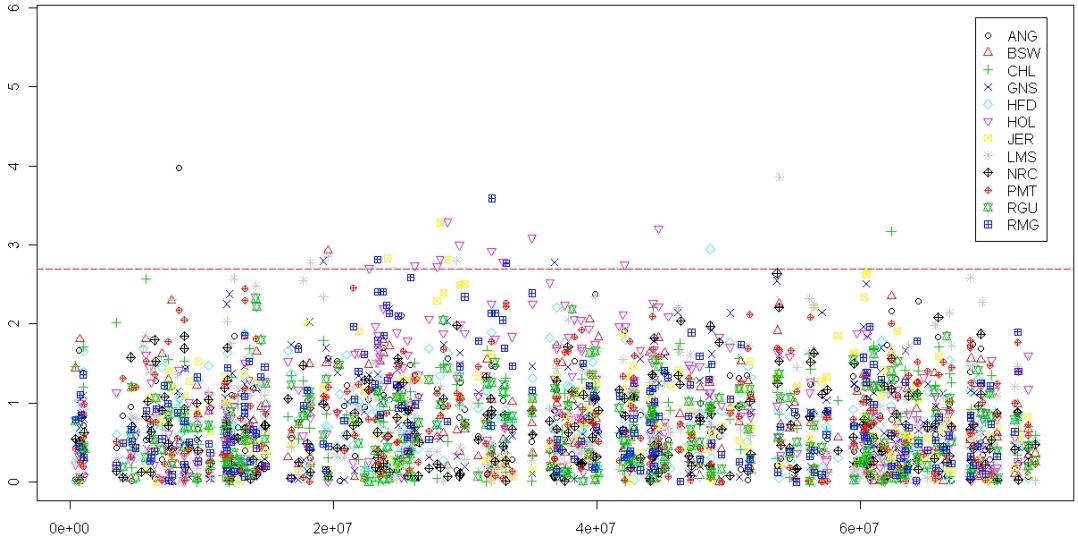


Figure D.20. Integrated extended haplotype homozygosity scores (|iHS|) on BTA20.

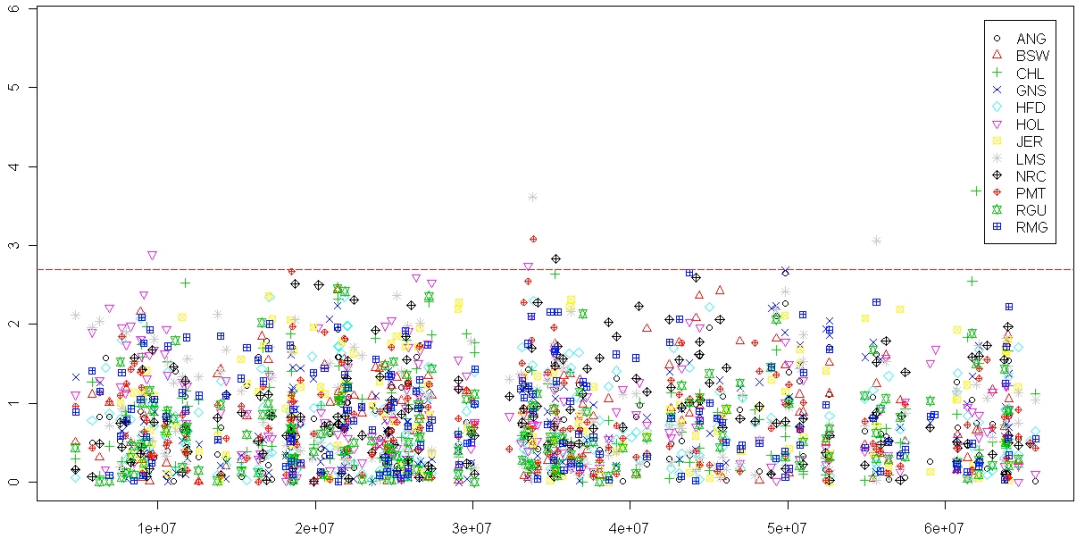


Figure D.21. Integrated extended haplotype homozygosity scores (|iHS|) on BTA21.

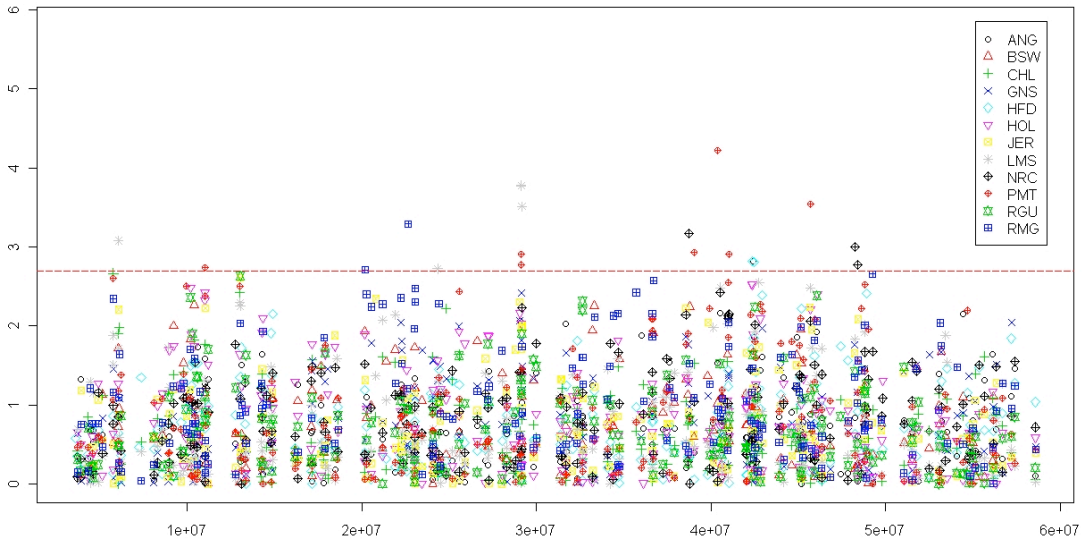


Figure D.22. Integrated extended haplotype homozygosity scores (|iHS|) on BTA22.

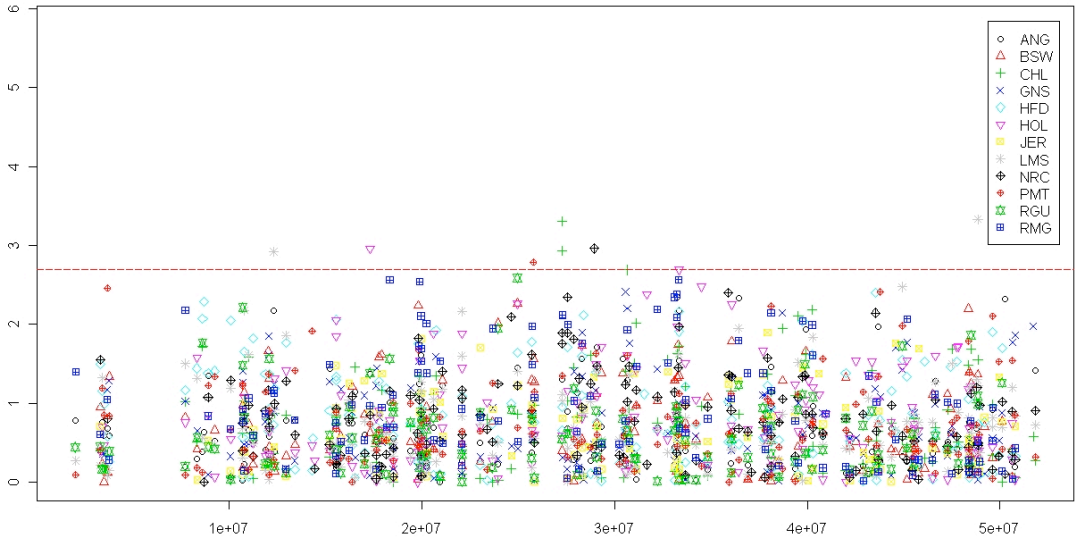


Figure D.23. Integrated extended haplotype homozygosity scores (|iHS|) on BTA23.

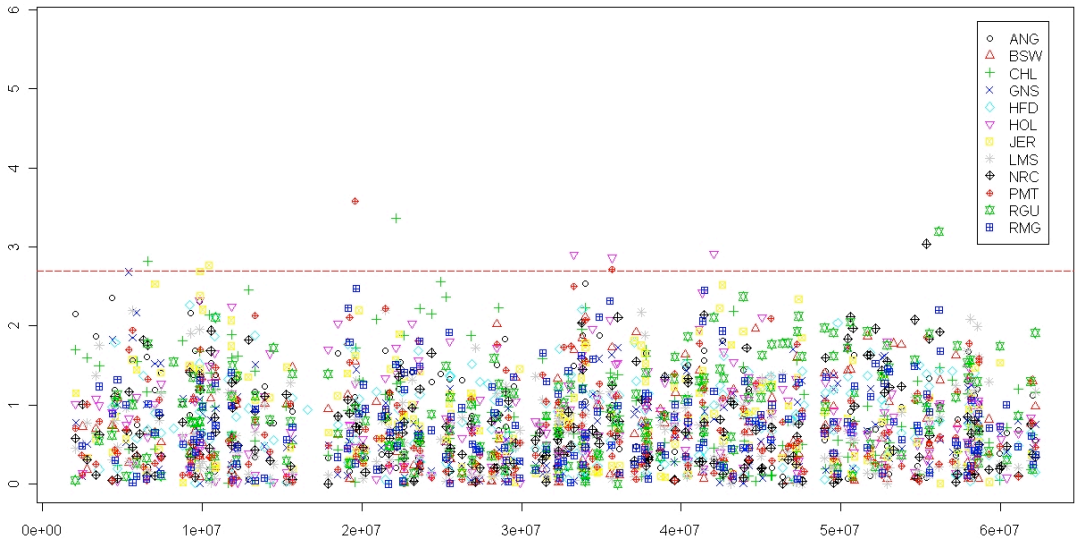


Figure D.24. Integrated extended haplotype homozygosity scores (|iHS|) on BTA24.

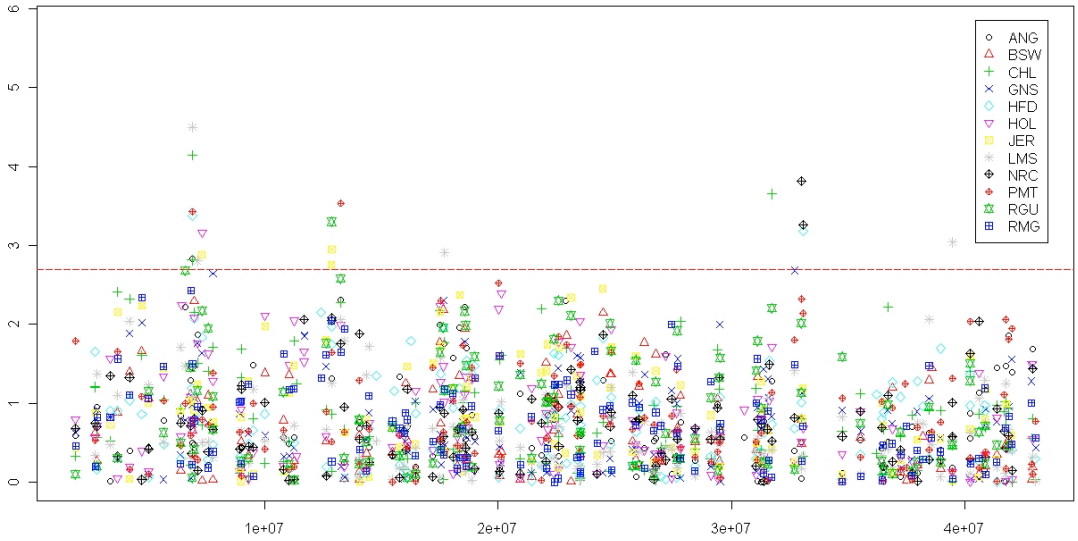


Figure D.25. Integrated extended haplotype homozygosity scores (|iHS|) on BTA25.

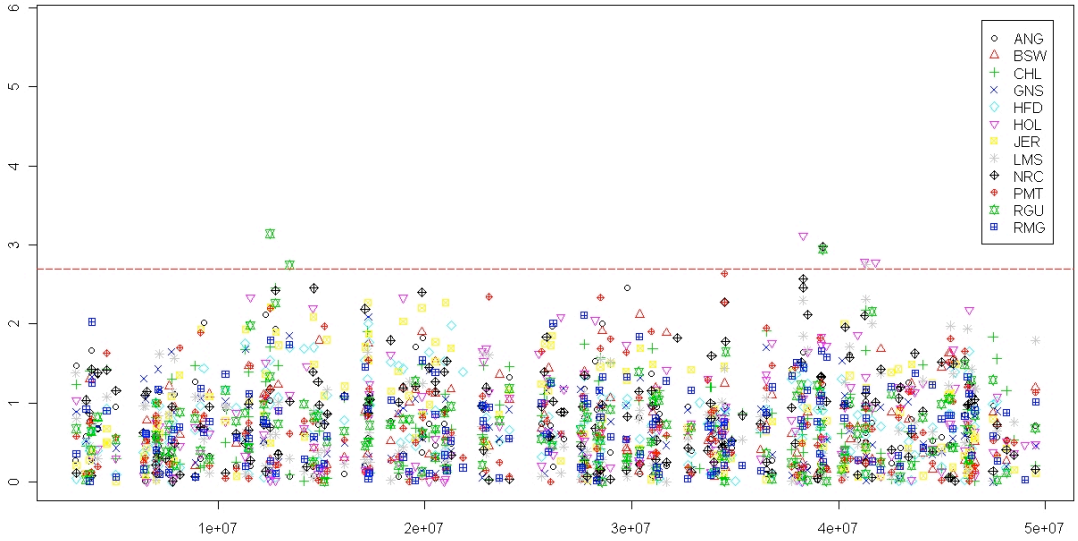


Figure D.26. Integrated extended haplotype homozygosity scores (|iHS|) on BTA26.

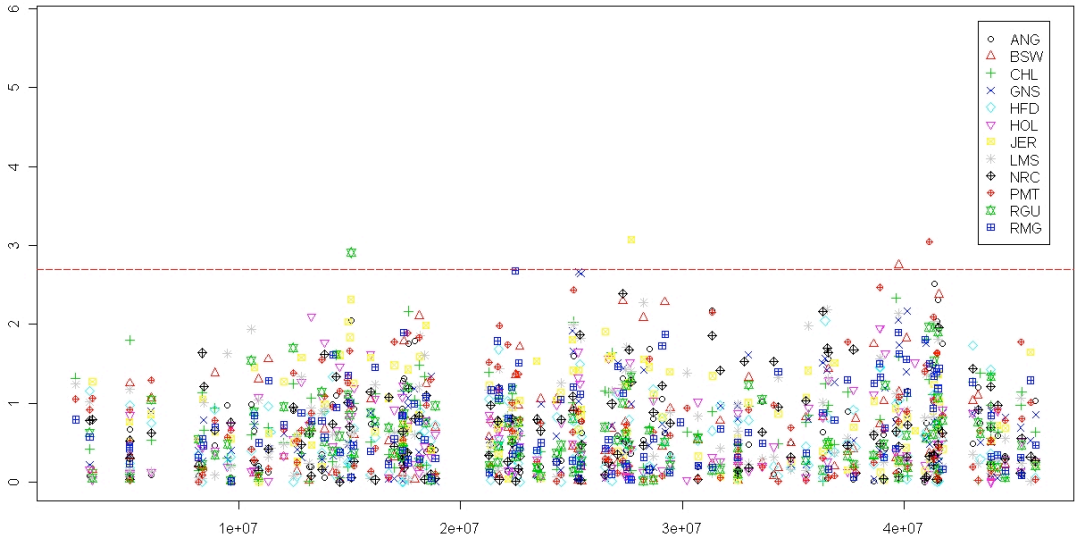


Figure D.27. Integrated extended haplotype homozygosity scores (|iHS|) on BTA27.

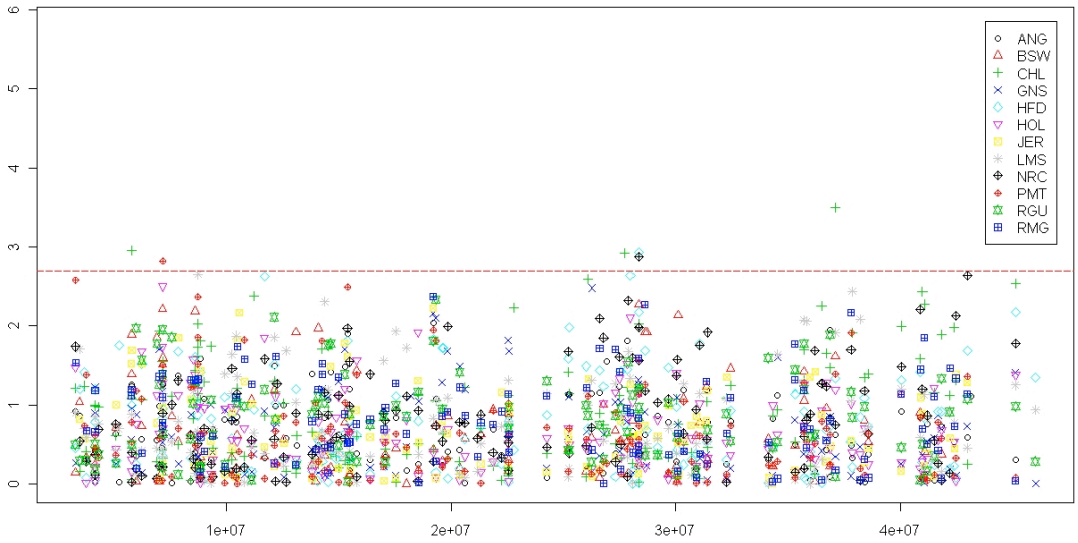


Figure D.28. Integrated extended haplotype homozygosity scores (|iHS|) on BTA28.

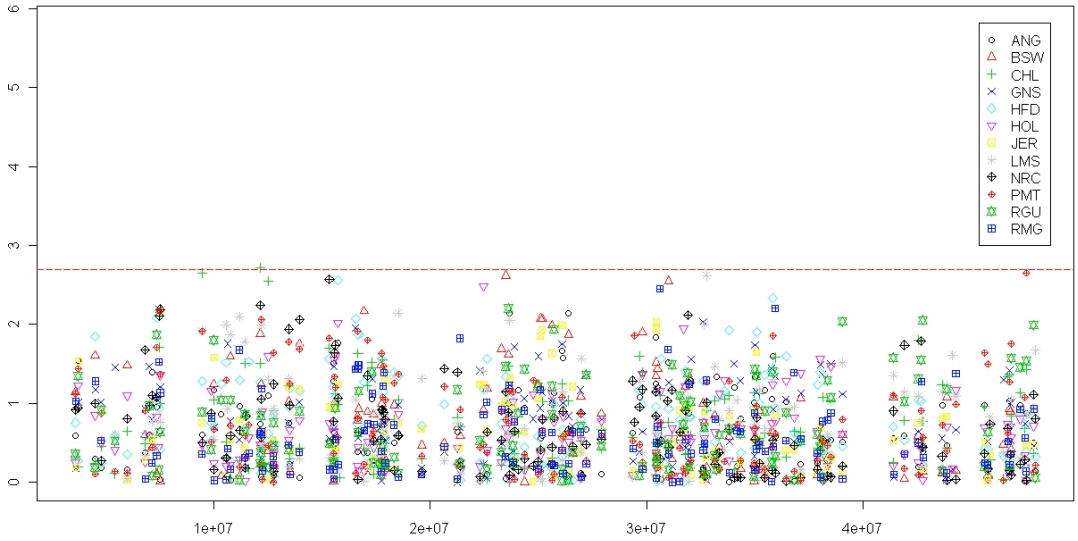


Figure D.29. Integrated extended haplotype homozygosity scores (|iHS|) on BTA29.

VITA

Name: Jung Woo Choi

Address: Animal Genomics Laboratory

2471 TAMU, College Station, TX 77843-2471

Email Address: choijungwoo@neo.tamu.edu

Education: B.S., Animal Science, Kang-Won National University, 2001

Ph.D., Animal Science, Texas A&M University, 2009