

# Detection and evaluation of selection signatures in sheep

Tiago do Prado Paim<sup>(1)</sup>, Patrícia Ianella<sup>(2)</sup>, Samuel Rezende Paiva<sup>(2)</sup>,  
Alexandre Rodrigues Caetano<sup>(2)</sup> and Concepta Margaret McManus Pimentel<sup>(3)</sup>

<sup>(1)</sup>Instituto Federal Goiano, Campus Iporá, Avenida Oeste, nº 350, Parque União, CEP 76200-000 Iporá, GO, Brazil. E-mail: tiago.paim@ifgoiano.edu.br <sup>(2)</sup>Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, W5 Norte (Final), Caixa Postal 02372, CEP 70770-917 Brasília, DF, Brazil. E-mail: patricia.ianella@embrapa.br, samuel.paiva@embrapa.br, alexandre.caetano@embrapa.br <sup>(3)</sup>Universidade de Brasília, Instituto de Biologia, Campus Universitário Darcy Ribeiro, Asa Norte, Caixa Postal 04508, CEP 70910-970 Brasília, DF, Brazil. E-mail: concepta@unb.br

**Abstract** – The recent development of genome-wide single nucleotide polymorphism (SNP) arrays made it possible to carry out several studies with different species. The selection process can increase or reduce allelic (or genic) frequencies at specific loci in the genome, besides dragging neighboring alleles in the chromosome. This way, genomic regions with increased frequencies of specific alleles are formed, characterizing selection signatures or selective sweeps. The detection of these signatures is important to characterize genetic resources, as well as to identify genes or regions involved in the control and expression of important production and economic traits. Sheep are an important species for these studies as they are dispersed worldwide and have great phenotypic diversity. Due to the large amounts of genomic data generated, specific statistical methods and softwares are necessary for the detection of selection signatures. Therefore, the objectives of this review are to address the main statistical methods and softwares currently used for the analysis of genomic data and the identification of selection signatures; to describe the results of recent works published on selection signatures in sheep; and to discuss some challenges and opportunities in this research field.

**Index terms:** *Ovis aries*, analysis software, animal genetic resources, genetic improvement, genomics, molecular genetics, SNP markers.

## Detecção e avaliação de assinaturas de seleção em ovinos

**Resumo** – O recente desenvolvimento de painéis de “single nucleotide polymorphisms” (SNPs) distribuídos ao longo do genoma possibilitou a realização de diversos trabalhos com diferentes espécies. O processo seletivo promove o aumento ou a diminuição da frequência alélica (ou gênica) em loci específicos do genoma, além de promover o arrasto dos alelos próximos no cromossomo. Desta forma, são formadas regiões do genoma com o aumento na frequência de determinados alelos na população, o que caracteriza a assinatura de seleção. A detecção destas assinaturas é importante para a caracterização dos recursos genéticos, bem como a identificação de genes ou regiões envolvidos no controle e na expressão de características de importância produtiva e econômica. Os ovinos são uma importante espécie para estes estudos, uma vez que encontram-se amplamente dispersos em diferentes ambientes e apresentam grande diversidade fenotípica. Devido à grande quantidade de dados gerados nas análises genômicas, são necessários métodos estatísticos e programas específicos para a detecção de assinaturas de seleção. Assim, os objetivos deste artigo de revisão são apresentar os principais métodos estatísticos e os programas atualmente utilizados para análise de dados genômicos e a detecção de assinaturas de seleção; descrever os resultados dos recentes trabalhos publicados sobre assinaturas de seleção em ovinos; e discutir alguns desafios e oportunidades nesta área de pesquisa.

**Termos para indexação:** *Ovis aries*, programas de análise, recursos genéticos animais, melhoramento genético, genoma, genética molecular, marcadores SNP.

## Introduction

Genomes from different species of animals, plants, insects, and microorganisms have been sequenced since the publication of the human genome sequence (Venter et al., 2001). This was facilitated by the technological revolution in sequencing methodologies observed in

the last decade, as well as by the identification of single nucleotide polymorphisms (SNPs) in great numbers (more than one million in some species) spread throughout the entire genome. Due to the intrinsic characteristics of these polymorphisms, it was possible to develop automated processes for genotyping of identified SNPs at a relatively low cost.



SNPs are usually biallelic in nature and can be genotyped with different platforms. Data generated with different platforms from different labs can be easily and accurately combined, which was very difficult to achieve with earlier technologies, such as microsatellites, and resulted in high error rates. Therefore, SNPs are currently the preferred markers for the majority of genomic studies (Grasso et al., 2014). There are many commercial platforms available, which have been optimized for high efficiency, robustness, and speed of data generation, achieving excellent cost-benefits, as, for example, the 777K SNP chip for cattle (BovineHD Genotyping BeadChip, Illumina, Inc., San Diego, CA, USA) and the 600K SNP chip for sheep (Ovine Infinium HD SNP BeadChip, Illumina, Inc., San Diego, CA, USA).

The domestication and selection processes of domestic animals occurred during a relative short period (last ~10.000 years), leading to a wide phenotypic diversity in traits related to production and type, with great differences observed between breeds (Grasso et al., 2014). According to Andersson (2012), animal and plant domestication is the most extensive genetic experiment that has been carried out by humans, resulting in huge shifts in the frequency of mutations that affect phenotypic traits.

Selection processes promote changes in allelic frequencies in specific loci depending on the selection pressures and objectives. Chromosome regions surrounding advantageous alleles are swept during this process, in a phenomenon termed hitchhiking effect (Fay & Wu, 2000). The process results in genomic regions with elevated homozygosity in the population, known as selection signatures or selective sweeps (Haas & Payseur, 2016) (Figure 1).

Chromosome crossing over and recombination processes promote the breakage of inherited chromosome segments, reducing the size of homozygous chromosome segments across generations (Kelley & Swanson, 2008). Therefore, the length of chromosome segments swept due to the presence of a positive allele is inversely proportional to number of selection generations, consequently affecting the size of the fragments in a homozygous state.

Boman et al. (2011) postulated that hitchhiking can result in the introduction of undesirable traits in a population, as detrimental alleles in other genes inside the favorable haplotype may be passed on

across generations. According to these authors, studies of selective sweeps could not only identify alleles under selection, but also provide information about deleterious genes selected altogether on a certain haplotype.

The correct experimental design and recording of phenotypic data (such as body measures and production indices) are important to any genomics study. Presently, high-density genotyping data are relatively easy to generate, but data management and statistical analyses of such large datasets are a challenge (Cadzow et al., 2014).

This review addresses the main statistical methods and softwares currently used for the analysis of genomic data and the identification of selection signatures; describes the results of recent works published on selection signatures in sheep; and discusses some challenges and opportunities in this research field.

### Selection signatures

Selection processes affect patterns of genetic variation in the neighborhood of selected loci; therefore, resulting variation patterns in these regions differ from the expected for neutral markers. The approach used to identify these patterns is known as hitchhiking mapping (Schlötterer, 2006). Overall, major differences between populations are attributed to a low number of segregating sites or to some sites with frequency deviations (low or high frequencies of derived alleles, not wild) and/or differentiated linkage disequilibrium structures (Boitard et al., 2009).

Mean effective population sizes, number of generations under selection, recombination rates, relative age of neutral linked alleles, bottleneck, and founder effects can affect the identification of selection signatures (Utsunomiya et al., 2015). The type and intensity of selection (Gouveia et al., 2014), as well as the selection coefficients applied (Kelley & Swanson, 2008), are factors that also directly affect the detection of these signatures. Both natural and artificial selection processes can act in three ways: positive, purifying, and balancing; each corresponding to a form of response in which selection pressures act differently to alter allelic and genotypic frequencies (Oleksyk et al., 2010). Positive selection occurs when newly arisen mutations confer selective advantages. Purifying selection, also known as negative or background selection, happens

when new mutations are disadvantageous and, therefore, tend to be eliminated. Balancing selection acts to maintain the polymorphism in a population, and can be observed when heterozygotes have selective advantages or when alleles are favored at different time intervals, for example (Gouveia et al., 2014).

Statistical tests developed to detect selective sweeps are based on neutral theory, and presume that the allele frequencies found in the genome are maintained across generations if loci do not have any effect on the traits under selection pressure and also that new mutations arise randomly in the genome. Changes in the allele frequency spectrum, as a new-arisen mutation, out of the pattern of the rest of the genome or increases in the frequency of one specific allele are supposed to be a signal of selection at the respective loci.

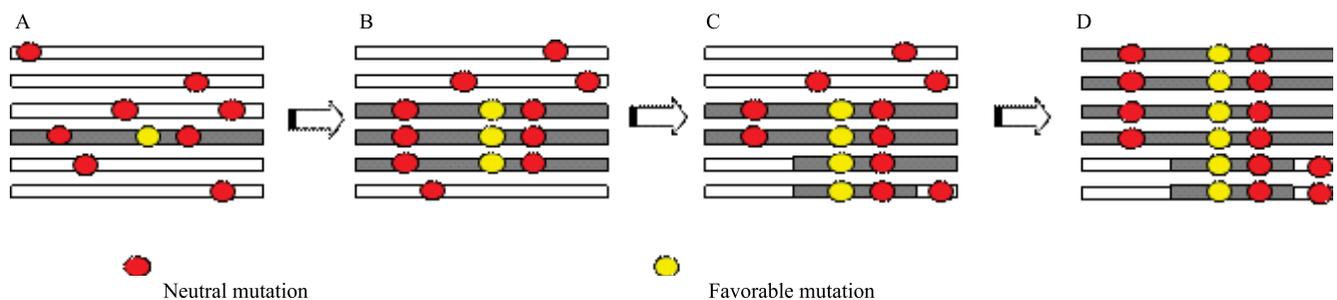
Natural demographic events, such as genetic drift, bottlenecks, expansion/subdivision of populations and migrations, can violate some assumptions of neutral theory, eventually leading to observed genomic signals that are similar to selection signatures (Gouveia et al., 2014). In general, statistical tests assume that high selection pressures are the main explanation for statistically significant results. However, the existence of other factors related mainly to sampling effects or unknown sample sub-structuring can artificially modify allelic frequencies, resulting in false positives (Cadzow et al., 2014). Therefore, for these authors, prior knowledge of the evolutionary histories of the populations included in a study and, especially, of the samples in the analysis is essential to evaluate the obtained results.

According to Moradi et al. (2012), demographic events and the bias analysis change the frequency pattern throughout the whole genome in a similar way, while selection events change allelic frequencies only in loci under selection and close regions. Data derived from millions of SNPs from a large number of individuals in a population (> 1,000) can help overcome these factors and generate great opportunities to distinguish between effects derived from population structure, positive selection, and bias analysis, leading to decreases in false positive results.

### Detection methods of selective sweeps

Several methods are proposed for the detection of genomic regions under selection. The choice of an adequate method depends on the time scale in which selection is expected to happen, the number of populations expected in the study, and the type of selection to which the population was subjected to (Gouveia et al., 2014). The most commonly used methods can be grouped into four main categories, based on: substitution rates, such as the MKT,  $K_a/K_s$ , and  $d_N/d_S$  tests; frequency spectrum, including the Tajima's D and Fay and Wu's H tests; linkage disequilibrium, represented mainly by the LRH, EHH, XP-EHH, Rsb, and iHS tests; and differences between populations, as, for example, the  $F_{ST}$ , LKT, LSBL, and hapFLK tests (Oleksyk et al., 2010; Gouveia et al., 2014).

The methods based on substitution rates assume that synonymous substitutions (silence shifts, without alterations in the sequence of amino acids)



**Figure 1.** Selective sweep with recombination: A, six chromosomal regions with neutrally segregating alleles (red), as well as one adaptive allele (yellow) in the gray chromosome; B, snapshot of the region during the sweep, showing the decrease in the variability of the region around the favorable allele; C, another snapshot during the advance of the selective process, where crossing chromosomes can be seen; and D, the result after the complete sweep with the fixation of favorable alleles in the region.

are selectively neutral and that the non-synonymous substitutions (amino acid shifts) are potentially selectable (Hohenlohe et al., 2010). If the non-synonymous substitution rate inside a gene differs significantly from that of the synonymous one, it is a signal that selection occurred. Consequently, in a neutral situation, the coding sequence of a gene shows the following relationship:  $d_N/d_S=1$ , where  $d_N$  is the non-synonymous substitution rate and  $d_S$  is the synonymous substitution rate; when there is a positive selection,  $d_N/d_S>1$ , and a negative selection,  $d_N/d_S<1$  (Yang et al., 2010).

The methods based on the change in the frequency spectrum consider two distinct approaches: spectrum shifts, as, for example, grouping of rare alleles in regions, which may be explored by Tajima's D statistics; or shifts in the frequency of ancestral and derived alleles, assuming that ancestral alleles are known, which may be obtained by Fay and Wu's H test (Oleksyk et al., 2010). Tajima's D statistics (Tajima, 1989) is useful in the identification of signatures related to a high number of alleles in a low-to-medium frequency (Cadzow et al., 2014). The Fay e Wu's H test, in turn, is useful to detect evidence of a more recently positive selection, mainly for alleles in a medium-to-high frequency (Sabeti et al., 2006), complementing Tajima's D and other methods.

The patterns of linkage disequilibrium are the focus of many tests to detect various types of selection. The genome-wide haplotype maps generated with different methods are used to identify evidences of the selective process across time. The extended haplotype homozygosity (EHH) is defined as the probability that two randomly chosen chromosomes carrying the same haplotype (as assayed by homozygosity at all SNPs) are identical by descent for the entire interval from the core region to point x (Sabeti et al., 2002). If the core allele is under selection, then the EHH will be near one throughout a long distance in the surrounding regions of the SNP. The concept of EHH has been applied in various studies with domestic species (Qanbari et al., 2014; Randhawa et al., 2014; Wei et al., 2015; Zhu et al., 2015).

According to Gautier & Vitalis (2012), the EHH test can be strongly influenced by the demographic history of the population in analysis, yielding a high number of false positives. For this reason, Voight et al. (2006) proposed a test based on the integral of

the observed decay of the EHH, named as integrated EHH (iHH), aiming to minimize eventual effects of unknown demographic factors. These same authors defined another statistical test – integrated haplotype homozygosity score (iHS) – that uses the iHH logarithm ratio calculated in the derived and ancestral allele of the focal SNP; in this way, the obtained value is less affected by the demographic history of the population. The iHS is standardized using the average and standard deviation of all SNPs with a similar allele frequency. This method has great potential in selection signature studies because it provides a standardized measure of the EHH decay around the derived allele in relation to the ancestral allele (Fariello et al., 2013). Regions in which homozygosity decreases slowly in relation to the derived allele – with greater homozygosity than expected in relation to the ancestral allele – are indicative of selection at that locus.

The estimates for the identification of selection signatures based only on data within a population have low resolution power when the frequency of the markers (SNPs) linked to the positive selected allele is already high ( $>0.7$ ). Tang et al. (2007) created a procedure, termed EHH site specific (EHHS or XP-EHH), to compare the EHH profile between populations by computing a weighted average for each SNP in each population.

Fagny et al. (2014) observed that the statistics based on the EHH had a higher capacity to detect the effects of the selection signatures on a wide range of allelic frequencies obtained by resequencing data of the entire genome, independently of the demographic history of the population evaluated. In this case, the SNP density available for the detection of estimates is maximum, which could affect the results. However, the methods based on linkage disequilibrium had a weak detection power for historical signatures of ancestral selection, i.e., that finished thousands of generations ago, because the linkage disequilibrium between SNPs is broken relatively quickly over time (Chen et al., 2010).

The  $F_{ST}$  has been widely used to detect selection signatures when there are two or more populations in the dataset (Fariello et al., 2013, 2014; McRae et al., 2014; Benjelloun et al., 2015; Xu et al., 2015; Zhao et al., 2015). This method adopts the difference in allelic frequency between populations to infer directional selection in one population in relation to another (Sabeti et al., 2006).

It should be noted that selection timing has a huge impact on the capability of each method to detect the selection signature. The methods based on absolute frequencies and frequency differences between populations are better to detect long-term events, i.e., with more than 1,000 generations, since they depend on the accumulation of mutations around the causative mutation. In situations where the adaptive advantage is small – mainly if it is recessive –, the time required for the frequency of this allele to increase to the point of detection is much higher, which decreases the detection power of these methods. The methods based on linkage disequilibrium had a higher detection power when a new mutation arised due to an adaptive advantage or to a previous variation exposed to a new environment that provides a favorable selective pressure; therefore, there is an increase in the frequency of this new allele, but without fixation, in the population (Fagny et al., 2014).

Considering each individual method has limitations and can cause some biases, combined methods have been proposed in recent years, as for example XP-CLR (Chen et al., 2010), hidden Markov models (Boitard et al., 2009), and Pool-hmm (Boitard et al., 2012, 2013). The main idea is to explore the strength of each method and minimize biases, avoiding false positive or negative results.

Grossman et al. (2010) obtained greater precision in identifying selection signature regions with the use of composite multiple signals (CMS), compared with XP-EHH,  $F_{ST}$ , iHH, iHS, and  $\Delta DAF$ ; this score was developed in a study of the differences between the frequencies of the derived allele in the non-selected population and in the putative selected population. The CMS combines the results of various tests for multiple signals of selection and increases the resolution of the result up to 100 times. Fariello et al. (2013) used another method, named as hapFLK, focusing on the difference between the haplotype frequencies between populations and accounting for the hierarchical structure of the sampled population. Ferrer-Admetlla et al. (2014) proposed the use of the number of segregating sites by length ( $nS_L$ ), which is statistically based on the length of haplotypes, for the detection of signatures in genomic data of a unique population.

## Softwares for the analysis of genomic data

The genomic data pattern differs between platforms from different chip manufacturers, which hampers the exchange and integration of data. Moreover, the softwares generally require specific formats for input files (Cadzow et al., 2014). Therefore, the researcher needs some computing abilities to format the data from banks that are frequently very large (up to 700,000 columns). Nicolazzi et al. (2015) highlight the chaotic situation faced by every researcher due the lack of data standardization. In this context, software have been developed specifically for data transformation, such as PGDSpider (Table 1), which is a basic data conversion tool for connecting population genetics and genomics software.

Besides data formatting, another difficulty faced by researchers is the identification of the correct computational tool to reach their objectives. Currently, the creation of specific pipelines for each study is being investigated. Cadzow et al. (2014), for example, showed a collection of scripts capable of implementing each step of the identification of selective sweeps.

The detection of selective sweeps using some of the cited methods generally begins by managing data into the required format, which may involve imputation, i.e., the estimation of missing genotypes. Sometimes, the detection of haplotypes with or without linkage disequilibrium pruning is also necessary to reduce the number of markers and avoid bias.

Some of the software available to perform all the required steps or at least a great part of them, include: PLINK (Purcell et al., 2007), JMP Genomics (SAS Institute Inc., Cary, NC, USA), and Golden Helix SNP & Variation Suite (SVS, Golden Helix, Bozeman, MT, USA). JMP Genomics and SVS are excellent softwares with several tools for data management, analysis, and visualization. Recently, an imputation tool was added to SVS. However, SVS and JMP Genomics are licensed programs that require payment by all users. Conversely, PLINK is a free software widely used and with a high number of tools, but requires familiarity with the command line interface and sometimes it is necessary to use other software for data visualization, such as Haploview (Broad Institute, Cambridge, MA, USA). More recently, gPLINK was launched, which is a java-based software package that creates a friendly way to integrate results with Haploview.

Generally, the software choice is guided by the data analysis method that the researchers desire to apply. Sometimes the name of the software and the name of the method is the same, as is the case for XP-CLR, REHH, and hapFLK (Table 1). XP-CLR, also known

as the cross-population composite likelihood ratio test, is a method based on the differentiation of multilocus allele frequencies between two populations. Other examples of software include OmegaPlus, which is a scalable implementation of the omega-statistics based

**Table 1.** Softwares for analysis of genomic data from formatting until analysis of population genomics and selective sweeps.

Function	Name	OS <sup>(1)</sup>	License <sup>(2)</sup>	Link or source
SNP data management <sup>(3)</sup>	PLINK 1.07	W/L/M	F	<a href="http://zzz.bwh.harvard.edu/plink/download.shtml">http://zzz.bwh.harvard.edu/plink/download.shtml</a>
	PLINK 1.90	W/L/M	F	<a href="https://www.cog-genomics.org/plink2">https://www.cog-genomics.org/plink2</a>
	snpQC	W/L/M	OS	<a href="http://www-personal.une.edu.au/~cgondro2/snpQC.htm">http://www-personal.une.edu.au/~cgondro2/snpQC.htm</a>
	JMP Genomics	W/L/M	L	<a href="http://www.jmp.com/software/genomics/">http://www.jmp.com/software/genomics/</a>
	Golden Helix SNP & Variation Suite	W/L/M	L	<a href="http://www.goldenhelix.com/SNP_Variation/index.html">http://www.goldenhelix.com/SNP_Variation/index.html</a>
	Progeny	W/L	L	<a href="http://www.progenygenetics.com/clinical/trial">http://www.progenygenetics.com/clinical/trial</a>
	BC/GENOME	L	L	<a href="http://www.bcplatforms.com/news/product-category/academia/#bcgenome-2">http://www.bcplatforms.com/news/product-category/academia/#bcgenome-2</a>
	SNPpy	L	OS	<a href="https://bitbucket.org/faheem/snpypy">https://bitbucket.org/faheem/snpypy</a>
	GBrowse	L/M	OS	<a href="http://gmod.org/wiki/GBrowse">http://gmod.org/wiki/GBrowse</a>
	IGV	W/L/M	OS	<a href="http://www.broadinstitute.org/igv/">http://www.broadinstitute.org/igv/</a>
	PGDSpider	W/L/M	OS	<a href="http://www.cmpg.unibe.ch/software/PGDSpider/">http://www.cmpg.unibe.ch/software/PGDSpider/</a>
	fcGENE	W/L	OS	<a href="http://sourceforge.net/projects/fcgene/">http://sourceforge.net/projects/fcgene/</a>
	Python	W/L/M	F	<a href="https://www.python.org/">https://www.python.org/</a>
	Multicore <sup>(4)</sup>	W/L/M	F	<a href="https://cran.r-project.org/src/contrib/Archive/multicore/">https://cran.r-project.org/src/contrib/Archive/multicore/</a>
Imputation	FImpute	L/M	F	<a href="http://www.aps.uoguelph.ca/~msargol/fimpute/">http://www.aps.uoguelph.ca/~msargol/fimpute/</a>
	Beagle	L/W/M	F	<a href="http://faculty.washington.edu/browning/beagle/beagle.html">http://faculty.washington.edu/browning/beagle/beagle.html</a>
	IMPUTE2	L/W/M	F	<a href="https://mathgen.stats.ox.ac.uk/impute/impute_v2.html">https://mathgen.stats.ox.ac.uk/impute/impute_v2.html</a>
	MACH	L/W/M	F	<a href="http://www.sph.umich.edu/csg/abecasis/MACH/tour/imputation.html">http://www.sph.umich.edu/csg/abecasis/MACH/tour/imputation.html</a>
	PedImpute	L/M	OS	<a href="http://dekoppel.eu/pedimpute/">http://dekoppel.eu/pedimpute/</a>
	ALPHAPHASE	W/L/M	F	<a href="https://sites.google.com/site/hickeyjohn/alphaphase">https://sites.google.com/site/hickeyjohn/alphaphase</a>
Population genomics and selective sweeps	FINDHAP	L	OS	<a href="http://aipl.arsusda.gov/software/findhap/">http://aipl.arsusda.gov/software/findhap/</a>
	Sweed	L	OS	<a href="http://sco.h-its.org/exelixis/web/software/sweed/index.html">http://sco.h-its.org/exelixis/web/software/sweed/index.html</a>
	Arlequin	W/L/M	F	<a href="http://cmpg.unibe.ch/software/arlequin35/">http://cmpg.unibe.ch/software/arlequin35/</a>
	Selscan	W/L/M	F	<a href="https://github.com/szpiech/selscan">https://github.com/szpiech/selscan</a>
	VCFtools	L/M	F	<a href="http://vcftools.sourceforge.net/">http://vcftools.sourceforge.net/</a>
	BayeScan	L/M	F	<a href="http://cmpg.unibe.ch/software/BayeScan/index.html">http://cmpg.unibe.ch/software/BayeScan/index.html</a>
	ADMIXTURE	L/M	F	<a href="https://www.genetics.ucla.edu/software/admixture/">https://www.genetics.ucla.edu/software/admixture/</a>
	fastStructure	W/L/M	F	<a href="http://rajanil.github.io/fastStructure/">http://rajanil.github.io/fastStructure/</a>
	LFMM	W/L/M	F	<a href="http://membres-timc.imag.fr/Olivier.Francois/lfmm/index.htm">http://membres-timc.imag.fr/Olivier.Francois/lfmm/index.htm</a>
	MatSAM	W	F	<a href="http://www.econogene.eu/software/sam/">http://www.econogene.eu/software/sam/</a>
	DIYABC	W/L/M	F	<a href="http://www1.montpellier.inra.fr/CBGP/diyabc/">http://www1.montpellier.inra.fr/CBGP/diyabc/</a>
	popABC	W/M/L	F	<a href="https://code.google.com/p/popabc/">https://code.google.com/p/popabc/</a>
	OmegaPlus	W/L	F	<a href="http://www.exelixis-lab.org/software.html">http://www.exelixis-lab.org/software.html</a>
	Pool-HMM	W/L	F	<a href="https://qgp.jouy.inra.fr">https://qgp.jouy.inra.fr</a>
	XP-CLR	W/L	F	<a href="https://reich.hms.harvard.edu/software">https://reich.hms.harvard.edu/software</a>
	hapFLK	W/L	F	<a href="https://forge-dga.jouy.inra.fr/projects/hapflk/files">https://forge-dga.jouy.inra.fr/projects/hapflk/files</a>
Sweep	W/L	F	<a href="http://www.broadinstitute.org/mpg/sweep/">http://www.broadinstitute.org/mpg/sweep/</a>	
REHH <sup>(4)</sup>	W/L	F	<a href="http://cran.r-project.org/">http://cran.r-project.org/</a>	
SweepFinder	W/L	F	Huber et al. (2016)	
VariScan	W/L/M	F	<a href="http://www.ub.edu/softevol/variscan/">http://www.ub.edu/softevol/variscan/</a>	

<sup>(1)</sup>OS, operational system: W, Windows; L, Linux, and M, MacOS. <sup>(2)</sup>License: OS, open source; F, free for all users; and L, licensed, requiring payment by all users. <sup>(3)</sup>Several softwares of data management can be used in multiple categories, such as, for example, PLINK, JMP Genomics, Golden Helix SNP & Variation Suite. <sup>(4)</sup>Package of the R software. Source: adapted from Nicolazzi (2015) and Cadzow et al. (2014).

on linkage disequilibrium (Kim & Nielsen, 2004), and Sweep (Broad Institute, Cambridge, MA, USA), which uses the long range haplotype test to look for alleles of high frequency with long-range linkage disequilibrium.

Softwares for the detection of selective sweeps also have specific tools for different types of studies. MatSAM (Table 1), for example, can be used to associate molecular markers and environmental variables, and is adopted for landscape genomic studies. Latent factor mixed models (LFMM) allow to simultaneously estimate the effects of environmental factors and neutral genetic structures on allele frequencies; additionally, computing time is reasonably fast, making this method attractive for studies with whole genomes or subsets of large random batches of SNPs in parallel (Rellstab et al., 2015). Pool-HMM aims at estimating allele frequencies and detecting selective sweeps, using next-generation sequence data from a sample of pooled individuals from the same population.

Some R packages are useful for these selective sweep analyses, as Multicore and REHH (Table 1). The Multicore package can be used for data management, and the REHH package to perform selective sweep detection by the method of EHH. These could be interesting options for a researcher familiarized with the R language.

Other points that could limit the use of a specific software are the operational system and the computational time required. Most software are available in Unix platform (as Linux), some of them have a version for Windows, but few are available for use in MacOSX (Table 1). The ADMIXTURE software, for example, uses the same statistical model as Structure, but calculates estimates much more rapidly through a fast numerical optimization algorithm. Therefore, ADMIXTURE is more useful in large datasets, yielding similar results in a shorter time. Sweed is also a faster version of SweepFinder.

### Selective sweeps in sheep

Several studies have been successful in identifying selection signatures in humans and bovines (Voight et al., 2006; Yang et al., 2010; Druet et al., 2013; Qanbari et al., 2014). In other species, such as swine, poultry, and sheep, the low value of one adult animal implies that the genomic research must be conducted in a

different way. For swine and poultry, for example, highly specialized nuclei have been used, which aggregated value to the evaluated animals. However, sheep production chains worldwide do not have the same level of specialization and investment in genetic breeding, which is attributed to the wide distribution of sheep in small flocks. Despite this, in recent years, some studies have been performed to identify selective sweeps in the species (Table 2), considering that sheep production represents a great proportion of agricultural production in many countries, such as Australia, New Zealand, and those in North Europe and in the Mediterranean (Moioli et al., 2013).

Since their domestication, approximately 8,000–9,000 years ago, sheep have been used by humans for wool, mutton, and milk production (McManus et al., 2010). The adaptation of the species to a high variety of geographical and climatic conditions and its specialization for specific traits, such as the production of meat, milk or pelt, has resulted in a wide phenotypic diversity. The breeding for wool production started only 2,000 years after domestication (Gutiérrez-Gil et al., 2014), and the first phenotypic modifications in sheep were registered in an illustration 3,000 years before Christ, which showed leg length reduction, tail elongation, and horn format alterations (Gutiérrez-Gil et al., 2014). These widely diverse phenotypic traits, resulting from millenary breeding history, turn sheep into an important source for studies of selective sweeps.

Boman et al. (2011) showed that selection based on progeny tests was able to induce a fast change in allele frequency, even for balanced and wide selection. The 3'-UTR mutation in the myostatin gene (c.\*1232G>A), previously found affecting muscularity in Texel animals, was also observed in the Norwegian White Sheep (NWS) population. This mutation has increased in allele frequency, from 0.31 in 1990 to 0.82 in 2006 in the NWS breed. A higher increase was verified in genetic values for weight estimated by the best linear unbiased prediction (Blup) method since 1991 and after the adoption of the carcass evaluation system in 1996 (Boman et al., 2011). Fariello et al. (2013) identified a selection signature in the 17-Mb region of chromosome 2 in Texel animals from Germany, New Zealand, and Scotland, which can be related to the mutation in the myostatin gene (GDF-8), located in the center of the region.

**Table 2.** Studies of selective sweeps in sheep carried out in the last five years.

Reference	Population or breed (number of individuals)	Objectives	Methods and softwares applied <sup>(1)</sup>
Fariello et al. (2013)	Sheep HapMap	To validate the hapFLK method to detect selective sweeps	hapFLK, FLK, $F_{ST}$ , and hap $F_{ST}$
Fariello et al. (2014)	Sheep HapMap	To confirm the selective sweeps found by $F_{ST}$ and identify new ones	hapFLK and FLK
Gutiérrez-Gil et al. (2014)	Sheep HapMap (5 milk breeds and 5 non-milk breeds)	To identify selective sweeps related to milk production	Pairwise $F_{ST}$ , observed heterozygosity, and regression analysis to detect asymptotic patterns of heterozygosity
Grasso et al. (2014)	Merino (110), Corriedale (108), and Creola (10)	To identify genetic diversity within and between three sheep breeds	fastStructure, PCA, and $F_{ST}$
Lv et al. (2014)	32 autochthonous breeds (1,224)	To characterize genetic effects of climatic adaptation, identifying related selective sweeps	Arlequin, PLINK, MatSAM, LFMM, SmartPCA, fastStructure, and Sweep
McRae et al. (2014)	Romney (180) and Perendale (149)	To identify selective sweeps, within and between two breeds, related to resistance or susceptibility to gastrointestinal nematodes	$F_{ST}$ , Peddrift, fastPHASE, EHH (Sweep v1.1), XP-EHH, and iHS (Pritchard)
Moioli et al. (2013)	Altamura (100)	To identify regions that affect milk production	Random animal effect and Fisher's exact test (SAS, (SAS Institute, Inc., Cary, NC, USA)
Moioli et al. (2016)	Sarda breed (100)	To identify candidate genes for immune response and relationship with paratuberculosis resistance	Effect of allelic substitution (SAS Institute, Inc., Cary, NC, USA)
Moradi et al. (2012)	Zel (47) and Lori-Bakhtiari (47) breeds, as well as Sheep HapMap (7 breeds)	To perform selective sweeps between fat and thin-tail Iranian breeds and to compare divergent breeds for this trait in Sheep HapMap	PCA using the R software, $F_{ST}$ , homozygosity, and fastPHASE
Zhu et al. (2015)	German Mutton (89), Dorper (47), and Sunit (12)	To identify selective sweeps in chromosome X in three sheep breeds	PLINK, Beagle, iHS, and $F_{ST}$
Randhawa et al. (2014)	37 pooled breeds (1,489) and 36 horned breeds (1,290); 3 double-muscle breeds (149) and 71 normal muscle breeds (2,654)	37 pooled breeds (1,489) and 36 horned breeds (1,290); 3 double-muscle breeds (149) and 71 normal muscle breeds (2,654)	$F_{ST}$ , XP-EHH, DAF, and CSS
Wang et al. (2015)	White Dorper (100), fat-tailed Chinese Mongolian (61), and German Mutton Merino (161)	White Dorper (100), fat-tailed Chinese Mongolian (61), and German Mutton Merino (161)	PCA, pairwise $F_{ST}$ , LSBL, and $d_i$
Wei et al. (2015)	10 Chinese breeds (140)	10 Chinese breeds (140)	PCA, fastStructure, neighbor-joining-tree, $d_i$ , Rsb, pairwise $F_{ST}$ (Genepop), and fastPHASE
Gorkhali et al. (2016)	24 sheep from each of the following 4 Nepalese breeds: Bhyanglung, Baruwai, Kage, and Lampuchhre	24 sheep from each of the following four Nepalese breeds: Bhyanglung, Baruwai, Kage, and Lampuchhre	PCA, pairwise $F_{ST}$ , and $d_i$
Yang et al. (2016)	77 Chinese native sheep from 21 representative breeds	77 Chinese native sheep from 21 representative breeds	Runs of homozygosity, $F_{ST}$ , XP-EHH, and LFMM
Manunza et al. (2016)	370 animals from 11 Spanish breeds	370 animals from 11 Spanish breeds	$F_{ST}$ -outlier approach in the BayeScan software, hapFLK, and FLK
Liu et al. (2016)	8 sheep populations with 20 individuals from each one	8 sheep populations with 20 individuals from each one	$H_p$ and $F_{ST}$
Zhao et al. (2016)	Sunite (66), German Mutton (159), and Dorper (93)	Sunite (66), German Mutton (159), and Dorper (93)	REHH and XP-EHH
Wei et al. (2016)	Hu (12), Tong (15), large-tailed Han (15), Lop (15), Tibetan (14), Sichuan (14), and Nagqu (37)	To study the adaptive evolution of high-altitude sheep by analyzing seven breeds	$F_{ST}$ and XP-EHH
Purfield et al. (2017)	Belclare (658), Beltex (64), Charollais (665), Suffolk (784), Texel (489), and Vendean (629)	To quantify the genetic diversity in six commercial sheep breeds with the aim of identifying genomic regions that have been subjected to selection	Runs of homozygosity, $F_{ST}$ , and hapFLK
Gouveia et al. (2017)	Brazilian Creole (22), Morada Nova (22), and Santa Inês (45)	To identify genomic regions that may have been under selection and, therefore, may explain ecological and production differences among three Brazilian locally adapted sheep breeds	$F_{ST}$ , Rsb, and iHS
Yuan et al. (2017)	Hu (12), Tong (15), large-tailed Han (15), Lop (15), Tibetan (14), Sichuan (14), and Nagqu (37)	To identify genes associated with tail fat deposition in Chinese populations	$F_{ST}$ and hapFLK

<sup>(1)</sup>FLK, test for the detection of selection signatures based on the LK statistics; hap $F_{ST}$ , haplotype extension of the  $F_{ST}$  test;  $F_{ST}$ , Wright's fixation index; fastStructure, software package for using multi-locus genotype data to investigate population structure; PCA, principal component analysis; Arlequin, an integrated software for population genetics data analysis; PLINK, whole-genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner; MatSAM, spatial analysis software to detect candidate loci for selection; LFMM, latent factor mixed model; Sweep, software for large-scale analysis of haplotype structure in genomes for the primary purpose of detecting evidence of natural selection; Peddrift (Dodds & McEwan, 1997), a method that incorporates pedigree data in the analysis and compensates for genetic drift; fastPHASE, a software for haplotype reconstruction and estimation of the missing genotypes from population data; EHH, extended haplotype homozygosity; XP-EHH, cross-population extended haplotype homozygosity; iHS, integrated haplotype score; Beagle, a software package used for genotype calling and phasing, imputation of ungenotyped markers, and identity-by-descent segment detection; DAF, increase in derived allele frequency; CSS, composite selection signal; LSBL, locus-specific branch lengths;  $d_i$ , specific divergence of each locus to each breed; Rsb, across pairs of populations; BayeScan, software used to identify candidate loci under natural selection from genetic data, considering differences in allele frequencies between populations; hapFLK, test for the detection of selection signatures based on multiple population genotyping data focused on the different haplotype frequencies between populations;  $H_p$ , average pooled heterozygosity; and REHH, relative extended haplotype homozygosity.

Moradi et al. (2012), evaluating two Iranian breeds (Zel and Lori-Bakhtiari) with contrasting phenotypes for fat accumulation in the tail, identified regions located in chromosomes 5, 7, and X related to this trait. Lv et al. (2014) assessed selection by climatic conditions in sheep by combining molecular and environmental data. These authors chose 32 autochthone breeds from a data bank of 74 sheep breeds used in the Sheep HapMap Project (International Sheep Genomics Consortium, 2017) and identified genes related to climatic adaptation that are involved in energetic metabolism, hormonal, and self-immune regulation. Fariello et al. (2014), also using Sheep HapMap data, detected new selective sweeps associated with pigmentation, morphology, and productive traits. Specifically, two ancestral selective sweeps were next to genes (TROM8 and TSHR) whose functions (cold and photoperiod perception, respectively) seem to be relevant for selection response during the recent history of sheep domestication.

Gutiérrez-Gil et al. (2014) identified selection signatures for milk traits in sheep. Due to the difficulty in distinguishing between effects of true selection signature and demographic events (as expansion or bottlenecks), these authors used three different methods to detect asymptotic patterns of heterozygosity: pairwise  $F_{ST}$ , observed heterozygosity, and regression analysis; and, only the regions identified by pairwise  $F_{ST}$  and at minimal another method were considered as selection signature. These authors found six regions under positive selection in milk sheep, whereas Moioli et al. (2013), in a similar study, identified two genes, Palmdelphin (PALMD) and Ring finger protein 145 (RFP145), as related to milk production in sheep.

McRae et al. (2014) found 16 genome regions related to resistance to gastrointestinal nematodes, including genes involved in chitinase activity and cytokine response. Several selection signals identified by these authors were cited for the first time, and only two regions, chromosome 7 (CSAP35E-MCM149; OAR7: 44.018.971- 81.694.614) and 25 (0,4-40,7 Mb; 6,6-44 Mb; OARv2.0) were correlated with quantitative trait loci (QTLs) previously reported for this trait. This result reinforces the theory that parasite resistance, as well as most quantitative traits, are under the control of many loci with small effects (McRae et al., 2014).

Despite the high number of recent works aiming to identify selection signatures in sheep, great part of the

phenotypic and genotypic diversity of the species has not yet been assessed. There are some breeds, genetic groups, and ecotypes of sheep adapted to specific regions, i.e., environmental conditions, and these can be an important source of molecular information for the study of sheep physiology and marker identification, which would be useful in animal breeding programs.

### Challenges and opportunities

Studies of selective sweeps can be highly complex depending on the genomic heterogeneity determined by mutations and the genetic architecture of the traits under selection, as well as on the evolutionary history of the evaluated populations due to the influence of phenomena such as drift, selection, recombination, and migration. Haas & Payseur (2016), studying sweeps for natural selection, i.e., genome-wide scans for natural selection, recommended that information about crucial markers of genomic diversity be used to calibrate the patterns for an entire genome. For example, the estimation of recombination rates and deleterious mutations to each locus could be used to adjust the model of purifying selection to the polymorphism level throughout the genome. In many species, the recombination rates are highly correlated with the distance to the centromere; therefore, even if the recombination rate is not available for the species, it is possible to use the distance to the centromere as an adjustment factor for the relative recombination rate. Another recommendation of these authors is to measure the consequences of mutation heterogeneity, recombination, selection, and genetic architecture for the genomic patterns of diversity using simulations for the entire chromosome. The authors argue that knowledge of potential selective agents facilitates the interpretation of selection signatures and that new methods focused on the correlation between environmental variables and genetic variation could improve the results of these studies.

Fariello et al. (2014) highlighted the need for sequencing data in large scale and high resolution, in order to allow the precise identification of causative mutations. These authors also pointed out the importance of recording phenotypic data to identify biological processes. New genomic approaches, including high density SNP chips, entire genome sequencing, and transcriptome studies, are an

opportunity to find selection signals in the genome (Gutiérrez-Gil et al., 2014; McRae et al., 2014).

Moioli et al. (2013, 2016) demonstrated a new strategy based on genotyping of divergent animals that is able to detect genes and mutations directly related to the target trait. These authors concluded that this strategy is particularly useful in sheep, because gene detection and characterization is more imperative than genomic selection for this species due the low value of each individual and worldwide distribution of the flocks, mainly in low-input systems.

The identification of genomic regions that are under the influence of both natural and artificial selection can help identify genetic and biological bases of economically important traits that are segregating within or between breeds. The elevated phenotypic diversity levels observed in sheep breeds worldwide provide an interesting opportunity for the identification of selection signatures and characterization of the functions of specific haplotypes and genes, especially for traits related to environmental adaptation and resistance or tolerance to diseases and parasites, which are important traits for sheep production.

There are few examples of genomic selection in sheep, particularly in Australia and New Zealand (Brito et al., 2017). However, these do not represent the reality of the species in other parts of the world. Sheep farming is closely related to low-input systems with small flocks worldwide, and, consequently, few private companies or farmers have enough resources to conduct animal improvement programs with genomic-assisted methods. Therefore, the strategies to apply genomic information into sheep production need to be different from those that have been extensively applied, especially in dairy cattle breeding.

Studies with sheep data should make use of these new analysis techniques, which demand lower numbers of animals and may allow to evaluate specific phenotypes and also keep overall study costs at a minimum. Studies to identify selection signatures provide an opportunity to develop genomic knowledge in sheep, because the resources required fit with the specific aforementioned limitations. The knowledge of genomic regions associated with specific traits can be applied in small herds, aiming to: increase yield, through traits related to the number of offspring; decrease costs, via parasite resistance; or aggregate value to products, including changes in the nutritional

composition of products, increasing economic feasibility. The main challenges for sheep research are the development and application of cost effective strategies and techniques for genotyping and selecting animals from commercial herds using the knowledge acquired from studies of selective sweeps.

## Acknowledgments

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), for scholarship granted.

## References

- ANDERSSON, L. How selective sweeps in domestic animals provide new insight into biological mechanisms. **Journal of Internal Medicine**, v.271, p.1-14, 2012. DOI: 10.1111/j.1365-2796.2011.02450.x.
- BENJELLOUN, B.; ALBERTO, F.J.; STREETER, I.; BOYER, F.; COISSAC, E.; STUCKI, S.; BENBATI, M.; IBNELBACHYR, M.; CHENTOUF, M.; BECHCHARI, A.; LEEMPOEL, K.; ALBERTI, A.; ENGELEN, S.; CHIKHI, A.; CLARKE, L.; FLICEK, P.; JOOST, S.; TABERLET, P.; POMPANON, F. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. **Frontiers in Genetics**, v.6, art.107, 2015. DOI: 10.3389/fgene.2015.00107.
- BOITARD, S.; KOFLER, R.; FRANÇOISE, P.; ROBELIN, D.; SCHLÖTTERER, C.; FUTSCHIK, A. Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. **Molecular Ecology Resources**, v.13, p.337-340, 2013. DOI: 10.1111/1755-0998.12063.
- BOITARD, S.; SCHLÖTTERER, C.; FUTSCHIK, A. Detecting selective sweeps: a new approach based on hidden Markov models. **Genetics**, v.181, p.1567-1578, 2009. DOI: 10.1534/genetics.108.100032.
- BOITARD, S.; SCHLÖTTERER, C.; NOLTE, V.; PANDEY, R.V.; FUTSCHIK, A. Detecting selective sweeps from pooled next-generation sequencing samples. **Molecular Biology and Evolution**, v.29, p.2177-2186, 2012. DOI: 10.1093/molbev/mss090.
- BOMAN, I.A.; KLEMETSDAL, G.; NAFSTAD, O.; BLICHFELDT, T.; VAGE, D.I. Selection based on progeny testing induces rapid changes in *myostatin* allele frequencies – a case study in sheep. **Journal of Animal Breeding and Genetics**, v.128, p.52-55, 2011. DOI: 10.1111/j.1439-0388.2010.00879.x.
- BRITO, L.F.; CLARKE, S.M.; MCEWAN, J.C.; MILLER, S.P.; PICKERING, N.K.; BAIN, W.E.; DODDS, K.G.; SARGOLZAEI, M.; SCHENKEL, F.S. Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. **BMC Genetics**, v.18, art.7, 2017. DOI: 10.1186/s12863-017-0476-8.
- CADZOW, M.; BOOCOOCK, J.; NGUYEN, H.T.; WILCOX, P.; MERRIMAN, T.R.; BLACK, M.A. A bioinformatics workflow

- for detecting signatures of selection in genomic data. **Frontiers in Genetics**, v.5, art.293, 2014. DOI: 10.3389/fgene.2014.00293.
- CHEN, H.; PATTERSON, N.; REICH, D. Population differentiation as a test for selective sweeps. **Genome Research**, v.20, p.393-402, 2010. DOI: 10.1101/gr.100545.109.
- DODDS, K.G.; MCEWAN, J.C. Calculating exact probabilities of allele frequency differences in divergent selection lines. **Proceedings of the Association for the Advancement of Animal Breeding and Genetics**, v.12, p.556-560, 1997.
- DRUET, T.; PÉREZ-PARDAL, L.; CHARLIER, C.; GAUTIER, M. Identification of large selective sweeps associated with major genes in cattle. **Animal Genetics**, v.44, p.758-762, 2013. DOI: 10.1111/age.12073.
- FAGNY, M.; PATIN, E.; ENARD, D.; BARREIRO, L.B.; QUINTANA-MURCI, L.; LAVAL, G. Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets. **Molecular Biology and Evolution**, v.31, p.1850-1868, 2014. DOI: 10.1093/molbev/msu118.
- FARIELLO, M.I.; BOITARD, S.; NAYA, H.; SANCRISTOBAL, M.; SERVIN, B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. **Genetics**, v.193, p.929-941, 2013. DOI: 10.1534/genetics.112.147231.
- FARIELLO, M.-I.; SERVIN, B.; TOSSER-KLOPP, G.; RUPP, R.; MORENO, C.; CRISTOBAL, M.S.; BOITARD, S. Selection signatures in worldwide sheep populations. **PLoS ONE**, v.9, e103813, 2014. DOI: 10.1371/journal.pone.0103813.
- FAY, J.C.; WU, C.-I. Hitchhiking under positive darwinian selection. **Genetics**, v.155, p.1405-1413, 2000.
- FERRER-ADMETLLA, A.; LIANG, M.; KORNELIUSSEN, T.; NIELSEN, R. On detecting incomplete soft or hard selective sweeps using haplotype structure. **Molecular Biology and Evolution**, v.31, p.1275-1291, 2014. DOI: 10.1093/molbev/msu077.
- GAUTIER, M.; VITALIS, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. **Bioinformatics**, v.28, p.1176-1177, 2012. DOI: 10.1093/bioinformatics/bts115.
- GORKHALI, N.A.; DONG, K.; YANG, M.; SONG, S.; KADER, A.; SHRESTHA, B.S.; HE, X.; ZHAO, Q.; PU, Y.; LI, X.; KIJAS, J.; GUAN, W.; HAN, J.; JIANG, L.; MA, Y. Genomic analysis identified a potential novel molecular mechanism for high-altitude adaptation in sheep at the Himalayas. **Scientific Reports**, v.6, art.29963, 2016. DOI: 10.1038/srep29963.
- GOUVEIA, J.J. de S.; PAIVA, S.R.; MCMANUS, C.M.; CAETANO, A.R.; KIJAS, J.W.; FACÓ, O.; AZEVEDO, H.C.; ARAUJO, A.M. de; SOUZA, C.J.H. de; YAMAGISHI, M.E.B.; CARNEIRO, P.L.S.; LÔBO, R.N.B.; OLIVEIRA, S.M.P. de; SILVA, M.V.G.B. da. Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. **Livestock Science**, v.197, p.36-45, 2017. DOI: 10.1016/j.livsci.2017.01.006.
- GOUVEIA, J.J. de S.; SILVA, M.V.G.B. da; PAIVA, S.R.; OLIVEIRA, S.M.P. de Identification of selection signatures in livestock species. **Genetics and Molecular Biology**, v.37, p.330-342, 2014. DOI: 10.1590/S1415-47572014000300004.
- GRASSO, A.N.; GOLDBERG, V.; NAVAJAS, E.A.; IRIARTE, W.; GIMENO, D.; AGUILAR, I.; MEDRANO, J.F.; RINCÓN, G.; CIAPPESONI, G. Genomic variation and population structure detected by single nucleotide polymorphism arrays in Corriedale, Merino and Creole sheep. **Genetics and Molecular Biology**, v.37, p.389-395, 2014. DOI: 10.1590/S1415-47572014000300011.
- GROSSMAN, S.R.; SHYLAKHTER, I.; KARLSSON, E.K.; BYRNE, E.H.; MORALES, S.; FRIEDEN, G.; HOSTETTER, E.; ANGELINO, E.; GARBER, M.; ZUK, O.; LANDER, E.S.; SCHAFFNER, S.F.; SABETI, P.C. A composite of multiple signals distinguishes causal variants in regions of positive selection. **Science**, v.327, p.883-886, 2010. DOI: 10.1126/science.1183863.
- GUTIÉRREZ-GIL, B.; JOSE ARRANZ, J.; PONG-WONG, R.; GARCIA-GÁMEZ, E.; KIJAS, J.; WIENER, P. Application of selection mapping to identify genomic regions associated with dairy production in sheep. **PLoS ONE**, v.9, e94623, 2014. DOI: 10.1371/journal.pone.0094623.
- HAASL, R.J.; PAYSEUR, B.A. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. **Molecular Ecology**, v.25, p.5-23, 2016. DOI: 10.1111/mec.13339.
- HOHENLOHE, P.A.; PHILLIPS, P.C.; CRESKO, W.A. Using population genomics to detect selection in natural populations: key concepts and methodological considerations. **International Journal of Plant Sciences**, v.171, p.1059-1071, 2010. DOI: 10.1086/656306.
- HUBER, C.D.; DEGIORGIO, M.; HELLMANN, I.; NIELSEN, R. Detecting recent selective sweeps while controlling for mutation rate and background selection. **Molecular Ecology**, v.25, p.142-155, 2016. DOI: 10.1111/mec.13351.
- INTERNATIONAL SHEEP GENOMICS CONSORTIUM. **Sheep HapMap and Animal Resources**. Available at: <<http://www.sheephapmap.org/hapmap.php>>. Accessed on: Aug. 28 2017.
- KELLEY, J.L.; SWANSON, W.J. Positive selection in the human genome: from genome scans to biological significance. **Annual Review of Genomics and Human Genetics**, v.9, p.143-160, 2008. DOI: 10.1146/annurev.genom.9.081307.164411.
- KIM, Y.; NIELSEN, R. Linkage disequilibrium as a signature of selective sweeps. **Genetics**, v.167, p.1513-1524, 2004. DOI: 10.1534/genetics.103.025387.
- LIU, Z.; JI, Z.; WANG, G.; CHAO, T.; HOU, L.; WANG, J. Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. **BMC Genomics**, v.17, art.863, 2016. DOI: 10.1186/s12864-016-3212-2.
- LV, F.-H.; AGHA, S.; KANTANEN, J.; COLLI, L.; STUCKI, S.; KIJAS, J.W.; JOOST, S.; LI, M.-H.; MARSAN, P.A. Adaptations to climate-mediated selective pressures in sheep. **Molecular Biology and Evolution**, v.31, p.3324-3343, 2014. DOI: 10.1093/molbev/msu264.
- MANUNZA, A.; CARDOSO, T.F.; NOCE, A.; MARTÍNEZ, A.; PONS, A.; BERMEJO, L.A.; LANDI, V.; SÁNCHEZ, A.;

- JORDANA, J.; DELGADO, J.V.; ADÁN, S.; CAPOTE, J.; VIDAL, O.; UGARTE, E.; ARRANZ, J.J.; CALVO, J.H.; CASELLAS, J. AMILLS, M. Population structure of eleven Spanish ovine breeds and detection of selective sweeps with BayeScan and hapFLK. **Scientific Reports**, v.6, art.27296, 2016. DOI: 10.1038/srep27296.
- MCMANUS, C.; PAIVA, S.R.; ARAÚJO, R.O. de. Genetics and breeding of sheep in Brazil. **Revista Brasileira de Zootecnia**, v.39, p.236-246, 2010. Suplemento especial. DOI: 10.1590/S1516-35982010001300026.
- MCRAE, K.M.; MCEWAN, J.C.; DODDS, K.G.; GEMMELL, N.J. Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. **BMC Genomics**, v.15, art.637, 2014. DOI: 10.1186/1471-2164-15-637.
- MOIOLI, B.; D'ANDREA, S.; DE GROSSI, L.; SEZZI, E.; DE SANCTIS, B.; CATILLO, G.; STERI, R.; VALENTINI, A.; PILLA, F. Genomic scan for identifying candidate genes for paratuberculosis resistance in sheep. **Animal Production Science**, v.56, p.1046-1055, 2016. DOI: 10.1071/AN14826.
- MOIOLI, B.; SCATÀ, M.C.; STERI, R.; NAPOLITANO, F.; CATILO, G. Signatures of selection identify loci associated with milk yield in sheep. **BMC Genetics**, v.14, art.76, 2013. DOI: 10.1186/1471-2156-14-76.
- MORADI, M.H.; NEJATI-JAVAREMI, A.; MORADI-SHAHRBABA, M.; DODDS, K.G.; MCEWAN, J.C. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. **BMC Genetics**, v.13, art.10, 2012. DOI: 10.1186/1471-2156-13-10.
- NICOLAZZI, E.L.; BIFFANI, S.; BISCARINI, F.; OROZCO-TERWENGEL, P.; CAPRERA, A.; NAZZICARI, N.; STELLA, A. Software solutions for the livestock genomics SNP array revolution. **Animal Genetics**, v.46, p.343-353, 2015. DOI: 10.1111/age.12295.
- OLEKSYK, T.K.; SMITH, M.W.; O'BRIEN, S.J. Genome-wide scans for footprints of natural selection. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v.365, p.185-205, 2010. DOI: 10.1098/rstb.2009.0219.
- PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P.I.W.; DALY, M.J.; SHAM, P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v.81, p.559-575, 2007.
- PURFIELD, D.C.; MCPARLAND, S.; WALL, E.; BERRY, D.P. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. **PLoS ONE**, v.12, e0176780, 2017. DOI: 10.1371/journal.pone.0176780.
- QANBARI, S.; PAUSCH, H.; JANSEN, S.; SOMEL, M.; STROM, T.M.; FRIES, R.; NIELSEN, R.; SIMIANER, H. Classic selective sweeps revealed by massive sequencing in cattle. **PLoS Genetics**, v.10, e1004148, 2014. DOI: 10.1371/journal.pgen.1004148.
- RANDHAWA, I.A.S.; KHATKAR, M.S.; THOMSON, P.C.; RAADSMA, H.W. Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. **BMC Genetics**, v.15, art.34, 2014. DOI: 10.1186/1471-2156-15-34.
- RELLSTAB, C.; GUGERLI, F.; ECKERT, A.J.; HANCOCK, A.M.; HOLDEREGGER, R. A practical guide to environmental association analysis in landscape genomics. **Molecular Ecology**, v.24, p.4348-4370, 2015. DOI: 10.1111/mec.13322.
- SABETI, P.C.; REICH, D.E.; HIGGINS, J.M.; LEVINE, H.Z.P.; RICHTER, D.J.; SCHAFFNER, S.F.; GABRIEL, S.B.; PLATKO, J.V.; PATTERSON, N.J.; MCDONALD, G.J.; ACKERMAN, H.C.; CAMPBELL, S.J.; ALTSHULER, D.; COOPER, R.; KWIATKOWSKI, D.; WARD, R.; LANDER, E.S. Detecting recent positive selection in the human genome from haplotype structure. **Nature**, v.419, p.832-837, 2002. DOI: 10.1038/nature01140.
- SABETI, P.C.; SCHAFFNER, S.F.; FRY, B.; LOHMUELLER, J.; VARILLY, P.; SHAMOVSKY, O.; PALMA, A.; MIKKELSEN, T.S.; ALTSHULER, D.; LANDER, E.S. Positive natural selection in the human lineage. **Science**, v.312, p.1614-1620, 2006. DOI: 10.1126/science.1124309.
- SCHLÖTTERER, C. Hitchhiking mapping: limitations and potential for the identification of ecologically important genes. In: EISENHABER, F. (Ed.). **Discovering biomolecular mechanisms with computational biology**. Boston: Springer, 2006. p.117-125. DOI: 10.1007/0-387-36747-0\_10.
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, v.123, p.585-595, 1989.
- TANG, K.; THORNTON, K.R.; STONEKING, M. A new approach for using genome scans to detect recent positive selection in the human genome. **PLoS Biology**, v.5, p.1587-1602, 2007. e171. DOI: 10.1371/journal.pbio.0050171.
- UTSUNOMIYA, Y.T.; PEREZ O'BRIEN, A.M.; SONSTEGARD, T.S.; SÖLKNER, J.; GARCIA, J.F. Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. **Frontiers in Genetics**, v.6, art.36, 2015. DOI: 10.3389/fgene.2015.00036.
- VENTER, J.C.; ADAMS, M.D.; MYERS, E.W.; LI, P.W.; MURAL, R.J.; SUTTON, G.G.; SMITH, H.O.; YANDELL, M.; EVANS, C.A.; HOLT, R.A.; GOCAYNE, J.D.; AMANATIDES, P.; BALLEW, R.M.; HUSON, D.H.; WORTMAN, J.R.; ZHANG, Q.; KODIRA, C.D.; ZHENG, X.H.; CHEN, L.; SKUPSKI, M.; SUBRAMANIAN, G.; THOMAS, P.D.; ZHANG, J.; GABOR MIKLOS, G.L.; NELSON, C.; BRODER, S.; CLARK, A.G.; NADEAU, J.; MCKUSICK, V.A.; ZINDER, N.; LEVINE, A.J.; ROBERTS, R.J.; SIMON, M.; SLAYMAN, C.; HUNKAPILLER, M.; BOLANOS, R.; DELCHER, A.; DEW, I.; FASULO, D.; FLANIGAN, M.; FLOREA, L.; HALPERN, A.; HANNENHALLI, S.; KRAVITZ, S.; LEVY, S.; MOBARRY, C.; REINERT, K.; REMINGTON, K.; ABU-THREIDEH, J.; BEASLEY, E.; BIDDICK, K.; BONAZZI, V.; BRANDON, R.; CARGILL, M.; CHANDRAMOULISWARAN, I.; CHARLAB, R.; CHATURVEDI, K.; DENG, Z.; DI FRANCESCO, V.; DUNN, P.; EILBECK, K.; EVANGELISTA, C.; GABRIELIAN, A.E.; GAN, W.; GE, W.; GONG, F.; GU, Z.; GUAN, P.; HEIMAN, T.J.; HIGGINS, M.E.; JI, R.R.; KE, Z.; KETCHUM, K.A.; LAI, Z.; LEI, Y.; LI, Z.; LI, J.; LIANG, Y.; LIN, X.; LU, F.; MERKULOV, G.V.; MILSHINA, N.; MOORE, H.M.; NAIK, A.K.; NARAYAN, V.A.; NEELAM, B.; NUSSKERN, D.; RUSCH,

- D.B.; SALZBERG, S.; SHAO, W.; SHUE, B.; SUN, J.; WANG, Z.; WANG, A.; WANG, X.; WANG, J.; WEI, M.; WIDES, R.; XIAO, C.; YAN, C.; YAO, A.; YE, J.; ZHAN, M.; ZHANG, W.; ZHANG, H.; ZHAO, Q.; ZHENG, L.; ZHONG, F.; ZHONG, W.; ZHU, S.; ZHAO, S.; GILBERT, D.; BAUMHUETER, S.; SPIER, G.; CARTER, C.; CRAVCHIK, A.; WOODAGE, T.; ALI, F.; AN, H.; AWE, A.; BALDWIN, D.; BADEN, H.; BARNSTEAD, M.; BARROW, I.; BEESON, K.; BUSAM, D.; CARVER, A.; CENTER, A.; CHENG, M.L.; CURRY, L.; DANAHAR, S.; DAVENPORT, L.; DESILETS, R.; DIETZ, S.; DODSON, K.; DOUP, L.; FERRIERA, S.; GARG, N.; GLUECKSMANN, A.; HART, B.; HAYNES, J.; HAYNES, C.; HEINER, C.; HLADUN, S.; HOSTIN, D.; HOUCK, J.; HOWLAND, T.; IBEGWAM, C.; JOHNSON, J.; KALUSH, F.; KLINE, L.; KODURU, S.; LOVE, A.; MANN, F.; MAY, D.; MCCAWLEY, S.; MCINTOSH, T.; MCMULLEN, I.; MOY, M.; MOY, L.; MURPHY, B.; NELSON, K.; PFANNKOCH, C.; PRATTS, E.; PURI, V.; QURESHI, H.; REARDON, M.; RODRIGUEZ, R.; ROGERS, Y.H.; ROMBLAD, D.; RUHFEL, B.; SCOTT, R.; SITTER, C.; SMALLWOOD, M.; STEWART, E.; STRONG, R.; SUH, E.; THOMAS, R.; TINT, N.N.; TSE, S.; VECH, C.; WANG, G.; WETTER, J.; WILLIAMS, S.; WILLIAMS, M.; WINDSOR, S.; WINN-DEEN, E.; WOLFE, K.; ZAVERI, J.; ZAVERI, K.; ABRIL, J.F.; GUIGÓ, R.; CAMPBELL, M.J.; SJOLANDER, K.V.; KARLAK, B.; KEJARIWAL, A.; MI, H.; LAZAREVA, B.; HATTON, T.; NARECHANIA, A.; DIEMER, K.; MURUGANUJAN, A.; GUO, N.; SATO, S.; BAFNA, V.; ISTRAIL, S.; LIPPERT, R.; SCHWARTZ, R.; WALENZ, B.; YOOSEPH, S.; ALLEN, D.; BASU, A.; BAXENDALE, J.; BLICK, L.; CAMINHA, M.; CARNES-STINE, J.; CAULK, P.; CHIANG, Y.H.; COYNE, M.; DAHLKE, C.; MAYS, A.; DOMBROSKI, M.; DONNELLY, M.; ELY, D.; ESPARHAM, S.; FOSLER, C.; GIRE, H.; GLANOWSKI, S.; GLASSER, K.; GLODEK, A.; GOROKHOV, M.; GRAHAM, K.; GROPMAN, B.; HARRIS, M.; HEIL, J.; HENDERSON, S.; HOOVER, J.; JENNINGS, D.; JORDAN, C.; JORDAN, J.; KASHA, J.; KAGAN, L.; KRAFT, C.; LEVITSKY, A.; LEWIS, M.; LIU, X.; LOPEZ, J.; MA, D.; MAJOROS, W.; MCDANIEL, J.; MURPHY, S.; NEWMAN, M.; NGUYEN, T.; NGUYEN, N.; NODELL, M.; PAN, S.; PECK, J.; PETERSON, M.; ROWE, W.; SANDERS, R.; SCOTT, J.; SIMPSON, M.; SMITH, T.; SPRAGUE, A.; STOCKWELL, T.; TURNER, R.; VENTER, E.; WANG, M.; WEN, M.; WU, D.; WU, M.; XIA, A.; ZANDIEH, A.; ZHU, X. The sequence of the human genome. *Science*, v.291, p.1304-1351, 2001. DOI: 10.1126/science.1058040.
- VOIGHT, B.F.; KUDARAVALLI, S.; WEN, X.; PRITCHARD, J.K. A map of recent positive selection in the human genome. *PLoS Biology*, v.4, p.446-458, 2006. e72. DOI: 10.1371/journal.pbio.0040072.
- WANG, H.; ZHANG, L.; CAO, J.; WU, M.; MA, X.; LIU, Z.; LIU, R.; ZHAO, F.; WEI, C.; DU, L. Genome-wide specific selection in three domestic sheep breeds. *PLoS ONE*, v.10, p. e0128688, 2015. DOI: 10.1371/journal.pone.0128688.
- WEI, C.; WANG, H.; LIU, G.; WU, M.; CAO, J.; LIU, Z.; LIU, R.; ZHAO, F.; ZHANG, L.; LU, J.; LIU, C.; DU, L. Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics*, v.16, art.194, 2015. DOI: 10.1186/s12864-015-1384-9.
- WEI, C.; WANG, H.; LIU, G.; ZHAO, F.; KIJAS, J.W.; MA, Y.; LU, J.; ZHANG, L.; CAO, J.; WU, M.; WANG, G.; LIU, R.; LIU, Z.; ZHANG, S.; LIU, C.; DU, L. Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. *Scientific Reports*, v.6, art.26770, 2016. DOI: 10.1038/srep26770.
- XU, L.; BICKHART, D.M.; COLE, J.B.; SCHROEDER, S.G.; SONG, J.; VAN TASSELL, C.P.; SONSTEGARD, T.S.; LIU, G.E. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Molecular Biology and Evolution*, v.32, p.711-725, 2015. DOI: 10.1093/molbev/msu333.
- YANG, J.; BENYAMIN, B.; MCEVOY, B.P.; GORDON, S.; HENDERS, A.K.; NYHOLT, D.R.; MADDEN, P.A.; HEATH, A.C.; MARTIN, N.G.; MONTGOMERY, G.W.; GODDARD, M.E.; VISSCHER, P.M. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, v.42, p.565-569, 2010. DOI: 10.1038/ng.608.
- YANG, J.; LI, W.-R.; LV, F.-H.; HE, S.-G.; TIAN, S.-L.; PENG, W.-F.; SUN, Y.-W.; ZHAO, Y.-X.; TU, X.-L.; ZHANG, M.; XIE, X.-L.; WANG, Y.-T.; LI, J.-Q.; LIU, Y.-G.; SHEN, Z.-Q.; WANG, F.; LIU, G.-J.; LU, H.-F.; KANTANEN, J.; HAN, J.-L.; LI, M.-H.; LIU, M.-J. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Molecular Biology and Evolution*, v.33, p.2576-2592, 2016. DOI: 10.1093/molbev/msw129.
- YUAN, Z.; LIU, E.; LIU, Z.; KIJAS, J.W.; ZHU, C.; HU, S.; MA, X.; ZHANG, L.; DU, L.; WANG, H.; WEI, C. Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Animal Genetics*, v.48, p.55-66, 2017. DOI: 10.1111/age.12477.
- ZHAO, F.; MCPARLAND, S.; KEARNEY, F.; DU, L.; BERRY, D.P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics Selection Evolution*, v.47, p.1-12, 2015. DOI: 10.1186/s12711-015-0127-3.
- ZHAO, F.; WEI, C.; ZHANG, L.; LIU, J.; WANG, G.; ZENG, T.; DU, L. A genome scan of recent positive selection signatures in three sheep populations. *Journal of Integrative Agriculture*, v.15, p.162-174, 2016. DOI: 10.1016/S2095-3119(15)61080-2.
- ZHU, C.; FAN, H.; YUAN, Z.; HU, S.; ZHANG, L.; WEI, C.; ZHANG, Q.; ZHAO, F.; DU, L. Detection of selection signatures on the X chromosome in three sheep breeds. *International Journal of Molecular Sciences*, v.16, p.20360-20374, 2015. DOI: 10.3390/ijms160920360.

Received on July 20, 2016 and accepted on August 28, 2017