Review Article

Identification of selection signatures in livestock species

João José de Simoni Gouveia[1,2], Marcos Vinicius Gualberto Barbosa da Silva[3], Samuel Rezende Paiva[4]
and Sônia Maria Pinheiro de Oliveira[5]

[1]*Colegiado Acadêmico de Zootecnia, Universidade Federal do Vale do São Francisco, Petrolina, PE, Brazil.*
[2]*Programa de Doutorado Integrado em Zootecnia, Universidade Federal do Ceará, Fortaleza, CE, Brazil.*
[3]*Embrapa Gado de Leite, Juiz de Fora, MG, Brazil.*
[4]*Embrapa Sede, Secretaria de Relações Internacionais, Brasília, DF, Brazil.*
[5]*Departamento de Zootecnia, Universidade Federal do Ceará, Fortaleza, CE, Brazil.*

Abstract

The identification of regions that have undergone selection is one of the principal goals of theoretical and applied evolutionary genetics. Such studies can also provide information about the evolutionary processes involved in shaping genomes, as well as physical and functional information about genes/genomic regions. Domestication followed by breed formation and selection schemes has allowed the formation of very diverse livestock breeds adapted to a wide variety of environments and with special characteristics. The advances in genomics in the last five years have enabled the development of several methods to detect selection signatures and have resulted in the publication of a considerable number of studies involving livestock species. The aims of this review are to describe the principal effects of natural/artificial selection on livestock genomes, to present the main methods used to detect selection signatures and to discuss some recent results in this area. This review should be useful also to research scientists working with wild animals/non-domesticated species and plant biologists working with breeding and evolutionary biology.

Keywords: artificial selection, domestic animals, selective sweep.

Received: July 23, 2013; Accepted: February 27, 2014.

Introduction

Selection tends to cause specific changes in the patterns of variation among selected loci and in neutral loci linked to them. These genomic footprints left by selection are known as selection signatures and can be used to identify loci subjected to selection (Kreitman, 2000). The recent availability of genomic information on domestic animal species and the development of improved statistical tools make the identification of these footprints in a given species possible (International Chicken Genome Sequencing Consortium, 2004; The Bovine Genome Sequencing and Analysis Consortium, 2009; The International Sheep Genomics Consortium, 2010; Groenen *et al.*, 2012; Dong *et al.*, 2013).

The identification of selection signatures is currently one of the principal interests of evolutionary geneticists because it can provide information ranging from basic knowledge about the evolutionary processes that are shaping genomes to functional information about genes/genomic regions (Nielsen, 2001, 2005; Schlötterer, 2003). For example, if a region that was not previously identified as con-tributing to any special trait in mapping experiments is targeted by selection in a specific population, then this information could lead to an initial inference about the functional characteristics of that region. This approach could also lead to the identification of genes related to ecological traits (*e.g.*, genes related to tropical adaptation) that are difficult to identify through laboratory experiments and may also be useful in corroborating quantitative trait loci (QTL) mapping experiments in production animals. The final and certainly most ambitious aim of these studies is to identify the causal mutations that confer a selective advantage in a specific population or species (Nielsen, 2001; Schlötterer, 2003; Hayes *et al.*, 2008).

Domestication greatly changed the morphological and behavioral characteristics of modern domestic animals and, along with breed formation and selection schemes for improving the production of specific products or achieving a morphological/behavioral standard, allowed the formation of very diverse modern breeds (Diamond, 2002; Toro and Mäki-Tanila, 2007; Flori *et al.*, 2009). These features, along with extensive knowledge about genomic regions that affect economically important traits and recent advances in the field of genomics, provide an excellent opportunity for identifying loci subjected to selection and for the

Send correspondence to Marcos Vinicius G.B. da Silva. Rua Eugênio do Nascimento, 610, Dom Bosco, 36038-330, Juiz de Fora, MG, Brazil. E-mail: marcos.vb.silva@embrapa.br.

validation of new methods developed to detect selection signatures (Hayes *et al.*, 2008; Flori *et al.*, 2009).

In this review, we describe the effects of natural/artificial selection on genomes, summarize the main methods of detecting the footprints of selection and, finally, indicate and discuss studies aimed at detecting selection signatures in livestock.

Natural and Artificial Selection

Natural selection is a phenomenon driven by the environment in which individuals with specific genotypes have a differential capacity for contributing to the next generation's gene pool (Falconer and Mackay, 1996; Templeton, 2006; Driscoll *et al.*, 2009). Natural selection could basically act in three ways: positive selection, purifying selection (also known as negative or background selection) and balancing selection. Each form of selection is a response to environmental pressure and acts differentially to alter the allelic and genotypic frequencies (Harris and Meyer, 2006; Oleksyk *et al.*, 2010).

Positive selection occurs when a newly arisen mutation has a selective advantage over other mutations and, therefore, increases in frequency in the population (Kaplan *et al.*, 1989). In purifying selection, the disadvantageous variants that appear in the population tend to be removed, thereby maintaining the functional integrity of DNA sequences (Charlesworth *et al.*, 1993). Balancing selection occurs when polymorphism is favored, leading to increased genetic variability. Several biological processes can be grouped in this type of selection, *e.g.*, overdominant selection (in which the heterozygote has a selective advantage), frequency dependent selection (in which different alleles are favored at different time intervals) and temporally or spatially heterogeneous selection (Charlesworth, 2006).

In contrast to natural selection, artificial selection (also called selective breeding) is a human-mediated process in which the gene pool of the next generation does not depend exclusively (or necessarily) on fitness components, but also on traits chosen by humans. Artificial selection can be classified as unconscious selection or methodical selection – the former occurs when there is no long-term objective, and this has been suggested as the cause of the early domestication process. The second occurs when a standard or objective drives the choice of parents for the next generation. Despite these differences and considering that the time frame in which these changes occur is often considerably different, the genetic consequences of natural and artificial selection are essentially the same (Avise and Ayala, 2009; Driscoll *et al.*, 2009; Gregory, 2009).

Selection Signatures

The occurrence of selection creates departures from the neutral theory expectations in the patterns of molecular variation. Each form of selection causes specific changes in the selected loci and in neutral loci linked to them (Kreitman, 2000). When positive selection operates in a newly arisen allele that has a selective advantage it tends to increase in frequency in the population and carries linked neutral alleles along with it. This phenomenon is known as the hitchhiking effect or selective sweep (Maynard-Smith and Haigh, 1974; Charlesworth, 2007). The selective sweep reduces the heterozygosity of regions surrounding the selected locus (Kaplan *et al.*, 1989; Kim and Stephan, 2002) and introduces a skew in the site frequency spectrum (SFS) because of an excess of rare variants in the selected region (Braverman *et al.*, 1995; Kim and Stephan, 2002).

An increase in the average linkage disequilibrium (LD) leading to long haplotypes is also expected in the region surrounding the selected site (Kim and Stephan, 2002). As LD decays and high frequency neutral alleles become fixed in the population after fixation of the selected mutation, this selection signature vanishes rapidly (Przeworski, 2002; Kim and Nielsen, 2004; McVean, 2007). Thus, a high frequency derived allele surrounded by a long-range LD is indicative of a recent selective sweep (Sabeti *et al.*, 2002; Voight *et al.*, 2006). In addition, the levels of within-population diversity tend to decrease while the between-population levels of diversity tend to increase in the region surrounding the selected locus (Beaumont, 2005; Storz, 2005). Furthermore, the number of nonsynonymous substitutions per nonsynonymous site (d_N) tends to be higher than the number of synonymous substitutions per synonymous site (d_S) (Nei, 2005; Harris and Meyer, 2006).

The model of selective sweep in which a newly arisen allele with a strong selective advantage increases quickly in frequency until reaching fixation is known as "hard sweep". In contrast, when the selected allele is part of existent genetic variation, it causes a "soft sweep" in which the footprint left by selection tends to be less pronounced and the frequency of the selected allele at the beginning of the selected phase is the crucial factor influencing the selective sweep (Przeworski *et al.*, 2005; Pritchard *et al.*, 2010).

Balancing selection favors the maintenance of polymorphism (Harris and Meyer, 2006; Oleksyk *et al.*, 2010). The persistence of the same alleles for a long time is known as long-term balancing selection and, in addition to maintaining polymorphism in the selected locus, it also tends to increase diversity in tightly linked neutral sites; if the region under selection has low recombination rates then it generally also has longer coalescence times than other regions (Charlesworth, 2006). In the presence of long-term balancing selection, the within-population diversity levels tend to increase and the between-population levels of diversity tend to decrease (Navarro and Barton 2002; Charlesworth *et al.*, 2003; Charlesworth, 2006), leading to reduced inbreeding coefficient (*FST*) values among populations compared to neutral expectations (Beaumont, 2005; Storz, 2005). However, in some cases, the *FST* levels may be

higher than expected by neutrality (Beaumont, 2005; Charlesworth, 2006).

When negative (background) selection occurs, the novel variants are disadvantageous and are consequently removed from the population, along with neutral variations linked to them (Innan and Stephan, 2003). If the recombination rate in the region is restricted or the population is highly inbred then background selection reduces the variability around the eliminated sites (Charlesworth *et al.*, 1993, 1995; Andolfatto, 2001; Stephan, 2010). An excess of low frequency alleles is also observed in small to moderately sized populations (Charlesworth *et al.*, 1993, 1995) and the number of nonsynonymous substitutions per non-synonymous site tends to be lower than the number of synonymous substitutions per synonymous site (Nei, 2005; Harris and Meyer, 2006). However, in regions with normal recombination rates, or when inbreeding is restricted, no reduction in variability is observed (Charlesworth *et al.*, 1993, 1995; Stephan, 2010). Furthermore, background selection does not cause a marked bias in the frequency spectrum (Charlesworth *et al.*, 1993, 1995; Kim and Stephan, 2000; Andolfatto, 2001; Stephan, 2010).

Selection signatures can be influenced by several factors. For example, the type of selection, the relative age of the neutral linked alleles, the strength of selection and the recombination rate (Braverman *et al.*, 1995; Kaplan *et al.*, 1989; Kim and Stephan, 2002; Charlesworth, 2007; McVean, 2007). Recognition of the molecular footprints left by different types of selection is a crucial task in identifying genomic regions subjected to selection. In this case, the neutral theory serves as the backbone for the statistical tests developed to detect selection signatures. However, in natural populations, some assumptions of the neutral theory can be violated (*e.g.*, population expansion, subdivision and bottlenecking) and this can lead to signals that mimic the footprints of selection. The interaction of different types of selection and interaction between selection and demographic factors can bias the footprints left in the genome (Barton, 1998; Kim and Stephan, 2000; Kreitman, 2000; Charlesworth *et al.*, 2003; Harris and Meyer, 2006; Toro and Mäki-Tanila, 2007). Because of this, it is worth noting that in studies designed to detect selection signatures in livestock a considerable high rate of false positives is expected as a result of genetic drift and founder effect, both of which were particularly important during the development of livestock breeds (Petersen *et al.*, 2013).

Methods for Detecting Selected Loci

The methods proposed for detecting selected loci can be classified in different ways (Harris and Meyer, 2006; Oleksyk *et al.*, 2010). Based on the main variables that affect the patterns of molecular variation left by selection, Hohenlohe *et al.* (2010) proposed a decision tree designed to identify the most appropriate method for each case. This decision tree is based primarily on the time scale in which selection can occur, but also considers other factors (*e.g.*, the number of populations in the study, mode of selection, etc.) and can be used by researchers in studies designed to detect selection signatures.

Tests based on synonymous and non-synonymous substitution rates

When the coding sequences of orthologous genes of interest are compared, it is expected that under neutral evolution, $d_N/d_S = 1$. When positive selection is in effect, $d_N/d_S > 1$, and under negative selection, $d_N/d_S < 1$. Differences in d_N/d_S are also expected among lineages when selection is in effect (Yang, 1998). Several methods have been proposed to estimate d_N and d_S (Nei, 2005). These methods were initially approximations based on the comparison of two sequences (Nei and Gojobori, 1986). More recently, maximum likelihood estimates from multiple alignments that account for transition/tranversion rate bias, codon usage bias, selective restraints at the protein level (Goldman and Yang, 1994), and variable d_N/d_S among sites and among lineages have been proposed (Nielsen and Yang, 1998; Yang *et al.*, 2000; Yang, 2002; Yang and Nielsen, 2002; O'Brien *et al.*, 2009). Hypothesis testing can be done using a likelihood ratio test that compares the model (assuming neutrality) with alternative models (Yang, 1998; Yang and Nielsen, 1998; Yang *et al.*, 2000, 2005). Packages such as MEGA (Tamura *et al.*, 2007) and PAML (Yang, 2007) implement the d_N/d_S selection tests.

Tests based on the frequency spectrum

The θ parameter can be estimated from DNA sequences in several ways, and comparison of the different estimates of θ is the basis for some tests aimed at identifying selected regions (Tajima, 1989; Fu and Li, 1993; Fu, 1996, 1997). Tajima (1989) proposed a test based on the difference between $\hat{\theta}_\pi$ (the average number of nucleotide differences) and $\hat{\theta}_S$ (the number of segregating sites along the DNA sequence) because the presence of selection tends to alter the value of $\hat{\theta}_\pi$ while that of $\hat{\theta}_S$ tends to remain unaffected (Tajima, 1989; Hartl and Clark, 2010). The proposed statistic (Tajima's *D*) corresponds to the standardized difference between $\hat{\theta}_\pi$ and $\hat{\theta}_S$ (Tajima, 1989; Harris and Meyer, 2006). Under neutrality, the value of *D* tends to be zero. Positive and negative selection tend to reduce heterozygosity and cause an excess of rare variants surrounding the selected locus, leading to $D < 0$ (Kaplan *et al.*, 1989; Tajima, 1989; Charlesworth *et al.*, 1993, 1995; Braverman *et al.*, 1995; Andolfatto, 2001; Kim and Stephan, 2002; Stephan, 2010). In contrast, long-term balancing selection increases the diversity around the selected locus, leading to $D > 0$ (Tajima, 1989; Navarro and Barton, 2002; Charlesworth, 2006).

Several other tests for detecting selection based on the excess of rare alleles have been developed (Fu and Li,

1993; Fu, 1996, 1997). However, the results of these tests do not always have a straightforward biological interpretation because in some situations it is impossible to differentiate between positive and negative selection (Tajima, 1989; Harris and Meyer, 2006), and also because these tests are sensitive to demography (Tajima, 1989; Charlesworth *et al.*, 1993; Fu and Li, 1993; Fu, 1996, 1997). While a reduction in heterozygosity and an excess of rare variants are not necessarily a specific pattern left by selection, an excess of derived variants (non-ancestral allele determined by an outgroup) has been identified as a unique feature produced by positive selection (Fay and Wu, 2000). To access this feature, Fay and Wu (2000) proposed a statistic called Fay and Wu's *H* that is calculated as the difference between $\hat{\theta}_\pi$ and $\hat{\theta}_H$ (where $\hat{\theta}_H$ is an estimator of θ weighted by the homozygosity of the derived alleles). When positive (but not negative) selection acts, the value of $\hat{\theta}_H$ tends to increase because of an excess of derived alleles, leading to $H < 0$. Thus, in contrast with Tajima's and Fu and Li's statistics, Fay and Wu's *H* allows the distinction between positive and negative selection (Fay and Wu, 2000).

The decrease in variability caused by positive selection tends to be broken by recombination events. Consequently, "valleys" of reduced heterozygosity have been suggested to be footprints of recent hitchhiking events. The depth and extent of the "valleys" is influenced by several factors, such as the strength of selection, recombination rates and effective population size. Because of this, Kim and Stephan (2002) proposed a composite likelihood approach for detecting positive selection in a recombining chromosome. The test is based on the expected number of sites where the derived allele is part of a given frequency interval in the population. More recently, extensions of these tests based on the frequency spectrum around a selective sweep have been proposed. These new methods can deal with genomic data and account for the ascertainment bias (Nielsen *et al.*, 2005; Kelley *et al.*, 2006; Williamson *et al.*, 2007).

Tests based on linkage disequilibrium

Exploitation of the LD patterns is the focus of several tests for detecting selection (Sabeti *et al.*, 2002, 2007; Kim and Nielsen 2004; Voight *et al.*, 2006; Kimura *et al.*, 2007). However, these signatures tend to be transient since the recombination tends to quickly break down this LD as soon as the selected locus reaches fixation (Przeworski, 2002; Kim and Nielsen, 2004; McVean, 2007). Sabeti *et al.* (2002) proposed an approach referred to as the long-range haplotype (*LRH*) test to detect recent selective sweeps by focusing on the relationship between the allele frequency and the LD level surrounding it.

This test starts with identification of the core haplotypes (through genotyping a set of single nucleotide polymorphisms (SNPs) in a region so small that recombination

may not occur). Subsequently, other SNPs at increasing distances from the core haplotypes are analyzed to evaluate the decay of LD according to distance (Sabeti *et al.*, 2002). The LD is measured at increasing distances from the core haplotypes through calculation of the extended haplotype homozygosity (*EHH*), which is the probability that two chromosomes carrying a specific core haplotype are homozygous for the whole region from the core to a distance x (Sabeti *et al.*, 2002). The relative *EHH* (*REHH*) is then calculated to compare the decay of *EHH* of one specific core haplotype to the decay of *EHH* of all the other core haplotypes combined. To test for selection, REHH and the frequency for each core haplotype is compared to REHH and the frequency of the other core haplotypes. Positive selection is inferred if one core haplotype has a combination of high REHH and high frequency in the population (Sabeti *et al.*, 2002).

An extension of the *LRH* test was proposed by Voight *et al.* (2006). This test is referred to as the *iHS* (integrated haplotype score) and was designed to work on a genomic scale using information from dense SNP chips. The *iHS* value can be defined simply as a measure of how unusual the haplotypes around an SNP are, compared to the genome (Voight *et al.*, 2006). In this approach, each SNP is treated as a core SNP and the test starts with calculation of the *EHH* for each core SNP. As SNPs are biallelic loci, each core SNP can be ancestral or derived. For the test, the integral of the observed decay of EHH from a core SNP until EHH reaches 0.05 is computed (the area under the curve in an *EHH vs.* distance plot). This value is referred to as the integrated *EHH* (*iHH*) and is identified as iHH_A or iHH_D, depending on whether it was computed from the ancestral or the derived allele of the core SNP. This value is then standardized to allow direct comparisons among different SNPs regardless of allele frequencies (Voight *et al.*, 2006).

Hussin *et al.* (2010) proposed a method based on the haplotype allelic classes (*HAC*). This measure can be defined as the count of allelic differences between the reference allelic class and the individual haplotypes in the sample. The statistic proposed is referred to as *Svd*, with positive values suggesting positive selection (Hussin *et al.*, 2010).

The *LRH* and *iHS* tests rely on the frequencies of alleles at core SNP and therefore have reduced power for detecting selection when the selected allele has reached fixation. To deal with situations in which the selected allele is fixed in one population but remains polymorphic in others, *LRH*-derived tests based on pairwise comparisons among populations have been proposed (Kimura *et al.*, 2007; Sabeti *et al.*, 2007; Tang *et al.*, 2007). The *XP-EHH* statistic can be defined as the normalized log-ratio between I_A and I_B, where I_A is the integral of the observed decay of *EHH* from a core SNP to an SNP X (which has an *EHH* value as close as possible to 0.04 in both populations) in population A, and I_B is the analogous measure in population

B (Sabeti *et al.*, 2007). The *ln(Rsb)* statistic proposed by Tang *et al.* (2007) is very similar to *XP-EHH*. The main difference between them is that the former calculates the *EHH* based on the status of each core SNP allele and the latter calculates the *EHH* based on the core SNP site (Sabeti *et al.*, 2007; Tang *et al.*, 2007).

Tests based on population differentiation

The estimation of *FST* from multiple loci and comparison of these values with its neutral expectations is the basis of several tests aimed at identifying selection (Lewontin and Krakauer, 1973; Bowcock *et al.*, 1991; Vitalis *et al.*, 2001, 2003; Beaumont and Balding, 2004; Foll and Gaggiotti, 2008; Excoffier *et al.*, 2009; Bonhomme *et al.*, 2010). The first effort in this direction was proposed by Lewontin and Krakauer (1973). They suggested that the *FST* estimated from several loci under neutrality must show small heterogeneity; however, if selection is acting on some of them then the estimates of *FST* tend to vary widely. The Lewontin and Krakauer test involves comparison between the variance of *FST* estimated from the data and the expected variance of *FST* under neutrality through a variance ratio test (Lewontin and Krakauer, 1973).

Lewontin and Krakauer's test was severely criticized soon after publication because of the assumptions they made in estimating the variance of *FST* under neutrality (Nei and Maruyama, 1975; Robertson, 1975). To avoid the effects of population structure, Bowcock *et al.* (1991) suggested the use of a null distribution obtained by calculating an *FST* distribution using simulations that take into account the populations phylogenetic history. More recently, models capable of generating the null distribution of *FST* that are robust to population history and structure (recent divergence and growth, isolation by distance and heterogeneous levels of gene flow between populations) have been proposed (Beaumont and Nichols, 1996; Beaumont and Balding, 2004; Foll and Gaggiotti, 2008; Excoffier *et al.*, 2009) and implemented in freely distributed softwares such as BayesFST (Beaumont and Balding, 2004), BayeScan (Foll and Gaggiotti, 2008) and Arlequin (Excoffier *et al.*, 2009). The methods proposed by Beaumont and Nichols (1996) and Excoffier *et al.* (2009) are computationally feasible, but the presence of some complex demographic histories can lead to important biases. On the other hand, Markov chain Monte Carlo (MCMC) based methods (Beaumont and Balding, 2004; Foll and Gaggiotti, 2008) efficiently accommodate some departures from model assumptions but are computationally very intensive.

Another way to avoid the effects of demography is to perform pairwise comparisons between populations (Tsakas and Krimbas, 1976). Based on this idea, Vitalis *et al.* (2001) proposed a simple model of population divergence from which they obtained the joint distribution of population-specific estimators of branch length which were used to construct the confidence interval. This approach seems to be robust against departures from model assumptions and also tends to remove the bias introduced by unknown population structure. However, the pairwise comparison tends to reduce the power of the test because information from other populations is discarded (Tsakas and Krimbas, 1976; Vitalis *et al.*, 2001). This analysis is implemented in the software DetSel 1.0 (Vitalis *et al.*, 2003).

The foregoing discussion has shown that there are currently several approaches for detecting footprints left by selection. Each of these approaches can capture specific patterns of molecular variation. The use of a combination of alternative approaches for detecting selection signals is an interesting strategy that has been suggested as a means of increasing the reliability of these studies. However, the success of one test and failure of another does not exclude the region of interest from having been subjected to selection since different tests can focus on different signals left by selection or look for different time scales in which the selection can act (Hohenlohe *et al.*, 2010; Oleksyk *et al.*, 2010).

Selection signatures in livestock

Domestication has resulted in considerable changes in the morphology and behavior of livestock species. In the early stages of domestication, unconscious selection for behavioral traits was applied. This early stage was followed by methodical selection in which specific traits were selected based on goals (Diamond, 2002; Gregory, 2009).

The development of specialized breeds, improved to produce specific products or to reach a morphological standard, increased the differences between domesticated animals and their wild relatives and also generated an enormous variety of different populations, with specific traits related to their specialization. Some of these traits are controlled by several interacting genes with minor effects. This creates an exceptional opportunity to gain knowledge of the molecular basis of these traits, particularly since most economically important traits in livestock are quantitative (Andersson and Georges, 2004).

The identification of genes targeted by selection in livestock can help to find and prove causal mutations in regions previously identified by QTL mapping experiments and can reveal genes related to ecological traits (*e.g.*, genes related to tropical adaptation) that are difficult to find experimentally. Furthermore, these studies can help to identify the genes or gene networks that contribute to the same trait but that were selected differentially between breeds; they can also unveil genes responsible for genetic correlations and the domestication process (Schlötterer, 2003; Hayes *et al.*, 2008; Ojeda *et al.*, 2008; Flori *et al.*, 2009; MacEachern *et al.*, 2009).

Signatures associated with domestication and early breed development

In some wild species, the expression both of eumelanin and phaeomelanin pigments is related to a camou-

flaged coat color. During domestication, non-camouflaged coat patterns were selected because of their direct effect on animal husbandry and also because these patterns may have been used as markers associated with improved individuals, or because of cultural preferences (Fang *et al.*, 2009; Wiener and Wilkinson, 2011).

The melanocyte stimulating hormone receptor gene (*MC1R*) influences the production of eumelanin and phaeomelanin pigments (Werth *et al.*, 1996; Kijas *et al.*, 1998; Fang *et al.*, 2009; Li *et al.*, 2010b) and is under selection in domestic cattle (Flori *et al.*, 2009; Stella *et al.*, 2010) and pig (Fang *et al.*, 2009; Li *et al.*, 2010b; Amaral *et al.*, 2011) breeds. Other genes that influence coat color pattern were also suggested to be under selection in domestic species. Selection signatures around the V-Kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*) have been reported for cattle (Stella *et al.*, 2010; Wiener *et al.*, 2011), pigs (Fontanesi *et al.*, 2010; Amaral *et al.*, 2011) and sheep (Kijas *et al.*, 2012). The melanocyte protein 17 precursor (*PMEL17*), also known as the Silver gene (*SILV*), is suggested to be under selection in some cattle breeds (Gautier *et al.*, 2009; Wiener *et al.*, 2011).

The presence/absence of horns is another important feature in breed definition in some livestock species. Recently, the relaxin-like receptor 2 (*RXFP2*) gene was associated with this trait (Johnston *et al.*, 2011), and a SNP surrounding this gene showed a strong selection signal in an analysis involving 74 sheep breeds. In cattle, the region surrounding the polled locus was shown to be under selection, although the gene responsible for this trait was not mapped (Drögemüller *et al.*, 2005; Li *et al.*, 2010a; Stella *et al.*, 2010). Behavioral changes, such as a reduction in fear and anti-predator responses and an increase in sociability, are believed to be important reflections of animal domestication (Diamond, 2002; Amaral *et al.*, 2011; Wiener and Wilkinson, 2011). Indeed, several studies in livestock suggest selection signatures surrounding genes related to nervous system development and function (The Bovine HapMap Consortium, 2009; Gautier *et al.*, 2009; Stella *et al.*, 2010; Amaral *et al.*, 2011).

Cattle

Modern bovine breeds can basically be grouped into two major types, the taurine and indicine groups. Within each group, several breeds have been developed, and there is considerable intra- and inter-group variability in productive (milk yield and quality, meat production), morphological (coat color, presence/absence of horns) and adaptive (disease resistance, heat tolerance) traits (The Bovine HapMap Consortium, 2009). Several genome-wide studies focusing on different approaches and using different sets of breeds have sought for selection signatures in bovines (Prasad *et al.*, 2008; Barendse *et al.*, 2009; Flori *et al.*, 2009; Gautier *et al.*, 2009; Hayes *et al.*, 2009; MacEachern *et al.*, 2009; The Bovine HapMap Consortium, 2009; Li *et al.*,

2010a; Qanbari *et al.*, 2010, 2011; Stella *et al.*, 2010; Wiener *et al.*, 2011; Hosokawa *et al.*, 2012).

Various studies in beef cattle using approaches such as differences in allele frequencies, *iHS* and *FST* have found selection signals in the centromeric region of BTA14 (Hayes *et al.*, 2009; The Bovine HapMap Consortium, 2009; Wiener *et al.*, 2011), a region involved in the control of marbling and fatness traits (Barendse, 1999; Moore *et al.*, 2003; Thaller *et al.*, 2003; Casas *et al.*, 2005; Pannier *et al.*, 2010; Veneroni *et al.*, 2010). An increase in intramuscular fat percentage in Australian Angus in recent years, together with a significant effect of this region on fat traits, may corroborate with the selection signature found in these studies (Hayes *et al.*, 2009).

The double muscled phenotype has been selected in some beef breeds and mutations in the Growth Differentiation Factor 8 (also known as myostatin or *GDF-8*) gene are related to this phenotype (Bellinge *et al.*, 2005). A decrease in heterozygosity around this gene has been demonstrated in double muscled breeds (Wiener *et al.*, 2003; Wiener and Gutiérrez-Gil, 2009) and an increase in LD (measured using the *iHS* approach) has been reported in this region (The Bovine HapMap Consortium, 2009).

Using the *FST* approach, a selection signature was found in the median region of BTA2 (Barendse *et al.*, 2009; The Bovine HapMap Consortium, 2009; Qanbari *et al.*, 2011). This region was associated with feed efficiency and intramuscular fat in beef breeds (Barendse *et al.*, 2007, 2009) and contains the R3H Domain Containing 1 (R3HDM1) and Zinc Finger, RAN Binding Domain Containing 3 (*ZRANB3*) genes, which have been suggested to be involved in feed efficiency (Barendse *et al.*, 2009; The Bovine HapMap Consortium, 2009).

Chromosome BTA6 harbors at least three QTLs that affect milk traits (Khatkar *et al.*, 2004; Ogorevc *et al.*, 2009; Weikard *et al.*, 2012) and these regions have been suggested to be under selection in dairy breeds (Hayes *et al.*, 2008; Barendse *et al.*, 2009; The Bovine HapMap Consortium, 2009; Qanbari *et al.*, 2010; Schwarzenbacher *et al.*, 2012). The first region contains the ATP-binding cassette, Sub-family G (WHITE), Member 2 (*ABCG2*) gene that was previously related to milk yield and quality traits (Cohen-Zinder *et al.*, 2005; Olsen *et al.*, 2007; Cole *et al.*, 2009; Weikard *et al.*, 2012). The second region contains the Peroxisome Proliferator-Activated Receptor Gamma, Coactivator 1 Alpha (*PPARGC1A*) gene that mediates the expression of genes involved in adipogenesis, gluconeogenesis and oxidative metabolism (Weikard *et al.*, 2005; Ogorevc *et al.*, 2009) and the third region contains the Casein Cluster associated with milk and protein yield (Boettcher *et al.*, 2004; Nilsen *et al.*, 2009; Sodeland *et al.*, 2011).

The increase in allele frequency differences between meat and dairy cattle and the high linkage disequilibrium in dairy breeds (using *EHH* and *iHS* methods) suggest that the

region surrounding DGAT1 is under selection (Hayes *et al.*, 2009; Qanbari *et al.*, 2010; Hosokawa *et al.*, 2012; Schwarzenbacher *et al.*, 2012). This gene is suggested to be responsible for a QTL with a major effect on milk fat percentage (Grisart *et al.*, 2002; Khatkar *et al.*, 2004; Cole *et al.*, 2009; Hayes *et al.*, 2010; Jiang *et al.*, 2010).

At least two QTLs affecting milk traits are located in the BTA20 chromosome. The first QTL was mapped surrounding the Growth Hormone Receptor Gene (*GHR*) and has a marked effect on protein percentage and a minor effect on fat percentage and milk yield, while the second overlaps the Prolactin Receptor (*PRLR*) and affects protein and fat yield (Blott *et al.*, 2003; Khatkar *et al.*, 2004; Schnabel *et al.*, 2005; Viitala *et al.*, 2006; Cole *et al.*, 2009; Ogorevc *et al.*, 2009; Jiang *et al.*, 2010). These regions are under selection (Flori *et al.*, 2009; Hayes *et al.*, 2009; The Bovine HapMap Consortium, 2009; Qanbari *et al.*, 2010, 2011; Stella *et al.*, 2010; Wiener *et al.*, 2011).

Some studies have shown the presence of QTLs affecting milk fat and protein traits in the region surrounding the Signal Transducer and Activator of Transcription 1 (*STAT1*) gene. This gene has been implicated in mammary gland development and is associated with milk, fat and protein yield in Holstein cattle (Cobanoglu *et al.*, 2006). Two studies comparing allele frequency differences between beef and dairy cattle suggested a selection signal in the region surrounding this gene (Hayes *et al.*, 2009; Hosokawa *et al.*, 2012).

The region surrounding the Sialic Acid Binding Ig-Like Lectin 5 (SIGLEC-5) and Zinc Finger Protein 577 (ZNF577) genes was shown to be associated with Net Merit and several related traits, such as conformation, longevity and calving ease in Holstein cattle (Cole *et al.*, 2009). Based on findings using the *iHS* approach, this region was suggested to be under selection in Holstein cattle and, although these traits were not the main objective in breeding improvement programs, a weak selection against unfavorable alleles may be responsible for this signature (Qanbari *et al.*, 2011).

Several other regions have been suggested to be under selection in cattle, but the genes under selection cannot be proposed for most of them. Functional analysis of these regions reveals the presence of genes involved in the gonadotropic and somatotropic axes, muscle development, growth, nervous system development and immune response (Barendse *et al.*, 2009; Flori *et al.*, 2009; Gautier *et al.*, 2009; The Bovine HapMap Consortium, 2009; Qanbari *et al.*, 2010, 2011; Stella *et al.*, 2010; Wiener *et al.*, 2011).

Pigs

Pig domestication occurred independently multiple times in diverse locations across Eurasia approximately 9000 years ago (Larson *et al.*, 2005). Domestic pig species are found in a wide range of environments and show extensive variation in morphological, behavioral and ecological

characteristics (Larson *et al.*, 2005; Chen *et al.*, 2007). The use of this species in very different production systems and environmental conditions around the globe has resulted in an enormous variety of breeds, each one harboring adaptations to special conditions. Currently, most pig production systems are based on five breeds (Large White, Duroc, Landrace, Hampshire and Pietrain) that have been subjected to intense artificial selection focused on productivity traits. Moreover, there is a considerable number of related species and wild individuals that can be used to infer some aspects of selection (Chen *et al.*, 2007).

The increase in muscle mass and decrease in fat content in pigs has been subject to strong selective pressure in commercial pig populations and is related to a substitution in intron 3 of the Insulin-Like Growth Factor 2 (*IGF2*) gene (Van Laere *et al.*, 2003). Using Tajima's *D*, Ojeda *et al.* (2008) identified a selection signature in the IGF2 gene in three breeds (Pietrain, Hampshire and Duroc) that are commonly used as sire lines, and have been selected for growth and meat leanness. The Melanocortin 4 Receptor (*MC4R*) gene related to growth and fatness traits has also been suggested to be under selection in pigs (Rubin *et al.*, 2012; Onteru *et al.*, 2013).

An intronic substitution in the Estrogen Receptor (ESR) gene has been associated with litter size in pigs (Rothschild *et al.*, 1996; Short *et al.*, 1997). Although some studies have reported divergent results (Muñoz *et al.*, 2007), this marker has been used by the pig breeding industry in Marker Assisted Selection (Dekkers, 2004). Recently, Bonhomne *et al.* (2010) suggested that this gene is under selection in the Large White breed.

Functional analysis of regions under positive selection in pig breeds has identified genes involved in development of the nervous system and muscle, growth, pigmentation, metabolism, visual/odor perception, immune and inflammatory responses and reproduction (Amaral *et al.*, 2011; Rubin *et al.*, 2012; Esteve-Codina *et al.*, 2013).

Sheep and goats

Sheep and goats were the first domesticated livestock species approximately 9000 years ago. The wide distribution of these species is a reflection of their adaptability to different environments and this has resulted in enormous morphological variation among populations (Diamond, 2002; Gentry *et al.*, 2004; Naderi *et al.*, 2008; Chessa *et al.*, 2009; Kijas *et al.*, 2009). Since their domestication, sheep have been selected for meat, wool and milk production (Chessa *et al.*, 2009; Kijas *et al.*, 2009).

Kijas *et al.* (2012) performed a genome scan based on *FST* to detect selection signatures in a panel of 2819 individuals from 74 sheep breeds. Thirty-one regions showed selection signals and contained genes related to coat color, bone morphology, growth and reproduction traits. This analysis revealed a strong peak of differentiation surrounding the Growth Differentiation Factor 8 (*GDF-8*) gene

when Texel individuals were compared with all other breeds (Kijas *et al.*, 2012). In addition, Clop *et al.* (2006) showed a reduction in the variability of microsatellites surrounding this gene upon comparing hyper-muscled Texels with other sheep breeds. The region surrounding *GDF-8* was associated with QTLs for carcass traits in the Texel breed (Johnson *et al.*, 2005) and a point in the 3' UTR of this gene was suggested to be the causal mutation affecting extreme muscling in Texel individuals (Clop *et al.*, 2006).

Moradi *et al.* (2012) performed a genome scan with approximately 50K SNPs to search for signatures of divergent selection in a comparison between fat and thin-tailed sheep breeds; their study identified at least three regions (OAR5, OAR7 and OARX chromosomes) that have undergone selection. Interestingly, most of the regions identified by Moradi *et al.* (2012) intersect with QTLs for carcass traits. Improvement in the sheep genome annotation will facilitate the search for and validation of candidate genes related to these traits.

Horses

Horse domestication appears to have occurred 6000 years ago and was central to the development of human history. The major attraction for domestication of this species was probably its ability to run fast for long distances, but its importance as a source of meat may also have been an important factor. The domestic horse shows marked variation in morphological traits, including shape, size, colours and gait (Bowling and Ruvinsky, 2000; Levine, 2005).

Thoroughbred horses have been selected for athletic performance traits and this has led to individuals with extreme phenotypes related to anaerobic and aerobic metabolic capabilities. A genome scan aimed at identifying putative regions under selection in this breed (based on a combination of reduced heterozygosity and increased population differentiation) revealed the presence of genes related to phosphoinositide 3-kinase (*PI3K*) and insulin-signalling pathways, oxidative stress, energy regulation, adipocyte differentiation and muscle regulation and development. These functions are directly related to the main focus of selection in these breeds, namely, racetrack performance (Gu *et al.*, 2009). Among the genes suggested to be under selection in Thoroughbred horses, the Pyruvate Dehydrogenase Kinase, isozyme 4 (*PDK4*) gene has been associated with racing performance phenotypes (Hill *et al.*, 2010).

Petersen *et al.* (2013) identified strong signal differentiation around the myostatin (*GDF-8*) gene in a comparison of the American Paint Horse and Quarter Horse with other breeds. This gene was also associated with muscle fiber type proportions in these breeds.

Another important trait for particular horse breeds is their ability to perform alternate gaits. Recently, it was shown that the gene Doublesex and Mab-3 Related Transcription Factor 3 (*DMRT3*) is involved with this trait in several breeds (Andersson *et al.*, 2012). In addition, the region encompassing this gene was suggested (based on population differentiation) to be under selection in several breeds that has been selected for alternative gaits (identified as a breed-defining characteristic) (Petersen *et al.*, 2013).

Conclusions

Domestication and artificial selection processes have definitely shaped livestock genomes. The identification of candidate regions as being under selection can help researchers understand the molecular mechanisms involved in adaptation and may also be useful in identifying regions associated with important traits that are under selection.

References

Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB and Groenen MAM (2011) Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS One 6:e14782.

Andersson L and Georges M (2004) Domestic-animal genomics: Deciphering the genetics of complex traits. Nat Rev Genet 5:202-212.

Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin C-J, Patra K, Arnason T, Wellbring L, Hjalm G, *et al.* (2012) Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature 488:642-646.

Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev 11:635-641.

Avise JC and Ayala FJ (2009) In the light of evolution III: Two centuries of Darwin. Proc Natl Acad Sci USA 106(suppl. 1):9933-9938.

Barendse W (1999) Assessing lipid metabolism. International patent application PCT/AU98/00882, International Patent Publication WO 99/23248.

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W and Thomas MB (2007) A validated whole-genome association study of efficient food conversion in cattle. Genetics 176:1893-1905.

Barendse W, Harrison BE, Bunch RJ, Thomas MB and Turner LB (2009) Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10:e178.

Barton NH (1998) The effect of hitch-hiking on neutral genealogies. Genet Res 72:123-133.

Beaumont MA (2005) Adaptation and speciation: What can *FST* tell us? Trends Ecol Evol 20:435-440.

Beaumont MA and Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969-980.

Beaumont MA and Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B 263:1619-1626.

Bellinge RHS, Liberles DA, Iaschi SPA, O'Brien PA and Tay GK (2005) Myostatin and its implications on animal breeding: A review. Anim Genet 36:1-6.

Blott S, Kim JJ, Moisio S, Schmidt-Kuntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D, *et al.* (2003) Molecular dissection of a quantitative trait locus: A phenyl-alanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163:253-266.

Boettcher PJ, Caroli A, Stella A, Chessa S, Budelli E, Canavesi F, Ghiroldi S and Pagnacco G (2004) Effects of casein haplotypes on milk production traits in Italian Holstein and Brown Swiss cattle. J Dairy Sci 87:4311-4317.

Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S and SanCristobal M (2010) Detecting selection in population trees: The Lewontin and Krakauer Test extended. Genetics 186:241-262.

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK and Cavalli-Sforza LL (1991) Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. Proc Natl Acad Sci USA 88:839-843.

Bowling AT and Ruvinsky A (2000) Genetic aspects of domestication, breeds and their origins. In: Bowling AT and Ruvinsky A (eds) The Genetics of the Horse. CABI Publishing, Wallingford, pp 25-51.

Braverman JM, Hudson RR, Kaplan NL, Langley CH and Stephan W (1995) The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140:783-796.

Casas E, White SN, Riley DG, Smith TPL, Brenneman RA, Olson TA, Johnson DD, Coleman SW, Bennett GL and Chase Jr CC (2005) Assessment of single nucleotide polymorphisms in genes residing on chromosomes 14 and 29 for association with carcass composition traits in *Bos indicus* cattle. J Anim Sci 83:13-19.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2:e64.

Charlesworth B (2007) A hitch-hiking guide to the genome: A commentary on `The hitch-hiking effect of a favourable gene' by John Maynard Smith and John Haigh. Genet Res 89:389-390.

Charlesworth B, Morgan MT and Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289-1303.

Charlesworth D, Charlesworth B and Morgan MT (1995) The pattern of neutral molecular variation under the background selection model. Genetics 141:1619-1632.

Charlesworth B, Charlesworth D and Barton NH (2003) The effects of genetic and geographic structure on neutral variation. Annu Rev Ecol Evol Syst 34:99-125.

Chen K, Baxter T, Muir WM, Groenen MA and Schook LB (2007) Genetic resources, genome mapping and evolutionary genomics of the pig (*Sus scrofa*) Int J Biol Sci 3:153-165.

Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, *et al.* (2009) Revealing the history of sheep domestication using retrovirus integrations. Science 324:532-536.

Clop A, Marcq F, Takeda H, Pirottini D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen J-M, Eychenne F, *et al.* (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet 38:813-818.

Cobanoglu O, Zaitoun I, Chang YM, Shook GE and Khatib H (2006) Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. J Dairy Sci 89:4433-4437.

Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, Lee JH, Drackley JK, Band MR, Hernandez AG, Shani M, *et al.* (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res 15:936-944.

Cole JB, VanRaden PM, O'Connell JR, Van Tassell CP, Sonstegard TS, Schnabel RD, Taylor JF and Wiggans GR (2009) Distribution and location of genetic effects for dairy traits. J Dairy Sci 92:2931-2946.

Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. J Anim Sci 82:E313-E328.

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. Nature 418:700-707.

Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, Tosser-Klop G, Wang J, Yang S, Liang J, *et al.* (2013) Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). Nat Biotech 31:135-141.

Driscoll, CA, MacDonald DW and O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. Proc Natl Acad Sci USA 106(suppl. 1):9971-9978.

Drogemuller C, Wohlke A, Momke S and Distl O (2005) Fine mapping of the polled locus to a 1-Mb region on bovine chromosome 1q12. Mamm Genome 16:613-620.

Esteve-Codina A, Paudel Y, Ferretti L, Raineri E, Megens H-I, Silió L, Rodríguez MC, Groenen MAM, Ramos-Onsins SE and Pérez-Enciso M (2013) Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. BMC Genom 14:148.

Excoffier L, Hofer T and Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity 103:285-298.

Falconer DS and Mackay TFC (1996). Introduction to Quantitative Genetics. 4th edition. Longman, Essex, 478 pp.

Fang M, Larson G, Soares Ribeiro H, Li N and Andersson L (2009) Contrasting mode of evolution at a coat color locus in wild and domestic pigs. PLoS Genet 5:e1000341.

Fay JC and Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, Foulley JL and Gautier M (2009) The genome response to artificial selection: A case study in dairy cattle. PLoS One 4:e6595.

Foll M and Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977-993.

Fontanesi L, Alessandro ED, Scotti E, Liotta L, Crovetti A, Chiofalo V and Russo V (2010) Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. Anim Genet 41:478-492.

Fu YX (1996) New statistical tests of neutrality for DNA samples from a population. Genetics 143:557-570.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Fu YX and Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693-709.

Gautier M, Flori L, Riebler A, Jaffrézic F, Laloé D, Gut I, Moazami-Goudarzi K and Foulley JL (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics 10:e550.

Gentry A, Clutton-Brock J and Groves CP (2004) The naming of wild animal species and their domestic derivatives. J Archaeol Sci 31:645-651.

Goldman N and Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725-736.

Gregory TR (2009) Artificial selection and domestication: Modern lessons from Darwin's enduring analogy. Evo Edu Outreach 2:5-27.

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, *et al.* (2002) Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res 12:222-231.

Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rotschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, *et al.* (2012) Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491:393-398.

Gu J, Orr N, Park SD, Katz LM, Sulimova G, MacHugh DE and Hill EW (2009) A genome scan for positive selection in thoroughbred horses. PLoS One 4:e5767.

Harris EE and Meyer D (2006) The molecular signature of selection underlying human adaptations. Am J Phys Anthropol 131(S43):89-130.

Hartl DL and Clark AG (2010) Princípios de Genética de Populações. 4ª edição. Artmed, Porto Alegre, 660 pp.

Hayes BJ, Lien S, Nilsen H, Olsen HG, Berg P, MacEachern S, Potter S and Meuwissen THE (2008) The origin of selection signatures on bovine chromosome 6. Anim Genet 39:105-111.

Hayes BJ, Chamberlain AJ, MacEachern S, Savin K, McPartlan H, MacLeod I, Sethuraman L and Goddard ME (2009) A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. Anim Genet 40:176-184.

Hayes B, Pryce J, Chamberlain AJ, Bowman PJ and Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet 6:e1001139.

Hill EW, Gu J, McGivney BA and MacHugh DE (2010) Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. Anim Genet 41(Suppl 2):56-63.

Hohenlohe PA, Phillips PC and Cresko WA (2010) Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. Int J Plant Sci 171:1059-1071.

Hosokawa D, Ishii A, Yamaji K, Sasazaki S, Oyama K and Mannen H (2012) Identification of divergently selected regions between Japanese Black and Holstein cattle using bovine 50k SNP array. Anim Sci J 83:7-13.

Hussin J, Nadeau P, Lefebvre JF and Labuda D (2010) Haplotype allelic classes for detecting ongoing positive selection. BMC Bioinformatics 11:e65.

Innan H and Stephan W (2003) Distinguishing the hitchhiking and background selection models. Genetics 165:2307-2312.

International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695-716.

Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y and Zhang Q (2010) Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5:e13661.

Johnson PL, McEwan JC, Dodds KG, Purchas RW and Blair HT (2005) A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. J Anim Sci 83:1988-2000.

Johnston SE, McEwan JC, Pickering NK, Kijas JW, Beraldi D, Pilkington JG, Pemberton JM and Slate J (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Mol Ecol 20:2555-2566.

Kaplan NL, Hudson RR and Langley CH (1989) The "Hitch-hiking Effect" revisited. Genetics 123:887-899.

Kelley JL, Madeoy J, Calhoun JC, Swanson W and Akey JM (2006) Genomic signatures of positive selection in humans and the limits of outlier approaches. Genome Res 16:980-989.

Khatkar MS, Thomson PC, Tammen I and Raadsma HW (2004) Quantitative trait loci mapping in dairy cattle: Review and meta-analysis. Genet Sel Evol 36:163-190.

Kijas JMH, Wales R, Törnsten A, Chardon P, Moller M and Andersson L (1998) Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. Genetics 150:1177-1185.

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, *et al.* (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One 4:e4668.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, Cristobal MS, Servin B, McCulloch R, Whan V, Gietzen K, *et al.* (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol 10:e1001258.

Kim Y and Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. Genetics 167:1513-1524.

Kim Y and Stephan W (2000) Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics 155:1415-1427.

Kim Y and Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160:765-777.

Kimura R, Fujimoto A, Tokunaga K and Ohashi J (2007) A practical genome scan for population-specific strong selective sweeps that have reached fixation. PLoS One 2:e286.

Kreitman M (2000) Methods to detect selection in populations with applications to the human. Annu Rev Genomics Hum Genet 01:539-559.

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, *et al.* (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307:1618-1621.

Levine MA (2005) Domestication and early history of the horse. In: Mills DS and McDonnell SM (eds) The Domestic Horse:

The Origins, Development and Management of its Behavior. Cambridge University Press, Cambridge, pp 5-22.

Lewontin RC and Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175-195.

Li M-H, Iso-Touro T, Laurén H and Katanen J (2010a) A microsatellite-based analysis for the detection of selection on BTA1 and BTA20 in northern Eurasian cattle (*Bos taurus*) populations. Genet Sel Evol 42:32.

Li J, Yang H, Li JR, Li HP, Ning T, Pan XR, Shi P and Zhang YP (2010b) Artificial selection of the melanocortin receptor 1 gene in Chinese domestic pigs during domestication. Heredity 105:274-281.

MacEachern S, Hayes BJ, McEwan J and Goddard M (2009) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in domestic cattle. BMC Genomics 10:e181.

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

McVean G (2007) The structure of linkage disequilibrium around a selective sweep. Genetics 175:1395-1406.

Moore SS, Li C, Basarab J, Snelling WM, Kneeland J, Murdoch B, Hansen C and Benkel B (2003) Fine mapping of quantitative trait loci and assessment of positional candidate genes for backfat on bovine chromosome 14 in a commercial line of *Bos taurus*. J Anim Sci 81:1919-1925.

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG and McEwan JC (2012) Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BMC Genetics 13:e10.

Muñoz G, Ovilo C, Estellé J, Silió L, Fernández A and Rodriguez C (2007) Association with litter size of new polymorphisms on ESR1 and ESR2 genes in a Chinese-European pig line. Genet Sel Evol 39:195-206.

Naderi S, Rezaei HR, Pompanon F, Blum MGB, Negrini R, Naghash HR, Balkiz Ö, Mashkourg M, Gaggiotti OE, Ajmone-Marsane P, *et al.* (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. Proc Natl Acad Sci USA 105:17659-17664.

Navarro A and Barton NH (2002) The effects of multilocus balancing selection on neutral variability. Genetics 161:849-863.

Nei M (2005) Selectionism and neutralism in molecular evolution. Mol Biol Evol 22:2318-2342.

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

Nei M and Maruyama T (1975) Letters to the editors: Lewontin-Krakauer test for neutral genes. Genetics 80:395.

Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. Heredity 86:641-647.

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39:197-218.

Nielsen R and Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG and Bustamante C (2005) Genomic scans for selective sweeps using SNP data. Genome Res 15:1566-1575.

Nilsen H, Olsen HG, Hayes B, Sehested E, Svendsen M, Nome T, Meuwissen T and Lien S (2009) Casein haplotypes and their association with milk production traits in Norwegian Red cattle. Genet Sel Evol 41:24.

O'Brien JD, Minin VN and Suchard MA (2009) Learning to count: Robust estimates for labeled distances between molecular sequences. Mol Biol Evol 26:801-814.

Ogorevc J, Kunej T, Razpet A and Dovc P (2009) Database of cattle candidate genes and genetic markers for milk production and mastitis. Anim Genet 40:832-851.

Ojeda A, Huang LS, Ren J, Angiolillo A, Cho IC, Soto H, Lemús-Flores C, Makuza SM, Folch JM and Pérez-Encizo M (2008) Selection in the making: A worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. Genetics 178:1639-1652.

Oleksyk TK, Smith MW and O'Brien J (2010) Genome-wide scans for footprints of natural selection. Philos Trans R Soc Lond B 365:185-205.

Olsen HG, Nilsen H, Hayes B, Berg PR, Svendsen M, Lien S and Meuwissen THE (2007) Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. BMC Genetics 8:e32.

Onteru SK, Gorbach DM, Young JM, Garrick DJ, Dekkers JCM and Rothschild MF (2013) Whole genome association studies of residual feed intake and related traits in the pig. PLoS ONE 8:e61756.

Pannier L, Mullen AM, Hamill RM, Stapleton PC and Sweeney T (2010) Association analysis of single nucleotide polymorphisms in DGAT1, TG and FABP4 genes and intramuscular fat in crossbred *Bos taurus* cattle. Meat Sci 85:515-518.

Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns MM, Borges AS, *et al.* (2013) Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS Genet 9:e1003211.

Prasad A, Schnabel RD, McKay SD, Murdoch B, Stothard P, Kolbehdari D, Wang Z, Taylor JF and Moore SS (2008) Linkage disequilibrium and selection signatures on chromosomes 19 and 29 in beef and dairy cattle. Anim Genet 39:597-605.

Pritchard JK, Pickrell JK and Coop G (2010) The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol 20:R208-R215.

Przeworski M (2002) The signature of positive selection at randomly chosen loci. Genetics 160:1179-1189.

Przeworski M, Coop G and Wall JD (2005) The signature of positive selection on standing genetic variation. Evolution 59:2312-2323.

Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR and Simianer H (2010) A genome-wide scan for signatures of recent selection in Holstein cattle. Anim Genet 41:377-389.

Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, Thaller G and Simianer H (2011) Application of site and haplotype frequency based approaches for detecting selection signatures in cattle. BMC Genomics 12:e318.

Robertson A (1975) Letters to the editors: Remarks on the Lewontin-Krakauer test. Genetics 80:396.

Rothschild M, Jacobson C, Vaske D, Tuggle C, Wang L, Short T, Eckardt G, Sasaki S, Vincent A, Mclaren D, *et al.* (1996) The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc Natl Acad Sci USA 93:201-205.

Rubin C-J, Megens H-J, Barrio AM, Maqboolc K, Sayyabc S, Schwochowc D, Wanga C, Carlborgd Ö, Jerna P, Jørgensen CB, *et al.* (2012) Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci USA 109:19529-19536.

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449:913-919.

Schlötterer C (2003) Hitchhiking mapping – functional genomics from the population genetics perspective. Trends Genet 19:32-38.

Schnabel RD, Sonstegard TS, Taylor JF and Ashwell MS (2005) Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim Genet 36:408-416.

Schwarzenbacher H, Dolezal M, Flisikowski K, Seefried F, Wurmser C, Schlötterer C and Fries R (2012) Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. BMC Genomics 13:e48.

Short TH, Rothschild MF, Southwood OI, McLaren DG, De Vries A, Van Der Steen H, Eckardt GR, Tuggle CK, Helm J, Vaske DA, *et al.* (1997) Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. J Anim Sci 75:3138-3142.

Sodeland M, Grove H, Kent M, Taylor S, Svendsen M, Hayes BJ and Lien S (2011) Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. BMC Genetics 12:e70.

Stella A, Ajmone-Marsan P, Lazzari B and Boettcher P (2010) Identification of selection signatures in cattle breeds selected for dairy production. Genetics 185:1451-1461.

Stephan W (2010) Genetic hitchhiking *vs.* background selection: The controversy and its implications. Phil Trans R Soc B 365:1245-1253.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol 14:671-668.

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

Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599.

Tang K, Thornton KR and Stoneking M (2007) A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol 5:e171.

Templeton AR (2006). Population Genetics and Microevolutionary Theory. 1st edition. John Wiley and Sons, New Jersey, 705 pp.

Thaller G, Kühn C, Winter A, Ewald G, Bellmann O, Wegner J, Zühlke H and Fries R (2003). DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle. Anim Genet 34:354-357.

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.

The International Sheep Genomics Consortium (2010) The sheep genome reference sequence: A work in progress. Anim Genet 41:449-453.

Toro M and Mäki-Tanila A (2007) Genomics reveals domestication history and facilitates breed development. In: Oldenbroek K (ed) Utilisation and Conservation of Farm Animal Genetic Resources. Wageningen Academic Publishers, Wageningen, pp 75-102.

Tsakas S and Krimbas CB (1976) Testing the heterogeneity of F values: A suggestion and a correction. Genetics 84:399-401.

Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, *et al.* (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425:832-836.

Veneroni GB, Meirelles SL, Grossi DA, Gasparin G, Ibelli AMG, Tizioto PC, Oliveira HN, Alencar MM and Regitano LCA (2010). Prospecting candidate SNPs for backfat in Canchim beef cattle. Genet Mol Res 9:1997-2003.

Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Maki-Tanila A, Georges M and Vilkki J (2006) The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. Genetics 173:2151-2164.

Vitalis R, Dawson K and Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. Genetics 158:1811-1823.

Vitalis R, Dawson K, Boursot P and Belkhir K (2003) DetSel 1.0: A computer program to detect markers responding to selection. J Hered 94:429-431.

Voight BF, Kudaravalli S, Wen X and Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:446-458.

Weikard R, Khün C, Goldhammer T, Freyer G and Schwerin M (2005) The bovine PPARGC1A gene: Molecular characterization and association of an SNP with variation of milk fat synthesis. Physiol Genomics 21:1-13.

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

Werth LA, Hawkins, GA, Eggen A, Petit E, Elduque C, Kreigesmann B and Bishop MD (1996) Melanocyte stimulating hormone receptor (MC1R) maps to bovine chromosome 18. J Anim Sci 74:262.

Wiener P and Gutiérrez-Gil B (2009) Assessment of selection mapping near the myostatin gene (GDF-8) in cattle. Anim Genet 40:598-608.

Wiener P and Wilkinson S (2011) Deciphering the genetic basis of animal domestication. Proc R Soc Lond B 278:3161-3170.

Wiener P, Burton D, Ajmone-Marsan P, Dunner S, Mommens G, Nijman IJ, Rodellar C, Valentini A and Williams JL (2003) Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus. Heredity 90:350-358.

Wiener P, Edriss MA, Williams JL, Waddington D, Law A, Woolliams JA and Gutiérrez-Gil B (2011) Information content in genome-wide scans: Concordance between patterns of genetic differentiation and linkage mapping associations. BMC Genomics 12:65.

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD and Nielsen R (2007) Localizing recent adaptive evolution in the human genome. PLoS Genet 3:e90.

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568-573.

Yang Z (2002) Inference of selection from multiple species alignments. Curr Opin Genet Dev 12:688-694.

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.

Yang Z and Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409-418.

Yang Z and Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908-917.

Yang Z, Nielsen R, Goldman N and Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.

Yang Z, Wong WSW and Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107-1118.

Associate Editor: Alexandre Rodrigues Caetano